Plan diagnosis with agents

Nico Roos [a] Cees Witteveen[b c]

[a] IKAT, Universiteit Maastricht, P.O. Box 616, NL-6200 MD Maastricht, e-mail: roos@cs.unimaas.nl
[b] Faculty EEMCS Delft University of Technology, P.O. Box 5031, NL-2600 GA Delft, e-mail: witt@ewi.tudelft.nl
[c] Center for Mathematics and Computer Science, P.O. Box 94079, NL-1090 GB Amsterdam, e-mail: C.Witteveen@cwi.nl

Abstract

We discuss the application of model based diagnosis to (multi-)agent based planning. First, we model a plan as a system to be diagnosed, taking into account the behavioural relations that might exist between several instances of actions occurring in a plan. Then we consider the case of a spatially distributed agent planning system where different agents are responsible for a part of the total plan.

1 Introduction

Model-Based Diagnosis (MBD) [1, 2, 4] is a well-known technique to infer abnormalities of (internal) components of a given system S from the input-output behaviour of S. To this end a model of S is given where the possible behaviours of each of the components and the relations between the components have been specified. Usually, for each component c at least two different *health states* are distinguished: a *normal* state and an *abnormal* state[1]. For each health state a specific behaviour of the component c is specified. The diagnostic engine is triggered whenever there is a discrepancy between the output —as predicted from the model and the input observations— and the actually observed output. The result of the diagnostic process is an assignment of health states to the components, called a *diagnosis* such that the actually observed output is *consistent* with this state qualification or can be *explained* by the state qualification. Usually, one requires the diagnosis to be specified in such a way that the number of components behaving abnormally is minimized.

Our contribution in this paper is an extension of MBD to both single agent and multi-agent *planning* systems. First, we introduce a notion of *plan diagnosis* in a single agent system, adapting MBD to deal with a *plan* as a system to be diagnosed. Here, the main idea is that by applying model based diagnosis, observations during plan execution can be used *(i)* to infer which already executed actions (the components of a plan) have to be qualified as failed and *(ii)* to predict which future actions will fail and whether the goals of the plan still will be achieved by executing the remaining part of the plan. An important difference with classical system diagnosis is that here we have to take into account that in

[1] In a more general set-up, we often distinguish a partially ordered set of health states.

a plan several related instances of actions occur and the health state of a subset of these instances might be used to infer the health state of other instances related to them.

Secondly, we concentrate on multi-agent plan diagnosis. Here, the agents together are assumed to execute a joint plan. This plan is partitioned over the agents. Each agent is responsible for the execution of its sub-plan and has to respect the dependencies with sub-plans of other agents. Concepts from classical MBD are extended to cover plan diagnosis for multi-agent systems.

2 Plan Based Diagnosis

States and Goals Unlike classical MBD, in plan-based diagnosis the model is not a description of an underlying system but a *plan* of an agent. For our purposes, we prefer to take an object or *resource-based* view on the world[2], assuming that, for the planning problem at hand, the world can be described by a set $Obj = \{o_1, o_2, \ldots, o_n\}$ of objects, their respective *domains* S_i and and their (current) values $v_i \in S_i$. A state of the world σ then simply is an element of the set $S_1 \times S_2 \times \ldots \times S_n$. The value $v_j \in S_j$ of the object o_j in the state σ will be denoted by $\sigma(j)$. It will not always be possible to give a complete state description . Therefore, we introduce a *partial state* $\pi \in S_{i_1} \times S_{i_2} \times \ldots \times S_{i_k}$, where $1 \leq i_1 < \ldots < i_k \leq n$. We will use $O(\pi)$ to denote the set of objects $\{o_{i_1}, o_{i_2}, \ldots, o_{i_k}\} \subseteq Obj$ specified in π. The value of $o_j \in O(\pi)$ will be denoted by $\pi(j)$. Partial states can be ordered with respect to their information content: π is said to be contained in π', denoted by $\pi \sqsubseteq \pi'$, iff $O(\pi) \subseteq O(\pi')$ and $\pi(j) = \pi'(j)$ for every $o_j \in O(\pi)$. The value of an object $o_j \in O$ not occurring in a partial state π is said to be undefined (w.r.t. π).

