INFOMLSAI Logics for Safe AI
Mock exam answers

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.

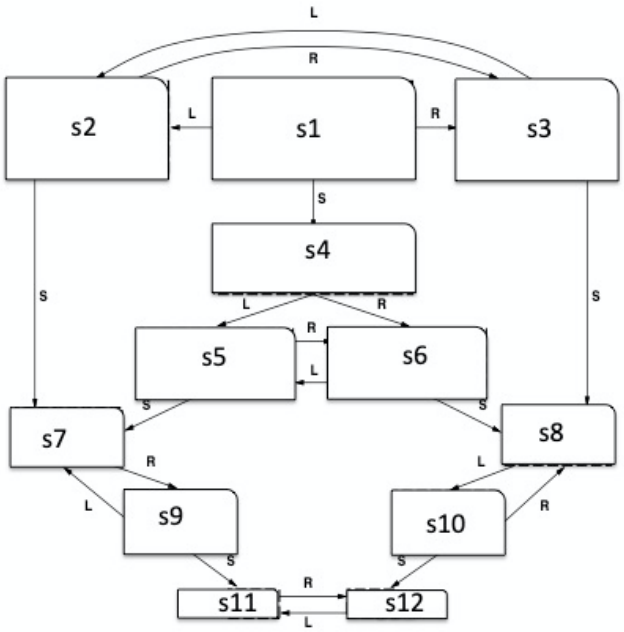

Figure 1: State transition system

(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. (10 marks)

Answer Fp. The formula is not true on all paths. For example it is not on the path $\lambda = s_1 s_2 s_3 s_2 s_3 \ldots$. There is no i on this path such that $\lambda[i, \infty] \models p$.

(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. (10 marks)

Answer $EXEXEXp$. No, this formula is not true, since there is no path of length 3 to a state where p holds (it is not possible in one step to reach a state satisfying $EXEXp$ which is a state from which it is possible in one step to reach a state satisfying EXp which is a state from which it is possible in one step to reach a state satisfying p).

(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. (10 marks)

It means that there is a path where in every state $\neg p$ holds. An example is $\lambda = s_1 s_2 s_3 s_2 s_3 \ldots$ where in every state $\neg p$ holds.

(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. (10 marks)

Answer It means that there is a path to a state satisfying $AG\,p$ where in every state on the path apart from the last one, $\neg p$. An example of such a path is $s_1, s_2, s_7, s_9, s_{11}$. In s_{11}, $AG\,p$ (on all paths in all states p) holds: the only path from s_{11} is $s_{11}, s_{12}, s_{11}, s_{12}, \ldots$ where p holds in every state.

(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). (10 marks)

Answer Let us call the transition system M.
$[p]_M \leftarrow \{s_{11}, s_{12}\}$ (because $\mathcal{V}(p) = \{s_{11}, s_{12}\}$)
Computing $[EG\,p]_M$:
$Q_1 \leftarrow \{s_1, \ldots, s_{12}\}$; $Q_2 \leftarrow \{s_{11}, s_{12}\}$;
$Q_1 \leftarrow \{s_{11}, s_{12}\}$; $Q_2 \leftarrow pre_\exists(Q_1) \cap Q_1 = \{s_{11}, s_{12}\}$
$Q_2 \subseteq Q_1$ so $[EG\,p]_M = \{s_{11}, s_{12}\}$
Computing $[EXEG\,p]_M$:
$[EXEG\,p]_M = pre_\exists([EG\,p]_M) = \{s_9, s_{10}, s_{11}, s_{12}\}$

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.

(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. (10 marks)

Answer $St = \{w_1, \ldots, w_8\}$; $\sim_1 = St^2$,
(assuming states are numbered left to right top to bottom)
$\mathcal{V}(\mathsf{inA}) = \{w_1, w_3, w_5, w_7\}$
$\mathcal{V}(\mathsf{inB}) = \{w_2, w_4, w_6, w_8\}$
$\mathcal{V}(\mathsf{cleanA}) = \{w_4, w_8\}$
$\mathcal{V}(\mathsf{cleanB}) = \{w_5, w_7\}$

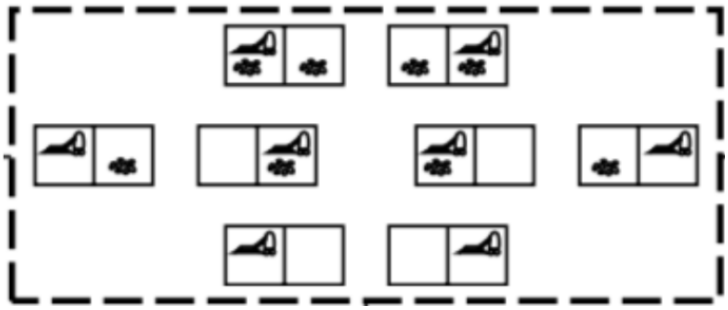

Figure 2: Vacuum cleaner with no sensors

(ii) Explain why in one agent case distributed knowledge and common knowledge coinside (why $D_{\{1\}}\varphi$ is true if and only if $C_{\{1\}}\varphi$ is true). (10 marks)

