22

Centralized vs Distributed Fault Localization

I. Katzela [1] *CTR-Columbia University* [2], *New York, NY 10027-6699, USA, tel: (212)854-7378, e-mail: irene@ctr.columbia.edu*

A.T. Bouloutas [3] *First Bank of Boston, Boston MA 0216, USA, tel: (617)434-0534 .*

S.B. Calo IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA, tel:(914)784-7514 , e-mail:calo@watson.ibm.com

Abstract

In this paper we compare the performance of fault localization schemes for communication networks. Our model assumes a number of management centers, each responsible for a logically autonomous part of the whole telecommunication network. We briefly present three different fault localization schemes: namely, "Centralized", "Decentralized" and "Distributed" fault localization, and, we compare their performance with respect to the computational effort each requires and the accuracy of the solution that each provides.

1 INTRODUCTION

Usually, a single fault in a large network results in a number of alarms, and it is not always easy to identify the primary source(s) of failure. The problem of fault management becomes even worse when several faults occur coincidentally in the telecommunication network. The fault management process can be divided into three stages: alarm correlation, fault identification, and testing. The first two stages, usually referred to as the fault localization process, correlate the fault indications (alarms) received from the managed objects and propose various fault hypotheses. In the third stage each of the proposed hypotheses is tested in order to localize the fault precisely. The fault localization process is important because the speed and accuracy of the fault management process are heavily dependent on it.

In the past a number of researchers addressed the problem of fault localization in communication networks (Bouloutas, 1992), (Wang, 1993), (Shroff, 1989), (Riese, 1991) [4]. Most of the proposed methods focus on centralized algorithms for fault localization. However, the growth in size and complexity of communication networks may require the partitioning of the management environment into a number of management domains in order to meet organizational and performance requirements. This transition from a centralized management paradigm to a distributed one will require the development of distributed algorithms for fault localization. A distributed fault management approach will be able to shield parts of the network management system from information that is not locally useful, a very impor-

[1] Work done during the author's internship at the IBM T. J. Watson Research Center, NY, Summer 93.
[2] CTR - Center for Telecommunications Research
[3] Work done while the author was with the IBM T. J. Watson Research Center, NY.
[4] Additional references in the area can be found in (Katzela, 1993)

A. S. Sethi et al. (eds.), *Integrated Network Management IV*
© Springer Science+Business Media Dordrecht 1995

tant function as management centers tend to overflow with information. However, problems that involve objects in more than one domain will have to be resolved collectively by many domain managers in a distributed fashion. This introduces a number of problems that make the design of distributed fault management solutions a challenging task.

This paper is organized as follows: in Section 2 we define the problem of distributed fault localization and present a suitable model for the system; in Section 3 we present three different approaches for distributed fault localization; in section 4 we compare the proposed approaches with respect to the computational effort each requires and the accuracy of the solution each provides; and finally, section 5 concludes the paper. with a summary of the results.

2 MODEL OF THE SYSTEM

We assume a distributed approach to managing communication networks (Katzela, 1993). Each communication network is partitioned into a number of static, disjoint, logically autonomous management domains. Each domain is managed by a single management process [5] which is responsible for it. The managed objects in a domain may or may not be visible to managers of other domains. Hence, each manager has a limited view of the status of other management domains; and, has partial and incomplete information about the state of the network. However, managers from different domains can communicate and exchange information about the status of their domains using a peer to peer type of communication. Each domain manager has adequate knowledge about adjacent managers and domains so that communication can be established and messages passed.

Faults manifest themselves as alarms and alerts that are emitted by the managed objects affected by the fault(s). Alarms are communicated to the managed object's domain manager, which is responsible for identifying the primary source(s) of failure and eventually correcting the fault(s). Each received alarm represents the fault from the point of view of the object that emitted the alarm, and therefore corresponds to partial information about the fault. It is the responsibility of the fault localization system to collect all these partial views of a fault, correlate the information, and infer the real cause(s) of the fault. It is not unusual for an alarm to appear in a managed object belonging to a particular management domain and to indicate a fault in another managed object in a different domain. Since alarms cross management domains, management centers have to collaborate in order to infer the real state of the system.

