INFOMLSAI Logics for Safe AI
Coursework 4

Coursework released:	14 June 2021, on Blackboard
Coursework due:	23:59 25 June 2021, on Blackboard
Submission format:	a folder containing a pdf and .ispl file, one per group

Please do the coursework in groups of 2-3 people. Submit a single zipped folder on Blackboard for your group. The folder should contain the pdf file and the ispl file. Please name the folder group-X, where X is the number of your group, and state in the pdf file and comments in the ispl file the names of the members of the group.

CW4-1 Describe formally (list states, agents, actions, transition function, etc.) a CGS for the simplified bank robber example. It should be the same as M_{robb} in Figure 6.1 in the reader on p.89, but without the state q_0 and the guard agent (see Figure 1 below). The required CGS M^{-}_{robb} should have one agent, the robber r, who is uncertain whether the code for the vault is 0 or 1. The available actions are still to try to enter 0 or 1, and the results are the same as in M_{robb}: if the code is correct, the vault opens, and if not, an alarm sounds. For propositions, use access (for the vault opens), code0 which is true in q_1 and q_3, and code1 which is true in q_2 and q_4. (1 mark)

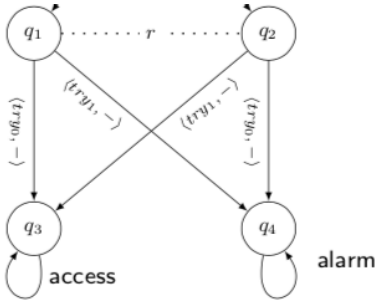

Figure 1: M^{-}_{robb} (only one agent: ignore the second action $-$ in transition labels $\langle try_0, - \rangle$ and $\langle try_1, - \rangle$)

CW4-2 Is it true in M^{-}_{robb}, q_1 under the ATL with perfect information semantics that the robber agent has a memoryless strategy to enforce access in the next state?

In other words, does it hold that $M^-_{robb}, q_1 \models_{Ir} \langle\langle r \rangle\rangle X \,\mathsf{access}$? Explain your answer. (1 mark)

CW4-3 Does it hold under the ATL with perfect information that $M^-_{robb}, q_1 \models_{Ir} K_r \langle\langle r \rangle\rangle X \,\mathsf{access}$? Explain your answer. (1 mark)

CW4-4 Does it hold under ATL$_{ir}$ semantics that $M^-_{robb}, q_1 \models_{ir} \langle\langle r \rangle\rangle X \mathsf{access}$? Explain your answer. (1 mark)

CW4-5 Does it hold under ATL$_{ir}$ semantics that $M^-_{robb}, q_1 \models_{ir} K_r \langle\langle r \rangle\rangle X \,\mathsf{access}$ Explain your answer. (1 mark)

CW4-6 Implement M^-_{robb} in ISPL. You can assume that the environment has the following variables: `code` (values 0 or 1) and `vault` (values `closed, open, alarm`) with the obvious evolution function (the vault opens if the robber enters the correct code and sounds alarm if the code is not correct). The robber can observe the vault but not the code. The initial states are q_1 and q_2 (where the vault is closed). In the Evaluation, access is true if the vault is open. Check the formulas from CW4-2 – CW4-5 under the default semantics and under the uniform strategy semantics. (Hint: in MCMAS, to ask for a uniform strategy, you need to use -uniform option.) Submit the ISPL encoding with the two formulas, and witness strategies under default and uniform semantics, if any. (1 mark)

CW4-7 How would you say in ATL with imperfect information and knowledge operators that the robber does not know what the code is (whether it is 0 or 1)? Hint: if you are checking whether the formula is true in your ISPL encoding, put EF in front of your formula, because MCMAS does not seem to like pure propositional or epistemic formulas (let's say it is a feature). (1 mark)

CW4-8 How would you say in ATL with imperfect information and knowledge operators that the robber will eventually know whether the code was 0 or 1? Is this formula true in M^-_{robb}, q_1 under ATL$_{ir}$ semantics? (1 mark)

CW4-9 The idea of checking for uniform strategies in MCMAS as follows [1], p.11. Suppose an ISPL encoding M is given, and the formula we are processing is $\langle\langle A \rangle\rangle \gamma$. Each agent in A has a protocol (assignment of sets of actions to states). Generate all possible 'uniform' models M' from M, where the protocol of each agent in A is restricted to be a singleton and contains the same action for each extended local state of the agent. So, for example, in M^-_{robb} above, since the robber has a choice of try_0 and try_1 in q_1 and q_2, for checking $\langle\langle r \rangle\rangle X \,\mathsf{access}$ there will be two models generated, one where the robber's protocol only contains try_0 in q_1 and q_2, and one where the robber only has try_1 in q_1 and q_2. Each of the new models essentially corresponds to a possible uniform strategy (since each agent has the same single action in all indistinguishable states). Then the formula $\langle\langle A \rangle\rangle \gamma$ is checked using the standard ATL with perfect information model checking algorithm in each M'. If in one of them the property evaluates to true, then there is a uniform strategy to satisfy the formula. Otherwise the formula is false.

Does this approach correspond to the ATL_{ir} semantics? Explain why. If it does not, give an example where it would give a different answer. (1 mark)

CW4-10 Give a model checking algorithm under ATL_{ir} semantics for a language containing propositional variables, booleans, and formulas $\langle\langle a\rangle\rangle X^2\varphi$ where a is a single agent the truth definition for $\langle\langle a\rangle\rangle X^2\varphi$ is $M, q \models \langle\langle a\rangle\rangle X^2\varphi$ iff there is a memoryless uniform strategy s_a for a such that for all paths λ in $\bigcup_{q' \sim_a q} out(q', s_a)$, $M, \lambda[2] \models \varphi$. (So the strategy guarantees to enforce φ in two steps from any state indistinguishable from q.) What is the complexity of your algorithm as a function of the transition relation and formula size?

References

[1] Alessio Lomuscio, Hongyang Qu, and Franco Raimondi. MCMAS: an open-source model checker for the verification of multi-agent systems. *Int. J. Softw. Tools Technol. Transf.*, 19(1):9–30, 2017.