Towards Verifying a Blocks World for Teams GOAL Agent

Appendix

Alexander Birch Jensen
DTU Compute - Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, Building 324, DK-2800 Kongens Lyngby, Denmark
aleje@dtu.dk

APPENDIX A

We consider here the first proof step: it reflects that the agent should first move from its initial position to the room containing the red block.

For the sake of readability, we introduce the shorthand φ_1 **ensures** ψ_1 for the first proof step (1):

(1) $\mathsf{B}color(b_a, red, r_1) \wedge \mathsf{B}in(r_0) \wedge \neg\mathsf{B}holding(b_a) \wedge \mathsf{G}collect(red)$
ensures
$\mathsf{B}color(b_a, red, r_1) \wedge \underline{\mathsf{B}in(r_1)} \wedge \neg\mathsf{B}holding(b_a) \wedge \mathsf{G}collect(red)$

Effect of *goTo*(r_1)

We need to prove that $goTo(r_1)$ gives the desired state, i.e. the following Hoare triple must be satisfied:

$\{\varphi_1 \wedge \neg\psi_1\}\ enabled(goTo(r_1)) \triangleright do(goTo(r_1))\ \{\psi_1\}$

By applying the rule for conditional actions we are required to prove the formula:

$(\varphi_1 \wedge \neg\psi_1 \wedge \neg enabled(goTo(r_1))) \longrightarrow \psi_1$

We rewrite φ_1 and ψ_1 to their full definitions and immediately realize that the left-hand side is unsatisfiable. Additionally, we need to prove the Hoare triple:

$\{\varphi_1 \wedge \neg\psi_1 \wedge enabled(goTo(r_1)\}\ goTo(r_1)\ \{\psi_1\}$

For all actions a different from `goTo`, we supply the frame axiom

$\{\mathsf{B}in(X)\}\ a\ \{\mathsf{B}in(X)\}$

which captures that only the action `goTo` can change the agent's belief about its current position. Furthermore, for all actions a' different from `pickUp` or `putDown`, we supply the frame axiom

$\{\neg\mathsf{B}holding(X)\}\ a'\ \{\neg\mathsf{B}holding(X)\}$

which states that only the actions `pickUp` and `putDown` can change the agent's belief about blocks it is holding. Lastly, for all actions a'' different from `putDown`, we supply the frame axiom

$\{\mathsf{G}collect(red) \wedge \mathsf{B}color(b_a, red, r_1)\}\ a'\ \{\mathsf{G}collect(red) \wedge \mathsf{B}color(b_a, red, r_1)\}$

stating that the goal to collect a red block, and the information about the block, is unchanged by those actions. It may perhaps seem odd that picking up a block does not immediately change the belief about the position of the block, and this in a formalization detail we will not delve further into here.

We weaken the precondition and strengthen the postcondition by the consequence rule (and the invariant *inv-in*). We now need to prove the Hoare triple:

$\{\mathsf{B}color(b_a, red, r_1) \wedge \mathsf{B}in(r_0) \wedge \neg\mathsf{B}in(r_1) \wedge\ \neg\mathsf{B}holding(b_a) \wedge \mathsf{G}collect(red)\}$
$goTo(r_1)$
$\{\mathsf{B}color(b_a, red, r_1) \wedge \mathsf{B}in(r_1) \wedge \neg\mathsf{B}in(r_0) \wedge\ \neg\mathsf{B}holding(b_a) \wedge \mathsf{G}collect(red)\}$

By the conjunction rule we split the proof into subproofs of the following three Hoare triples:

$$\{\mathsf{B}in(r_0) \wedge \neg\mathsf{B}in(r_1)\}\ goTo(r_1)\ \{\mathsf{B}in(r_1) \wedge \neg\mathsf{B}in(r_0)\}$$

$$\{\neg\mathsf{B}holding(b_a)\}\ goTo(r_1)\ \{\neg\mathsf{B}holding(b_a)\}$$

$$\{\mathsf{G}collect(red) \wedge \mathsf{B}color(b_a, red, r_1)\}\ goTo(r_1)\ \{\mathsf{G}collect(red) \wedge \mathsf{B}color(b_a, red, r_1)\}$$

The first Hoare triple is the effect axiom for $goTo(r_1)$ that we derived in the transformation and the last two are frame axioms that we supplied above.

We need to show that other actions do not change the mental state since we only have a single trace. This plays well into our mutually exclusive decision rules. We will show it merely for a single action for completeness.

Non-effect of *goTo*(r_0)

To prove φ_1 **ensures** ψ_1, every action a should satisfy the Hoare triple $\{\varphi_1 \wedge \neg\psi_1\}\ a\ \{\varphi_1 \vee \psi_1\}$. For $goTo(r_0)$ we should thus prove:

$$\{\varphi_1 \wedge \neg\psi_1\}\ enabled(goTo(r_0)) \triangleright do(goTo(r_0))\ \{\varphi_1 \vee \psi_1\}$$

By the rule for conditional actions we need to assert the truth of the formula

$$(\varphi_1 \wedge \neg\psi_1 \wedge \neg enabled(goTo(r_0)) \longrightarrow (\varphi_1 \vee \psi_1)$$

which is easy to prove as the left-hand side of the implication is only true when φ_1 is true, which in turn guarantees the truth of the disjunction on the right-hand side. Furthermore, we must prove the Hoare triple:

$$\{\varphi_1 \wedge \neg\psi_1 \wedge enabled(goTo(r_0))\}\ goTo(r_0)\ \{\varphi_1 \vee \psi_1\}$$

By the consequence rule we weaken the precondition and strengthen the postcondition:

$$\{\mathsf{B}color(b_a, red, r_1) \wedge \mathsf{B}in(r_0) \wedge\ \neg\mathsf{B}holding(b_a) \wedge \mathsf{G}collect(red)\}$$
$$goTo(r_0)$$
$$\{\mathsf{B}color(b_a, red, r_1) \wedge \mathsf{B}in(r_0) \wedge\ \neg\mathsf{B}holding(b_a) \wedge \mathsf{G}collect(red)\}$$

Essentially, the Hoare triple above states that the action has no effect on the mental state. The rule for infeasible actions allows us to prove this if we can instead prove the formula (using the $enabled(goTo(r_0)$ equivalence):

$$\mathsf{B}color(b_a, red, r_1) \wedge \mathsf{B}in(r_0) \wedge \neg\mathsf{B}holding(b_a) \wedge \mathsf{G}collect(red) \longrightarrow \neg(\mathsf{B}(holding(b_a) \wedge \neg in(r_0)))$$

It is trivial to show that the formula is valid by considering the possible truth values of conjuncts on both sides of the implication.

Sketch of Remaining Proof Steps

We will merely sketch the remaining proof obligations: For the remainder of proof step (1), the proofs for other non-enabled actions are analogous as a result of mutually exclusive decision rules. For the proofs of steps (2)–(5), they are structurally similar to (1) but each for another enabled action and requiring additional invariants to be proved.