Note that throughout the paper we assume that faults affect only managed objects but do not affect managers, information transfer processes of management systems, or other parts of the Telecommunication Management Network (TMN). This is a reasonable assumption and it stems from the fact that usually TMN systems have much stricter reliability requirements than the rest of the network.

Before we proceed it is essential to examine the structure of the alarms. Each alarm is characterized by the *domain of the alarm*, which is defined as the set of all independent

[5] In this paper we will use the terms management process, manager and management center interchangeably.

managed objects that could have caused the alarm - in other words, all the managed objects that might be at fault. Note that the domain of an alarm should not be confused with the domain of a management center. The domain of an alarm depends both on the semantics of the alarm and the topology of the communication network. It is the management centers' responsibility to find the domain of a received alarm before proceeding to the fault localization process.

3 FAULT LOCALIZATION APPROACHES

Assume that at a given moment in time a number of alarms appear in a communication network. The objective is to design algorithms that are able to find the *"best"* explanation of the received alarms, i.e., the managed object or the set of managed objects that could have been at fault and caused the alarms. In principle all the managed objects that appear in the domain of a received alarm constitute a possible explanation of it. If two or more alarms share an intersection, these alarms should be examined together because it is more probable that they are caused by the same set of faults. That is the reason why we introduce the notion of a *cluster of alarms.* A *cluster of alarms* is defined as a set of alarms that have intersecting domains. Note that a cluster may span more than one management domain as the alarms that comprise the cluster span more than one domain (Katzela, 1993).

Each cluster of alarms may have a number of explanations. The fault localization algorithm should be able to choose the *"best"* (most probable) among the possible ones. One way to find the most probable explanation would be to associate a probability of failure with each managed object. Then the *"best"* explanation would be the set of managed objects whose combined probability of failure is maximum. Instead of assigning a probability of failure to each managed object we can associate an *"information cost"* which is defined as the negative of the logarithm of the probability of failure for the managed object. For independent faults, the information costs are equivalent to probabilities, but working with them has certain advantages. If we choose to work with information costs then the *"best"* explanation (most probable) will be the set of managed objects whose sum of information costs is minimum (Bouloutas, 1992).

Before we proceed to examine distributed fault localization approaches, we assume that there exists a centralized algorithm that is able to find the most likely errors in a set of managed objects, given a set of alarms and the information costs associated with each managed object. As was shown in (Katzela, 1994) the fault localization problem is NP-Complete. Thus, in general there is no polynomial algorithm that gives the exact solution. One could then either construct a polynomial algorithm that gives an approximate solution, or a polynomial algorithm that gives the correct solution with some probability. In (Katzela, 1993), we present a possible probabilistic algorithm which finds an exact solution if the number of faults in the system is less than k, a parameter. We represent this algorithm by $G_p(A, N, k)$ where A is the received alarm cluster, N is the set of managed objects associated with A, and k is the maximum number of concurrent failures that can be identified by the algorithm. The probabilistic algorithm fails to give a solution with probability Q which is equal to the probability that there are more than k concurrent faults in the system. Hence:

$$Q = P_r(\text{algorithm fails}) \;=\; P_r(> k \text{ faults in the network}) \leq 1 - \sum_{i=0}^{k} b(i; N, p) \qquad (1)$$

Each managed object has a probability of failure assigned to it, p is the maximum of all such probabilities for the managed objects associated with the received alarm cluster A, and $b(i; N, p)$ is the probability of N Bernoulli trials, with probability p of success and i successes.

Since each management center can resolve problems that affect managed objects in its domain we will only examine the case where faults affect objects in many management domains. For simplicity we assume that the entire network consists of two domains. A generalization of the results for multiple domains is discussed in (Katzela, 1993). We also, without loss of generality, assume that there is only one cluster of alarms that crosses the boundary between domains.

3.1 Centralized Localization

The first approach, namely *Centralized Localization*, assumes the existence of a central manager that oversees all the domain managers and has a global view of the network. Problems that affect more than one domain could be resolved directly by the central manager as if there were no domain managers. In other words, if the received alarm cluster spans more than one domain, then the domain managers take no action and the central manager uses a centralized algorithm like $G_p(A, N, k)$ to identify the failure. The *Centralized Localization* approach always guarantees to output the optimum explanation of the received alarms. It fails to output an explanation of the received alarms with probability Q given by (1).

