12

Optimal Replacement and Inspection of Stochastically Failing Equipment

R. RADNER AND D. W. JORGENSON
University of California, Berkeley

1. Introduction

Much progress has been made in treating certain problems in decision-making under uncertainties that arise in cases of inventory control with stochastic demands (see [2]). Similar problems arise in the maintenance of stochastically failing equipment. In this paper we investigate the structure of optimal policies for replacement and inspection of such equipment. The models considered are of a general class that might be called "preparedness models" (after Savage [16]). A piece of equipment is kept in storage for use in case of emergency. The objective of decisions taken is to maintain the equipment in a state of operational readiness. The distinctive feature of preparedness models is that the actual state of the equipment at any time can be ascertained only by inspection. If the equipment is found to have failed, repair or replacement follows inspection.

We may distinguish two sources of uncertainty in preparedness problems: It is impossible to predict with certainty when a failure will occur, and at any given time the state of the equipment—whether good or failed—is unknown unless inspection is undertaken. These features of preparedness models are shared by a much broader class of problems; for example, in a continuous manufacturing process the quality of the output may be affected by failure of the equipment, but such failure is detected only when the output is inspected.

The times at which inspections and replacements are made comprise the set of decision variables. In this paper we determine optimal policies that characterize the times at which actions are to be taken to maintain the equipment in a state of operational readiness as much of the time as possible.

Research done at The RAND Corporation. Thanks are due K. J. Arrow for helpful comments and criticism, and the Institute of Business and Economic Research, University of California, Berkeley, for preparation of the final manuscript.

Reprinted from STUDIES IN APPLIED PROBABILITY AND MANAGEMENT SCIENCE edited by Arrow, Karlin and Scarf. Published by Stanford University Press, Stanford, California.

We consider optimal policies for both single-part equipment and equipment made up of more than one part. In the latter cases we are led to the consideration of "opportunistic" maintenance policies, in which the timing of maintenance on one part may depend upon the state of another part. We utilize known information about the structure of optimal policies to simplify computation.

1.1 Review of the Literature. We first review briefly previous work on the maintenance of stochastically failing equipment. The development of the economic theory of replacement has paralleled that of inventory theory. (For a review of the historical development of inventory theory see [2, chapter 1].) The first problems considered were purely deterministic. In such problems the purchase price of the equipment and its salvage value as a function of its age are taken as given; a stream of revenues and costs is associated with operation of the equipment without replacement. The object of the replacement policy is to maximize the average flow of profit over the life of the investment; maximum profit is obtained when the decline in average costs of procurement (purchase price less salvage value) is balanced by the decline in operating profits as the age of the machine increases. Where the time span is unlimited, the problem may be simplified considerably, since replacement serves as a regeneration point of the investment process, so that maximization of average profit over a single cycle of operations from one replacement to the next is sufficient for maximization of profit over the entire investment. The optimal policy for single-part equipment is described by one parameter, the time to replacement. Policies that can be described by such a parameter may be called *periodic*. A full summary of available results for the deterministic case is given by Alchian [1].

The second stage in the development of replacement theory was consideration of a model in which equipment is subject to random failures, but the state of the equipment is always known with certainty. The operating characteristics of a policy of replacing (repairing) the equipment when it has failed were first described by Lotka [11]. A policy of preventive maintenance for the same model was considered by Campbell [4]. (See also [18].) The policy is to replace the equipment after a sufficiently long time has passed since the last replacement, whether or not the equipment has failed. Such a policy is of interest if the cost of replacement after failure is greater than the cost of replacement before failure. A plausible example is the replacement of all lamps in a street lighting system or in an office building. The cost per lamp of replacing all lamps at once is much less than the cost of replacing each lamp as it fails. The cost of the additional lamps required for preventive maintenance must be balanced against the cost of the additional failures that occur if replacement is postponed. This problem has been rediscovered many times [13, 17, 19]. Detailed results are found in [3, 5, 8, and 17]; a summary of the literature may be found in [10].

A variant of this problem is the case of an equipment composed of two identical modules. If one module fails, it must be replaced. The cost of replacing the second module at the same time as the first is considerably less than that of replacing the second module when it fails. Sasieni [15] has proposed a class of policies which we investigate below under the name of (n, N) policies. An (n, N) policy is described by two parameters; the first is the time after which both modules are replaced if one fails; the second is the time after which both modules are replaced if neither has failed. Dreyfus has devised a numerical example for which this policy is not optimal [6]. The original model considered by Lotka has been extended in another direction by Palm [14]. Where there are several machines subject to random failure and the rate at which machines may be repaired is fixed, queues of machines waiting for repair may be formed. The problem is to choose the number of repairmen so as to minimize the combined costs of lost machine time and of repair facilities. (For a summary of this work, see [13, chapter 11].)

The stochastic models of replacement (repair) discussed above are not preparedness models, since—although it is assumed that the time to failure of each unit is a random variable—the actual state of the equipment at any time is known with certainty. The literature on preparedness models is relatively recent and has developed largely under the impetus of military applications. Models that combine some aspects of the preparedness problem with features of preventive maintenance problems have been discussed by Savage [16] and Drenick [5]. In Savage's model, the cost of inspection and repair (these actions are always taken together) is an increasing function of the age of the equipment and depends on the state of the world, that is, whether or not there is an emergency. The time to the next emergency is a random variable; if an emergency occurs, inspection and repair must be undertaken at once, at a cost that is higher than the cost of the same inspection and repair under non-emergency conditions. Repair has the effect of renewing the process, so that preventive maintenance may be undertaken to reduce the costs of inspection and repair when the emergency occurs. The analogy to preventive maintenance is provided by calling failure in operation an emergency. A second preparedness model, considered by Drenick, involves inspection of a part during preventive maintenance; the part may be replaced even if it is good, provided that the cost of replacement is less than the likelihood of failure over the next interval between routine maintenance actions multiplied by the cost of repairing the part after failure in operation. Inspection and replacement models that are closely related to preparedness models have been considered in the theory of surveillance sampling [17]. A lot is held in stock; the lot deteriorates with time, and the problem is to inspect the lot and to replace it by a new lot when sufficient deterioration has occurred. Results for this problem are confined to descriptions of the consequences of various sampling and replacement plans. Finally, the problem of the optimal inspection interval for a single part characterized by exponential times to failure for essentially the same model as that considered in section 2, below, has been discussed (see Kamins [9] and the references listed there). A method for describing the consequences of alternative schemes for inspection and replacement of parts in a complex equipment has been devised by McGlothlin and Bean [12].

1.2 Structure of Replacement and Inspection Problems. To characterize the model of replacement and inspection used in this paper, we may begin by discussing the distribution of times to failure for the equipment. We assume first that the equipment is always in one of two possible states—good or failed. The probability $R(t)$ that the equipment is good at time t is called the *reliability function* for the equipment. We assume that the reliability function $R(t)$ has the properties $R(0) = 1$, $R(\infty) = 0$, $R(t)$ is twice differentiable, and $R'(t) < 0$. An important example that clearly satisfies these conditions is the exponential reliability function

$$R(t) = \exp(-\lambda t) .$$

For any reliability function, the failure rate is defined as $\rho(t) = -R'(t)/R(t)$. The failure rate for the exponential reliability function is the constant λ. For complex equipment with more than one stochastically failing module, we assume that the equipment is good only if all components are good. The reliability function of the system $R(t)$ may be computed from the reliability functions of the parts. Where $R_i(t)$ is the reliability function of the ith part, we have

$$R(t) = \prod_i R_i(t) .$$

Time may be considered to be discrete or continuous. For single-part equipment three alternative actions are possible at each point of time: do nothing, inspect, or replace. Associated with each action is a loss, time down, i.e., the time actually spent in inspection or replacement (repair). We assume that inspection time H and replacement time K are fixed constants. For the case of equipment with two parts, considered in sections 3 and 4 below, time for replacement of the ith part is K_i, and K_3 denotes the time required to replace parts 1 and 2 simultaneously. We assume that $K_1, K_2 \leqq K_3 \leqq K_1 + K_2$, so that the time required to replace both parts simultaneously is greater than or equal to the time required to replace either, and less than or equal to the time required to replace both separately. We further assume that time down is the only loss associated with inspection and replacement actions.

