Verification within the KARO Agent Theory: The reduction approach

Ullrich Hustadt[1] and Renate A. Schmidt[2]

[1] Department of Computer Science, University of Liverpool, UK, U.Hustadt@csc.liv.ac.uk
[2] Department of Computer Science, University of Manchester, UK, schmidt@cs.man.ac.uk

The spread of computer technology has meant that more dynamic, complex and distributed computational systems are being developed. Understanding and managing such complexity is notoriously difficult and it is to support this that the agent-based systems paradigm was introduced. In order to reason about agent-based systems, a number of theories of rational agency have been developed, for example the BDI (Rao and Georgeff 1991) and KARO (van Linder, van der Hoek, and Meyer 1996, 1998) frameworks

The KARO framework (for Knowledge, Abilities, Results and Opportunities) falls within a tradition of the use of *modal logic(s)* for describing the behaviour of intelligent agents. As such it builds upon an even longer tradition of philosophical logic where modal logic is employed to describe intensional notions such as knowledge, belief, time, obligation, desire, etc. These very notions are deemed to be appropriate for specifying, designing and implementing intelligent agents (Weiss 1999).

Formally, the language of the KARO framework is defined over three primitive types: (i) a set of countably infinite atomic propositional variables, (ii) a set Ag of agent names (a finite subset of the positive integers), and (iii) a set $\mathsf{Ac_{at}}$ of (countably infinite) atomic action variables.

Formulae are defined inductively as follows. (i) $\top$ is an atomic propositional formula; (ii) $\varphi \vee \psi$ and $\neg\varphi$ are propositional formula provided so are φ and ψ; (iii) $\mathbf{K}_i\varphi$ (knowledge), $\langle \mathsf{do}_i(\alpha)\rangle\varphi$ (achievement of results by actions), $\mathbf{A}_i\alpha$ (ability), $\mathbf{O}_i\alpha$ (opportunity), $\mathbf{W}^s_i\varphi$ (selected wish), and $\Diamond_i\varphi$ (implementability) are propositional formulae, provided i is an agent name, α is an action formula and φ is a propositional formula; (iv) `id` (skip) is an atomic action formula; (v) $\alpha \vee \beta$ (non-deterministic choice), $\alpha\,;\beta$ (sequencing), $\varphi!$ (confirm), $\alpha^{(n)}$ (bounded repetition), and $\alpha^\star$ (unbounded repetition) are action formulae, provided α and β are action formulae, φ is a propositional formula, and n is a natural number. Implicit connectives include $\bot$, $\wedge$, $\rightarrow$, $\ldots$ for propositional formulae, the duals of $\mathbf{K}_i$ and $\langle \mathsf{do}_i(\alpha)\rangle$ (denoted by $[\mathsf{do}_i(\alpha)]$), as well as $\mathbf{P}_i(\alpha,\varphi) = \langle \mathsf{do}_i(\alpha)\rangle\varphi \wedge \mathbf{A}_i\alpha$.

The semantics of the core KARO logic is based on *interpretations* $\mathcal{M} = (W, V, D, I, M)$, where (i) W is a non-empty set of worlds, (ii) V maps propositional variables to subsets of W, (iii) for every $i \in \mathsf{Ag}$ and every $a \in \mathsf{Ac_{at}}$, D contains a binary relation $r_{(i,a)}$ on W and a subset $c_{(i,a)}$ of W, (iv) I contains an equivalence relation K_i on W for each agent $i \in \mathsf{Ag}$, and (v) M contains a serial relation relation W_i on W for each agent $i \in \mathsf{Ag}$. The relations r and sets c are extended to $\mathsf{Ag} \times \mathsf{Ac}$-sorted relations r^* and sets c^* in a way standard for dynamic logic. The semantics of well-formed formulae of the KARO logic is then defined in a standard way for the fusion of PDL, S5, and KD, e.g. we have

$$
\begin{array}{ll}
\mathcal{M}, w \models [\mathsf{do}_i(\alpha)]\varphi & \text{iff } \forall v \in W\ ((w,v) \in r^*_{(i,\alpha)} \rightarrow \mathcal{M}, v \models \varphi) \\
\mathcal{M}, w \models \mathbf{A}_i\alpha & \text{iff } w \in c^*_{(i,\alpha)} \\
\mathcal{M}, w \models \mathbf{O}_i\alpha & \text{iff } \mathcal{M}, w \models \langle \mathsf{do}_i(\alpha)\rangle\top \\
\mathcal{M}, w \models \mathbf{W}^s_i\varphi & \text{iff } \forall v \in W\ ((w,v) \in W_i \rightarrow \mathcal{M}, v \models \varphi) \\
\mathcal{M}, w \models \mathbf{K}_i\varphi & \text{iff } \forall v \in W\ ((w,v) \in K_i \rightarrow \mathcal{M}, v \models \varphi) \\
\mathcal{M}, w \models \mathbf{P}_i(\alpha,\varphi) & \text{iff } w \in c^*_{(i,\alpha)} \wedge \exists v \in W\ ((w,v) \in r^*_{(i,\alpha)} \wedge \mathcal{M}, v \models \varphi)
\end{array}
$$

