Probability in the Engineering and Informational Sciences, **16**, 2002, 339–349. Printed in the U.S.A.

OPTIMAL MAINTENANCE AND OPERATION OF A SYSTEM WITH BACKUP COMPONENTS

RHONDA RIGHTER
Department of Operations and Management Information Systems
Santa Clara University
Santa Clara, CA 95053
E-mail: RRighter@scu.edu

We consider a system with heterogeneous unreliable components that requires only one component to be turned on in order for it to operate. Repair workers may have different skills and may be unavailable for random periods of time. The problem is to determine a usage and repair policy to maximize system availability. We give conditions under which the optimal usage policy is to always use, or turn on, the component with the shortest repair time, and the optimal repair policy is to always repair the most reliable component (with the smallest failure rate). We fully characterize the optimal policy when there are only two components. Our system is equivalent to a closed system with multiple single-server queues, where the objective is to minimize server idle time at one of the queues.

1. INTRODUCTION

Many practical systems require a higher availability than is possible with individual unreliable components, so redundancy is introduced. It may not be necessary or cost-effective to require all of the components to be equally reliable or to train repair workers to repair all of the possible components. We consider a system with heterogeneous components in which we must determine both usage and repair policies. The usage policy tells us which of the operational components to use at any given time. For each repair worker, the repair policy determines which of the failed components to repair, among the components that the worker is trained to repair.

Our system requires one working component for the system to be operational; n components are available. The component currently in the system, or turned on, is

© 2002 Cambridge University Press 0269-9648/02 $12.50

https://doi.org/10.1017/S0269964802163054 Published online by Cambridge University Press

subject to failure, but other components that are in working order remain so until they are used in the system (cold standby). When component i is turned on, its up time, or time to failure, is exponentially distributed with failure rate λ_i. We initially assume that there is a single repair worker to repair failed components, although we later consider the case in which different components may be assigned to different repair workers. The repair time of component i is exponentially distributed with repair rate μ_i. All repair and failure times are independent. Components may be swapped in and out of the system (turned on and off) preemptively without cost, and repair of components may be preempted without cost. At any point in time, we must decide which of the working components to use and which of the failed components to repair, so as to maximize the system availability.

Rabinowitz, Goren, and Mehrez [7] studied a related deterministic problem. In their problem, there are only two components and the schedule of on periods and maintenance periods must be done on a cyclical basis. They give references to several applications areas, including military applications (such as operating remotely piloted vehicles and real-time data acquisition and processing) and applications in communications, signal processing, surveillance, and manufacturing. Also, see their article for references to other work on the deterministic problem.

Earlier work on the stochastic model has assumed warm standby, so all working components are subject to failure and, generally, the system is a series system, so all components must be working for the system to be working. This is the classic stochastic repairman or machine-interference problem, and the only decision is which of the failed components to repair. The SFR (smallest failure rate) policy, which always repairs the component with the smallest failure rate, has been shown to be optimal for a variety of assumptions and objectives by Smith [9], Derman, Lieberman, and Ross [2], Nash and Weber [6], Katehakis and Derman [4], Courcoubetis, Varaiya, and Walrand [1], Frostig [3], Righter [9], and Koole and Vrijenhoek [5].

We show for our model, in which only one working component is required for the system to function and other working components may be turned off, that LRR (largest-repair-rate first) is optimal for choosing which of the working components to turn on, as long as the repair policy also gives priority to components with larger repair rates; that is, we should always turn on, or use, the component with the shortest expected repair time among the working components. For the general model, the optimal usage and optimal repair policies are more complicated and depend on both repair and failure rates. We completely characterize the policy for systems with two components, for both discounted and average cost optimality. For more general systems, we show that when LRR is consistent with SFR, the optimal policy repairs the component with smallest failure rate among the nonworking components; that is, the optimal policy is LRR for both the usage policy and the repair policy. We also extend our results to random environments and systems with multiple, possibly differently skilled, repair workers.

