

K.B. Lakshmanan^a, Daniel J. Rosenkrantz^b, S.S. Ravi^{b,*}

^a Department of Computer Science, State University of New York, Brockport, NY 14420-2933, USA

^b Department of Computer Science, State University of New York, Albany, NY 12222-0001, USA

Received April 1997; revised July 1998

Communicated by G. Ausiello

Abstract

In this paper, we consider systems that can be modeled as directed acyclic graphs such that nodes represent components of the system and directed edges represent fault propagation between components. Some components can be equipped with alarms that ring when they detect faulty (abnormal) behavior. We study algorithms that attempt to minimize the number of alarms to be placed so that a fault at any single component can be detected and uniquely diagnosed. We first show that the minimization problem is intractable, i.e., NP-hard, even when restricted to three level graphs in which all nodes have outdegree two or less. We present optimal algorithms for three special classes of graphs – tree structured graphs, single-entry single-exit series-parallel graphs and two level graphs. We then present a polynomial-time approximation algorithm for the general case which guarantees that the ratio of the number of alarms placed to the optimum required is within a factor that is logarithmic in the number of nodes in the graph. Moreover, by showing a reduction from the minimum dominating set problem to the minimum alarm set problem, we argue that this performance guarantee is tight to within a constant factor. Finally, we demonstrate the connection between the minimum alarm set problem and the minimum test collection problem, and prove similar results. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Alarm placement; Approximation algorithms; Fault diagnosis; Minimum test collection; NP optimization problems

1. Introduction

Since the pioneering work of Mayeda and Ramamoorthy [15] and Preparata et al. [16], graph models of systems have been employed quite extensively in the study of fault diagnosis and fault tolerance [14]. There are, however, a number of differences in

[☆] Research supported by NSF Grant CCR-97-34936. A brief version of the paper appeared in the Proceedings of the 1995 Pacific Rim International Symposium on Fault Tolerant Systems, December 1995.

* Corresponding author.

E-mail address: ravi@cs.albany.edu (S.S. Ravi).

what the graphs really represent in the various studies. In this paper, we are interested in *operative diagnosis* of faults arising in a wide variety of systems such as chemical plants, aircrafts, and medical diagnosis [10, 11, 18–20]. In these applications, the system under consideration consists of a number of components, some of which may become faulty. The fault at a component will result in the faulty or abnormal behavior of not only that component but also a few others. This manifestation is called *fault propagation*. Some components can be monitored for their abnormal behavior using sensors or *alarms*. Our interest is in algorithms for the placement of alarms that permit detection and unique identification of faulty components.

Formally, let $G = (V, E)$ be a directed graph that models the fault propagation characteristic of the system under consideration. That is, each node $v \in V$ of the graph represents a system component, and the directed edge $(u, v) \in E$ represents that a fault at u will propagate to v . If there is a directed path from u to w , the fault at u will also propagate to w along this path. Consider an alarm that is attached to a node v to observe its faulty or abnormal behavior. The alarm will ring if v is faulty or if some other node u is faulty and the fault propagates to v . In other words, a fault at a node will cause every reachable alarm to ring.

In this paper, we consider only single faults. We are interested in placing alarms on the nodes so that a fault at any single node can be detected and the faulty node uniquely identified. A fault can be detected only if at least one alarm rings. The *syndrome* for fault diagnosis is the set of ringing alarms. We say that a set of alarms allows *unique single fault diagnosis* if a fault can be detected and the faulty node correctly identified, provided there is only one faulty node. Such an alarm set is also said to be a solution for the *alarm placement problem*. The *minimum alarm set problem* is to find a solution for the alarm placement problem that requires the fewest number of alarms. In other words, every solution for the alarm placement problem is a feasible solution for the minimum alarm set problem. In the standard notation employed in [4], our optimization problem can be stated as follows:

Instance: A directed acyclic graph $G = (V, E)$.

Solution: A set of alarm nodes $A \subseteq V$ that allows unique single fault diagnosis, i.e., detect and uniquely identify any single faulty node.

Measure: Cardinality of the alarm set A , i.e., $|A|$.

Example 1. Consider the fault propagation graph shown in Fig. 1. Nodes 4 and 5 require an alarm as they are of outdegree 0 and a fault at either of these two nodes will not be detected unless they are equipped with alarms. In fact, these two alarms by themselves are sufficient to detect a single fault. However, they are not sufficient to uniquely identify a faulty node. For example, a single fault at nodes 1, 2 or 3 will cause both alarms 4 and 5 to ring. Consider the alarm set $\{2, 3, 4, 5\}$. The sets of ringing alarms for a single fault at nodes 1–5 are $\{2, 3, 4, 5\}$, $\{2, 4, 5\}$, $\{3, 4, 5\}$, $\{4\}$ and $\{5\}$, respectively. Thus, for every faulty node, the set of ringing alarms is nonempty and is distinct. The alarm sets $\{1, 2, 4, 5\}$, $\{1, 3, 4, 5\}$, $\{2, 3, 4, 5\}$ and $\{1, 2, 3, 4, 5\}$ are all

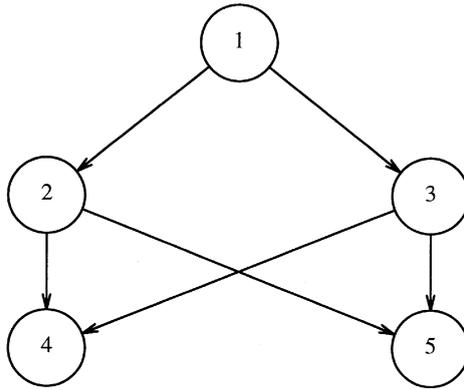


Fig. 1. Fault propagation graph for Example 1.

solutions for the alarm placement problem. The first three are also optimal solutions for the minimum alarm set problem.

Several practical systems for which the graph model under consideration is applicable can be seen in [10, 11, 18–20]. In [18, 19], Rao investigated several fault diagnosis algorithms and their complexities. He also presented NP-completeness results for the alarm placement and multiple fault diagnosis problems. The intractability result for the alarm placement problem in this paper is based on a reduction from the vertex cover problem – different from that in [19]. Our reduction allows proof of intractability of the problem even for very simple directed graphs.

In Sections 2 and 3, we present some formal definitions and preliminary results. In Section 4, we show that the alarm placement problem is intractable, i.e., it is NP-complete, even when restricted to three level graphs in which all nodes have outdegree two or less. The implication of the NP-completeness result is that algorithms that operate in polynomial time, in terms of the number of nodes or edges in the graph, are not likely to be found for the alarm placement problem. For practical engineering systems with modular design, the associated fault propagation graph is likely to be sparse with simpler structure. We therefore focus on graphs that exhibit special structure – tree structured graphs, single-entry single-exit series-parallel graphs and two level graphs, and present optimal algorithms. In Section 6, we present a polynomial-time approximation algorithm for the general case that guarantees that the ratio of the number of alarms placed to the optimum required is at most $0.31 + 2 \ln n$, where n is the number of nodes in the graph. While this logarithmic ratio may appear to be too large for practical applications, we argue that polynomial-time approximation algorithms with better ratio bounds are unlikely to be found, by showing a reduction from the dominating set problem to the alarm placement problem. Finally, in Section 7, we demonstrate the connection between the minimum alarm set problem and the minimum test collection problem studied in [8] for fault diagnosis in the structural model originated by Mayeda and Ramamoorthy [15].

