Design of diagnostic structures using genetic algorithm

Abstract. The article deals with problems encountered during designing diagnostic structures (DS) type PMC and BGM for networks with a fixed logical structure (topology). The paper presents an adaptive method that uses a genetic algorithm, which allows the designing of the most inexpensive DS for a given topology.

Streszczenie. W artykule rozpatrzono problemy występujące przy projektowaniu struktur opiniowania diagnostycznego (OD) dla sieci komputerowych o ustalonej strukturze logicznej (topologii). Przedstawiono adaptacyjną metodę, wykorzystującą algorytm genetyczny, która umożliwia projektowanie najtańszych struktur OD dla danej topologii. (Projektowanie struktur diagnostycznej przy wykorzystaniu algorytmu genetycznego)

Keywords: system diagnosis, diagnostic opinion structures, PMC and BGM models, genetic algorithm, design.

Słowa kluczowe: diagnostyka systemowa, struktury opinowania diagnostycznego, modele PMC i BGM, algorytm genetyczny, projektowanie.

Introduction

As computer networks are expanding into areas requiring high dependability, self-diagnosability is becoming a feature of major importance in distributed computing systems. The increase of the dependability of distributed computer systems could be achieved by using effective methods of diagnosis and reconstruction of the system status after the failure - essentially, methods that utilize results of computers mutual testing. Systems capable to identify faulty units that belong to them are called self-diagnosable. An analysis of diagnostic capabilities of distributed systems is based on system-level fault diagnosis models [1], [2]. It is assumed that network units (computers or cluster of computers) are capable of testing themselves and other units. The process of identifying faulty network units is called diagnosis of the network. A network is said to be $t$-diagnosable if all faulty nodes (units) can be detected without replacement, provided that the number of faulty nodes does not exceed $t$.

One of a major issue encountered in designing and operating computer networks and processors with fault-tolerance is the definition of a proper diagnostic structure. A reasonable designed diagnostic structure is a prerequisite for achieving a highly efficient process for identifying faulty nodes. One of the most known and used diagnostic systems is the so-called method of diagnostic opinion [1, 2, 3, 4]. Inferring on the system’s efficiency takes place, in this case, on the basis test results of all the possible conducted tests, carried out by distinguished system units.

A diagnostic structure (DS) of the system is coherent digraph (oriented unigraph) $G = \langle E, U \rangle$ without a loop ($\langle e^c, e^d \rangle \in E \cup e \in E$), in which the arc $\langle e^c, e^d \rangle$ indicates that the node $e^c$ of the network shows an opinion (based on the test results) on the reliability state of the node $e^d$. A structure capable of identifying $t$ faulty elements and containing a minimal number of arcs is called an optimal $t$-diagnosable structure [3].

The diagnosability of the network is the maximum number of faulty nodes that can be identified by the diagnostic system. The diagnosability of the network depends on the diagnostic structure $G$ and a fault model, which determines expected test results once the node has become faulty. In the PMC model [1], a faulty node performing a test on another node will report unreliable results, and a fault-free node performing a test on another node will always produce correct test results. In [2] Barsi, Grandoni, and Maestrini introduced a model (known from now as the BGM model) in which a faulty node would always test faulty regardless of the condition of the testing processor.

In many situations, for selecting a rational structure it is necessary to take into account the general cost of certain reciprocal testing of telecommunication network elements, which in the general case cannot be considered equal. Then we can talk about the need to determine the most inexpensive structure, i.e. such, for which the total generalized cost of testing has a minimum value.

In the literature there are known analytical methods of determining PMC and BGM diagnostic structures DS [5, 6, 7, 8, 9]. In the studies [7, 8] presented was the adaptive method for determining the most inexpensive DOS. Proposed in [8] the adaptive algorithm is based on the decomposition of the tasks performed in two sequential subtasks: establishing the component of strong consistency and determining of the acyclic part of the structure. The limitation set out in the [8] algorithm operates on the cost of mutual testing of units specified for the full graph. In the work [10], is presented a system-level fault identification algorithm, using a parallel genetic algorithm, for diagnosing faulty nodes in large heterogeneous systems. The assumptions concerning test outcomes are the same as in the PMC model.

In the practice of exploitation of computer networks it is necessary to determine the most inexpensive DS for its established logical network structure (topology), which may take place after each network reconfiguration made after the diagnosis of particular faulty network nodes. This paper presents an adaptive design method of DSs taking into account the network topology. In addition, a generalization of the method has been proposed, allowing the determination of both PMC and BGM diagnostic structures.

Basic terminology

In order to clarify the terminology, basic terms have been presented below that are used further on in this paper.

