INFOMLSAI Logics for Safe AI

Mock Exam

This is an open-book examination

You can use your notes, the textbook, lecture slides and recordings etc.

This is an individual assessment. Under no circumstances are you to discuss any aspect of this assessment with anyone; nor are you allowed to share this document, ideas or solutions with others using email, social media, instant messaging, websites, or any other means. Your attempts at these questions must be entirely your own work.

Marks available for sections of questions are shown in brackets in the right-hand margin. This exam is marked out of 50.

The suggested time to answer both questions is about 1 hour.
You can write your answers using word processing software, and also include within the document scanned or photographed portions that you have written by hand, or write it all by hand and scan or photograph it to produce a single PDF. Use the standard naming convention for your document: YourSolisId.pdf. Write your Solis ID number at the top of each page of your answers. Do not include your name.

Q1 Consider the state transition systems in Figure 1. Suppose proposition p holds in states s_{11} and s_{12} and in no other state.

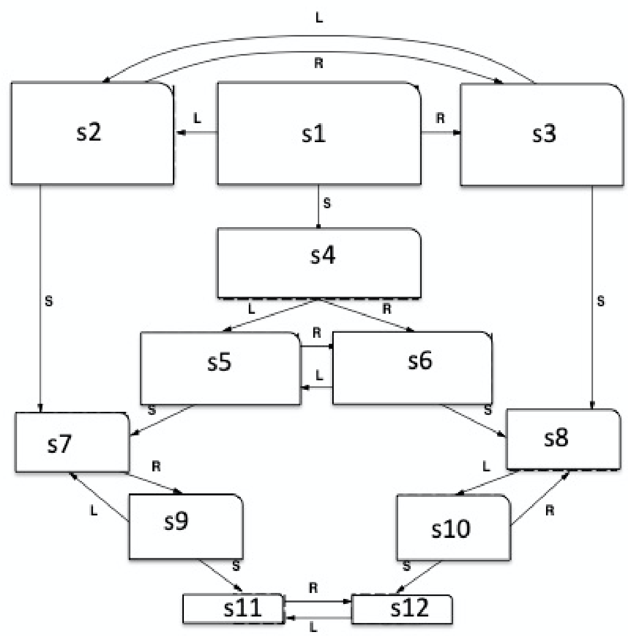

Figure 1: State transition system

(i) Express in LTL: at some point in the future p holds. Is this formula true on all paths starting in s_1? Explain why with reference to the truth definition for LTL. (5 marks)

(ii) Express in CTL: it is possible in three steps to reach a state where p holds. Is this formula true in s_1? Explain why with reference to the truth definition for CTL. (5 marks)

(iii) What does the formula $EG\,\neg p$ mean? Is it true in s_1? Explain why with reference to the truth definition for CTL. (5 marks)

(iv) What does the formula $E\neg p\,U\,AG\,p$ mean? Is it true in s_1? Explain why with reference to the truth definition for CTL. (5 marks)

(v) Trace the global model checking algorithm for formula $EXEG\,p$ on this state transition system. Use the algorithm as presented in Lecture 2/2 (slides 7-10). (5 marks)

Q2 Consider the vacuum cleaner domain from Russell and Norvig's textbook. In Figure 2, there is a representation of states that the agent considers possible when it has no sensors at all.

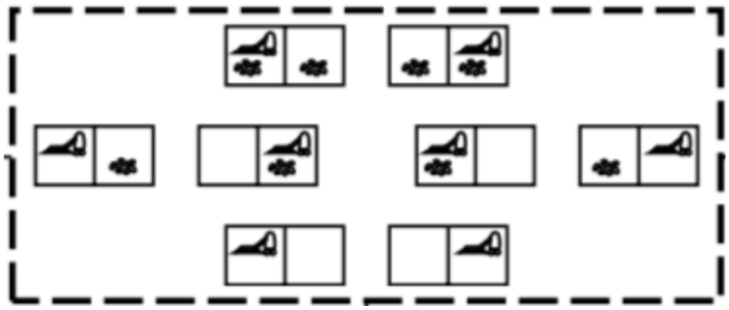

Figure 2: Vacuum cleaner with no sensors

(i) Represent this as a Kripke model with states (possible worlds) $w_1, \ldots, w_8$, one agent (so one indistinguishability relation $\sim_1$) and propositions inA, inB, cleanA, cleanB. inA is true when the agent is in the room on the left, and inB when the agent is in the room on the right. (5 marks)

(ii) Explain why, in the one agent case, distributed knowledge and common knowledge coincide (why $D_{\{1\}}\varphi$ is true if and only if $C_{\{1\}}\varphi$ is true). (5 marks)

(iii) Express in epistemic logic: agent 1 knows that it does not know whether room A is clean. (5 marks)

(iv) Is the formula from **(iii)** above true in all states in the model? Justify your answer referring to the truth definition for epistemic logic formulas in Kripke models. (5 marks)

(v) Figure 3 describing vacuum cleaner domain from Russell and Norvig's textbook demonstrates how an agent acquires knowledge by executing actions. For example, executing L (left) makes sure that the agent knows it is in room A, because this action always results in moving to or staying in room A. Executing S (suck) makes sure that the agent knows that the room where it is located is clean. Suppose you model the system depicted in Figure 3 as an interpreted system. How many local states of the agent are there? How many states of the environment? How many global states? (5 marks)

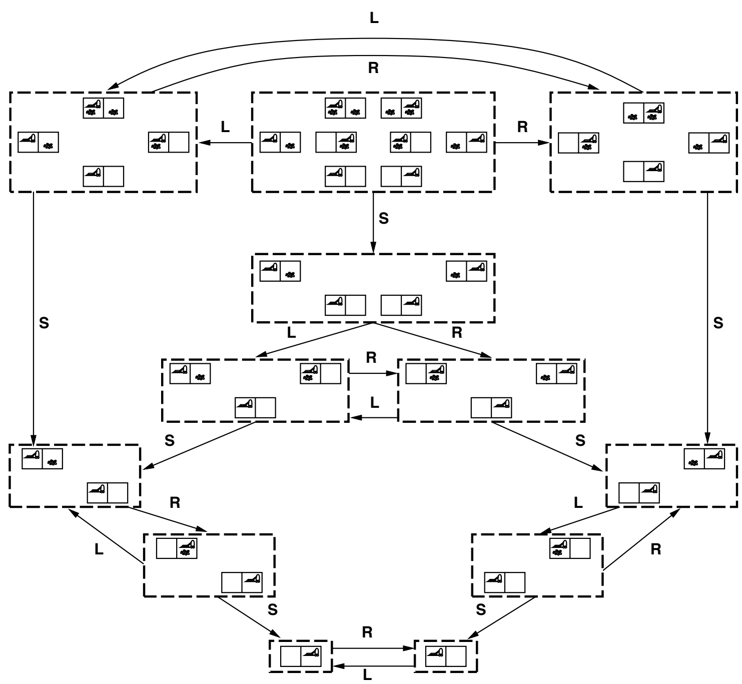

Figure 3: Vacuum cleaner world domain from Russell and Norvig's textbook.