Continuous-Time Stochastic Games with Time-Bounded Reachability

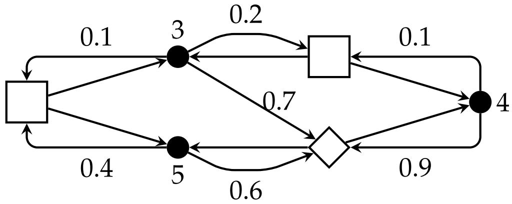
T. Brázdil, V. Forejt, J. Krčál, J. Křetínský, A. Kučera
Faculty of Informatics, Masaryk University, Botanická 68a, 60200 Brno, Czech Republic
{brazdil,forejt,krcal,xkretins,kucera}@fi.muni.cz

ABSTRACT. We study continuous-time stochastic games with time-bounded reachability objectives. We show that each vertex in such a game has a *value* (i.e., an equilibrium probability), and we classify the conditions under which optimal strategies exist. Finally, we show how to compute optimal strategies in finite uniform games, and how to compute ε-optimal strategies in finitely-branching games with bounded rates (for finite games, we provide detailed complexity estimations).

1 Introduction

Markov models are widely used in many diverse areas such as economics, biology, or physics. More recently, they have also been used for performance and dependability analysis of computer systems. Since faithful modeling of computer systems often requires both *randomized* and *non-deterministic* choice, a lot of attention has been devoted to Markov models where these two phenomena co-exist, such as *Markov decision processes* and *stochastic games*. The latter model of stochastic games is particularly apt for analyzing the interaction between a system and its environment, which are formalized as two *players* with antagonistic objectives (we refer to, e.g., [10, 5, 11] for more comprehensive expositions of results related to games in formal analysis and verification of computer systems). So far, most of the existing results concern *discrete-time* Markov decision processes and stochastic games, and the accompanying theory is relatively well-developed (see, e.g., [9, 4]).

In this paper, we study *continuous-time stochastic games (CTGs)* and hence also *continuous-time Markov decision processes (CTMDPs)* with time-bounded reachability objectives. Roughly speaking, a CTG is a finite or countably infinite graph with three types of vertices—controllable vertices (boxes), adversarial vertices (diamonds), and actions (circles). The outgoing edges of controllable and adversarial vertices lead to the actions that are *enabled* at a given vertex. The outgoing edges of actions lead to controllable or adversarial vertices, and every edge is assigned a positive probability so that the total sum of these probabilities is equal to 1. Further, each action is assigned a positive real *rate*. A simple finite CTG is shown below.



A game is played by two players, $\Box$ and $\Diamond$, who are responsible for selecting the actions (i.e., resolving the non-deterministic choice) in the controllable and adversarial vertices, respectively. The selection is timeless, but performing a selected action takes time which is exponentially distributed (the parameter is the rate of a given action). When a given action is finished, the next vertex is chosen randomly according to the fixed probability distribution over the outgoing edges of the action. A *time-bounded reachability objective* is specified by a set T of target vertices and a time bound $t > 0$. The goal of player $\Box$ is to maximize the probability of reaching a target vertex before time t, while player $\Diamond$ aims at minimizing this probability.

Note that events such as component failures, user requests, message receipts, exceptions, etc., are essentially history-independent, which means that the time between two successive occurrences of such events is exponentially distributed. CTGs provide a natural and convenient formal model for systems exhibiting these features, and time-bounded reachability objectives allow to formalize basic liveness and safety properties of these systems.

Previous work. Although the practical relevance of CTGs with time-bounded reachability objectives to verification problems is obvious, to the best of our knowledge there are no previous results concerning even very basic properties of such games. A more restricted model of uniform CTMDPs is studied in [2, 7]. Intuitively, a uniform CTMDP is a CTG where all non-deterministic vertices are controlled just by one player, and all actions are assigned the same rate. In [2], it is shown that the maximal and minimal probability of reaching a target vertex before time t is efficiently computable up to an arbitrarily small given error, and that the associated strategy is also effectively computable. An open question explicitly raised in [2] is whether this result can be extended to all (not necessarily uniform) CTMDP. In [2], it is also shown that time-dependent strategies are more powerful than time-abstract ones, and this issue is addressed in greater detail in [7] where the mutual relationship between various classes of time-dependent strategies in CTMDPs is studied. Furthermore, in [1] reward-bounded objectives in CTMDPs are studied.

Our contribution is twofold. Firstly, we examine the *fundamental properties* of CTGs, where we aim at obtaining as general (and tight) results as possible. Secondly, we consider the associated *algorithmic issues*. Concrete results are discussed in the following paragraphs.

Fundamental properties of CTGs. We start by showing that each vertex $\hat{v}$ in a CTG with time-bounded reachability objectives has a *value*, i.e., an *equilibrium probability* of reaching a target vertex before time t. The value is equal to $\sup_\sigma \inf_\pi \mathcal{P}^{\sigma,\pi}_{\hat{v}}(\mathit{Reach}^{\leq t}(T))$ and $\inf_\pi \sup_\sigma \mathcal{P}^{\sigma,\pi}_{\hat{v}}(\mathit{Reach}^{\leq t}(T))$, where σ and π range over all time-abstract strategies of player $\Box$ and player $\Diamond$, and $\mathcal{P}^{\sigma,\pi}_{\hat{v}}(\mathit{Reach}^{\leq t}(T))$ is the probability of reaching T before time t in a play obtained by applying the strategies σ and π. This result holds for *arbitrary* CTGs which may have countably many vertices and actions. This immediately raises the question whether each player has an *optimal* strategy which achieves the outcome equal to or better than the value against every strategy of the opponent. We show that the answer is negative in general, but an optimal strategy for player $\Diamond$ is guaranteed to exist in *finitely-branching* CTGs, and an optimal strategy for player $\Box$ is guaranteed to exist in *finitely-branching* CTGs with *bounded rates* (see Definition 2). These results are tight, which is documented by appropriate counterexamples. Moreover, we show that in the subclasses of CTGs just mentioned, the players have also optimal CD strategies (a strategy is CD if it is deterministic and "counting", i.e., it only depends on the number of actions in the history of a play, where actions with the same rate are identified). Note that CD strategies still use infinite memory and in general they do not admit a finite description. A special attention is devoted to finite uniform CTGs, where we show a somewhat surprising result—both players have *finite memory optimal strategies* (these finite memory strategies are deterministic and their decision is based on "bounded counting" of actions; hence, we call them "BCD").

