A Temporal Logic for Stochastic Multi-Agent Systems

Wojciech Jamroga

Department of Informatics, Clausthal University of Technology, Germany
wjamroga@in.tu-clausthal.de

Abstract. Typical analysis of Markovian models of processes refers only to the expected utility that can be obtained by the process. On the other hand, modal logic offers a systematic method of characterizing processes by combining various modal operators. A multivalued temporal logic for Markov chains and Markov decision processes has been recently proposed in [1]. Here, we discuss how it can be extended to the multi-agent case. We relate the resulting logic to existing (two-valued) logics of strategic ability, and present fixpoint characterizations for some natural combinations of strategic and temporal operators.

Keywords: temporal logic, multi-agent system, Markov decision process.

1 Introduction

There are many different models of agents and multi-agent systems; however, most of them follow a similar pattern. First of all, they include information about possible situations (states of the system) that defines relations between states and their external characteristics (essentially, "facts of life" that are true in these states). Second, they provide information about relationships between states (e.g., possible transitions between states). Models that share this structure can be, roughly speaking, divided into two classes. *Qualitative models* provide no numerical measures for these relationships. *Quantitative models* assume that relationships are measurable, and provide numerical information about the degrees of relations. In [1], we explored analogies between transition systems and Markovian models in order to provide a more expressive language for reasoning about, and specification of agents in stochastic environments. In [2], we tentatively extended the framework to the multi-agent case. Here, we present some formal results on the multi-agent version of the language.

Analysis of quantitative process models is usually based on the notion of expected reward. On the other hand, logical approaches are most often concerned with "limit properties" like the existence of an execution path that displays a specific temporal pattern. We believe that both kinds of properties are interesting and worth using to describe processes. For instance, besides the expected value of cumulative future reward, we can ask of the maximal (or minimal) cumulative reward. Or, we might be concerned with the expected value of minimal guaranteed reward etc. A typical analysis of multi-agent Markov decision processes is even more constrained, as we assume that all the agents in the system cooperate to achieve a common goal (i.e., maximize their common expected cumulative reward). Our extension allows to study the outcomes that can be obtained by *various* groups of agents.

The roots of our proposal can be traced back to multivalued logics on one hand (e.g., fuzzy logics [3] and probabilistic logics [4, 5]), and (crisp) modal logics of probability [6–8] on the other. A closer inspiration comes from multi-valued modal logics [9–13]. Of the latter, [11–13] are particularly relevant, as they define multi-valued versions of temporal logic. Still, the version of Markov Temporal Logic proposed here is (to our best knowledge) the first multivalued logic for reasoning about strategic abilities of agents in stochastic multi-agent systems.

We begin by recalling the basic idea of Markov Temporal Logic (MTL) from [1] (Section 2). The remaining sections present the original contribution of the paper: the syntax and semantics of the multi-agent MTL was only presented at a workshop with informal proceedings [2], and the theoretical results (relationship to ATL$*$, fixpoint properties) are entirely new.

2 Markov Temporal Logic

In this section we recall the idea of Markov Temporal Logic (MTL) from [1]. The logic allows for flexible reasoning about outcomes of agents acting in stochastic environments. The core of the logic is called MTL_0, and addresses outcomes of Markov chains. Intuitively, MTL_0 can be seen as a quantitative analogue of the branching-time logic CTL* [14].

2.1 Basic Models: Markov Chains

Typically, a Markov chain [15, 16] is a directed graph with probabilistic transition relation. In our definition, we include also a device for assigning states with utilities and/or propositional values. This is done through *utility fluents* which generalize atomic propositions in modal logic in the sense that they can take both numerical and qualitative truth values.

Definition 1 (Domain of truth values). *A* domain $D = \langle U, \top, \bot, ^{-} \rangle$ *consists of: (1) a set* $U \subseteq \mathbb{R}$ *of* utility values *(or simply* utilities*); (2) special values* $\top, \bot$ *standing for the logical truth and falsity, respectively;* $\hat{U} = U \cup \{\top, \bot\}$ *will be called the* extended utility set*; and, finally, (3) a complement function* $^{-} : \hat{U} \to \hat{U}$*. A domain should satisfy the conditions specified in [1], omitted here for lack of space.*

Definition 2 (Markov chain). *A* Markov chain *over domain* $D = \langle U, \top, \bot, ^{-} \rangle$*, and a set of utility fluents* Π *is a tuple* $M = \langle St, \tau, \pi \rangle$*, where:*

- St *is a set of* states *(we will assume that the set is finite and nonempty throughout the rest of the paper);*

– $\tau : St \times St \to [0,1]$ *is a* stochastic transition relation *that assigns each pair of states* q_1, q_2 *with a probability* $\tau(q_1, q_2)$ *that, if the system is in* q_1*, it will change its state to* q_2 *in the next moment. For every* $q_1 \in St$*,* $\tau(q_1, \cdot)$ *is assumed to be a probability distribution, i.e.* $\sum_{q \in St} \tau(q_1, q) = 1$.
  By abuse of notation, we will sometimes write $\tau(q)$ *to denote the set of states accessible in one step from* q*, i.e.* $\{q' \mid \tau(q, q') > 0\}$.
– $\pi : \Pi \times St \to \hat{U}$ *is a* valuation of utility fluents.

A *run* in Markov chain M is an infinite sequence of states $q_0 q_1 \dots$ such that each q_{i+1} can follow q_i with a non-zero probability. The set of runs starting from state q is denoted by $\mathcal{R}_M(q)$.[1] Let $\lambda = q_0 q_1 ...$ be a run and $i \in \mathbb{N}_0$. Then: $\lambda[i] = q_i$ denotes the ith position in λ, and $\lambda[i..\infty] = q_i q_{i+1} \dots$ denotes the infinite subpath of λ from position i on.

