A Protocol for Multi-Agent Diagnosis with Spatially Distributed Knowledge

Nico Roos
Universiteit Maastricht,
IKAT,
P.O.Box 616,
6200 MD Maastricht.

Annette ten Teije
Utrecht University,
ICS,
P.O.Box 80.089,
3508 TB Utrecht.

Cees Witteveen
Delft University of Technology,
ITS,
P.O.Box 5031,
2600 GA Delft.

ABSTRACT

In a large distributed system it is often infeasible or even impossible to perform diagnosis using a single model of the whole system. Instead, several spatially distributed local models of the system have to be used to detect possible faults. Traditional diagnostic tools, however, are not suitable to deal with such a set of spatially distributed local models.

A Multi-Agent System of diagnostic agents, where each agent has a model[1] of a subsystem, may be proposed as a solution for establishing global diagnoses of large distributed systems. Unfortunately, establishing a global minimal diagnosis is NP-Hard, even if every agent is able to determine local minimal diagnoses in polynomial time. Moreover, communication overhead in establishing a global diagnosis will be high: unless $P = NP$ a super-polynomial number of messages between the agents will be required for establishing a global diagnosis.

In this paper we present a protocol that overcomes this complexity issue by exchanging diagnostic precision for enables agents to determine local minimal diagnoses that are consistent with global diagnoses. Moreover, the protocol ensures that no agent acquires knowledge of global diagnoses. The protocol does not guarantee that a combination of the agents' local minimal diagnoses is also a global minimal diagnosis. However, for every global minimal diagnosis, there is a combination of local minimal diagnoses.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent Systems*

General Terms

Algorithms, Theory, Verification

[1]Here, we focus on Model-Based Diagnosis.

AAMAS'03, July 14–18, 2003, Melbourne, Australia.

Keywords

Model-Based Diagnosis

1. INTRODUCTION

A classical diagnostic tool can be viewed as a *single diagnostic agent* having a model of the whole system to be diagnosed. There are, however, several reasons why in some applications such a single agent approach may be inappropriate. First of all, if the system is a large physically distributed system, e.g. a modern telecommunication network, there may be not enough time to compute a diagnosis centrally and to communicate all observations. Secondly, if the structure of the system is dynamic, e.g. AGV systems driving in a platoon, the system may change too fast to maintain an accurate global model of the system over time. Finally, sometimes a central model is simply undesirable. For example, if the system is distributed over different legal entities and one entity does not wish other entities to have a detailed model of its part of the system. For such systems, a *distributed* approach of multiple diagnostic agents can offer a solution, as was shown in [12].

In such multi-agent based systems, the model (knowledge) of the system can be distributed over the agents in two principally different ways[2] (cf. [5]): *(i) spatially distributed*: knowledge of system behavior is distributed over the agents according to the spatial distribution of the system's components, and *(ii) semantically distributed*: knowledge of system behavior is distributed over the agents according to the type of knowledge involved, e.g. a model of the electrical and a model of the thermodynamical behavior of the system. For both types of distributions it can be shown that a multi-agent system is able to establish the same global diagnoses as a single diagnostic agent having the combined knowledge of all agents [12].

In this paper we will focus on a spatial distribution of knowledge over the diagnostic agents. First, we review in section 2 the well known definitions of Model Based Diagnosis [10, 7]. In section 3, we formalize multi-agent diagnosis and present some formal results. We analyze the problems involved in multi-agent diagnosis in section 4. Section 5 describes a protocol for establishing a diagnosis which is used in the experiments reported on in section 6. Section 7 concludes the paper.

Remark For simplicity, we do not consider time, though the definition can be extended to dynamic systems. We

[2]Combinations are, of course, also possible.

also do not consider formulations based on Discrete Event Systems [3, 9]. These formulations emphasize more the dynamical aspects of the systems on an abstract level, and especially the occurrence of failure events. As far as failure events can be related to fault modes, discrete event systems can be viewed as a special case of our approach.

2. THE DIAGNOSTIC SETTING

A system to be diagnosed is a tuple $S = (C, M, Id, Sd, Ctx, Obs)$ where C is a set of components, $M = \{M_c \mid c \in C\}$ is a specification of possible fault modes per component, Id is a set of (identifiers of) connection points between components, Sd is the system description, Ctx is a specification of input values of the system that are determined outside the system by the environment and Obs is a set of observed values of the system. A component in C has a normal mode $nor \in M_c$, one general fault mode $ab \in M_c$, and possibly several specific fault modes. We assume that all components have *inputs* and *outputs*.[3]

The *system description* $Sd = Str \cup Beh$ consists of a structural description Str and a behavioral description Beh of the components. The structural description Str consists of instances of the form $p = in(x, c)$ or $p = out(x, c)$ where x is an input or output identification of a component c and $p \in Id$ is a connection point identifier. A connection point $p \in Id$ is connected to at most one output of some component; i.e. if $p = out(x, c)$ and $p = out(y, c')$, then $x = y$ and $c = c'$. A connection point p has a value $value(p)$ which is determined by the output of a component or a system input.