A goal G of an agent is specified as a set of partial states $G = \{g_1, g_2, \ldots, g_m\}$ over the domains $S_1, S_2, \ldots, S_n$, at least one of which the agent wants to bring about, i.e., a goal G is said to be *satisfied by a partial state* π if there exists a $g_i \in G$ such that $g_i \sqsubseteq \pi$.

Actions An *action* can be viewed as a function that replaces the values of a subset of the objects in Obj by other values, dependent upon the values of a (possibly different) subset of objects. Hence, every action a can be modeled as a (partial) function $f_a : S_{i_1} \times \ldots \times S_{i_k} \rightarrow S_{j_1} \times \ldots \times S_{j_l}$, where $1 \leq i_1 < \ldots < i_k \leq n$ and $\{j_1, \ldots, j_l\} \subseteq \{i_1, \ldots, i_k\}$. That is, all objects whose value domains occur in $ran(f_a)$, denoted by $ran_O(a) = \{o_{i_1}, \ldots, o_{i_k}\}$, are contained in the analogously defined set of objects $dom_O(a) = \{o_{j_1}, \ldots, o_{j_l}\}$. The function specification f_a constitutes the *normal* behaviour of the action, denoted by f_a^{nor}. The *abnormal* behaviour of a *broken action* is specified by the function $f_a^{ab} : S_{i_1} \times \ldots \times S_{i_k} \rightarrow S_{j_1} \times \ldots \times S_{j_l}$, where $f_a^{ab}(s_{j_1}, s_{j_2}, \ldots, s_{j_l}) = (\top, \top, \ldots, \top)$. Here, $\top$ denotes an arbitrary value of the corresponding object o_i in the domain S_i. We assume that this choice of values as the result of a broken action always is observable, i.e., for every $(s_{i_1}, s_{i_2}, \ldots, s_{i_k}) \in dom(f_a)$, $f_a^{ab}(s_{i_1}, s_{i_2}, \ldots, s_{i_k}) \neq f_a^{nor}(s_{i_1}, s_{i_2}, \ldots, s_{i_k})$: there is always at least one value distinguishing f_a^{ab} from f_a^{nor}.

Behavioural rules A plan P is conceived as a partially ordered set A of *instances* of actions. The main difference with components in a classical system is that often in

[2] In contrast to the conventional approach to state-based planning, cf. [3].

a plan the behaviour of different instances of actions are closely related. For example, suppose that we have a plan for carrying luggage from a depot to a number of waiting planes. Such a plan will contain several instances of a drive action pertaining to the same carrier. Suppose that we detect that a drive action behaves abnormally because of malfunctioning of the carrier. Then it is reasonable to assume that instances of the same drive action that occur in the plan *after* this instance can be predicted to behave abnormally, too. To capture such behavioural relations between related instances, we specify a set of *rules*. Each such a rule is of the form $\{a_{i_1}, \ldots, a_{i_k}\} \rightarrow \{a_{j_1}, \ldots, a_{j_l}\}$, expressing that whenever a set $\{a_{i_1}, \ldots, a_{i_k}\}$ of instances of actions occurs as a subset of the set of instances actions qualified as behaving abnormally, it is inferred that from that time on all instances $a_{j_1}, \ldots, a_{j_l}$ will be qualified as behaving abnormally, too. A set Φ_A of such rules[3] is said to specify a *behavioural theory* for the set of instances A.

Given a subset $Ab \subseteq A$ of abnormal instances of actions, the set of immediate behavioural consequences of Ab using Φ is defined as the set $BC_\Phi(Ab) = \bigcup\{\beta \mid \alpha \rightarrow \beta \in \Phi \mid \alpha \subseteq Ab\}$. We say that Φ is *behaviourally closed* if the following condition holds: Whenever Φ contains two rules $\alpha \rightarrow \beta$ and $\gamma \rightarrow \delta$ such that $\beta \cap \gamma \neq \varnothing$, then Φ also contains a rule $\alpha' \rightarrow \beta'$ such that $\alpha' \subseteq \alpha \cup (\gamma - \beta)$ and $\beta \cup \delta \subseteq \beta'$.