2.2 Logical Operators as Minimizers and Maximizers

Note that – when truth values represent utility of an agent – temporal operators "sometime" and "always" have a very natural interpretation. "Sometime p" ($\Diamond p$) can be rephrased as "p is achievable in the future". Thus, under the assumption that agents want to obtain as much utility as possible, it is natural to view the operator as maximizing the utility value along a given temporal path. Similarly, "always p" ($\Box p$) can be rephrased as "p is guaranteed from now on". In other words, $\Box p$ asks for the minimal value of p on the path. On a more general level, every universal quantifier is essentially a minimizer of truth values, while existential quantifiers can be seen as maximizers. Thus, $\mathsf{E}\gamma$ ("there is a path such that γ") maximizes the utility specified by γ across all paths that can occur; likewise, $\mathsf{A}\gamma$ ("for all paths γ") minimizes the value of γ across paths. Also, disjunction and conjunction can be seen as a maximizer and a minimizer: $\varphi \lor \psi$ reads easily as "the utility that can be achieved through φ or ψ", while $\varphi \land \psi$ reads as "utility guaranteed by both φ and ψ".

2.3 MTL$_0$: A Logic of Markov Chains

Operators of MTL$_0$ include path quantifiers $\mathsf{E}, \mathsf{A}, \mathsf{M}$ for the maximal, minimal, and average outcome of a set of temporal paths, respectively, and temporal operators $\Diamond, \Box, \mathsf{m}$ for the maximal, minimal, and average outcome along a given path.[2] Propositional operators follow the same pattern: $\lor, \land, \oplus$ refer to maximization, minimization, and weighted average of outcomes obtained from different utility channels or related to different goals. Finally, we have the "defuzzification" operator $\preccurlyeq$, which provides a two-valued interface to the logic. $\varphi_1 \preccurlyeq \varphi_2$ yields "true" if the outcome of φ_1 is less or equal to φ_2, and "false" otherwise. Among other advantages, it allows to define the classical computational problems of validity, satisfiability and model checking for MTL.

---

[1] If the model is clear from the context, the subscripts will be omitted.

[2] We allow to discount future outcomes with a discount factor c. Also, we introduce the "until" operator $\mathcal{U}$, which is more general than $\Diamond$.

Let $Bool(\omega) = \neg\omega \mid \omega\wedge\omega \mid \omega\oplus_c\omega \mid \omega \preccurlyeq \omega$ denote quasi-Boolean combinations of formulae of type ω. The syntax of MTL$_0$ can be defined by the following production rules:

$$\varphi ::= p \mid Bool(\varphi) \mid \mathsf{E}\gamma \mid \mathsf{M}\gamma,$$
$$\gamma ::= \varphi \mid Bool(\gamma) \mid \bigcirc_c \gamma \mid \Box_c\gamma \mid \gamma\,\mathcal{U}_c\,\gamma \mid \mathsf{m}_c\gamma,$$

where $p \in \Pi$ is a utility fluent, and c is a discount factor such that $0 < c \le 1$. Additionally, we define $\varphi_1 \cong \varphi_2 \equiv (\varphi_1 \preccurlyeq \varphi_2) \wedge (\varphi_2 \preccurlyeq \varphi_1)$. Boolean constants T, F ("true", "false"), disjunction, and the "sometime" temporal operator $\Diamond$ are defined in the standard way. The following shorthands are used for discount-free versions of temporal operators: $\bigcirc \equiv \bigcirc_1$, $\Diamond \equiv \Diamond_1$, $\Box \equiv \Box_1$, $\mathcal{U} \equiv \mathcal{U}_1$.

Example 1. Let r be a utility fluent that represents the immediate reward at each state. The following MTL$_0$ formulae define some interesting characteristics of a process: $\mathsf{Mm}_{0.9}\mathsf{r}$ (expected average reward with time discount 0.9), $\mathsf{Am}_{0.9}\mathsf{r}$ (guaranteed average reward with the same discount factor), $\mathsf{M}\Box\mathsf{r}$ (expected minimal undiscounted reward), and $\mathsf{A}\Diamond\mathsf{r}$ (guaranteed maximal reward).

The main idea behind MTL$_0$ is that formulae can refer to both quantitative utilities and qualitative truth values. Thus, we treat complex formulae as fluents, just like the atomic utility fluents from Π, through a valuation function that assigns formulae with extended utility values from $\hat{U}$. Let $M = \langle St, \tau, \pi\rangle$ be a Markov chain over domain $D = \langle U, \top, \bot, \overline{}\rangle$ and a set of utility fluents Π. The valuation function $[\cdot]$ is defined below.

- $[p]_{M,q} = \pi(p, q)$, for $p \in \Pi$;
- $[\neg\varphi]_{M,q} = \overline{[\varphi]_{M,q}}$;
- $[\varphi_1 \wedge \varphi_2]_{M,q} = \min([\varphi_1]_{M,q}, [\varphi_2]_{M,q})$;
- $[\varphi_1 \oplus_c \varphi_2]_{M,q} = (1-c)\cdot[\varphi_1]_{M,q} + c\cdot[\varphi_2]_{M,q}$;
- $[\varphi_1 \preccurlyeq \varphi_2]_{M,q} = \top$ if $[\varphi_1]_{M,q} \le [\varphi_2]_{M,q}$ and $\bot$ else;
- $[\mathsf{E}\gamma]_{M,q} = \sup\{[\gamma]_{M,\lambda} \mid \lambda \in \mathcal{R}(q)\}$;
- The Markovian path quantifier $\mathsf{M}\gamma$ produces the expected truth value γ across all the possible runs, cf. [16] for the formal construction;
- $[\varphi]_{M,\lambda} = [\varphi]_{M,\lambda[0]}$;
- $[\neg\gamma]_{M,\lambda}$, $[\gamma_1 \wedge \gamma_2]_{M,\lambda}$, $[\gamma_1 \oplus_c \gamma_2]_{M,\lambda}$, $[\gamma_1 \preccurlyeq \gamma_2]_{M,\lambda}$: analogous to Boolean combinations of "state formulae" φ;
- $[\bigcirc_c \gamma]_{M,\lambda} = c\cdot[\gamma]_{M,\lambda[1..\infty]}$;
- $[\Box_c\gamma]_{M,\lambda} = \inf_{i=0,1,\dots}\{c^i[\gamma]_{M,\lambda[i..\infty]}\}$;
- $[\gamma_1\,\mathcal{U}_c\,\gamma_2]_{M,\lambda} = \sup_{i=0,1,\dots}\big\{\min(\ \min_{0\le j<i}\{c^j[\gamma_1]_{M,\lambda[j..\infty]}\},\ c^i[\gamma_2]_{M,\lambda[i..\infty]})\big\}$;
- The Markovian temporal operator m_c produces the average discounted reward along the given run:

$$[\mathsf{m}_c\gamma]_{M,\lambda} = \begin{cases} (1-c)\sum_{i=0}^{\infty} c^i[\gamma]_{M,\lambda[i\dots\infty]} & \text{if } c<1 \\ \frac{\limsup_{i\to\infty}\frac{1}{i+1}\sum_{j=0}^{i}[\gamma]_{M,\lambda[j\dots\infty]}+\liminf_{i\to\infty}\frac{1}{i+1}\sum_{j=0}^{i}[\gamma]_{M,\lambda[j\dots\infty]}}{2} & \text{if } c=1 \end{cases}$$

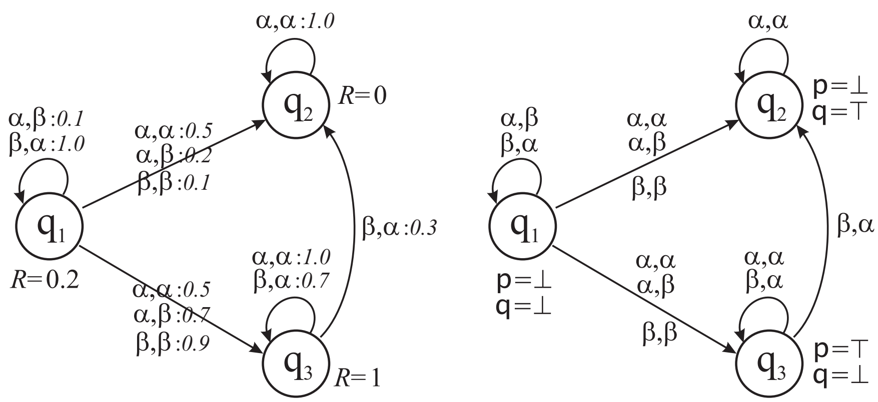

Fig. 1. (A) Simple MMDP with two agents; (B) Simple concurrent game structure

3 Reasoning about Stochastic Multi-Agent Processes

Strategic abilities were already considered in MTL_1, the version of Markov Temporal Logic for reasoning about Markov decision processes [1]. In consequence, MTL_1 can be seen as a quantitative analogue of the *single-agent* fragment of ATL$*$ [17] with memoryless strategies. In the more general case, a system can include multiple agents/processes, interacting with each other. To address their properties, a family of operators $\langle\langle A\rangle\rangle$ can be used, parameterized with groups of agents A. Intuitively, $\langle\langle A\rangle\rangle\varphi$ refers to how much agents A can "make out of" φ by following their best joint policy. This yields a language similar to the alternating-time temporal logic ATL$*$ from [17], albeit with strategic operators separated from path quantifiers.

Markov decision processes [18, 19] extend Markov chains with an explicit action structure: transitions are generated by actions of an (implicit) decision maker. *Multi-agent Markov decision processes* (MMDP) [20] extend Markov decision processes to the multi-agent setting: transitions are now labeled by combinations of agents' actions. We observe the similarity between MMDP's and concurrent game structures which are the models of ATL$*$ (cf. Figure 1).

As models for our multi-agent MTL, we will use a refinement of MMDP's similar to the version of Markov chains presented in Section 2.1. The semantics of $\langle\langle A\rangle\rangle\varphi$ is based on maximization of the value of φ with respect to A's joint strategies. We assume that the opponents play a strategy that minimizes φ most. This way, operator $\langle\langle A\rangle\rangle$ corresponds to the maxmin of the two-player game where A is the (collective) maximizer, and the rest of agents fills in the role of the (collective) minimizer. Note that such a semantics entails that the opponents of A must also play only *memoryless* (i.e., Markovian) strategies.

3.1 MTL$_2$: Syntax

Let $\mathbb{A}\mathrm{gt}$ be the set of all agents. MTL$_2$ adds to MTL$_0$ a family of operators $\langle\langle A\rangle\rangle$, one for each group of agents $A \subseteq \mathbb{A}\mathrm{gt}$. Formally, the syntax of MTL$_2$ is given by the following grammar:

$$\vartheta ::= p \mid Bool(\vartheta) \mid \langle\langle A\rangle\rangle\varphi,$$
$$\varphi ::= \vartheta \mid Bool(\varphi) \mid \mathsf{E}\gamma \mid \mathsf{M}\gamma,$$
$$\gamma ::= \varphi \mid Bool(\gamma) \mid \bigcirc_c \gamma \mid \Box_c\gamma \mid \gamma\,\mathcal{U}_c\,\gamma \mid \mathsf{m}_c\gamma.$$

An example formula of MTL$_2$ is $\langle\langle 1,2\rangle\rangle\mathsf{Amr}$ which makes agents 1 and 2 maximize the guaranteed average reward r with respect to their available policies.

3.2 MTL$_2$: Semantics

The semantics of MTL$_2$ is defined for a version of multi-agent Markov decision processes that incorporates qualitative as well as quantitative atomic properties of states.