Following the usual practice in replacement and maintenance problems, we assume that any machine (or part) that is replaced (repaired) is as good as new at the end of the replacement action. The point at which replacement ends serves as a point of regeneration of the process. The effect of inspection is not so straightforward. Here we assume that if a part is inspected and found to be good, the distribution of times to failure is the conditional distribution beginning at the start of inspection, given that the equipment has not failed at that point. (This amounts to assuming that the part does not age during inspection.) In the case of an exponential reliability function, the conditional distribution is also exponential with the same rate of failure, so that inspection, like replacement, serves as a point of regeneration of the process. This property of the exponential distribution, referred to as lack of memory or the Markovian property, is not shared by any other reliability function.

Since the distinctive feature of preparedness models is the information structure associated with inspection of randomly failing equipment, it is important to describe this structure in detail. We assume first that the state of an equipment is determined with certainty by inspection. If the part is good at the beginning of inspection, there are two possibilities: the equipment survives the stress of inspection and is declared good; the equipment fails under inspection stress and is declared failed. The likelihood of failure during inspection is assumed constant and equal to $1 - \sigma$; hence the probability that the equipment survives inspection if it is initially good is σ. If the equipment has failed at the beginning of inspection, it is declared failed. In any case, once the equipment is declared failed, replacement (or repair) is undertaken. Postponing replacement results in time down in addition to that required for replacement, since once the equipment has failed it remains failed until replaced or repaired. It would not be difficult to take account of possible errors in inspection, but we ignore such errors in the following discussion.

Two possible modes of inspection may be distinguished. Equipment may be inspected at distinct points of time, in which case the state of the equipment is known with certainty only at the times of inspection; or equipment may be inspected continuously, or *monitored*. In this latter case, the state of the equipment is known with certainty at all times. For single-part equipment, actual operation of the equipment may constitute monitoring, in which case the theory of replacement is equivalent to the theory of preventive maintenance described above. For complex equipment, some parts may be monitored so that their state is always known with certainty, while others are inspected only at distinct points of time. Finally, certain parts may be inaccessible to inspection unless removed from the equipment.

The object of replacement and inspection decisions is to maximize the time that the equipment is actually in operational readiness, that is, not failed. This criterion may be made precise in at least two alternative ways: (1) Maximize the long-run average proportion of the time that the equipment is good; (2) maximize the discounted total time good, where time good is discounted at some rate α. We denote the proportion of time good by G and the discounted total time good by g; both G and g are functions of the sequence of actions taken over the lifetime of the equipment. An immediate consequence of either criterion is that if equipment is declared failed in inspection, replacement actions are undertaken immediately.

1.3 Survey of Results. The first class of problems we consider involves a single piece of stochastically failing equipment. Since replacement results in regeneration of the equipment, any optimal replacement policy (without inspection) is strictly periodic. The optimal period between replacements is easily determined; we shall show that for an optimal replacement interval the reliability of the equipment when replacement is started is equal to the average proportion of time good during the cycle. A similar relationship can be found for inspection in the case of an exponential distribution of times to failure. The optimal policy is periodic in that there is a fixed average time between inspections that is equal to the time to inspection plus inspection time plus replacement time multiplied by the probability of not passing inspection. For the case of an exponential reliability function an optimal policy consists of either a series of inspections, with replacement after inspection if the equipment has failed, or a series of replacements without previous inspection. This characterization of optimal policies follows from the fact that the equipment is "as good as new" after each action, whether the action is replacement or inspection. If inspection is chosen before replacement without inspection, it will be chosen again after the same amount of time has elapsed; the same is obviously true in the reverse case.

The second class of problems to be discussed involves a piece of equipment with two modules, one of which is monitored. In the first case we assume that the part that is not monitored is inaccessible to inspection; i.e., that the only possible action that may be chosen as an alternative to doing nothing is replacement of the part. If the time required for replacement of both parts simultaneously is less than the sum of the times required for replacement of each part separately, it may be possible to reduce time down for replacement by pursuing a policy of opportunistic replacement.[1] The central feature of such opportunistic policies is that the action to be taken with respect to the nonmonitored part is conditional on the state of the monitored part. It will be shown that provided the distribution of times to failure of the monitored part is exponential, the optimal policy for replacement of the nonmonitored part has a structure denoted (n, N). That is, if the monitored part has not failed, the nonmonitored part is replaced if a certain time N has elapsed; if the monitored part has failed, the nonmonitored part is replaced if a certain time n has elapsed. Furthermore, $0 \leqq n \leqq N$, where the second inequality reduces to an equality if and only if the time required for replacement of both parts simultaneously is equal to the sum of the times required for replacement of each part separately. This result holds for any reliability function for the nonmonitored part.

In the second case to be discussed, it is assumed that the nonmonitored part is never replaced without prior inspection. The only action that may be chosen as an alternative to doing nothing is to inspect the part; if the part has failed, it is replaced immediately. As in the case of a nonmonitored part that is inaccessible to inspection, it may be possible to reduce time down for inspection and replacement by pursuing a policy of opportunistic replacement. Provided that the distribution of times to failure of both the monitored and nonmonitored parts is exponential, the optimal policy has an (n, N) structure.

Apparently, policies of the (n, N) type play a role in the theory of replace-

[1] This term was suggested by A. Sweetland.

ment similar to the role played by (s, S) policies in inventory theory. The difference between times n and N increases with an increase in economies of scale in replacement activities, where economies of scale are experienced if the time K_3 required for replacement when both parts are replaced simultaneously is less than the sum $(K_1 + K_2)$ of the times required for replacement of each part individually. The difference $(K_1 + K_2 - K_3)$ plays the same role as that played by a set-up cost or order cost in inventory theory. As the difference $(K_1 + K_2 - K_3)$ tends to zero, the difference $(N - n)$ between the maintenance times also tends to zero.

Although the optimality of (n, N) policies is demonstrated here only under rather special conditions, such policies may provide a useful guide to action in a wider variety of circumstances. The structure of (n, N) policies is simple, and the optimal values of their parameters are relatively easy to compute. For these reasons, (n, N) policies should be easy to administer. Of course these remarks apply as well to the special case of periodic policies.

2. Replacement and Inspection of Single-Part Equipment

2.1 Replacement for an Arbitrary Distribution of Times to Failure. The first problem to be discussed is optimal replacement policy for single-part equipment with an arbitrary distribution of times to failure. Since replacement (or repair) serves as a regeneration point of the process, the optimal policy is strictly periodic, and it suffices to maximize time good over a single cycle. If N is the length of time between the end of one repair and the beginning of another, average reliability over the cycle is given by

$$G(N) = \frac{\int_0^N R(t)\,dt}{N + K}\,, \tag{2.1}$$

where K is time spent in repair or time required for replacement. In the following discussion it is assumed that K is strictly positive.