The set $Beh = \bigcup_{c \in C} Beh_c$ specifies the behavior for each component $c \in C$. The behavior description Beh_c of a component describes the component's behavior for each (fault) mode in M_c, possibly with the exception of $ab \in M_c$. In this specification, the predicate $mode(c, m)$ is used to denote the mode $m \in M_c$ of a component c. For each instance $mode(c, m)$, Beh_c specifies a behavioral description of the form: $mode(c, m) \rightarrow \Phi$ where $m \in M_c$.[4] The expression Φ describes the component's behaviour given its mode $m \in M_c$.

The context Ctx describes the values of system inputs $Id_{in} = \{p \in Id \mid \forall x, c : (p = out(x, c)) \notin Str\}$ that are determined by the environment, so Ctx consists of instances of the form $value(p) = v$ where v is the value of a connection point $p \in Id_{in}$.

Finally, the set of observations Obs describes the values of those connection points that are observed (measured) by the diagnostic agent. It therefore also consists of instances of the form $value(p) = v$ where v is a value of a connection point $p \in Id$.

A *candidate diagnosis* is a set D of instances of the predicate $mode(c, m)$ such that for every component $c \in C$ there is exactly one mode in $m \in M_c$ such that $mode(c, m) \in D$. We define a *diagnosis* D as a candidate diagnosis meeting some additional constraints. Combining the two well-known types of diagnoses, *consistency based* [8, 10] and *abductive* [1], we use the following, more general, definition (cf [2]):

[3]This assumption is not valid for every system. It is, however, possible to transform most systems to a system consisting of components with only inputs and outputs (see for instance [4]).

[4]Note that we may use a single description for a class of components. Instances of this description must imply the form of description give here.

DEFINITION 1. *Let $S = (C, M, Id, Sd, Ctx, Obs)$ be the system to be diagnosed and let $\mid\sim$ to denote the possibly limited reasoning capabilities of a diagnostic system.*[5] *Moreover, let Obs_{con}, $Obs_{\text{abd}} \subseteq Obs$ be subsets of observations and let D be a candidate diagnosis. Then D is a diagnosis for S iff*

1. $D \cup Sd \cup Ctx \mid\sim \bigwedge_{\varphi \in Obs_{\text{abd}}} \varphi$,
2. $D \cup Sd \cup Ctx \cup Obs_{\text{con}} \not\mid\sim \bot$.

The number of diagnoses can be quite high, exponential in the worst case. In case of consistency based diagnosis, we can characterize the set of diagnoses using a small number of *minimal diagnoses*: a diagnosis D is a minimal diagnosis if for no other diagnosis D',

$$\{mode(c, nor) \mid mode(c, nor) \in D\} \subset \{mode(c, nor) \mid mode(c, nor) \in D'\}.$$

3. MULTI AGENT DIAGNOSIS

The knowledge distribution over multiple agents defines a division of a system into several subsystems. If knowledge is *spatially* distributed, the set of components C is partitioned over the agents. So, agent A_i has knowledge about a set C_i of components and $C = \biguplus_{i=1}^{m} C_i$ where m is the number of agents. This results in the following distribution of knowledge: $Beh_i = \{\xi \in Beh \mid \xi = (mode(c, m) \rightarrow \Phi), c \in C_i\}$, $Str_i = \{(p = in(x, c)) \in Str \mid c \in C_i\} \cup \{(p = out(x, c)) \in Str \mid c \in C_i\}$ and $Obs_i = \{(value(p) = v) \in Obs \mid (p = out(x, c)) \in Str, c \in C_i\}$. Note that we do not have to split up the context Ctx.

By distributing knowledge, i.e. Beh_i and Str_i over the agents, we loose the knowledge about the connections between components managed by different agents. Therefore, we provide each agent A_i with information about connection points that connect to components managed by other agents and we split the set connection points into *relative* inputs In_i and outputs Out_i of the agent's subsystem. Here, $In_i = \{p \in Id \mid \{p = in(x, c), p = out(y, c')\} \subseteq Str, c \in C_i, c' \notin C_i\}$ and $Out_i = \{p \in Id \mid \{p = out(x, c), p = in(y, c')\} \subseteq Str, c \in C_i, c' \notin C_i\}$. Hence, $S_i = (C_i, M, Id, Sd_i, Ctx, Obs_i, In_i, Out_i)$ is a subsystem to be diagnosed by the agent. A candidate diagnosis of the subsystem S_i is denoted by D_i.