Algorithms. We show that for finite CTGs, ε-optimal strategies for both players are computable in $|V|^2 \cdot |A| \cdot bp \cdot (|\mathcal{R}| + 1)^{\mathcal{O}((\max \mathcal{R}) \cdot t + \ln \frac{1}{\varepsilon})}$ time, where $|V|$ and $|A|$ is the number of vertices and actions, resp., bp is the maximum bit-length of transition probabilities and rates (we assume that rates and the probabilities in distributions assigned to the actions are represented as fractions of integers encoded in binary), $|\mathcal{R}|$ is the number of rates, $\max \mathcal{R}$ is the maximal rate, and t is the time bound. This solves the open problem of [2] (in fact, our result is more general as it applies to finite CTGs, not just to finite CTMDPs). Actually, the algorithm works also for *infinite-state* CTGs as long as they are finitely-branching, have bounded rates, and satisfy some natural "effectivity assumptions" (see Corollary 14). For example, this is applicable to the class of infinite-state CTGs definable by pushdown automata (where the rate of a given configuration depends just on the current control state), and also to other automata-theoretic models. Finally, we show how to compute the optimal BCD strategies for both players in finite uniform CTGs.

Due to space constraints, proofs are omitted here. Full proofs can be found in [3]. In the following we just try to indicate basic ideas behind the proofs. This is not always possible, because some arguments are tricky and occasionally we also rely on relatively advanced calculations. Nevertheless, the results themselves should be easy to understand.

2 Definitions

In this paper, the sets of all positive integers, non-negative integers, rational numbers, real numbers, non-negative real numbers, and positive real numbers are denoted by $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{R}^{\geq 0}$, and $\mathbb{R}^{>0}$, respectively. Let A be a finite or countably infinite set. A *probability distribution* on A is a function $f : A \to \mathbb{R}^{\geq 0}$ such that $\sum_{a \in A} f(a) = 1$. The *support* of f is the set of all $a \in A$ where $f(a) > 0$. A distribution f is *rational* if $f(a) \in \mathbb{Q}$ for every $a \in A$, *positive* if $f(a) > 0$ for every $a \in A$, and *Dirac* if $f(a) = 1$ for some $a \in A$. The set of all distributions on A is denoted by $\mathcal{D}(A)$. A *σ-field* over a set Ω is a set $\mathcal{F} \subseteq 2^{\Omega}$ that includes Ω and is closed under complement and countable union. A *measurable space* is a pair $(\Omega, \mathcal{F})$ where Ω is a set called *sample space* and $\mathcal{F}$ is a σ-field over Ω whose elements are called *measurable sets*. A *probability measure* over a measurable space $(\Omega, \mathcal{F})$ is a function $\mathcal{P} : \mathcal{F} \to \mathbb{R}^{\geq 0}$ such that, for each countable collection $\{X_i\}_{i \in I}$ of pairwise disjoint elements of $\mathcal{F}$, $\mathcal{P}(\bigcup_{i \in I} X_i) = \sum_{i \in I} \mathcal{P}(X_i)$, and moreover $\mathcal{P}(\Omega) = 1$. A *probability space* is a triple $(\Omega, \mathcal{F}, \mathcal{P})$, where $(\Omega, \mathcal{F})$ is a measurable space and $\mathcal{P}$ is a probability measure over $(\Omega, \mathcal{F})$. Given two measurable sets $X, Y \in \mathcal{F}$ such that $\mathcal{P}(Y) > 0$, the *conditional probability* of X under the condition Y is defined as $\mathcal{P}(X \mid Y) = \mathcal{P}(X \cap Y)/\mathcal{P}(Y)$. We say that a property $A \subseteq \Omega$ holds *for almost all* elements of a measurable set Y if $\mathcal{P}(Y) > 0$, $A \cap Y \in \mathcal{F}$, and $\mathcal{P}((A \cap Y) \mid Y) = 1$.

In our next definition we introduce continuous-time Markov chains (CTMCs). The literature offers several equivalent definitions of CTMCs (see, e.g., [8]). For purposes of this paper, we adopt the variant where transitions have discrete probabilities and the rates are assigned to states.

DEFINITION 1. *A* continuous-time Markov chain (CTMC) *is a tuple $\mathcal{M} = (M, \rightarrow, Prob, \mathbf{R}, Init)$, where M is a finite or countably infinite set of* states, *$\rightarrow \subseteq M \times M$ is a* transition relation *such that every $s \in M$ has at least one outgoing transition, Prob is a function which to each $s \in M$ assigns a positive probability distribution over the set of its outgoing transitions,* $\mathbf{R}$ *is a function which to each $s \in M$ assigns a positive real* rate, *and Init is the* initial probability distribution *on M.*

We write $s \xrightarrow{x} s'$ to indicate that $s \rightarrow s'$ and $Prob(s)(s \rightarrow s') = x$. A *time-abstract path* is a finite or infinite sequence $u = u_0, u_1, \ldots$ of states such that $u_{i-1} \rightarrow u_i$ for every $1 \leq i < length(u)$, where $length(u)$ is the length of u (the length of an infinite sequence is ∞). A *timed path* (or just *path*) is a pair $w = (u, t)$, where u is a time-abstract path and $t = t_1, t_2, \ldots$ is a sequence of positive reals such that $length(t) = length(u)$. We put $length(w) = length(u)$, and for every $0 \leq i < length(w)$, we usually write $w(i)$ and $w[i]$ instead of u_i and t_i, respectively.

