Models and Methods for Plan Diagnosis

Nico Roos[1] and **Cees Witteveen**[2]

Abstract. We consider a model-based diagnosis approach to the diagnosis of plans. Here, a plan executed by some agent(s) is considered as a system to be diagnosed. We introduce a simple formal model of plans and plan execution where it is assumed that the execution of a plan can be monitored by making partial observations of plan states. These observations of plan states are used to compare them with predicted states based on (normal) plan execution. Deviations between observed and predicted states can be explained by qualifying some plan steps in the plan as behaving abnormally. A diagnosis is a subset of plan steps qualified as abnormal that can be used to restore the compatibility between the predicted and the observed partial state. In contrast to model-based diagnosis, where minimum and minimal diagnoses are preferred, we argue that in plan-based diagnosis maximum informative diagnoses should be preferred. These are diagnoses that make the strongest predictions with respect to partial states to be observed in the future. We show that in contrast to minimum diagnoses, finding a (minimal) maximum informative diagnosis can be achieved in polynomial time. Finally, we show how our approach can be extended in a simple way to deal with diagnosis based on iterative timed observations.

1 Introduction

With an increasing complexity of plans, the possibility that something goes wrong during their execution increases correspondingly. No wonder then that more attention is paid to the development of *robust* plans. One way to enhance robustness is to perform plan diagnosis in order to identify the causes of failures, to predict future failures and, if possible, to prevent failures to occur.

In this paper we will concentrate on *internal* failure sources and leave external failure sources such as the environment, failures of executing agents as in [1] or incompatibilities between agents as in [9, 10] for future research. In particular, we will confine ourselves to the identification of failing *actions* as the only source of plan failure.[3] In a multi-agent planning systems, for example, identification of such failing actions can be used to identify agents responsible for executing these actions.

One of the main goals of this paper then is to show how a plan consisting of a partially ordered set of instances of actions can be viewed as a system to be diagnosed and how a diagnosis can be established using *partial observations* of a plan in progress. Distinguishing between normal and abnormal execution of actions in a plan, we then introduce a plan diagnosis as a set of instances of actions qualified as abnormal to explain the deviations between expected plan states and observed plan states. Although plan diagnosis conceived in this way seems to be a rather straightforward application of MBD to plans, we have to recognize some differences, too: First of all, we want to deal with *partial observations* in time of a planning system. That is, taking a plan in execution, in many applications it is simply not possible to observe all the effects of the actions occurring in the plan that have to be executed at some time t. Therefore, we will have to deal with partial observations of (the results of) plan execution at several time steps. Secondly, in diagnoses of systems we usually distinguish only input and output observations. Using the system description S together with some normality assumptions, we can predict the output from the input observation and compare the predicted output with the observed output to establish a diagnosis. In plan diagnosis, this idea of restricting observations to the start and the end of plan execution would severely limit the fruitfulness of diagnosing plans: instead, we would like to apply diagnosis using an arbitrary *sequence* of (partial) observations in order to establish a diagnosis. Thirdly, while in standard MBD usually subset-minimal diagnoses, or within them *minimum (cardinality)* diagnoses, are preferred, we will show that this focusing on minimality of diagnoses alone is not sufficient for plan diagnosis. We would prefer diagnoses that maximize the similarity between predicted and observed plan states. Such diagnoses we will call *maximum informative* diagnoses.

In order to present a general approach to plan diagnosis and to state results that are valid for planning systems in general, we do not adhere to special planning formalisms. As the reader will notice, this framework is general enough to cover the main features of well-known planning formalisms like STRIPS [7].

Results The results obtained in this paper are threefold. First of all we present a formal framework for plan diagnosis that enables us to define exactly how observations of a plan in execution can be used to derive a plan diagnosis. We show that establishing a plan diagnosis comes down to finding a subset of actions in a plan such that if these actions are qualified as abnormal, the observed plan states are compatible with predicted plan states. Secondly, after introducing minimal maximum informative diagnoses (abbreviated as *mini-maxi diagnoses*) as a special kind of diagnoses that have to be preferred above the well-known subset-minimal and minimum diagnoses known from model-based diagnosis, we show that in contrast to minimum diagnoses, mini-maxi diagnoses can be computed efficiently. Thirdly, we extend the framework to plan diagnosis based on iterative partial observations and we show how this case can be reduced to establishing diagnoses with a simple pair of observations.[4]

[1] Department of Computer Science, Universiteit Maastricht, P.O.Box 616, NL-6200 MD Maastricht, email: roos@cs.unimaas.nl

[2] Faculty EEMCS, Delft University of Technology, P.O.Box 5031, NL-2600 GA Delft, email: C.Witteveen@tudelft.nl

[3] If the plan is correctly specified, every error in the plan execution process becomes manifest in the incorrect behavior of one or more instances of actions occurring in the plan. Whether or not we should be satisfied with the identification of one or more of such failing actions, a diagnostic process that identifies a set of actions that can be shown to be responsible for the abnormalities observed is a useful analysis on its own.

[4] An earlier version of the framework has appeared in [15] without the discussion of maximum informative diagnoses, algorithms and iterative obser-

Organization This paper is organized as follows. In the next section, we place our approach into perspective by discussing some related approaches to plan diagnosis. Then we introduce a simple formal framework for representing states, actions and plans. In Section 3, we introduce the main concepts of plan-based diagnosis, while Section 4 formalizes plan-based diagnosis, introduces the idea of maximum informative diagnosis. In this section, we also discuss an efficient algorithm to find minimal maximum informative diagnoses. In Section 6, we extend our framework to diagnosis with a sequence of observations and Section 7 concludes this paper with a brief outlook to future research. Note that due to lack of space, all proofs of results have been omitted.

2 Related research

In this section we briefly discuss some other approaches to plan diagnosis we already alluded to in the introduction.

Birnbaum et al. [1] apply MBD to *planning agents* relating health states of agents to *outcomes* of their planning activities. They do not take into account abnormalities that can be attributed to actions in a plan as a separate source of errors. In contrast to their approach, we do not take into account abnormalities of the executing agents, but exclusively focus upon the detection of abnormal actions in the plan. As we already remarked above, we feel that such an approach focusing upon actions as the immediate factors underlying abnormal plan behavior should precede more elaborate failure analyses.

Another approach that directly applies model-based diagnosis to plan execution has been proposed in De Jonge et al. [5]. Here, the authors focus on agents each having an individual plan, and where conflicts between these plans may arise (e.g. if they require the same resource). Diagnosis is applied to determine those factors — not limited to actions — that are accountable for *future* conflicts.

Kalech and Kaminka [9, 10, 11, 12] apply *social diagnosis* in order to find the cause of an anomalous plan execution. They consider hierarchical plans consisting of so-called *behaviors*. Such plans do not prescribe a (partial) execution order on a set of actions. Instead, based on its observations and beliefs, each agent chooses the appropriate behavior to be executed. Each behavior in turn may consist of primitive actions to be executed, or of a set of other behaviors to choose from. Social diagnosis then addresses the issue of determining what went wrong in the joint execution of such a plan by identifying the disagreeing agents and the causes for their selection of incompatible behaviors (e.g., belief disagreement, communication errors). This approach might complement our approach when conflicts not only arise as the consequence of faulty actions, but also as the consequence of different selections of sub-plans in a joint plan.

