Playing Optimally on Timed Automata with Random Delays*

Nathalie Bertrand[1,2] and Sven Schewe[2]

[1] Inria Rennes Bretagne Atlantique, France
[2] University of Liverpool, UK

Abstract. We marry continuous time Markov decision processes (CTMDPs) with stochastic timed automata into a model with joint expressive power. This extension is very natural, as the two original models already share exponentially distributed sojourn times in locations. It enriches CTMDPs with timing constraints, or symmetrically, stochastic timed automata with one conscious player. Our model maintains the existence of optimal control known for CTMDPs. This also holds for a richer model with two players, which extends continuous time Markov games. But we have to sacrifice the existence of simple schedulers: polyhedral regions are insufficient to obtain optimal control even in the single-player case.

1 Introduction

Control problems have been widely investigated in the verification community as a generalisation of the original model-checking problem. Rather than checking whether a system satisfies a property, the goal is to control the system such that it fulfils a desired property. In a framework where timing constraints are essentials, the system can be modelled using timed automata [1], and timed games have been introduced to solve the control problem [2].

Another popular model for systems with nondeterministic choices and real-time aspects is the one of continuous time Markov decision processes (CTMDPs), where the real-time aspects are governed by probability distributions, while the nondeterministic choices are resolved by a scheduler. Time-bounded reachability requires that a goal region should be reached within some time-bound, and the objective is then to build a scheduler that maximises the probability of these executions. This problem has recently received a lot of attention for CTMDPs [5, 8, 15, 18, 16, 13]. A more fundamental question than the quest for the construction or approximation of optimal schedulers is the question of their existence.

In this paper, we introduce a variant of timed games where delays are randomised rather than being nondeterministic. This model of *timed automata Markov decision processes* (TAMDPs) extends the probabilistic semantics for

* This work was supported by the Engineering and Physical Science Research Council grant EP/H046623/1 'Synthesis and Verification in Markov Game Structures' and a Leverhulme Trust Visiting Fellowship.

timed automata from [3, 4] with nondeterminism. Our model also forms an extension of CTMDPs, roughly, by adding timing constraints to the firability of transitions. We consider the time-bounded reachability problem and provide a positive answer to the fundamental question of the existence of optimal schedulers. This result immediately extends to a more general setting with two players, where controlled and adversarial nondeterminism coexist.

The structure of the optimal schedulers is, however, involved. We show that it does not suffice to consider regions or, more generally, to divide the space defined by the relevant clock values into polyhedra, to obtain optimal control.

2 Preliminaries

2.1 Timed automata

We recall here basics on timed automata, from [1], that will be useful for this paper. Timed automata are extension of finite automata with real-valued variables (called clocks) that all evolve at the same speed. Clocks can be tested and reset to 0.

For a given finite set of clocks X, a *valuation* $v : X \to \mathbb{R}_{\geq 0}$ maps every clock to a non-negative real. A *guard* over X is a finite conjunction of constraints $x \sim c$, with $\sim \in \{<, \leq, =, \geq, >\}$, for a clock $x \in X$ and an integer $c \in \mathbb{N}$. Given a guard g over X and a valuation $v \in \mathbb{R}^X_{\geq 0}$, we write $v \models g$ whenever v satisfies the constraints expressed by g, and define $[\![g]\!] = \{v \mid v \models g\}$. The set of all possible guards over X is denoted $\mathcal{G}(X)$. For $v \in \mathbb{R}^X_{\geq 0}$ a valuation and $t \in \mathbb{R}_{\geq 0}$, $v + t$ denotes the valuation defined by $v + t(x) = v(x) + t$ for every $x \in X$. Moreover, if $X' \subseteq X$ is a subset of clocks and v a valuation, $v_{[X' \leftarrow 0]}$ denotes the valuation that agrees with v on $X \setminus X'$ and is equal to 0 for all clocks in X'.