2. Graph preprocessing

2.1. Condensation

If the directed graph that models the fault propagation characteristic of the system under consideration contains a cycle, a fault at any one of those nodes will propagate to every other node in the cycle, making identification of the faulty node impossible by any algorithm. Hence, we assume that the fault propagation graph has been *condensed* by replacing each strongly connected component of the graph by a single node. Such a condensation of a directed graph with respect to strongly connected components produces an acyclic graph [7]. The condensation can be carried out in linear time in the number of nodes and edges in the original graph.

2.2. Transitive reduction

For the alarm placement problem, we are only interested in knowing whether there is path from one node to another in the graph. A directed graph G^t is said to be a *transitive reduction* of the directed graph G provided there is a directed path from node u to node v in G^t if and only if there is such a path in G , and that there is no graph with fewer edges than G^t satisfying the above property [1]. The transitive reduction of a directed acyclic graph is unique and the time complexity of obtaining the transitive reduction is the same as that of computing the transitive closure.

3. Preliminary results

Let $G=(V,E)$ be a condensed, transitively reduced, directed acyclic graph with $|V|=n$. We begin this section with a number of definitions of interest. We measure the *length* of a path by the number of edges in the path.

Definition 1. The *level of a node* in a directed acyclic graph G is one more than the length of the longest path from that node to a node of outdegree 0.

Definition 2. A directed acyclic graph G is said to be a *t level graph* if t is the highest level of any node.

Definition 3. An alarm x is said to be *reachable* from a faulty node a if there is a directed path from a to x in G .

Definition 4. An alarm x *distinguishes* the set of faulty nodes $\{a,b\}$ if and only if x is reachable from a but not from b , or vice versa.

Definition 5. An alarm set A distinguishes the set of faulty nodes $\{a,b\}$ if and only if there exists an alarm $x \in A$ such that x distinguishes $\{a,b\}$.

Definition 6. An alarm set A allows unique single fault diagnosis if and only if for every faulty node $a \in V$, the set of reachable alarms is nonempty, and for every pair of faulty nodes $a, b \in V$, A distinguishes $\{a, b\}$.

In every acyclic directed graph G , there is at least one node of outdegree 0 [7]. A fault at a node of outdegree 0 does not propagate, and hence an alarm is required in that node in order to achieve fault detection. Besides, every node in the graph G has a directed path to at least one node of outdegree 0 [7]. Thus, placing alarms in every node of outdegree 0 ensures that the set of ringing alarms is nonempty for every faulty node. Further, if node u is of outdegree 1 with the only outgoing edge (u, v) , every alarm node reachable from v is also reachable from u . Only an alarm at u can distinguish the set of faulty nodes $\{u, v\}$. Hence, in any solution for the alarm placement problem, nodes of outdegree 0 and 1 must be equipped with alarms. However, these alarms may not be sufficient for unique single fault diagnosis. On the other hand, since the directed graph is acyclic, for every pair $a, b \in V$, either a or b is not reachable from the other. Therefore, alarms at all nodes of the graph is one solution for the alarm placement problem.

Lemma 1. Let A^* be an optimal solution for the minimum alarm set problem on the graph G . Then

$$\lceil \log_2(n+1) \rceil \leq |A^*| \leq n.$$

Moreover, these bounds are tight, i.e., there exist graphs with n nodes for which these bounds are met.

Proof. Since the set of ringing alarms for each faulty node should be nonempty and distinct from the set of ringing alarms for other faulty nodes, the lower bound follows from information theoretic considerations. Now consider a two level graph with k nodes of outdegree 0 at Level 1, each of which must be equipped with an alarm. At Level 2, introduce one node for every combination of 2 or more Level 1 nodes and add edges from these new nodes to the corresponding Level 1 nodes. The number of Level 2 nodes is $2^k - k - 1$. This graph meets the lower bound for the number of alarms. The upper bound of n for the number of alarms follows from the fact that alarms at all nodes of the graph is one solution for the alarm placement problem. A graph consisting of a directed path of n nodes meets this bound. \square

4. NP-completeness

As seen in the previous section, every node of outdegree 0 or 1 must be equipped with an alarm in any solution for the alarm placement problem. In this section, we will show that the *decision theoretic* version of the alarm placement problem is NP-complete, even when restricted to three level graphs in which all nodes have

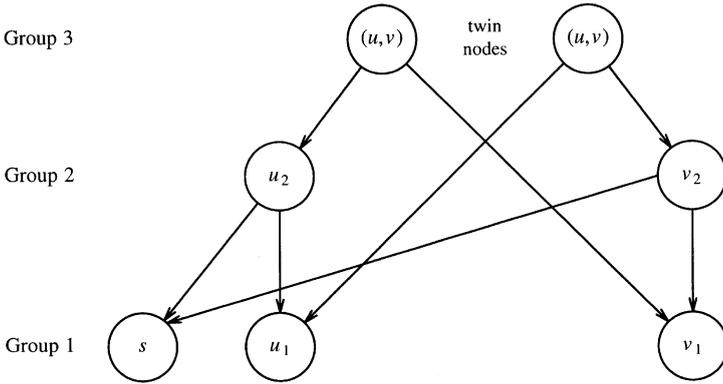


Fig. 2. Graph structure for reduction from the vertex cover problem.

outdegree 2 or less. Specifically, we show the following problem to be NP-complete: Given a directed acyclic graph $G = (V, E)$ and a positive integer k , does there exist a solution to the alarm placement problem that uses k or fewer alarms? For this result, we employ a reduction from the NP-complete *vertex(node) cover* problem [6], which can be stated as follows: Given an undirected graph $G' = (V', E')$ and a positive integer k , is there a vertex cover of size k or less for G' , i.e., a subset $C \subseteq V'$ with $|C| \leq k$ such that for each edge $(u, v) \in E'$ at least one of u and v belongs to C ?