Infinite paths are also called *runs*. The set of all runs in $\mathcal{M}$ is denoted $Run_{\mathcal{M}}$, or just Run when $\mathcal{M}$ is clear from the context. A *template* is a pair (u, I), where $u = u_0, u_1, \ldots$ is a finite time-abstract path and $I = I_0, I_1, \ldots$ a finite sequence of non-empty intervals in $\mathbb{R}^{\geq 0}$ such that $length(u) = length(I)$. Every template (u, I) determines a *basic cylinder $Run(u, I)$* consisting of all runs w such that $w(i) = u_i$ for all $0 \leq i < length(u)$, and $w[j] \in I_j$ for all $0 \leq i < length(u) - 1$. To $\mathcal{M}$ we associate the probability space $(Run, \mathcal{F}, \mathcal{P})$ where $\mathcal{F}$ is the σ-field generated by all basic cylinders $Run(u, I)$ and $\mathcal{P} : \mathcal{F} \rightarrow \mathbb{R}^{\geq 0}$ is the unique probability measure on $\mathcal{F}$ such that

$$\mathcal{P}(Run(u, I)) = Init(u_0) \cdot \prod_{i=0}^{length(u)-2} Prob(u_i)(u_i \rightarrow u_{i+1}) \cdot \left(e^{-\mathbf{R}(u_i)\cdot \sup(I_i)} - e^{-\mathbf{R}(u_i)\cdot \inf(I_i)} \right)$$

Note that if $length(u) = 1$, the "big product" above is empty and hence equal to 1.

Now we formally define continuous-time games, which generalize continuous-time Markov chains by allowing not only probabilistic but also *non-deterministic* choice. Continuous-time games also generalize the model of continuous-time Markov decision processes studied in [2, 7] by splitting the non-deterministic vertices into two disjoint subsets of *controllable* and *adversarial* vertices, which are controlled by two "players" with antagonistic objectives. Thus, one can model the interaction between a system and its environment.

DEFINITION 2. *A* continuous-time game (CTG) *is a tuple $G = (V, A, \mathbf{E}, (V_\Box, V_\Diamond), \mathbf{P}, \mathbf{R})$ where V is a finite or countably infinite set of* vertices, *A is a finite or countably infinite set of* actions, $\mathbf{E}$ *is a function which to every $v \in V$ assigns a non-empty set of actions* enabled *in v, $(V_\Box, V_\Diamond)$ is a partition of V,* $\mathbf{P}$ *is a function which assigns to every $a \in A$ a probability distribution on V, and* $\mathbf{R}$ *is a function which assigns a positive real* rate *to every $a \in A$.*

We require that $V \cap A = \emptyset$ and use N to denote the set $V \cup A$. We say that G is finitely-branching *if for each $v \in V$ the set $\mathbf{E}(v)$ is finite (note that $\mathbf{P}(a)$ for a given $a \in A$ can still have an infinite support.) We say that G has* bounded rates *if $\sup_{a \in A} \mathbf{R}(a) < \infty$, and that G is* uniform *if* $\mathbf{R}$ *is a constant function. Finally, we say that G is* finite *if both V and A are finite.*

If $V_\Box$ or $V_\Diamond$ is empty (i.e., there is just one type of vertices), then G is a *continuous-time Markov decision process (CTMDP).* Technically, our definition of CTMDP is slightly different from the one used in [2, 7], but the difference is only cosmetic. The two models are equivalent in a well-defined sense (a detailed explanation is included in [3]). Also note that $\mathbf{P}$ and

$\mathbf{R}$ associate the probability distributions and rates directly to actions, not to pairs of $V \times A$. This is perhaps somewhat non-standard, but leads to simpler notation (since each vertex can have its "private" set of enabled actions, this is no restriction).

A *play* of G is initiated in some vertex. The non-deterministic choice is resolved by two players, $\Box$ and $\Diamond$, who select the actions in the vertices of $V_\Box$ and $V_\Diamond$, respectively. The selection itself is timeless, but some time is spent by performing the selected action (the time is exponentially distributed with the rate $\mathbf{R}(a)$), and then a transition to the next vertex is chosen randomly according to the distribution $\mathbf{P}(a)$. The players can also select the actions *randomly*, i.e., they select not just a single action but a *probability distribution* on the enabled actions. Moreover, the players are allowed to play differently when the same vertex is revisited. We assume that both players can see the history of a play, but cannot measure the elapsed time.

Let $\odot \in \{\Box, \Diamond\}$. A *strategy* for player $\odot$ is a function which to each $wv \in N^*V_\odot$ assigns a probability distribution on $\mathbf{E}(v)$. The sets of all strategies for player $\Box$ and player $\Diamond$ are denoted by Σ and Π, respectively. Each pair of strategies $(\sigma, \pi) \in \Sigma \times \Pi$ together with an initial vertex $\hat{v} \in V$ determine a unique *play* of the game G, which is a CTMC $G(\hat{v}, \sigma, \pi)$ where N^*A is the set of states, the rate of a given $wa \in N^*A$ is $\mathbf{R}(a)$ (the rate function of $G(\hat{v}, \sigma, \pi)$ is also denoted by $\mathbf{R}$), and transitions exist only between states of the form wa and $wava'$, where $wa \xrightarrow{x} wava'$ iff one of the following conditions is satisfied:

- $v \in V_\Box$, $a' \in \mathbf{E}(v)$, and $x = \mathbf{P}(a)(v) \cdot \sigma(wv)(a') > 0$
- $v \in V_\Diamond$, $a' \in \mathbf{E}(v)$, and $x = \mathbf{P}(a)(v) \cdot \pi(wv)(a') > 0$

The initial distribution is determined as follows:

- $Init(\hat{v}a) = \sigma(\hat{v})(a)$ if $\hat{v} \in V_\Box$ and $a \in \mathbf{E}(\hat{v})$;
- $Init(\hat{v}a) = \pi(\hat{v})(a)$ if $\hat{v} \in V_\Diamond$ and $a \in \mathbf{E}(\hat{v})$;
- in the other cases, *Init* returns zero.

Note that the set of states of $G(\hat{v}, \sigma, \pi)$ is infinite. Also note that all states reachable from a state $\hat{v}a$, where $Init(\hat{v}a) > 0$, are alternating sequences of vertices and actions. We say that a state w of $G(\hat{v}, \sigma, \pi)$ *hits* a vertex $v \in V$ if v is the last vertex which appears in w (for example, $v_1a_1v_2a_2$ hits v_2). Further, we say that w hits $T \subseteq V$ if w hits some vertex of T. From now on, the paths (both finite and infinite) in $G(\hat{v}, \sigma, \pi)$ are denoted by Greek letters $\alpha, \beta, \ldots$. Note that for every $\alpha \in Run_{G(\hat{v},\sigma,\pi)}$ and every $i \in \mathbb{N}_0$ we have that $\alpha(i) = wa$ where $wa \in N^*A$.