Definition 1 (Timed automaton). *A* timed automaton *is a tuple* $\mathcal{A} = (L, X, E)$ *where*

- L *is a finite set of* locations*,*
- X *is a finite set of* clocks*, and*
- $E \subseteq L \times \mathcal{G}(X) \times 2^X \times L$ *is a finite set of* edges*.*

The semantics of a timed automaton $\mathcal{A} = (L, X, E)$ is given in terms of an infinite-state transition system $\mathcal{T} = (L \times \mathbb{R}^X_{\geq 0}, \to, \mathbb{R}_{\geq 0} \times E)$, where the relation $\to$ is exactly composed of transitions $(\ell, v) \xrightarrow{t,e} (\ell', v')$ such that the edge $e = (\ell, g, X', \ell') \in E$ satisfies $v + t \models g$ and $v' = (v + t)_{[X' \leftarrow 0]}$. A *run* of $\mathcal{A}$ is a finite sequence of transitions $\rho = (\ell_0, v_0) \xrightarrow{t_0,e_0} (\ell_1, v_1) \xrightarrow{t_1,e_1} \cdots (\ell_n, v_n)$. We denote the last state (ℓ_n, v_n) of run ρ by $\mathsf{last}(\rho)$ and the value $\sum_{i=0}^{n-1} t_i$ is called the *total duration* of ρ. We write $\mathsf{Runs}(\mathcal{A})$ for the set of all runs of $\mathcal{A}$.

In order to encompass CTMDPs in our TAMDP model defined in the next subsection, we first extend timed automata with discrete probabilities. In probabilistic timed automata, introduced in [14], edges do not result in the reset of a fixed set of clocks and lead to a fixed location, but rather yield a distribution $\delta \in \mathsf{Dist}(2^X \times L)$ over resets and locations.

Definition 2 (Probabilistic timed automaton). *A* probabilistic timed automaton *is a tuple $\mathcal{A} = (L, X, E)$ where L and X are as for a timed automaton and $E \subseteq L \times \mathcal{G}(X) \times \mathsf{Dist}(2^X \times L)$ is a finite set of* probabilistic edges.

We write $(\ell, v) \xrightarrow{t,e,p} (\ell', v')$ if from state $(\ell, v+t)$ and assuming probabilistic edge e is selected, the next state is (ℓ', v') with probability p. The rest of the definitions is unchanged. In the sequel, we will consider *symbolic paths*, that is, special sets of runs in probabilistic timed automata. Given a prefix run $\rho \in \mathsf{Runs}(\mathcal{A})$, a sequence of edges $e_0, \cdots, e_n$, together with probabilisties $p_0, \cdots, p_n$, and a time-bound T, the finite symbolic path $\pi(\rho, e_0, p_0 \cdots, e_n, p_n, T)$ is defined by:

$$\pi(\rho, e_0, p_0 \cdots, e_n, p_n, T) = \{\rho \xrightarrow{t_0,e_0,p_0} (\ell_1, v_1) \xrightarrow{t_1,e_1,p_1} \cdots (\ell_n, v_n) \mid \sum_{i=0}^{n-1} t_i \leq T\}.$$

2.2 MDP model for timed automata

A probabilistic semantics for timed automata [3, 4], also referred to as stochastic timed automata, was introduced to address the problem of 'unrealistic' sets of paths, where unrealistic is identified with paths that have a very low probability (in particular 0-sets). Informally, the semantics of a stochastic timed automaton consists of an infinite-state infinitely-branching Markov chain (whose underlying graph is a timed transition system $\mathcal{T}$), where transitions between states are governed by the following: first, a delay is sampled randomly among possible delays, and second, an enabled transition is chosen randomly among enabled ones.

For technical convenience—and following [6]—we require our (probabilistic) timed automata to be reactive. A (probabilistic) timed automaton $\mathcal{A} = (L, X, E)$ is called *reactive* if, for each state $s = (\ell, v)$ of $\mathcal{A}$, there is an edge $e \in E$ such that $s \xrightarrow{0,e} s'$ for some state s' of $\mathcal{A}$. In words, we require that every state is the source of some edge, such that $\mathcal{A}$ never blocks. A (probabilistic) timed automaton can easily be made reactive by adding a self loop for all clock valuations where no guard of any transition leaving a state is satisfied.

A natural way of incorporating some control in stochastic timed automata is to have nondeterministic, rather than randomised, decisions among enabled actions. From a state $s = (\ell, v)$, first a delay t is sampled in $\mathbb{R}_{\geq 0}$, and then the controller chooses which probabilistic edge to fire from state $(\ell, v+t)$ among the possible ones.

We thus define the model of *timed automata Markov decision processes* (TAMDPs for short).