Lesser et al. [2, 8] also apply diagnosis to (multi-agent) plans. Their research concentrates on the use of a *causal model* that can help an agent to refine its initial diagnosis of a failing *component* (called a *task*) of a plan. As a consequence of using such a causal model, the agent would be able to generate a new, situation-specific, plan that is better suited to pursue its goal. While their approach in its ultimate intentions (establishing anomalies in order to find a suitable plan repair) comes close to our approach, their approach to diagnosis concentrates on specifying the exact causes of the failing of one single *component* (task) of a plan. Diagnosis is based on observations of a component without taking into account the consequences of failures of such a component with respect to the remaining plan. In our approach, instead, we are interested in applying MBD-inspired methods to *detect* plan failures. Such failures are based on observations during plan execution and may concern individual components of the plan. Furthermore, we do not only concentrate on identifying failing components themselves, but also on the consequences of these failures for the future execution of plan elements.

vations.

3 Preliminaries

3.1 Plans as Systems

We consider plan-based diagnosis as a simple extension of the model-based diagnosis (MBD) approach [3, 4, 14], where the model is not a description of an underlying physical system but a *plan* of one or more agents. To keep this model simple and general, we will keep the plan representation details minimal.

3.1.1 States

By executing plans we change a part of the world. Therefore, before we discuss plans, we need to introduce a simple state-based view on the world. We assume that for the planning problem at hand, the world can be described by a set $Var = \{v_1, v_2, \ldots, v_n\}$ of variables and their respective *value domains* D_i. A *complete state of the world* σ then is a value assignment $\sigma : Var \rightarrow \bigcup_{i=1}^{n} D_i$ to the variables. Slightly abusing terminology, we simply denote a complete state by an n-tuple $\sigma = (\sigma(v_1), \ldots, \sigma(v_n)) \in D_1 \times D_2 \times \ldots \times D_n$. A *partial state* is an element $\pi \in D_{i_1} \times D_{i_2} \times \ldots \times D_{i_k}$, where $1 \leq k \leq n$ and $1 \leq i_1 < \ldots < i_k \leq n$. We use $Var(\pi)$ to denote the set of variables $\{v_{i_1}, v_{i_2}, \ldots, v_{i_k}\} \subseteq Var$ specified in such a partial state π. The value $\sigma(v_j)$ of variable $v_j \in Var(\pi)$ will be denoted by $\pi(v_j)$. The value of a variable $v_j \in Var$ not occurring in a partial state π is said to be *undefined* (or *unpredictable*) in π, denoted by $\perp$. Including $\perp$ in every value domain D_i allows us to consider every partial state π as an element of $D_1 \times D_2 \times \ldots \times D_n$.

Partial states can be ordered with respect to their information content: Given values d and d', we say that d' is at least as informative as d, abbreviated as $d \leq d'$, iff $d = \perp$ or $d = d'$. The containment relation $\sqsubseteq$ between partial states is the point-wise extension of $\leq$: π is said to be contained in π', denoted by $\pi \sqsubseteq \pi'$, iff $\forall v \in Var[\pi(v) \leq \pi'(v)]$.

An important notion in plan diagnosis is the notion of *compatibility* between partial states. Intuitively, two states π and π' are said to be compatible if there is no essential disagreement about the values assigned to variables in the two states, i.e., in principle they could characterize the same state of the world. More exactly, compatibility implies that for every $v \in Var$, either $\pi(v) = \pi'(v)$ or at least one of the values $\pi(v)$ and $\pi'(v)$ is undefined:

Definition 1 (compatibility relation) *Two partial states π and π' are said to be compatible, denoted by $\pi \approx \pi'$, if $\forall v \in Var[\ \pi(v) \leq \pi'(v)$ or $\pi'(v) \leq \pi(v)\ ]$.*

If two partial states π_1 and π_2 are compatible, their information content can be *fused* to obtain a new partial state $\pi = \pi_1 \sqcup \pi_2$ that contains them both: $\pi = \pi_1 \sqcup \pi_2$ holds iff $\forall v \in Var[\pi(v) = \max_{\leq}\{\pi(v), \pi'(v)\}]$.

3.1.2 Actions, Plan operators and Plan Steps

In the preceding sections we used to term "actions" in a rather informal way. From now on, we make a distinction between *actions*, *plan operators* and *plan steps*. First of all, an *action* refers to an activity that results in some change of the (current) state of the world, such as loading a vehicle or assembling components. A *plan operator* refers to a description of such an action in a plan. More exactly, a plan operator o is a function mapping partial states to partial states by replacing the values of a subset $V_o \subseteq Var$ by other values (dependent upon the values of another set $V'_o \supseteq V_o$ of variables). Hence, every plan operator o can be modeled as a (partial) function $f_o : D_{i_1} \times \ldots \times D_{i_k} \rightarrow D_{j_1} \times \ldots \times D_{j_l}$, where $1 \leq i_1 < \ldots < i_k \leq n$ and $\{j_1, \ldots, j_l\} \subseteq \{i_1, \ldots, i_k\}$. The variables whose value domains occur in $dom(f_o)$ will be denoted by $dom_{Var}(o) = \{v_{i_1}, \ldots, v_{i_k}\}$ and, likewise, $ran_{Var}(o) = \{v_{j_1}, \ldots, v_{j_l}\}$. It is required that $ran_{Var}(o) \subseteq dom_{Var}(o)$, i.e., the function f_o associated with a plan operator is *range-restricted*. This functional specification f_o constitutes the *normal* behavior of the plan operator o, also denoted by f_o^{nor}.

Example Figure 1 depicts two states σ_0 and σ_1 (the white boxes) each characterized by the values of four variables v_1, v_2, v_3 and v_4. The partial states π_0 and π_1 (the gray boxes) characterize a subset of variables in a (complete) state. Plan operators are used to model state changes. The domain of the plan operator o is the subset $\{v_1, v_2\}$, denoted by the arrows pointing to o. The range of o is the subset $\{v_1\}$, which is denoted by the arrow pointing from o. Finally, the dashed arrow denotes that the value of variable v_2 is not changed by operator causing the state change.

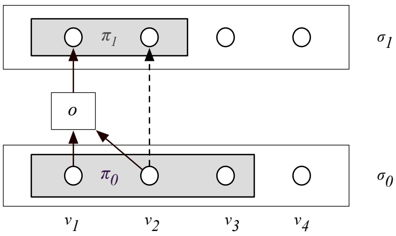

Figure 1. Plan operators, states and partial states

■

A plan operator o may be used at several places in a plan. A specific occurrence of o is called a *plan step* mapping a specific partial state into another partial state. A plan step s as an occurrence of o then describes a specific function application of the function specified by the operator o at a specific place in the plan. Therefore, given a set $\mathcal{O}$ of plan operators, we consider a set $S = inst(\mathcal{O})$ of *instances* of plan operators in $\mathcal{O}$, called the set of plan steps. A plan step will be denoted by a small roman letter s_i. The plan operator o the instance s_i was instantiated from is denoted by $op(s_i)$. If $op(s_i) = o$, the instance s_i is said to be of *type o*.

