Multi-Agent Diagnosis with semantically distributed knowledge

Nico Roos [a] Annette ten Teije [b] Cees Witteveen [c]

[a] Universiteit Maastricht, IKAT, P.O.Box 616, 6200 MD Maastricht.
[b] Free University, Faculty of Sciences, De Boelelaan 1081A, 1081 HV Amsterdam.
[c] Delft University of Technology, ITS, PO Box 5031, 2600 GA Delft.

Abstract

We consider the problem of finding a commonly agreed upon diagnosis for errors observed in a system monitored by a number of different expert agents, each having its own specialized view on the system. That is, the expert agents have to agree on one or more diagnoses based on their specialized views of the system. Reaching an agreement is complicated by the two factors: (*i*) different specialisms need not distinguish the same fault modes of a component and (*ii*) knowledge of different specialisms need not be correct in some cases. This paper analyzes these problems and presents protocols that enable the agents to deal with these issues.

1 Introduction

A traditional diagnostic tool can be viewed as a single *diagnostic agent* having a model of the whole system to be diagnosed. In some applications, however, such a single agent approach is infeasible or at least undesirable. For example, the integration of knowledge into one model of the system is infeasible if the system is too large, too dynamic or distributed over different legal entities. Integration is undesirable if it concerns the combination of knowledge from different fields of expertise. In this latter case, where knowledge is called to be *semantically distributed*[1] [4], it would be better to introduce specialized agents communicating about anomalies detected.

The introduction of specialized (expert) agents immediately raises the problem how to reach an agreement on the cause of observed anomalies. As was pointed out in [7, 8], assuming a fixed maximum number of broken components, there exists a polynomial time protocol for reaching an agreement between the agents in case of a semantic knowledge distribution. This protocol is rather straight forward. A more difficult situation arises if the knowledge of some agents is *incomplete* in the sense that the agents have no behavioral knowledge about some fault modes, or if the knowledge of some agents is *incorrect* in the sense that the agents have incompatible knowledge about the behaviors of components. In this paper, we will address both issues.

This paper is organized as follows. Section 2 specifies the diagnostic setting, which is extended to multi-agent diagnosis in section 3. Section 4 introduces protocols for multi-agent diagnosis and section 5 concludes the paper.

[1] Besides a semantic knowledge distributed, we also distinguish a *spatial knowledge distribution*: knowledge of system behavior is distributed over the agents according to the spatial distribution of the system's components. The latter has been discussed in [9].

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 p 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 *in-* and *outputs*.[2]

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 in- or an output identification of a component c and $p \in Id$ is a connection point identifier[3]. Of course, 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 has a value, which is determined by the output of a component or a system input. The function $value(p)$ denotes the value of the connection point.

The set $Beh = \bigcup_{c \in C} Beh_c$ specifies a 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 $ab \in M_c$; i.e. $mode(c, ab) \rightarrow \top$. 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 set Ctx describes the values of system inputs $cId = \{p \in Id \mid \forall x, c : (p = out(x, c)) \notin Str\}$ that are determined by the environment. Ctx consists of instances of the form $value(p) = v$ where $p \in cId$ is a connection point and v is a value.

Finally, the set 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 $p \in Id$ is a connection point and v is a value.

A *candidate diagnosis* is a set D of instances of the predicate $mode(,)$ such that for every component $c \in C$ there is exactly one mode in $m \in M_c$ such that $mode(c, m) \in D$. A *diagnosis* is defined as follows:

Definition 1 *Let $S = (C, M, Sd, Ctx, Obs)$ 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}, $Obs_{\mathrm{abd}} \subseteq Obs$ be subsets of observations and let D be a candidate diagnosis. Then D is a diagnosis for S iff*

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

[2] This assumption is not valid in every system. We can, however, transform most systems to a system consisting components with only inputs and outputs (see for instance [3]).

[3] A connection between components is modeled by *connection point* that is shared by one or more inputs and an output. Note that a physical connection should be modeled by component.

[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.