Definition 3 (Timed automaton MDP). *A* timed automaton Markov decision process *is a tuple $\mathcal{M} = (L, X, E, \Lambda)$, where $\mathcal{A} = (L, X, E)$ is a reactive probabilistic timed automaton and $\Lambda : L \to \mathbb{R}_{\geq 0}$ is a rate function.*

The semantics of a TAMDP is an infinite-state infinitely branching Markov decision process, whose states (resp. edges) are states (resp. edges) of the underlying probabilistic timed automaton $\mathcal{A}$. From a state $s = (\ell, v)$, the sojourn time in location ℓ follows an exponential distribution with rate $\Lambda(\ell)$ and some intermediary state $(\ell, v+t)$ is reached, where nondeterministically an edge $e \in E$ enabled in $(\ell, v+t)$ fires. The formal semantics of TAMDPs will be detailed further in the next subsection when defining probability measures associated with (a restricted class of well-behaved) schedulers. Note that runs of $\mathcal{M}$ coincide with runs of its underlying probabilistic timed automaton, thus we still write $\mathsf{Runs}(\mathcal{M})$ for the runs of $\mathcal{M}$. Also, in analogy to the common terminology of CTMDPs, we sometimes refer to edges (especially when an edge is selected) as *actions*.

2.3 Comparison with existing models

CTMDPs. Continuous-time Markov decision processes form a restricted class of TAMDPs where the underlying probabilistic timed automaton has no clock and is thus a finite automaton. For CTMDPs, it was proven in [16] that optimal control exists for time-bounded reachability, and that optimal schedulers can be taken from a restricted class of finitely representable schedulers.

Stochastic timed games. In stochastic timed games [7], locations are partitioned into locations owned by three players, a reachability player (who has a time-bounded reachability objective), a safety player (who has the opposite time-bounded safety objective), and an environment player (who makes random moves). In a location of the reachability or safety player, the respective player decides both the sojourn time and the edge to fire, whereas in the environment's locations, the delay as well as the edge are chosen randomly. For this model, it was shown that, assuming there is a single player and the underlying timed automaton has only one clock, the existence of a strategy for a reachability goal almost-surely (resp. with positive probability) is PTIME-complete (resp. NLOGSPACE-complete). Moreover, for two-player games, quantitative questions are undecidable.

Stochastic real-time games. In stochastic real-time games [9], states of the arena are partitionned into environment nodes —where the behaviour is similar to CTMDPs— and control nodes —where one player chooses a distribution over actions, which induces a probability distribution for the next state. For this game model, objectives are given by deterministic timed automata (DTA), and the goal for player 0 is to maximize the probability that a play satisfies the objective. The main result concerns qualitative properties, and states that if player 0 has an almost-sure winning strategy, then she has a simple one, that can be described by a DTA.

Markovian timed automata. The model closest to ours is the one of Markovian timed automata (MTA), that, similar to our TAMDPs, consist in an extension of timed automata with exponentially distributed sojourn time. MTA were first introduce as an intermediate model to model-check CTMCs or CTMDPs against deterministic timed automata specifications [10, 11]. In the recent paper [12] approximations techniques are provided for the optimal time-(un)bounded reachability probabilities in MTA. In comparison, we focus on the existence of optimal schedulers/strategies for the same problems.

2.4 Schedulers for TAMDPs

Intuitively, a scheduler is responsible for choosing which edge to fire among enabled ones after a random delay has been sampled. To make its decision, the scheduler has access to the all history of the play so far. Formally:

Definition 4 (Scheduler). *Let $\mathcal{M} = (L, X, E, \Lambda)$ be a timed automaton Markov decision process. A* scheduler *for $\mathcal{M}$ is a function $\sigma : \mathsf{Runs}(\mathcal{M}) \times \mathbb{R}_{\geq 0} \to \mathsf{Dist}(E)$ such that for every e_n with $\sigma(s_0 \xrightarrow{t_0, e_0, p_0} s_1 \cdots s_n, t_n, p_n)(e_n) > 0$, there exists a state s_{n+1} with $s_n \xrightarrow{t_n, e_n, p_n} s_{n+1}$.*

Prior to defining a meaningful class of schedulers for TAMDPs, let us first recall the notions of deterministic and memoryless schedulers. A scheduler σ for $\mathcal{M}$ is *deterministic* if it only makes pure decisions: for all $\rho \in \mathsf{Runs}(\mathcal{M})$ and all $t \in \mathbb{R}_{\geq 0}$, $\sigma(\rho, t)$ is a Dirac distribution (*i.e.*, there exists an $e \in E$ such that $\sigma(\rho, t)(e) = 1$). A scheduler σ is *memoryless*[3] if, for all $\rho, \rho' \in \mathsf{Runs}(\mathcal{M})$ with equal total duration and such that $\mathsf{last}(\rho) = \mathsf{last}(\rho')$, $\sigma(\rho, t) = \sigma(\rho', t)$ whatever the delay $t \in \mathbb{R}_{\geq 0}$.