3.2 Decentralized Localization

The second approach, namely *Decentralized Localization*, assumes the existence of a central manager that oversees all the domain managers. Problems that affect more than one domain could be resolved in a collaborative way between the central manager and the domain managers. Unlike the first approach, the second one does not require extensive involvement of the central manager.

Assume that m of the L received alarms cross the boundary between the domains. These m alarms might have been produced by a fault in either domain and there is no a-priori information as to whether an alarm that crosses the boundary is explained by a fault in the first domain or by a fault in the second domain, or both. There are 2^m possible explanations for these m alarms depending on whether faults in domain one or domain two explain the alarms. Each domain manager calculates, using perhaps the $G_p(A, N, k)$ algorithm, 2^m optimum solutions, one for each of the possible explanations of the m alarms that cross the boundary.The central manager receives the 2^m optimum solutions from the two domain managers and finds the compatible ones. Two partial solutions are compatible if all the alarms received by both domains are explained in the final solution. Then the central manager selects the compatible global solution of minimum information cost. As one can easily verify, the above described procedure is able to identify the optimum solution given

that the two management domains can find the optimum solution for managed objects and alarms in their respective domains (Katzela, 1993).

Finally, since each domain manager uses the probabilistic algorithm for fault identification there is a probability Q' (which will be calculated in section 4.1) that the *Decentralized Localization* fails to output a solution.

3.3 Distributed Localization

The third approach, namely *Distributed Localization* , does not assume the existence of a central manager. The area of the network is divided into two domains, each managed solely by a single domain manager. This strategy tries to find the faults from the point of view of each of the two domain managers without the use of a central manager. Let us examine the problem from the point of view of domain manager one. For each alarm that crosses the boundary, domain manager one would like to associate a probability that this alarm is explained by *Domain Two.* One way to do that would be to represent all the managed objects that belong to *Domain Two* and are associated with the alarm by a *proxy node.* Failure of the proxy node would indicate that some managed objects in *Domain Two* have failed. If one could associate a probability of failure with the proxy node, then the management proces of *Domain One* could use the probabilistic algorithm $G_p(A, N, k)$ to solve a centralized problem in a new expanded domain that includes the managed objects in its domain plus m proxy nodes, one for each alarm that crosses the boundary between the two domains. Here all the alarms that cross the boundary are treated as regular alarms. Once the algorithm has output the optimum solution there will be some alarms that are explained by regular nodes in *Domain One* and some alarms that are explained by *proxy nodes.* The alarms that are explained by *proxy nodes* are the alarms that are not explained by *Domain One* and are hopefully explained by *Domain Two.* The global solution is the one that includes all the regular nodes that appear in the optimum solutions of the two domain managers.

The exact probability of failure for a proxy node is difficult to find since it depends on all the managed objects and all the alarms in the cluster. Thus, the exact calculation of the probability of a proxy node is an NP-complete problem (Katzela, 1993). The best we can achieve is an estimation of the probability of failure for a *proxy node.* The estimated probability for a *proxy node* differs from the exact value by an estimation error. As a result of the estimation errors the distributed localization approach does not always guarantee an optimum global solution (Katzela, 1993).

4 PERFORMANCE COMPARISON

The objective of any fault management process is to minimize the time to localize a fault. The time to localize a fault is the sum of the time to propose possible hypotheses of the fault and the time to do testing in order to verify these hypotheses. Thus, we should minimize the time to perform fault identification and the time to perform testing. The time to identify the fault is affected by the identification algorithm. Hence, the first objective is to minimize the time complexity of the identification algorithm. The second objective is to minimize the time of testing. The time of testing is affected by the number of managed

objects that need to be tested which is equal to the number of proposed hypotheses by the identification algorithm. If the management process is able to identify correctly the source of failure, the minimum number of tests is required. Thus, minimizing the number of tests is equivalent to minimizing the number of fault hypotheses, or equivalently, maximizing the accuracy of the fault identification algorithm. Thus, the performance measures of interest are the *time complexity* of the identification algorithm and the *accuracy* of the solution that the identification algorithm provides.