To maximize the average proportion of time good during the cycle, we differentiate $G(N)$ twice:

$$G'(N) = \frac{R(N)}{N + K} - \frac{G(N)}{N + K}\,, \tag{2.2}$$

$$G''(N) = \frac{R'(N)}{N + K} - 2\frac{G'(N)}{N + K}\,. \tag{2.3}$$

The first-order condition for a maximum, $G'(N) = 0$, is therefore

$$R(N) = G(N) \tag{2.4}$$

or

$$(N + K)R(N) - \int_0^N R(t)\,dt = 0\,. \tag{2.5}$$

We demonstrate the existence of a solution of (2.5) by observing that the left-hand side of (2.5) is positive for $N = 0$ (since $K > 0$) and is negative for large N (since $R(N)$ decreases monotonically to zero and $NR(N) - \int_0^N R(t)\,dt$ is a decreasing function of N that equals zero at $N = 0$). At any solution $\hat{N}$ of (2.5) we have from (2.3)

$$G''(\hat{N}) = \frac{R'(\hat{N})}{N + K} < 0 ,$$

so that any such solution gives a maximum; moreover, it follows that there is only one solution.

Note that by (2.4), at the best replacement time the reliability of the equipment equals the average time good during the cycle. As an illustration we compute optimal time to replacement for the exponential distribution of times to failure. In the exponential case $R(t) = \exp(-\lambda t)$, but at the maximum $G(N) = R(N)$; hence

$$\exp(-\lambda N) = \frac{\int_0^N \exp(-\lambda t)\,dt}{N + K} = \frac{(1/\lambda)[1 - \exp(-\lambda N)]}{N + K} ,$$

which is equivalent to

$$\text{(2.6)} \qquad \frac{\exp(\lambda N)}{\lambda} - N = \frac{1}{\lambda} + K ,$$

a transcendental equation to be solved for N. The term on the right-hand side may be interpreted as the mean time to failure plus the time required for replacement. A quadratic approximation to $\exp(\lambda N)$ gives the formula $N = \sqrt{2K/\lambda}$, which may be useful as an initial value for iterative computation of a solution to the transcendental equation.

In the general case, the transcendental equation has the form

$$R(N) = G(N) = \frac{\int_0^N R(t)\,dt}{N + K} .$$

Using this implicit form, we examine the effects of changes in the parameters on the optimal value $\hat{N}$:

$$\text{(2.7)} \qquad \frac{\partial \hat{N}}{\partial K} = \frac{-R(\hat{N})}{(\hat{N} + K)R'(\hat{N})} = \frac{1}{(\hat{N} + K)\rho(\hat{N})} > 0 ,$$

where $\rho(N)$ is the instantaneous failure rate at time N. Thus the optimal time to replacement increases with an increase in the time required for replacement.

For the exponential case, the distribution of times to failure involves a single parameter λ, the failure rate ($1/\lambda$ is the mean time to failure). In this case (2.7) reduces to $\partial \hat{N}/\partial K = 1/(\hat{N} + K)\lambda$. In addition, we can verify from (2.6) that

$$\frac{N}{\lambda} = \frac{K - N[\exp(\lambda N) - 1]}{\lambda[\exp(\lambda N - 1)]} < 0 ,$$

since $G(N) = R(N)$ implies that

$$N + K = \exp(\lambda N)\int_0^N \exp(-\lambda t)\,dt < \exp(\lambda N)N\,;$$

so that $K - N[\exp(\lambda N) - 1] < 0$. Thus the optimal time to replacement increases with any increase in the mean time to failure (decrease in the rate of failure), in the case of an exponential reliability function.

2.2 Inspection for the Exponential Case. If the distribution of times to failure is exponential, inspection of equipment has the same renewal property as replacement, and essentially the same technique may be employed to obtain an optimal interval of inspection. Any optimal inspection policy will be periodic in the sense that the time between the end of one inspection (and repair of failure that has occurred) and the beginning of the next is fixed. For inspection, the proportion of time good to total time in a single cycle is given by

$$G(N) = \frac{\int_0^N \exp(-\lambda t)\,dt}{N + H + [1 - \sigma\exp(-\lambda N)]K}\,, \tag{2.8}$$

where N, λ, and K are defined as before, H is the time required for inspection, and σ is the probability that the equipment will not fail if it is good at the beginning of inspection.

To maximize $G(N)$ with respect to N, we differentiate twice as follows:

$$G'(N) = \frac{\exp(-\lambda N)[N + H + (1-\sigma)K] - (1/\lambda)[1 - \exp(-\lambda N)]}{\{N + H + [1 - \sigma\exp(-\lambda N)]K\}^2}\,,$$

$$G''(N) = \frac{-\lambda\exp(-\lambda N)[N + H + (1-\sigma)K]}{\{N + H + [1 - \sigma\exp(-\lambda N)]K\}^2} - G'(N)\frac{1 + \sigma\lambda\exp(-\lambda N)K}{N + H + [1 - \sigma\exp(-\lambda N)]K}\,.$$

Any solution to the first equation satisfies the transcendental equation

$$\frac{\exp(\lambda N)}{\lambda} - N = \frac{1}{\lambda} + H + (1-\sigma)K\,, \tag{2.9}$$

which has a solution for any $H + (1-\sigma)K > 0$. Uniqueness of the solution to this equation is assured by the fact that $G'(N) = 0$ implies that

$$G''(N) = -\frac{\lambda\exp(-\lambda N)[N + H + (1-\sigma)K]}{\{N + H + [1 - \sigma\exp(-\lambda N)]K\}^2} < 0\,.$$

A quadratic approximation to $\exp(\lambda N)$ gives the formula

$$N = \sqrt{\frac{2[H + (1-\sigma)K]}{\lambda}}\,.$$

It is interesting that (2.9) can be expressed equivalently as

$$G(N) = \frac{R(N)}{1 - \sigma K R'(N)}\,, \tag{2.10}$$

which should be compared with (2.4). Using the implicit form of the transcendental equation, we observe that

$$\frac{\partial N}{\partial H} = -\frac{\exp(-\lambda N)}{G''(N)\{N + H + [1 - \sigma \exp(-\lambda N)]K\}^2} > 0$$

and

$$\frac{\partial N}{\partial K} = -\frac{\exp(-\lambda N)(1-\sigma)}{G''(N)\{N + H + [1 - \sigma \exp(-\lambda N)]K\}^2} > 0 ,$$

so that the optimal inspection interval N increases with any increase in the time required for inspection H, or with the time required for replacement K. Second, changes in N with respect to changes in the likelihood σ of surviving inspection given that the equipment is good before checkout may be calculated as follows:

$$\frac{\partial N}{\partial \sigma} = \frac{\exp(-\lambda N_K)}{G''(N)\{N + H + [1 - \sigma \exp(-\lambda N)]K\}^2} < 0 .$$

Finally, the change in N with respect to the failure rate λ is

$$\frac{\partial N}{\partial \lambda} = \frac{(N/\lambda) - (1/\lambda^2)[1 - \exp(-\lambda N)]}{G''(N)\{N + H + [1 - \sigma \exp(-\lambda N)]K\}^2} < 0 ,$$

since

$$N > \int_0^N \exp(-\lambda t)\, dt = \frac{1}{\lambda}[1 - \exp(-\lambda N)] ,$$

which implies that

$$\frac{N}{\lambda} - \frac{1}{\lambda^2}[1 - \exp(-\lambda N)] > 0 .$$

The optimal time to inspection decreases with any increase in the likelihood that the equipment survives inspection, given that it has not failed when inspection begins. Moreover, the optimal time to inspection decreases with any increase in the failure rate. Equivalently, the optimal time to inspection increases with any increase in the mean time to failure.