The diagnosis of one agent. Each agent A_i in the multi-agent system has to diagnose the subsystem $S_i = (C_i, M, Id, Sd_i, Ctx, Obs_i, In_i, Out_i)$. This can be viewed a single agent diagnosis if values of the inputs and outputs of the subsystem are known. Let us use V_i to denote the set of value assignments $value(p) = v$ to the inputs, where $p \in In_i$. So, V_i is the local context of the subsystem S_i that is determined by the outputs of other subsystems. We therefore extend Definition 1 to the diagnosis of subsystems.

DEFINITION 2. *Let $S_i = (C_i, M, Id, Sd_i, Ctx, Obs_i, In_i, Out_i)$ be a subsystem to be diagnosed, let V_i be a (partial) description of the values of the connection points In_i and let D_i be a candidate diagnosis for S_i. Then D_i is a diagnosis for S_i iff D_i is a diagnosis for $(C_i, M, Id, Sd_i, Ctx \cup V_i, Obs_i)$.*

[5]I.e $\{\varphi \mid \Sigma \mid\sim \varphi\} \subseteq \{\varphi \mid \Sigma \vdash \varphi\}$.

The diagnosis of multiple agents. Given multiple diagnostic agents, an important question is how the diagnoses of the agents relate to the diagnoses of a single (omniscient) agent that has complete knowledge of the system description and the observations. When addressing this question we assume throughout the paper that *there are no conflicts between the knowledge of the different agents.* That is, there always exists a diagnosis D such that $D \cup Sd \cup Cxt \cup Obs$ is consistent.

The following propositions show how multi-agent diagnosis and single agent diagnosis (w.r.t. the same global system S) are related.

PROPOSITION 1. *Let $S_1, ..., S_k$ be the subsystems making up the system S. Moreover, let D be a single agent diagnosis of S.*

Then $V_i = \{(value(p) = v) \mid p \in In_i,\ D \cup Sd \cup Ctx \mid\sim (value(p) = v)\}$ is the local context of S_i that is determined by the other subsystems S_j, and $D_i = \{mode(c, s) \mid c \in C_i, mode(c, s) \in D\}$ is a diagnosis of S_i. [12]

PROPOSITION 2. *Let $S_1, ..., S_k$ be the subsystems that make up the system S and let the local context V_i of S_i describe the values of connection points in In_i that must be determined by the other subsystems S_j, and let D_i be a diagnosis of S_i determined by agent A_i given V_i.*

Then, $D = \bigcup_{i=1}^{k} D_i$ is a single-agent diagnosis if

1. *D is a candidate diagnosis,*
2. *$D_i \cup Sd_i \cup Ctx \cup V_i \mid\sim (value(p) = v)$, and*
3. *for every $p \in Out_i$, $p \in In_j$ and $(value(p) = v) \in V_j$. [12]*

Complexity. If knowledge is spatially distributed, each agent manages a different part of the system. The behavior of a subsystem managed by an agent depends on the behavior of the other subsystems. This makes it difficult to predict the behavior of the whole system, since the values of the connection points in Out_i depend on the local context V_i. The values specified by V_i, however, are determined by other subsystems S_j whose local context V_j may depend on the values of the connection points in Out_i. Because of these circular dependencies, predicting the behavior of the system is an NP-Hard problem: The Constraint Satisfaction Problem (CSP) can be easily reduced to this problem by using In_i to represent a variable, (Sd_i, D_i, Ctx, Obs_i) to represent a constraint and V_i to represent a variable assignment from the domain.

THEOREM 1. *Given a global candidate diagnosis D, predicting the values of all connection point is an NP-Hard problem. [12]*

To avoid solving such a hard problem for every candidate diagnosis, consistency based diagnosis and consistency based diagnosis with abductive explanation of normal observations are preferred. These approaches do not apply to fault models. Nevertheless, we still have to solve one NP-hard problem to predict the *normal* behavior of the system[6].

[6]We might, however, avoid predicting the normal behavior if, at some abstract level, we can assume default values for the connection points. This also reduces the amount of information exchange

The determination of a minimal diagnosis is also an NP-hard problem when knowledge is spatially distributed over the agents. This complexity issue is a direct result of the intrinsic complexity of Model-Based Diagnosis, as follows from the outline of the proof of the following theorem.

