1971-05

Optimal Design for System Reliability and Maintainability

Burton, Richard M.

http://hdl.handle.net/10945/41299
Optimal Design for System Reliability and Maintainability

RICHARD M. BURTON AND GILBERT T. HOWARD

Abstract—A complex system consisting of \( N \) modules that are logically interconnected for mission success is to be placed in the field for a fixed period of time. For some modules, standby units can be provided, for others this option is not available, but we must select from among several design alternatives differing in cost, weight, and reliability. The problem is to determine simultaneously the module designs and the numbers of standby units to maximize the system reliability, subject to cost and/or weight constraints. Other authors have considered a similar problem for a pure series interconnection of modules, but we permit the system to be any configuration of modules in series and/or parallel.

A dynamic programming model is presented for this problem. The notion of the generalized decomposition operator is used to develop a set of recursive relations. An example is included.

I. INTRODUCTION

The purpose of this paper is to present some generalizations to the essential spare parts problem in which it is desired to maximize, for given resources, the reliability of a system of modules connected in logical series. The decision variables are the numbers of spare parts that can be provided within each module. Black and Proschan [3] present a technique for finding the optimal solution to the spare parts problem. The generalizations presented here involve the logical structure of the system as well as the assumptions about the failure characteristics of the individual elements within the modules. The model permits various assumptions about the failure characteristics of the spare parts in a module. Redundant components may be hot (active), cold (passive), or warm. Further generalizations include treating the problem in which a module has an existing inventory of spare parts. Here, we distinguish two cases. In one case existing spare parts can be relinquished (sold) in exchange for additional resources that can then be utilized to provide spare parts elsewhere in the system. In the other case, no existing spares can be sold. Finally we discuss the questions of design, repair, and modernization of a module and demonstrate how to incorporate these considerations within the spare parts problem formulation. The system can be composed of modules in any logical series and/or logical parallel combination. A dynamic programming algorithm is used to find the optimal solution.

The literature that considers the essential spare parts problem is extensive. Black and Proschan [3] consider the problem where each component has an exponential failure probability density function (pdf), and they develop a procedure for optimizing the number of spare parts for maximum system reliability for a specific operating time at a given cost. Their procedure is optimal for any failure distribution having the monotone likelihood ratio property in the difference. The dynamic programming algorithm presented here does not have this restriction. Blitz [4] generalizes the Black and Proschan procedure to consider a nonlinear cost constraint. Proschan [12] summarizes previous work by Black and Proschan and extends the analysis to consider periodic replenishment. All of the above work assumes only one constraint, say, cost or weight. Proschan and Bray [13] develop an algorithm that optimizes the number of spare parts for multiple linear constraints.

The algorithm employed here was developed by Burton and Howard [5] for a related problem concerning optimal investment decisions for the modules in the mixed series and parallel system. Their algorithm is a generalization of the original dynamic programming algorithm developed by Bellman and Dreyfus [1] for a series-parallel system. After Bellman and Dreyfus, Kettelle [8] and Littschwager [9] also used dynamic programming to solve related problems, but the structure of the systems was restricted to the series-parallel case. Fyffe et al. [7] use dynamic programming with multiple constraints.

The methods presented in this paper are sufficiently general to consider any failure distribution function (although we shall also discuss some particular cases) and any series and parallel system configuration. The algorithm for maximizing system reliability can consider nonlinear constraints, but separability is required. This paper investigates various assumptions about the failure characteristics within the modules and delineates some cases where optimal designs can be obtained. The details of the solution method are reported in [5].

Section II gives a formal statement of the model. Section
III discusses some applications to military and civilian problems. Section IV is a discussion of how to determine the reliability functions for the modules under various assumptions about the component failures. In Section V, the dynamic programming algorithm is presented. Some generalizations are presented in Section VI, and an example in Section VII. Section VIII presents some comments on future research.