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

Remark In the literature two types of diagnoses are distinguished: *consistency based* [5, 6] and *abductive* [1] diagnosis. Both can be combined into one more general diagnostic definition [2]. This latter definition is used here.

3 Multi-agent diagnosis

A knowledge distribution over multiple agents induces a division of a system S into several subsystems. In the case of a *semantical* knowledge distribution, each agent A_i makes diagnosis of a different *aspect* of the system S. An aspect defines a system S_i of S consisting of a structural description Str_i and a behavioral description Beh_i. A component $c \in C$ has a specific behavior $mode(c, m) \rightarrow \Phi_i \in Beh_{c,i}$ for each fault mode $m \in M_c$ and each aspect i. Of course, given k different aspects, $Beh_c = \bigcup_{i=1}^{k} Beh_{c,i}$ and $\models (\Phi_1 \wedge ... \wedge \Phi_k) \leftrightarrow \Phi$ where Φ is the complete (single agent) behavior of mode m.

Without loss of generality, we may assume that the value of each output of a component is completely determined by the behavior with respect to one aspect. Therefore, also the structural description and the observations are distributed based on the aspect that determines the value of an output or requires the value of an input: Str_i and Obs_i.

By distributing knowledge, i.e. Beh_i and Str_i over the agents, we must provide agents with information about the components' inputs that (i) are needed for the components' behavioral description and that (ii) are determined by aspects that do not belong to the agent's expertise. Other agents must provide the agent i with the values of these inputs. In_i and Out_i will be used to denote the connection points the values of which are provided by other agents, respectively must be passed on to other agents. Hence, $S_i = (C_i, M, Id, Sd_i, Ctx, Obs_i, In_i, Out_i)$ is a subsystem to be diagnosed 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 must make a diagnosis of 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. We use the set V_i to denote value assignments $value(p) = v$, with $p \in In_i$, to the inputs. 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 $Obs_{\mathrm{con},i}$, $Obs_{\mathrm{abd},i} \subseteq Obs_i$ be subsets of the observations, and let V_i be a (partial) descriptions of the values of the connection points In_i. Finally, let D_i be a candidate diagnosis. Then D_i is a diagnosis for S_i iff*

$$D_i \cup Sd_i \cup Ctx \cup V_i \models \bigwedge_{\varphi \in Obs_{\mathrm{abd},i}} \varphi \text{ and } D_i \cup Sd_i \cup Ctx \cup V_i \cup Obs_{\mathrm{con},i} \not\models \bot.$$

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 agent that has complete knowledge of the system description and the observations. To answer this questions we assume *there are no conflicts between the knowledge of the different agents*; i.e. there always exists a diagnosis D such that: $D \cup Sd \cup Cxt \cup Obs$ is consistent. We need this assumption because single agent diagnosis requires consistent knowledge.

Proposition 1 [6] *Let $S_1, ..., S_k$ be the subsystems that make 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.*

Proposition 2 *Let $S_1, ..., S_k$ be the subsystems that make up the system S. Moreover, 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 D is a candidate diagnosis and if for every $i = 1, ...k$: $D_i \cup Sd_i \cup Ctx \cup V_i \mid\sim (value(p) = v)$ for every $p \in Out_i$, $p \in In_j$ and $(value(p) = v) \in V_j$.*

Note that a global diagnosis D is also a diagnosis of the agent A_i if an aspect i plays a role in every component $c \in C$.

The above propositions show that multi-agent diagnosis is possible. Note, however, that given a global candidate diagnosis D, predicting the values of all connection points is an NP-Hard problem [8]. When knowledge of the system is semantically distributed over the agents, often there are only a few connection points between the subsystems managed by different agents. Moreover, if the connections between subsystems do not form cycles, the distribution of knowledge over the agents does not contribute significantly to the time complexity of predicting the system's behavior given a diagnosis. Since usually, there are not many connections between different behavioral aspects of the system, in the remainder of this paper, we will assume that the prediction of the system's behavior is not an issue.

