Model Generation for PRS-like Agents

Wayne Wobcke,[1] Marc Chee[1] and Krystian Ji[1,2]

[1]School of Computer Science and Engineering
University of New South Wales
Sydney NSW 2052, Australia
{wobcke|marcchee|krystianj}@cse.unsw.edu.au

[2]National ICT Australia
University of New South Wales
Sydney NSW 2052, Australia
krystian.ji@nicta.com.au

ABSTRACT

We develop a sound foundation for model checking algorithms for the class of PRS-style BDI agents, by showing how a reachability graph for any given PRS-type agent can be constructed from the agent program, thus addressing a long-standing issue in the verification of BDI agents.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents*

General Terms

Theory, Verification

Keywords

Model checking, BDI agent architectures

1. INTRODUCTION

Although Rao and Georgeff [6] presented a model checking algorithm for their BDI logic, they did not address the central question of generating the underlying models from agent programs, an essential prerequisite to applying model checking techniques to agents of this class.

The main problem is to capture, in semantic structures, notions of belief and intention and their dynamics, and to systematically relate these notions to the operational behaviour of the agent interpreter, so agent program designers can reason logically about the mental states of the agents with confidence that this reasoning faithfully represents the underlying behaviour of the agent.

Recent approaches to applying model checking to classes of systems that include some characteristics of agency either encode beliefs and intentions as part of the states in model structures (e.g. Bordini *et al.* [1]) and so fail to provide a semantic account of these notions, or work only with simplified notions (e.g. the analogue of knowledge used in Raimondi and Lomuscio [5]).

AAMAS'05, July 25-29, 2005, Utrecht, Netherlands.

In this paper, we present an approach to representing the execution structures, including the agent's beliefs and intentions, for any PRS-type agent program, Georgeff and Lansky [3], in a reachability graph, as in SPIN, Holzmann [4]. The construction is based on our logical approach to Agent Dynamic Logic, an extension of PDL that includes modal operators for belief, goal and intention, Wobcke [7, 8]. The notion of intention is a more "rational" version of the operational notion used in PRS-type systems, though does not capture all the complexities of Bratman's theory [2].

In the reachability graph, beliefs are encoded using a relation b on states, and intentions are represented as those actions the agent will eventually perform successfully in any of a set of related alternative worlds (defined using a relation a on states). For any reachability graph G, let $\mathcal{I}(G)$ be the subgraph of G that contains only those transitions of G corresponding to successful action executions. The satisfaction conditions of modal formulae are as follows.

$$\sigma \models_G \mathsf{B}\alpha \quad \text{if } \sigma' \models_G \alpha \text{ whenever } b(\sigma, \sigma')$$
$$\sigma \models_G \mathsf{I}\pi \quad \text{if } \sigma' \models_{\mathcal{I}(G)} \mathsf{A}\Diamond do(\pi) \text{ whenever } a(\sigma, \sigma')$$
$$\sigma \models_G \mathsf{G}\gamma \quad \text{if } \sigma \models_G \mathsf{I}(achieve\ \gamma)$$

Here *achieve* γ is an "action" denoting achievement of a goal γ, and a state satisfies $do(\pi)$ if the subgraph emanating from that state is isomorphic to an execution tree in $\mathcal{R}_\pi$, the interpretation of π.

2. AN EXAMPLE

The reachability graph construction procedure can be illustrated through a simple "waypoint following" agent (the full version of the paper includes a general algorithm for the construction of a reachability graph for any PRS-like agent program). The agent has a simple task: it must visit four waypoints, numbered 1–4, whilst not running out of fuel. There is fuel at locations 1 and 3 though this knowledge is not explicit in the agent's beliefs. Rather, the agent is constructed to have refuelling plans that are triggered by *warn* events, using which the agent visits a fuel depot (at location 1 or location 3) and attempts a *refuel* action.

The agent always has knowledge of its position, represented as beliefs at_i and $\neg at_j$ $(j \neq i)$, and each $visit_i$ action includes a correct observation of the agent's position. The agent initially has no belief about the fuel level, but after a *refuel* action, correctly observes the state of the fuel tank, represented as a belief *full* or *empty*. The $visit_i$ and *refuel* actions always succeed, except that on occasion (here only at location 3) a *warn* event occurs. The initial configuration is that the agent starts at location 1 (and believes this) and