The classic machine-interference problem is equivalent to a finite-source queuing system or to a closed hub-and-spoke queuing system. The repair worker can be represented as a single-server exponential queue (the hub, or queue 0), where the

https://doi.org/10.1017/S0269964802163054 Published online by Cambridge University Press

service rate for job i at the hub is μ_i. When component i is working, it can be represented as a customer at queue i (a spoke) with an exponentially distributed service time with rate λ_i, so customer i is routed from queue i to the hub queue and back again. Customers wait for service only when they are at the hub, so the scheduling problem is to determine which customer to serve next at the hub. Maximizing the time that a series system is working is equivalent to maximizing server idleness at the hub.

Our model is also equivalent to a closed queuing system, but now there are only two single-server queues, one representing the repair worker as before, with service rates μ_i, and the second representing the system operation, with service rates λ_i. Customers are routed from the first queue to the second and back again. Now, customers at both queues must be scheduled, and our objective is equivalent to minimizing the idle time at the second server. The LRR policy corresponds to SEPT (shortest-expected-processing-time first) at server 1. Multiple repair workers can be represented as multiple single-server queues.

We first give some intuitively obvious properties of the optimal policy. We then completely characterize the optimal policy for two-component systems. Finally, we consider general systems with multiple components. Throughout, we will use increasing, larger, and so forth in the nonstrict sense.

2. BASIC PROPERTIES OF THE OPTIMAL POLICY

It is very easy to show the following with sample path arguments. Similar properties were shown for the deterministic two-component model by Rabinowitz et al. [7]. Parts (i), (ii), and (iv) of the lemma hold when repair and failure times have arbitrary distributions, as preemption is permitted and costs nothing, and part (iii) holds for exponential failures and repairs.

LEMMA 2.1: *The stochastically optimal policy*

(i) never idles the repair worker when there are nonworking components
(ii) never turns on more than one working component
(iii) never repairs a working component
(iv) and never leaves all working components turned off.

That is, for any policy that violates one of the above, there exists a policy that does not such that the system has a stochastically larger up time, or availability, during any time period.

3. THE OPTIMAL POLICY FOR A TWO-COMPONENT SYSTEM

When there are only two components, we can completely characterize the policy that maximizes the long-run proportion of time that the system is working, or the discounted time the system is working. Let us denote by (a, b) the policy that always uses component a when both components are working and always repairs component b when both components are down and is consistent with Lemma 2.1. We know

https://doi.org/10.1017/S0269964802163054 Published online by Cambridge University Press

from standard stochastic dynamic programming theory that such a stationary and deterministic policy is optimal for this simple, finite-state Markovian system.

3.1. Long-Run System Availability

We first consider the problem of maximizing the long-run average time the system is working (i.e., the long-run proportion of time that at least one component is working).

Consider policy (1,1). We define a renewal cycle as the cycle that begins and ends at the point when both components are working. Let b be the time until the end of the cycle when we start with *both* components working, let d be the time until the end of the cycle when we start with the system *down* (i.e., both components down), and let d_i, $i = 1,2$, be the time until the end of the cycle when we start with only component i down. Then we have the following system of equations:

$$b = \frac{1}{\lambda_1} + d_1,$$

$$d = \frac{1}{\mu_1} + d_2,$$

$$d_1 = \frac{1}{\lambda_2 + \mu_1} + \frac{\lambda_2}{\lambda_2 + \mu_1} d,$$

$$d_2 = \frac{1}{\lambda_1 + \mu_2} + \frac{\lambda_1}{\lambda_1 + \mu_2} d,$$

with solution

$$b = \frac{\lambda_2\mu_1\mu_2 + \mu_1^2\mu_2 + \lambda_1^2\lambda_2 + \lambda_1\mu_2\lambda_2 + \lambda_1\lambda_2\mu_1 + \lambda_1\mu_2\mu_1}{\lambda_1\mu_1\mu_2(\lambda_2 + \mu_1)},$$

