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 pdf file, one per group

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

CW4-1 Define a Concurrent Epistemic Game Structure (CEGS) $M_{chicken}$ for the following example. Two agents are moving towards each other. They are not sure whether they are in a country which drives on the left (q_0) or on the right (q_1). Each agent can execute actions $left$ and $right$. In state q_0, if both agents go left, they pass each other correctly. In q_1, if both agents go right, they pass each other correctly. For all other combinations of actions, there is a crash. List the states, agents, indistinguishability relations, actions, transition function, assignment. Use propositional variables lft for driving on the left and crash for there is a crash. Distinguish between states resulting from different combinations of actions and whether they are in a country that drives on the left or a country that drives on the right. The agents can observe whether a crash happened or not, and what the actions leading to it were. For example, if both agents went left and a crash happened, they both know that they were not in a left-driving country. (1 mark)

CW4-2 Is it true in $M_{chicken}, q_0$ under the independent combination of ATL semantics with epistemic semantics (no uniform strategies requirement) that agent 1 has a memoryless strategy to enforce $\neg\mathsf{crash}$ in the next state ($\langle\langle 1\rangle\rangle X \neg \mathsf{crash}$)? (1 mark)

CW4-3 Does it hold under ATL$_{ir}$ semantics that $M_{chicken}, q_0 \models_{ir} \langle\langle 1\rangle\rangle X \neg \mathsf{crash}$? Explain your answer. (1 mark)

CW4-4 Is it true in $M_{chicken}, q_0$ under the independent combination of ATL semantics with epistemic semantics (no uniform strategies requirement) that both agents together have a memoryless strategy to enforce $\neg\mathsf{crash}$ in the next state? Explain your answer. (1 mark)

CW4-5 Does it hold under ATL$_{ir}$ semantics that $M_{chicken}, q_0 \models_{ir} \langle\langle 1,2 \rangle\rangle X \neg \mathsf{crash}$? Explain your answer. (1 mark)

CW4-6 Is it true in $M_{chicken}, q_0$ under the independent combination of ATL semantics with epistemic semantics (no uniform strategies requirement) that both agents know that they have a memoryless strategy to enforce $\neg \mathsf{crash}$ in the next state ($K_1 \langle\langle 1,2 \rangle\rangle X \neg \mathsf{crash} \wedge K_2 \langle\langle 1,2 \rangle\rangle X \neg \mathsf{crash}$)? Explain your answer. (1 mark)

CW4-7 Does it hold under ATL$_{ir}$ semantics that $M_{chicken}, q_0 \models_{ir} K_1 \langle\langle 1,2 \rangle\rangle X \neg \mathsf{crash} \wedge K_2 \langle\langle 1,2 \rangle\rangle X \neg \mathsf{crash}$? Explain your answer. (1 mark)

CW4-8 How would you say in ATL$_{ir}$ that agent 1 can ensure that eventually it knows whether it is in a country that drives on the left or a country that drives on the right? Is this formula true in q_0? Explain your answer. (1 mark)

CW4-9 How would you say in ATL$_{ir}$ that it is inevitable that if in the next state there is no crash, then agent 1 knows whether it is in a country that drives on the left or a country that drives on the right? Is this formula true in q_0? Explain your 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 and $\langle\langle a \rangle\rangle X^2$ is a new modality which means 'reachable in two steps' (note that $\langle\langle a \rangle\rangle X^2$ is not definable in ATL$_{ir}$).

The truth definition for $\langle\langle a \rangle\rangle X^2 \varphi$ is:

$$M, q \models \langle\langle a \rangle\rangle X^2 \varphi \text{ iff there is a memoryless uniform strategy } s_a \text{ for } a \text{ such that for all paths } \lambda \text{ in } \bigcup_{q' \sim_a q} out(q', s_a),\ M, \lambda[2] \models \varphi.$$

(It requires that the strategy is guaranteed to enforce φ in two steps from any state indistinguishable from q.) What is the big O complexity of your algorithm as a function of the model size and formula size? (Note that we are not asking for the most efficient algorithm, just a correct one with correct complexity analysis.) (1 mark)

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.