II. MODEL

We consider a system of $N$ modules, all of whose interconnections are series and/or parallel. The system is intended to perform some specified mission over a fixed time period of length $t_0$. Let $p_i(n_i)$, $i = 1, \ldots, N$ be the reliability of the system where $p_i(n_i)$ is the reliability of the $i$th module and $n_i$ is the total number of components supplied for module $i$. The reliability of two modules $i$ and $j$ in logical series is $p_i(n_i)p_j(n_j)$. The two modules in parallel have reliability $1 - (1 - p_i(n_i))(1 - p_j(n_j))$. The system reliability function $P$ reflects the appropriate composition of series and parallel connections. The m-component vector of available resources is $b$ and the applicable constraints are $(n_1, \ldots, n_N) \in K$ and $n_i \in S_i$, $i = 1, \ldots, N$, where

$$K = \left\{ n_i : \sum_{i=1}^{N} g_{ki}(n_i) \leq b_k, \quad k = 1, \ldots, m \right\}.$$

The function $g_{ki}(n_i)$ gives the quantity of the $k$th resource consumed by providing a total of $n_i$ components for module $i$. The resources might be dollars, weight, or volume, and the constraints are on total cost, weight, or volume. A quantity discount on the purchase of spare parts can be reflected by the $g_{ki}$ functions being nonlinear. Other sets $K$ are permitted, e.g., multiplicative constraints, but the set given is appropriate for the problem considered here. The set $S_i$ constrains the individual decision $n_i$ and may be simply an upper or lower bound.

The problem is to select $n_i$, $i = 1, \ldots, N$, to maximize $P$ subject to $(n_1, \ldots, n_N) \in K$ and $n_i \in S_i$, $i = 1, \ldots, N$.

III. APPLICATIONS

Here we want to illustrate the importance of this model by relating it to a variety of real problems. The illustrations are not restricted to essential spare parts examples.

The problem of determining an appropriate spare parts list for a ship, or an electronic subsystem for a ship, can be formulated as a problem of the type discussed here. A ship is to be sent on a specified mission for a fixed period of time. The equipment on board can fail in the course of the mission, and the problem is to determine the composition of replacement items to put on board so that the probability of completing the mission is maximized. Of course, these items are limited by space and by cost considerations. The essential spare parts problem represents one extreme case in that other alternatives such as predeployment repair and modernization of a given module might be done in lieu of simply including replacement parts. The problem for a spacecraft is identical except it is likely that weight will be an additional limiting factor.

The same problem arises in the purchase of specialized systems such as new aircraft and ships, both commercial and military. The purchase of a new plane frequently includes replacement items. One reasonable goal would be to select the package of spare parts to maximize the reliability of the system over its planned operating life. The constraint here is a budget limitation.

Another application arises in countries where investment items (e.g., plant and equipment) can be imported tax free but repair parts can be imported subsequently only under extreme import taxes. One method of operating is to include some spare parts with the initial investment. Again the problem is to select that mix of spare parts that will maximize the system reliability.

IV. DEVELOPMENT OF THE MODEL

In this section we discuss the form of the functions $p_i(n_i)$ under various assumptions about the performance characteristics of the modules. For the standby redundancy situation several assumptions may be made about the failure characteristics of the modules in standby. A component whose failure probability in standby is zero is called a cold standby. Those whose failure pdf's are the same in standby as in service are called hot standby components. Other cases are called warm. Normally in the warm standby case the probability of failure is less than the failure probability when in service.

For the case of cold standbys, if the exponential failure pdf is assumed, the probability that at least $k_i$-out-of-$n_i$ components will function for time $t_0$ can be shown to be (for example, Benning [2])

$$p_i(n_i) = \sum_{j=0}^{n_i-k_i} \frac{e^{-k_i t_0}}{j!} \left( k_i t_0 \right)^j.$$

A specific case of cold standbys arises when $k_i = 1$. Black and Proschan [3] deal with the series-parallel case for the exponential failure pdf and develop a procedure for selecting the optimal spare parts kit.