2.3 Choice Between Replacement and Inspection. To analyze the choice between replacement and inspection, let us denote the average time good for the system in which replacement is undertaken at time N by $G_R(N)$, denote the average reliability when inspection is undertaken at time N by $G_I(N)$, and let N_R and N_I denote the corresponding best times. From (2.4) and (2.10) we have

$$G_R(N_R) = R(N_R) ,$$

$$G_I(N_I) = \frac{R(N_I)}{1 - \sigma K R'(N_I)} .$$

Replacement is better than inspection if $G_R(N_R) > G_I(N_I)$, i.e., if and only if

$$R(N_R) > \frac{R(N_I)}{1 - \sigma K R'(N_I)} ,$$

which, after some algebraic manipulation, reduces to

$$\exp(\lambda N_I) + \lambda\sigma K > \exp(\lambda N)R . \tag{2.11}$$

In particular, (2.11) is satisfied if $N_I > N_R$. By comparing (2.6) with (2.9), we can easily show that $N_I > N_R$ if and only if

$$H > \sigma K . \tag{2.12}$$

Thus this last condition is a sufficient, though not necessary, condition for replacement to be better than inspection.

3. Replacement of a Single Part in the Presence of a Monitored Part

3.1 Opportunistic Replacement. For systems with more than one part the possibility arises of allowing maintenance decisions with respect to one part to depend upon information about the states of other parts. In particular, it may be possible to take advantage of "economies of scale" in replacement times. In this section we investigate the simple case of a system with two parts, where one part is continuously monitored and the other cannot be inspected but can be replaced at any time. We show that if the monitored part has an exponential failure-time distribution, then the optimal replacement policy for the nonmonitored part is of the (n, N) type. We also indicate a method for computing the best values of n and N and the corresponding best average time good. In the next section we obtain similar results for the case in which the nonmonitored part is inspected prior to replacement.

3.2 The Model. Consider a two-part system such that at any time each part can be in one of two states, *good* or *failed*, and the system is good if and only if both parts are good. In this section we consider the case in which the first part is never inspected, and its state is not known between replacements. The second part, however, is subject to continuous inspection, so that failure is evident as soon as it occurs.

Thus the *information* about the system at any time when a replacement is not being made can be described by two variables: x, the time since the first part was last replaced, and y, which is 0 or 1 according as the second (monitored) part is or is not in a failed state. Also, whenever no replacement is being made four alternative *actions* are possible:

(0) do nothing,
(1) begin the replacement of part 1,
(2) begin the replacement of part 2,
(3) begin the replacement of both parts.

Each replacement action takes a certain amount of time, say K_i for action (i). We assume that $0 < K_1, K_2 \leqq K_3 \leqq K_1 + K_2$. There are economies of scale in replacement time if $K_3 < K_1 + K_2$.

Let $G(t)$ describe the actual state of the system at time t, so that $G(t)$ is 1 or 0 according as the system is good or failed at time t. During the time that a replacement is being made, $G(t)$ is, of course, equal to 0. On the other hand, even if at a given time between replacements the monitored item is known to be good, the value of $G(t)$ is uncertain; its expected value is the probability that part 1 is still good, which is equal to $R_1(x)$, where R_1 is the reliability function of part 1, and x is the time elapsed since part 1 was last replaced. In our discussion of replacement, we shall make no further assumption about the reliability function of part 1, except that it is absolutely continuous and decreasing from 1 to 0 as x increases from 0 to infinity. We make a more special assumption about the second (monitored) part, namely that its time-to-failure distribution is exponential. The important consequence of this assumption is that any point of time at which part 1 has just been replaced and part 2 is known to be good is a point of *renewal* for the process being studied.

It is obvious from these general considerations that part 2 should be replaced (either with or without replacement of part 1) if and only if it has just failed; i.e., (a) if $y = 1$, take action 0 or action 1; (b) if $y = 0$, take action 2 or action 3. In other words, the only decision remaining at any time is whether or not to replace part 1, and this decision will depend upon the age x of part 1 and whether or not part 2 has just failed.

3.3 The Functional Equation for the Optimal Policy. In this paper, the criterion for choosing among policies is the long-run average value of $G(t)$, or "average time good." It is mathematically convenient to treat this criterion as the limiting case of the criterion of the discounted integral of $G(t)$ from 0 to infinity, with, roughly speaking, a zero discount rate. More precisely, we use the fact that

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T G(t)\,dt = \lim_{\alpha\to 0} \alpha\int_0^\infty \exp(-\alpha t)\, G(t)\, dt\,, \tag{3.1}$$

where $\alpha > 0$ and the indicated limits exist.

Let $g(x, y)$ denote the maximum expected integral of the discounted value of $G(t)$ that is possible starting from a state of information (x, y). Following the usual technique of dynamic programming, we shall explore the properties of the optimal policy by examining a functional equation that characterizes the function g.

Thus, suppose that at time 0 part 1 already has age x, and that part 2 has not failed. Define the action $(0 \mid \Delta)$ as follows:

Do nothing until either time Δ or the first time that part 2 fails; if time Δ is reached first, take action (1); if part 2 fails before time Δ, take the better of actions (2) and (3). Let u denote the minimum of Δ and the time at which part 2 fails. Over the interval from 0 to u an expected amount of discounted time good accumulates:

$$\int_0^u \exp(-\alpha t)\, R_1(x+t)\, dt\,. \tag{3.2}$$

Furthermore, if $u < \Delta$, we have a state of information $(x + u, 0)$, whereas if $u = \Delta$, we have $(x + \Delta, 1)$. For $u < \Delta$, u has an exponential probability density with rate λ_2; the probability that $u = \Delta$ is $\exp(-\lambda_2\Delta)$. Hence the expected discounted time good that results from performing action $(0 \mid \Delta)$ and following an optimal policy thereafter is

$$g_0(x \mid \Delta) \equiv \int_0^{\Delta}\left[\int_0^u \exp(-\alpha t)\, R_1(x+t)\, dt + \exp(-\alpha u)\, g(x+u, 0)\right] \cdot \lambda_2 \exp(-\lambda_2 u)\, du + \left[\int_0^{\Delta} \exp(-\alpha t)\, R_1(x+t)\, dt + \exp(-\alpha\Delta)\, g_1\right] \exp(-\lambda_2\Delta)\,, \tag{3.3}$$

where

$$g_1 \equiv \exp(-\alpha K_1)\, g(0, 1)\,. \tag{3.4}$$

Note that g_1 is the expected discounted time good resulting from performing action (1) first and following an optimal policy thereafter.

We assume that neither part ages while the other is being replaced, i.e., that the system is "turned off" during maintenance. (Even if this assumption is not strictly true, the result will usually be a good approximation.)