As pointed out in [17], not all schedulers are meaningful, even in the restricted case of continuous-time Markov decision processes (CTMDPs). In particular, under some schedulers, the set of runs reaching a given location can be non-measurable. We follow the approach from [16] and consider a class of schedulers obtained as the completion of the class of cylindrical schedulers. We only report what is necessary in our context, and refer to [16] for the details of this construction.

Definition 5 (Cylindrical scheduler). *A scheduler σ for $\mathcal{M}$ and time-bound T is* cylindrical *if there exists a finite partition $\mathcal{I}$ of $[0, T]$ into intervals $I_0 = [0, T_0]$ and $I_{i+1} = (T_i, T_{i+1}]$ such that, for every pair of runs $\rho = (\ell_0, v_0) \xrightarrow{t_0, e_0, p_0} (\ell_1, v_1) \cdots (\ell_n, v_n)$ and $\rho' = (\ell_0, v_0') \xrightarrow{t_0', e_0, p_0} (\ell_1, v_1') \cdots (\ell_n, v_n')$ and every pair of delays $(t_n, t_n') \in \mathbb{R}_{\geq 0}$, as soon as, for all $0 \leq j \leq n$, t_j and t_j' belong to the same interval I_j, then $\sigma(\rho, t_n) = \sigma(\rho', t_n')$.*

In plain English, a scheduler is cylindrical for a partition $\mathcal{I}$ of $[0, T]$ if it takes the same decision for runs and delays that are equivalent with respect to $\mathcal{I}$.

The set of cylindrical schedulers can then be extended to *measurable* schedulers by defining a metric on cylindrical schedulers and then taking the limits of

[3] Note that our notion of memoryless scheduler is looser than the usual one: the scheduler can also base its decision on the elapsed time so far. This particularity is due to the kind of properties we consider, namely time-bounded reachability.

Cauchy-sequences of cylindrical schedulers with respect to that metric (see [16] for details).

Given a TAMDP $\mathcal{M}$, any measurable scheduler σ yields a probability measure, denoted $\mathbb{P}_\sigma$, over $\mathsf{Runs}(\mathcal{M})$ with a fixed initial state, or more generally a fixed initial run. Let us define $\mathbb{P}_\sigma$ over $\mathsf{Runs}(\mathcal{M}, \rho)$ initiated by a finite prefix $\rho \in \mathsf{Runs}(\mathcal{M})$, by first associating a measure with every finite symbolic path $\pi = \pi(\rho, e_0, p_0 \cdots e_n, p_n, T)$. For every time-bound $T \geq 0$, $\mathbb{P}_\sigma(\rho, T) = 1$ and inductively for $\pi = \pi(\rho, e_0, p_0 \cdots e_n, p_n, T)$ with $\rho \in \mathsf{Runs}(\mathcal{M})$ ending in location ℓ_0, scheduler σ assigns the following probability

$$\mathbb{P}_\sigma(\pi) = \int_{t=0}^{T} \sigma(\rho, t)(e_0) \cdot p_0 \cdot \mathbb{P}_\sigma(\pi(\rho_1, e_1, p_1 \cdots e_n, p_n, T - t)) \cdot \Lambda(\ell_0) \cdot e^{-\Lambda(\ell_0)t} \, \mathrm{d}t,$$

where $\rho_1 = \rho \xrightarrow{t, e_0, p_0} s_1$. Mapping $\mathbb{P}_\sigma$ can then be extended in a unique way into a probability measure over $\mathsf{Runs}(\mathcal{M}, \rho)$ equipped with the σ-algebra generated by symbolic paths starting with ρ.

Time-bounded reachability probability. In this paper, we are interested in time-bounded reachability probabilities. Let us introduce the time-bounded reachability probability problem. Given a TAMDP $\mathcal{M}$, an initial state (ℓ, v), a set of goal locations $G \subseteq L$, and a time-bound T, let $\mathsf{Reach}_{\mathcal{M}}(\ell, v, G, T)$ denote the set of runs of $\mathcal{M}$ that originate (ℓ, v) and reach the goal within T time-units:

$$\mathsf{Reach}_{\mathcal{M}}(\ell, v, G, T) = \{(\ell, v) = (\ell_0, v_0) \xrightarrow{t_0, e_0, p_0} (\ell_1, v_1) \cdots (\ell_n, v_n) \in \mathsf{Runs}(\mathcal{M}) \mid \exists i \leq n,\ \ell_i \in G \text{ and } \sum_{j<i} t_j \leq T\}.$$