Definition 1

Conditions are met for the $t$-diagnosable PMC diagnostic structure:

1. Necessary condition [4,11]:

   \[ |E| \geq 2 \cdot t + 1 \wedge (\forall e \in E; \mu(e) \geq t) \]

   where $|E|$ indicates the cardinality of set $E$ and $\mu(e)$ the input degree of node $e$, that is the number of elements, which tests the node $e$.

2. Sufficient condition [4,11]:

   \[ (\forall 0 \leq p \leq t - 1) \wedge (\forall E \subseteq C; |E'| = |E| - 2 \cdot t + p) \]

   where $t(E')$ indicates the set of consequents of $E'$ such that $t(E') \cap E' = \emptyset$.

Definition 2

Conditions are met for the $t$-diagnosable BGM diagnostic structure:
1. Necessary condition [4]:
\[(\forall e \in E; \mu^-(e) \geq t),\]
where \(\mu^-(e)\) indicates the input degree of node \(e \in E\).

2. Sufficient condition [2, 4]:
\[(\forall e', e'' \in E; \mu^-(e') = \mu^-(e'') = t) \land (e' \in \Gamma^{-1}(e'') \cap \Gamma(e'')),\]
\[(\exists e' \in \Gamma^{-1}(e'') \cap \Gamma(e'')); \Gamma^{-1}(e') \neq \Gamma^{-1}(e'')\),
where \(\Gamma(e)\) indicates the set of consequents of node \(e\) and \(\Gamma^{-1}(e)\) indicates the set of predecessors of node \(e\).

BGM class structure also includes the PMC class structure. However, in this case the BGM structure will be only understood as such DDS, which fulfill the precondition and sufficient condition for the BGM class structures, but are not PMC-type structures. In relation to the above-mentioned the pre-relaxation of the condition given earlier is sufficient for the form:
\[(\forall e', e'' \in E; \mu^-(e') = \mu^-(e'') = t) \land (\Gamma^{-1}(e') \cap \Gamma(e'') \neq \Gamma^{-1}(e')(e''))\]

**DS representation**

The general principle of the genetic algorithm (GA) used to support the generation of the diagnostic opinion structures for both PMC and BGM types is discussed in [7]. However, the algorithm presented in [7] operates on DSs in the form of stored bit vectors, without the inclusion of additional restrictions in the form of existing communication links. One of the ways of determining the heuristic method that is GA, to find solutions of a particular type (in this case structures based on a graph with defined edges), is the definition of chromosomes (bit vectors) so that they refer only to the permissible solutions (structures representing graphs in accordance with the given communication structure).

**Fig. 1. Converting the set of arcs to a binary vector**

Let us assume that the logical structure network (nodes and possible lines of communication between nodes) will be described in a ordinary graph \(G_N = \langle E, K \rangle\), where \(K\) is the set of edges representing possible paths of data transmission between nodes \((K \subseteq \{e', e'' \}; e', e'' \in E)\). Graph \(G_N = \langle E, K \rangle\) will be presented in a form of a neighbourhood matrix \(M_N\). The resulting matrix should be converted to an equivalent transition matrix \(M_T\), and then define the bijection \(T : U \rightarrow b\) converting the the set of arcs, indicated in matrix \(M_T\) in the elements of the bit vector \(b\). An example of such a transformation has been illustrated in Figure 1.

The resulting phenotype should be interpreted as a reference point for chromosome processed by the genetic algorithm, which will be used to design the diagnostic opinion structure. This means that all the chromosomes processed by the GA population will have the same number of genes as the bit vector, and only changes may affect the allele.

**Evaluation of Partial Solutions**

GA is an iterative algorithm where during each iteration all of the chromosomes representing partial solutions undergo assessment, as outlined in the figure below.

From the set of chromosomes, which have received the highest marks, the so-called family pool is drawn, based on which the next generation of chromosomes is created. Thus, the basic assumption is that assessments of individual partial solutions are reliable and show, which of them indicate improvement in the right direction. Due to the fact that mapping DS for bit vectors is clear, the evaluation of chromosomes (bit vectors) can be carried out without performing the reverse transformation, i.e. \(T^{-1} : b \rightarrow U\). For illustrative purposes this reverse transformation was presented on the example in Figure 3.

**Fig. 2. Genetic algorithm operation flowchart**

**Fig. 3. Bit vector transformation to a set of arcs**

Due to the fact that the evaluation of each chromosome is the most frequently performed operation at the time of the genetic algorithm, then it is about its effectiveness, thus the usefulness in solving real problems. In the remainder of this article presented were two adaptation functions defined respectively for PMC and BGM diagnostic opinion structures.