Expression (3.3) is extended to the limiting cases of $\Delta = 0$ and $\Delta = \infty$ in the obvious way:

$$g_0(x \mid \infty) \equiv \int_0^{\infty}\left[\int_0^u \exp(-\alpha t)\, R_1(x+t)\, dt + \exp(-\alpha u)\, g(x+u, 0)\right]\lambda_2 \exp(-\lambda_2 u)\, du\,, \tag{3.3a}$$

$$g_0(x \mid 0) \equiv g_1\,.$$

Following the usual technique of dynamic programming, we can partially characterize the function g:

$$g(x, 1) = \max\,[\sup_{\Delta \geq 0} g_0(x \mid \Delta), g_1]\,. \tag{3.5}$$

Turning now to the case $\dot{y} = 0$, we can consider the consequences of performing either action (2) or action (3); performance of either yields

$$g(x, 0) = \max\,[g_2(x), g_3]\,, \tag{3.6}$$

where

$$g_2(x) = \exp(-\alpha K_2)\, g(x, 1) \tag{3.7}$$

and

$$g_3(x) = \exp(-\alpha K_3)\, g(0, 1)\,. \tag{3.8}$$

Equations (3.3)-(3.8) together implicitly characterize the function g. Furthermore, once g is known, they immediately yield the optimal replacement policy, since they indicate which action gives the maximum value. Indeed, without actually solving these equations, we shall use them to show that an (n, N) policy is optimal.

3.4 Structure of the Optimal Policy. We shall now prove that there are numbers n and N such that

(a) $0 \leqq n \leqq N$ and $N > 0$;

(b) if $y = 1$, then action $(0 \mid N - x)$ is best for $0 \leqq x < N$, and action (1) is best for $x = N$;

(c) if $y = 0$, then action (2) is best for $0 \leqq x < n$, and action (3) is best for $n \leqq x \leqq N$.

First note that unless one of the parts is completely unreliable, $g(0, 1) > 0$, since $g_0(0 \mid \infty) > 0$. Note also that $g_0(x \mid \Delta)$ is continuous in Δ for each x, since R_1 is continuous, and that $g(x, y)$ is bounded by $\int_0^\infty \exp(-\alpha t)\, dt = 1/\alpha$.

Next we show that

$$g(0, 1) = g_0(0 \mid N) \tag{3.9}$$

for some N with $0 < N < \infty$. On the one hand, if $g(0, 1) = g_1$, then by (3.4) we have $g(0, 1) = \exp(-\alpha K_1) g(0, 1)$, which cannot be, since $g(0, 1) > 0$ and $K_1 > 0$. Moreover, it follows from (3.3a) that $N > 0$.

On the other hand, suppose that $N = \infty$; then by the "optimality principle" of dynamic programming, $g(x, 1) = g_0(x \mid \infty)$ for all x. In this case part 1 would be replaced only at the same time that part 2 is replaced. Starting with an age x of part 1, and $y = 1$, let T be the interval of time until the first replacement of parts 1 and 2 together. T would be a nonnegative random variable, possibly infinite, and $g(x, 1)$ would not exceed the expected value of

$$\int_0^T R_1(x + t) \exp(-\alpha t)\, dt + \exp(-\alpha T) g_3 ,$$

which, since R_1 is nonnegative and decreasing, would not exceed the expected value of

$$R_1(x) \int_0^T \exp(-\alpha t)\, dt + \exp(-\alpha T) g_3 .$$

Since T would be at least as large as an exponentially distributed variable with mean $1/\lambda_2$, this last expected value would not exceed

$$\frac{R_1(x)}{\alpha + \lambda_2} + \left(\frac{\lambda_2}{\alpha + \lambda_2}\right) g_3 ; \tag{3.10}$$

but (3.10) can be made less than g_3 by taking x sufficiently large, which contradicts the supposition that $N = \infty$. Finally, since

$$\lim_{\Delta \to \infty} g_0(x \mid \Delta) = g_0(x \mid \infty) ,$$

the value of $g(0, 1) = \sup_{\Delta \geqq 0} g_0(x \mid \Delta)$ must be attained at some positive finite Δ.

Taking N as in (3.9), we have by the optimality principle

$$g(x, 1) = g_0(x \mid N - x) \qquad (0 \leqq x < N) \tag{3.11}$$

and

(3.12) $$g(N, 1) = g_1 .$$

Since replacement of part 1 when part 2 is good renews the process, we shall never observe a value of x larger than N, provided of course that we start the process with $x \leqq N$. Thus, having specified N, we need only specify at what set of values of x, with $y = 0$, we take action (2), and at what set we take action (3).

To that end we first explore more fully the properties of $g(x, 1)$ for $0 \leqq x \leqq N$.

Define $\bar{R}_1$ by

(3.13) $$\bar{R}_1(x, u) = \int_0^u \exp(-\alpha t) R_1(x + t)\, dt .$$

By (3.3), (3.11), and (3.12), for $0 \leqq x \leqq N$ we have

(3.14) $$g(x, 1) = \int_0^{N-x} [\bar{R}_1(x, u) + \exp(-\alpha u) g(x + u, 0)]\lambda_2 \exp(-\lambda_2 u)\, du + \{\bar{R}_1(x, N - x) + \exp[-\alpha(N - x)] g_1\} \exp[-\lambda_2(N - x)] .$$

This expression shows that $g(x, 1)$ is continuous in x, since $g(x, 0)$ is bounded.

Turning now to the consideration of an optimal policy for $y = 0$, we first show that

(3.15) $$g(N, 0) = g_3 .$$

This follows from

$$g_2(N) = \exp(-\alpha K_2) g(N, 1) = \exp(-\alpha K_2) g_1$$
$$= \exp[-\alpha(K_1 + K_2)] g(0, 1) \leqq \exp(-\alpha K_3) g(0, 1) = g_3 .$$

We may have from (3.15) either $g_2(N) < g_3$ or $g_2(N) = g_3$. Consider first the former case. Let

$$n = \inf\{z \mid g_2(x) < g_3 \quad \text{for all} \quad z < x \leqq N\} .$$

From the continuity of $g(x, 1)$ and therefore of $g_2(x)$, it follows that $n < N$ and $g_2(n) = g_3$.

We shall now show that $g_2(x) \geqq g_3$ for $0 \leqq x \leqq n$ by showing that there is a function f such that

(3.16a) $$f(x) = g(x, 1) \qquad (n \leqq x \leqq N) ,$$

(3.16b) $$f(x) \leqq g(x, 1) \qquad (0 \leqq x \leqq n) ,$$

and

(3.16c) $f(x)$ is nonincreasing for $0 \leqq x \leqq n$.

By the optimality principle, for $0 \leqq x \leqq x + h \leqq N$ we have

(3.17) $$g(x, 1) = \int_0^h [\bar{R}_1(x, u) + \exp(-\alpha u) g(x + u, 0)]\lambda_2 \exp(-\lambda_2 u)\, du + [\bar{R}_1(x, h) + \exp(-\alpha h) g(x + h, 1)] \exp(-\lambda_2 h) .$$

Define f by

$$f(x) = \int_0^{N-x} [\bar{R}_1(x, u) + \exp(-\alpha u) g_3] \lambda_2 \exp(-\lambda_2 u)\, du$$
$$+ \{\bar{R}_1(x, N - x) + \exp[-\alpha(N - x)] g_1\} \exp(-\lambda_2)(N - x) .$$

It is easily verified that f satisfies (3.16a) and (3.16b); and, moreover, that for $0 \leqq x \leqq N$, $n \leqq x + h \leqq N$, and $h \geqq 0$ we have

$$f(x) = \int_0^h [\bar{R}_1(x, u) + \exp(-\alpha u) g_3] \lambda_2 \exp(-\lambda_2 u)\, du$$
$$+ [\bar{R}_1(x, h) + \exp(-\alpha h) f(x + h)] \exp(-\lambda_2 h) , \tag{3.18}$$

by making use of the identity

$$\bar{R}_1(x, u) \equiv \bar{R}_1(x, v) + \exp(-\alpha v) \bar{R}_1(x + v, u - v) .$$