Definition 3 (MMDP). *A* multi-agent Markov decision process *over domain $D = \langle U, \top, \bot, {}^{-}\rangle$ and a set of utility fluents Π is a tuple $\mathcal{M} = \langle \mathbb{A}\mathrm{gt}, St, \{Act_i\}_{i\in\mathbb{A}\mathrm{gt}}, \tau, \pi\rangle$, where: St, π are like in a Markov chain, $\mathbb{A}\mathrm{gt} = \{1,\ldots,k\}$ is the set of* agents*, Act_i is the set of* individual actions *of agent i, and $Act = \prod_{i\in\mathbb{A}\mathrm{gt}} Act_i$ is the space of* joint actions *(*action profiles*). $\tau : St \times Act \times St \to [0,1]$ is a stochastic transition relation; $\tau(q_1,\alpha,q_2)$ defines the probability that, if the system is in q_1 and the agents execute α, the next state will be q_2. For every $q \in St, \alpha \in Act$, we assume that either (1) $\tau(q,\alpha,q') = 0$ for all q' (i.e., α is not enabled in q), or (2) $\tau(q,\alpha,\cdot)$ is a probability distribution. Additionally, we define $act(q) = \{\alpha \in Act \mid \exists q'.\ \tau(q,\alpha,q') > 0\}$ as the set of enabled action profiles in q.*

For a joint action α, we define α^i to denote agent i's individual part in α, and we extend the notation to sets of joint actions and agents. Also, let $\mathcal{A}$ be a set of action profiles, and α a collective action of agents A. Then, $\mathcal{A}|\alpha = \{\beta \in \mathcal{A} \mid \beta^A = \alpha\}$ is the set of action profiles that include α.

A policy is a conditional plan that specifies future actions of an agent. Policies can be stochastic as well, thus allowing for randomness in the agent's play.

Definition 4. *An* individual strategy *(*policy*) of agent i is a function $s_i : St \times Act_i \to [0,1]$ that assigns each state q with a probability distribution over i's enabled actions $act(q)^i$. That is, $s(q,\alpha_i) \in [0,1]$ for all $q \in St, \alpha_i \in act(q)^i$, and $\sum_{\alpha_i\in act(q)^i} s(q,\alpha_i) = 1$. Values of $s(q,\alpha_i)$ for $\alpha_i \notin act(q)^i$ are irrelevant. The set of all i's strategies is denoted by Σ_i. A* collective strategy *s_A for team $A \subseteq \mathbb{A}\mathrm{gt}$ is simply a tuple of individual strategies, one per agent from A. The set of all A's collective strategies is given by $\Sigma_A = \prod_{i\in A}\Sigma_i$. The set of all* strategy profiles *in a model is given by $\Sigma = \Sigma_{\mathbb{A}\mathrm{gt}}$.*

For a collective strategy s, we define s^i as the i's individual part in s. We also extend the notation to sets of agents.

Definition 5. *Policy* $s \in \Sigma_A$ instantiates MMDP $\mathcal{M} = \langle \mathbb{A}\mathrm{gt}, St, \{Act_i\}_{i \in \mathbb{A}\mathrm{gt}}, \tau, \pi \rangle$ *to a simpler* MMDP $\mathcal{M} \dagger s = \langle \mathbb{A}\mathrm{gt} \setminus A, St, \{Act_i\}_{i \in \mathbb{A}\mathrm{gt} \setminus A}, \tau', \pi \rangle$ *with*

$$\tau'(q, \alpha, q') = \sum_{\alpha' \in (act(q)|\alpha)} \big(\prod_{i \in A} s^i(q, \alpha')\big)\ \tau(q, \alpha', q').$$

If $A = \mathbb{A}\mathrm{gt}$*, then* s *instantiates* $\mathcal{M}$ *to a Markov chain.*

The semantics of MTL$_2$ extends that of MTL$_0$ with the following clauses:

- $[p]_{\mathcal{M},q} = \pi(p, q)$, for $p \in \Pi$;
- $[\neg\vartheta]_{\mathcal{M},q}$, $[\vartheta_1 \wedge \vartheta_2]_{\mathcal{M},q}$, $[\vartheta_1 \oplus_c \vartheta_2]_{\mathcal{M},q}$, $[\vartheta_1 \preccurlyeq \vartheta_2]_{\mathcal{M},q}$: analogous as for "state formulae" φ;
- $[\langle\langle A \rangle\rangle \varphi]_{\mathcal{M},q} = \sup_{s \in \Sigma_A} \inf_{t \in \Sigma_{\mathbb{A}\mathrm{gt} \setminus A}} \{[\varphi]_{\mathcal{M} \dagger \langle s,t \rangle, q}\}$;

In order to keep consistent with qualitative logics of strategic ability, we assume that instantiation of an MMDP by a policy s is "soft" in the sense that nested strategic operators discard previous instantiations and instantiate the original model again: $[\langle\langle A \rangle\rangle \varphi]_{\mathcal{M} \dagger s, q} = [\langle\langle A \rangle\rangle \varphi]_{\mathcal{M},q}$.

Example 2. Consider the multi-agent Markov decision process from Figure 1A, consisting of two agents (1 and 2). If the agents cooperate, they can maximize the expected achievable reward quite successfully, as $[\langle\langle 1, 2 \rangle\rangle \mathsf{M} \Diamond R]_{q_1} = 0.9$ (best policy: both agents play β in q_1 with probability 1; the choices at other states are irrelevant). If agent 1 is to maximize the expected achievable reward on his own, against adversary behavior of agent 2, then he is bound to be less successful: $[\langle\langle 1 \rangle\rangle \mathsf{M} \Diamond R]_{q_1} = 0.6$. Also, in this case agent 1 should employ a different policy, namely play α in q_1 with probability 1.

4 Formal Results

The semantics of MTL, presented in the previous section, portrays it as a language of arithmetic expressions that can be used to define numerical characteristics of Markov processes. However, MTL can be also seen as a *logic*, i.e. a set of *sentences* that are true in some contexts, and false (at least to a degree) in others. This view allows us to use the conceptual apparatus of mathematical logic to study e.g. the expressivity of the language. Also, we can state interesting properties of the domain (multi-agent stochastic processes) through formulae of MTL. To this end, we first define what it means for a formula to be valid and/or satisfiable.

4.1 Levels of Truth

Since every domain must include a distinguished value for the classical (complete) truth, validity of formulae can be defined in a straightforward way.