---

1 Every PMC-type diagnostic opinion structure is also a proper BGM-type diagnostic opinion structure.
Fitness Function Design

In the case of PMC-type diagnostic structures the fitness function was determined based on the sufficient condition (2) in such a way as to give higher marks for partial solutions having a sufficient number of unique predecessors. The form of the elaborated function is as follows:

\[ f(G) = \sum_{p=0}^{t+1} \sum_{E \in Z_p} k(E') \]

where: \( E' \) – the subset of set \( E \) whereas DS is \( G =< E, U, Z_p > \), \( Z_p \) – such a set of arcs where \( \forall E' \subset Z_p: |E'| = |E| - 2 \cdot t + p, t + p \) – a parameter describing the diagnosability of DS, \( k(E') \) is the function defined as follows:

\[ k(E') = \begin{cases} 1, & \gamma'(E') = p, \gamma'(E') \neq p. \end{cases} \]

The adjustment function defined in the above-mentioned way ensures that the structures found by GA will be \( t \)-diagnosable; however, it leaves no control on the other properties. According to [4] the most desirable are optimal structures, since they are also economic structures, i.e. structures with the lowest cost connections. Therefore, it seems reasonable to modify the presented adaptation function in such a way so that the highest ratings obtain optimal DS in terms of diagnostics. This could be achieved by adding a segment to the adjustment function that increases the rating in the case when the tested solution shows typical properties of optimal structures. Based on [4] it must be assumed that the feature that characterizes the desired solution (optimal) is \( \forall e \in E; \mu'(e) = \mu^+(e) = t \), thus the fitness function can be modified to the following form:

\[ f(G) = \sum_{e \in E} g(e) + \sum_{p=0}^{t+1} \sum_{E \in Z_p} k(E') \]

where the function \( g(e) \) is defined as follows:

\[ g(e) = \begin{cases} 1, & \mu'(e) = \mu^+(e) = t, \mu'(e) \neq t \end{cases} \]

and \( \mu^+(e) \) indicates the node \( e \) output degree.

For the case when only the set of BGM-type diagnostic structures are considered, which are not simultaneously PMC-type structures. The fitness function should be constructed based on the previously mentioned dependence, which clearly shows that any two nodes of the graph \( G \) cannot have the same set of predecessors. The proposed adaptation function is as follows:

\[ f(G) = \sum_{e \in E} w(\Gamma^+(e), \Gamma^+(e')) \]

where \( \Gamma^+(e) \) - the set of node predecessors and \( w(\Gamma^+(e), \Gamma^+(e')) \) indicates the function defined in the following way:

\[ w(\Gamma^+(e), \Gamma^+(e')) \begin{cases} 1, & \gamma^+(e) \neq \gamma^+(e') \end{cases} \]

Just as with PMC-type structures, it is also in this case that the basic form of the adaptation function gives positive ratings to all \( t \)-diagnosable structures, but without taking into account their additional properties. In order to increase the efficiency of determining the BGM-type optimal structure for GA in accordance with the given topology, it is necessary to add fitness functions to the additional segment in the form of the earlier defined \( g(e) \) function. In this case, the fitness function shall take the following form:

\[ f(G) = \sum_{e \in E} g(e) + \sum_{e \in E \in U} w(\Gamma^+(e), \Gamma^+(e')) \]

It is worth noting that the fitness function should clearly state the better chromosome by giving it a higher rating [12]. In our solution the addition operation can raise doubts, which could lead to a situation, in which two structures with different properties are evaluated identically. However, this is an intentional action, because the test structures may equally satisfy various demands put forward earlier, and thus may be equally (in terms of diagnosis) useful during its evolution. Empirical studies have confirmed the validity of assumptions made.

Incorporating generalized tests costs

The adaptation function presented in previous chapters enabling the finding of \( t \)-diagnosable PMC and BGM-type DS are based on the assumption that all tests have the same cost. In order to take into account in the final solution the additional conditions specified in the form of a generalized cost of individual tests the previously presented adaptation function should be developed to the form presented below:

For PMC-type structures:

\[ f(G) = \sum_{p=0}^{t+1} \sum_{E \in Z_p} k(E') - \sum_{i=1}^{t} \sum_{u \in U_i} n(r(u)) \]

and for optimal structures

\[ f(G) = \sum_{e \in E} g(e) + \sum_{p=0}^{t+1} \sum_{E \in Z_p} k(E') - \sum_{i=1}^{t} \sum_{u \in U_i} n(r(u)) \]

For BGM-type structures:

\[ f(G) = \sum_{e \in E} w(\Gamma^+(e), \Gamma^+(e')) - \sum_{i=1}^{t} \sum_{u \in U_i} n(r(u)) \]

In the above formulas \( U_i \) - a set of tests carried out by \( i \)-th node (\( U_i \subset U \)) and \( r(u) \) - is the generalized cost of test \( u \). The additional segment in the form of: \( \sum_{u \in U_i} n(r(u)) \) represents total normalized costs of all nodes coming out of node \( i \).