Although $p_i(n_i)$ cannot be obtained in closed form for a general failure pdf, a recursive computational procedure can be used in the case $k_i = 1$ to obtain an approximation to $p_i(n_i)$, which can then be used with the solution proposed in Section V. Let $g_i(t)$ be the failure pdf for the components in module $i$, and let $R_i(t, n_i)$ be the probability that $n_i$ components are sufficient to operate module $i$ for time $t$. Suppose that the components are used in the module in the order $n_1, n_2-1, n_3-2, \ldots, 1$. We have

$$R_i(t, 1) = 1 - \int_0^t g_i(\tau) d\tau \quad (1)$$

and
\[ R_i(t, m) = R_i(t, 1) + \int_0^t R_i(t - \tau, m - 1)g(\tau) \, d\tau, \]

where \( 2 \leq m \leq n_i \). (2)

The recursive equation (2) gives the probability that \( m \) components are sufficient for module \( i \) at time \( t \). It is expressed as the probability that the \( m \)th component alone will function until time \( t \) plus the probability that the \( m \)th component will fail at \( \tau \) and the remaining \( m - 1 \) components will function from \( \tau \) until \( t \). A discrete approximation to (2) can be used to compute recursively, for each \( m \), an approximation to \( R_i(t, m) \), which provides a point on the \( p_i(n_i) \) function.

The hot standby case is easier to deal with than either the cold or the warm case. If we let \( p \) be the probability that each component in module \( i \) functions for time \( t \), then the probability that at least \( k_i \) out of the \( n_i \) identical components function for time \( t \) is given by the binomial distribution (for example, Chan [6])

\[ p_i(n_i) = \sum_{k_i=0}^{n_i} \binom{n_i}{k_i} p^{k_i} (1 - p)^{n_i - k_i}. \]

The case \( k_i = 1 \) leads to the familiar expression for the reliability of a parallel connection of \( n_i \) components

\[ p_i(n_i) = 1 - (1 - p)^{n_i}. \]

(Another possibility for the hot standby case is that in which any one component is sufficient for module operations but the various components are different. Burton and Howard [5] give an illustration of how to develop the reliability function in this case.)

For the warm standby case Polovko [11] shows that if the component failures are exponential with parameter \( \lambda_i \) and \( \bar{\lambda}_0 \) in standby and \( \bar{\lambda}_0 \) in service, the system reliability to time \( t_0 \) is given by

\[ p_i(n_i) = e^{-\bar{\lambda}_0 t_0} \left[ 1 + \sum_{k=1}^{n_i-1} \frac{a_k}{k!} (1 - e^{-\bar{\lambda}_0 t_0})^k \right], \]

where

\[ a_k = \prod_{j=0}^{k-1} \left( 1 + \frac{\bar{\lambda}_0}{\bar{\lambda}_i} \right). \]

A recursive relationship that may be used to compute \( p_i(n_i) \) for a general failure pdf is also given in Polovko [11].

V. Solution Procedure

For the case where there are no connections between the modules other than series and parallel connections, \( P \) can be rewritten as \( P = p_i O_{n_i-1} O_{n_i-2} \cdots O_1 P_i \) where \( O_i \) is the composition operator (Nemhauser [10] for further discussion) determining the way in which modules \( i \), \( i - 1 \), \( i - 2 \), \( i \) are logically connected to module \( i + 1 \). In the all-series case \( p_{i+1} O_i P_i \cdots O_1 P_1 = p_{i+1} (p_{i+1} O_i P_i \cdots O_1 P_1) \) and in the all-parallel case \( p_{i+1} O_i P_i \cdots O_1 P_1 = 1 - (1 - p_{i+1}) \).

\( (1 - p_{i+1} \cdots O_1 P_1) \). In other cases a renumbering of modules may be required. More detail is given in [5].

This model can be solved by dynamic programming using the recursive relationships

\[ f_i(X_i) = \max_{n_i \in N_i X_i} \{ p_i(n_i) \} \]

where \( X_n = b \) is the vector of initial resources, and the stage transformation is \( X_i-1 = X_i - g_i(n_i), i = 1, \ldots, N \). The vector \( g_i(n_i) = (g_{i1}(n_i), \ldots, g_{im}(n_i)) \) gives the quantity of each of the \( m \) resources consumed at stage \( i \) by using a total of \( n_i \) components for module \( i \).