We denote by $\mathcal{R}(G)$ the set of all rates used in G (i.e., $\mathcal{R}(G) = \{\mathbf{R}(a) \mid a \in A\}$), and by $\mathcal{H}(G)$ the set of all vectors of the form $\mathbf{i} : \mathcal{R}(G) \to \mathbb{N}_0$ satisfying $\sum_{r \in \mathcal{R}(G)} \mathbf{i}(r) < \infty$. When G is clear from the context, we write just $\mathcal{R}$ and $\mathcal{H}$ instead of $\mathcal{R}(G)$ and $\mathcal{H}(G)$, respectively. For every $\mathbf{i} \in \mathcal{H}$, we put $|\mathbf{i}| = \sum_{r \in \mathcal{R}} \mathbf{i}(r)$. For every $r \in \mathcal{R}$, we denote by $\mathbf{1}_r$ the vector of $\mathcal{H}$ such that $\mathbf{1}_r(r) = 1$ and $\mathbf{1}_r(r') = 0$ if $r' \neq r$. Further, for every $wx \in N^*N$ we define the vector $\mathbf{i}_{wx} \in \mathcal{H}$ such that $\mathbf{i}_{wx}(r)$ returns the cardinality of the set $\{j \in \mathbb{N}_0 \mid 0 \leq j < length(w), w(j) \in A, \mathbf{R}(w(j)) = r\}$ (Note that the last element x of wx is disregarded.) Given $\mathbf{i} \in \mathcal{H}$ and $wx \in N^*N$, we say that wx *matches* $\mathbf{i}$ if $\mathbf{i} = \mathbf{i}_{wx}$.

We say that a strategy τ is *counting (C)* if $\tau(wv) = \tau(w'v)$ for all $w, w' \in N^*$ such that $\mathbf{i}_{wv} = \mathbf{i}_{w'v}$. In other words, a strategy τ is counting if the only information about the history of a play w which influences the decision of τ is the vector $\mathbf{i}_{wv}$. Hence, every counting strategy τ can be considered as a function from $\mathcal{H} \times V$ to $\mathcal{D}(A)$, where $\tau(\mathbf{i}, v)$ corresponds to the value of $\tau(wv)$ for every wv matching $\mathbf{i}$. A counting strategy τ is *bounded counting (BC)* if there is $k \in \mathbb{N}$ such that $\tau(wv) = \tau(w'v)$ whenever $|w|, |w'| \geq k$. A strategy τ is *deterministic (D)* if $\tau(wv)$ is a Dirac distribution for all wv. Strategies that are not necessarily counting are called *history-dependent (H)*, and strategies that are not necessarily deterministic are called *randomized (R)*. Thus, we obtain the following six types of strategies: BCD, BCR, CD, CR, HD, and HR. The most general (unrestricted) type is HR, and the importance of the other types of strategies becomes clear in subsequent sections.

In this paper, we are interested in continuous-time games with *time-bounded reachability objectives*, which are specified by a set $T \subseteq V$ of *target* vertices and a *time bound* $t \in \mathbb{R}^{>0}$. The goal of player $\Box$ is to maximize the probability of reaching a target vertex before the time bound t, while player $\Diamond$ aims at minimizing this probability. Let $\hat{v}$ be the initial vertex. Then each pair of strategies $(\sigma, \pi) \in \Sigma \times \Pi$ determines a unique *outcome* $\mathcal{P}^{\sigma,\pi}_{\hat{v}}(\mathit{Reach}^{\leq t}(T))$, which is the probability of all $\alpha \in \mathit{Run}_{G(\hat{v},\sigma,\pi)}$ that visit T before time t (i.e., there is $i \in \mathbb{N}_0$ such that $\alpha(i)$ hits T and $\sum_{i=0}^{i-1} \alpha[i] \leq t$). A fundamental question (answered in Section 3) is whether continuous-time games with time-bounded reachability objectives have a *value*, i.e., a unique equilibrium outcome. We say that $\hat{v} \in V$ *has a value* if

$$\sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{\hat{v}}(\mathit{Reach}^{\leq t}(T)) = \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \mathcal{P}^{\sigma,\pi}_{\hat{v}}(\mathit{Reach}^{\leq t}(T))$$

If $\hat{v}$ has a value, then $\mathit{val}(\hat{v})$ denotes the *value of* $\hat{v}$ defined by the above equality. Further, if $\hat{v}$ has a value, it makes sense to define *ε-optimal* and *optimal* strategies in $\hat{v}$. Let $\varepsilon \geq 0$. We say that a strategy $\sigma \in \Sigma$ is an *ε-optimal maximizing strategy in* $\hat{v}$ (or just *ε-optimal in* $\hat{v}$) if

$$\inf_{\pi\in\Pi} \mathcal{P}^{\sigma,\pi}_{\hat{v}}(\mathit{Reach}^{\leq t}(T)) \;\geq\; \mathit{val}(\hat{v}) - \varepsilon\,,$$

and that a strategy $\pi \in \Pi$ is an *ε-optimal minimizing strategy in* $\hat{v}$ (or just *ε-optimal in* $\hat{v}$) if

$$\sup_{\sigma\in\Sigma} \mathcal{P}^{\sigma,\pi}_{\hat{v}}(\mathit{Reach}^{\leq t}(T)) \;\leq\; \mathit{val}(v) + \varepsilon$$

A strategy is *optimal* in $\hat{v}$ if it is 0-optimal in $\hat{v}$, and just *optimal* if it is optimal in every $\hat{v}$.

3 The Existence of Values and Optimal Strategies

In this section we first prove that every vertex in a CTG with time-bounded reachability objectives has a value. This result holds without any additional restrictions (i.e., for CTGs with possibly countable state-space and infinite branching degree). From this we immediately obtain the existence of ε-optimal strategies for both players for every $\varepsilon > 0$. Then, we study the existence of optimal strategies. We show that even though optimal minimizing strategies may not exist in infinitely-branching CTGs, they always exist in finitely-branching ones. As for optimal maximizing strategies, we show that they do not necessarily exist even in finitely-branching CTGs, but they are guaranteed to exist if a game is both finitely-branching and has bounded rates (see Definition 2).