3.1.3 Plans and Plan Execution

A plan is a tuple $P = \langle \mathcal{O}, S, < \rangle$ where $S \subseteq Inst(\mathcal{O})$ is a set of plan steps occurring in $\mathcal{O}$ and $(S, <)$ is a partial order. The partial order relation $<$ specifies an *execution relation* between plan steps: for each $s \in S$ it holds that s is executed immediately after all plan steps s' such that $s' < s$ have been finished. We will denote the *transitive reduction* of $<$ by $\ll$, i.e., $\ll$ is the smallest subrelation of $<$ such that the transitive closure $\ll^+$ of $\ll$ equals $<$.

Example Figure 3.1.3 gives an illustration of a plan. Arrows relate the objects a plan step uses as inputs and the objects it produces as its outputs to the plan step itself. In this plan, the direct execution relation is specified as $s_1 \ll s_3$, $s_2 \ll s_4$, $s_4 \ll s_5$ and $s_4 \ll s_6$. ■

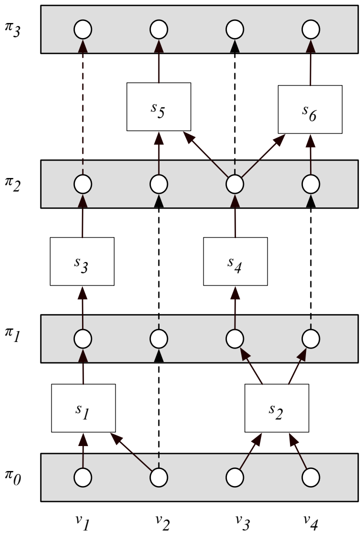

Figure 2. Plans and plan steps. Each state characterizes the values of four variables v_1, v_2, v_3 and v_4. States are changed by application of plan steps s_i for $i = 1, 2, \ldots, 6$.

Without loss of generality, we assume that very plan step $s \in S$ takes a unit of time to execute and the execution of the first plan step starts at time $t = 0$. Using this assumption and the definition of the execution relation $<$, the time t at which a plan step s will be executed is uniquely determined: Let $depth_P(s)$ be the depth of plan step s in plan $P = \langle \mathcal{O}, S, < \rangle$.[5] Then the time t_s at which the plan step s is executed is $t_s = depth_P(s)$ and s will be completed at time $t_s + 1$. Let P_t denote the set of plan steps s with $depth_P(s) = t$, let $P_{>t} = \bigcup_{t'>t} P_{t'}$, $P_{<t} = \bigcup_{t'<t} P_{t'}$ and let $P_{[t,t']} = \bigcup_{k=t}^{t'} P_k$.

Example Consider again Figure 3.1.3. In this plan, the depth of s_1 and s_2 is 0, the depth of s_3 and s_4 is 1, and the depth of s_5 and s_6 is 2. Therefore, $P_0 = \{s_1, s_2\}$, $P_1 = \{s_3, s_4\}$ and $P_2 = \{s_5, s_6\}$. ■

Given a state σ at some time t and the set P_t of plan steps to be executed at time t we want to be sure that the next state σ' at time $t + 1$ is uniquely defined. If P_t contains two plan steps s and s' with overlapping ranges, i.e., if $ran_{Var}(s) \cap ran_{Var}(s') \neq \varnothing$, the final result of a variable v occurring in this intersection is not uniquely defined in σ'. We therefore assume the following condition to hold:

Determinism: If P is a plan and s, s' are plan steps in P such that $ran_{Var}(s) \cap ran_{Var}(s') \neq \varnothing$ then $depth_P(s) \neq depth_P(s')$.

It is not difficult to see (and can be easily proven using the derivability relations to be discussed) that Determinism guarantees that a future plan state can be defined uniquely given a plan and a currently uniquely defined plan state.

[5] Here, $depth_P(s) = 0$ if $\{s' \in S \mid s' \ll s\} = \varnothing$ and $depth_P(s) = 1 + max\{depth_P(s') \mid s' \ll s\}$, else. If the context is clear, we omit the subscript P.

3.2 Qualifications

The correct execution of a plan step may fail either because of an inherent malfunctioning or because of a malfunctioning of an agent responsible for executing the action, or because of unknown external circumstances. In all these cases we would like to model the effects of executing such failing plan operators. Therefore, we introduce a set of *health modes* H_s for each plan step s. This set H_s contains at least the normal mode nor, the mode ab indicating the most general abnormal behavior, and possibly several other specific fault modes. The most general abnormal behavior of plan step s is specified by the function f_s^{ab}, where $f_s^{ab}(d_{i_1}, d_{i_2}, \ldots, d_{i_k}) = (\bot, \bot, \ldots, \bot)$ for every partial state $(d_{i_1}, d_{i_2}, \ldots, d_{i_k}) \in dom(f_o)$.[6] To keep the discussion simple, in the sequel we distinguish only the health modes nor and ab.

Let us assume, for the moment, that each plan step can be viewed as an independent component of a plan. To each plan step then s a health mode $h_s \in \{nor, ab\}$ can be assigned and the result is called a *qualified* plan. In establishing which part of the plan fails, we are only interested in those plan steps qualified as abnormal. Therefore, we define a qualified version P_Q of a plan $P = \langle \mathcal{O}, S, < \rangle$ as a tuple $P_Q = \langle \mathcal{O}, S, <, Q \rangle$, where $Q \subseteq S$ is the subset of plan steps qualified as abnormal (and therefore, $S - Q$ is the subset of plan steps qualified as normal).

Since a qualification Q corresponds to assigning the health mode ab to every plan step in Q and since $f_s^{ab}(d_{i_1}, d_{i_2}, \ldots, d_{i_k}) = (\bot, \bot, \ldots, \bot)$ for every plan step $s \in Q$ with $type(s) = o$, the results of anomalously behaving plan steps are unpredictable. Note that a "normal" plan P corresponds to the qualified plan $P_\varnothing$ and that in our context "undefined" is considered to be equivalent to "unpredictable".

3.3 Derivability relations induced by plan execution

Note that P on a given initial state π_0 will induce a sequence of states $\pi_0, \pi_1, \ldots, \pi_k$, where π_{t+1} is generated from π_t by applying the set of plan steps P_t to σ_t. To define this relation between partial states at different time points we denote a partial state π at a given time t by a tuple, also called a *timed state*, denoted by (π, t).

Let s be a plan step. We say that s is *enabled* in a state π if $dom_{Var}(s) \subseteq Var(\pi)$. Intuitively, we can predict the timed state $(\pi', t+1)$ using the timed state (π, t) and the set P_t of to be executed plan steps as follows:

1. whenever a variable v does not occur in the range of a plan step $s \in P_t$, its value in state π' is the same as its value in π, i.e., $\pi(v) = \pi'(v)$;
2. if the variable v occurs in the range of a normally qualified plan step s that is enabled in π, then $\pi'(v) = f_s^{nor}(\pi(v))$;
3. in all other cases, there is not sufficient information to predict the value of $\pi'(v)$, either because v occurs in the range of an abnormally qualified plan step s or v occurs in the range of some plan step $s \in P_t$ not enabled in π.