From (3.14) the derivative of $g(x, 1)$ with respect to x is continuous for $0 \leqq x \leqq N$, and hence the same is true of $f(x)$. Differentiating both sides of (3.18) with respect to x gives

$$f'(x) = \int_0^h \frac{\partial \bar{R}_1(x, u)}{\partial x} \lambda_2 \exp(-\lambda_2 u)\, du$$
$$+ \left[\frac{\partial \bar{R}_1(x, h)}{\partial x} + \exp(-\alpha h) f'(x + h)\right] \exp(-\lambda_2 h) . \tag{3.19}$$

But

$$\frac{\partial \bar{R}_1(x, u)}{\partial x} = \int_0^u \exp(-\alpha t) R_1'(x + t)\, dt \leqq 0 ,$$

since R_1 is nonincreasing. Also, $f'(x_0) < 0$ for some x_0 such that $n \leqq x_0 \leqq N$, because $f(N) = g(N, 1) < g(n, 1) = f(n)$. Hence, applying (3.19) with $h = x_0 - x$, we have $f'(x) \leqq 0$ for all $x \leqq x_0$, which shows that f satisfies condition (3.16c). Note that we get a strict inequality if R_1 is strictly decreasing.

The case $g_2(N) = g_3$ arises if $K_3 = K_1 + K_2$, as is seen from the proof of (3.15). The optimality of the (n, N) policy for this case, therefore, follows from considering it to be a limiting case of $K_3 < K_1 + K_2$.

3.5 Long-Run Average Time Good Under an (n, N) Replacement Policy. Once we have demonstrated the optimality of an (n, N) replacement policy, it remains to determine the best values of n and N and the corresponding average time good. It appears convenient to abandon the functional equation and proceed directly.

Under an (n, N) policy there is a natural cycle from the end of one replacement action to the end of the next. The long-run average time good per unit time is

$$G = \frac{T}{L} , \tag{3.20}$$

where T is the expected time good during the cycle and L is the expected cycle length.

We can express the expected cycle length as

$$L = n + E(U) + E(V) + E(W) , \tag{3.21}$$

where

U = time spent in replacement of part 2 between $x = 0$ and $x = n$;

V = time from n to next failure of part 2, or $(N - n)$, whichever is the smaller;

W = time spent in replacement at the end of the cycle.

It is easily seen that

$$E(U) = n\lambda_2 K_2 . \tag{3.22}$$

The random variable V has an exponential probability density with rate λ_2 from 0 to $(N - n)$, and a probability mass of $\exp[-\lambda_2(N - n]$ concentrated at $(N - n)$. The calculation of its expected value is routine and yields

$$E(V) = \frac{1}{\lambda_2}\{1 - \exp[-\lambda_2(N - n)]\} . \tag{3.23}$$

Finally, W is a random variable that takes either the value K_3 or the value K_1, and its expected value is

$$E(W) = K_3\{1 - \exp[-\lambda_2(N - n)]\} + K_1 \exp[-\lambda_2(N - n)] . \tag{3.24}$$

Summarizing (3.21)-(3.24), we obtain

$$L = n(1 + \lambda_2 K_2) + \frac{1}{\lambda_2} + K_3 - \left(\frac{1}{\lambda_2} + K_3 - K_1\right) \exp[-\lambda_2(N - n)] . \tag{3.25}$$

Note that the expected cycle length is independent of the reliability function of the nonmonitored part.

Turning now to the expected time good during the cycle, we have

$$T = E\left[\int_0^{n+V} R_1(t)\, dt\right] . \tag{3.26}$$

For the special case of an exponential part 1, we have

$$\begin{aligned} T &= E\left[\int_0^{n+V} \exp(-\lambda_1 t)\, dt\right] \\ &= E\left\{\frac{1}{\lambda_1} - \frac{1}{\lambda_1} \exp[-\lambda_1(n + V)]\right\} , \end{aligned}$$

which reduces to

$$T = \frac{1}{\lambda_1} - \frac{\exp(-\lambda_1 n)}{\lambda_1(\lambda_1 + \lambda_2)}\{\lambda_2 + \lambda_1 \exp[-(\lambda_1 + \lambda_2)(N - n)]\} . \tag{3.27}$$

Using (3.20), (3.25), and (3.27), we can compute the maximum of G and maximizing values of n and N by standard methods.

The following special cases are of interest.

(i) If $K_3 = K_1 + K_2$, then $n = N$; here part 1 can be treated independently of part 2, and the results of section 2 apply.

(ii) If $K_3 = K_2$, then $n = 0$; here part 1 is replaced whenever part 2 fails.

4. Inspection of a Single Part in the Presence of a Monitored Part

4.1 Opportunistic Inspection. We shall now consider a situation similar to that discussed in the previous section, except that we assume that part 1 is inspected before any action to replace it is begun. The inspection reveals the state of part 1, so that replacement is not made needlessly; but on the other hand, inspection may take time and may even cause the part to fail.

The same possibilities of opportunistic maintenance of part 1 occur here as in the case of replacement. Indeed, we shall show that an (n, N) policy is optimal if both parts have exponential distributions of time to failure. For nonexponential distributions, the optimal policies typically will not be of the simple (n, N) type.

It would be desirable to consider the case in which replacement is possible without, as well as with, prior inspection. We do not yet have a satisfactory analysis of this situation, but it is at least clear that optimal policies will typically be more complicated than simple (n, N) policies.

4.2 The Model. Consider now a system of two parts like the one discussed in the previous section, except that whenever replacement or inspection is not being made the following four alternative actions are possible:

(0) do nothing;

(1) begin the inspection of part 1, and if part 1 has failed, replace it;

(2) begin the replacement of part 2;

(3) begin the inspection of part 1, and if part 1 has failed, replace both parts.

In addition to the assumptions made in section 3, we assume here that

(i) inspection takes H units of time, which is *not* counted as time good;

(ii) if the part is good just prior to inspection, the probability is σ that it will survive inspection $(0 < \sigma \leqq 1)$;

(iii) both parts have exponential failure-time distributions, the failure rate for part i being denoted by λ_i. Just as in the previous section, if the monitored part has not failed, then either action (0) or action (1) is appropriate; if the monitored part has failed, then either action (2) or action (3) is appropriate.

It should be noted that when $H = \sigma = 0$, the case of inspection treated here reduces formally to the replacement case of section 3, although here we have the special exponential assumption for the nonmonitored part.