$$d = \frac{\lambda_1 + \mu_2 + \mu_1}{\mu_1\mu_2},$$

$$d_1 = \frac{\lambda_1\lambda_2 + \lambda_2\mu_2 + \lambda_2\mu_1 + \mu_1\mu_2}{\mu_1\mu_2(\lambda_2 + \mu_1)},$$

$$d_2 = \frac{\lambda_1 + \mu_1}{\mu_1\mu_2}.$$

Similarly, letting $\hat{b}$, $\hat{d}$, $\hat{d}_1$, and $\hat{d}_2$ be the total amount of time the system is working until the end of the cycle for the corresponding starting states, we have

$$\hat{b} = \frac{\lambda_2\mu_2 + \mu_1\mu_2 + \lambda_1\mu_2 + \lambda_1\lambda_2}{\lambda_1\mu_2(\lambda_2 + \mu_1)}.$$

https://doi.org/10.1017/S0269964802163054 Published online by Cambridge University Press

Thus, from standard renewal theory, the long-run proportion of time that the system is working under policy (1,1) is

$$p(1,1) := \frac{\hat{b}}{b} = \frac{\mu_1\mu_2(\lambda_1+\lambda_2)+\lambda_1\lambda_2\mu_1+\mu_1^2\mu_2}{\mu_1\mu_2(\lambda_1+\lambda_2)+\lambda_1\lambda_2(\mu_1+\mu_2)+\mu_1^2\mu_2+\lambda_1^2\lambda_2}.$$

A similar analysis gives us

$$p(1,2) = \frac{\mu_1\mu_2+\lambda_1\mu_2}{\mu_1\mu_2+\lambda_1\mu_2+\lambda_1\lambda_2}$$

and, by symmetry,

$$p(2,1) = \frac{\mu_1\mu_2+\lambda_2\mu_1}{\mu_1\mu_2+\lambda_2\mu_1+\lambda_1\lambda_2},$$

$$p(2,2) = \frac{\mu_1\mu_2(\lambda_1+\lambda_2)+\lambda_1\lambda_2\mu_2+\mu_1\mu_2^2}{\mu_1\mu_2(\lambda_1+\lambda_2)+\lambda_1\lambda_2(\mu_1+\mu_2)+\mu_1\mu_2^2+\lambda_1\lambda_2^2}.$$

Therefore, we prefer to use component 1 instead of component 2 when both are up if $\Delta_1 \geq 0$ and $\Delta_2 \geq 0$, where

$$\begin{aligned}\Delta_1 &:= p(1,1) - p(2,1)\\ &= \frac{\mu_1\mu_2(\lambda_1+\lambda_2)+\lambda_1\lambda_2\mu_1+\mu_1^2\mu_2}{\mu_1\mu_2(\lambda_1+\lambda_2)+\lambda_1\lambda_2(\mu_1+\mu_2)+\mu_1^2\mu_2+\lambda_1^2\lambda_2} - \frac{\mu_1\mu_2+\lambda_2\mu_1}{\mu_1\mu_2+\lambda_2\mu_1+\lambda_1\lambda_2}\\ &= \lambda_1\lambda_2\mu_1\mu_2\frac{\mu_1-\mu_2}{(\mu_1\mu_2+\lambda_1\lambda_2+\lambda_2\mu_1)(\mu_1\mu_2(\lambda_1+\lambda_2)+\lambda_1\lambda_2(\mu_1+\mu_2)+\mu_1^2\mu_2+\lambda_1^2\lambda_2)}\end{aligned}$$

and

$$\begin{aligned}\Delta_2 &:= p(1,2) - p(2,2)\\ &= \lambda_1\lambda_2\mu_1\mu_2\frac{\mu_1-\mu_2}{(\mu_1\mu_2+\lambda_1\lambda_2+\lambda_1\mu_2)(\mu_1\mu_2(\lambda_1+\lambda_2)+\lambda_1\lambda_2(\mu_1+\mu_2)+\mu_1\mu_2^2+\lambda_1\lambda_2^2)}.\end{aligned}$$

Thus, we prefer to use component 1 if and only if $\mu_1 \geq \mu_2$; that is, we should use the component with the fastest repair rate whenever it is working, regardless of the failure rates. (Of course, when $\mu_1 = \mu_2$, we are, in fact, indifferent.)