A simple example of a behaviourally closed set of rules is the following: Let $\simeq$ be an equivalence relation on A, where $a \simeq a'$ iff a and a' are instances of the same action, e.g., the driving action we mentioned above. Let A' be a union of some equivalence classes in $A/_\simeq$ specifying those types of actions where abnormal behaviour will be preserved. That is, if $a \in A'$ is detected as behaving abnormally, then every future similar instance $a' \simeq a$ will also behave abnormally. Then we can define a behaviourally closed set $\Phi = \{\{a\} \rightarrow \{a' : a' \simeq a\} \mid a \in A'\}$ of rules indicating that from now on every instance a' similar to an instance a characterized as abnormal will be evaluated as abnormal, too.

If a rule set is behaviourally closed, two or more successive applications of rules always can be obtained by applying just one single rule in Φ, i.e., the behavioural consequence operator BC_Φ needs to be applied just once to obtain all the behavioural consequences of a subset $Ab \subseteq A$ and the set Ab itself:

Proposition 1 *Let Φ be a behaviourally closed set of rules for a set of instances A and $Ab \subseteq A$. Define the inflationary operator*[4] *$BC^1_\Phi(Ab) = Ab \cup BC_\Phi(Ab)$. Then $BC^1_\Phi(Ab) = BC^1_\Phi(BC^1_\Phi(Ab))$.*

Plans A *plan* $P = \langle A, <, \Phi\rangle$ is a triple consisting of the set of actions A, a partial ordering $< \subseteq A \times A$ on the actions, and a set of behaviour rules Φ. The partial order $<$ specifies the dependencies between actions; e.g., $a < a'$ means that the action a must finish before a' may start. We will often denote the *transitive reduction* of $<$ by $\ll$, i.e., the transitive closure $\ll^+$ of $\ll$ equals $<$. Note that an action a may change the values of the objects in the domain $dom_O(a')$ of another action a'; i.e., $ran_O(a) \cap dom_O(a') \neq \varnothing$. Obviously, if a plan P allows an action a to be executed concurrently with an other action a', we would not allow that their domains and ranges do overlap. Therefore, we have the following requirement:

[3] Often we will omit the subscript A in Φ_A if the context of the set of actions is clear.

[4] A consequence operator T is called inflationary if T applied to A returns the set of all immediate consequences and A itself.

Concurrency Requirement: Let $P = \langle A, <, \Phi \rangle$ be a plan and $a, a' \in A$ two actions such that neither $a < a'$ nor $a' < a$ holds. Then both $ran_O(a) \cap dom_O(a') = \varnothing$ and $ran_O(a') \cap dom_O(a) = \varnothing$ should hold.

Plan execution For simplicity, when a plan is executed, we will assume that every action takes a unit of time to execute and we are allowed to observe the execution of a plan P at discrete times $t = 0, 1, 2, \ldots, k$ where k is the depth of the plan, i.e., the longest $<$-chain of actions occurring in P. Let $depth_P(a)$ be the depth[5] of action a in plan $P = (A, <, \Phi)$, i.e., $depth_P(a) = 0$ if $\{a' \mid a' \ll a\} = \varnothing$ and $depth_P(a) = 1 + max\{depth_P(a') \mid a' \ll a\}$, else. We assume that the plan starts to be executed at time $t = 0$ and that concurrency is completely specified by the plan, i.e., if $depth_P(a) = k$ then execution of a has been completed at time $t = k + 1$. Now all actions a with $depth(a) = 0$ are completed at time $t = 1$ and every action a with $depth(a) = k + 1$ will be started at time k and will be completed at time $k + 1$. The above specified *concurrency requirement* ensures that the concurrent execution of actions leads to a well-defined result.

The effect of the execution of plan P on a given (partial) state π at some time $t \geq 0$, denoted by (π, t), can be defined as follows: Let P_t denote the set of actions a with $depth(a) = t$ and let $P_{>t} = \bigcup_{t' > t} P_t$. Moreover, let the partial state π restricted to a given set O, denoted as $\pi \restriction O$, be defined as $\pi \restriction O = \pi'$ where $\pi' \sqsubseteq \pi$ and $O(\pi') = O \cap O(\pi)$. Now we say that $(\pi', t + 1)$ is (directly) generated by *normal* execution of P from (π, t), abbreviated by $(\pi, t) \rightarrow_P (\pi', t + 1)$, iff the following conditions hold:

1. $\pi' \restriction ran_O(a) = f_a^{nor}(\pi \restriction dom_O(a))$ for each $a \in P_t$ with $dom_O(a) \subseteq O(\pi)$, that is, the consequences of all actions a enabled in π should be visible in π';
2. $O(\pi') \cap ran_O(a) = \varnothing$ for each $a \in P_t$ with $dom_O(a) \not\subseteq O(\pi)$, that is, the values of objects modified by an action a are left undefined in π' if a is not enabled in π;
3. $\pi'(i) = \pi(i)$ for each $o_i \notin \bigcup_{a \in P_t} ran_O(a)$, that is, every object not in the range of an action at time t should remain unchanged.