4.3 The Functional Equation for the Optimal Policy. Using the notation and concepts of the previous section, we easily obtain the following functional equation for g:

$$g(x, 1) = \max\,[\sup_{\Delta \geq 0} g_0(x \mid \Delta),\ g_1(x)]\ , \tag{4.1}$$

where

$$g_0(x \mid \Delta) = \int_0^{\Delta} [\bar{R}_1(x, u) + \exp(-\alpha u)\, g(x + u, 0)]\lambda_2 \exp(-\lambda_2 u)\, du \\ + [\bar{R}_1(x, \Delta) + \exp(-\alpha\Delta)\, g_1(x + \Delta)] \exp(-\lambda_2 \Delta)\ , \tag{4.2}$$

$$g_1(x) = \exp(-\alpha H)\{\sigma R_1(x) + [1 - \sigma R_1(x)] \exp(-\alpha K_1)\}\, g(0, 1)\ , \tag{4.3}$$

$$\bar{R}_1(x, u) = \int_0^{u} \exp(-\alpha t) R_1(x + t)\, dt\ , \tag{4.4}$$

$$R_1(x) = \exp(-\lambda_1 x)\ ; \tag{4.5}$$

and

$$g(x, 0) = \max\,[g_2(x),\ g_3(x)]\ , \tag{4.6}$$

where

$$g_2(x) = \exp(-\alpha K_2)\, g(x, 1)\ , \tag{4.7}$$

$$g_3(x) = \exp(-\alpha H)\{\sigma R_1(x) \exp(-\alpha K_2) + [1 - \sigma R_1(x)] \exp(-\alpha K_3)\}\, g(0, 1)\ . \tag{4.8}$$

4.4 Optimality of an (n, N) Policy. We shall again prove that an (n, N) policy is optimal. The proof parallels that of section 3 very closely through the proof of (3.15) and the definition of n; hence this part will be omitted. The reader should bear in mind that whereas g_1 and g_3 were constants in section 3, they are here continuous nonincreasing functions of x. In section 3, part 1 was renewed by replacement; in this section it is renewed by inspection followed by replacement if the part has failed; the "renewal property" if the part has not failed follows from the exponential assumption.

We take up the proof, therefore, where it remains to show that action (2) is optimal when $y = 0$ and $0 \leqq x \leqq n$, i.e., that $g_2(x) \geqq g_3(x)$ for $x \leqq n$. As before, we consider the case $g_2(N) < g_3(N)$.

Making use of the explicit exponential form of $R_1(x)$, we find after some calculation that the counterpart of (3.14) is

$$g(x, 1) = h(x) + \int_0^{N-x} \exp(-\alpha u)\, g(x + u, 0)\lambda_2 \exp(-\lambda_2 u)\, du\ , \tag{4.9}$$

where

$$h(x) = \left(\frac{1}{\alpha + \lambda_1 + \lambda_2}\right) \exp(-\lambda_1 x) \\ + \left\{g_1(N) \exp[-(\alpha + \lambda_2)N] - \left(\frac{1}{\alpha + \lambda_1 + \lambda_2}\right) \cdot \exp[-(\alpha + \lambda_1 + \lambda_2)N]\right\} \exp[(\alpha + \lambda_2)x] \\ = B_1 \exp(-\lambda_1 x) + B_2 \exp(\alpha + \lambda_2)x\ . \tag{4.10}$$

Define $f(x)$ by

(4.11) $$f(x) = h(x) + \int_0^{N-x} \exp(-\alpha u)\, g_3(x+u)\lambda_2 \exp(-\lambda_2 u)\, du\,;$$

then $f(x) = g(x, 1)$ for $n \leqq x \leqq N$. Also, since $g(x, 0) \geqq g_3(x)$ for all x, it follows that $f(x) \leqq g(x, 1)$ for $0 \leqq x \leqq n$. Hence, by (4.7), it is sufficient to show that $\exp(-\alpha K_2) f(x) \geqq g_3(x)$ for $0 \leqq x \leqq n$, which we shall now do.

From (4.8) it follows that

(4.12) $$g_3(x) = F_0 + F_1 \exp(-\lambda_1 x)\,,$$

where

(4.13) $$F_0 = \exp[-\alpha(H + K_3)]\, g(0, 1)$$
$$F_1 = \sigma \exp(-\alpha H)[\exp(-\alpha K_2) - \exp(-\alpha K_3)]\, g(0, 1)\,.$$

Hence

$$\int_0^{N-x} \exp(-\alpha u)\, g_3(x+u)\lambda_2 \exp(-\lambda_2 u)\, du$$
$$= A_0 + A_1 \exp(-\lambda_1 x) + A_2 \exp[(\alpha + \lambda_2)x]\,,$$

where

$$A_0 = \frac{\lambda_2 F_0}{\alpha + \lambda_2}\,,$$

(4.14) $$A_1 = \frac{\lambda_2 F_1}{\alpha + \lambda_1 + \lambda_2}\,,$$

$$A_2 = -\lambda_2\left(\frac{F_0 \exp[-(\alpha + \lambda_2)N]}{\alpha + \lambda_2} + \frac{F_1 \exp[-(\alpha + \lambda_1 + \lambda_2)N]}{\alpha + \lambda_2 + \lambda_2}\right).$$

Hence

(4.15) $$f(x) = A_0 + (A_1 + B_1)\exp(-\lambda_1 x) + (A_2 + B_2)\exp[(\alpha + \lambda_2)x]\,,$$

and

(4.16) $$\exp(-\alpha K_2) f(x) - g_3(x) = [\exp(-\alpha K_2) A_0 - F_0]$$
$$+ [\exp(-\alpha K_2)(A_1 + B_1) - F_1]\exp(-\lambda_1 x)$$
$$+ \exp(-\alpha K_2)(A_2 + B_2)\exp(\alpha + \lambda_2)x\,.$$

It is sufficient to show that the coefficient of $\exp(-\lambda_1 x)$ in (4.16) is non-negative, since in this case either

(i) $[\exp(-\alpha K_2) f(x) - g_3(x)]$ is convex, in which case it is decreasing for $x \leqq n$ because it is decreasing to the right of n, or
(ii) $[\exp(-\alpha K_2) f(x) - g_3(x)]$ is nonincreasing for all x.

The coefficient of $\exp(-\lambda_1 x)$ in (4.16) is

$$C \equiv \exp(-\alpha K_2)(A_1 + B_1) - F_1\,,$$

which from (4.10) and (4.14) can be expressed as

(4.17) $$C = \frac{\exp(-\alpha K_2) - F_1\{\alpha + \lambda_1 + \lambda_2[1 - \exp(-\alpha K_2)]\}}{\alpha + \lambda_1 + \lambda_2}\,.$$

Let $\bar{g}$ denote the value of $g\ (0, 1)$ when $\sigma = 1$ and $H = 0$ (inspection is free). Then $g(0, 1) \leqq \bar{g}$. Moreover, $\bar{g}$ can be calculated explicitly, since the optimal policy in this case is clearly to monitor part 1 and to replace each part just when it fails (with probability 1, parts 1 and 2 are never replaced together).

LEMMA. $$\bar{g} = \frac{1}{\alpha + \lambda_1[1 - \exp(-\alpha K_1)] + \lambda_2[1 - \exp(-\alpha K_2)]}\ .$$

PROOF. Let z denote the time to the first failure; z has an exponential distribution with rate $(\lambda_1 + \lambda_2)$. The replacement time is K_i with probability $(\lambda_i/\lambda_1 + \lambda_2)$, $i = 1, 2$. Hence

$$\begin{aligned}\bar{g} = \int_0^\infty \Big[\int_0^z \exp(-\alpha t)\, dt + \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right) \exp[-\alpha(z + K_1)\bar{g}] \\ + \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right) \exp[-\alpha(z + K_2)]\, \bar{g}\Big](\lambda_1 + \lambda_2) \exp[-(\lambda_1 + \lambda_2)z]\, dz\ .\end{aligned}$$

Solving for $\bar{g}$ yields the lemma.