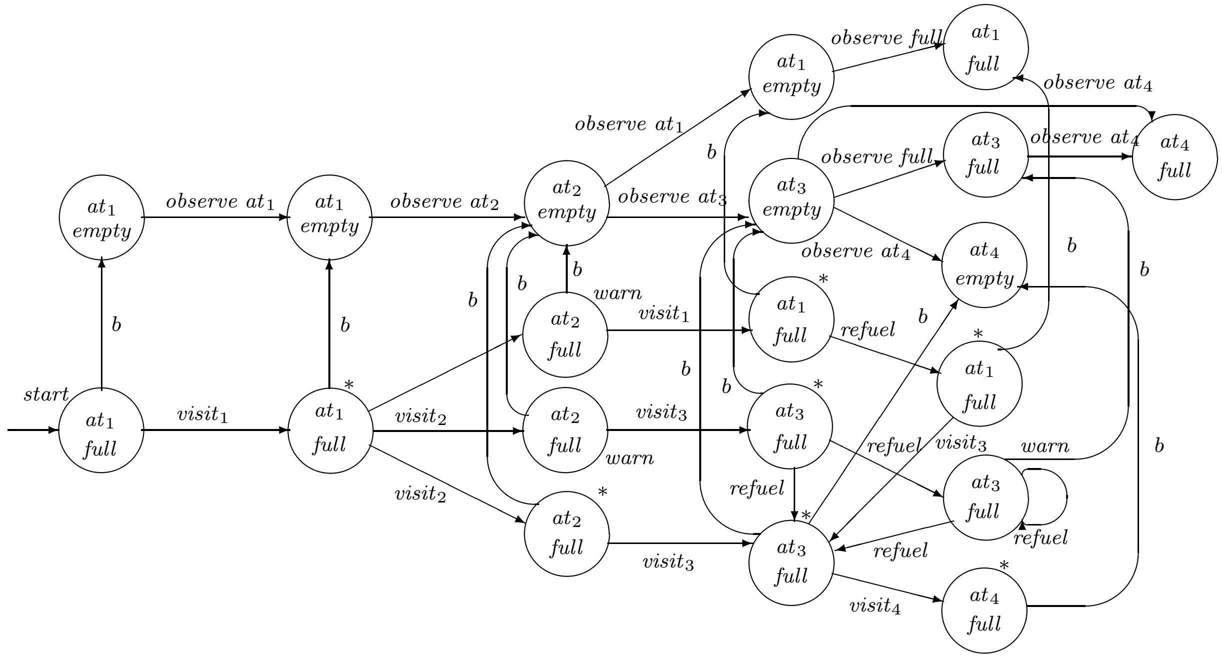

Figure 1: Waypoint Agent Reachability Graph

has a full fuel tank (though does not believe this). In the reachability graph shown in Figure 1, a and b-related links between a state and itself are omitted, and states resulting from successful executions are marked with $*$.

The following lists some examples of the kind of properties of the agent that could be tested using such a graph:

$\mathsf{A}\Box(\mathit{full} \Leftrightarrow \mathsf{B}\mathit{full})$

$\mathsf{A}\Diamond\, at_4$

$\mathsf{A}(\mathsf{I}\mathit{visit}_i\,\mathcal{U}\,at_i)$

if $\mathit{warn} \in \sigma$ then $\sigma \models \mathsf{A}(\mathsf{I}\mathit{refuel}\,\mathcal{U}\,\mathsf{B}\mathit{full})$

The first two formulae are not satisfied at the initial state of the graph, the second since the agent may become stuck handling an infinite sequence of fuel warnings. The third and fourth properties represent the persistence of intentions.

3. CONCLUSION

This paper has shown how the execution models of a family of BDI agents based on the PRS architecture, including notions of belief and intention and their dynamics, can be represented in reachability graphs, structures that are suitable for applying model checking algorithms to this class of agents. By clarifying the mental notions underlying this class of agents and by systematically relating these notions to the operational behaviour of the agent interpreter, a more robust methodology underpinning the verification of complex BDI agents is made possible.

4. ACKNOWLEDGEMENTS

This work is funded by an Australian Research Council Discovery Project Grant. Krystian Ji is supported by a UNSW International Postgraduate Research Scholarship and a scholarship from NICTA, National ICT Australia Ltd. National ICT Australia is funded through the Australian Government's *Backing Australia's Ability* initiative, in part through the Australian Research Council.

5. REFERENCES

[1] R. H. Bordini, M. Fisher, C. Pardavila and M. Wooldridge. Model Checking AgentSpeak. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 409–416, 2003.

[2] M. E. Bratman. *Intention, Plans, and Practical Reason.* Harvard University Press, Cambridge, MA, 1987.

[3] M. P. Georgeff and A. L. Lansky. Reactive Reasoning and Planning. In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, pages 677–682, 1987.

[4] G. J. Holzmann. The Model Checker SPIN. *IEEE Transactions on Software Engineering*, 23(5):279–295, 1997.

[5] F. Raimondi and A. Lomuscio. Verification of Multiagent Systems via Ordered Binary Decision Diagrams: An Algorithm and Its Implementation. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 630–637, 2004.

[6] A. S. Rao and M. P. Georgeff. A Model-Theoretic Approach to the Verification of Situated Reasoning Systems. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 318–324, 1993.

[7] W. R. Wobcke. Modelling PRS-like Agents' Mental States. In M. Ishizuka and A. Sattar, editors, *PRICAI 2002: Trends in Artificial Intelligence.* Springer-Verlag, Berlin, 2002.

[8] W. R. Wobcke. Model Theory for PRS-Like Agents: Modelling Belief Update and Action Attempts. In C. Zhang, H. W. Guesgen and W. K. Yeap, editors, *PRICAI 2004: Trends in Artificial Intelligence.* Springer-Verlag, Berlin, 2004.