Theorem 2. *The alarm placement problem for unique single fault diagnosis is NP-complete, even when restricted to three level graphs in which all nodes have outdegree 2 or less.*

Proof. Whether or not a given placement of alarms allows unique single fault diagnosis can be verified in polynomial time since this merely entails the computation of the set of alarms reachable from each node. Hence the above problem is in NP. Given an arbitrary undirected graph $G' = (V', E')$, our goal now is to construct a directed acyclic graph G on which the alarm placement problem can be studied. Consider a graph structure G illustrated in Fig. 2. For each node $v \in V'$ in G' , place one node v_1 in Group 1 and another node v_2 in Group 2. In addition, place an extra node s in Group 1. For each edge $(u, v) \in E'$ in G' , place two nodes – “twin” nodes – in Group 3. Introduce edges in G from each node v_2 to its corresponding node v_1 and to the special node s . For one of the twin nodes in Group 3 corresponding to the edge (u, v) , introduce edges to node u_2 in Group 2 and node v_1 in Group 1. For the other twin node corresponding to the edge (u, v) , introduce edges to node v_2 in Group 2 and node u_1 in Group 1. Thus, in G , all nodes in Group 1 are of outdegree 0, and all nodes in Groups 2 and 3 are of outdegree 2. Also, G is a three level graph which exhibits no reconvergent fanout.

Consider now the alarm placement for unique single fault diagnosis in G . Clearly every node in Group 1 requires an alarm. Once those alarms are placed, a fault in a

Group t node rings exactly t alarms in Group 1, for $1 \leq t \leq 3$. Thus nodes in different groups can be distinguished simply by the number of Group 1 alarms they ring. Moreover, the alarms in Group 1 allow the faulty node to be uniquely identified, except that the twin nodes in Group 3 corresponding to each edge cannot be distinguished.

We now claim that G' has a vertex cover of size k if and only if G has a solution of size $n + k + 1$ for the alarm placement problem. Suppose C is a vertex cover of size k for G' . Place alarms in all Group 1 nodes and in those Group 2 nodes in G that correspond to the nodes in the vertex cover C . Since each edge (u, v) is incident on at least one vertex in the cover, the corresponding alarm(s) in Group 2 will help distinguish the twin nodes in Group 3 corresponding to (u, v) . Conversely, suppose we have a solution to the alarm placement problem in G of size $n + k + 1$. Clearly $n + 1$ of these alarms are in Group 1. If any node in Group 3 has an alarm, the only purpose that alarm serves is to distinguish that node from its twin node and hence that alarm can be moved down to the node in Group 2 to which it or its twin is directly connected. In other words, every solution for the alarm placement problem can be suitably adjusted with no increase in the number of alarms so that alarms are placed only in Group 1 and Group 2 nodes. Since each pair of twin nodes corresponding to an edge is distinguished, the alarm nodes in Group 2 form a vertex cover in G' . \square

The implication of the NP-completeness proof above is that we are unlikely to find a polynomial-time algorithm for solving the problem exactly. There are two avenues to pursue: construct optimal algorithms for special classes of graphs and/or construct approximation algorithms for the general case. We pursue both approaches.

5. Optimal algorithms for special classes of graphs

Let $G = (V, E)$ be a condensed, transitively reduced, directed acyclic graph with $|V| = n$, and $|E| = e$. As seen already, nodes of outdegree 0 and 1 must be equipped with alarms in any solution for the alarm placement problem. In this section, we will show that placing alarms at all nodes of outdegree 0 and 1 is *sufficient* for unique single fault diagnosis, if G is a tree structured graph or a single-entry single-exit series-parallel graph. Thus, the optimal solution for the minimum alarm set problem can be found for these cases in $O(n + e)$ time. Also, we have already proved that the minimum alarm set problem is intractable even for three level graphs. In this section, we will present an optimal algorithm for two level graphs. The algorithm can be implemented to run in $O(n^2)$ time.

5.1. Tree structured graphs

We say that G is a *tree structured graph* if the undirected graph corresponding to G is acyclic. Note that G is not necessarily a rooted directed tree.

Theorem 3. *If G is a tree, placing alarms at all nodes of outdegree 0 and 1 is sufficient for unique single fault diagnosis in G .*

Proof. The proof is by contradiction. Suppose placing alarms at all nodes of outdegree 0 and 1 does not distinguish the set of faulty nodes $\{a, b\}$, i.e., a and b ring the same set of alarms. Given that there is no directed cycle in G , either a or b is not reachable from the other. Without loss of generality, assume that a is not reachable from b .

Case 1: a is of outdegree 0 or 1. In this case a will be equipped with an alarm and a fault at b cannot ring this alarm, contradicting the assumption that both a and b ring the same set of alarms.

Case 2: a is of outdegree 2 or more, and hence there is no alarm placed at a . Let (a, c_1) and (a, c_2) be two edges outgoing from a . Let d_i be an alarm reachable from c_i by a path of zero or more directed edges, for $i = 1, 2$. Since all nodes of outdegree 0 have alarms, d_i exist. Now consider the paths from a to d_1 and d_2 . These two paths are node disjoint, except for the common node a ; otherwise, there will be two paths from a to this common node, contradicting the assumption that G is a tree structured graph. Consider paths from b to d_1 and d_2 . Let e be the last common node in these two paths, i.e., the path from e to d_1 is node disjoint from the path from e to d_2 , except for the common node e . Note that there is no path from e to a and that e cannot lie in either of the two paths from a to d_1 or from a to d_2 . Let f be the first node where the path from a to d_1 meets the path from e to d_1 . Also let g be the first node where the path a to d_2 meets the path from e to d_2 . Then, the paths from a to f and g along with the paths from e to f and g form an undirected cycle, contradicting that G is a tree structured graph. \square

5.2. Single-entry single-exit series-parallel graphs

Single-entry single-exit connected series-parallel graphs, or SEC series-parallel for short, are directed acyclic graphs that have an ordered pair of special nodes called the *source* and the *sink*, or collectively as *terminals*. This class of graphs is defined recursively by the following rules:

- (i) A graph consisting of a source and a sink with a single edge from the source to the sink is a SEC series-parallel graph.
- (ii) If G_1 and G_2 are SEC series-parallel graphs with terminals (s_1, t_1) and (s_2, t_2) respectively, then the graph G obtained by merging node t_1 with s_2 is a SEC series-parallel graph with terminals (s_1, t_2) . G is said to be a series connection of G_1 and G_2 .
- (iii) If G_1 and G_2 are SEC series-parallel graphs with terminals (s_1, t_1) and (s_2, t_2) respectively, then the graph G obtained by merging node s_1 with s_2 and node t_1 with t_2 is a SEC series-parallel graph with terminals (s_1, t_1) . G is said to be a parallel connection of G_1 and G_2 .
- (iv) Any graph that cannot be obtained by the above rules is not a SEC series-parallel graph.

If G is a SEC series–parallel graph with terminals (s, t) , then every node in G is reachable from s , but s is not reachable from any other node. Moreover, no node in G , other than t itself, is reachable from t , but t is reachable from every node in G . Thus, a SEC series–parallel graph is clearly acyclic. If we restrict the parallel connection to only graphs with three or more nodes, then the resulting graph will also be transitively reduced. The following lemma is an easy consequence of these observations.

