Risk-Sensitive Decision-Theoretic Diagnosis

Mark A. Shayman [1]
Department of Electrical and Computer Engineering and
Institute for Systems Research
University of Maryland
College Park, MD 20742
Email: shayman@eng.umd.edu
Fax: (301) 314-9281

Emmanuel Fernandez-Gaucherand [2]
Department of Electrical and Computer Engineering
and Computer Science
University of Cincinnati
Cincinnati, OH 45221
Email: emmanuel@ececs.uc.edu

May 1, 2000; revised January 23, 2001

[1]Research supported in part by NSF grant ECS-9626399 and the Laboratory for Telecommunications Science under DoD contract MDA90497C3015

[2]Research supported in part by NSF grant NSF-INT 9602939 and by United Engineering Foundation grant 00/ER-09

Abstract

We consider the problem of determining the optimal sequence of tests for the discovery of a faulty component, where there is a random cost associated with testing a component. Our work is motivated by applications in telecommunications networks, e.g., location and isolation of faults (or intruders) in IP networks. A novel feature in our approach is that a risk-sensitive performance criterion is used in order to rank different competing schedules. Risk-sensitivity is incorporated through the use of an exponential utility function, and hence optimal schedules attain a trade-off between minimal expected costs and, e.g., a low variance about the achievable expected costs. We characterize optimal schedules both when the testing sequence is not subject to precedence constraints, and when it is subject to such constraints, given by an arbitrary partial order. For the case with precedence constraints, we show that our models can be analyzed via modular decompositions, as studied by Monma and Sidney,

1 Introduction

The motivation for the work presented here comes from the problem of fault management for communication networks. An important element in many approaches to fault management is *sequential testing* [19]. Based on available network management data, a set of components (hardware or software) is identified as containing the potential root cause of the failure. Then the suspect components are tested sequentially until the defective component is identified. For the resulting scheduling problem, it is typically assumed that that there is a single faulty component (the *mutually exclusive faults* (MEF) case), that the probability of component i being faulty is a known value p_i, and that there is a random cost C_i associated with testing it, and the goal is to minimize the expected sum of the testing costs. Under these assumptions, classical results apply and indicate that it is optimal to test in order of increasing ratios $E[C_i]/p_i$. This is sometimes referred to as the "C over p rule." There is a large literature on this problem and its extension to the case where there are precedence constraints on the testing sequence. See e.g., [5, 35, 15, 26, 7, 11, 17, 1, 32, 36, 16, 24]. Analogous results are available on the problem in which the assumption of mutually exclusive faults is replaced by the assumption of independent faults, and a sequence of components are tested until the first faulty component is discovered at which time testing stops. This problem is referred to as the *independent faults* (INF) problem. See e.g., [5, 27, 10, 22, 21, 13, 34, 23, 28]. The "C over p" rule has been applied in network fault management in, e.g., [19, 3]. In the diagnosis problems we consider, a test either identifies a faulty component or eliminates it from suspicion. Diagnosis problems in which tests reveal only partial information concerning faults are considered in [8].

In the above approaches, the objective is to minimize the average sum of the testing costs. This may make sense for diagnostic problems that will be repeated many times under the same conditions–i.e., with the same model–such as the diagnosis of engine failures in a particular brand of motorcycle considered in [24]. This is unlikely to be the case for the diagnosis of communication network faults. We present three arguments in support of the latter statement.

1. When a network anomaly is detected, various data about the status and performance of network elements is stored in the MIB's (Management Information Bases) and is available to the network management system. Using this information to instantiate the appropriate nodes in a Bayesian network, evidence propagation algorithms for Bayesian networks may be used to determine the p_i's [19, 20]. Since the performance data are constantly changing and the network configuration changes quite frequently, even if a fault in a particular component is repeated, it is likely that the p_i's will be different. Thus, an objective of minimizing average cost over many repetitions of the same fault scenario may not be appropriate.

2. Even apart from the above, the network manager may prefer to accept some degree of increased expected cost in order to reduce the variance of the cost; i.e., the manager may be *risk-averse*. The manager is therefore willing to perhaps pay a premium (in average), in order to avoid the uncertainty that, for normal operational ranges, the actual costs (e.g., testing time) incurred in repeated fault

management situations may vary widely (perhaps for the worse) from the predicted average optimal values.

3. Policies determined by optimizing a risk-sensitive criterion may perform better even relative to a risk-neutral criterion when parameter errors are present [12, 4].

The above concerns lead us to believe that the problem of optimal sequential fault diagnosis is ideally suited to analysis as a stochastic decision process with a *risk-sensitive optimality criterion*. This is achieved via the use of an exponential utility function, leading to an optimization process that yields risk (i.e., variance) sensitive scheduling rules; see [2], [6], [18], [25], [29], [38]. As mentioned previously, it is this choice of optimality criteria that differentiates our work from the available literature.