Let us relabel the components so that $\mu_1 \geq \mu_2$, and therefore it is optimal to use component 1 whenever it is working. We will prefer to repair component 1 when both components are down if and only if $\Delta_3 \geq 0$, where

$$\begin{aligned}\Delta_3 &:= p(1,1) - p(1,2)\\ &= \lambda_1\lambda_2\frac{\mu_1\mu_2(\mu_1-\mu_2)+(\lambda_1+\mu_2)(\lambda_2\mu_1-\lambda_1\mu_2)}{(\mu_1\mu_2+\lambda_1\lambda_2+\lambda_1\mu_2)(\mu_1\mu_2(\lambda_1+\lambda_2)+\lambda_1\lambda_2(\mu_1+\mu_2)+\mu_1^2\mu_2+\lambda_1^2\lambda_2)};\end{aligned}$$

https://doi.org/10.1017/S0269964802163054 Published online by Cambridge University Press

that is, we prefer to repair component 1 when

$$\mu_1\mu_2(\mu_1 - \mu_2) + (\lambda_1 + \mu_2)(\lambda_2\mu_1 - \lambda_1\mu_2) \geq 0.$$

The following theorem summarizes our results.

THEOREM 3.1: *For a two-component system, LRR is the optimal usage policy to maximize the long-run system availability; that is, it is optimal to use the working component with the largest repair rate, regardless of the failure rates. Let component 1 be the component with the largest repair rate, so* $\mu_1 \geq \mu_2$. *It is optimal to repair component 1 when both components are failed if and only if*

$$\mu_1\mu_2(\mu_1 - \mu_2) + (\lambda_1 + \mu_2)(\lambda_2\mu_1 - \lambda_1\mu_2) \geq 0.$$

A sufficient condition for the latter is $\lambda_1 \leq \lambda_2$, *or, more generally,* $\lambda_1\mu_2 \leq \lambda_2\mu_1$. *Thus, if the repair and failure rates are agreeable in the sense that the LRR ordering agrees with the SFR ordering, then SFR is the optimal repair policy.*

3.2. Discounted System Availability

Now we suppose that we have exponential discounting with rate α and we wish to maximize the long-run expected discounted system availability or, equivalently, to minimize the long-run expected discounted system unavailability. Redefining our earlier terms, let b, d, d_1, and d_2 represent the states when both components are working, both are down, component 1 alone is down, and component 2 alone is down, respectively. Let $V(x)$ be the discounted system *unavailability* under the optimal policy starting in state x, $x = b, d, d_1, d_2$, and let us scale our rate parameters so that $\alpha + \lambda_1 + \lambda_2 + \mu_1 + \mu_2 = 1$. Then we have the following dynamic programming equations, where we have used uniformization to assume there are (possibly dummy) state transitions according to a Poisson process with rate $\alpha + \lambda_1 + \lambda_2 + \mu_1 + \mu_2 = 1$.