For the rest of this section, we fix a CTG $G = (V, A, \mathbf{E}, (V_\Box, V_\Diamond), \mathbf{P}, \mathbf{R})$, a set $T \subseteq V$ of target vertices, and a time bound $t > 0$. Given $\mathbf{i} \in \mathcal{H}$ where $|\mathbf{i}| > 0$, we denote by $F_{\mathbf{i}}$ the probability distribution function of the random variable $\sum_{r\in\mathcal{R}} \sum_{i=1}^{\mathbf{i}(r)} X_i^{(r)}$ where all $X_i^{(r)}$ are mutually independent and each $X_i^{(r)}$ is an exponentially distributed random variable with the rate r (for reader's convenience, basic properties of exponentially distributed random variables are recalled in [3]). We also define $F_{\mathbf{0}}$ as a constant function returning 1 for every argument (here $\mathbf{0} \in \mathcal{H}$ is the empty history, i.e., $|\mathbf{0}| = 0$). In the special case when $\mathcal{R}$ is a singleton, we use F_ℓ and f_ℓ to denote $F_{\mathbf{i}}$ and $f_{\mathbf{i}}$ such that $\mathbf{i}(r) = \ell$, where r is the only element of $\mathcal{R}$. Further, given $\sim\, \in \{<, \leq, =\}$ and $k \in \mathbb{N}$, we denote by $\mathcal{P}_{\hat{v}}^{\sigma,\pi}(\mathit{Reach}_{\sim k}^{\leq t}(T))$ the probability of all $\alpha \in \mathit{Run}_{G(\hat{v},\sigma,\pi)}$ that visit T for the first time in the number of steps satisfying $\sim k$ and before time t (i.e., there is $i \in \mathbb{N}_0$ such that $i = \min\{j \mid \alpha(j) \text{ hits } T\} \sim k$ and $\sum_{i=0}^{i-1} \alpha[i] \leq t$).

The following theorem says that every vertex in a CTG with bounded reachability objectives has a value. Let us note that the powerful result of Martin [6] about weak determinacy of Blackwell games cannot be applied in this setting, at least not immediately. As we shall see, the ideas presented in the proof of Theorem 3 are useful also for designing an algorithm which for a given $\varepsilon > 0$ computes ε-optimal strategies for both players.

THEOREM 3. *Every vertex $v \in V$ has a value.*

Roughly speaking, Theorem 3 is proved in the following way. Given $\sigma \in \Sigma$, $\pi \in \Pi$, $\mathbf{j} \in \mathcal{H}$, and $u \in V$, we denote by $P^{\sigma,\pi}(u, \mathbf{j})$ the probability of all runs $\alpha \in \mathit{Run}_{G(u,\sigma,\pi)}$ such that for some $n \in \mathbb{N}_0$ the state $\alpha(n)$ hits T and matches $\mathbf{j}$, and for all $0 \leq j < n$ we have that $\alpha(j)$ does not hit T. Then we introduce two functions $\mathcal{A}, \mathcal{B} : \mathcal{H} \times V \to [0,1]$ where

$$\mathcal{A}(\mathbf{i}, v) = \sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \sum_{\mathbf{j}\in\mathcal{H}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(v, \mathbf{j}) \qquad \mathcal{B}(\mathbf{i}, v) = \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \sum_{\mathbf{j}\in\mathcal{H}} F_{\mathbf{i}+\mathbf{j}}(t) \cdot P^{\sigma,\pi}(v, \mathbf{j})$$

Intuitively, $\mathcal{A}(\mathbf{i}, v)$ and $\mathcal{B}(\mathbf{i}, v)$ give the "best" probability achievable by player $\Box$ and player $\Diamond$ in a vertex v, assuming that the history of a play matches $\mathbf{i}$. Hence, it suffices to prove that $\mathcal{A} = \mathcal{B}$, because then also $\mathcal{A}(\mathbf{0}, v) = \mathcal{B}(\mathbf{0}, v) = \mathit{val}(v)$, where $\mathbf{0}$ returns zero for every argument. The equality $\mathcal{A} = \mathcal{B}$ is obtained by demonstrating that both $\mathcal{A}$ and $\mathcal{B}$ are equal to the (unique) least fixed point of a monotonic function $\mathcal{V} : (\mathcal{H} \times V \to [0,1]) \to (\mathcal{H} \times V \to [0,1])$ defined as follows: for every $H : \mathcal{H} \times V \to [0,1]$, $\mathbf{i} \in \mathcal{H}$, and $v \in V$ we have that

$$\mathcal{V}(H)(\mathbf{i}, v) = \begin{cases} F_{\mathbf{i}}(t) & v \in T \\ \sup_{a\in\mathbf{E}(v)} \sum_{u\in V} \mathbf{P}(a)(u) \cdot H(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & v \in V_\Box \setminus T \\ \inf_{a\in\mathbf{E}(v)} \sum_{u\in V} \mathbf{P}(a)(u) \cdot H(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & v \in V_\Diamond \setminus T \end{cases}$$

The details are technical and can be found in [3].

Observe that due to Theorem 3, both players have ε-optimal strategies in every vertex v (for every $\varepsilon > 0$). This follows directly from the definition of $\mathit{val}(v)$ given in Section 2. Now we examine the existence of *optimal* strategies. We start by observing that optimal minimizing and optimal maximizing strategies do not necessarily exist, even if we restrict ourselves to games with finitely many rates (i.e., $\mathcal{R}(G)$ is finite) and finitely many distinct transition probabilities.

OBSERVATION 4. *Optimal minimizing and optimal maximizing strategies in continuous-time games with time-bounded reachability objectives do not necessarily exist, even if we restrict ourselves to games with finitely many rates (i.e., $\mathcal{R}(G)$ is finite) and finitely many distinct transition probabilities.*

However, if G is finitely-branching, then the existence of an optimal minimizing CD strategy can be established by adapting the construction used in the proof of Theorem 3. Observe that we do not require that G has bounded rates.

THEOREM 5. *If G is finitely-branching, then there is an optimal minimizing CD strategy.*

The issue with optimal maximizing strategies is slightly more complicated. First, we observe that optimal maximizing strategies do not necessarily exist even in finitely-branching games.