This dynamic programming model provides a computationally feasible means of dealing with realistic spare parts problems. The appropriate \( p_i(n_i) \) functions are first determined for each module and then for computational purposes a discrete approximation is made to each function.

The maximum system reliability subject to the resource constraints is computed recursively beginning with \( f_1(X_1) \), the maximum reliability that can be obtained from module one with resources \( X_i \) available. The solution obtained is globally optimal (to the discrete approximation) regardless of the form of \( p_i(n_i) \). Computational considerations restrict \( m \), the number of resources, to be small, say three or less. The number of computations rises exponentially with \( m \) but only linearly with \( N \) so that a very large number of modules can be considered in a single problem.

VI. Generalizations

We consider in this section some modifications of the original problem that can be easily dealt with.

A problem arising in system maintenance is that of determining the optimal package of spare parts to purchase with fixed resources, given that some spare parts are already
TABLE I
DATA FOR EXAMPLE

<table>
<thead>
<tr>
<th>Module</th>
<th>Cost per Module</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Cold standbys, exponential failures, ( \lambda_1 = 0.00501 ), ( S_1 = { n_1 : 0 \leq n_1 \leq 2 } )</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>Modernization and repair ( r_2(0) = 0.5 ), ( r_2(5) = 0.8 ), ( r_2(10) = 0.84 ), ( r_2(15) = 0.95 ), ( r_2(20) = 0.995 )</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2 out of ( n ), hot standbys, probability that one component operates for the entire time is 0.92.</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>Warm standbys, exponential failure ( 2 \lambda_2 = 0.10536 ) in service, ( 2 \lambda_2 = 0.01005 ) in standby.</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3 out of ( n ), cold standbys, exponential failures, ( \lambda_3 = 0.05129 ).</td>
</tr>
</tbody>
</table>

available, but the existing configuration is not optimal for the level of resource expenditure it represents, nor are the resources available sufficient to bring the system to an optimal condition. We consider two cases. In one, resale of existing spares is permitted; in the other, it is not.

If resale is not permitted, the problem can be formulated as the following. Maximize

\[
\{p_1(m_1 + n_1), p_2(m_2 + n_2), \ldots, p_N(m_N + n_N)\}
\]

with the constraints

\[
\sum_{i=1}^{N} g_i(n_i) \leq b
\]

\[n_i \geq 0, \quad i = 1, \ldots, N\]

where

\( m_i \) number of existing components of type \( i \),

\( n_i \) number of additional components to be provided to module \( i \),

\( g_i(n_i) \) cost for \( n_i \) components in module \( i \),

\( b \) budget available.

If resale of existing components is permitted at a price of \( h_i(q_i) \) for \( q_i \) units sold, the problem becomes the following. Maximize

\[
P(p_1(n_1 + m_1 - q_1), \ldots, p_N(n_N + m_N - q_N))
\]

with the constraints

\[
\sum_{i=1}^{N} g_i(n_i) \leq b + \sum_{i=1}^{N} h_i(q_i)
\]

\[0 \leq q_i \leq m_i
\]

\[n_i \geq 0, \quad i = 1, \ldots, N.\]

Another generalization deals with the possibility of system repair or modernization rather than providing redundancy to increase reliability. Redundancy may be impractical and predeployment servicing or replacement may be the only option available. In this case it is more convenient to let the decision variable be the resource expenditure \( d_i \), and let \( r_i(d_i) \) represent the reliability of the module, given an expenditure of \( d_i \) units of the resources, possibly repair time and money. The function \( r_i(d_i) \) may represent a module that has nonzero reliability even with no resource expenditure. Other possibilities are permitted; in fact, \( r_i(d_i) \) can be any real-valued function.

The possibility of replacing the entire module with new equipment of the same or different design can be considered by including the reliability of the new equipment as a point on the \( r_i(d_i) \) function. Another possibility is to expend the resources in a research and development effort to produce a better module. All the options available can be considered in obtaining \( r_i(d_i) \). We let

\[
r_i(d_i) \equiv \max \{ r_{i1}(d_i), r_{i2}(d_i), \ldots \}
\]

where \( r_{ij}(d_i) \) is the reliability that can be obtained for module \( i \) if alternative \( j \) is followed and \( d_i \) units of resource are used. See Fig. 1 where \( p_i(d_i) \) is indicated by the dashed line.