$$V(b) = \min\{\lambda_1 V(d_1) + (\lambda_2 + \mu_1 + \mu_2)V(b); \lambda_2 V(d_2) + (\lambda_1 + \mu_1 + \mu_2)V(b)\},$$

$$V(d) = \min\{1 + \mu_1 V(d_2) + (\lambda_1 + \lambda_2 + \mu_2)V(d);$$
$$1 + \mu_2 V(d_1) + (\lambda_1 + \lambda_2 + \mu_1)V(d)\},$$

$$V(d_1) = \lambda_2 V(d) + (\lambda_1 + \mu_2)V(d_1) + \mu_1 V(b),$$

$$V(d_2) = \lambda_1 V(d) + \mu_2 V(b) + (\lambda_2 + \mu_1)V(d_2).$$

Consider policy (1,1) of using and repairing component 1 whenever possible. We can obtain the discounted system unavailability under that policy, call it $V_{(1,1)}(x)$, from the set of above equations by taking the first of the two terms in the first two equations. Let policy (2-1,1) be the policy that, if the initial state is b, uses component 2 until the first state transition, and thereafter follows policy (1,1), and let $V_{(2-1,1)}(x)$ be its discounted system unavailability. From the principle of policy iteration, (1,1) cannot be the optimal policy if $\Delta < 0$, where

https://doi.org/10.1017/S0269964802163054 Published online by Cambridge University Press

$$\begin{aligned}\Delta &:= V_{(2-1,1)}(b) - V_{(1,1)}(b)\\ &= \lambda_2 V_{(1,1)}(d_2) + (\lambda_1 + \mu_1 + \mu_2)V_{(1,1)}(b) - \lambda_1 V_{(1,1)}(d_1)\\ &\quad - (\lambda_2 + \mu_1 + \mu_2)V_{(1,1)}(b)\\ &= \frac{\lambda_1\lambda_2(\mu_2 - \mu_1)}{D},\end{aligned}$$

where

$$\begin{aligned}D = {} & \lambda_1\lambda_2^2 - \lambda_1\mu_2 - 2\lambda_1\lambda_2 + \lambda_1 + \lambda_1\mu_2\lambda_2 - 1 + 2\lambda_2 + \mu_1 + 2\mu_2 - 2\mu_1\mu_2\\ & - 3\mu_2\lambda_2 - \mu_1\lambda_2 - \mu_2^2 - \lambda_2^2 + \mu_1\mu_2^2 + \mu_2\lambda_2^2 + \mu_2^2\lambda_2 + \mu_1\mu_2\lambda_2,\end{aligned}$$

and where the last equality for Δ comes from solving the set of four equations for $V_{(1,1)}(x)$. We can rewrite the denominator, D, using the fact that $\alpha = 1 - (\lambda_1 + \lambda_2 + \mu_1 + \mu_2)$ as follows:

$$D = -[\alpha(1 - \lambda_2) + \mu_1\mu_2](1 - \mu_2) - \lambda_1\lambda_2(1 - \lambda_2) \leq 0.$$

Thus, $\Delta < 0$ if and only if $\mu_2 > \mu_1$. Therefore, if $\mu_2 > \mu_1$ and it is optimal to always repair component 1, it cannot be optimal to use component 1. A similar argument shows that if $\mu_2 > \mu_1$ and it is optimal to always repair component 2, it cannot be optimal to use component 1. Hence, as in the average system availability case, it is optimal to use the component with the largest repair rate, regardless of the failure rates and the discount rate.

Now let us relabel the components so that $\mu_1 > \mu_2$ and it is optimal to use component 1 whenever possible. Let policy (1,2-1) be the policy that, if the initial state is d, repairs component 2 until the first state transition and thereafter follows policy (1,1), and let $V_{(1,2-1)}(x)$ be its discounted system unavailability. By comparing $V_{(1,1)}(d)$ and $V_{(1,2-1)}$, we have (after some algebra) that it is optimal to repair component 1 when

$$(\alpha + \lambda_1)[\mu_1(1 - \lambda_1) - \mu_2(1 - \lambda_2)] + \mu_1[\mu_1(\mu_2 - \lambda_1) - \mu_2(\mu_2 - \lambda_2)] \geq 0.$$

Summarizing, we have the following.