The complexity of the identification algorithm for each of the proposed approaches is a function of the number of nodes associated with the received alarm cluster, the number of alarms that cross domains, and the parameter k of the probabilistic algorithm that is the base of all the proposed localization schemes. On the other hand, the accuracy of each approach depends on the error in the estimations of the probabilities of failure for the managed objects associated with the received cluster of alarms.

4.1 Comparison between Centralized and Decentralized Fault Localization

Assume that the received cluster of alarms, A, is associated with N managed objects that may fail, each with probability p. We would like to compare the performance of the centralized versus the decentralized approach for this network setting. For the centralized approach, the central manager process should use the probabilistic algorithm $G_p(A, N, k)$. For the decentralized approach we assume that we partition the managed system in two domains namely *Domain One* (D_1) and *Domain Two* (D_2). Also we assume that there are m alarms that cross the boundaries between the two domains, and the N managed objects associated with the received cluster of alarms A are partitioned into N_1 objects in D_1 and N_2 in D_2, such that $N = N_1 + N_2$. According to the decentralized approach, each domain manager uses the probabilistic algorithm 2^m times for its area in order to find 2^m optimum solutions, one for every possible interpretation of the alarms that cross the boundary. In each case, D_1 will use $G_p(A_1, N_1, k_1)$ and D_2 will use $G_p(A_2, N_2, k_2)$ to identify the optimum solution. A_1 is the set of alarms that domain manager one takes into account in this instance, A_2 is the set of alarms that domain manager two takes into account in this instance, k_1 is the number of faults that manager one must localize, and k_2 is the number of faults that manager two must localize.

The selected performance measures for comparing these approaches are *accuracy* and *time complexity*. The *accuracy* performance measure has two aspects: The difference between the information cost of the proposed solution and the optimum one; and, the probability that the approach fails to give a solution. By design, both the centralized and the decentralized approaches give the optimum cost solution whenever they give a solution. Thus, we need to discuss only the second aspect of *accuracy*. The centralized approach fails to find a solution with probability Q, which is given by (1). The decentralized approach fails to find a solution with probability Q' which is:

$$\begin{aligned} Q' &= P_r(\text{Decentralized approach fails}) \overset{(1)}{=} Pr(> k_1 \quad \text{faults in } D_1) + \\ &\quad Pr(> k_2 \quad \text{faults in } D_2) - Pr(> k_1 \quad \text{faults in } D_1) \cdot Pr(> k_2 \quad \text{faults in } D_2) \quad (2) \end{aligned}$$

Regarding *time complexity*, the centralized approach has complexity which is bounded by $C_{cen} = O(N^k)$ and the decentralized approach by $C_{dec} = O(2^m \max(N_1^{k_1}, N_2^{k_2}) + 2^{2m})$.

It is obvious that the accuracy of the decentralized approach increases with an increase in the values of k_1 and k_2, which also leads to an undesireable increase in the time complexity of the approach. Suppose that we fix the accuracy of the two approaches and then compare them with respect to time complexity. For a given k (number of faults that the centralized approach can localize) the accuracy of the centralized approach is fixed. We need to calculate the values of k_1 and k_2 such that the two probabilities are equal - the probability that the decentralized approach fails and the probability of failure for the centralized approach, in order to achieve the same accuracy for both approaches. In addition the decentralized approach should be able to identify at least as many faults as the centralized one. Hence:

$$\begin{aligned} P_r(\text{Decentralized approach fails}) &\leq P_r(\text{Centralized approach fails}) \\ \text{and} \quad k_1 + k_2 &\geq k \end{aligned} \tag{3}$$

The unknowns in (3) are the parameters k_1 and k_2. Typically it is difficult to solve such a set of inequalities. In order to simplify our analysis we will propose an approximation for calculating the parameters k_1 and k_2.