VII. Example

A hypothetical example is solved to illustrate the generality of the procedure. Fig. 2 shows the system configuration considered in the example. Table I summarizes the failure characteristics and the cost information for each of the modules. Cold, hot, and warm standbys are considered in modules 1, 3, and 4, respectively. Modules 3 and 5 require that the number of functioning components be \( \geq k_i \) where \( k_3 = 2 \) and \( k_5 = 3 \). For module 2 modernization, replacement and repair options are available as discussed in the previous section. Only two modules of type 1 are available.

This example actually combines one of the generalizations discovered in Section VI with the basic model. For the second module the decision variable is the (discrete) level of investment \( d_2 \) while for the other modules the decision is the number of components to provide.

A budget of 40 was used for the example and the mission time was taken to be 1. The optimal solution is \( n_1^0 = 1 \), \( d_2^0 = 0 \), \( n_2^0 = 3 \), \( n_3^0 = 5 \), \( n_4^0 = 6 \), and the system reliability is 0.99. A module of type 1 is used although no allocation is made to module 2 whose reliability is 0.5 with no expenditure. The availability of only two modules of type 1 does not influence the solution. The budget of 40 is used completely.

VIII. Future Research

Some interesting areas of future research in optimal system design include the investigation of more general systems effectiveness criteria to reflect more than just the system reliability, consideration of more general system structures, and the investigation of system repair during mission time. The \( k \)-out-of-\( n \) module considers common spares for \( k \) components in series, but the allocation problem when common spares are permitted for nonseries components appears to be more difficult.

For the results of such research to be usable in large
systems, emphasis must be placed not only on theoretical developments but on computational methods as well.

REFERENCES


Available of Priority Standby Redundant Systems

J. A. BUZACOTT

Abstract—A priority standby system consisting of two repairable units is considered. One unit, the priority unit, is always in service except when it is failed. The standby unit is in service only for the duration of repair of the priority unit. Expressions are derived for the availability of such a system for both preemptive and nonpreemptive repair. The results assume reasonably general failure-time and repair-time distributions of the priority and standby units. The preemptive priority results are relatively insensitive to the form of the distributions.

Reader Aids:
Purpose: Widen state of the art
Special math needed for explanations: Renewal theory
Special math needed for results: Renewal theory
Results useful to: System designers and analysts

INTRODUCTION

In a recent note Osaki [1] considered a redundant system consisting of a repairable priority unit and a standby unit. Usually, the priority unit performs the desired system function. Failure of the priority unit the standby unit is switched into service. Meanwhile the priority unit is repaired. As soon as repair of the priority unit is completed it is switched back into service. While the standby unit is not in service it is assumed that it cannot fail, i.e., the case of “cold” standby is being considered.

Priority standby can be contrasted with usual standby with repair, which has been analyzed by Gnedenko [2] and Srinivasan [3]. In usual standby, switchover from one unit to the other does not occur until the operating unit fails, irrespective of which unit is operating. Gnedenko and Srinivasan used simple renewal theory arguments and obtained Laplace–Stieljes transforms of the time between successive switchovers. From these transforms the availability and mean time between failures of the system can be derived.

Osaki considered a priority system where the priority unit has general failure-time and repair-time distributions and the standby unit had an exponential failure-time distribution. He derived expressions for the Laplace–Stieljes transform of the time to first system failure, i.e., the time from an instant when both units are in the as-new condition to the first instant when the standby unit fails while the priority unit is under repair.

However, Osaki’s results are not sufficient to determine the availability of such a repairable priority system (except in the trivial case where the standby unit has a zero repair duration). The purpose of this paper is to show how a renewal theory approach used elsewhere by the author for the special case of production systems [4] can be generalized to give formulas for the availability of repairable priority standby systems. Both the priority and standby units are considered to have general but well-behaved failure-time and repair-time distributions.

Because the standby unit is repairable it is necessary to