Definition 6 (Levels of validity). *Let* $\mathcal{M}$ *be a multi-agent Markov decision process,* q *a state in* $\mathcal{M}$*, and* ϑ *a formula of* MTL$_2$*. Then:*

- ϑ *is* true in $\mathcal{M}, q$ *(written $\mathcal{M}, q \models \vartheta$) iff* $[\vartheta]_{\mathcal{M},q} = \top$.
- ϑ *is* valid in $\mathcal{M}$ *(written $\mathcal{M} \models \vartheta$) iff it is true in every state of $\mathcal{M}$.*
- ϑ *is* valid for multi-agent Markov decision processes *(written $\models \vartheta$) iff it is valid in every* MMDP $\mathcal{M}$.
- *Additionally, for path formulae γ, we can say that γ holds on run λ in* MMDP $\mathcal{M}$ *(written $\mathcal{M}, \lambda \models \gamma$) iff* $[\gamma]_{\mathcal{M},\lambda} = \top$.

The notion of validity helps to express general properties of stochastic multi-agent systems in a neat logical way. Moreover, Definition 6 allows to define the typical decision problems for MTL_2 in a natural way:

- Given a formula ϑ, the *validity problem* asks if $\models \vartheta$;
- Given a formula ϑ, the *satisfiability problem* asks if there are $\mathcal{M}, q$ such that $\mathcal{M}, q \models \vartheta$;
- Given a model $\mathcal{M}$, state q and formula ϑ, the *model checking problem* asks if $\mathcal{M}, q \models \vartheta$.

For example, we can search for a model in which agent a can guarantee the average reward r to be at least 0.6 in the long run by solving the satisfiability problem for formula $0.6 \preccurlyeq \langle\langle a \rangle\rangle \mathsf{Am}\,\mathsf{r}$.

We consider model checking the most important of the three problems, since in the analysis of a stochastic system the domain specification is usually given by a procedural representation (rather than axiomatic theory). Some work on model checking multi-valued temporal logics has been reported in [11, 12]. Perhaps even more importantly, computing approximate "solutions" of MDP's is one of the central issues studied by the Markov community. Integration of the two approaches seems a very promising (and exciting) path for future research.

4.2 Concurrent Game Structures as MMDP's. Correspondence between $\mathbf{MTL_2}$ and ATL*

Multi-agent Markov decision processes can be seen as generalizations of concurrent game structures [17], in which quantitative information is added through non-classical values of atomic statements and probabilities of transitions. Conversely, concurrent game structures can be seen as a subclass of MMDP's with all fluents assuming only classical truth values.

Definition 7. *Let $\mathcal{M}$ be an* MMDP*. Formula φ is* propositional *in $\mathcal{M}$ iff it can take only the values of $\top, \bot$, i.e., $[\varphi]_{\mathcal{M},q} \in \{\top, \bot\}$ for all $q \in St$. A* concurrent game structure *is an* MMDP *with only propositional fluents.*

This way, we obtain the class of models that are used for qualitative alternating-time logics, i.e. ATL and ATL$*$. Of course, when interpreting formulae of qualitative ATL/ATL$*$, one must as well ignore the probabilities that are present in Markov decision processes. Note also that the semantics of the original ATL/ATL$*$ uses the "history-based" notion of a strategy (i.e., strategies assign choices to *histories* rather than single states), while our MTL_2 is underpinned by a much weaker notion of *memoryless* (or positional) strategies. This makes the two logics formally incomparable. However, we can show that MTL$_2$ strictly generalizes the memoryless version of ATL$*$. The latter was studied in [21] under the acronym of ATL$_{\mathrm{Ir}}*$ (ATL with Perfect **I**nformation and imperfect **r**ecall), and we will use the name here.

Proposition 1. *Let $\mathcal{M}$ be a transition system, and φ a formula of ATL$_{\mathrm{Ir}}*$. Moreover, let φ' be the result of replacing every occurrence of $\langle\!\langle A \rangle\!\rangle$ with $\langle\!\langle A \rangle\!\rangle \mathsf{A}$ in φ for all $A \subseteq \mathbb{A}\mathrm{gt}$. Then, $\mathcal{M}, q \models_{\mathrm{MTL}_2} \varphi'$ iff $\mathcal{M}, q \models_{\mathrm{ATL}_{\mathrm{Ir}}*} \varphi$.*

Proof (sketch). Let σ_A denote the set of *deterministic* memoryless strategies of group A.[3] The proof follows by induction on the structure of φ; here, we only sketch the induction step for the most important case, namely $\varphi \equiv \langle\!\langle A \rangle\!\rangle_{Ir}\gamma$. We recall from [21] the semantics of $\langle\!\langle A \rangle\!\rangle_{Ir}$: let $out_{\mathcal{M}}(q, s)$ be the set of paths in $\mathcal{M}$ that can result from execution of strategy s from state q on; then, $\mathcal{M}, q \models \langle\!\langle A \rangle\!\rangle_{Ir}\gamma$ iff there is $s \in \sigma_A$ such that for every $\lambda \in out_{\mathcal{M}}(q, s)$ we have $\mathcal{M}, \lambda \models \gamma$.