Lemma 4. *If G is a SEC series–parallel graph with terminals (s, t) , then*

- (i) *t is of outdegree 0 and hence will require an alarm in any solution for the alarm placement problem.*
- (ii) *Since t is reachable from every node in G , at least one other alarm must be reachable from every other node in any solution for the alarm placement problem.*
- (iii) *If G has three or more nodes, three or more alarms are necessary in any solution for the alarm placement problem.*

Lemma 5. *Let A_i be a solution for the alarm placement problem on a SEC series–parallel graph $G_i = (V_i, E_i)$, for $i = 1, 2$. Then $(A_1 - \{t_1\}) \cup A_2$ is a solution for the alarm placement problem on the graph G which is a series connection of G_1 and G_2 .*

Proof. The proof is by contradiction. Suppose the alarm placement in G does not distinguish the set of faulty nodes $\{a, b\}$, i.e., a and b ring the same set of alarms.

Case 1: $a \in V_2$ and $b \in V_2$. Since the series connection does not change the set of ringing alarms for any faulty node in G_2 , including s_2 , this contradicts the assumption that A_2 is a solution for the alarm placement problem on G_2 .

Case 2: $a \in V_1 - \{t_1\}$ and $b \in V_2$. By Lemma 4(ii), a will ring at least one alarm in $A_1 - \{t_1\}$, which b cannot. This contradicts the assumption that a and b ring the same set of alarms.

Case 3: $a \in V_1 - \{t_1\}$ and $b \in V_1$. After the series connection, every faulty node in G_1 will ring all the alarms in G_2 , in addition to the same set of alarms they rang before the connection, except for t_1 . Since the set of ringing alarms is modified in exactly the same way for every faulty node in G_1 , this contradicts the assumption that A_1 is a solution for the alarm placement problem on G_1 . \square

Lemma 6. *Let A_i be a solution for the alarm placement problem on a SEC series–parallel graph $G_i = (V_i, E_i)$, for $i = 1, 2$, each with three or more nodes. Then $(A_1 - \{s_1\}) \cup (A_2 - \{s_2, t_2\})$ is a solution for the alarm placement problem on the graph G which is a parallel connection of G_1 and G_2 .*

Proof. The proof is by contradiction. Suppose the alarm placement in G does not distinguish the set of faulty nodes $\{a, b\}$, i.e., a and b ring the same set of alarms. Since both G_1 and G_2 have three or more nodes, by Lemma 4(iii), each must have at least one non-terminal alarm node. After the parallel connection, the merged source s_1 is the only node in G that can ring those non-terminals alarms in both G_1 and G_2 . Thus s_1 can always be uniquely diagnosed.

Case 1: $a \in V_1 - \{s_1\}$ and $b \in V_1 - \{s_1\}$. Since the parallel connection does not change the set of ringing alarms for any faulty node in $V_1 - \{s_1\}$, this contradicts the assumption that A_1 is a solution for the alarm placement problem on G_1 .

Case 2: $a \in V_2 - \{s_2\}$ and $b \in V_2 - \{s_2\}$. Since the parallel connection does not change the set of ringing alarms for any faulty node in $V_2 - \{s_2\}$, except changing t_2 to t_1 , this contradicts the assumption that A_2 is a solution for the alarm placement problem on G_2 .

Case 3: $a \in V_1 - \{s_1\}$ and $b \in V_2 - \{s_2\}$. By Lemma 4(ii), a will ring at least one alarm in $A_1 - \{t_1\}$. This alarm cannot be at s_1 since s_1 is not reachable from a . Thus, a will ring at least one alarm in $A_1 - \{t_1, s_1\}$, which b cannot. This contradicts the assumption that a and b ring the same set of alarms. \square

Theorem 7. *If G is a transitively reduced SEC series–parallel graph, placing alarms at all nodes of outdegree 0 and 1 is sufficient for unique single fault diagnosis in G .*

Proof. Obtain a solution for the alarm placement problem on G by constructing G itself through a sequence of series and parallel connections, starting from the fundamental building block for SEC series–parallel graphs – an edge from the source to the sink node. This graph requires an alarm at both nodes, since their outdegrees are 0 and 1. Also, since G is transitively reduced, we require parallel connection of two graphs only if each has three or more nodes. Lemmas 5 and 6 assure us that neither a series connection nor a parallel connection will introduce new alarms. Moreover, whenever a parallel connection of two graphs is made, the alarms, if any, at their source nodes can be removed. Since a node of outdegree 2 or more can be formed only by a parallel connection of two graphs, this constructive proof assures us that a node of outdegree 2 or more does not require an alarm in any solution for the alarm placement problem. \square

5.3. Two level graphs

In a two level graph, all nodes at Level 1 are of outdegree 0. Each node at Level 2 has one or more edges to Level 1 nodes. The proof that the following $O(n^2)$ algorithm computes an optimal solution for the minimum alarm set problem is fairly obvious.

OPT-TWO-LEVEL(G)

- (i) Place alarms at all Level 1 nodes.
- (ii) Place alarms at all Level 2 nodes of outdegree 1.
- (iii) Partition the remaining Level 2 nodes into equivalence classes based on the set of reachable Level 1 nodes. That is, nodes a and b belong to the same equivalence class if the set of reachable Level 1 nodes is the same for both a and b .
- (iv) For each equivalence class of Level 2 nodes, place alarms in all nodes but one.

Thus, if an equivalence class contains k nodes, place alarms at $k - 1$ of them.

END OPT-TWO-LEVEL

6. Approximation algorithm for general case

In this section, we employ a set covering model for choosing the set of alarm nodes that distinguishes every pair of faulty nodes. We also invoke the well-known greedy heuristic for solving the set covering problem as one of the steps in our approximation algorithm. See the detailed analysis of GREEDY-SET-COVER in [3]. In particular, note Corollary 37.5 and Exercise 37.3-3 in [3]. Let $G = (V, E)$ be a condensed, transitively reduced, directed acyclic graph with $|V| = n$, and $|E| = e$. Let A^* be an optimal solution for the minimum alarm set problem on G . Consider the following approximation algorithm APPROX-ALARM-SET that computes A_{approx} as an approximation for A^* .