Formally, this relation is defined as follows:

Definition 2 *We say that $(\pi', t+1)$ is (directly) generated by execution of the Q-qualified plan P_Q from (π, t), abbreviated by $(\pi, t) \rightarrow_{Q;P} (\pi', t+1)$, iff for every $v \in Var$ the following conditions hold:*

1. *if $v \notin \bigcup_{s \in P_t} ran_{Var}(s)$ then $\pi'(v) = \pi(v)$;*
2. *if $v \in \bigcup_{s \in P_t - Q} ran_{Var}(s)$ then $\pi'(v) = f_s^{nor}(\pi)(v)$;*
3. *else $\pi'(v) = \bot$.*

It is easy to see that thanks to Determinism, the immediate derivability relation $\rightarrow_{Q;P}$ is well-defined and deterministic:

Proposition 1 *Let P_Q be a qualified plan and let (π, t) a timed state. Then $(\pi, t) \rightarrow_{Q;P} (\pi', t+1)$ and $(\pi, t) \rightarrow_{Q;P} (\pi'', t+1)$ implies $\pi'' = \pi'$.*

We extend this direct derivability relation to a general derivability relation in a straightforward way:

Definition 3 *For arbitrary values of $t \leq t'$ we say that (π', t') is (directly or indirectly) generated by execution of P_Q from (π, t), denoted by $(\pi, t) \rightarrow^*_{Q;P} (\pi', t')$, iff the following conditions hold:*

1. *if $t = t'$ then $\pi' = \pi$;*
2. *if $t' = t+1$ then $(\pi, t) \rightarrow_{Q;P} (\pi', t')$;*
3. *if $t' > t+1$ then there must exists some state $(\pi'', t'-1)$ such that $(\pi, t) \rightarrow^*_{Q;P} (\pi'', t'-1)$ and $(\pi'', t'-1) \rightarrow_{Q;P} (\pi', t')$.*

Note that $(\pi, t) \rightarrow^*_{\varnothing;P} (\pi', t')$ denotes the normal execution of a normal plan $P_\varnothing$. Such a normal plan execution will also be denoted by $(\pi, t) \rightarrow^*_P (\pi', t')$.

Using these definitions, it is not difficult to show that for every $0 \leq t \leq k$, the timed state (π', t) where $(\pi, 0) \rightarrow^*_{Q;P} (\pi', t)$ is uniquely defined if $<$ satisfies the Determinism requirement.

Example Figure 3 gives an illustration of an execution of a plan with abnormal plan steps. Suppose plan step s_3 is abnormal and generates a result that is unpredictable ($\bot$). Given the qualification $Q = \{s_3\}$ and the partially observed state π_0 at time point $t = 0$, we predict the partial states π_i as indicated in Figure 3, where $(\pi_0, t_0) \rightarrow^*_{Q;P} (\pi_i, t_i)$ for $i = 1, 2, 3$.

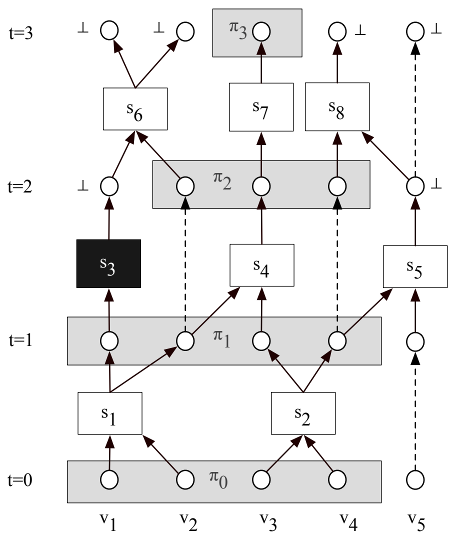

Figure 3. Plan execution with an abnormal plan step (s_3).

Note that since the value of v_1 and of v_5 cannot be predicted at time $t = 2$, the result of plan step s_6 and of plan step s_8 cannot be predicted and π_3 contains only the value of v_3. ■

[6] This definition implies that the behavior of abnormal steps is essentially unpredictable.

4 Observations and Diagnoses

To establish plan diagnosis in our framework we need to make *observations*. Our framework provides a natural candidate for representing such observations: an observation $obs(t)$ at time t is a timed state (π, t) where π is a partial state. This implies that we do not require observations to specify a complete state. Suppose we have an observation $obs(t) = (\pi, t)$ and an observation $obs(t') = (\pi', t')$ at some later time $t' > t \geq 0$ during the execution of the plan P. We would like to use these observations to infer the health modes of the plan steps occurring in P. First, assuming a normal execution of P, we can predict the partial state of the world at a time point t' given the observation $obs(t)$: if all plan steps behave normally, we predict the timed state $(\pi'_{\varnothing}, t')$ such that $obs(t) \rightarrow^*_P (\pi'_{\varnothing}, t')$. Such a prediction has to be compared with the actual observation $obs(t') = (\pi', t')$ made at time t'. It is easy to see whether the predicted state and the observed state match: in that case we should be able to find a state σ such that both the observed state π' and the predicted state $\pi'_{\varnothing}$ both are contained in σ, that is, $\pi' \sqsubseteq \sigma$ and $\pi'_{\varnothing} \sqsubseteq \sigma$. By definition of compatibility, such a π' can only exist if $\pi'_{\varnothing}$ and π' are *compatible* states, i.e. if $\pi' \approx \pi'_{\varnothing}$ holds.[7] If this is not the case, the execution of some plan steps must have gone wrong and we have to determine a qualification Q such that the predicted state π'_Q derived using Q is compatible with π'. Hence, we have the following straight-forward extension of the diagnosis concept in MBD to plan diagnosis (cf. [4]):

Definition 4 *Let $P = \langle \mathcal{O}, S, < \rangle$ be a plan with observations $obs(t) = (\pi, t)$ and $obs(t') = (\pi', t')$, where $t < t' \leq depth(P)$ and let $obs(t) \rightarrow^*_{Q;P} (\pi'_Q, t')$ be a derivation using P_Q. Then Q is said to be a* plan diagnosis *of $\langle P, obs(t), obs(t') \rangle$ iff $\pi' \approx \pi'_Q$.*

In order to be able to establish a diagnosis, we simply assume that for every variable v there exists at least some plan step s and some time $t \leq t'' \leq t'$ such that $s \in P_{t''}$ and $v \in ran_{Var}(s)$.