We therefore assume that risk sensitivity is given by an exponential (dis-) utility function $U_\gamma(x) = (\mathrm{sgn}\gamma)e^{\gamma x},\ \gamma \in \Re,\ \gamma \neq 0$. The above utility function can be explained as modeling the behavior of a "decision maker" with a constant level of risk aversion, as given by the parameter γ [30], [25]. This can be a good approximation over a range of operation, and furthermore leads to a decomposable (multiplicative) "exponential of sum of costs" criterion amenable to analysis via dynamic programming [2], [9], [6], [25], [38]. Another viewpoint is just to consider a Taylor expansion of $E[e^{\gamma \mathcal{C}}]$ about $E[\mathcal{C}]$, where $\mathcal{C}$ represents the sum of aggregated random costs. Up to a first approximation, the latter leads to an objective for optimization composed of the sum of expected costs (the standard or *risk-neutral* criterion) and the *variance of* $\mathcal{C}$ multiplied by the risk sensitivity parameter γ. Hence, if $\gamma > 0$ the decision maker is said to be *risk-averse* in that the objective for the optimization seeks to minimize both variance (uncertainty measuring risks involved) and the expected costs. If $\gamma < 0$, then variance is actually seen as a desirable feature, and the decision maker is said to be *risk-seeking*. Finally, if $\gamma = 0$, the standard expected costs criterion is recovered, which cannot distinguish between two scheduling rules with different variances but equal expected costs, a risk insensitive or *risk-neutral* situation.

In the sequel, we are able to obtain explicit characterization of the optimal testing sequence in the unconstrained case. In the more general case where the sequence is constrained by an arbitrary partial order, we show that our model leads to modular decompositions of the optimality criterion, and therefore is amenable to study following work by Monma and Sidney [28]. At the root of our results is an interchange argument. That such an argument proves useful is somewhat surprising since scheduling problems that can be solved by interchange arguments in the risk-neutral case are not necessarily amenable to such solution in the case of risk-sensitive criterion [2].

An abridged description of this work was presented in [31].

2 Mutually Exclusive Faults Without Precedence Constraints

We view the fault management process as a hierarchical process, consisting of two phases. In the first phase, the agent searches its domain for evidence of anomalous behavior. If evidence of such behavior is detected in a subdomain, a second phase begins in which the network element(s) in the subdomain are tested to isolate the specific fault. The phase one problem corresponds to the following generic search problem: Given a set of n components which fail *independently*, each of which may be tested with random testing cost to determine whether or not it is faulty, determine the "optimal" order in which the components should be tested. We further assume that once a faulty component is found, the search stops; this has been called a *satisficing search* by Simon and Kadane [34]. We refer to this as the *independent fault* (INF) problem. The phase two problem corresponds to a different generic search problem: Given a set of n components *with at most one in failure*, determine the "optimal" order in which the components should be tested. We refer to this as the *mutually exclusive fault* (MEF) problem.

We begin by considering the problem of risk-sensitive sequential diagnosis for the MEF problem without precedence constraints, and subsequently we will briefly comment on the INF problem, the solution of which follows similar arguments as the former. Let C_i be a random variable representing the cost of testing component i given that it is faulty, and let D_i be a random variable representing the cost of testing component i given that it is not faulty. We assume that conditional on the state of component i, the testing cost of component i is independent of the states and testing costs of all other components. We allow for the possibility that the true fault may not be one of the n hypothesized faults; thus $\sum_{i=1}^{n} p_i$ may be less than 1.

What we refer to as a "component" may itself be a complex entity or even a subsystem. Thus, what we refer to as a "test" of a particular component may itself refer to a *sequence* of tests. If there is a fault in the component, the sequence stops when the fault is discovered. If there is no fault, the entire test sequence is performed before the component is exonerated from suspicion. Thus, for such a diagnostic process the average cost $E[D_i]$ associated with a negative result will generally be larger than the average cost $E[C_i]$ associated with a positive result. This observation also applies to other potential applications such as searching for a hostile entity in a collection of geographic regions. To exonerate a region (negative test result) requires a complete search of the region, while finding the hostile entity within the region (positive test result) may only require a partial search.

We may also choose to include in C_i the cost associated with repairing the faulty component. In the risk-neutral case, the contribution of the repair cost to the objective function is independent of the ordering and hence has no effect on the optimal schedule. However, we show in the sequel that this is not true in the risk-sensitive case. If repair costs are included, obviously it is possible that $E[C_i]$ may exceed $E[D_i]$.

In other applications, a component subsystem may itself be structured hierarchically as a tree. The leaves represent the 'atomic' elements of the subsystem. The root node represents the entire subsystem, while an interior node represents that part of the subsystem comprised of all its descendant leaves. It may be possible to test the portion of the subsystem corresponding to a particular node. A negative test result indicates that this portion is fault-free and completes the investigation of that portion of the system. On the other hand, a positive test result indicates that the portion contains a fault; in this case, the portion is further tested to localize the fault by proceeding in a *depth-first* manner. (This type of algorithm occurs in a number of different applications; in multiple-access communications, it is sometimes referred to as the Adaptive Tree Walk Protocol [37].) In this approach, the average cost $E[D_i]$ associated with a negative result will generally be smaller than the average cost $E[C_i]$ associated with a positive result. The exponential utility function used in our risk-sensitive optimality criterion also has some relation to scheduling problems which involve minimizing completion times when (exponential) discounting is employed, as illustrated in the following remark.