A single agent approach is based on the implicit assumption that an agent has complete and consistent knowledge of a component's behavior given its known behavioral modes. Without this assumption, a single agent cannot make a diagnosis using Definition 1. However, when knowledge is semantically distributed, this assumption need not be valid. Therefore, we must study the consequences of incomplete and incorrect knowledge on establishing a global diagnosis.

Agents with incomplete knowledge When agents look at different aspects of a component, they may not have the same detailed knowledge for every aspect. Concerning the electrical aspects of an integrated circuit for instance, an agent may distinguish many specialized fault modes for which knowledge concerning the thermodynamic aspects of the circuit is lacking. Hence, for a component c an agent A_i may only have behavioral knowledge for some of the component's fault modes $M_{c,i} \subseteq M_c$.

The lack of knowledge about a component's behavior for some fault modes raises a problem: the agents may not be able to reach an agreement. To overcome this problem an agent A_i may just assume a behavior for each fault mode $m \in (M_c - M_{c,i})$. The question is, which behaviors can validly be assumed? If the behavior of a less specific fault mode would be known, this behavior may be used. Since a set of fault modes $M_{c,i}$ always contains the normal mode nor and the least specific fault mode ab (even if no behavior of ab is known), we may assume the existence of a hierarchy of fault modes ordered with respect to specificity. We call such a hierarchy and *abstraction hierarchy*.

[6] The proofs are omitted because of lack of space.

Definition 3 *Let c be a component with M_c as its set of behavior modes. An abstraction hierarchy on M_c is a strict partial order $\succ$ defined on $M_c - \{nor\}$ where the intuitive meaning of $m \succ m'$ is that m is more specific than m' and ab is the unique least specific element in the hierarchy, i.e. for all $m \in M_c - \{nor, ab\}$: $m \succ ab$.*

A more specific mode implies a more specific description of the faulty behavior of the component. Therefore, the following requirement must hold.

> For every $m, m' \in M_{c,i}$: if $m \succ m'$, $mode(c, m) \rightarrow \Phi \in Beh_{c,i}$ and $mode(c, m') \rightarrow \Phi' \in Beh_{c,i}$, then $\models \Phi \rightarrow \Phi'$.

Moreover, we assume that for any component c, $mode(c, ab) \rightarrow \top$ holds. That is, there is no behavioral description for the fault mode ab.

Definition 4 *Let $\Phi_{i,\mathrm{nor}}$ be the normal behavior with respect to aspect i of a component c and let Cst be a set of formulas describing the physical constraints of the world.*[7]

An abstraction hierarchy is complete *iff for each a fault mode m_0, if m_0 is not a most specific fault mode, then there is a set of fault modes $m_1, ..., m_\ell$ such that $m_j \succ m_0$ for $j \geq 1$, $mode(c, m_j) \rightarrow \Phi_{i,j} \in Beh_{c,i}$ and $Cst \models (\Phi_{i,0} \wedge \neg\Phi_{i,\mathrm{nor}}) \leftrightarrow (\Phi_{i,1} \vee ... \vee \Phi_{i,\ell})$.*

The abstraction hierarchy on the fault modes defines a similar abstraction hierarchy on the diagnoses.

Definition 5 *Let D, D' be two candidate diagnoses. D is at least as specific as D', $D \succeq D'$, iff for every $mode(c, m) \in D$ there is a $mode(c, m') \in D'$ such that $m \succeq m'$.*

Note that agents that wish to give a best possible explanation for the observed anomalies, should focus on the *most specific* diagnosis. Whether the agents only determine the most specific diagnoses depends on the type of diagnosis they use; i.e. the choice for the sets Obs_{abd} and Obs_{con}.

Proposition 3 *Pure abductive diagnosis produces only the most specific diagnoses.*

Pure consistency based diagnosis also returns every less specific diagnosis.