THEOREM 3.2: *For a two-component system, LRR is the optimal usage policy to maximize the expected discounted system availability; that is, it is optimal to use the working component with the largest repair rate, regardless of the failure rates or the discount rate. Let component 1 be the component with the largest repair rate, so* $\mu_1 \geq \mu_2$. *It is optimal to repair component 1 when both components are failed if and only if*

$$(\alpha + \lambda_1)[\mu_1(1 - \lambda_1) - \mu_2(1 - \lambda_2)] + \mu_1[\mu_1(\mu_2 - \lambda_1) - \mu_2(\mu_2 - \lambda_2)] \geq 0.$$

A sufficient condition for the latter is $\lambda_1 \leq \lambda_2$. *Thus, if the repair and failure rates are agreeable in the sense that LRR ordering agrees with SFR ordering, then SFR is the optimal repair policy.*

https://doi.org/10.1017/S0269964802163054 Published online by Cambridge University Press

4. MORE THAN TWO COMPONENTS

We now show that whenever the repair policy agrees with LRR, then LRR is still optimal even when there are more than two components. By agreeing with LRR, we mean that components with higher repair rates have a higher priority for repair. We will use a sample-path argument, which will allow us to show the stronger result that for any policy that does not use LRR for its usage policy, there exists a policy that uses LRR for its usage policy such that the latter policy has stochastically more system up time in any time interval.

THEOREM 4.1:

(i) If repairs are done according to the LRR policy, then the LRR policy is the stochastically optimal usage policy; that is, the system will work for a stochastically longer time in any finite time period under the LRR usage policy than under any other usage policy.

(ii) When failure and repair rates are agreeable in the sense that components with larger repair rates also have smaller failure rates, so $\mu_1 \geq \cdots \geq \mu_n$ *and* $\lambda_1 \leq \cdots \leq \lambda_n$*, then the LRR policy is the stochastically optimal usage and repair policy.*

PROOF: Our proof of (i) is similar in structure to the proof of Theorem 2.2 in [8], so we concentrate on the differences.

We use uniformization, so that potential events, or decision points, occur according to a Poisson process with rate $\sum \lambda_i + \sum \mu_i$. We suppose that there are a finite number of decision points, N, and use induction on N, where $N = 0$ is trivial. Suppose the result holds for N and consider $N + 1$. Assume that we have a decision point at time 0. For simplicity, we will denote the time of the lth decision point after time 0 by time l.

Let π be a policy that disagrees with LRR for its first decision and agrees with LRR thereafter. It is sufficient to construct a policy π' that agrees with LRR for its first decision and such that the system up time is stochastically larger under π' than under π. (See [8].) We may also assume that π agrees with the properties of Lemma 2.1.

Suppose that, at time 0, π uses working component j when component i is also working and component $i \neq j$ has the largest repair rate among all working components. Let π' use component i at time 0. If the event at time 1 is neither the potential failure of components i or j, then, at that point, both systems are in the same state, and letting π' agree with π and coupling all the random variables directly thereafter, we have that the system up times are the same under both policies. Therefore, let us suppose that the event at time 1 is the potential failure of component i or component j. Because π agrees with LRR from time 1 on, it will use component i until it fails (possibly with interruptions when components with larger repair rates are working) before it uses component j again. Let σ be the time of the first potential failure of either component i or j that occurs while π is using component i. If, at time 1,

https://doi.org/10.1017/S0269964802163054 Published online by Cambridge University Press

component j under policy π fails, let R_j be the time at which it is repaired under policy π, as long as it is less than σ; otherwise, let it equal σ. Let R_i be the minimum of σ and the time at which component i would be repaired if it were down at time 1 and if, between times 1 and R_j, the repair worker repaired i instead of j whenever j was being repaired under policy π and did not repair i whenever j was not repaired under π. Let R_k be defined in the same way as R_i, where k is the component that π would repair instead of j if j does not fail under π at time 1 until time R_j. Note that $\mu_j \geq \mu_k$, because we assume that π follows the LRR repair policy. Then, because $\mu_i \geq \mu_j \geq \mu_k$, we can couple R_i, R_j, and R_k so that $R_i \leq R_j \leq R_k$ with probability 1. Between times 1 and R_i, let π' use component j whenever π uses component i, and if, at time 1, component i under policy π' fails, let π' treat component i as if it were component j under the LRR policy for its repair decisions. Otherwise, let π' agree with π. Note that between times 0 and R_i the system is up under both policies.