APPROX-ALARM-SET(G)

- (i) For every node $v \in V$, compute $F(v)$, the set of nodes reachable from v .
- (ii) $R \leftarrow \{v: v \in V \text{ and outdegree of } v \text{ is } 0 \text{ or } 1\}$
- (iii) $X \leftarrow \{\{a, b\}: a, b \in V \text{ and } R \text{ does not distinguish } \{a, b\}\}$
- (iv) For every node $v \in V - R$, compute $S_v = \{\{a, b\}: \{a, b\} \in X, \text{ and } v \text{ distinguishes } \{a, b\}\}$.
- (v) Let C be the collection of subsets of X computed in Step (iv). Compute a *set cover* for X , i.e., a subcollection $C' \subseteq C$ such that each element in X belongs to at least one member of C' , using the greedy heuristic [3].
- (vi) For each set $S \in C'$, choose a node v such that $S = S_v$. Let this node be denoted by $\alpha(S)$. $H \leftarrow \{\alpha(S): S \in C'\}$
- (vii) $A_{\text{approx}} \leftarrow R \cup H$

END APPROX-ALARM-SET

Step (iv) computes at most n sets, each with at most $n(n-1)/2$ elements, and hence can be executed in $O(n^3)$ time. Step (v) can be executed in $O(n^3)$ time. See the analysis of GREEDY-SET-COVER in p.976 and Exercise 37.3-3 of [3]. All other steps can also be executed in $O(n^3)$ time. Thus APPROX-ALARM-SET runs in $O(n^3)$ time.

Theorem 8. *Algorithm APPROX-ALARM-SET produces a feasible solution for the minimum alarm set problem.*

Proof. The proof is by contradiction. First, observe that all nodes of outdegree 0 are included in A_{approx} . Hence a single fault can always be detected. Suppose A_{approx} does not distinguish the set of faulty nodes $\{a, b\}$.

Case 1: $\{a, b\} \notin X$. Then, there exists an $r \in R$ that distinguishes $\{a, b\}$. Since $r \in A_{\text{approx}}$, this contradicts the assumption that A_{approx} does not distinguish $\{a, b\}$.

Case 2: $\{a, b\} \in X$. Then, there exists an $S_v \in C'$ such that $\{a, b\} \in S_v$, i.e., v distinguishes $\{a, b\}$. Since $v \in A_{\text{approx}}$, this contradicts the assumption that A_{approx} does not distinguish $\{a, b\}$. \square

Definition 7. APPROX-ALARM-SET has a ratio bound of $\rho(n)$ if for every graph G with n nodes,

$$\frac{|A_{\text{approx}}|}{|A^*|} \leq \rho(n).$$

Theorem 9. Algorithm APPROX-ALARM-SET has a ratio bound of $0.31 + 2 \ln n$.

Proof. Observe that A^* is an optimal solution for the minimum alarm set problem on G if and only if $A^* - R$ is an optimal solution for the set cover problem in Step (v) of APPROX-ALARM-SET. The greedy heuristic for the set cover problem on a set X has a ratio bound of $1 + \ln |X|$. See Corollary 37.5 in [3]. In our case, $|X| \leq n(n-1)/2$, and hence H approximates $A^* - R$ within a ratio of $0.31 + 2 \ln n$. Thus APPROX-ALARM-SET has a ratio bound of $0.31 + 2 \ln n$. \square

Thus, the ratio of the number of alarms placed by APPROX-ALARM-SET to the optimum required is within $0.31 + 2 \ln n$ for any graph G . While this ratio may appear to be too large for practical applications, we now show that polynomial-time algorithms which improve the ratio bound by more than a constant factor are not likely to be found. Consider the *dominating set* problem [6] which can be stated as follows: Given a connected undirected graph $G' = (V', E')$, determine a dominating set for G' , i.e., a subset $D \subseteq V'$ such that for all $u \in V' - D$ there is a $v \in D$ for which $(u, v) \in E'$. The *minimum dominating set problem* is to find a dominating set of smallest size. The optimization problem is known to be NP-hard and hence approximation algorithms with guaranteed ratio bounds are of interest for this problem also.

Consider the following reduction from the dominating set problem to the alarm placement problem. Given an arbitrary connected undirected graph $G' = (V', E')$ with m nodes, our goal now is to construct a directed acyclic graph G on which the alarm placement problem can be studied. First, consider the graph structure G illustrated in Fig. 3(a). For each node $v \in V'$ in G' , place one node v_1 in Group 1 and another node v_2 in Group 2. In addition, place an extra node s in Group 1. Also, for each node $v \in V'$ in G' , place “twin” nodes v_{31} and v_{32} in Group 3. Introduce edges in G from each node v_2 to its corresponding node v_1 and to the special node s . Suppose $N(v)$ is the set of neighbors of v in G' . Introduce edges from v_{31} to v_2 in Group 2 and w_1 in Group 1 for each $w \in N(v)$. For its twin node v_{32} , introduce edges from v_{32} to v_1 in Group 1 and w_2 in Group 2 for each $w \in N(v)$.

Consider now the alarm placement for unique single fault diagnosis in G . Clearly every node in Group 1 requires an alarm. Once those alarms are placed, a fault in a Group t node rings exactly t alarms in Group 1, for $t = 1, 2$. A fault in a Group 3 node will cause 3 or more alarms to ring in Group 1. Thus nodes in different groups can be distinguished simply by the number of Group 1 alarms they ring. Faults in Groups 1 and 2 nodes can also be uniquely diagnosed. However, the twin nodes, e.g., v_{31} and v_{32} , cannot be distinguished. Moreover, if $\{u\} \cup N(u) = \{v\} \cup N(v)$, u_{31} , u_{32} , v_{31} , and v_{32} all ring the same set of alarms, and hence cannot be distinguished.

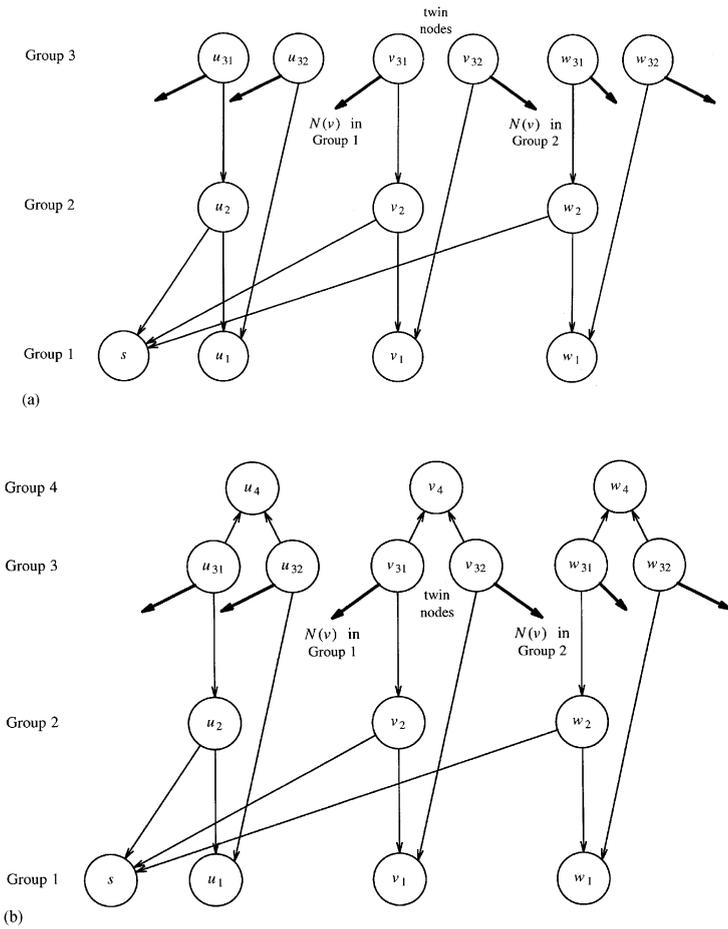


Fig. 3. Graph structure for reduction from the dominating set problem. (a) Initial structure. (b) Intermediate structure with Group 4 nodes. (c) Final structure with encoded Groups 1 and 4 nodes.