For arbitrary values of $t \leq t'$ we say that (π', t') is (directly or indirectly) generated by *normal* execution of P from (π, t), denoted by $(\pi, t) \rightarrow_P^* (\pi', t')$, iff the following conditions hold: ***(i)*** if $t = t'$ then $\pi' = \pi$; ***(ii)*** if $t' = t + 1$ then $(\pi, t) \rightarrow_P (\pi', t')$; ***(iii)*** if $t' > t + 1$ then there must exists some state $(\pi'', t' - 1)$ such that $(\pi, t) \rightarrow_P^* (\pi'', t' - 1)$ and $(\pi'', t' - 1) \rightarrow_P (\pi', t')$.

Qualifications In order to predict the result of a plan execution, we introduce the notion of a *qualified* plan. A qualified version P_Q of a plan $P = (A, <, \Phi)$ is a tuple $P_Q = (A, <, \Phi, Q)$, where $Q \subseteq A$ is the set of *abnormally* behaving actions.[6] For such a subset Q we define the execution relation $\rightarrow_{Q;P}$ as follows: $(\pi', t+1)$ is (directly) generated by execution of P where actions in Q are behaving abnormally from (π, t), abbreviated by $(\pi, t) \rightarrow_{Q;P} (\pi', t + 1)$, if for all $i = 1, \ldots, n$ we have

1. $\pi' \restriction ran_O(a) = f_a^{nor}(\pi \restriction dom_O(a))$ for each $a \in P_t - Q$ with $dom_O(a) \subseteq O(\pi)$,
2. $O(\pi') \cap ran_O(a) = \varnothing$ for each $a \in Q$,

[5] If the context is clear, we often will omit the subscript P in referring to the depth of an action a.

[6] Hence, $A - Q$ is the set of normally executed actions.

3. $O(\pi') \cap ran_O(a) = \varnothing$ for each $a \in P_t$ with $dom_O(a) \not\subseteq O(\pi)$, and

4. $\pi'(i) = \pi(i)$ for each $o_i \notin \bigcup_{a \in P_t} ran_O(a)$.

Note that this leaves free the values of $\pi'(i)$ for those objects affected by abnormal actions $a \in P_t \cap Q$. Furthermore, by definition, $\rightarrow_{\varnothing;P} = \rightarrow_P$. The reflexive transitive closure $\rightarrow^*_{Q;P}$ is defined analogously.

Diagnosis Suppose that we have a (possibly partial) observation $obs(t) = (\pi, t)$ of the state of the world π at time point t and an observation $obs(t') = (\pi', t')$ at time point $t' > t$ during execution of the plan P. Then, assuming a normal execution of the plan P we can predict the state of the world at a time point t' given the observation $obs(t)$: if all actions behave normally, we should have $obs(t) \rightarrow^*_{\varnothing;P} (\pi'', t')$ where π'' satisfies $(\pi' \restriction O(\pi'')) \sqsubseteq \pi''$. If not, the execution of some actions must have gone wrong. In that case, we would like to determine which action may have failed. To this end, we may apply the standard definition of MBD (cf. [2]):

Definition 1 *Let $P = (A, <, \Phi)$ be a plan, let $Beh = \bigcup_{a \in A}(f_a^{nor} \cup f_a^{ab})$ be the behavioural description of the actions A, let $obs(t) = (\pi, t)$ and $obs(t') = (\pi', t')$ with $t < t'$ be two (partial) observations, and let $obs(t) \rightarrow^*_{\varnothing;P} (\pi'', t')$ be the normal execution of a plan. Moreover, let Q be a qualification and let $obs(t) \rightarrow^*_{Q;P} (\pi''', t')$ be the plan execution given this qualification. The qualification Q is an O-plan-diagnosis of $\langle P, Beh, obs(t), obs(t') \rangle$ iff*