Answer $M, q \models D_{\{1\}}\varphi$ iff for every q' such that $q \sim^D_{\{1\}} q'$, $M, q' \models \varphi$. Since $\sim^D_{\{1\}} = \cap_i \in \{1\} \sim_i$, $\sim^D_{\{1\}} = \sim_1$, so $M, q \models D_{\{1\}}\varphi$ iff $M, q \models K_1\varphi$. Similarly, $\sim^C_{\{1\}}$ is a transitive closure of $\sim^E_{\{1\}}$ which is $\cup_i \in \{1\} \sim_i = \sim_1$ and $\sim_1$ is already a transitive relation, so $\sim^C_{\{1\}}$ is the same as $\sim_1$ and $M, q \models C_{\{1\}}\varphi$ iff $M, q \models K_1\varphi$. Since both $D_{\{1\}}\varphi$ and $C_{\{1\}}\varphi$ are equivalent to $K_1\varphi$, $D_{\{1\}}\varphi$ is true if and only if $C_{\{1\}}\varphi$ is true.

(iii) Express in epistemic logic: agent 1 knows that it does not know whether room A is clean. (10 marks)

Answer $K_1(\neg K_1 \mathsf{cleanA} \wedge \neg K_1 \neg \mathsf{cleanA})$

(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. (10 marks)

Answer yes, it is true in all states. In all states the agent considers possible at least one state where cleanA is true, and at least one state where it is false. So in every state $\neg K_1 \mathsf{cleanA}$ is true and $\neg K_1 \neg \mathsf{cleanA}$ is true. So in every state in the model $\neg K_1 \mathsf{cleanA} \wedge \neg K_1 \neg \mathsf{cleanA})$ is true. $K_1\varphi$ is true if in all indistinguishable states φ is true. Since for every state, in all indistinguishable states $\neg K_1 \mathsf{cleanA} \wedge \neg K_1 \neg \mathsf{cleanA})$ is true, the formula $K_1(\neg K_1 \mathsf{cleanA} \wedge \neg K_1 \neg \mathsf{cleanA})$ is true in every state.

(v) Figure 3 describing vacuum cleaner domain from Russell and Norvig's textbook demonstrates how 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? (10 marks)

Answer Interpreted system: there are 12 local states of the agent (each corresponding to a dashed box in Figure 3). There are 8 possible states of the environment (corresponding to $w_1, \ldots, w_8$ from **(i)**). So there are $12 \times 8 = 96$ global states (some of them unreachable because agent's knowledge will not correspond to the state of the environment, for example the environment state where both rooms are dirty is incompatible with s_{11} or s_{12} where the agent knows that they are clean).

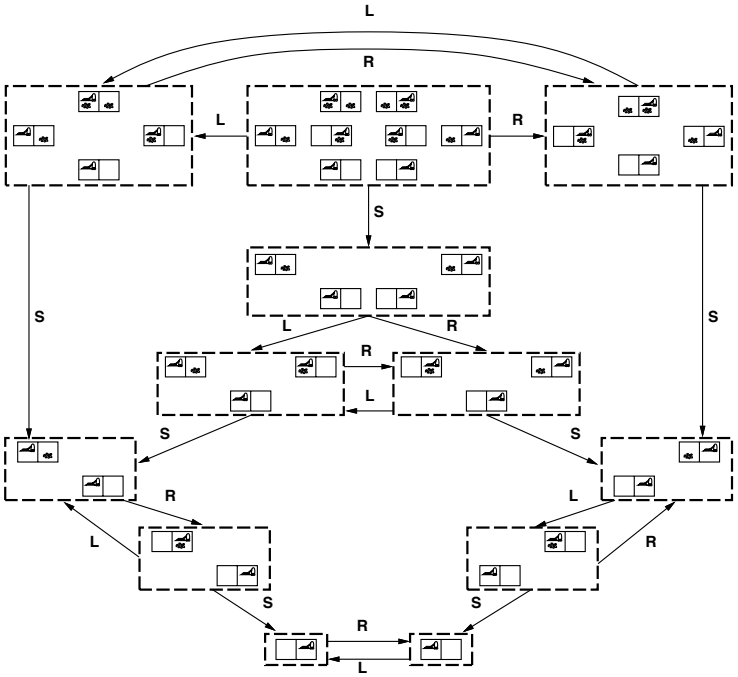

Figure 3: Vacuum cleaner world domain from Russell and Norvig's textbook.