Example Consider again Figure 3 and suppose that we did not know that plan step s_3 was abnormal and that we observed $obs(0) = ((d_1, d_2, d_3, d_4), 0)$ and $obs(3) = ((d'_1, d'_3, d'_5), 3)$. Using the normal plan derivation relation starting with $obs(0)$ we will predict a state $\pi'_{\varnothing}$ at time $t = 3$ where $\pi'_{\varnothing} = (d''_1, d''_2, d''_3)$. If everything is ok, the values of the variables predicted as well as observed at time $t = 3$ should correspond, i.e. we should have $d'_j = d''_j$ for $j = 1, 3$. If, for example, only d'_1 would differ from d''_1, then we could qualify s_6 as abnormal, since then the predicted state at time $t = 3$ using $Q = \{s_6\}$ would be $\pi'_Q = (d''_3)$ and this partial state agrees with the predicted state on the value of v_3. ■

Remark Note that for all variables in $Var(\pi') \cap Var(\pi'_Q)$, the qualification Q provides an *explanation* for the observation π' made at time point t'. Hence, for these variables the qualification provides an *abductive diagnosis* [3]. For all observed variables in $Var(\pi') - Var(\pi'_Q)$, no value can be predicted given the qualification Q. Hence, by declaring them to be unpredictable, possible conflicts with respect to these variables if a normal execution of all plan steps is assumed, are resolved. This corresponds with the idea of a *consistency-based diagnosis* [14]. ■

4.1 Maximal informative diagnoses

On intuitive grounds, one would prefer, like in MBD, *smaller* diagnoses above larger ones. One of the intuitions behind this preference

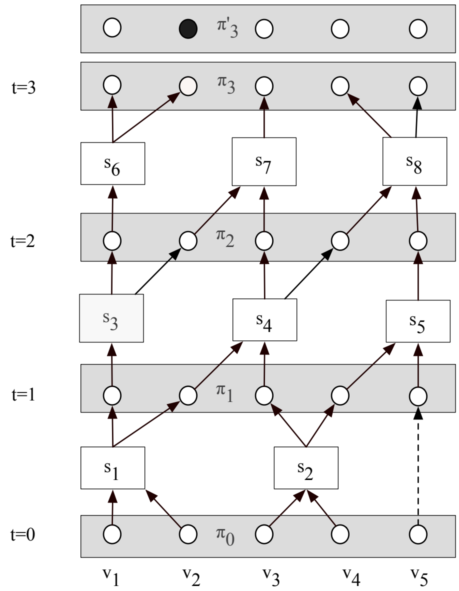

Figure 4. Plan execution with an observation deviating from the expected observation, as indicated by the black dot.

is that, normally, we expect a plan to deliver correct results. Any deviation from this normality assumption should be as small as possible and we prefer a qualification that does not contain more actions qualified as abnormal than necessary. This, like in MBD, would be an obvious reason to prefer *subset-minimal* diagnoses and especially *minimum* diagnoses among the set of minimal diagnoses. These notions can be easily defined in our framework as follows: Given plan observations $\langle P, (\pi, t), (\pi', t') \rangle$, a qualification Q is said to be

1. a *(subset) minimal plan diagnosis* if for every plan diagnosis Q' such that $Q' \subseteq Q$, it holds that $Q = Q'$.
2. a *minimum plan diagnosis* if for every plan diagnosis Q', it holds that $|Q| \leq |Q'|$.

Example Consider the plan depicted in Figure 4. Suppose $obs(0) = (\pi_0, 0)$ and $obs(3) = (\pi'_3, 3)$ and π'_3 equals π_3 except that there is a deviation in the value of v_2 at time $t = 3$ (as indicated by the black dot). Note that there are three possible minimal diagnoses that are also minimum diagnoses : $Q_1 = \{s_1\}$, $Q_2 = \{s_3\}$ and $Q_3 = \{s_6\}$. Let π'_{Q_i} denote the state derived at time $t = 3$ by using Q_i as a qualification. Then $Var(\pi'_{Q_1}) = \varnothing$, $Var(\pi'_{Q_2}) = \{v_4, v_5\}$ and $Var(\pi'_{Q_3}) = \{v_3, v_4, v_5\}$, so these partial states predicted differ in their information content.

■

Example 4.1 shows that, in general, minimum or minimal diagnoses might considerably differ in their predictive power. For example, if we take D_1 as a diagnosis, the values of all variables predicted at time $t = 3$ will be undefined, while taking D_3 as a diagnosis, only v_1 and v_2 are undefined. Hence, it seems that minimality as the single criterion to choose a suitable diagnosis is not sufficient. Intuitively, besides minimizing the *number of abnormal plan steps*, we would prefer those diagnoses Q that generate a state π'_Q that minimizes the *number of undefined values*. We call such diagnoses *maximally informative diagnoses*:

Definition 5 (maxi-diagnosis) *Given plan observations $\langle P, (\pi, t), (\pi', t') \rangle$, a diagnosis Q is said to be a* maximally

[7] See the definition in the preliminaries.

informative plan diagnosis, *abbreviated* maxi-diagnosis, *if there exists no diagnosis Q' such that $Var(\pi_Q) \subset Var(\pi_{Q'})$.*

Note that for a given state π, $Var(\pi)$ is the set of variables defined in π.

Remark Analogous to the distinction between minimal diagnoses and minimum diagnoses, we could introduce the notion of a *maximum* informative diagnosis as a diagnosis Q for which there exists no diagnosis Q' such that $|Var(\pi_Q)| < |Var(\pi_{Q'})|$. Unlike minimal and minimum diagnoses, however, it turns out that every *maximally* informative diagnosis is also a *maximum* informative diagnosis, i.e., there is no distinction between the subset-maximal en the cardinality-maximal notions of informative diagnoses. In fact, we can show an even stronger result: given some plan observations $\langle P, (\pi, t), (\pi', t')\rangle$, for every two maxi-diagnoses Q and Q', it holds that $Var(\pi_Q) = Var(\pi_{Q'})$, i.e., they are equally informative in a strict sense. ■

Such maxi-diagnoses however, do not necessarily be *subset minimal* diagnoses. By combining the two criteria, however, we obtain a qualification that is able to achieve compatibility with the observations, being as exact in its predictions as possible, without considering too many actions as behaving abnormally. We therefore define a *minimal maximally-informative* diagnosis as follows:

Definition 6 (mini-maxi diagnosis) *Given plan observations $\langle P, (\pi, t), (\pi', t')\rangle$, a diagnosis Q is said to be a* minimal maximally informative plan diagnosis, *abbreviated as* mini-maxi diagnosis, *if* (i) *Q is a maxi-diagnosis and* (ii) *there exists no maxi-diagnosis Q' such that $Q' \subset Q$.*

4.2 Finding maxi-diagnoses

Finding minimum diagnoses is computationally hard, even in our simple framework. Surprisingly, however, finding maxi-diagnoses and even finding mini-maxi-diagnoses is tractable. We will first give an intuitive description of an efficient procedure to find a (mini-) maxi diagnosis and then give a polynomial algorithm for finding a mini-maxi diagnosis.

Suppose we have plan observations $\langle P, (\pi, t), (\pi', t')\rangle$. To determine a maxi-diagnosis, we first determine the *disagreement set* $Dis_\varnothing$ of all those variables whose values are defined in both the observed state π' and the predicted state $\pi'_\varnothing$ at time t' but differ:

$$Dis_\varnothing = \{v \in Var \mid [\pi'_\varnothing(v) \neq \pi'(v) \wedge (\pi'(v) > \bot) \wedge (\pi'_\varnothing(v) > \bot)]\}$$

Next, we collect all plan steps s at time $t' - 1$ such that there exists a variable $v \in ran_{Var}(s) \cap Dis_\varnothing$. By the determinism requirement, two different plan steps s and s' occurring in some set P_t cannot have a variable in common in their range, hence for every $v \in Dis_\varnothing$ there is at most one plan step $s_v \in P_{t'-1}$ such that $v \in ran_{Var}(s)$. Then we remove all variables v that occur in the range of the plan steps just selected from the disagreement set. For $i = 2, 3, \ldots$, we iteratively select new plan steps at times $t' - i$ having a variable in their range that also occurs in the disagreement set and we remove these variables until the disagreement set is empty. It is not difficult to see that this procedure generates the set $Q_{max} = \{s_v \mid v \in Dis_\varnothing\}$ where s_v is the latest plan step in the plan causing the value v to occur in the disagreement set. It can be easily proven that Q_{max} is a maxi-diagnosis.