THEOREM 2. *Determining a minimal diagnosis D is an NP-hard problem, even if each agent is able to determine all its local minimal diagnoses in polynomial time.*

To prove the theorem, we reduce a classical single agent diagnostic problem to a multi-agent diagnostic problem. Given a system S, partition the system into subsystems S_i such that every S_i consists of exactly one component diagnosed by agent A_i. Clearly, each agent A_i is able to determine all its local minimal diagnoses in polynomial time. If, however, the agents could determine, through collaboration, a global minimal diagnosis in polynomial time, this would immediately imply the existence of a polynomial algorithm solving the NP-hard single agent diagnostic problem.

Note that the above theorem does not hold if knowledge is semantically distributed over the agents. In this case, the agents are able to find a minimal diagnosis of the system in polynomial time [12].

The complexity of multi-agent diagnosis does not tell us what the communication overhead of the multi-agent system will be. In fact, one agent could collect the information of all the other agents, determine the diagnoses and distribute the results. This approach requires only polynomial number of messages. However, if agents may only send message in which they *(1)* claim that an input determined by another subsystem is either correct or incorrect, *(2)* reject such a claim, *(3)* accept a claim or *(4)* retract a claim, then a super-polynomial number of messages is needed in order to determine a minimal diagnosis.

COROLLARY 1. *Suppose that agents are only allowed to exchange messages in which they (1) claim that an input determined by another subsystem is either correct or incorrect, (2) reject such a claim, (3) accept a claim or (4) retract a claim.*

Then, in order to determine a minimal diagnosis, agents have to exchange a super-polynomial number of messages.

Note that no single agent is able to derive a global diagnosis of the system based on these messages. Therefore, agents have to establish a minimal diagnosis in collaboration.

Now, suppose that a polynomial number of messages would suffice for determining a minimal diagnosis. Then agents are able to adapt their local diagnoses based on these messages and can determine all local minimal diagnoses in polynomial time with only a polynomial number of messages. But then a minimal diagnosis of the system can be determined in polynomial time, contradicting[7] Theorem 2. Hence, a super-polynomial number of messages is required.

4. ANALYSIS

After observing abnormal behavior of the system, the agents must make a global diagnosis. In order to do so, each agent must make a local diagnosis in which it also takes into consideration the correctness of those inputs of its subsystem that are determined by other agents. Therefore, we must extend a candidate diagnosis D_i of agent A_i

[7]under the assumption that $P \neq NP$.

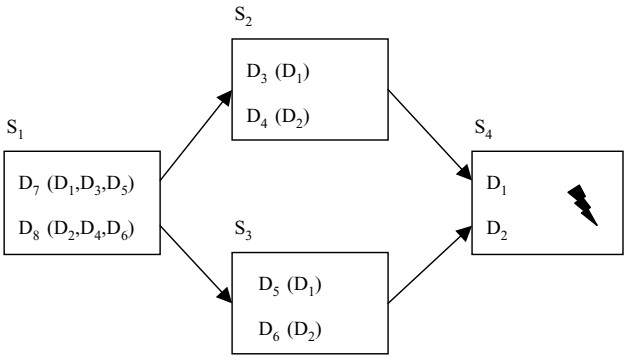

Figure 1: Combining local diagnoses.

with *correctness assumptions* Ca_i about the systems inputs. For every input $p \in In_i$, Ca_i contains either the proposition $correct(p)$ or $\neg correct(p)$. The *conditional context* Cc_i will be used to describe inputs of a subsystem S_i, i.e. the local context of the subsystem determined by other subsystems *conditional* to these correctness assumptions, i.e.,

$$Cc_i = \{correct(p) \leftrightarrow (value(p) = v) \mid value(p) \in V_i\}.$$

If in its local diagnosis (D_i, Ca_i), agent A_i assumes that one of its inputs is incorrect, the agent must communicate this information to an other agent A_j determining the input. Next, agent A_j may treat this information as an observation of one of its outputs, and adapt its local diagnosis accordingly.

There are three problems connected with this approach. The first problem concerns the determination of a minimal diagnosis. As we have seen in the previous section, if knowledge is spatially distributed over the agents, determining a minimal diagnosis is an NP-hard problem leading to a combinatorial explosion in the inter-agent communication.

The second problem concerns the occurrence of circular dependencies. Suppose that agent A_i blames an observed anomaly on one of its inputs determined by the output of subsystem of agent A_j. Agent A_j in its turn may blame the error in the output determining the input of S_i on one of its own inputs. If this input is determined by an output of the subsystem of agent A_i, we may have a cycle of blames that supports itself. Clearly, local diagnoses that constitutes such cycles of blames do not represent a valid diagnosis of the system.