<u>**"MTL$_2$ $\Rightarrow$ ATL$_{\mathrm{Ir}}*$":**</u> Let $\mathcal{M}, q \models_{\mathrm{MTL}_2} \langle\!\langle A \rangle\!\rangle \mathsf{A}\gamma$. Then, $[\langle\!\langle A \rangle\!\rangle \mathsf{A}\gamma]_{\mathcal{M},q} = \top$, and so $\sup_{s\in\Sigma_A} \inf_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \inf_{\lambda\in\mathcal{R}_{\mathcal{M}\dagger\langle s,t\rangle}(q)} [\gamma]_{\mathcal{M}\dagger\langle s,t\rangle,\lambda} = \top$; let s^* be a strategy that maximizes the above expression. Note that all the state subformulae of γ will be in fact evaluated in the original MMDP $\mathcal{M}$, so we get that $\inf_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \inf_{\lambda\in\mathcal{R}_{\mathcal{M}\dagger\langle s^*,t\rangle}(q)} [\gamma]_{\mathcal{M},\lambda} = \top$. Thus, $\forall_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \forall_{\lambda\in\mathcal{R}_{\mathcal{M}\dagger\langle s^*,t\rangle}(q)} [\gamma]_{\mathcal{M},\lambda} = \top$, and by the induction hypothesis we obtain that $\forall_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \forall_{\lambda\in\mathcal{R}_{\mathcal{M}\dagger\langle s^*,t\rangle}(q)} \mathcal{M}, \lambda \models_{\mathrm{ATL}_{\mathrm{Ir}}*} \gamma$. Now we observe that if $s \in \Sigma_A$ is a randomized strategy and $\lfloor s \rfloor \in \sigma_A$ is any determinization of s then $\mathcal{R}_{\mathcal{M}\dagger\langle\lfloor s\rfloor,t\rangle}(q) \subseteq \mathcal{R}_{\mathcal{M}\dagger\langle s,t\rangle}(q)$, so also for $\lfloor s^* \rfloor$ we have that $\forall_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \forall_{\lambda\in\mathcal{R}_{\mathcal{M}\dagger\langle\lfloor s^*\rfloor,t\rangle}(q)} \mathcal{M}, \lambda \models_{\mathrm{ATL}_{\mathrm{Ir}}*} \gamma$. Finally, we take t to be the uniform randomized strategy of $\mathbb{A}\mathrm{gt}\setminus A$ since it does not remove any paths from the model: $\mathcal{R}_{\mathcal{M}\dagger\langle\lfloor s^*\rfloor,uniform\rangle}(q) = out_{\mathcal{M}}(q, \lfloor s^* \rfloor)$. In consequence, $\forall_{\lambda\in out_{\mathcal{M}}(q,\lfloor s^*\rfloor)} \mathcal{M}, \lambda \models_{\mathrm{ATL}_{\mathrm{Ir}}*} \gamma$, which concludes this part of the proof.

<u>**"ATL$_{\mathrm{Ir}}*$ $\Leftarrow$ MTL$_2$":**</u> Let $\mathcal{M}, q \models_{\mathrm{ATL}_{\mathrm{Ir}}*} \langle\!\langle A \rangle\!\rangle_{Ir}\gamma$. Then, $\exists_{s\in\sigma_A} \forall_{\lambda\in out_{\mathcal{M}}(q,s)} \mathcal{M}, \lambda \models_{\mathrm{ATL}_{\mathrm{Ir}}*} \gamma$. We take such s. By induction, $\forall_{\lambda\in out_{\mathcal{M}}(q,s)} \mathcal{M}, \lambda \models_{\mathrm{MTL}_2} \gamma$. Take any $t \in \Sigma_{\mathbb{A}\mathrm{gt}\setminus A}$, then $\mathcal{R}_{\mathcal{M}\dagger\langle s,t\rangle}(q) \subseteq out_{\mathcal{M}}(q, s)$, and hence also $\forall_{\lambda\in\mathcal{R}_{\mathcal{M}\dagger\langle s,t\rangle}(q)} \mathcal{M}, \lambda \models_{\mathrm{MTL}_2} \gamma$. As $\sigma_A \subseteq \Sigma_A$, we finally get that $\exists_{s\in\Sigma_A} \forall_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \forall_{\lambda\in\mathcal{R}_{\mathcal{M}\dagger\langle s,t\rangle}(q)} \mathcal{M}, \lambda \models_{\mathrm{MTL}_2} \gamma$. In consequence, $\sup_{s\in\Sigma_A} \inf_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \inf_{\lambda\in\mathcal{R}_{\mathcal{M}\dagger\langle s,t\rangle}(q)} [\gamma]_{\mathcal{M}\dagger\langle s,t\rangle,\lambda} = \top$, which concludes the proof.

Proposition 2. *There is a transition system M with states q, q' which cannot be distinguished by any formula of ATL$*$ nor ATL$_{\mathrm{Ir}}*$, and can be distinguished by a formula of* MTL$_2$.

Proof. Consider the transition system in Figure 2, which can be seen as a concurrent game structure with a single agent ($\mathbb{A}\mathrm{gt} = \{1\}$) and a single action that can be executed ($Act = \{\alpha\}$). Note that states q_1, q_2 are bisimilar under CTL* bisimulation, so the same CTL* properties hold in both states (cf. e.g. [22]). Since the agent cannot make any real choices, both ATL$*$ and ATL$_{\mathrm{Ir}}*$ have no more

[3] Recall that Σ_A is the set of all (possibly randomized) memoryless strategies of A.

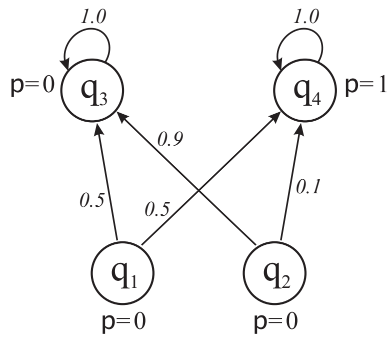

Fig. 2. MTL$_2$ vs. ATL$*$: probabilities matter!

distinguishing power for this model as CTL*, and hence the same properties of ATL$*$ (resp. ATL$_{\mathrm{Ir}}*$) hold in q_1, q_2 as well.

On the other hand, we have that $[\langle\langle 1\rangle\rangle\mathsf{Mmp}]_{q_1} = 0.5 = [\langle\langle 1\rangle\rangle\mathsf{Em}_{0.5}\mathsf{p}]_{q_1}$, and $[\langle\langle 1\rangle\rangle\mathsf{Mmp}]_{q_2} = 0.1 \neq 0.5 = [\langle\langle 1\rangle\rangle\mathsf{Em}_{0.5}\mathsf{p}]_{q_2}$. Thus, for $\varphi \equiv (\langle\langle 1\rangle\rangle\mathsf{Mmp} \cong \langle\langle 1\rangle\rangle\mathsf{Em}_{0.5}\mathsf{p})$, we have $q_1 \models \varphi$ and $q_2 \not\models \varphi$ (and even $q_2 \models \neg\varphi$).