In order to obtain a mini-maxi diagnosis, we have to refine this procedure slightly. Firstly, let us introduce the notion of a scope of a plan step s. Intuitively, the scope of a plan step s contains all variables v that will become undefined sooner or later if s is qualified as abnormal. The scope $scope(s)$ is inductively defined as follows: *(i)* $ran_{Var}(s) \subseteq scope(s)$ and *(ii)* if $\exists s'[scope(s) \cap dom_{Var}(s') \neq \varnothing$ then $scope(s') \subseteq scope(s)$. Now, in the above procedure to generate a maxi-diagnosis, if we simply add a set S_i of new plan steps belonging to $P_{t'-i}$ to the already selected set of plan steps S, some of the plan steps s occurring in S_i might contain variables v' in their scope that also occur in the domain of plan steps $s' \in S$ already selected. That implies $scope(s) \supseteq scope(s')$: hence, adding such a plan step s makes the inclusion of the previously added plan steps s' superfluous. Therefore, at each iteration step, we remove such redundant plan steps s' to obtain a mini-maxi diagnosis.

The following algorithm (see Algorithm 1 states an iterative procedure to obtain a mini-maxi diagnosis Q_{max}:

Algorithm 1 Algorithm to compute mini-maxi diagnoses

Require: plan observations $\langle P, (\pi, t), (\pi', t')\rangle$
Ensure: a mini-maxi informative diagnosis Q_{max}
  Let $Dis_0 = Dis_\varnothing$ and let $Q_{max} = \varnothing$;
  $i := 0$
  while $Dis_0 \neq \varnothing$ **do**
    $i := i + 1;$
    $S_i := \{s \in P_{t'-i} \mid \exists v \in Dis_0[v \in ran_{Var}(s)]\};$
    $Q_i := \{s \in Q_{max} \mid \exists s' \in S_i[s \in scope(s')]\};$
    $Q_{max} := (Q_{max} - Q_i) \cup S_i;$
    $Dis_0 := Dis_0 - \bigcup_{s \in Q_{max}} ran_{Var}(s)$
  end while
  return Q_{max}

Example Consider again the plan execution depicted in Figure 4. Given *obs*(0) and *obs*(3) and a deviation in the value of s_2 at time $t = 3$, we determine the disagreement set $Dis = \{s_2\}$. After selecting s_6 as a plan step to be included in the diagnosis, the disagreement set is empty. Hence, $D = \{s_6\}$ is a maxi-diagnosis. ■

5 Diagnosing a sequence of observations

Until now we discussed the diagnosis of a plan P using (simple) plan observations: we considered diagnoses based on two observations of P at different time points $t < t'$. Considering the plan P as a system to be diagnosed, there is a direct correspondence between MBD and plan diagnosis: the observation *obs*(t) at the earliest time point corresponds to observing the inputs of the system, while the observations *obs*(t') at the latest time point corresponds to observing the outputs. Plan diagnosis, however is not limited to making observations at two different points of time. For it may happen that during the execution of a plan we are able to make a sequence of $k > 2$ observations at some specific time points $t_1 < t_2 < \ldots < t_k$.

In this section we will adapt the definition of a plan diagnosis to such a sequence of observations. We will first make a careful analysis of the adaptations to be made by discussing a simple example.

Example Consider the plan P as depicted in Figure 5 (a). There are three observations (π_0, t_0), (π_1, t_1) and (π_2, t_2). Using the observation (π_0, t_0) and assuming no faulty plan steps, we predict the partial state $(\pi'_{0,1}, 1)$ at time $t = 1$ as depicted in Figure 5 (b). Note that this predicted state is compatible with the observed state (π_1, t_1).

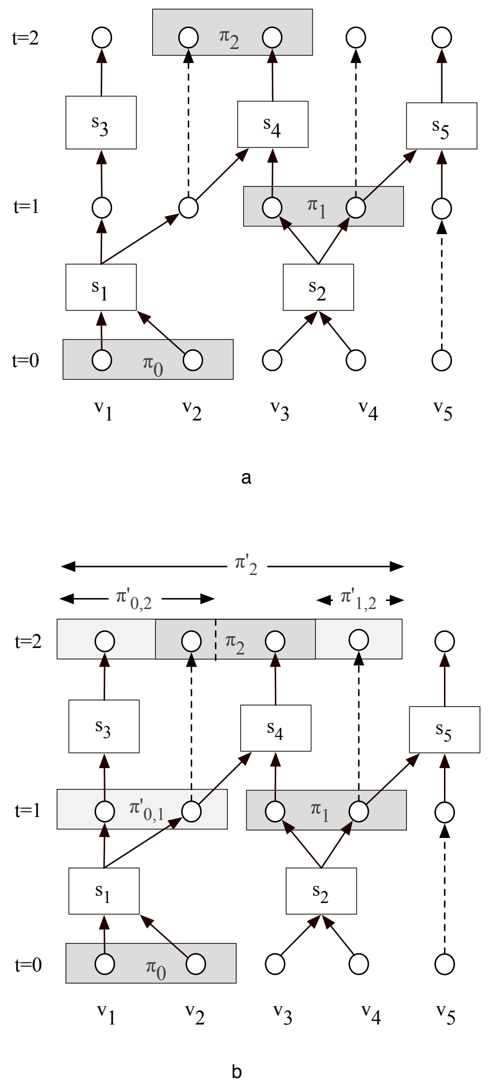

Figure 5. A plan with a sequence of three observations $(\pi_0, 0)$, $(\pi, 1)$ and $(\pi_2, 2)$ (a) and the predictions (b) that can be derived using these observations.

Using the same observation (π_0, t_0), we also predict an observed state $(\pi'_{0,2}, 2)$ at time $t = 2$ where only the variables v_1 and v_2 are defined. Let us suppose that this prediction is compatible with the observed state π_2 at time $t = 2$.

The observation (π_1, t_1) can also be used to obtain information about the state of the plan at time $t = 2$. In this case, however, using (π_1, t_1) the empty partial state $\pi'_{1,2} = (\bot, \ldots, \bot)$ is predicted. This state, by definition, is compatible with any prediction or observation made at time $t = 2$. Therefore, we could conclude that the fusion $\pi'_{0,2} \sqcup \pi'_{1,2} \sqcup \pi_2$ represents the total information that can be derived from both observations at time $t = 2$, assuming that the plan is executed correctly and that the prediction $\pi'_{0,2}$ is compatible with the observation π_2.