From (4.13) and the lemma it follows that

$$\begin{aligned}F_1 &\leqq \frac{\sigma \exp(-\alpha H)[\exp(-\alpha K_2) - \exp(-\alpha K_3)]}{\alpha + \lambda_1[1 - \exp(-\alpha K_1)] + \lambda_2[1 - \exp(-\alpha K_2)]} \\ &\leqq \frac{\exp(-\alpha K_2)[1 - \exp(-\alpha K_1)]}{\alpha + \lambda_1[1 - \exp(-\alpha K_1)] + \lambda_2[1 - \exp(-\alpha K_2)]}\ ,\end{aligned}$$

since $0 < \sigma \leqq 1$, $0 < \exp(-\alpha H) \leqq 1$ and $K_3 \leqq K_1 + K_2$. Hence, from this and (4.17), it suffices to show that

$$1 - \frac{[1 - \exp(-\alpha K_1)]\{\alpha + \lambda_1 + \lambda_2[1 - \exp(-\alpha K_2)]\}}{\alpha + \lambda_1[1 - \exp(-\alpha K_1)] + \lambda_2[1 - \exp(-\alpha K_2)]} > 0\ ,$$

which is easily verified.

4.5 Long-Run Average Time Good Under an (n, N) Inspection Policy. As in the case of replacement, we can compute the discounted time good from the functional equations, or we can compute the long-run average time good directly. For the latter the details are obvious but tedious, and we present here only the results.

The expected cycle length can be shown to be

$$L = n(1 + \lambda_2 K_2) + \frac{1}{\lambda_2}\{1 - \exp[-\lambda_2(N - n)]\} + E(W)\ , \tag{4.18}$$

where

$$\begin{aligned}E(W) &= (1 - \sigma)K_1 P_0 + K_1 P_1 + [(1 - \sigma)K_3 + \sigma K_2]P_2 + K_3 P_3 + H\ , \\ P_0 &= \exp[-\lambda_1 N - \lambda_2(N - n)]\ , \\ P_1 &= \exp[-\lambda_2(N - n)] - P_0\ , \\ P_2 &= \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)[\exp(-\lambda_1 n - P_0)]\ , \\ P_3 &= 1 - \exp[-\lambda_2(N - n)] - P_2\ .\end{aligned} \tag{4.19}$$

The expected time good during a cycle is the same as in the case of replacement:

$$T = \frac{1}{\lambda_1} - \frac{\exp(-\lambda_1 n)}{\lambda_1(\lambda_1 + \lambda_2)}\{\lambda_2 + \lambda_1 \exp[-(\lambda_1 + \lambda_2)(N - n)]\}$$

$$T = \frac{1}{\lambda_1}\left(1 - \frac{\lambda_2 \exp(-\lambda_1 n) + \lambda_1 \exp[-\lambda_1 N - \lambda_2(N - n)]}{\lambda_1 + \lambda_2}\right). \quad (4.20)$$

The following special cases are of interest:

(i) If $H = \sigma = 0$, then we have the replacement case of section 3.

(ii) If $K_1 + K_2 = K_3$, then $n = N$.

(iii) If $H = 0$ and $\sigma = 1$, then $n = N = 0$; i.e., the first part is also monitored.

From $K_3 = K_2$ we cannot conclude that $n = 0$, since inspection takes time. However, in this case the model of this section would appear inappropriate, as it rules out the possibility of replacement without inspection.

5. Conclusion

In preparedness-type models of single-part equipment set up to maximize the proportion of the time that the system is in a ready state, optimal replacement policies are periodic, and the optimal period is easily computed. The same is true of inspection policies if the part has an exponential distribution of time to failure.

In the case of equipment with several parts, the possibility arises of opportunistic maintenance to take advantage of "economies of scale" in maintenance actions. In the simple cases we have discussed of equipment with two parts, one of which is monitored, the optimal policies are of the (n, N) type, which includes the periodic type as a special case.

Models with more than two parts appear to present serious analytical difficulties if one wishes to take account of economies of scale. It is likely, furthermore, that optimal policies in such cases would typically be too complicated for practical use. Therefore, policies of the (n, N) type may provide good practical solutions to problems involving many-part systems. Much work remains, of course, in developing techniques for optimizing even within this simple class.

References

[1] ALCHIAN, A. A. *Economic Replacement Policy*, The RAND Corporation, Santa Monica, Calif., Report R-225, April 12, 1952.

[2] ARROW, K. J., S. KARLIN, and H. SCARF. *Studies in the Mathematical Theory of Inventory and Production*, Stanford, Calif.: Stanford Univ. Press, 1958.

[3] BARLOW, R. and L. HUNTER. Optimum Preventive Maintenance Policies, *Operations Res.*, 1960, **8**(1), 90-100.

[4] CAMPBELL, N. R. The Replacement of Perishable Members of a Continually Operating System, *J. Roy. Stat. Soc.*, 1941, **7**(suppl.), 110-30.

[5] DRENICK, R. F. Mathematical Aspects of the Reliability Problem, *J. Soc. for Indust. and Appl. Math.*, 1960, **8**(1), 125-49.

[6] DREYFUS, S. *A Note on an Industrial Replacement Process*, The RAND Corporation, Santa Monica, Calif., Paper P-1045, March 27, 1957.

[7] FELLER, W. *An Introduction to Probability Theory and its Application*, vol. 1, 2d ed., New York: Wiley, 1957.

[8] FLEHINGER, B. J. Systems Reliability as a Function of System Age: Effects of Intermittent Component Usage and Periodic Maintenance, *Operations Res.* 1960, **8**(1), 30-44.

[9] KAMINS, M. *Determining Checkout Intervals for Systems Subject to Random Failures*, The RAND Corporation, Santa Monica, Calif., Research Memorandum RM-2578, June 15, 1960.

[10] KLEIN, M., and L. ROSENBERG. Deterioration of Inventory and Equipment, *Naval Res. Logist. Quart.*, 1960, **7**(1), 49-62.

[11] LOTKA, A. J. A Contribution to the Theory of Self-Renewing Aggregates with Special Reference to Industrial Replacement, *Ann. Math. Stat.*, 1939, **10**(1), 1-25.

[12] MCGLOTHLIN, W. H., and E. BEAN. *An Analytical Model for Developing Optimal Ballistic Missile Maintenance Policies*, The RAND Corporation, Santa Monica, Calif., Paper P-1696, May 13, 1959.

[13] MORSE, P. M. Maintenance of Equipment, Chapter 11 in *Queues, Inventories, and Maintenance*, New York: Wiley, 1958, pp. 157-79.

[14] PALM, C. The Distribution of Repairmen in Servicing Automatic Machinery (in Swedish), *Industridninger Norden*, 1947, **75**, 75-80, 90-94, 119-23.

[15] SASIENI, M. W. A Markov Chain Process in Industrial Replacement, *Operational Res. Quart.*, 1956, **7**(4), 148-54.

[16] SAVAGE, I. R. Cycling, *Naval Res. Logist. Quart.*, 1956, **3**(3), 163-75.

[17] SOLOMON, H., and C. DERMAN. The Development and Evaluation of Surveillance Sampling Plans, *Management Sci.*, 1958, **5**(1), 72-88.

]18] WEISS, G. H. On the Theory of Replacement of Machinery with a Random Failure Time, *Naval Res. Logist. Quart.*, 1956, **3**(4), 279-94.

[19] WELKER, E. L. *Relationship Between Equipment Reliability Preventive Maintenance Policy and Operating Costs*, ARINC Research Corporation, Washington, D.C., ARINC Research Monograph No. 7, February 13, 1959.

[20] WELKER, E. L., and C. E. BRADLEY. *A Model for Scheduling Maintenance Utilizing Measures of Equipment Performance*, ARINC Research Corporation, Washington, D. C., ARINC Research Monograph No. 8, October 1, 1959.