Note that one can easily express $\mathsf{Reach}_{\mathcal{M}}(\ell, v, G, T)$ as a countable union of symbolic paths starting in (ℓ, v), and it is thus legal to consider its probability under measurable schedulers. The maximum time-bounded reachability probability problem consists in maximising the probability of $\mathsf{Reach}_{\mathcal{M}}(\ell, v, G, T)$ among measurable schedulers, and we write

$$\mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T) = \sup_{\sigma} \ \mathbb{P}_\sigma(\mathsf{Reach}_{\mathcal{M}}(\ell, v, G, T)).$$

A natural question is whether optimal schedulers exist *at all*, that is, whether the supremum of the time-bounded reachability probability, $\mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T)$, is taken for some scheduler. If this is the case, it is worth knowing if simple (e.g., cylindrical, region-based, or, more generally, polyhedral) optimal schedulers exist. In the remainder of the paper, we establish the existence of optimal schedulers for the time-bounded reachability probability problem for TAMDPs, and show that polyhedral schedulers are not sufficient.

3 Optimal schedulers for TAMDPs

In this section, we establish the existence of optimal schedulers for the maximum time-bounded reachability probability problem in timed automata Markov decision processes. In order to do so, we start by providing lower bounds for $\mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T)$ by allowing, in addition to the time-bound, only a fixed number of steps to reach the goal, and then show that these lower bounds are sharp.

We consider the optimal probability to reach the goal G from (ℓ, v) within T time-units, with the additional constraint that it should be in no more than N discrete steps. This probability, which we denote $\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T)$, optimises the probability of the following set of runs:

$$\mathsf{Reach}^N(\ell, v, G, T) = \{(\ell, v) = (\ell_0, v_0) \xrightarrow{t_0, e_0, p_0} (\ell_1, v_1) \cdots (\ell_N, v_N) \in \mathsf{Runs}(\mathcal{M}) \mid \exists i \le N,\ \ell_i \in G \text{ and } \sum_{j<i} t_j \le T\}.$$

That is, we have $\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T) = \sup_\sigma\ \mathbb{P}_\sigma(\mathsf{Reach}^N_{\mathcal{M}}(\ell, v, G, T))$.

For all $N \in \mathbb{N}$, $\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T)$ is obviously a lower bound for $\mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T)$. Moreover, $\Big(\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T)\Big)_{N \in \mathbb{N}}$ is non-decreasing, and, as we shall see, the sequence converges to the ordinary optimum $\mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T)$. In order to prove this, we start with the following simple lemma:

Lemma 1. *For all $\varepsilon > 0$ there exists an $M \in \mathbb{N}$ such that, for all $N \ge M$ and all measurable schedulers σ, $\mathbb{P}_\sigma\Big(\mathsf{Reach}(\ell, v, G, T) \smallsetminus \mathsf{Reach}^N(\ell, v, G, T)\Big) < \varepsilon$ and $\mathbb{P}_\sigma\Big(\mathsf{Reach}(\ell, v, G, T)\Big) - \mathbb{P}_\sigma\Big(\mathsf{Reach}^N(\ell, v, G, T)\Big) < \varepsilon$.*

Proof. Let $\lambda = \max_{\ell \in L} \Lambda(\ell)$ be the maximal transition rate in $\mathcal{M}$. Given $\varepsilon > 0$, we choose M such that $\sum_{k=M}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} < \varepsilon$. Under any measurable scheduler and assuming all locations have rate λ, the number of steps taken within T time-units is Poisson distributed, and the likelihood to perform M or more steps is bounded from above by $\sum_{k=M}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$, and therefore smaller than ε. Of course, this upper bound also applies to the case where rates are smaller or equal λ in all locations. The set of all runs in $\mathcal{M}$ performing M or more steps obviously contains $\mathsf{Reach}(\ell, v, G, T) \smallsetminus \mathsf{Reach}^N(\ell, v, G, T)$ for every $N \ge M$, so we conclude that $\mathbb{P}_\sigma\Big(\mathsf{Reach}(\ell, v, G, T) \smallsetminus \mathsf{Reach}^N(\ell, v, G, T)\Big) < \varepsilon$.