The third problem concerns the combination of local diagnoses. Suppose that we have four different subsystems each managed by a different agent, as illustrated in figure 1. In subsystem S_4 an anomaly is observed. The agent managing S_4 establishes two diagnoses D_1 and D_2. In both diagnoses, subsystems S_2 and S_3 are blamed for the observed anomaly. Based on diagnosis D_1 the agent of S_2 derives the diagnosis D_3 and the agent of S_3 the diagnosis D_5. Similarly, based on diagnosis D_2 the diagnoses D_4 and D_6 are derived. If in each of these diagnoses subsystem S_1 is blamed for the problem, then the agent managing S_1 must make a diagnosis either based on the diagnosis D_3 and D_5 or on the diagnosis D_4 and D_6. It must therefore consider D_3 and D_5 as well as D_4 and D_6 because they are based on D_1 and D_2, respectively. Clearly, the agent may not derive a diagnosis based on, for instance, D_4 and D_5.

5. THE PROTOCOL

We wish to design a protocol that will enable each agent to determine all its local minimal diagnoses such that each local minimal diagnosis is consistent with a global diagnosis of a single agent having the combined knowledge of all agents. *None of the local agents should, however, be able to determine a global diagnosis and preferably not even be able to determine the subsystem causing the observed anomalies.*

Since diagnoses can be derived from conflict sets [10, 8, 7], and since conflict sets contain the dependencies needed for handling the problem with loops described in the previous section, we propose a protocol based on determining local conflict sets. A conflict set is a set of assumptions that cannot be correct given the current observations of the system. In the absence of fault models for the components, such an assumption states that a component behaves normally. In that case, a conflict set is a subset Ξ of $\{mode(c, nor) \mid c \in C\}$ such that $\Xi \cup Sd \cup Cxt \cup Obs$ is inconsistent. Every diagnosis can be derived from the conflict sets by selecting an assumption of each conflict set and by subsequently stating that this assumption is incorrect.

To determine the diagnoses, it suffices to consider minimal conflict sets. The number of minimal conflict sets can, however, be exponential in $|Ctx \cup Obs|$. To reduce this number, we limit ourselves to minimal sets of correctness assumptions that are needed in order to *causally* predict the values of observed connection points. We will call such sets *dependency sets*. Such a dependency set that enables us to causally predict the value of a connection point is called a conflict set if the predicted value does not correspond with the observed value.

Since each connection point has exactly one dependency set[8], the number of conflict sets will be equal to the number of connection points that are observed to be incorrect. This reduction of the number of conflict sets also reduces the diagnostic precision. That is, the number of diagnoses increases. On the other hand, we can also reduce the number of diagnosis by using the dependencies set to apply abductive explanation of *normal* observations [11]. That is, components that belong to a dependency set of a connection point that is observed to be correct, are assumed to function normally. This assumption is justified, if the probability that one fault is compensated by another is sufficiently small. A dependency set of a connection point that is observed to be correct is called a *confirmation set*.

Since an agent only knows the components that belong to the subsystem it manages, the correctness assumptions of a dependency set determined by an agent may only concern the local components and the inputs of the subsystem.

DEFINITION 3. *A dependency set of connection point $p \in Id$ of a subsystem S_i is defined as the* smallest *set $Dep(p) \subseteq \{mode(c, nor) \mid c \in C\} \cup \{correct(p) \mid p \in In_i\}$ such that*

$$Dep(p) \cup Sd_i \cup Cxt \cup (Obs_i \setminus (value(p) = v)) \mathrel{|\!\sim} (value(p) = v).$$

Here, $|\!\sim$ is restricted in such a way that the output value of a component can only be derived from the component's input values.

The agents have to determine the local parts of the global conflict (and confirmation) sets using the local dependency

[8]provided that no component behaves like a switch [11]

sets. When an agent A_i observes the value of a connection point $p \in C_i$, the agent knows that the assumption $\{mode(c, nor) \mid mode(c, nor) \in Dep(p)\}$ in the local dependency set $Dep(p)$ belongs to the global conflict or confirmation set depending on the observed value of p. Next, for each input q of the subsystem such that $correct(q) \in Dep(p)$, agent A_i must inform agent A_j managing the subsystem that determines the value of q about the status of the connection point q. The status is *correct* if $Dep(p)$ is part of a global confirmation set and is *possibly incorrect* if $Dep(p)$ is part of a global conflict set.