Approximate Calculation for k_1 and k_2

As an approximation we assume that the number of faults each domain manager should localize is proportional to the number of managed objects in the alarm cluster that belong to its domain. This assumption is valid and stems from the fact that the managed objects in the system fail independently. Hence :

$$\frac{k_1}{k_2} = \frac{N_1}{N_2} \overset{(3)}{\Rightarrow} k_1 + k_1 \frac{N_2}{N_1} \geq k \Rightarrow k_1 \geq \lceil k \frac{N_1}{N} \rceil \tag{4}$$

We approximate the constraint that the probability of failure for the decentralized approach should be less than or equal to the probability of failure for the centralized approach with the following two requirements:

$$\begin{aligned} P_r(> k_1 \text{ faults in } D_1) &\leq \frac{N_1}{N} P_r(\text{Centralized approach fails}) \\ P_r(> k_2 \text{faults in} D_2) &\leq \frac{N_2}{N} P_r(\text{Centralized approach fails}) \end{aligned} \tag{5}$$

The proposed approximation has decomposed the original complex problem in (3) into two simpler problems, one for each domain. Without loss of generality it is sufficient to solve the problem only for D_1. The results are equivalent for D_2.

The new problem for domain one can be stated as follows: *Given a probability of failure for the decentralized approach, what is the value of the parameter k_1 that domain manager one should use in the application of the probabilistic algorithm, so that the following two inequalities hold?*

$$\begin{aligned} P_r(> k_1 \text{ faults in } D_1) &\leq \frac{N_1}{N} P_r(\text{Centralized approach fails}) \\ k_1 &\geq \lceil k \frac{N_1}{N} \rceil \end{aligned} \tag{6}$$

Which is equivalent to the system:

$$\begin{aligned} \overset{(1)}{\Rightarrow} \sum_{i=k_1+1}^{N_1} b(i;N_1,p) &\leq \frac{N_1}{N} \sum_{i=k+1}^{N} b(i;N,p) \\ k_1 &\geq \lceil k\frac{N_1}{N} \rceil \end{aligned} \tag{7}$$

The system of inequalities in (7) is still difficult to solve. We would like to simplify it and find a closed form solution for k_1. In order to simplify (7) we should find a simpler expression for $\sum_{i=k_1+1}^{N_1} b(i;N_1,p)$. The form of the expression depends on whether k_1 is $\overset{<}{>}$ $\lceil (N_1+1)p \rceil$ (Katzela, 1993). Table 1 summarizes the formulas for estimating k_1 in each case.

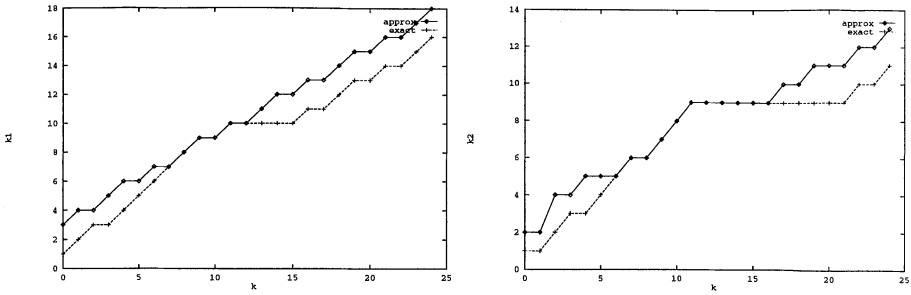
Table 1 Formulas for estimating the parameter k_1

$\lceil (N_1+1)p \rceil < \lceil k\frac{N_1}{N} \rceil$	$k_1 > \max\left[\lceil \log_\beta \left(\frac{\frac{N_1}{N}\sum_{i=k+1}^{N} b(i;N,p)}{\Theta} \right) \rceil, \lceil k\frac{N_1}{N} \rceil \right]$
$\lceil (N_1+1)p \rceil > \lceil k\frac{N_1}{N} \rceil$, N_1 large	$k_1 > \max\left[\lceil (\eta - \frac{\frac{N_1}{N}\sum_{i=k+1}^{N} b(i;N,p)}{b(\lceil (N_1+1)p \rceil;N_1,p)} \rceil, \lceil k\frac{N_1}{N} \rceil \right]$
$\lceil (N_1+1)p \rceil > \lceil k\frac{N_1}{N} \rceil$, N_1 small	$k_1 > \max\left[\lceil 2\eta - \frac{\frac{N_1}{N}\sum_{i=k+1}^{N} b(i;N,p)}{b(\lceil (N_1+1)p \rceil;N_1,p)} \rceil, \lceil k\frac{N_1}{N} \rceil \right]$
$\beta = \frac{p}{1-p}$, $\eta = (N+1)p+1$	$\Theta = \binom{N}{N/2}(1-p)^{N+1}\eta$

The formulas in Table 1 provide an overestimation of the value of k_1. Similarly we can calculate k_2. We simply substitute, in the appropriate formulas in Table 1, N_2 for N_1 and k_2 for k_1.