Proposition 4 *Let $S_1, ..., S_k$ be the subsystems that make up the system S and let the abstraction hierarchy of fault modes be complete. Moreover, let D be a most specific diagnosis of S. Then there exists a set of most specific diagnoses $D_1, ..., D_k$ for respectively $S_1, ..., S_k$ such that $D = \bigcup_{i=1}^{k} D_i$.*

The behavioral description $Beh_{c,i}$ of a component with respect to an aspect need not specify a behavior for each fault mode in M_c. In order for the agent to establish a global diagnosis, the missing behaviors have to be added. The following assumption serves this purpose.

Assumption *A fault mode m of a component c for which an agent has no behavioral knowledge, has the same behavior as the most specific mode $m' \in M_{c,i}$ such that $m \succeq m'$.*

The assumption extends the behavioral description, making the behavioral knowledge of every fault mode of every component complete for all aspects. Hence, the results of propositions 1 and 2 apply.

[7] Some abnormal behaviors of a component need not be physically possible. For these behaviors, a complete hierarchy need not contain a fault modes describing them.

Agents with incorrect knowledge Agents lacking knowledge about behavior modes is not the only problem that may arise in a multi-agent system. Knowledge of agents may in some situation be incorrect leading to inconsistencies between local diagnoses. Hence, the agents will not be able to agree on a global diagnosis.

A robust multi agent system should be able to handle situations in which inconsistencies between local diagnoses arise. One possibility, which has been proposed in [10], is the use of voting. However, if agents look at different aspects of the system, voting offers no solution. Moreover, voting requires the communication of all local diagnoses of all agents. The number of these diagnoses may be exponential in the number of components.

The abstraction hierarchy on the fault modes also makes it possible to handle inconsistencies. When agents cannot agree on a most specific diagnosis, they may investigate whether they can resolve the inconsistency by looking at less specific diagnoses. If the agents apply pure consistency based diagnosis, such a diagnosis always exits since there is no behavioral description for the fault mode ab. Hence, one or more global diagnoses always exists but these global diagnoses need not correspond with most specific diagnoses of individual agents. An agent may determine several more specific diagnoses which cannot be diagnoses according to other agents. Especially if the abstraction hierarchy of fault modes is complete, knowledge of the agents must be inconsistent.

Proposition 5 *Let $S_1, ..., S_k$ be the subsystems that make up the system S and let the abstraction hierarchy of fault modes be complete. Moreover, let D be a most specific diagnosis of S. Then, the knowledge of two agents A_i and A_j is inconsistent if agent A_i has a diagnosis $D_i \succ D$ and for every diagnosis $D_i' \succ D$ there exist no diagnosis D_j established by agent A_j such that $\{mode(c, m) \in D_i' \mid c \in C_j\} \subseteq D_j$.*

A difficult issue is comparing the quality of the diagnoses. Clearly, the most specific diagnoses are preferred. There is, however, another way in which diagnoses can be distinguished. Assuming that the abstraction hierarchy of fault modes is complete, given two consistency based diagnoses D and D', the diagnosis D may abductively explain an observation φ while a diagnosis D' may not. Clearly, diagnoses that give an better explanation should be preferred.

Definition 6 *A diagnosis D gives a better explanation than diagnosis D' iff* $\{\varphi \in Obs \mid D' \cup Sd \cup Cxt \mid\sim \varphi\} \subseteq \{\varphi \in Obs \mid D \cup Sd \cup Cxt \mid\sim \varphi\}$.

4 Protocols for establishing a diagnostic agreement

The agents may determine a global diagnosis by first determining all fault modes $M_c = \bigcup_{i=1}^{m} M_{c,i}$ as well as the abstraction hierarchy $\succ$ on M_c for each component c, and subsequently exchanging all their local diagnoses. The first step is straight forward and will not be discussed here because of space limitations. The second step is more problematic. The number of diagnoses to be exchanged between the agents can be quite high and can be exponential in the number of component is the worst case. In order to control the complexity, agents should focus on diagnoses in which a minimal number of components, with respect to $\subseteq$, are broken.

Since a local minimal diagnosis need not be a global minimum diagnosis, the agent proposing the diagnosis needs to receive feedback when a proposed diagnosis is rejected

by other agents. Subsequently, the agent can generate a new diagnosis taking into account the diagnoses that have been rejected.