The agent A_j that receives information about the status of an output q of its subsystem knows that the assumptions $\{mode(c, nor) \mid mode(c, nor) \in Dep(q)\}$ in the local dependency set $Dep(q)$ belong to a global conflict or confirmation set depending on the status information it receives. Since it is possible that agent A_j receives status information of several outputs $q_1, ..., q_m$ of its subsystem based on one and the same observed connection point p, the assumptions $\{mode(c, nor) \mid mode(c, nor) \in Dep(q_x)\}$ in the local dependency set $Dep(q_x)$ of all these outputs belong to the same global conflict or confirmation set depending on the status information. Hence, in order to enable agent A_j to combine these sets of assumptions into a local part of a global conflict or confirmation set, an agent must provide with the status information of an input of its subsystem, an identification of the observed connection point p on which the status information is based. In order to guarantee anonymity, a randomly generated number can be used for this purpose.

Another important issue is the handling of circular dependencies between subsystems. Agents must detect such loops of dependencies when determining the local parts of global conflict and confirmation sets. An agent A_j may receive status information about an output q of its subsystem that depends on the status information of an input r, which agent A_j has sent to the agent A_h managing the subsystem that determines the value of r. Clearly, if $correct(r) \in Dep(q)$, agent A_j should not send again status information to agent A_h about r in order to avoid an infinite loop.

The protocol presented in figure 2 enables the agents to determine the local parts of the global conflict and confirmation sets. Using its local conflict and confirmation sets, an agent can determine its local diagnoses. The agent may choose any approach that enables it to derive the local diagnoses using the local conflict and confirmation sets.

The correctness of the protocol follows from the following propositions.

PROPOSITION 3. *Consider all sets of assumptions in any set Sk respectively Sb determined by an agents. The combination of the sets having the same identification of an observed connection point p results in the global conflict respectively confirmation set of the connection point p. Moreover, for each global conflict set and for each global confirmation set such local sets in Sk respectively Sb exist.*

PROPOSITION 4. *The number of massages that will exchanged in the worst case is bounded by $|Id|^2$.*

Assuming that every component has at least one output, the protocol has a worst case time complexity of $\mathcal{O}(|Id|^3)$.

Using the protocol presented in figure 2, each agent can determine all local diagnoses using the local conflict and confirmation set. Assuming that an anomaly caused by one component cannot be compensated by an anomaly caused another component, the agent may remove from each local conflict set those components that occur in one of its local confirmation sets. All minimal local diagnosis can then be derived from the resulting conflict sets. The combination of the local minimal diagnoses of different agent forms a global diagnosis of the whole system provided that not a cycle of blames arises. E.g., agent A_1 blames the cause of a problem on the subsystem managed by agent A_2 in a local diagnosis, while agent A_2 blames the cause of the problem on the subsystem managed by agent A_1 in its local diagnosis. Moreover, if no cycle of blames occurs in combining the local diagnoses, the resulting global diagnosis need not be a minimal global diagnosis of the system. However, the opposite, as stated by the following theorem, does hold.

THEOREM 3. *For each global minimal diagnosis D based on global conflict (and confirmation) sets, there are local minimal diagnoses (D_i, Ca_i) based on local conflict sets in Sk (and confirmation sets in Sb) such that $D = \bigcup_i D_i$.*

```
Protocol of agent A_i
  for each connection point p that is observed or is in Out_i do
    determine the dependency set Dep(p);
    rid(p) := randomly generated identification;
  end;
  Sb := ∅;
  Sk := ∅;
  for for each observed connection point p do
    if value of p is correct then
      Sb := Sb ∪ {(Dep(p), rid(p))}
    else
      Sk := Sk ∪ {(Dep(p), rid(p))};
  for each (X, id) ∈ Sb do
    for each correct(q) ∈ X do
      send (q, 'correct', id) to agent A_j with q ∈ Out_j;
  for each (X, id) ∈ Sk do
    for each correct(q) ∈ X do
      send (q, 'possibly incorrect', id) to
        agent A_j with q ∈ Out_j;
  repeat
    for each (p, status, id) received from agent A_j do
      if status = 'possibly incorrect' then
        if (Y, id) ∈ Sk then
          X := Dep(p) − Y;
          Sk := (Sk − {(Y, id)}) ∪ {(Dep(p) ∪ Y, id)};
        else
          X := Dep(p);
          Sk := Sk ∪ {(Dep(p), id)};
        end; [ if ]
        for each correct(q) ∈ X do
          send (q, 'possibly incorrect', id) to
            agent A_j with q ∈ Out_j;
      else if status = 'correct' then
        if (Y, id) ∈ Sb then
          X := Dep(p) − Y;
          Sb := (Sb − {(Y, id)}) ∪ {(Dep(p) ∪ Y, id)};
        else
          X := Dep(p);
          Sb := Sb ∪ {(Dep(p), id)};
        end; [ if ]
        for each correct(q) ∈ X do
          send (q, 'correct', id) to
            agent A_j with q ∈ Out_j;
      end; [ if ]
  until no more changes;
  return Sb and Sk;
end.
```

Figure 2: A protocol for multi-agent diagnosis

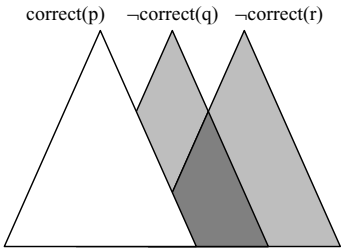

Figure 3: Focusing on likely broken components.