As an example, consider a network scenario where the received alarm cluster A is associated with $N = 100$ managed objects. Each of these objects has a probability of failure $p = 0.01$. We also assume that these N objects are partitioned into the two domains so that $\frac{N_1}{N_2} = \frac{3}{2}$. As is shown in Figure 1, by using the formulas in Table 1, we get k_1 and k_2 close to their exact values. Thus, the overestimation of k_1 and k_2 is very small and the approximation works satisfactorily. A similar behavior, small overestimation, is observed for different values of p. Finally as we stated earlier, the decentralized approach should be able to localize, in total, an equal or larger number of faults. As we can observe from the curves in Figure 1, sometimes the decentralized approach has to localize as much as twice the number of faults as the centralized approach in order to achieve the same probability of failure. Such an increase in the parameters k_1 and k_2 results in increased complexity for the decentralized approach.

A specific problem instance, N_1 managed objects in domain one and N_2 managed objects in domain two, could occur with probability $P_r(N;N_1,N_2)$. It is easy to show that the



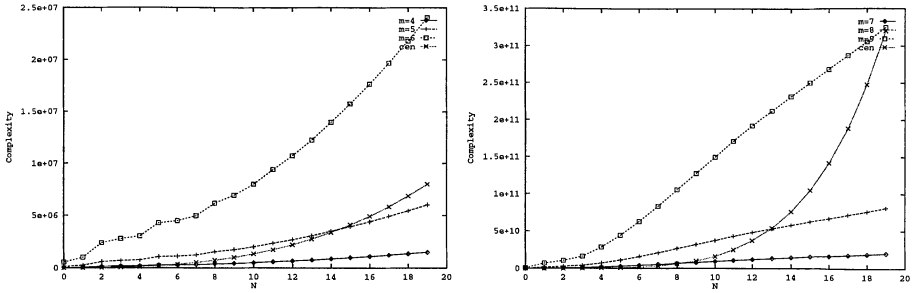
(a) N =100, p = 0.01 (b) N = 100, p = 0.01

Figure 1: Estimation and exact values of k_1, k_2 for different values of k

average complexity per domain manager per problem instance is: $\frac{C_{dec}^{aver}}{problem} = \sum_{i=1}^{N} P_r(N; i; N - i)i^{k_i}$, where k_i is the number of faults that the domain manager can localize in this specific instance. For fixed probability of failure for the decentralized approach, k_i can be calculated by use of Table 1. In principle, the probability of a specific problem scenario, $P_r(N; N_1, N_2)$, could follow any distribution. The value of $P_r(N; N_1, N_2)$ for a specific problem instance indicates how *likely* it is that this problem will be encountered. The concept of *likely* is related to the probability distribution that adequately represents the set of problems to be encountered. Such a probability distribution is difficult to define and analyze. Let us assume for simplicity that the partitioning of N managed objects associated with the received alarm cluster between the two domains is done randomly. Then the probability distribution of the problems to be encountered is a generalized Bernoulli distribution. Thus $P_r(N; N_1, N_2) = \frac{N!}{N_1!N_2!} = \frac{N!}{N_1!(N-N_1)!}$ and the average time complexity of the algorithm will be:

$$C_{dec}^{aver} = 2^m \sum_{i=1}^{N} \frac{N!}{i!(N-i)!} \max\left(i^{k_{i1}}, (N-i)^{k_{i2}}\right) + 2^{2m} = 2^{2m} \sum_{i=0}^{\lceil \frac{N}{2} \rceil} \frac{N!}{i!(N-i)!} i^{k_i} + 2^{2m} \quad (8)$$