Remark 1 In the special case where $C_i = D_i$, $\forall i$ and $\gamma < 0$ (risk-seeking decision maker), the MEF problem without precedence constraints is mathematically equivalent to the total weighted exponential completion time problem [14, 28]. To obtain this equivalence, p_i and C_i are redefined to be the weight and processing time, respectively, for job i.

Let $\psi(i;\gamma) := E[e^{\gamma C_i}]$, and let $\overline{\psi}(i;\gamma) := E[e^{\gamma D_i}]$. Thus, $\psi(i;\gamma)$ and $\overline{\psi}(i;\gamma)$ are the moment generating functions of C_i and D_i, respectively. In the sequel, we generally suppress the dependence on γ.

Let $a = (a_1, \ldots, a_m)$ be a permutation of an m-element subset, denoted $\{a\}$, of $\{1, \ldots, n\}$. We refer to a as a schedule and indicate the number of elements in the schedule by $|a|$. If $|a| = n$, we refer to a as a complete schedule. For two (disjoint) schedules $a = (a_1, \ldots, a_m)$ and $b = (b_1, \ldots, b_\ell)$ such that $m + \ell \leq n$, we denote the concatenated schedule as

$$ab := (a_1, \ldots, a_m, b_1, \ldots, b_\ell).$$

To simplify notation, we will generally assume that $\gamma > 0$–i.e., the decision maker is risk-averse. However, we will indicate the analogous results for the case $\gamma < 0$ as well. Let F_i denote the event that component i is faulty. Denote the random cost as

$$\mathcal{C}_i = I_{F_i} C_i + (1 - I_{F_i}) D_i,$$

and for $a = (a_1, \ldots, a_m)$, the utility of the aggregated cost is computed as

$$\mathcal{C}(a) = exp\{\gamma \sum_{i=1}^{m} \mathcal{C}_{a_i}\}. \tag{1}$$

In the event that $I_{F_{a_i}} = 1$, i.e., the $a_i th$ component is faulty, then

$$E[\mathcal{C}(a) | I_{F_{a_i}} = 1] = \psi(a_i) \prod_{j=1}^{i-1} \overline{\psi}(a_j), \tag{2}$$

since testing would terminate once the fault is found. Hence, for $|a| = m > 1$, the expected exponential risk-sensitive cost $V(a)$ for the schedule a is given by

$$V(a) = \sum_{i=1}^{m} p_{a_i} \psi(a_i) \prod_{j=1}^{i-1} \overline{\psi}(a_j) + (1 - \sum_{i=1}^{m} p_{a_i}) \prod_{i=1}^{m} \overline{\psi}(a_i), \tag{3}$$

where the first term comes from (2) and the independence assumptions, and the second term accounts for the probability that none of the $|a| = m$ tested components are faulty. In the special case where $|a| = 1$ and $a = (i)$, we have

$$V(i) = p_i \psi(i) + (1 - p_i) \overline{\psi}(i). \tag{4}$$

If we want to emphasize the dependence on the risk-sensitivity parameter γ, we will use the notation $V(a;\gamma)$ instead of $V(a)$.

The following result plays a crucial role in the sequel.

Lemma 1 *Suppose that $\{a\} = \{b\}$. Then $V(a) \leq V(b)$ if and only if*

$$\sum_{i=1}^{m} p_{a_i} \psi(a_i) \prod_{j=1}^{i-1} \overline{\psi}(a_j) \leq \sum_{i=1}^{m} p_{b_i} \psi(b_i) \prod_{j=1}^{i-1} \overline{\psi}(b_j). \tag{5}$$

Proof: The result follows directly from (3). Let $m = |a| = |b|$. Since $\{a\} = \{b\}$,

$$(1 - \sum_{i=1}^{m} p_{a_i}) \prod_{i=1}^{m} \overline{\psi}(a_i) = (1 - \sum_{i=1}^{m} p_{b_i}) \prod_{i=1}^{m} \overline{\psi}(b_i). \tag{6}$$

■

Let

$$\hat{V}(a) := \sum_{i=1}^{m} p_{a_i} \psi(a_i) \prod_{j=1}^{i-1} \overline{\psi}(a_j). \tag{7}$$

As a consequence of Lemma 1, when comparing schedules containing the same set of elements, it is equivalent to base the comparison on the function $\hat{V}$ rather than on V. This greatly simplifies the analysis.

Let $\overline{\psi}(a) := \prod_{i=1}^{|a|} \overline{\psi}(a_i)$. It follows easily from the definitions that

$$\overline{\psi}(ab) = \overline{\psi}(a)\overline{\psi}(b), \tag{8}$$

$$\hat{V}(ab) = \hat{V}(a) + \overline{\psi}(a)\hat{V}(b). \tag{9}$$

The following proposition contains the key interchange argument used subsequently to obtain the optimal scheduling rules.