However, in this way we did not use all the information available at time $t = 1$ to make a prediction for the state of the plan at time $t = 2$. For example, we are not able to detect whether the value of v_3 deviates from the prediction that can be made if we systematically *combine* the predictions using both the observations π_0 and π_1. For example, since the predicted state $\pi'_{0,1}$ and the observed state π_1 are compatible, the total state information available at time $t = 1$ is the fused state $\pi'_{0,1} \sqcup \pi_1$. From this latter state we are able to predict the partial state π'_2 at time $t = 2$ where $Var(\pi'_2) = \{v_1, v_2, v_3, v_4\}$ and therefore, we could detect whether π_2 at variable v_3 is compatible with this prediction. We conclude that we have to carefully combine all the derivations made from previous observations with the current observed state information to make predictions for the state of the plan at a future time. ■

To model the case where a sequence of $k > 2$ observations is made, we consider a plan $P = \langle \mathcal{O}, S, < \rangle$ with a sequence $Obs = (obs(t_1), \ldots, obs(t_k))$ of observations where $obs(t_i) = (\pi_i, t_i)$ for $i = 1, \ldots, k$ and $t_1 < t_2 < \ldots < t_k \leq depth(P)$.

Let us first consider, given a plan P and such a sequence of observations Obs, the constraints a diagnosis $Q \subseteq S$ has to satisfy. Consider the first observation (π_1, t_1). From this observation we can make predictions $\pi'_{1,i}$ for all time points t_i, with $i > 1$, using the derivations

$$(\pi_1, t_1) \rightarrow^*_{Q;P} (\pi'_{1,i}, t_i).$$

Clearly, since Q is assumed to be a diagnosis of P using Obs, we should require $\pi'_{1,i} \approx \pi_i$ for all $1 < i \leq k$.

Now consider the second time point t_2. Note that the total state information available at time t_2 consists of the observed partial state π_2 and the predicted partial state $\pi'_{1,2}$ compatible with it. Hence, the total information available at t_2 is represented by the fused state $\pi_2 \sqcup \pi'_{1,2}$. Using this fused state and the qualification Q we can make predictions $\pi'_{2,i}$ for the partial states π_i observed at t_i for $i > 2$:

$$(\pi_2 \sqcup \pi'_{1,2}, t_2) \rightarrow^*_{Q;P} (\pi'_{2,i}, t_i)$$

Again, since Q is a diagnosis, all these predictions $\pi'_{2,i}$ should be compatible with π_i for all $2 < i \leq k$, i.e., it should hold that $\pi'_{2,i} \approx \pi_i$ for all $2 < i \leq k$.

Proceeding inductively, assume that predictions $\pi'_{h,i}$ have been made using all information available at times t_h where $h = 1, 2, \ldots i-1$. Then the predictions $\pi'_{i,j}$, where $j = i+1, \ldots k$, can be obtained as follows: The total information available at time t_i is $\pi_i \sqcup \pi'_{1,i} \sqcup \ldots \sqcup \pi'_{i-1,i}$. We can make predictions $\pi'_{i,j}$ for all times t_j where $j = i+1, \ldots, k$ using the derivations:

$$(\pi_i \sqcup \pi'_{1,i} \sqcup \ldots \sqcup \pi'_{i-1,i}, t_i) \rightarrow^*_{Q;P} (\pi'_{i,j}, t_j)$$

The representation of the total information available at time t_i can be simplified, since it turns out that $\pi_i \sqcup \pi'_{1,i} \sqcup \ldots \sqcup \pi'_{i-1,i} = \pi_i \sqcup \pi'_{i-1,i}$. This can be seen as follows:

For $1 \leq h < i - 1$ it holds that

$$(\pi_h \sqcup \pi'_{1,h} \sqcup \ldots \sqcup \pi'_{h-1,h}, t_h) \rightarrow^*_{Q;P} (\pi'_{h,i-1}, t_{i-1})$$

as well as

$$(\pi_h \sqcup \pi'_{1,h} \sqcup \ldots \sqcup \pi'_{h-1,h}, t_h) \rightarrow^*_{Q;P} (\pi'_{h,i}, t_i)$$

Since the derivability relation is deterministic, it therefore must hold that for $h = 1, \ldots, i-2$,

$$(\pi'_{h,i-1}, t_{i-1}) \rightarrow_{Q;P} (\pi'_{h,i}, t_i) \quad (1)$$

Since we have

$$\pi'_{h,i-1} \sqsubseteq \pi_{i-1} \sqcup \pi'_{1,i-1} \sqcup \ldots \sqcup \pi'_{h,i-1} \sqcup \ldots \sqcup \pi'_{i-2,i-1} \quad (2)$$

and

$$(\pi_{i-1} \sqcup \pi'_{1,i-1} \sqcup \ldots \sqcup \pi'_{i-2,i-1}, t_{i-1}) \rightarrow^*_{Q;P} (\pi'_{i-1,i}, t_i) \quad (3)$$

it follows from equations 1, 2, 3 and $\sqsubseteq$-preserving properties of the derivability relation that, for $h = 1, \ldots, i-2$,

$$\pi'_{h,i} \sqsubseteq \pi'_{i-1,i}.$$

Hence we obtain

$$\pi_i \sqcup \pi'_{1,i} \sqcup \ldots \sqcup \pi'_{i-1,i} = \pi_i \sqcup \pi'_{i-1,i}.$$

Therefore, at time t_i we only need to make a prediction $\pi'_{i,i+1}$ for time t_{i+1} using the derivation

$$(\pi_i \sqcup \pi'_{i-1,i}, t_i) \rightarrow^*_{Q;P} (\pi'_{i,i+1}, t_{i+1})$$

This line of reasoning underlies the following definition of a diagnosis using a sequence of observations:

Definition 7 *Let $P = \langle \mathcal{O}, S, < \rangle$ be a plan with a sequence $Obs = (\,obs(t_1) = (\pi_1, t_1), ..., obs(t_k) = (\pi_k, t_k)\,)$ of observations, where $t_1 < t_2 < ... < t_k \leq depth(P)$. Then the qualification $Q \subseteq S$ is said to be a plan diagnosis of P using Obs iff there exist partial states $\pi'_{i,i+1}$ for $1 \leq i < k$ such that*

1. $(\pi_1, t_1) \rightarrow^*_{Q;P} (\pi'_{1,2}, t_2)$,
2. $(\pi_i \sqcup \pi'_{i-1,i}, t_i) \rightarrow^*_{Q;P} (\pi'_{i,i+1}, t_{i+1})$ *for every* $2 \leq i < k$ *and*
3. $\pi'_{i,i+1} \approx \pi_{i+1}$ *for every* $1 \leq i < k$.

By slightly changing this definition, we can make a closer connection between the definition of a diagnosis based on a sequence of observations and the definition of a diagnosis based on a pair of observations. To this end, given the sequence of observations *Obs*, the qualification Q and Definition 7, we construct a new sequence of observations $(obs^*(t_1), \ldots, obs^*(t_k))$ as follows:

1. $obs^*(t_1) = obs(t_1)$;
2. for $i = 2, \ldots, k$, $obs^*(t_i) = (\pi_i \sqcup \pi'_i, t_i)$, where (π'_i, t_i) satisfies $obs^*(t_{i-1}) \rightarrow^*_{Q;P} (\pi'_i, t_i)$.

Now we can establish the following connection between diagnosis based on a sequence of observations and diagnosis based on a pair of observations:

Proposition 2 *Q is a diagnosis of P using $Obs = (obs(t_1), \ldots, obs(t_k))$ iff for $i = 1, \ldots k-1$, Q is a diagnosis of the pair of observations $(P, obs^*(t_i), obs(t_{i+1}))$.*

It is not difficult to adapt the idea of mini-maxi diagnoses to a sequence of observations. To construct such a diagnosis Q_{max}, it suffices to construct the separate qualifications $Q_{max,1}, \ldots, Q_{max,k}$ as follows:

Note that this algorithm makes use of Proposition 2 to compute the resulting mini-maxi diagnosis using an algorithm developed for diagnosis based on a pair of observations.