The above example shows that a proper notion of bisimulation for Markov decision processes must take into account transition probabilities.

4.3 State-Based Formulae and Bellman Equations

"ATL without star" (or "vanilla ATL") is the most often used variant of alternating-time temporal logic, mainly due to the complexity of its model checking problem and the fact that its semantics can be defined entirely in relation to states. "Vanilla" ATL can be seen as a syntactic restriction of ATL*, in which every temporal modality is preceded by exactly one path quantifier. In this section, we consider a similar syntactic restriction on MTL$_2$; we call it state-based MTL$_2$.

Definition 8. State-based MTL$_2$ *(sMTL$_2$ in short) is given as follows:*

$$\vartheta ::= p \mid Bool(\vartheta) \mid \langle\langle A\rangle\rangle\varphi,$$
$$\varphi ::= \mathsf{E}\gamma \mid \mathsf{M}\gamma,$$
$$\gamma ::= \bigcirc_c \vartheta \mid \Box_c\vartheta \mid \vartheta\,\mathcal{U}_c\,\vartheta \mid \mathsf{m}_c\vartheta.$$

Proposition 3 presents fixpoint characterizations for most modalities of sMTL$_2$. Note that the last validity from the list is in fact a modal formulation of *Bellman equation*, which is the basic law used in analysis of Markov decision processes. The other formulae can be seen as variants of the equation for non-standard analysis based on minimal/maximal rather than average rewards. The results from [12] suggest that $\langle\langle A\rangle\rangle\mathsf{M}\Box_c$ and $\langle\langle A\rangle\rangle\mathsf{M}\,\mathcal{U}_c$ do not have fixpoint characterizations, but this remains to be formally proven.

Proposition 3. *The following formulae of* sMTL$_2$ *are valid:*

– $\langle\!\langle A\rangle\!\rangle \mathsf{E}\Box_c\varphi \cong \varphi \wedge \langle\!\langle A\rangle\!\rangle \mathsf{E}\bigcirc_c \langle\!\langle A\rangle\!\rangle \mathsf{E}\Box_c\varphi$;
– $\langle\!\langle A\rangle\!\rangle \mathsf{A}\Box_c\varphi \cong \varphi \wedge \langle\!\langle A\rangle\!\rangle \mathsf{A}\bigcirc_c \langle\!\langle A\rangle\!\rangle \mathsf{A}\Box_c\varphi$;
– $\langle\!\langle A\rangle\!\rangle \mathsf{E}\varphi_1 \mathcal{U}_c \varphi_2 \cong \varphi_2 \vee \varphi_1 \wedge \langle\!\langle A\rangle\!\rangle \mathsf{E}\bigcirc_c \langle\!\langle A\rangle\!\rangle \mathsf{E}\varphi_1 \mathcal{U}_c \varphi_2$;
– $\langle\!\langle A\rangle\!\rangle \mathsf{A}\varphi_1 \mathcal{U}_c \varphi_2 \cong \varphi_2 \vee \varphi_1 \wedge \langle\!\langle A\rangle\!\rangle \mathsf{A}\bigcirc_c \langle\!\langle A\rangle\!\rangle \mathsf{A}\varphi_1 \mathcal{U}_c \varphi_2$;
– $\langle\!\langle A\rangle\!\rangle \mathsf{Em}_c\varphi \cong \varphi \oplus_c \langle\!\langle A\rangle\!\rangle \mathsf{E}\bigcirc \langle\!\langle A\rangle\!\rangle \mathsf{Em}_c\varphi$;
– $\langle\!\langle A\rangle\!\rangle \mathsf{Am}_c\varphi \cong \varphi \oplus_c \langle\!\langle A\rangle\!\rangle \mathsf{A}\bigcirc \langle\!\langle A\rangle\!\rangle \mathsf{Am}_c\varphi$;
– $\langle\!\langle A\rangle\!\rangle \mathsf{Mm}_c\varphi \cong \varphi \oplus_c \langle\!\langle A\rangle\!\rangle \mathsf{M}\bigcirc \langle\!\langle A\rangle\!\rangle \mathsf{Mm}_c\varphi$.

Proof (sketch). We will sketch the proof of the first validity; the others can be proved in an analogous way.

Let $L = [\langle\!\langle A\rangle\!\rangle \mathsf{E}\Box_c\varphi]_{\mathcal{M},q}$ and $R = [\varphi \wedge \langle\!\langle A\rangle\!\rangle \mathsf{E}\bigcirc_c \langle\!\langle A\rangle\!\rangle \mathsf{E}\Box_c\varphi]_{\mathcal{M},q}$. It is easy to see that $R = \min([\varphi]_{\mathcal{M},q},\ c\cdot \sup_{s\in\Sigma_A} \inf_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \sup_{q'\in\tau_{\mathcal{M}\dagger\langle s,t\rangle}(q)} \sup_{s'\in\Sigma_A} \inf_{t'\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \{[\mathsf{E}\Box_c\varphi]_{\mathcal{M}\dagger\langle s',t'\rangle,q'}\})$. Moreover, by [1, Proposition 8], we get that $L = \min([\varphi]_{\mathcal{M},q},\ c\cdot \sup_{s\in\Sigma_A} \inf_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \sup_{q'\in\tau_{\mathcal{M}\dagger\langle s,t\rangle}(q)} \{[\mathsf{E}\Box_c\varphi]_{\mathcal{M}\dagger\langle s,t\rangle,q'}\})$. Thus, in order to prove $L = R$, it is sufficient to prove that

$$\sup_{s\in\Sigma_A} \inf_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \sup_{q'\in\tau_{\mathcal{M}\dagger\langle s,t\rangle}(q)} \{[\mathsf{E}\Box_c\varphi]_{\mathcal{M}\dagger\langle s,t\rangle,q'}\}$$
$$= \sup_{s\in\Sigma_A} \inf_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \sup_{q'\in\tau_{\mathcal{M}\dagger\langle s,t\rangle}(q)} \sup_{s'\in\Sigma_A} \inf_{t'\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \{[\mathsf{E}\Box_c\varphi]_{\mathcal{M}\dagger\langle s',t'\rangle,q'}\}.$$