Case 1: $R_i < \sigma$. In this case, at time R_i both components i and j will be working under π', whereas under π, component i will be working and component j may be either down or up, and all other components will be in the same state under the two policies (because R_k is greater than R_i, and, except for components i, j, and k, both policies repair the same components). From time R_i on, we couple all random variables directly under the two policies (so we no longer use R_j, R_k, or σ, which is all right because our system is memoryless). Let π' agree with π (and LRR) from time R_i on, except that π' idles the repair worker when π repairs component j and j is working under π', and π' uses component j when all other components are down and component j is down under π. As soon as component j is repaired under π or it fails under π', the two systems will be in the same state, and from then on, letting π' agree with LRR and coupling everything directly, they will have the same system up times. Until that time, if j fails at time 1 under π, there may be times when the system is up under π' but not under π. The reverse will never be true. Finally, a sample-path argument shows that a nonidling LRR repair policy yields a stochastically larger system up time than π' and, therefore, than π.

Case 2: $\sigma = R_i$. By definition, π was using component i and π' was using component j between times $\sigma - 1$ and σ, and the event at time σ is a potential failure for components i or j. In this case, if either component i or j fails by time σ in either system, it will still be failed at time σ. We couple the potential i or j failures at times 1 and σ under the two policies so that the same event occurs at time 1 under π as at time σ under π' and vice versa. Then component i will be down under π at time σ (because it fails at time σ) if and only if it is down under π' at time σ (because it failed at time 1 and has not yet been repaired), and similarly for component j; that is, at time σ both systems will be in the same state, and from then on, they will have the same system up time. Before time σ, both systems will be working.

Now let us consider (ii). The above argument for (i) again shows that LRR is the optimal usage policy (now we can assume that π follows LRR for repairs as well as for its usage policy from time 1 on by induction). The argument to show that LRR is

https://doi.org/10.1017/S0269964802163054 Published online by Cambridge University Press

also the optimal repair policy is very similar to (i), interchanging the roles of "use" and "repair." Now we suppose π repairs component j at time 0 when failed component i has a larger repair rate, and thereafter follows LRR for both its usage and repair decisions, and let π' repair i at time 0. Again we may assume that the event at time 1 is a potential repair of either component i or j, and we define σ as the time of the first potential repair event that occurs while π is repairing component i. Because $\mu_i \geq \mu_j$, we can couple the events at times 1 and σ so that one of the following five situations occur: (i) the component that is being repaired at times 1 and σ completes repair at both those times under both policies, (ii) the component that is being repaired at times 1 and σ does not complete repair at either time under both policies, (iii) a component is repaired at time 1 but not at time σ under both policies, (iv) a component is repaired at time σ but not at time 1 under both policies, and (v) component i is repaired at time 1 and no component is repaired at time σ under policy π', and component i is repaired at time σ and no component is repaired at time 1 under policy π. The argument otherwise proceeds very much as in (i) and is left to the reader. ■

5. EXTENSIONS

The reason we need to assume that the repair strategy agrees with LRR in our proof of part (i) of Theorem 4.1 is so that "component k" could be assumed to have a stochastically larger repair time than component i in Case 1. If there are only two components, or if all components are working, that assumption is unnecessary. Also, if there are $n-1$ or n repair workers, then we can couple the repairs of all the other components except i and j so that they are the same under both policies, and again the assumption is unnecessary. Indeed, the case of n repair workers is analogous to the classic machine-interference problem.