1. $\pi' \restriction O = \pi'' \restriction O$,

2. *π''' is consistent with π'; i.e., for every $o_j \in O(\pi') \cap O(\pi''')$: $\pi'(j) = \pi'''(j)$.*

Choosing $O = \{o_j \mid o_j \in O(\pi') \cap O(\pi''), \pi'(j) = \pi''(j)\}$ in 1. gives the strongest diagnostic definition.[7] Choosing $O = \varnothing$ results in consistency-based diagnosis [4].

In MBD normally either numerical *minimum* or subset *minimal* diagnoses are considered. Here, we will even further refine the set of diagnoses by taking the behavioural rules Φ into consideration.

Behavioural Rules and Diagnosis Using the rule set Φ in establishing a diagnosis, we could reason as follows: If a qualification Q has been established as a diagnosis, some of the actions executed at some time t' and detected as abnormal could easily occur as the behavioural consequences of other abnormally behaving and earlier executed actions. Hence, instead of a diagnosis Q, what has to be established is a minimum set Q' of abnormally behaving actions such that

1. the diagnosis Q can be generated using the rules in Φ from Q',

2. no action in Q' is the behavioural consequence of another action in Q' and

3. Q' is a minimum set of actions satisfying 1. and 2.

In a certain sense, the set Q' has to be considered as a set of *causes* of the abnormal behaviour under the set of rules Φ.

[7] Note, however, that if one broken action may repair the effects of another broken action, this form of diagnosis may overlook a correct plan-diagnosis.

Definition 2 *Let Q be a standard diagnosis of $\langle P, Beh, obs(t), obs(t')\rangle$. We say that $Q' \subseteq Q$ is a* minimum causal explanation *of Q using the rule set Φ if the following conditions hold:*

1. *$Q = Q' \cup BC_\Phi(Q')$, i.e., Q' and its set of immediate Φ-consequences generate the diagnosis Q;*
2. *$Q' = Q'_t \cup Q'_{t+1} \cup \ldots \cup Q'_{t'}$ where $Q'_{t+i} = P_{t+i} \cap Q'$, that is Q' can be split in (disjoint) sets of abnormalities established at time points $t, t+1, \ldots t'$;*
3. *$Q'_{t+i} \cap BC_\Phi(Q'_t \cup \ldots \cup Q'_{t+i-1}) = \varnothing$ for $i > 0$, i.e., no set of causes occurring at time $t+i$ can be explained by abnormalities occurring at some earlier time under the rule set Φ.*
4. *Q' is a smallest set of abnormalities such that the conditions above do hold.*

Prediction of plan results Except for playing a role in establishing causal explanations, the set of behavioural consequences also plays a major role in the prediction of the results of the plan. The main purpose of plan diagnosis of course is to *predict* whether the goal G still can be reached given the observations $obs(t)$ during plan execution. This prediction can be based on the diagnosis Q and the behavioural consequences of it established at time t as follows: Let Q be the set of instances of actions qualified as abnormal as the result of a diagnosis at time t. Let $Q_{pred} \subseteq A$ be defined as the set $Q_{pred} = BC_\Phi(Q) \cap P_{>t}$, i.e., the set of to be executed actions that using Φ and Q can be inferred to behave abnormally, too. Then the predicted output at time $t' > t$ equals the state τ such that $obs(t) \rightarrow^{(t'-t)}_{Q_{pred};P} (\tau, t')$. Suppose that $t' = depth(P)$, i.e., the plan has been executed completely, then we can check whether or not for some state $\sigma \in G$: $\sigma \sqsubseteq \tau$. If such a state does not exist, we infer that given the abnormalities inferred, the goal G of the plan will not be achieved.

3 Multi-agent diagnosis of multi-agent plans

A group of collaborating agents often have to coordinate their actions. If the problem addressed by the agents is complex, planning is required, resulting in a distributed multi-agent plan.