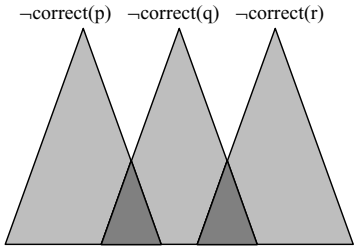

Figure 4: Multiple focuses.

The above presented protocol enables agents to determine local conflict and confirmation sets from which they can derive their local diagnoses. In deriving the local diagnoses, the agents should take into account that no component of a local conflict set needs to be broken if the local conflict set is the result of an observation in another subsystem. As follows from Theorem 3, for every minimal global diagnosis there is a combination of minimal local diagnoses that together forms the global diagnosis.

In the experiments reported on in the next section, we used a focusing approach [11] instead of determining local diagnoses. This approach determines the most likely broken components in a global conflict set; the focus. One can prove that the components of a focus are those components belong to no confirmation set and occur in the highest number of other conflict sets. The correctness of this focusing process is based on the assumption that (1) the probability that a component is broken is every small, and (2) the anomaly caused by one broken component cannot be compensated by the anomaly caused by another broken component. If the second assumption does not hold, focuses must be determined differently [11]. Figure 3 gives an illustration of the focusing process. The figure show one confirmation set and two conflict sets.

Note that because of the focuses, we loose again some diagnostic precision. In figure 4, we have three conflict sets resulting in three focuses which are represented by the *two* dark gray triangles in the figure. If a component in the left focus (the left dark gray triangle) is broken, one of the components in the conflict set of r must also be broken. In that case we should not prefer a component in the right focus (the right dark gray triangle) to be broken over another component in the conflict set of r. Similarly, if a component in the right focus is broken, one of the components in the conflict set of p must also be broken.

The main advantage of the focusing approach is that the approach has a polynomial time complexity and that we can determine additional measurements in order to reduce the number of diagnoses without actually determining the diagnoses. Moreover, diagnoses that were ignored by the focusing process will be considered after making the additional measurements.

An agent cannot determine by itself whether the focus of a local conflict set is a part of the focus of the corresponding global conflict set. Therefore, the agent must inform the other agents about the number of local conflict sets it has used to determine a focus. An agent removes the focus of a conflict set upon receiving information that another agent has used more conflict sets to determine a focus of a conflict set with the same identification.

In our implementation, agents informed each other about the number of conflict sets used to determine a focus through broadcasting these numbers together with identifications of the observed connection points. Instead of broadcasting the agents could also propagate this information. Then the agents must make sure that no loops occur in the propagation process.

6. EXPERIMENTS

The performance of the proposed protocol has been validated though experiments. In the experiments, we have measured the communication overhead of 1000 generated systems consisting of 100 components distributed over 10 agents. The number of broken components varied from 1 to 10 and the number of observation points from 10 to 100 with step size 10. The broken components and the observation points were selected randomly. The step size of 10 was chosen to guarantee that each agent has on average 1 to 10 observation points. For each combination of the number of broken components and the number of observation points, 10 instance were generated.

In each experiment, a system was generated by randomly distributing the components over a 'physically' area 1 by 1 unit and by uniformly distribution a group of agents over the same area. Components were assigned to their most nearby agent thereby forming the subsystems managed by the agents. Each component in the generated system had two inputs and one output. Each of the two inputs of a component c was connected to an output. The output was selected by choosing a value δ and by using it as an approximation of the length of the connection between the input and the output. That is, we search for a component d having a distance to c closest to the value δ.

We assumed a very simple behavior for the components. A broken component gives a wrong output value. Moreover, if one of the inputs of a component has a wrong value, then the output of the component will also have a wrong value. We also assume that the agents predict the behavior of the system starting from the observed values of the connection points. So, if a connection point is observed to have a wrong value, then the inputs values of the components that are determined by this connection point, are assumed to be correct because a new prediction is made. It is necessary to predict the behaviors of components using the observation in order to handle circular dependencies between components. If the behavior of the system is not predicted using the observation made, the agents can observe the outputs of every component in a cycle without being able to determine which components in the cycle are broken.