We are interested in investigating the conditions under which $C_{dec}^{aver} < C_{cen}$. The average complexity of the decentralized approach depends on the values of N, k, p and m. The parameter m is the one which has the greatest effect on the complexity of the decentralized approach. As we can easily observe from Figure 2, for fixed k and p the time complexity of the decentralized approach remains less than that of the centralized approach up to a certain value of m. For example in Figure 2(a) , for $k = 3$ the complexity of the decentralized approach remains strictly less than that of the centralized approach when $m \leq 4$. Similarly in Figure 2(b) for $k = 5$, the same behavior is observed for $m \leq 7$. Finally, as we could see in Figure 2(b) for the same value of k the allowed number of alarms m that can cross domains so that the complexity of the decentralized approach is less than that of the centralized one increases with a decrease in the probability p. For example for $k = 5$ and $p = 0.1$ in Figure 2(b) the allowed maximum m is $m = 7$, for $p = 0.01$ the maximum m is $m = 5$ in Figure 3(a), for $p = 0.001$ in Figure 3(b) $m = 2$.



(a) k=3, p = 0.1 (b) k=5, p = 0.1

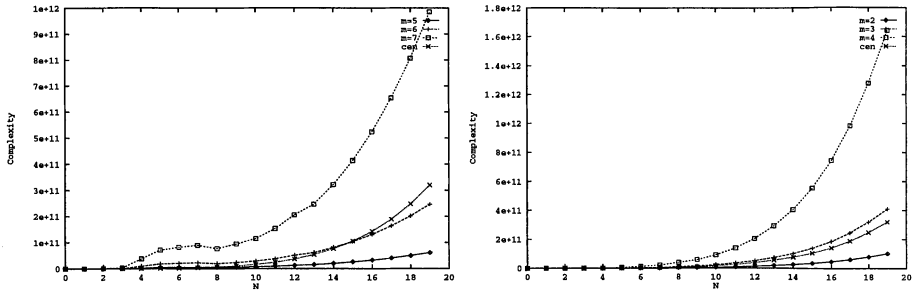
Figure 2: Average Complexity of Decentralized Approach versus N, for different m and k and the same p. The curve *cen* represents the complexity of the Centralized Approach vs N.

4.2 Comparison between Centralized and Distributed Fault Identification

It is easy to show that the complexity of the distributed approach is always considerably less than the complexity of the centralized approach ($max(N_1^{k_1}, N_2^{k_2}) << N^k$). Hence, it remains to compare the approaches with respect to accuracy. The first aspect of accuracy is the probability of failure of the localization schemes to output a solution. Again we can select k_1 and k_2 for the two domains of the distributed approach so that the probabilities of failure for the centralized and distributed approaches are the same. The corresponding values of k_1, k_2 can be calculated by the use of Table 1.

A possible solution for a received alarm cluster is characterized by its information cost which is the sum of the information costs of the managed objects that are included in the solution. Most of the time there are more than one possible solutions. All of the localization approaches discussed in the previous sections select among all the possible solutions the one which has the minimum information cost. The deviation of any solution from the optimum one is characterized by the difference in the information cost of the solution from the information cost of the optimum (minimum cost) solution. Unlike the centralized and the decentralized localization approaches, the distributed localization does not always guarantee that it can find the optimum solution. This deviation from the optimum solution stems from errors in the approximation of the probability of failure for the proxy nodes.

The exact probability of failure for a proxy node (and thus the exact information cost for the node) is difficult to find since it depends on all the managed objects and all the alarms in the cluster. The best we can achieve is an estimation of the information cost of a proxy node. Thus, the information cost of the proxy node differs from its exact value by an estimation error. The introduction of estimation errors might cause a difference between the solution proposed by the distributed identification algorithm and the optimum one which is given by the decentralized and the centralized approaches. It is of interest to find a bound for the difference between the information cost of the distributed solution and the optimum one. Also it is of interest to investigate how sensitive the distributed solution is to changes



(a) k=5, p = 0.01 (b) k=5, p = 0.001

Figure 3: Average Complexity of Decentralized Approach versus N, for different p, m and the same k. The curve *cen* represents the complexity of the Centralized Approach versus N.

in the information cost of the proxy nodes.