OBSERVATION 6. *Optimal maximizing strategies in continuous-time games with time-bounded reachability objectives may not exist, even if we restrict ourselves to finitely-branching games.*

Now we show that if G is finitely-branching *and* has bounded rates, then there is an optimal maximizing CD strategy. To achieve that, we introduce the notion of *k-step optimal* strategies, which optimize the outcome in finite plays of length k. Observe that, due to Theorem 3, for all $k \in \mathbb{N}$ and $v \in V$ we have that $\sup_{\sigma\in\Sigma} \inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma,\pi}(Reach^{\leq t}_{\leq k}(T)) = \inf_{\pi\in\Pi} \sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi}(Reach^{\leq t}_{\leq k}(T))$. We use $val^k(v)$ to denote the *k-step value* defined by this equality, and we say that strategies $\sigma^k \in \Sigma$ and $\pi^k \in \Pi$ are *k-step optimal* if for all $v \in V$, $\pi \in \Pi$, and $\sigma \in \Sigma$ we have $\inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma^k,\pi}(Reach^{\leq t}_{\leq k}(T)) = \sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi^k}(Reach^{\leq t}_{\leq k}(T)) = val^k(v)$. The existence and basic properties of k-step optimal strategies are stated in our next lemma.

LEMMA 7. *If G is finitely-branching and has bounded rates, then we have the following:*

1. *For all $\varepsilon > 0$, $k \geq (\sup \mathcal{R})te^2 - \ln \varepsilon$, $\sigma \in \Sigma$, $\pi \in \Pi$, and $v \in V$ we have that*

$$\mathcal{P}_v^{\sigma,\pi}(Reach^{\leq t}(T)) - \varepsilon \;\leq\; \mathcal{P}_v^{\sigma,\pi}(Reach^{\leq t}_{\leq k}(T)) \;\leq\; \mathcal{P}_v^{\sigma,\pi}(Reach^{\leq t}(T))$$

2. *For every $k \in \mathbb{N}$, there are k-step optimal BCD strategies $\sigma^k \in \Sigma$ and $\pi^k \in \Pi$. Further, for all $\varepsilon > 0$ and $k \geq (\sup \mathcal{R})te^2 - \ln \varepsilon$ we have that every k-step optimal strategy is also an ε-optimal strategy.*

If G is finitely-branching and has bounded rates, one may be tempted to construct an optimal maximizing strategy σ by selecting those actions that are selected by infinitely many k-step optimal BCD strategies for all $k \in \mathbb{N}$ (these strategies are guaranteed to exist by Lemma 7 (2)). However, this is not so straightforward, because the distributions assigned to actions in finitely-branching games can still have an infinite support. Intuitively, this issue is overcome by considering larger and larger finite subsets of the support so that the total probability of all of the infinitely many omitted elements approaches zero. Hence, a proof of the following theorem is somewhat technical.

THEOREM 8. *If G is finitely-branching and has bounded rates, then there is an optimal maximizing CD strategy.*

3.1 Optimal Strategies in Finite Uniform CTGs

In this subsection, we restrict ourselves to finite uniform CTGs and prove that both players have *optimal BCD strategies* in such games. Roughly speaking, the result is obtained by showing that optimal CD strategies (which are guaranteed to exist by Theorem 5 and Theorem 8) can be safely redefined into *greedy* strategies after performing a finite (and effectively computable) number of steps. Greedy strategies try to maximize/minimize the probability of reaching T in as few steps as possible, and hence they can ignore the history of a play. Hence, the original optimal CD strategies become stationary after a finite number of steps, which means that they are in fact BCD. We also show that this result is tight in the sense that optimal BCD strategies do not necessarily exist in uniform CTGs with infinitely many states. In Section 4, we use these results to design an algorithm which *computes* the optimal BCD strategies in finite uniform games.

In this subsection, we assume that the previously fixed CTG G is finite and that $\mathbf{R}(a) = r > 0$ for all $a \in A$. We start by introducing greedy strategies.

DEFINITION 9. *A strategy $\sigma \in \Sigma$ is* greedily maximizing *if for all $v \in V$ and $\sigma' \in \Sigma$ one of the following two conditions is satisfied:*

- *For all $i \in \mathbb{N}_0$ we have* $\inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma,\pi}(Reach_{\leq i}^{\leq\infty}(T)) = \inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma',\pi}(Reach_{\leq i}^{\leq\infty}(T))$.
- *There is $i \in \mathbb{N}_0$ such that* $\inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma,\pi}(Reach_{\leq i}^{\leq\infty}(T)) > \inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma',\pi}(Reach_{\leq i}^{\leq\infty}(T))$ *and for all $j < i$ we have* $\inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma,\pi}(Reach_{\leq j}^{\leq\infty}(T)) = \inf_{\pi\in\Pi} \mathcal{P}_v^{\sigma',\pi}(Reach_{\leq j}^{\leq\infty}(T))$.

Similarly, $\pi \in \Pi$ is greedily minimizing *if for all $v \in V$ and $\pi' \in \Pi$ one of the following conditions holds:*

- *For all $i \in \mathbb{N}_0$ we have* $\sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi}(Reach_{\leq i}^{\leq\infty}(T)) = \sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi'}(Reach_{\leq i}^{\leq\infty}(T))$.
- *There is $i \in \mathbb{N}_0$ such that* $\sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi}(Reach_{\leq i}^{\leq\infty}(T)) < \sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi'}(Reach_{\leq i}^{\leq\infty}(T))$ *and for all $j < i$ we have* $\sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi}(Reach_{\leq j}^{\leq\infty}(T)) = \sup_{\sigma\in\Sigma} \mathcal{P}_v^{\sigma,\pi'}(Reach_{\leq j}^{\leq\infty}(T))$.

A strategy τ is stationary *if τ is deterministic and $\tau(wv)$ depends just on v for every vertex v.*

Note that time plays no role in greedily maximizing/minimizing strategies. Our next lemma reveals that greedy *stationary* strategies exist and can be effectively computed in polynomial time in finite CTGs.

LEMMA 10. *There is a greedily maximizing stationary strategy σ_g, and a greedily minimizing stationary strategy π_g. Moreover, the strategies σ_g and π_g are computable in polynomial time.*

Now we can state the main theorem of this subsection.