The second claim follows from $\mathbb{P}_\sigma\Big(\mathsf{Reach}(\ell, v, G, T)\Big) - \mathbb{P}_\sigma\Big(\mathsf{Reach}^N(\ell, v, G, T)\Big) = \mathbb{P}_\sigma\Big(\mathsf{Reach}(\ell, v, G, T) \smallsetminus \mathsf{Reach}^N(\ell, v, G, T)\Big)$. □

We can now establish that $\Big(\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T)\Big)_{N \in \mathbb{N}}$ converges to the optimum:

Lemma 2. $\lim_{N \to \infty} \mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T) = \mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T)$.

Proof. The '$\le$' direction is simple: for all schedulers σ and all $N \in \mathbb{N}$, $\mathbb{P}_\sigma\Big(\mathsf{Reach}^N(\ell, v, G, T)\Big) \le \mathbb{P}_\sigma\Big(\mathsf{Reach}(\ell, v, G, T)\Big)$ trivially holds, and consequently $\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T) \le \mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T)$.

For the '$\ge$' direction, we show that, for all $\varepsilon > 0$, $\lim_{N \to \infty} \mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T) \ge \mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T) - 2\varepsilon$. Let $\varepsilon > 0$. On one hand, we can always choose a scheduler σ such that $\mathbb{P}_\sigma\Big(\mathsf{Reach}(\ell, v, G, T)\Big) > \mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T) - \varepsilon$. On the other hand, applying Lemma 1, there exists $M \in \mathbb{N}$ such that, for every $N \geq M$, $\mathbb{P}_\sigma\Big(\mathsf{Reach}(\ell, v, G, T)\Big) - \mathbb{P}_\sigma\Big(\mathsf{Reach}^N(\ell, v, G, T)\Big) < \varepsilon$. As a consequence, for N large enough, $\mathbb{P}_\sigma\Big(\mathsf{Reach}^N(\ell, v, G, T)\Big) > \mathbb{P}_\sigma\Big(\mathsf{Reach}(\ell, v, G, T)\Big) - \varepsilon > \mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T) - 2\varepsilon$, and thus $\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T) > \mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T) - 2\varepsilon$. □

From now on, we focus on the under-approximation $\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T)$, in order to prove the existence of optimal schedulers for $\mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T)$.

Lemma 3. *For every state (ℓ, v) in $\mathcal{M}$, the sequence $\Big(\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T)\Big)_{N\in\mathbb{N}}$ is characterized inductively by:*

$$\mathsf{Opt}^0_{\mathcal{M}}(\ell, v, G, T) = 0 \textit{ if } \ell \notin G, \tag{1}$$

$$\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T) = 1 \textit{ if } \ell \in G \textit{ for all } N \in \mathbb{N}\textit{, and otherwise} \tag{2}$$

$$\mathsf{Opt}^{N+1}_{\mathcal{M}}(\ell, v, G, T) = \int_0^T \max_{e\in E} \sum_{(\ell,v)\xrightarrow{t,e,p}(\ell',v')} p \cdot \mathsf{Opt}^N_{\mathcal{M}}(\ell', v', G, T-t) \cdot \Lambda(\ell) \cdot e^{-\Lambda(\ell)t} \mathrm{d}t. \tag{3}$$

Equation (3), stating that optimality is memoryless, is the only non obvious one.

Proof. The correctness of Equation (3) can be shown by a simple inductive proof over N. The base case, for $N = 0$, clearly holds since all the schedulers return the same probability.

For the induction step, assume the equation holds up to N. Then, for $N + 1$ step-bounded reachability, the scheduler has to make a decision what action to choose from (ℓ, v) if a discrete action occurs after delay t. By induction hypothesis, this is to optimise the outcome in case of having $T - t$ time-units and N steps left. □

Lemma 4. $\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T) \in [0, 1]$, $\mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T) \in [0, 1]$*, and* $\mathsf{Opt}^N_{\mathcal{M}}(\ell, v + t, G, T - t)$ *and* $\mathsf{Opt}_{\mathcal{M}}(\ell, v + t, G, T - t)$ *are uniformly continuous in t and v.*

Proof. First, it is easy to see that $\mathsf{Opt}^N_{\mathcal{M}}(\ell, v, G, T) \in [0, 1]$, for all parameters. Taking the limit when N tends to infinity, this also holds for $\mathsf{Opt}_{\mathcal{M}}(\ell, v, G, T)$.

Let n be the number of clocks and $\|\cdot\|$ be any norm on valuations[4]. We now prove by induction on N that $\mathsf{Opt}^N_{\mathcal{M}}(\ell, v+t, G, T-t)$ is uniformly continuous in t and v. Obviously, $\mathsf{Opt}^0_{\mathcal{M}}(\ell, v + t, G, T - t)$ is constant (for fixed ℓ and G) and thus uniformly continuous in v and t.