Multi-agent plans A multi-agent plan is a partition of a plan $P = (A, <, \Phi)$ over a group of agents Ag such that agent i is responsible for the execution of the sub-plan $P_i = (A_i, I_i, O_i <_i, \Phi_i)$ where $A = \biguplus_i A_i$, $I_i = \{a \in (A - A_i) \mid a \ll a', a' \in A_i\}$, $O_i = \{a \in (A - A_i) \mid a' \ll a, a' \in A_i\}$, $<_i = (< \cap ((A_i \cup I_i) \times A_i)) \cup (< \cap (A_i \times (A_i \cup O_i)))$ and $\Phi = \bigcup_i \Phi_i$. Note that the set of actions I_i provide input for the plan of agent i and the set of actions O_i of other agents receive output from the plan of agent i. Here, agent i has to synchronize its action with other agents.

Multi-agent plan execution To enable agents to collectively determine the effect of executing the plan P using only their local knowledge, requires that agents communicate information about the state of the world. Since we use an object oriented view of the world (*i*) agents only have to communicate in situations where they must synchronize their actions with actions, I_i and O_i, of other agents, and (*ii*) they only have to communicate the values of the objects required by the actions $a \in O_i$ of another agent.

Proposition 2 *If the value of an object o required by an action $a \in A_i$ is determined by an action $a' \in A_j$ with $i \neq j$, then $a' \ll a$, $a' \in I_i$ and $a \in O_j$.*

To derive the partial state π' at time t' of a sub-plan $P_i = (A_i, I_i, O_i <_i, \Phi_i)$ given an partial state π at time t, we must take into account the effect of the actions $a \in I_i$ such that for some $a' \in A_i$ with $depth_P(a') < t'$, $a \ll a'$. Let $(\pi^{a_1}, t_1), ..., (\pi^{a_k}, t_k)$ with $\{a_j \in I_i \mid depth_P(a_i) < t'\}$, $t_j = depth_P(a_j)$ and $O(\pi^{a_j}) \subseteq ran_O(a_j) \cap \bigcup_{a' \in A_i | a_j \ll a'} dom_O(a')$ describe the known values of the objects in $ran_O(a)$ that are relevant for agent i. Then we say that (π', t') is generated from (π, t) by the execution of the plan of agent i given an qualification $Q_i \subseteq A_i$, for $i = 1, 2, ..., n$, abbreviated by $(\pi, t), (\pi^{a_1}, t_1), ..., (\pi^{a_k}, t_k) \rightarrow^*_{Q_i;P_i} (\pi', t')$. Note that the value of an object o in $O(\pi)$ may (indirectly) be determined by one or more actions $a \in I_i$. We assume that there are no inconsistencies between the value of o in π and the value of o that follows form the action in I_i if this information is available in $(\pi^{a_1}, t_1), ..., (\pi^{a_k}, t_k)$.

The multi agent plan presented here is closely related to the spatially distributed system of connected components discussed in [5]. In [5] it was pointed out that predicting the behaviour of such a system is in general an NP-Hard problem. The underlying reason was the possible existence of cyclic dependencies between components. In a (multi-agent) plan similar cycles between actions do not occur. A cycle would imply that the input of an action depends on the output of a future action. In fact predicting the execution of a (multi-agent) plan can be done in linear time.

The diagnosis of a sub-plan An agent executing a sub-plan may make partial observation of the local state of the world at different times t and t' with $t < t'$. Agent i can predict the expected state of the world at time t' using the knowledge of its local plan and information received from other agents about the expected effects of action in I_i. If agent i notices a difference between the expected and the observed state of the world at t, diagnosis is required. Since an observed discrepancy may be caused by the execution of another sub-plan, agent i may also receive information from other agents about the expected values of objects in $dom_O(a)$ with $a \in O_i$.

Definition 3 *Let $P_i = (A_i, I_i, O_i <_i, \Phi_i)$ be the sub-plan of agent i and let $obs(t)$ and $obs(t')$ with $t < t'$ be two observations. Moreover let $(\pi^{a_1}, t_1), ..., (\pi^{a_k}, t_k)$ with $t_j < t'$ be the communicated information about the expected effects of $a_j \in I_i$ and let $(\pi^{a'_1}, t'_1), ..., (\pi^{a'_k}, t'_l)$ with $t < t'_j$ be the communicated information about the expected effects of plan P_i for the actions $a'_j \in O_i$.*

Then a local diagnosis is a plan-diagnosis according to Definition 1 of

$$\langle P, Beh, \{obs(t), (\pi^{a_1}, t_1), ..., (\pi^{a_k}, t_k)\}, \{obs(t'), (\pi^{a'_1}, t'_1), ..., (\pi^{a'_k}, t'_l)\}\rangle.$$

Multi-agent plan-diagnosis An important question is whether the combined diagnoses of the individual agents lead to the same set of global diagnoses of the whole plan P. The following two propositions show that this is the case.