THEOREM 11. *Let σ_g be a greedily maximizing stationary strategy, and π_g a greedily minimizing stationary strategy. Let σ be an optimal maximizing CD strategy, and π an optimal minimizing CD strategy. Then for all sufficiently large $k \in \mathbb{N}$ we have that BCD strategies $\sigma' \in \Sigma$ and $\pi' \in \Pi$ defined by*

$$\sigma'(i,v) = \begin{cases} \sigma(i,v) & \text{if } i < k; \\ \sigma_g(v) & \text{otherwise.} \end{cases} \qquad \pi'(i,v) = \begin{cases} \pi(i,v) & \text{if } i < k; \\ \pi_g(v) & \text{otherwise.} \end{cases}$$

are optimal. Moreover, if all transition probabilities in G are rational, then σ' and π' are optimal for all $k \geq rt(1 + m^{|A|^2\cdot|V|^2})$, where m is the maximal denominator of transition probabilities.

A natural question is whether Theorem 11 can be extended to infinite-state uniform CTGs. The question is answered in our next observation.

OBSERVATION 12. *Optimal BCD strategies do not necessarily exist in uniform infinite-state CTGs, even if they are finitely-branching and use only finitely many distinct transition probabilities.*

4 Algorithms

Now we present algorithms which compute ε-optimal BCD strategies in finitely-branching CTGs with bounded rates and optimal BCD strategies in finite uniform CTGs. In this section, we assume that all rates and distributions used in the considered CTGs are *rational*.

4.1 Computing ε-optimal BCD strategies

For the rest of this subsection, let us fix a CTG $G = (V, A, \mathbf{E}, (V_\Box, V_\Diamond), \mathbf{P}, \mathbf{R})$, a set $T \subseteq V$ of target vertices, a time bound $t > 0$, and some $\varepsilon > 0$. For simplicity, let us first assume that G is finite; as we shall see, our algorithm does not really depend on this assumption, as long as the game is finitely-branching, has bounded rates, and its structure can be effectively generated (see Corollary 14). Let $k = (\max \mathcal{R})te^2 - \ln(\frac{\varepsilon}{2})$. Then, due to Lemma 7, all k-step optimal strategies are $\frac{\varepsilon}{2}$-optimal.

For every $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, and for every $v \in V$, our algorithm computes an action $C(\mathbf{i}, v) \in \mathbf{E}(v)$ which represents the choice of the constructed ε-optimal BCD strategies $\sigma_\varepsilon \in \Sigma$ and $\pi_\varepsilon \in \Pi$. That is, for every $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, and for every $v \in V_\Box$, we put $\sigma_\varepsilon(\mathbf{i}, v)(C(\mathbf{i}, v)) = 1$, and for the other arguments we define σ_ε arbitrarily so that σ_ε remains a BCD strategy. The strategy π_ε is induced by the function C in the same way.

To compute $C(\mathbf{i}, v)$, our algorithm uses a family of probabilities $R(\mathbf{i}, u)$ of reaching T from u before time t in at most $k - |\mathbf{i}|$ steps using the strategies σ_ε and π_ε and assuming that the history matches $\mathbf{i}$. Actually, our algorithm computes the probabilities $R(\mathbf{i}, u)$ only up to a sufficiently small error so that the actions chosen by C are "sufficiently optimal" (i.e., the strategies σ_ε and π_ε are ε-optimal, but they are not necessarily k-step optimal for the k chosen above). Our algorithm works in two phases:

1. For $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, compute a number $\ell_\mathbf{i}(t) > 0$ such that $\frac{|F_\mathbf{i}(t) - \ell_\mathbf{i}(t)|}{F_\mathbf{i}(t)} \leq \frac{\varepsilon^{2|\mathbf{i}|+1}}{2^{2|\mathbf{i}|+1}}$. For every $a \in A$ and $u \in V$, compute a floating point representation $\mathbf{p}(a)(u)$ of $\mathbf{P}(a)(u)$ satisfying $\frac{|\mathbf{P}(a)(u) - \mathbf{p}(a)(u)|}{\mathbf{P}(a)(u)} \leq \frac{\varepsilon^{2k+1}}{2^{2k+1}}$.
2. Compute (in a bottom up fashion) the functions R and C as follows: Given $\mathbf{i} \in \mathcal{H}$, where $|\mathbf{i}| \leq k$, and $v \in V$, we have that

$$R(\mathbf{i}, v) = \begin{cases} \ell_\mathbf{i}(t) & \text{if } v \in T \\ 0 & \text{if } v \notin T \text{ and } |\mathbf{i}| = k \\ max_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{p}(a)(u) \cdot R(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & \text{if } v \in V_\Box \setminus T \text{ and } |\mathbf{i}| < k \\ min_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{p}(a)(u) \cdot R(\mathbf{i} + \mathbf{1}_{\mathbf{R}(a)}, u) & \text{if } v \in V_\Diamond \setminus T \text{ and } |\mathbf{i}| < k \end{cases}$$

   For all $|\mathbf{i}| < k$ and $v \notin T$, we put $C(\mathbf{i}, v) = a$ where a is an action that realizes the maximum (or minimum).

In [3] we show that the strategies σ_ε and π_ε are indeed ε-optimal. Complexity analysis of the algorithm reveals the following (bp denotes the maximum bit-length of $\mathbf{P}(a)(v)$ and rates, assuming that we represent $\mathbf{P}(a)(v)$ and rates as fractions of integers encoded in binary).

THEOREM 13. *Assume that G is finite. Then for every $\varepsilon > 0$ there are ε-optimal BCD strategies $\sigma_\varepsilon \in \Sigma$ and $\pi_\varepsilon \in \Pi$ computable in time $|V|^2 \cdot |A| \cdot bp \cdot (|\mathcal{R}| + 1)^{\mathcal{O}((\max \mathcal{R}) \cdot t + \ln \frac{1}{\varepsilon})}$.*

Note that our algorithm needs to analyze only a finite part of G. Hence, it also works for infinite games which satisfy the conditions formulated in the next corollary.