Let us show the uniform continuity of $\mathsf{Opt}^N_{\mathcal{M}}(\ell, v + t, G, T - t)$ in t for all $N \in \mathbb{N}$. Assume $|t - t'| < \varepsilon$, and, w.l.o.g., $t < t'$. Observe that

$$\begin{aligned}
&\mathsf{Opt}^{N+1}_{\mathcal{M}}(\ell, v + t, G, T - t) \\
&\quad= \int_0^{T-t} \max_{e\in E} \sum_{(\ell,v+t)\xrightarrow{\tau,e,p}(\ell',v')} p \cdot \mathsf{Opt}^N_{\mathcal{M}}(\ell', v', G, T - t - \tau) \cdot \Lambda(\ell) \cdot e^{-\Lambda(\ell)\tau} \mathrm{d}\tau \\
&\quad= \int_t^{T} \max_{e\in E} \sum_{(\ell,v)\xrightarrow{\tau,e,p}(\ell',v')} p \cdot \mathsf{Opt}^N_{\mathcal{M}}(\ell', v', G, T - \tau) \Lambda(\ell) e^{-\Lambda(\ell)(\tau-t)} \mathrm{d}\tau\,.
\end{aligned}$$

[4] Recall that all norms over $\mathbb{R}^n$ are equivalent, so the choice of $\|\cdot\|$ is arbitrary.

Thus

$$\begin{aligned}
&\Big|\mathsf{Opt}^{N+1}_{\mathcal{M}}(\ell, v+t, G, T-t) - \mathsf{Opt}^{N+1}_{\mathcal{M}}(\ell, v+t', G, T-t')\Big| \\
&= \Big| \int_t^T \max_{e \in E} \sum_{(\ell,v)\xrightarrow{\tau,e,p}(\ell',v')} p \cdot \mathsf{Opt}^{N}_{\mathcal{M}}(\ell', v', G, T-\tau) \Lambda(\ell) e^{-\Lambda(\ell)(\tau-t)} \mathrm{d}\tau \\
&\qquad - \int_{t'}^T \max_{e \in E} \sum_{(\ell,v)\xrightarrow{\tau,e,p}(\ell',v')} p \cdot \mathsf{Opt}^{N}_{\mathcal{M}}(\ell', v', G, T-\tau) \Lambda(\ell) e^{-\Lambda(\ell)(\tau-t')} \mathrm{d}\tau \Big| \\
&\le \int_t^{t'} \max_{e \in E} \sum_{(\ell,v)\xrightarrow{\tau,e,p}(\ell',v')} p \cdot \mathsf{Opt}^{N}_{\mathcal{M}}(\ell', v', G, T-\tau) \Lambda(\ell) e^{-\Lambda(\ell)(\tau-t)} \mathrm{d}\tau \\
&\qquad + \int_{t'}^T \max_{e \in E} \sum_{(\ell,v)\xrightarrow{\tau,e,p}(\ell',v')} p \cdot \mathsf{Opt}^{N}_{\mathcal{M}}(\ell', v', G, T-\tau) \Lambda(\ell) e^{-\Lambda(\ell)(\tau-t)} \big|1 - e^{-\Lambda(\ell)(t-t')}\big| \mathrm{d}\tau \\
&\le \int_t^{t'} \Lambda(\ell) e^{-\Lambda(\ell)(\tau-t)} \mathrm{d}\tau + \big|1 - e^{-\Lambda(\ell)(t-t')}\big| \int_{t'}^T \Lambda(\ell) e^{-\Lambda(\ell)(\tau-t)} \mathrm{d}\tau \\
&= (1 - e^{-\Lambda(\ell)(t'-t)}) + \big|1 - e^{-\Lambda(\ell)(t-t')}\big| (e^{-\Lambda(\ell)(t'-t)} - e^{-\Lambda(\ell)(T-t)}) \\
&\le (1 - e^{-\Lambda(\ell)(t'-t)}) + \big|1 - e^{-\Lambda(\ell)(t-t')}\big| = e^{-\Lambda(\ell)(t-t')} - e^{-\Lambda(\ell)(t'-t)} \\
&\le e^{\Lambda(\ell)\varepsilon} - e^{-\Lambda(\ell)\varepsilon},
\end{aligned}$$

and we can conclude the uniform continuity of $\mathsf{Opt}^{N+1}_{\mathcal{M}}(\ell, v+t, G, T-t)$ in t.