One unusual operator of KARO logic is the *implementability operator* $\Diamond_i\varphi$ whose semantics is defined as follows.

$$
\mathcal{M}, w \models \Diamond_i\varphi \qquad \text{iff } \exists k \in \mathbb{N}\ \exists a_1, \ldots, a_k \in \mathsf{Ac_{at}}\ \mathcal{M}, w \models \mathbf{P}_i(a_1; \ldots ; a_k, \varphi)
$$

Recently, we started to examine proof methods for the KARO framework. One approach we have investigated is based on the reduction approach, in particular, the translation of modal formulae to first-order logic combined with first-order reasoning. In (Hustadt et al. 2001) we describe a decision procedure for the satisfiability problem in a fragment of KARO logic consisting of three components: (i) a normalisation function which reduces complex action formulae to atomic action subformulae, (ii) a particular translation of normalised formulae into a fragment of first-order logic, and (iii) a resolution-based decision procedure for this fragment. The particular fragment of KARO logic we consider, called *core KARO logic*, excludes the star operator $\_^\star$ and the implementability operator.

For this fragment we are able to define a normalisation function mapping any formula φ to a normal form $\varphi\downarrow$, such that $\varphi\downarrow$ and φ are logically equivalent and $\varphi\downarrow$ contains no non-atomic action formulae.

The particular translation we use has only recently been proposed by de Nivelle (1999) and can be seen as a special case of the T-encoding introduced by Ohlbach (1998). It allows for conceptually simple decision procedures for extensions of $\mathsf{K4}$ by ordered resolution without any reliance on loop checking or similar techniques.

Using certain structual transformation techniques, translated KARO formulae can be embedded into a number of solvable clausal classes, for example, the classes $\mathcal{S}^+$ (Fermüller et al. 2001) and DL^* (de Nivelle, Schmidt and Hustadt 2000). A decision procedures for DL^* can be formulated in the resolution framework of Bachmair and Ganzinger (2001) using an ordering refinement of resolution.

The exclusion of the star operator from the core KARO logic is due to the fact that this operator is not first-order definable and, therefore, is outside the scope of the approach sketched above. The exclusion of the implementability operator is partly due to an ongoing discussion about the most appropriate semantics for this operator. Note that the semantics of the implementability operator, as defined above, involves a quantification over arbitrary atomic actions, that is, a second-order quantification over arbitrary binary relations. It might be more appropriate if we only quantify over a fixed set of atomic actions, for example, those atomic actions occurring in the KARO formula under consideration. The exclusion of the implementability operator from the core KARO logic is also due to the fact that it poses similar problems as the star operator does for the reduction approach, that is, the problem introduced by the existential quantification over the natural numbers involved in the semantical definition of the implementability operator.

Future work will focus on overcoming these limitations. One approach is to incorporate some form of inductive reasoning into the resolution calculus. An alternative approach is to consider the combination of the reduction approach with a specialised calculus for PDL. In Hustadt, Dixon, Fisher, and Schmidt (2000) we present a combination of clausal temporal resolution (restricted to a linear time temporal logic) and the reduction approach. Similarly, it might be possible to combine a calculus for PDL with the reduction approach to the other modal logic incorporated into KARO logic.

References

L. Bachmair and H. Ganzinger (2001). Resolution theorem proving. In J. A. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning*. Elsevier.

H. de Nivelle (1999). Translation of S4 into GF and 2VAR. Manuscript.

H. de Nivelle, R. A. Schmidt, and U. Hustadt (2000). Resolution-based methods for modal logics. *Logic Journal of the IGPL*, 8(3):265–292.

C. Fermüller, A. Leitsch, U. Hustadt, and T. Tammet (2001). Resolution decision procedures. In J. A. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning*. Elsevier.

U. Hustadt, C. Dixon, R. A. Schmidt, and M. Fisher. Normal forms and proofs in combined modal and temporal logics. In *Proc. FroCoS'2000*, volume 1794 of *LNAI*, pages 73–87. Springer, 2000.

U. Hustadt, C. Dixon, R. A. Schmidt, M. Fisher, J.-J. Meyer, and W. van der Hoek (2001). Verification within the KARO Agent Theory. In *Proc. First Goddard Workshop on Formal Approaches to Agent-Based Systems*. Springer.

H. J. Ohlbach (1998). Combining Hilbert style and semantic reasoning in a resolution framework. In *Proc. CADE-15*, volume 1421 of *LNAI*, pages 205–219. Springer.

A. S. Rao and M. P. Georgeff (1991). Modeling agents withing a BDI-architecture. In *Proc. KR-91*, pages 473–484. Morgan Kaufmann.

B. van Linder, W. van der Hoek, and J.-J. Ch. Meyer (1996). How to motivate your agents. In *Intelligent Agents II*, volume 1037 of *LNAI*. Springer.

B. van Linder, W. van der Hoek, and J.-J. Ch. Meyer (1998). Formalizing abilities and opportunities of agents. *Fundamenta Informaticae*, 34(1,2):53–101.

G. Weiß, editor (1999). *Multiagent systems: A modern approach to distributed artificial intelligence*. MIT Press.