For a given network let I^* be the information cost of the optimum solution; I_{cen}, I_{dec} and I_{dis} the information costs of the solutions that the centralized, decentralized, and distributed localization schemes give, respectively. As we know, $I_{cen} = I_{dec} = I^*$. The objective is to find how close the solution that the distributed approach gives is to the optimum one. In other words we would like to find a bound for $\mid I^* - I_{dis} \mid$. As it was analyzed in (Katzela, 1993), the $\mid I^* - I_{dis} \mid$ is bounded as follows:

$$2 \cdot m \cdot e_{min} \leq \mid I^* - I_{dis} \mid \leq 2 \cdot m \cdot e_{max} \tag{9}$$

Where m is the number of alarms that cross the boundary between the two domains and e_{max}, e_{min} are the minimum and maximum probability estimation errors for all the proxy nodes in both domains. The difference between the information costs in the two approaches depends on the number of alarms m that cross the boundary between domains and the minimum and maximum probability estimation errors for all the proxy nodes in both domains. As it was shown in (Katzela, 1993) the estimation error for the weights of a proxy node is expected to be small. Hence, for small values of m , which is the expected case, the difference $\mid I^* - I_{dis} \mid$ will be also small. Thus, although the decentralized approach does not always guarantee to output the optimum solution, it provides a solution which has information cost close to the information cost of the optimum one.

5 CONCLUSIONS

In this paper we compare the performance between a number of fault localization approaches suitable for a distributed fault management environment. The three proposed methods are namely the *Centralized*, *Decentralized* and *Distributed* Fault Localization approaches. As measures of comparison we used the *accuracy* of the solution and the *complexity* of the identification process that each approach employs. Our comparison proved that the decentralized approach generally has considerably less complexity than the centralized approach, and can

provide the same or better solution accuracy. Also the distributed localization approach was proved to have the least complexity of all three schemes in all network settings, but it can not always guarantee an optimum solution. However, as was shown in the previous section, it provides a solution which is almost as accurate as the solution provided by the other two approaches.

5 REFERENCES

Bouloutas, A., Calo, S. and Finkel A. (1992) Alarm Correlation and Fault Identification in Communication Networks. *IBM Technical Report*, **RC 17967**.

Katzela, I., Bouloutas, A. and Calo, S. (1993) Comparison of Distributed Fault Identification Schemes in Communication Networks. *IBM Technical Report*, **RC 19656**.

Katzela, I. and Schwartz, M. (1994) Schemes for Fault Identification in Communication Networks. *CTR Technical Report*, **CU/CTR/TR 362-49-09**.

Riese, M. (1991) Model Based Diagnosis of Networks: Problem Characterization and Survey. *OEGAI-91 Workshop on Model Based Reasoning.*

Shroff, N. and Schwartz, M. (1989) Fault Detection/Identification in the Linear Lightwave Networks *CTR Technical Report*, **CU/CTR/TR 243-91-24**.

Wang, C. and Schwartz, M. (1993) Identification of Faulty Links in Dynamic-Routed Networks *IEEE JSAC*, **11**, 1449-60 .

5 BIOGRAPHY

Seraphin B. Calo received the M.S., M.A., and PhD. degrees from Princeton University, Princeton, New Jersey, in 1971, 1975, and 1976, respectively. Since 1977 he has been a Research Staff Member in the IBM Research Division at the Thomas J. Watson Research Center, Yorktown Heights, New York. He has worked and published in the areas of queueing theory, data communication networks, multi-access protocols, satellite communications, expert systems, and complex systems management. Dr. Calo joined the Systems Analysis Department in 1987, and is currently Manager of Systems Applications. This research group is involved in studies of architectural issues in the design of complex software systems, and the application of advanced technologies to systems management problems. Dr. Calo is involved with IEEE symposia related to networks and computer systems, and was instrumental in establishing the IEEE International Workshop on Systems Management.

Irene Katzela received the Diploma in Electrical Engineering from the National Technical University of Athens, Greece, in 1990 and the M.S. and MPhil degree from Columbia University, New York in 1993 and 1994 respectively. Currently she is working towards her PhD degree, in the area of fault management, at Columbia University. Since 1991 she is a Graduate Research Assistant at the Center for Telecommunication Research at Columbia University. Her other research interests include network management, design and verification of protocols and wireless networking. She is a student member of IEEE and a member of the National Technical Chambers of Greece.