Proposition 1 *If $|i| = |j| = 1$, then*

$$\hat{V}(aijd) \leq \hat{V}(ajid) \Leftrightarrow \frac{\overline{\psi}(i) - 1}{p_i \psi(i)} \leq \frac{\overline{\psi}(j) - 1}{p_j \psi(j)}. \tag{10}$$

Proof: Iterating (9), we obtain

$$\hat{V}(abcd) = \hat{V}(a) + \overline{\psi}(a)\hat{V}(b) + \overline{\psi}(a)\overline{\psi}(b)\hat{V}(c) + \overline{\psi}(a)\overline{\psi}(b)\overline{\psi}(c)\hat{V}(d). \tag{11}$$

Hence, using (8) and (11), we obtain

$$\hat{V}(abcd) - \hat{V}(acbd) = \overline{\psi}(a)\{\hat{V}(c)[\overline{\psi}(b) - 1] - \hat{V}(b)[\overline{\psi}(c) - 1]\}, \tag{12}$$

and thus

$$\hat{V}(abcd) \leq \hat{V}(acbd) \quad \Leftrightarrow \quad \frac{\hat{V}(c)}{\overline{\psi}(c) - 1} \leq \frac{\hat{V}(b)}{\overline{\psi}(b) - 1} \quad \Leftrightarrow \quad \frac{\overline{\psi}(b) - 1}{\hat{V}(b)} \leq \frac{\overline{\psi}(c) - 1}{\hat{V}(c)},$$

from which the result follows with $b = i$ and $c = j$. ■

Theorem 1 *Let $\gamma \neq 0$. For the MEF problem with no precedence constraints, a complete schedule $t = (t_1, \ldots, t_n)$ is optimal if and only if*

$$\frac{\text{sgn}(\gamma)(\overline{\psi}(t_i) - 1)}{p_{t_i} \psi(t_i)} \leq \frac{\text{sgn}(\gamma)(\overline{\psi}(t_j) - 1)}{p_{t_j} \psi(t_j)}, \tag{13}$$

whenever $i < j$.

Proof: For the case $\gamma > 0$, the result follows immediately by applying an interchange argument, based on Proposition 1. The analogous result for the case $\gamma < 0$ is derived similarly. Note that if $\gamma < 0$, then $\psi(i) < 1$ and $\overline{\psi}(i) < 1$. ■

Thus, from Theorem 1, we conclude that the components should be tested in increasing order of the ratios

$$\frac{\operatorname{sgn}(\gamma)(E[e^{\gamma D_i}] - 1)}{p_i E[e^{\gamma C_i}]}.$$

Remark 2 If a component is tested and found *not faulty*, this information apparently could be used in a closed-loop decision rule. For the MEF problem, the exoneration of a particular component causes the probabilities p_i of the remaining components to be rescaled by a constant factor–i.e., renormalized. Since this renormalization leaves the ordering of the ratios in (13) unchanged, it follows that the open-loop schedule specified in Theorem 1 indeed gives the optimal policy.

The following result shows that the classical ratio test for the risk-neutral problem is recovered when the absolute value of the risk-sensitivity parameter is small.

Corollary 1 *Suppose that the ratios $\{E[D_i]/p_i\}_{i=1}^n$ are distinct. Then there exists $\epsilon > 0$ such that if $|\gamma| < \epsilon$ then there is a unique optimal complete schedule $t = (t_1, \ldots, t_n)$ for the MEF problem, and it is specified by the condition that whenever $i < j$, then*

$$\frac{E[D_{t_i}]}{p_{t_i}} < \frac{E[D_{t_j}]}{p_{t_j}}. \tag{14}$$

Proof: For small $|\gamma|$, $\psi(i) \approx 1 + \gamma E[C_i]$ and $\overline{\psi}(i) \approx 1 + \gamma E[D_i]$. Using these approximations, the assertion follows from Theorem 1. ■

Remark 3 From Corollary 1 we see that if the risk-sensitivity parameter is sufficiently small, the costs $\{C_i\}$ associated with *positive* tests are irrelevant to the determination of the optimal testing order. In particular this is the case in the risk-neutral problem–i.e., in the limit as $\gamma \to 0$. However, from Theorem 1, it follows that for the general risk-sensitive problem the costs associated with positive tests as well as costs associated with negative tests are relevant to the optimal testing order, a rather important distinction with respect to the risk-neutral case, i.e., the "C over p rule". Note from (13) that the larger C_i, the higher priority for component i (for $\gamma > 0$), i.e., it may be too risky to delay testing of such a component. In particular this indicates that a risk-averse decision maker should test components that have high repair or replacement costs early.

Next, we consider the structure of the optimal schedule for $|\gamma| >> 0$. For this discussion, we make the simplifying assumption that for each i, the random variables C_i and D_i are constants with values c_i and d_i, respectively.