COROLLARY 5.1: *For two components, or when there are at least $n-1$ repair workers, regardless of the repair strategy, LRR is the stochastically optimal usage strategy. For more than two components and one repair worker, it is stochastically optimal to always use the component with the largest repair rate whenever all components are working.*

A careful look at our proof of Theorem 4.1 shows that the results also hold in a Markov random environment, in which the failure and repair rates depend on the state of the environment, as long as the LRR ordering of components is the same regardless of the state. Let the environmental state X change according to a Markov process, and assume that when $X = x$, component i has failure rate $\lambda_i(x)$ and repair rate $\mu_i(x)$. For (i) to hold, we need that the order of the repair rates for the components does not change in different environments, $\mu_1(x) \geq \mu_2(x) \geq \cdots \geq \mu_n(x)$ for all x, so the LRR policy is the same policy in all environmental states. For (ii) to hold, we also need $\lambda_1(x) \leq \lambda_2(x) \leq \cdots \leq \lambda_n(x)$ for all x. In particular, our results hold when the system may not need to operate for certain random periods of time (during

https://doi.org/10.1017/S0269964802163054 Published online by Cambridge University Press

which times all failure rates are zero) and when the repair worker may be unavailable for random periods of time.

We can also have multiple repair workers, with differing abilities. For example, components 1 and 2 may have their own dedicated repair workers, and components 3 and 4 may be repaired by a third repair worker. Then, the LRR policy is optimal for choosing between using components 3 and 4 and for choosing between using components 1 and 2. Moreover, components 1 and 2 should have priority over other components with smaller failure rates, because for components with the same repair rates, those with dedicated repair workers will generally be repaired more quickly. More generally, we have the following theorem. Its proof is an easy extension of the proof of Theorem 4.1, and is left to the reader.

THEOREM 5.2: *Suppose there are multiple repair workers and each worker can repair a subset of the components, where the subsets are mutually exclusive and exhaustive. Also, suppose that repair workers that are responsible for more than two components decide which to repair according to LRR. Then, among working components that are assigned to the same repair worker, it is optimal to choose which one to use according to LRR, and components with dedicated repair workers should have priority for usage over all other components with smaller repair rates.*

Acknowledgment

The author was supported by the Breetwor Fellowship of Santa Clara University.

References

1. Courcoubetis, C., Varaiya, P., & Walrand, J. (1984). Invariance in resource-sharing systems. *Journal of Applied Probability* 21: 777–785.
2. Derman, C., Lieberman, G.J., & Ross, S.M. (1980). On the optimal assignment of servers and a repairman. *Journal of Applied Probability* 17: 577–581.
3. Frostig, E. (1993). Optimal policies for machine repairmen problems. *Journal of Applied Probability* 30: 703–715.
4. Katehakis, M.N. & Derman, C. (1984). Optimal repair allocation in a series a system. *Mathematics of Operations Research* 9: 615–623.
5. Koole, G. & Vrijenhoek, M. (1996). Scheduling a repairman in a finite source system. *Mathematics and Methods of Operations Research* 44: 333–344.
6. Nash, P. & Weber, R.R. (1982). Dominant strategies in stochastic allocation and scheduling problems. In M.A.H. Dempster (ed.), *Deterministic and stochastic scheduling*. Dordrecht: D. Reidel, pp. 343–353.
7. Rabinowitz, G., Goren, S., & Mehrez, A. (2000). Scheduling two machines that require multiple types of maintenance, for a single operation. *European Journal of Operational Research* 127: 546–564.
8. Righter, R. (1996). Optimal policies for scheduling repairs and allocating heterogeneous servers. *Journal of Applied Probability* 33: 536–547.
9. Smith, D.R. (1978). Optimal repair of a series system. *Operations Research* 26: 653–662.

https://doi.org/10.1017/S0269964802163054 Published online by Cambridge University Press