Proposition 3 *Let the qualification Q be a plan-diagnosis of $\langle P, Beh, obs(t), obs(t')\rangle$. Moreover, let $(\pi^{a_1}, t_1), ..., (\pi^{a_k}, t_k)$ be the (predicted) values of the objects $O(\pi^{a_j}) \subseteq ran_O(a_j) \cap \bigcup_{a' \in A_i | a_j \ll a'} dom_O(a')$ with $a_j \in I_i$ and $t \leq depth_P(a_j) \leq t'$, and let*

$(\pi^{a'_1}, t'_1), ..., (\pi^{a'_k}, t'_l)$ *be the (predicted) values of the objects* $O(\pi^{a'_j}) \subseteq dom_O(a'_j) \cap \bigcup_{a' \in A_i | a' \ll a'_j} ran_O(a')$ *with* $a'_j \in O_i$ *and* $t \leq depth_P(a'_j) \leq t'$.

Then the qualification $Q_i = Q \cap A_i$ *is a plan-diagnosis of*

$$\langle P, Beh, \{obs(t), (\pi^{a_1}, t_1), ..., (\pi^{a_k}, t_k)\}, \{obs(t'), (\pi^{a'_1}, t'_1), ..., (\pi^{a'_k}, t'_l)\}\rangle.$$

Proposition 4 *Let the qualification* Q_i *be a plan-diagnosis of*

$$\langle P, Beh, \{obs(t), (\pi^{a_1}, t_1), ..., (\pi^{a_k}, t_k)\}, \{obs(t'), (\pi^{a'_1}, t'_1), ..., (\pi^{a'_k}, t'_l)\}\rangle$$

with $a_j \in I_i$, $t \leq depth_P(a_j) \leq t'$, $a'_j \in O_i$ *and* $t \leq depth_P(a'_j) \leq t'$.

Then the qualification $Q = \bigcup_i Q_i$ *is a plan-diagnosis of* $\langle P, Beh, obs(t), obs(t')\rangle$ *iff*

- $(\pi^{a_1}, t_1), ..., (\pi^{a_k}, t_k)$ *are the (predicted) values of the objects* $O(\pi^{a_j}) \subseteq ran_O(a_j) \cap \bigcup_{a' \in A_i | a_j \ll a'} dom_O(a')$ *with* $a_j \in I_i$ *and* $t \leq depth_P(a_j) \leq t'$*; and*
- $(\pi^{a'_1}, t'_1), ..., (\pi^{a'_k}, t'_l)$ *are the (predicted) values of the objects* $O(\pi^{a'_j}) \subseteq dom_O(a'_j) \cap \bigcup_{a' \in A_i | a' \ll a'_j} ran_O(a')$ *with* $a'_j \in O_i$ *and* $t \leq depth_P(a'_j) \leq t'$.

To determine a global plan diagnosis without any agent having complete knowledge of the multi-agent plan or even the global diagnosis itself, the protocol presented in [5] can be applied. This protocol enables agents to efficiently determine a diagnosis. The communication overhead of the protocol is linear in the product the number of agents and of the number of action determining the observed object at time t' (cf. [5]).

4 Conclusion

We have presented a new object-oriented model for describing multi-agent plans consisting of (related) instances of actions. This model enables agents to find causes for discrepancies between the predicted and observed effects of a plan-execution by applying techniques developed for multi-agent model-based diagnosis. Moreover, we have extended the diagnostic theory enabling the prediction of future failure of actions.

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. E. Fikes and N. Nilsson. STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving, *Artificial Intelligence*, **5**, 2:189–208, 1971.

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

[5] N. Roos, A. ten Teije, and C. Witteveen. A protocol for multi-agent diagnosis with spatially distributed knowledge. In *AAMAS 2003*, pages 655–661, 2003.