Corollary 2 *Suppose that $C_i = c_i$ and $D_i = d_i$, $\forall i$. Suppose that the complete schedule $t = (t_1, \ldots, t_n)$ is such that if $i < j$, then one of the following conditions holds:*

(a) $d_{t_i} - c_{t_i} < d_{t_j} - c_{t_j}$.

(b) $d_{t_i} - c_{t_i} = d_{t_j} - c_{t_j}$ *and* $p_{t_j} < p_{t_i}$.

(c) $d_{t_i} - c_{t_i} = d_{t_j} - c_{t_j}$ *and* $p_{t_j} = p_{t_i}$ *and* $c_{t_i} < c_{t_j}$.

Then for all sufficiently large $\gamma > 0$, t is the unique optimal complete schedule for the MEF problem.

Proof: From Proposition 1 it follows that $\hat{V}(aijb) < \hat{V}(ajib)$ if and only if

$$p_j e^{\gamma c_j}(e^{\gamma d_i} - 1) < p_i e^{\gamma c_i}(e^{\gamma d_j} - 1). \tag{15}$$

By considering (15) for $\gamma >> 0$ and using an interchange argument, the result follows. ■

In particular, this result shows that if the differences $\{d_i - c_i\}$ are distinct, then a very risk-averse decision maker should order the tests so that the differences between the cost of a negative test and positive test are increasing. On the other hand, if there are two tests for which these differences are equal, the test with higher probability of success (i.e., testing the component with higher failure rate) should be scheduled first. If both the differences and probabilities are equal, the test with the smaller cost should be scheduled first. In the special case in which the cost of negative tests is the same as that of positive tests, and in which the success probabilities are all distinct, the tests should be scheduled in order of decreasing success probability. Therefore, hypersensitivity to risk induces myopia in the decision maker's behavior.

Corollary 3 *Suppose that $C_i = c_i$ and $D_i = d_i$, $\forall i$. Suppose that the complete schedule $t = (t_1, \ldots, t_n)$ is such that if $i < j$, then one of the following conditions holds:*

(a) $c_{t_i} < c_{t_j}$.

(b) $c_{t_i} = c_{t_j}$ *and* $p_{t_j} < p_{t_i}$.

(c) $c_{t_i} = c_{t_j}$ *and* $p_{t_j} = p_{t_i}$ *and* $d_{t_i} < d_{t_j}$.

Then for all $\gamma < 0$ with $|\gamma|$ sufficiently large, t is the unique optimal complete schedule for the MEF problem.

Proof: Follows easily from Proposition 1 and Theorem 1, e.g., reversing the inequality in (15). ■

Example 1 Suppose there are three components with fault probabilities $p_1 = 0.3$, $p_2 = 0.4$, $p_3 = 0.2$. Note that there remains a probability of 0.1 that the failure is not in any of the three suspected components. Suppose that the costs of negative tests are $D_1 = 2$, $D_2 = 3$, $D_3 = 2$ and the costs of positive tests are $C_1 = 1$, $C_2 = 2$, $C_3 = 3$. Thus negative tests are more costly than positive tests for components 1 and 2, but the opposite is true for component 3. Thus, it is more costly to exonerate components 1 and 2 than it is to confirm failure and repair them; the opposite is true for component 3. For example, component 3 might have a substantial repair cost.

Since $D_1/p_1 = 6.67$, $D_2/p_2 = 7.50$, $D_3/p_3 = 10.00$, it follows from Corollary 1 that for small $|\gamma|$, the optimal schedule is $(1, 2, 3)$. In particular this is the case in the limit as $\gamma \to 0$–i.e., the risk-neutral case. Note that the costs associated with positive tests, and hence repair costs, have no bearing on this conclusion.

Since, $c_1 < c_2 < c_3$, it follows from Corollary 3 that for sufficiently negative values of γ–i.e., for a sufficiently risk-seeking decision maker–the optimal schedule is $(1, 2, 3)$. In fact, numerical investigation confirms that this is the optimal schedule for all $\gamma \leq 0$.

Since $d_3 - c_3 < d_1 - c_1 = d_2 - c_2$ and $p_1 < p_2$, it follows from Corollary 2 that for sufficiently positive values of γ–i.e., for a sufficiently risk-averse decision maker–the optimal schedule is $(3, 2, 1)$. In fact, numerical investigation based on Theorem 1 reveals that the optimal schedule is $(1, 2, 3)$ for $-\infty < \gamma \leq 0.1869$, $(1, 3, 2)$ for $[0.1870, 0.2027]$, $(3, 1, 2)$ for $[0.2028, 0.2644]$, and $(3, 2, 1)$ for $[0.2645, \infty)$.

Remark 4 Suppose that instead of assuming mutually exclusive faults, we assume that the states of the n components (faulty or not faulty) are independent: the *independent faults* (INF) problem. Similarly as the comments in Remark 2, if a component is tested and found *not faulty*, this information apparently could be used in a closed-loop decision rule. However, for the INF problem, this type of gathered information does not change the individual probabilities for the remaining components; thus an open-loop schedule is optimal and an interchange argument is applicable. Let $q_i = 1 - p_i$, the probability that component i is not faulty. Using a derivation similar to that for Theorem 1 gives the following.