COROLLARY 14. *Let G be a finitely-branching game with bounded rates and let $v \in V$. Assume that the vertices and actions of G reachable from v in a given finite number of steps are effectively computable, and that an upper bound on rates is also effectively computable. Then for every $\varepsilon > 0$ there are effectively computable BCD strategies $\sigma_\varepsilon \in \Sigma$ and $\pi_\varepsilon \in \Pi$ that are ε-optimal in v.*

4.2 Computing optimal BCD strategies in uniform finite games

For the rest of this subsection, we fix a finite uniform CTG $G = (V, A, \mathbf{E}, (V_\Box, V_\Diamond), \mathbf{P}, \mathbf{R})$ where $\mathbf{R}(a) = r > 0$ for all $a \in A$. Let $k = rt(1 + m^{|A|^2 \cdot |V|^2})$ (see Theorem 11).

The algorithm works similarly as the one of Section 4.1, but there are also some differences. Since we have just one rate, the vector $\mathbf{i}$ becomes just a number i. Similarly as in Section 4.1, our algorithm computes an action $C(i, v) \in \mathbf{E}(v)$ representing the choice of the constructed optimal BCD strategies $\sigma_{max} \in \Sigma$ and $\pi_{min} \in \Pi$. By Lemma 11, every optimal strategy can, from the k-th step on, start to behave as a fixed greedy stationary strategy, and we can compute such a greedy stationary strategy in polynomial time. Hence, the optimal BCD strategies σ_{max} and π_{min} are defined as follows:

$$\sigma_{max}(i, v) = \begin{cases} C(i, v) & \text{if } i < k; \\ \sigma_g(v) & \text{otherwise.} \end{cases} \qquad \pi_{min}(i, v) = \begin{cases} C(i, v) & \text{if } i < k; \\ \pi_g(v) & \text{otherwise.} \end{cases}$$

To compute the function C, our algorithm uses a table of symbolic representations of the (precise) probabilities $R(i, v)$ (here $i \le k$ and $v \in V$) of reaching T from v before time t in at most $k - i$ steps using the strategies σ_{max} and π_{min} and assuming that the history matches i.

The function C and the family of all $R(i, v)$ are computed (in a bottom up fashion) as follows: For all $0 \le i \le k$ and $v \in V$ we have that

$$R(i, v) = \begin{cases} F_i(t) & \text{if } v \in T \\ \sum_{j=0}^{\infty} F_{i+j}(t) \cdot \mathcal{P}_v^{\sigma_g, \pi_g}(Reach_{=j}^{\le\infty}(T)) & \text{if } v \notin T \text{ and } i = k \\ max_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot R(i+1, u) & \text{if } v \in V_\Box \setminus T \text{ and } i < k \\ min_{a \in \mathbf{E}(v)} \sum_{u \in V} \mathbf{P}(a)(u) \cdot R(i+1, u) & \text{if } v \in V_\Diamond \setminus T \text{ and } i < k \end{cases}$$

For all $i < k$ and $v \in V$, we put $C(i, v) = a$ where a is an action maximizing or minimizing $\sum_{u \in V} \mathbf{P}(a)(u) \cdot R(i+1, u)$, depending on whether $v \in V_\Box$ or $v \in V_\Diamond$, respectively. The effectivity of computing such an action (this issue is not trivial) is discussed in the proof of the following theorem.

THEOREM 15. *The BCD strategies σ_{max} and π_{min} are optimal and effectively computable.*

5 Conclusions, Future Work

We have shown that vertices in CTGs with time bounded reachability objectives have a value, and we classified the subclasses of CTGs where a given player has an optimal strategy. We also proved that in finite uniform CTGs, both players have optimal BCD strategies. Finally, we designed algorithms which compute ε-optimal BCD strategies in finitely-branching CTGs with bounded rates, and optimal BCD strategies in finite uniform CTGs.

There are several interesting directions for future research. First, we can consider more general classes of strategies that depend on the elapsed time (in our setting, strategies are time-abstract). In [2], it is demonstrated that time-dependent strategies can achieve better results than the time-abstract ones. Further, [7] shows the power of time-dependent strategies differs when the player knows only the time consumed by the last action, or the complete timed history of a play. It is not immediately clear whether Theorem 3 still holds for time-dependent strategies, and whether it makes sense to think about optimal strategies in this setting. Second, a generalization to semi-Markov processes and games, where arbitrary (not only exponential) distributions are considered, would be desirable. Another interesting open problem is the existence of optimal BCD strategies in (not necessarily uniform) games.

References

[1] C. Baier, B.R. Haverkort, H. Hermanns, and J.-P. Katoen. Reachability in continuous-time Markov reward decision processes. In E. Graedel, J. Flum, and T. Wilke, editors, *Logic and Automata: History and Perspectives*, volume 2 of Texts in Logics and Games, pages 53–72. Amsterdam University Press, 2008.

[2] C. Baier, H. Hermanns, J.-P. Katoen, and B.R. Haverkort. Efficient computation of time-bounded reachability probabilities in uniform continuous-time Markov decision processes. *TCS*, 345:2–26, 2005.

[3] T. Brázdil, V. Forejt, J. Krčál, J. Křetínský, and A. Kučera. Continuous-time stochastic games with time-bounded reachability. Technical report FIMU-RS-2009-09, Faculty of Informatics, Masaryk University, Brno, 2009.

[4] J. Filar and K. Vrieze. *Competitive Markov Decision Processes*. Springer, 1996.

[5] E. Grädel, W. Thomas, and T. Wilke. *Automata, Logics, and Infinite Games*. Number 2500 in LNCS. Springer, 2002.

[6] D.A. Martin. The determinacy of Blackwell games. *JSL*, 63(4):1565–1581, 1998.

[7] M. Neuhäußer, M. Stoelinga, and J.-P. Katoen. Delayed nondeterminism in continuous-time Markov decision processes. In *Proceedings of FoSSaCS 2009*, volume 5504 of *LNCS*, pages 364–379. Springer, 2009.

[8] J.R. Norris. *Markov Chains*. Cambridge University Press, 1998.

[9] M.L. Puterman. *Markov Decision Processes*. Wiley, 1994.

[10] W. Thomas. Infinite games and verification. In *Proceedings of CAV 2003*, volume 2725 of *LNCS*, pages 58–64. Springer, 2003.

[11] I. Walukiewicz. A landscape with games in the background. In *Proceedings of LICS 2004*, pages 356–366. IEEE, 2004.