Algorithm 2 Computing a mini-maxi diagnosis based on a sequence of observations

Require: a plan P with a sequence Obs of observations $obs(t_1) = (\pi_1, t_1), \ldots, obs(t_k) = (\pi_k, t_k)$ where $t_1 < t_2 < \ldots < t_k \leq depth(P)$.
Ensure: a mini-maxi diagnosis Q_{max}.
1: Find a mini-maxi diagnosis $Q_{max,1}$ for $(P, (\pi_{t_1}, t_1), (\pi_{t_2}, t_2))$ using Algorithm 1 and compute the predicted state $\pi'_{max,1}$ using $Q_{max,1}$;
2: $i := 2$;
3: **while** $i < k$ **do**
4: Find a mini-maxi diagnosis $Q_{max,i}$ for $(P, (\pi_{t_i} \sqcup \pi'_{max,i}, t_i), (\pi_{t_{i+1}}, t_{i+1}))$ using Algorithm 1 and compute the corresponding predicted state $(\pi'_{max,i+1}, t_{i+1})$ using $Q_{max,i}$;
5: **end while**
6: return $Q_{max} := \bigcup_i Q_{max,i}$

6 Conclusion

We have presented a simple formal framework to specify an executable plan and we have defined the notion of a diagnosis using partial observations of a plan in execution. We based our analysis of plans and observations upon a model-based diagnosis approach and considered a plan as a description of a system that can be observed and can be used to make predictions about its (future) behavior.

Using this framework, we derived a definition for a plan diagnosis as a set of abnormally qualified plan steps that are able to derive a partial plan state *compatible* with an observed partial plan state. In contrast to model-based diagnosis, where minimal and minimum diagnoses are aimed for, we have shown that minimality in plan diagnosis not always leads to the results we prefer. The reason is that making observations of plans is not completely comparable to making observations of input-output behavior of systems in model-based diagnosis. Often we make observations during plan execution and would like to make predictions of future outcomes of plan execution based on a plan diagnosis established so far. That implies that *predictions* about future behavior are as important as explanations of already observed behavior. In order to make powerful predictions, we argued that we should therefore aim at *maximal informative* diagnoses.

We showed that in contrast to minimum diagnosis, a minimal maximum informative diagnosis can be found efficiently, although maximum informative diagnoses of minimum size are difficult to compute.

Finally, we extended our approach to diagnosis with iterative observations, showing that in such cases both the general definition of what constitutes a diagnosis as well as the computation of maximum informative diagnoses can be reduced to their counterparts discussed for the simple case where only two successive observations are involved.

Current work can be extended in several ways. We mention three possible extensions:

First of all, we could improve our current notion of diagnosis by taking into account the difference between plan operators and plan steps. In some cases it could be useful to make a distinction between establishing diagnoses at the plan step level and diagnoses at the plan operator level. For example, if instances of a driving action (i.e. plan steps) pertain to a plan operator that refers to the use of one single vehicle and all these instances are qualified as being abnormal, there is sufficient reason to believe that the vehicle itself (the plan operator) is faulty. Such a distinction requires the inclusion of *causal rules* linking different plan steps to each other. By means of such causal rules the number of plan steps qualified as abnormal often can be significantly reduced.

Secondly, going beyond plan operators, we could improve the diagnostic model to include a model of the executing agent(s) that is involved in executing one or more plan steps. In particular we need to consider cases where the agent might evolve through several abnormal states. We suspect the resulting model to be related to diagnosis in Discrete Event Systems [6, 13].

Thirdly, we hope to extend our current approach by including methods for *plan repair* in the context of the inferred agent's current (abnormal) state. Such methods especially seem to be useful in the context of iterative observations as discussed in the final part of this paper.

As a final remark, thanks to an anonymous reviewer, we were pointed out that there exist some analogies between our approach to plan diagnosis and approaches to software debugging where the underlying system is a piece of software that is modeled by a dependency-based model such as e.g. [16]. Certainly, it would be a fruitful idea to explore the possibilities for applying the ideas developed in this paper into the software debugging area.

ACKNOWLEDGEMENTS

This research is supported by the Technology Foundation STW, applied science division of the Dutch Science Foundation (NWO) and the Technology Programme of the Ministry of Economic Affairs. Project DIT5780: Distributed Model Based Diagnosis and Repair.

REFERENCES

[1] L. Birnbaum, G. Collins, M. Freed, and B. Krulwich, 'Model-based diagnosis of planning failures', in *AAAI 90*, pp. 318–323, (1990).
[2] N. Carver and V.R. Lesser, 'Domain monotonicity and the performance of local solutions strategies for cdps-based distributed sensor interpretation and distributed diagnosis.', *Autonomous Agents and Multi-Agent Systems*, **6**(1), 35–76, (2003).
[3] L. Console and P. Torasso, 'Hypothetical reasoning in causal models', *International Journal of Intelligence Systems*, **5**, 83–124, (1990).
[4] L. Console and P. Torasso, 'A spectrum of logical definitions of model-based diagnosis', *Computational Intelligence*, **7**, 133–141, (1991).
[5] F. de Jonge and N. Roos, 'Plan-execution health repair in a multi-agent system', in *PlanSIG 2004*, (2004).
[6] 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).
[7] R. E. Fikes and N. Nilsson, 'Strips: A new approach to the application of theorem proving to problem solving', *Artificial Intelligence*, **5**, 189–208, (1971).
[8] B. Horling, B. Benyo, and V. Lesser, 'Using Self-Diagnosis to Adapt Organizational Structures', in *Proceedings of the 5th International Conference on Autonomous Agents*, pp. 529–536. ACM Press, (2001).
[9] M. Kalech and G. A. Kaminka, 'On the design ov social diagnosis algorithms for multi-agent teams', in *IJCAI-03*, pp. 370–375, (2003).
[10] M. Kalech and G. A. Kaminka, 'Diagnosing a team of agents: Scaling-up', in *AAMAS 2004*, (2004).
[11] M. Kalech and G. A. Kaminka, 'Diagnosing a team of agents: Scaling-up', in *AAMAS-05*, (2005).
[12] G. A. Kaminka, 'Detecting disagreements in large-scale multi-agent teams', *JAAMAS*, (2006). To appear.
[13] Y. Pencolé, M. Cordier, and L. Rozé, 'Incremental decentralized diagnosis approach for the supervision of a telecommunication network', in *DX01*, (2001).
[14] R. Reiter, 'A theory of diagnosis from first principles', *Artificial Intelligence*, **32**, 57–95, (1987).
[15] N. Roos and C. Witteveen, 'Diagnosis of plans and agents', in *Multi-Agent Systems and Applications IV: CEEMAS 2005, LNCS 3690*, pp. 357–366, (2005).
[16] F. Wotawa, 'On the relationship between model-based debugging and program slicing', *Artif. Intell.*, **135**(1-2), 125–143, (2002).