Theorem 2 *Let $\gamma \neq 0$. For the INF problem with no precedence constraints, a complete schedule $t = (t_1, \ldots, t_n)$ is optimal iff whenever $i < j$, then*

$$\frac{\operatorname{sgn}(\gamma)(q_{t_i}\overline{\psi}(t_i) - 1)}{p_{t_i}\psi(t_i)} \leq \frac{\operatorname{sgn}(\gamma)(q_{t_j}\overline{\psi}(t_j) - 1)}{p_{t_j}\psi(t_j)}. \tag{16}$$

From Theorem 2, analogues of Corollaries 1 and 2 are easily obtained.

3 Mutually Exclusive Faults With Precedence Constraints

We extend the results in Section 2 to include the situation when there may be precedence constraints among different tests, expressed as a partial order on the set of possible tests. By virtue of Lemma 1 and the functional equations (8) and (9), we are able to show a decomposition result for our performance objective, and hence we demonstrate that the modular decomposition theory of Monma and Sidney [28] can be applied to our problem. Decomposition results for various scheduling problems under precedence constraints have been described by a number of authors, including (among others) Sidney [32, 33], Kadane and Simon [23], and Glazebrook and Gittens [14].

Firstly, we recall the relevant definitions and results from [28]. Let $S = \{1, \ldots, n\}$ be a set of jobs to be sequenced. Let f be a *cost function* that assigns a cost $f(u)$ to the schedule u. The precedence relation constraints are represented by a (directed) *precedence graph* $G = (S, R)$, where the nodes in S correspond to the jobs and an arc $(i, j) \in R$ corresponds to the precedence constraint, denoted $i \to j$, that job i must precede job j. A complete schedule t is *feasible* if it is consistent with R; it is *optimal* if it minimizes $f(\cdot)$ over the set of all feasible complete schedules. Precedence subgraphs $G_1 = (S_1, R_1)$ and $G_2 = (S_2, R_2)$, with S_1, S_2 disjoint, are *in parallel* if whenever $i \in G_1$ and $j \in G_2$, then $(i, j), (j, i) \notin R$. An *initial* (*terminal*) subset of (S, R) is a subset $S' \subseteq S$ such that if $i \in S'$, then every predecessor (successor) of i is also in S'.

The following definitions extend the interchange argument approach used in the previous section to the context in [28]. The sequencing function f is said to have the *strong adjacent sequence interchange (strong ASI) property* if there exists a *transitive* "preference" relation $\preceq$ defined on all pairs of schedules such that

$$b \preceq c \Leftrightarrow f(abcd) \leq f(acbd), \ \forall a, d. \tag{17}$$

The sequencing function f is said to have the *strong series network decomposition (strong SND) property* if whenever $\{b\} = \{c\}$, then

$$f(b) \leq f(c) \Rightarrow f(abd) \leq f(acd), \ \forall a, d. \tag{18}$$

$(f, \preceq)$ is said to have the *consistency property* if whenever $\{b\} = \{c\}$, then

$$f(b) \leq f(c) \Rightarrow b \preceq c. \tag{19}$$

The preference relation $\preceq$ is extended to sets by saying that $B \preceq C$ if and only if $b \preceq c$, where b and c are the minimum cost feasible permutations of B and C, respectively. The set $B \subseteq S$ is called *p-minimal* in $G = (S, R)$ if B is a nonempty initial set in G and $B \preceq C$ for every nonempty initial set C in G. A p-minimal set for which no proper subset is p-minimal is called *p^*-minimal.*

Monma and Sidney [28] presented the following algorithm.

Decomposition Algorithm:

Step 0 (Initialize) Set v to be the empty permutation.

Step 1 (Decompose) (a) If $S = \emptyset$ then stop; v is an optimal permutation. (b) Else, find a p^*-minimal set U in $G = (S, R)$.

Step 2 (Sequence) Find an optimal permutation u for U. Set $v := vu$. Delete U from G and go to Step 1.

Theorem 3 *[28] Assume that the strong ASI, strong SND, and consistency properties hold. Then the Decomposition Algorithm produces only optimal permutations, and every optimal permutation can be produced by the Decomposition Theorem.*

Next, we show how to define a sequencing function for our problem that exhibits the decomposition properties required to apply the results in [28]. Suppose that we use $\hat{V}$ as our sequencing function. It follows from (8), (9) and (12) that for the strong ASI property to hold, the preference relation would need to be defined so that

$$b \preceq c \Leftrightarrow \frac{\overline{\psi}(b) - 1}{\hat{V}(b)} \leq \frac{\overline{\psi}(c) - 1}{\hat{V}(c)}. \tag{20}$$

However, if the preference relation is so defined, then the conditions $\hat{V}(b) \leq \hat{V}(c)$, $\{b\} = \{c\}$ would imply that $c \preceq b$. This is the opposite of what is required for the consistency property (19) to hold.