In order to uniquely diagnose Group 3 faulty nodes, we continue the reduction by introducing some more nodes in Group 4 as follows: For each $v \in V'$, introduce v_4 and include edges from the twin nodes v_{31} and v_{32} to v_4 . Fig. 3(b) illustrates this intermediate structure. All the new nodes are of outdegree 0 and hence will require alarms. A fault in a Group 4 node will cause exactly one Group 4 alarm to ring, and hence can be uniquely identified. The Group 4 nodes will now help distinguish u_{31} and u_{32} from v_{31} and v_{32} . At this stage only the twin nodes cannot be distinguished, i.e., v_{31} and v_{32} cannot be distinguished, for each $v \in V'$. We need additional alarms.

On the other hand, it is possible to reduce the number of alarm nodes in Groups 1 and 4 substantially and achieve the same precision in diagnosis. Consider the fault in a Group 2 node v_2 . It is recognized by the ringing of its unique Group 1 alarm node v_1 ,

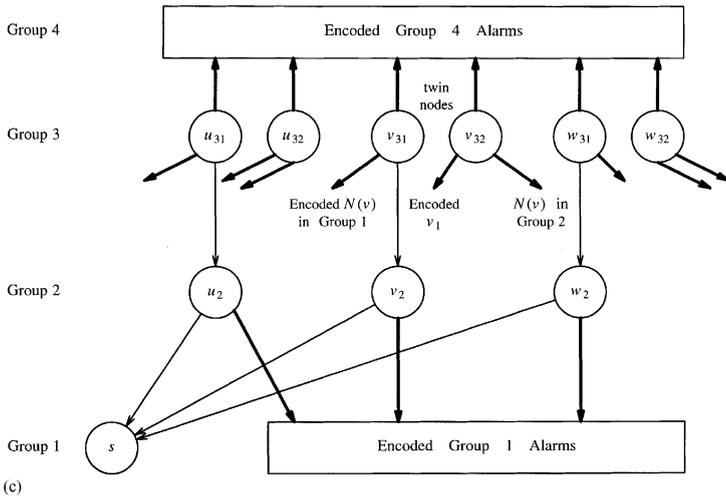


Fig. 3. (Continued)

along with s . This amounts to a unary encoding of the identity of the Group 2 faulty node when the status of all Group 1 alarms, excluding s , is viewed as a binary vector. Instead, we can encode the identity of the m Group 2 nodes in binary – from 1 to m . We now replace the Group 1 nodes other than s with $\lceil \log_2(m + 1) \rceil$ alarm nodes in Group 1, and introduce edges from the Group 2 nodes to the alarm nodes suitably so that each faulty node in Group 2 rings exactly those alarms which correspond to the positions of bit 1 in its binary representation. By encoding the identity of the Group 2 nodes from 1 to m , rather than from 0 to $m - 1$, we ensure that s and at least one more alarm will ring for each Group 2 faulty node. Observe that Group 3 nodes can still be distinguished from Group 2 nodes since they now ring Group 4 nodes also. The edges leading from Group 3 nodes to the binary encoded Group 1 nodes should also be suitably adjusted. For example, consider the edges from v_{31} to the Group 1 alarm nodes. Form the union of all Group 1 binary encoded alarm nodes, except s , that will ring for each one of the faulty nodes w_2 where $w \in N(v)$. Then introduce edges from v_{31} to each of the alarm nodes in the union just formed. Observe that Group 4 nodes can also binary encode the identity of the Group 3 nodes in a similar manner with $\lceil \log_2(m + 1) \rceil$ alarms.

The final structure of G with $3m + \ell$ nodes, where $\ell = 1 + 2\lceil \log_2(m + 1) \rceil$ is illustrated in Fig. 3(c). Of these, ℓ are alarm nodes in Groups 1 and 4 combined. In this final form of G , a fault in a Group 1 node causes exactly one Group 1 alarm to ring. A fault in a Group 2 node causes s and one or more encoded Group 1 alarms to ring. A fault in a Group 3 node causes s , one or more encoded Group 1 alarms and one or more encoded Group 4 alarms to ring. A fault in a Group 4 node causes exactly one Group 4 alarm to ring. Moreover, as already pointed out, we need additional alarms only to distinguish the twin nodes. For this, it is sufficient to consider solutions that

place alarms in Group 2 nodes. If any node in Group 3 has an alarm, the only purpose that alarm serves is to distinguish that node from its twin and hence that alarm can be moved down to any node in Group 2 to which it or its twin is directly connected. In other words, every solution for the alarm placement problem can be suitably adjusted with no increase in the number of alarms so that no alarms are placed in Group 3 nodes.

We now claim that G' has a dominating set of size k if and only if G has a solution of size $k + \ell$ for the alarm placement problem. Suppose D is a dominating set of size k for G' . Place alarms in all Groups 1 and 4 nodes, and in Group 2 nodes in G that correspond to the nodes in the dominating set D . Since each node v is adjacent to at least one node in the dominating set, the corresponding alarm(s) in Group 2 will distinguish the two nodes v_{31} and v_{32} . Conversely, suppose we have a solution to the alarm placement problem in G of size $k + \ell$, with no alarms in Group 3 nodes. Clearly, ℓ of these alarms are in Groups 1 and 4. Since each pair of twin nodes is distinguished, the alarm nodes in Group 2 form a dominating set in G' .

Lemma 10. *Suppose the minimum alarm set problem has a polynomial-time approximation algorithm with a ratio bound of $\rho(n)$ on directed graphs with n nodes, where ρ is a monotonically increasing positive function. Let α be an arbitrary positive constant. Then, the minimum dominating set problem has a polynomial-time approximation algorithm with a ratio bound of $(1 + 1/\alpha)\rho(4q\lceil\alpha\log_2 q\rceil) - (1/\alpha)$ on graphs with q nodes.*

Proof. Let $G' = (V', E')$ be an undirected graph with q nodes. Consider the following polynomial-time approximation algorithm for the dominating set problem.

Let $N_0 = 2^{3(\alpha+4)}$. Since α is a constant, so is N_0 . If $q \leq N_0$, we can find an optimal dominating set for G' in polynomial time, since G' has only a constant number of nodes. So, we may assume that $q > N_0$. The steps of the approximation algorithm are as follows.