The generation of new minimal diagnoses can be improved if agents supply the reasons for rejecting a proposed diagnosis. When agent A_1 proposes a partial diagnosis D_1, agents $A_2, ..., A_k$ might reject the diagnosis because some (combination of) modes is inconsistent with its observations. Let $R_i \subseteq D_1$ be such (a combination of) modes. Then R_i is a smallest subset of D_1 such that: $R_i \cup Sd_i \cup Ctx \cup Obs_i \mid\sim \perp$ for $2 \leq i \leq k$.

Note that an agent A_i might determine more than one smallest subset R_i. If SR_i is the set of all smallest subsets R_i, agent A_1 can use this information $TR = \bigcup_{2 \leq i \leq k} SR_i$ as a set of constraints in its search for a next diagnosis. It may not select a new diagnosis D_1' containing any $R_i \in TR$ as a subset.

The following simple protocol shows how the agents may proceed. To gain robustness, eventually, always one of the agents takes the initiative to establishes the global diagnoses. In the protocol, the agent that takes the initiative is agent A_1.

Agent	**Action**
A_1	$TR := \varnothing$;
A_1	finished := false;
A_1	while not finished do
A_1	generate the next most specific minimal diagnosis D_1 of S_1
	such that for no $R \in TR$: $R \subseteq D_1$;
A_1	finished := not diagnosis_found;
A_1	while diagnosis_found, for $i := 2$ to k do
A_1	send 'propose D_1' to A_i;
A_i	receive 'propose D_1' from A_1;
A_i	determine a most specific local diagnosis D_i of S_i such that $D_1 \succeq D_i$;
A_i	if a diagnosis D_i exists then;
A_i	send 'accept D_i' to A_1;
A_i	else
A_i	send 'reject SR_i' to A_1;
A_i	end;
A_1	if received 'accept D_i' from A_i then
A_1	$D_1 := D_i$;
A_1	else if received 'reject SR_i' from A_i then
A_1	$TR := TR \cup SR_i$;
A_1	diagnosis_found := false;
A_1	end;
A_1	end;
A_1	if diagnosis_found then
A_1	store D_1;
A_1	end;
A_1	end;

5 Conclusion

In this paper, we analyzed the problem of multi-agent diagnosis when knowledge is semantically distributed over the agents. Especially the case that the agents' knowledge concerning the faulty behavior of some components, is incorrect has been considered. A solution based on an abstraction hierarchy on the fault modes has been proposed and a protocol for determining the global diagnoses with a minimal number of broken components has been given.

An important question for further research is how to order the global minimal diagnoses. A minimal diagnosis in which the knowledge of the agents can be assumed to be correct gives a better explanation of the observed anomalies than a diagnosis that is less specific in order to deal with incorrectness in agents' knowledge. It is not obvious, however, how to order diagnoses implying that the knowledge of some agent cannot be correct in the current situation. Should, for instance, a diagnosis in which all agents assume that a component is broken though they do no agree on the fault mode, be better than a diagnosis in which some agents assume the component to be broken and agree on th fault mode and the other agent assume the component is not broken?

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] J. J. v. 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.

[4] P. Frohlich, I. de Almeida Mora, W. Nejdl, and M. Schroeder. Diagnostic agents for distributed systems. In J.-J. C. Meyer and P.-Y. Schobbens, editors, *Formal Models of Agents. ESPRIT Project ModelAge Final Report Selected Papers. LNAI 1760*, pages 173–186. Springer-Verlag, 2000.

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

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

[7] N. Roos, A. ten Teije, A. Bos, and C. Witteveen. Multi-agent diagnosis: an analysis. In *BNAIC'01*, 2001.

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

[9] N. Roos, A. ten Teije, and C. Witteveen. A protocol for multi-agent diagnosis with spatially distributed knowledge. In *AAMAS 2003*, 2003.

[10] M. Schroeder and G. Wagner. Distributed diagnosis by vivid agents. In *Proceedings of the first conference on Autonomous Agents*, pages 268–275, 1997.