In order to obtain a preference relation that satisfies the consistency property (as well as the other requisite properties), we replace the sequencing function $\hat{V}$ with the sequencing function $\hat{V}'$ defined as follows: For any schedule $a = (a_1, \dots, a_m)$, define the *reverse* of a, denoted a', by $a' = (a_m, \dots, a_1)$. Then define

$$\hat{V}'(a) = \hat{V}(a'). \tag{21}$$

Also, given a precedence graph $G = (S, R)$, let $G' = (S, R')$ be the *reverse* of G–i.e., $(i,j) \in R \Leftrightarrow (j,i) \in R'$. Then a complete schedule t satisfies the precedence constraints specified by the precedence graph G if and only if t' satisfies the precedence constraints specified by the reversed precedence graph G'. These definitions lead to the following equivalency result, the proof of which is immediate and therefore omited.

Lemma 2 *A schedule t is an optimal solution to the sequencing problem associated with the sequencing function $\hat{V}$ and precedence graph G if and only if t' is an optimal solution to the problem associated with the sequencing function $\hat{V}'$ and precedence graph G'.*

Thus, from the above result, it suffices to work with $\hat{V}'$ instead of $\hat{V}$. In place of (9) and (12) we have

$$\begin{aligned} \hat{V}'(ab) &= \hat{V}'(b) + \overline{\psi}(b)\hat{V}'(a), & (22)\\ \hat{V}'(abcd) - \hat{V}'(acbd) &= \{\hat{V}'(b)[\overline{\psi}(c) - 1] - \hat{V}'(c)[\overline{\psi}(b) - 1]\}\overline{\psi}(d). & (23) \end{aligned}$$

It follows that

$$\hat{V}'(abcd) \le \hat{V}'(acbd) \Leftrightarrow \frac{\overline{\psi}(c) - 1}{\hat{V}'(c)} \le \frac{\overline{\psi}(b) - 1}{\hat{V}'(b)}. \tag{24}$$

Thus, the strong ASI property will hold provided we use $\hat{V}'$ as our sequencing function, and we define the preference relation by

$$b \preceq c \Leftrightarrow \frac{\overline{\psi}(c) - 1}{\hat{V}'(c)} \le \frac{\overline{\psi}(b) - 1}{\hat{V}'(b)}. \tag{25}$$

It is straightforward to verify that with this definition, the strong SND property and the consistency property also hold. Note that the numerators in (25) have the same sign as that of the risk-sensitivity parameter γ.

We summarize the above results as follows.

Proposition 2 *For the sequencing function $\hat{V}'$ and associated preference relation (25), each of the strong ASI, strong SND, and consistency properties hold.*

From Theorem 3 and Proposition 2 we obtain the following result.

Theorem 4 *For the MEF problem with $\gamma \neq 0$ and precedence graph G, a complete schedule t is optimal if and only if it is the reverse of a schedule produced by the Decomposition Algorithm with precedence graph G' and preference relation (25).*

Example 2 Consider Example 1 but suppose we add the precedence constraint $2 \to 3$–i.e., that component 2 must be tested prior to component 3. Obviously, the schedule $(1,2,3)$ remains optimal for $-\infty < \gamma < 0.1869$. However, it is not obvious which schedule is optimal for larger values of γ since none of $(1,3,2)$, $(3,1,2)$, $(3,2,1)$ are feasible. A direct numerical investigation reveals that $(1,2,3)$ is in fact optimal for $-\infty < \gamma < 0.2119$, and $(2,3,1)$ is optimal for $0.2120 \le \gamma < \infty$.

Remark 5 Using analogous arguments, it is possible to show that the decomposition results are also applicable to the INF problem with precedence constraints.

Remark 6 The properties introduced in [28] and used here, i.e., the strong ASI, strong SND and consistency properties, guarantee optimality of sequences obtained via modular decompositions; see [28] and references therein. Although p^*-minimal sets can be found in polynomial time for some problems, to our knowledge this is an open issue for problems with exponential criteria [28].

References

[1] D. Adolphson and T. C. Hu. Optimal linear ordering. *SIAM J. Appl. Math.*, 25:403–423, 1973.

[2] G. Avila-Godoy and E. Fernandez-Gaucherand. Exponential risk-sensitive optimal scheduling. In *Proceedings of the 36th IEEE Conference on Decision and Control*, pages 3958–3963, San Diego, CA, December 1997.

[3] J. Baras, H. Li, and G. Mykoniatis. Integrated, distributed fault management for communication networks. Technical Report 98-10, Center for Satellite and Hybrid Communication Networks, University of Maryland, College Park, MD, 1998.

[4] J. S. Baras and M. R. James. Robust and risk-sensitive output feedback control for finite state machines and hidden markov models. *J. Math. Sys., Estimation and Control*, to appear.

[5] R. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, New Jersey, 1957.

[6] D. Bertsekas. *Dynamic Programming and Optimal Control, Volume 1.* Athena Scientific, Belmont, MA, 1995.