Evaluating the usefulness of the algorithm

The above-presented adaptation method of designing DS for networks with a given communication structure was implemented in the C language. In order to validate its usefulness a series of tests were conducted designed to assess the effect of the number of nodes and edges (at a fixed parameter \( i \) of the initial graph (representing the network, for which the DS is designed) for the time of determining the structure. All tests were performed on a Pentium 2 GHz class computer using GA parameters presented in Table 1.

<table>
<thead>
<tr>
<th>Tab. 1. Genetic algorithm parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameter</strong></td>
</tr>
<tr>
<td>Population</td>
</tr>
<tr>
<td>Mutation</td>
</tr>
<tr>
<td>Number of predecessors</td>
</tr>
<tr>
<td>Selection</td>
</tr>
<tr>
<td>Crossing</td>
</tr>
<tr>
<td>Crossing probability</td>
</tr>
</tbody>
</table>
Therefore, the solution space exploration was, in which the number of nodes ranged from 5 to 9, respectively, from 2^5 to 2^9. Note that the experiment s were carried out for the matrix that can be specified in the transition matrix (diagonal directly from the maximum number of opinion combinations 166). The obtained results of the tests clearly show that the developed method is linearly dependent on the number of edges (Fig. 4) and exponentially on the number of nodes (Fig. 5). It is worth noting that the number of nodes is linked to the minimum number of edges appearing in graph G by property (1), (2) for the PMC structures and properties (3), (4) for the BGM structures. These properties make it clear that for the r-diagnosable structure the input degrees of all nodes must be greater than or equal to r, and in the case of structures with the lowest possible generalized cost the following should occur: 

∀e ∈ E: μ^r(e) = μ^r(e) = t → ∀e ∈ E: μ^r(e) = 2t.

Note that the experiments were carried out for the structures, in which the number of nodes ranged from 5 to 9. Therefore, the solution space exploration was, respectively, from 2^20 to 2^22. This dependence results directly from the maximum number of opinion combinations that can be specified in the transition matrix (diagonal matrix A^r(y−1), which is assumed to contain just zeros).

The exact transition of the number of nodes in the proposed structure (from 5 to 9 nodes) to the size of solutions search space is summarized in the table.

The study was conducted for extreme cases where it was possible to determine only a few r-diagnosable DS with a given topology. It is typically encountered that the number of possible connections between nodes is much higher, and therefore the effectiveness of the proposed method increases as the number of nodes increases.

### Table 2. Dependence of the number of possible solutions with the number of nodes

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>Number of possible solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2^20</td>
</tr>
<tr>
<td>6</td>
<td>2^30</td>
</tr>
<tr>
<td>7</td>
<td>2^42</td>
</tr>
<tr>
<td>8</td>
<td>2^56</td>
</tr>
<tr>
<td>9</td>
<td>2^72</td>
</tr>
</tbody>
</table>

### Conclusion

The method presented in the paper allows the determination of the most inexpensive r-diagnosable diagnostic structures of type PMC and BGM, taking into account the existing interconnection network topology and general costs of assigned tests. It can be applied in practice, when a new network diagnostic structure needs to be determined after reconfiguration, for example, due to diagnosing specific nodes being faulty.

The results based on the simulation experiments show that the developed algorithm works both for regular and irregular structures.

Application of the proposed heuristics gives satisfactory results at permissible (in exploitation of communication networks in practice) search times for a solution.

**Scientific study financed from education funds in 2008-2011 as a research project no. O N514 051135.**

### REFERENCES


**Authors:** mgr inż. Łukasz Strzelecki, Wojskowa Akademia Techniczna, Instytut Teleinformatyki i Automatyki, ul. Kaliskiego 2, 00-908 Warszawa 49, E-mail: l.strzelecki@wat.edu.pl; dr inż. Zbigniew Zieliński, Wojskowa Akademia Techniczna, Instytut Teleinformatyki i Automatyki, ul. Kaliskiego 2, 00-908 Warszawa 49, E-mail: z.zielinski@ita.wat.edu.pl