The difference between the sides of the equation is that in the left hand side optimal strategies s, t are chosen once (at state q), while in the right hand side strategies are re-evaluated after each step. Let $s*$ be a strategy of A that optimizes L, and let us take s and s' in R to be the same as $s*$ in L. We observe that $\inf_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \sup_{q'\in\tau_{\mathcal{M}\dagger\langle s*,t\rangle}(q)} \{[\mathsf{E}\Box_c\varphi]_{\mathcal{M}\dagger\langle s*,t\rangle,q'}\}$ is indeed equal to $\inf_{t\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \sup_{q'\in\tau_{\mathcal{M}\dagger\langle s*,t\rangle}(q)} \inf_{t'\in\Sigma_{\mathbb{A}\mathrm{gt}\setminus A}} \{[\mathsf{E}\Box_c\varphi]_{\mathcal{M}\dagger\langle s*,t'\rangle,q'}\}$. Thus, we obtain that A have at least as good options in R as in L, and hence $L \le R$.

For the other direction, note that s, t in R are only relevant wrt the agents' actions in state q (later s', t' will be used). By unfolding R, we obtain an infinite sequence of collective action profiles $s_n(q_n), t_n(q_n)$ which maximize (over A's actions) and minimize (over $\mathbb{A}\mathrm{gt}\setminus A$'s actions) the value of $\mathsf{E}\Box_c\varphi$ in the next step. Now we observe that, when the system returns to state q, the same strategies s, t will be again optimal for the respective parties since the same expression will be maximized/minimized. Thus, the sequence of action profiles can be combined into a single pair of memoryless strategies $s*, t*$, which maximizes/minimizes $\mathsf{E}\Box_c\varphi$ as good as the original sequence of strategies. In consequence, also $R \le L$.

5 Conclusions

We extend the Markov Temporal Logic MTL from [1] to handle Markovian models with multiple agents acting in parallel. In terms of formal results, we show that the resulting logic strictly embeds $\mathrm{ATL}_{\mathrm{Ir}}*$, i.e., alternating-time temporal logic with memoryless strategies. We also present fixpoint characterizations for some natural combinations of strategic, path, and temporal operators, that can be seen as analogues of Bellman equation. The characterizations enable computing the truth values of many MTL_2 formulae by solving sets of simple equations.

Acknowledgements. The research was partially conducted within the Polish development project no. O R000024 04.

References

1. Jamroga, W.: A temporal logic for Markov chains. In: Proceedings of AAMAS'08. (2008) 697–704
2. Jamroga, W.: A temporal logic for multi-agent MDP's. In: Proceedings of Workshop on Formal Models and Methods for Multi-Robot Systems. (2008) 29–34
3. Zadeh, L.: Fuzzy sets. Information and Control **8**(3) (1965) 338–353
4. Nilsson, N.J.: Probabilistic logic. Artificial Intelligence **28**(1) (1986) 71–87
5. Ruspini, E., Lowrance, J., Strat, T.: Understanding evidential reasoning. Artificial Intelligence **6**(3) (1992) 401–424
6. Hansson, H., Jonsson, B.: A logic for reasoning about time and reliability. Formal Aspects of Computing **6**(5) (1994) 512–535
7. Aziz, A., Singhal, V., Brayton, R.K., Sangiovanni-Vincentelli, A.L.: It usually works: The temporal logic of stochastic systems. In: Proceedings of CAV. Volume 939 of LNCS. (1995) 155–165
8. Halpern, J.Y.: A logical approach to reasoning about uncertainty: a tutorial. In Arrazola, X., Korta, K., Pelletier, F.J., eds.: Discourse, Interaction, and Communication. Kluwer (1998) 141–155
9. Godefroid, P., Huth, M., Jagadeesan, R.: Abstraction-based model checking using modal transition systems. In: Proceedings of CONCUR. Volume 2154 of LNCS. (2001) 426–440
10. Easterbrook, S., Chechik, M.: A framework for multi-valued reasoning over inconsistent viewpoints. In: International Conference on Software Engineering. (2001) 411–420
11. Konikowska, B., Penczek, W.: Model checking for multi-valued computation tree logic. In Fitting, M., Orlowska, E., eds.: Beyond Two: Theory and Applications of Multiple Valued Logic. (2003) 193–210
12. de Alfaro, L., Faella, M., Henzinger, T., Majumdar, R., Stoelinga, M.: Model checking discounted temporal properties. Theoretical Computer Science **345** (2005) 139–170
13. Lluch-Lafuente, A., Montanari, U.: Quantitative μ-calculus and CTL based on constraint semirings. Electr. Notes Theor. Comput. Sci. **112** (2005) 37–59
14. Emerson, E.A.: Temporal and modal logic. In van Leeuwen, J., ed.: Handbook of Theoretical Computer Science. Volume B. Elsevier Science Publishers (1990) 995–1072
15. Markov, A.: Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga. Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete **2**(15) (1906) 135–156
16. Kemeny, J.G., Snell, L.J., Knapp, A.W.: Denumerable Markov Chains. Van Nostrand (1966)
17. Alur, R., Henzinger, T.A., Kupferman, O.: Alternating-time Temporal Logic. Journal of the ACM **49** (2002) 672–713
18. Bellman, R.: A Markovian decision process. Journal of Mathematics and Mechanics **6** (1957) 679–684
19. Bellman, R.: Dynamic Programming. Princeton University Press (1957)
20. Boutilier, C.: Sequential optimality and coordination in multiagent systems. In: Proceedings of IJCAI. (1999) 478–485
21. Schobbens, P.Y.: Alternating-time logic with imperfect recall. Electronic Notes in Theoretical Computer Science **85**(2) (2004)
22. Moller, F., Rabinovich, A.: On the expressive power of CTL*. In: Proceedings of LICS'99. (1999) 360–369