- (i) Determine whether G' has a dominating set of size at most 4. If so, output the smallest such set and halt. (Obviously, this step runs in polynomial time.)
- (ii) Construct another undirected graph G'' consisting of $\beta = \lceil\alpha\log_2 q\rceil$ disjoint copies of G' .
- (iii) Starting with the undirected graph G'' , produce the directed acyclic graph G described in the reduction above.
- (iv) Run the polynomial-time alarm set approximation algorithm on G , and obtain the nodes of G'' corresponding to the Group 2 alarm nodes in the approximate alarm set solution. Let A'' denote the resulting set of nodes of G'' .
- (v) Partition A'' into $A'_1, A'_2, \dots, A'_\beta$, where A'_i contains only nodes from the i th copy of G' in G'' ($1 \leq i \leq \beta$). Return a set A'_j of minimum size among these subsets.

To begin with, note that G'' has $m = \beta q$ nodes. Further, if k' and k'' denote the sizes of minimum dominating sets for G' and G'' , respectively, then $k'' = k'\beta$. Let $\ell = 1 + 2\lceil\log_2(m + 1)\rceil$. We make an observation that follows in a straightforward manner from the definitions of N_0, β and ℓ .

Observation. For all $q > N_0$, the following inequalities hold: (a) $\log_2 q > 3 \log_2 \beta$ (b) $3 \log_2 m > \ell$.

If the algorithm halts after Step (i), we would once again have an optimal dominating set for G' . So, we assume that the algorithm did not halt after Step (i). Consequently, $k' > 4$. Using this fact, we can show that the parameters k'' and ℓ satisfy the following inequality.

Claim. For all $q > N_0$, $k'' > \alpha \ell$.

Proof. Since $k' > 4$, we have $k'' = k' \beta > 4\beta$. Now,

$$\begin{aligned}
 k'' &> 4\beta \\
 &\geq 4\alpha \log_2 q \quad (\text{since } \beta = \lceil \alpha \log_2 q \rceil) \\
 &> 3\alpha \log_2 q + 3\alpha \log_2 \beta \quad (\text{Part (a) of Observation}) \\
 &= 3\alpha \log_2(\beta q) \\
 &= \alpha(3 \log_2 m) \quad (\text{since } m = \beta q) \\
 &> \alpha \ell \quad (\text{Part (b) of Observation})
 \end{aligned}$$

This completes the proof of claim. \square

We now prove that the approximation algorithm indeed provides the performance guarantee indicated in the statement of the lemma. By the claim above, $k'' > \alpha \ell$. Any minimum alarm set for the constructed graph G is of size $k'' + \ell$. When the approximation algorithm with ratio bound $\rho(n)$ is run on G , it produces a solution with at most $\rho(n)(k'' + \ell)$ alarms. The number of Group 2 alarm nodes in this approximate solution is $\rho(n)k'' + (\rho(n) - 1)\ell$, which is less than $((1 + 1/\alpha)\rho(n) - (1/\alpha))k''$, since $k'' > \alpha \ell$. Note that G has $n = 3m + \ell$ nodes. For $m \geq 10$, it follows that $\ell < m$ and $n = 3m + \ell < 4m$. Therefore, the approximation algorithm for the minimum dominating set problem has a ratio bound of $(1 + 1/\alpha)\rho(4m) - (1/\alpha)$ on G'' .

In addition, given any dominating set of size x for G'' , the method given in Step (v) above constructs a dominating set of size at most $\lceil x/\beta \rceil$ for G' . Therefore, given any dominating set D'' for G'' , where $|D''| \leq \delta k''$ for some factor δ , Step (v) can be used to produce a dominating set D' for G' , with $|D'| \leq \delta k'$. Hence, the approximation algorithm for the minimum dominating set problem has a ratio bound of $(1 + 1/\alpha)\rho(4q \lceil \alpha \log_2 q \rceil) - (1/\alpha)$ on graphs with q nodes. \square

There are several negative results indicating that approximation algorithms with constant or even logarithmic ratio bounds are not likely to exist for the minimum set cover, the minimum hitting set and the minimum dominating set problems [2, 5, 12, 13]. These problems are closely related, and the nonapproximability results of any one apply to

others too. Specifically, the following results are known for the minimum dominating set problem.

Lemma 11 (Arora et al. [2] and Feige [5]). *Let $q = |V'|$ be the number of nodes in the undirected graph G' .*

- (i) *There exists a constant $\varepsilon > 0$ such that no polynomial-time approximation algorithm for the minimum dominating set problem achieves a ratio bound of $1 + \varepsilon$, unless $P = NP$.*
- (ii) *There exists no polynomial-time approximation algorithm for the minimum dominating set problem with a ratio bound of $(1 - \varepsilon) \ln q$ for any $\varepsilon > 0$, unless $NP \subset DTIME[q^{\log_2 \log_2 q}]$.*

Since α can be arbitrarily large, the results for the minimum alarm set problem follow from Lemmas 10 and 11.

Theorem 12. *Let $n = |V|$ be the number of nodes in the condensed, transitively reduced, directed acyclic graph G .*

- (i) *There exists a constant $\varepsilon > 0$ such that no polynomial-time approximation algorithm for the minimum alarm set problem achieves a ratio bound of $1 + \varepsilon$, unless $P = NP$.*
- (ii) *There exists no polynomial-time approximation algorithm for the minimum alarm set problem with a ratio bound of $(1 - \varepsilon) \ln n$ for any $\varepsilon > 0$, unless $NP \subset DTIME[n^{\log_2 \log_2 n}]$.*

7. Minimum test collection problem

In this section we explore the connection between the minimum alarm set problem and the minimum test collection problem studied in [8]. The problem can be stated formally as follows [4]:

Instance: A collection C of subsets of a finite set S .

Solution: A subcollection $C' \subseteq C$ such that for each pair of distinct elements $x_1, x_2 \in S$ there is some set $c \in C'$ that contains exactly one of x_1 and x_2 .

Measure: Cardinality of the subcollection, i.e., $|C'|$.

The minimum test collection problem appears to be a fundamental problem and has been used to solve a number of other problems related to fault diagnosis under the model proposed originally by Mayeda and Ramamoorthy [15]. The minimum test collection problem is known to be NP-hard. An approximation algorithm with a ratio bound of $0.31 + 2 \ln |S|$ is presented in [9], by reducing it to the *set cover problem*. We will now demonstrate a reduction from the minimum alarm set problem to the minimum test collection problem to show that approximation algorithms with better ratio bounds are not likely to exist for this problem also.

Let $G=(V,E)$ be a condensed, transitively reduced, directed acyclic graph with $|V|=n$. For each $v \in V$, compute $T(v)$, the set of nodes from which v is reachable. For the minimum test collection problem, let $S=V$. Also let $C = \{T(v): v \in V\}$.