[7] W. Black. Discrete sequential search. *Information and Control*, 8:156–162, 1965.

[8] D. A. Castanon. Optimal search strategies in dynamic hypothesis testing. *IEEE Transactions on Systems, Man and Cybernetics*, 25:1130–1138, 1995.

[9] R. Cavazos-Cadena and E. Fernández-Gaucherand. Controlled markov chains with risk-sensitive criteria: average cost, optimality equations, and optimal solutions. *Mathematical Methods of Operations Research*, 49:299–324, 1999.

[10] B. V. Dean. Stochastic networks in research planning. In Yovits et al., editor, *Research Program Effectiveness*, chapter 12. Gordon and Breach, New York, 1966.

[11] D. C. Denby. Minimum downtime as a function of reliability and priority in a component repair. *Journal of Industrial Engineering*, 18:436–439, 1967.

[12] P. Dupuis, M.R. James, and I. Peterson. Robust properties of risk-sensitive control. Technical Report 98-15, Lefschetz Center for Dynamical Systems, Brown University, Providence, RI, August 1998.

[13] M. R. Garey. Optimal sequencing with precedence constraints. *Discrete Mathematics*, 4:37–56, 1973.

[14] K.D. Glazebrook and J.C. Gittens. On single-machine scheduling under precedence constraints and linear or discounted costs. Technical report, Mathematical Institute, University of Oxford, 1979.

[15] H. Greenberg. Optimum test procedure under stress. *Operations Research*, 12:689–692, 1964.

[16] G. J. Hall. Sequential search with random overlook probabilities. *Annals of Statistics*, 4:807–816, 1976.

[17] W. A. Horn. Single-machine job sequencing with treelike precedence ordering and linear delay penalties. *SIAM J. Appl. Math.*, 23:189–202, 1972.

[18] A. R. Howard and J. E. Matheson. Risk-sensitive markov decision processes. *Management Sciences*, 18:356–369, 1972.

[19] J-F. Huard and A.A. Lazar. Fault isolation based on decision-theoretic troubleshooting. Technical Report 442-96-08, Center for Telecommunications Research, Columbia University, New York, NY, 1996.

[20] F.V. Jensen. *An Introduction to Bayesian Networks.* Springer, New York, 1996.

[21] W. B. Joyce. Organization of unsuccessful r&d programs. *IEEE Transactions on Engineering Management*, 18:57–65, 1971.

[22] J. B. Kadane. Quiz show problems. *Journal of Mathematical Analysis and Applications*, 27:609–623, 1969.

[23] J. B. Kadane and H. A. Simon. Optimal strategies for a class of constrained sequential problems. *Annals of Statistics*, 5(2):237–255, 1977.

[24] J. Kalagnanam and M. Henrion. A comparison of decision analysis and expert rules for sequential diagnosis. In *Uncertainty in Artificial Intelligence 4*, pages 271–281. Elsevier, 1990.

[25] S. I. Marcus, E. Fernández-Gaucherand, D. Hernández-Hernández, S. Coraluppi, and P. Fard. Risk sensitive markov decision processes. In C. I. Byrnes, B. N. Datta, D. S. Gilliam, and C. F. Martin, editors, *Systems & Control in the Twenty-First Century*, pages 263–279. Birkhauser, 1996.

[26] D. Matula. A periodic optimal search. *Amer. Math. Monthly*, 71:15–21, 1964.

[27] L. G. Mitten. An analytic solution to the least cost testing sequence problem. *Journal of Industrial Engineering*, 11:17, 1960.

[28] C. L. Monma and J. B. Sidney. Optimal sequencing via modular decomposition: Characterization of sequencing functions. *Mathematics of Operations Research*, 12(1):22–31, 1987.

[29] M. Pinedo. *Scheduling: Theory, Algorithms, and Systems.* Prentice Hall, Englewood Cliffs, NJ, 1995.

[30] J. W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32:122–136, 1964.

[31] M. A. Shayman and E. Fernandez-Gaucherand. Risk-sensitive decision-theoretic troubleshooting. In *Proceedings of Allerton Conference on Communication, Control and Computing*, Allerton, IL, September 1999.

[32] J. B. Sidney. Decomposition algorithms for single-machine sequencing with preference relations and deferral costs. *Operations Research*, 23:283–298, 1975.

[33] J. B. Sidney. A decomposition algorithm for sequencing with general precedence constraints. *Mathematics of Operations Research*, 6:190–204, 1981.

[34] H. A. Simon and J. B. Kadane. Optimal problem-solving search: All-or-none solutions. *Artificial Intelligence*, 6:235–247, 1975.

[35] O. Staroverov. On a searching problem. *Theory Probability Appl.*, 8:184–187, 1963.

[36] L. D. Stone. *Theory of Optimal Search.* Academic Press, New York, 1975.

[37] A. Tanenbaum. *Computer Networks.* Prentice Hall PTR, Upper Saddle River, NJ, 1996.

[38] P. Whittle. *Risk-sensitive Optimal Control.* Wiley, New York, 1990.