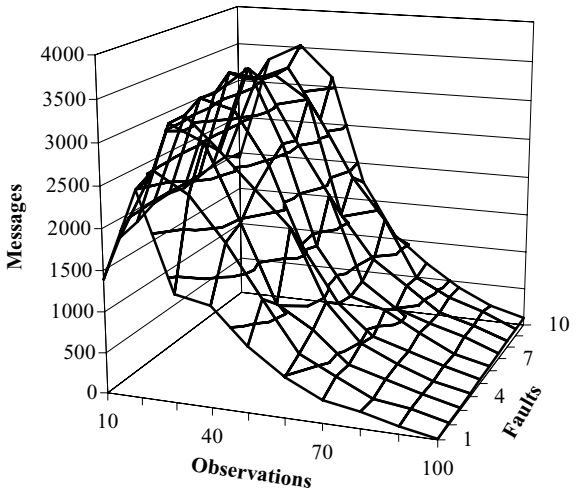

Figure 5: The communication overhead.

As a consequence of the way the system is generated, the system will contain many circular dependencies. These circular dependencies make the diagnostic process harder. Especially circular dependencies involving multiple agents result in a large communication overhead.

For each of the 1000 generated systems, the number of messages that were exchanged between the agents by the protocol were measured. This also includes the number of messages needed to determine the focuses, as described at the end of the previous section. Figure 5 presents the results if δ, the length of a connection between an input and an output, is randomly chosen from the interval $[0, 1]$. Other ways of choosing a value δ, in which we preferred local connection or limited the length of a connection, resulted in a lower communication overhead. The reduction in the communication overhead was no more than a factor 2, while the shape of the graph representing the number of messages, was similar to the graph of figure 5 for all the different ways of choosing δ.

The results of experiments show that the communication overhead is rather high. This is a consequence of the way a system has been generated which leads to a large number of circular dependencies in the system. These circular dependencies make the diagnostic process harder, especially circular dependencies involving multiple agents. Since realistic systems will probably contain less circular dependencies as the systems generated in the experiments, we expect that the communication overhead of these systems will be much lower.

7. CONCLUSION

Multi-agent diagnosis of spatially distributed subsystems is an NP-hard problem. This does not only imply a high time complexity, but also a high communication overhead. Especially the latter makes it infeasible to establish a diagnosis in a distributed way.

In this paper we presented a protocol that overcomes the complexity by exchanging diagnostic precision for a low communication overhead. The protocol lets agents free in choosing which algorithm to use for local diagnosis provided that the algorithm can derive a diagnosis using conflict sets.

The behavior of the algorithm has been verified by experiments. In the experiments, the agent apply a diagnostic approach that is based on focusing on likely broken components. This approach requires some additional information to be exchanged between the agents. The number of messages that agents exchange in running the protocol is rather high. This high number of messages is caused by the high number of circular dependencies between subsystems managed by different agents. We expect that for practical problems the number of circular dependencies will be much lower, thereby reducing the number of messages that the agents have to exchange.

8. REFERENCES

[1] L. Console and P. Torasso. Hypothetical reasoning in causal models. *International Journal of Intelligence Systems*, 5:83–124, 1990.

[2] L. Console and P. Torasso. A spectrum of logical definitions of model-based diagnosis. *Computational Intelligence*, 7:133–141, 1991.

[3] R. Debouk, S. Lafortune, and D. Teneketzis. Coordinated decentralized protocols for failure diagnosis of discrete-event systems. *Journal of Discrete Event Dynamical Systems: Theory and Application*, 10:33–86, 2000.

[4] J. J. van Dixhoorn. Bond graphs and the challenge of a unified modelling theory of physical systems. In F. E. Cellier, editor, *Progress in Modelling & Simulation*, pages 207–245. Academic Press, 1982.

[5] P. Frohlich, I. de Almeida Mora, W. Nejdl, and M. Schroeder. Diagnostic agents for distributed systems. In J.-J. Ch. Meyer and P.-Y. Schobbens, editors, *Formal Models of Agents. LNAI 1760*, pages 173–186. Springer-Verlag, 2000.

[6] J. de Kleer. Focusing on probable diagnosis. In *AAAI-91*, pages 842–848, 1991.

[7] J. de Kleer, A.K. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56:197–222, 1992.

[8] J. de Kleer and B. C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32:97–130, 1987.

[9] Y. Pencolé, M. Cordier, and L. Rozé. Incremental decentralized diagnosis approach for the supervision of a telecommunication network. In *DX01*, 2001.

[10] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.

[11] N. Roos. Efficient model-based diagnosis. *Intelligent System Engineering*, pages 107–118, 1993.

[12] N. Roos, A. ten Teije, A. Bos, and C. Witteveen. An analysis of multi-agent diagnosis. In *AAMAS 2002*, pages 986-987, 2002.