Suppose A is a feasible solution of size k for the minimum alarm set problem. Suppose v in A distinguishes the set of faulty nodes $\{a,b\}$. Then $T(v)$ includes either a or b , but not both. Thus, the collection of k subsets $T(v)$ for each $v \in A$ is a feasible solution for the minimum test collection problem. Conversely, suppose C' is a feasible solution for the minimum test collection problem. Then the set of nodes v such that $T(v)$ belongs to the collection C' distinguishes every pair of faulty nodes. However, the alarm set formed now does not guarantee that every single fault will be detected, i.e., the set of ringing alarms is nonempty for every faulty node.

We therefore modify the reduction so that S includes another element which does not appear in any subset in the collection C . The effect of this change is to ensure that each $v \in V$ is a member of at least one subset in the collection C' formed as a feasible solution for the minimum test collection problem. In other words, the set of ringing alarms is nonempty for every faulty node. A feasible solution for the minimum alarm set problem still leads to a feasible solution of the same size for the minimum test collection problem as before. The following results are an obvious consequence of the reduction shown above and Theorem 12.

Lemma 13. *Suppose the minimum test collection problem has a polynomial-time approximation algorithm with a ratio bound of $\rho(m)$, where $m=|S|-1$, and ρ is a monotonically increasing positive function. Then the minimum alarm set problem has a polynomial-time approximation algorithm with a ratio bound of $\rho(n)$ on directed graphs with n nodes.*

Theorem 14. *Let $m=|S|-1$.*

- (i) *There exists a constant $\varepsilon > 0$ such that no polynomial-time approximation algorithm for the minimum test collection problem achieves a ratio bound of $1 + \varepsilon$, unless $P=NP$.*
- (ii) *There exists no polynomial-time approximation algorithm for the minimum test collection problem with a ratio bound of $(1 - \varepsilon) \ln m$ for any $\varepsilon > 0$, unless $NP \subset DTIME[m^{\log_2 \log_2 m}]$.*

8. Concluding remarks

In this paper, we consider systems that can be modeled as directed acyclic graphs such that nodes represent components and directed edges represent fault propagation between components. This model has wide applicability, since faults and hence their propagation can be interpreted in a number of ways depending on the system considered – failure of components, errors in computation, diseases, etc. Therefore, the problem of alarm placement so that faulty components can be detected and uniquely diagnosed

is of practical importance. In this paper, we first showed that the alarm minimization problem is intractable, i.e., NP-hard, even when restricted to three level graphs in which all nodes have outdegree two or less. We then focused on three special classes of graphs, and presented optimal algorithms for these classes. We also presented a polynomial-time approximation algorithm for the general case that guarantees that the ratio of the number of alarms placed to the optimum required is within $0.31 + 2 \ln n$, where n is the number of nodes in the graph. Moreover, by showing a reduction from the dominating set problem to the alarm placement problem, we argued that it is unlikely that there exists a polynomial-time algorithm that approximates the optimal number of alarms within a ratio of $(1 - \varepsilon) \ln n$ for any $\varepsilon > 0$. Finally, we demonstrated the connection between the minimum alarm set problem and the minimum test collection problem, and proved similar results. Study of approximation algorithms for *system-level fault diagnosis* appears to be a valuable area for future research. One of the referees of this paper has pointed out that the approximation algorithm proposed in Section 6 has striking similarities to one proposed in [17] for a different problem in the context of distributed detection networks.

Acknowledgements

The authors thank the anonymous referees for many constructive comments that greatly improved the presentation of this paper. In particular, [17] was brought to their attention by one of the referees.

References

- [1] A. Aho, M.R. Garey, J.D. Ullman, The transitive reduction of a directed graph, *SIAM J. Comput.* 1 (1972) 131–137.
- [2] S. Arora, C. Lund, R. Motwani, M. Sudan, M. Szegedy, Proof verification and intractability of approximation problems, *Proc. 33rd IEEE FOCS*, 1992, pp. 14–23.
- [3] T.H. Cormen, C.E. Leiserson, R.L. Rivest, *Introduction to Algorithms*, MIT Press, Cambridge, MA, 1990.
- [4] P. Crescenzi, V. Kann, A compendium of NP optimization problems, [ftp.nada.kth.se, pub/documents/Theory/Viggo-Kann/compendium.ps](http://ftp.nada.kth.se/pub/documents/Theory/Viggo-Kann/compendium.ps), 1995.
- [5] U. Feige, A threshold of $\ln n$ for approximating set cover, *Proc. 28th Ann. ACM STOC*, Philadelphia, PA, 1996, pp. 314–318.
- [6] M.R. Garey, D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, CA, 1979.
- [7] F. Harary, R.Z. Norman, D. Cartwright, *Structural Models: An Introduction to the Theory of Directed Graphs*, Wiley, New York, 1965.
- [8] T. Ibaraki, T. Kameda, S. Toida, NP-Complete diagnosis problems on system graphs, *Trans. IECE Japan* E-62 (1979) 81–88.
- [9] V. Kann, On the approximability of NP-complete optimization problems, Ph.D. Thesis, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, 1992.
- [10] M. Kokawa, S. Miyazaki, S. Shingai, Fault location using digraph and inverse direction search with applications, *Automatica* 19 (1983) 729–736.
- [11] M. Kokawa, S. Shingai, Failure propagating simulation and nonfailure paths search in network systems, *Automatica* 18 (1982) 335–342.

- [12] P.G. Kolaitis, M.N. Thakur, Approximation properties of NP minimization classes, Proc. 6th Ann. Conf. on Structures in Complexity Theory, 1991, pp. 353–366.
- [13] C. Lund, M. Yannakakis, On the hardness of approximating minimization problems, JACM 41 (1994) 960–981.
- [14] M. Malek, K.Y. Liu, Graph theory models in fault diagnosis and fault tolerance, Design Automation Fault-Tolerant Comput. 3 (1979) 155–169.
- [15] W. Mayeda, C.R. Ramamoorthy, Distinguishability criteria in oriented graphs and its application to computer diagnosis-I, IEEE Trans. Circuit Theory 16 (1969) 448–454.
- [16] F.P. Preparata, G. Metze, R.T. Chien, On the connection assignment problem of diagnosable systems, IEEE Trans. Comput. 16 (1967) 848–854.
- [17] N.S.V. Rao, Computational complexity issues in synthesis of simple distributed detection networks, IEEE Trans. Systems Man and Cybernet 21 (1991) 1071–1081.
- [18] N.S.V. Rao, Expected-value analysis of two single fault diagnosis algorithms, IEEE Trans. Comput. 42 (1993) 272–280.
- [19] N.S.V. Rao, Computational complexity issues in operative diagnosis of graph-based systems, IEEE Trans. Comput. 42 (1993) 447–457.
- [20] J. Shiozaki, H. Matsuyama, E. O’Shima, M. Ira, An improved algorithm for diagnosis of system failures in the chemical process, Comp. Chem. Eng. 9 (1985) 285–293.