For the uniform continuity in v, we start with the induction hypothesis

$$\begin{aligned}
\forall \ell \in L\ \forall G \subseteq L\ \forall \varepsilon > 0\ \exists \delta > 0\ \forall v, w \in \mathbb{R}^n_{\ge 0}.\ \|v - w\| < \delta \\
\Rightarrow \Big|\mathsf{Opt}^{N}_{\mathcal{M}}(\ell, v, G, T) - \mathsf{Opt}^{N}_{\mathcal{M}}(\ell, w, G, T)\Big| < \varepsilon.
\end{aligned}$$

With such ε and δ, and letting $\lambda = \max\{\Lambda(\ell) \mid \ell \in L\}$, we will show that if $\|v - w\| < \varepsilon$ then $|\mathsf{Opt}^{N+1}_{\mathcal{M}}(\ell, v, G, T) - \mathsf{Opt}^{N+1}_{\mathcal{M}}(\ell, w, G, T)| < \lambda(T\varepsilon + n\lceil T\rceil\delta)$. This will be sufficient to establish the induction step by taking δ' for a given ε in the same way as choosing δ for $\lambda(n+1)\lceil T\rceil\varepsilon$. In order to do so, let us define

$$I_v^w = \Big\{\tau \in [0, T] \mid v + \tau \text{ and } w + \tau \text{ do } \textit{not} \text{ satisfy the same guards}\Big\}.$$

Observe that, provided $\|v - w\| < \delta$, the 'length' $\int_{\tau \in I_v^w} \mathrm{d}\tau$ of I_v^w is bounded by $n\lceil T\rceil\delta$. Indeed, $v + t$ and $w + t$ can only have different enabled transitions if for one of the clocks x the two valuations disagree on $x < c$ (for $c \in \mathbb{N}$), and the interval over which they differ is bounded in length by $\|v - w\|$. The number of such intervals is itself bounded by $\lceil T\rceil$ for clock x. Finally we obtain the bound $n\lceil T\rceil\delta$ when considering the n clocks.

Let us now turn to the induction step. First we assume without loss of generality that $\delta < \varepsilon$. (Obviously, δ can always be chosen this way.) We then get:

$$
\begin{aligned}
&\left|\mathsf{Opt}^{N+1}_{\mathcal{M}}(\ell,v,G,T)-\mathsf{Opt}^{N+1}_{\mathcal{M}}(\ell,w,G,T)\right| = \big|\int_{t=0}^{T}\Lambda(\ell)e^{-\Lambda(\ell)t}\\
&\quad\Big(\max_{e\in E}\sum_{(\ell,v)\xrightarrow{t,e,p}(\ell',v')} p\;\mathsf{Opt}^{N}_{\mathcal{M}}(\ell',v',G,T{-}t) - \max_{e\in E}\sum_{(\ell,w)\xrightarrow{t,e,p}(\ell',w')} p\;\mathsf{Opt}^{N}_{\mathcal{M}}(\ell',w',G,T-t)\Big)\mathrm{d}t\big|\\
&\le \lambda\int_{t=0}^{T}\big|\max_{e\in E}\sum_{(\ell,v)\xrightarrow{t,e,p}(\ell',v')} p\;\mathsf{Opt}^{N}_{\mathcal{M}}(\ell',v',G,T-t)\\
&\qquad\qquad - \max_{e\in E}\sum_{(\ell,w)\xrightarrow{t,e,p}(\ell',w')} p\;\mathsf{Opt}^{N}_{\mathcal{M}}(\ell',w',G,T-t)\big|\mathrm{d}t\\
&\le \lambda\Big(\int_{[0,T]\setminus I^w_v}\max_{e\in E}\sum_{(\ell,v)\xrightarrow{t,e,p}(\ell',v')} p\;\big|\mathsf{Opt}^{N}_{\mathcal{M}}(\ell',v',G,T-t)-\mathsf{Opt}^{N}_{\mathcal{M}}(\ell',w',G,T-t)\big|\mathrm{d}t\\
&\qquad\qquad + \int_{I^w_v}\mathrm{d}t\Big) \qquad\qquad\qquad\qquad (\mathsf{Opt}^{N}_{\mathcal{M}} \text{ is in } [0,1])\\
&\le \lambda\Big(\int_{[0,T]\setminus I^w_v}\varepsilon\mathrm{d}t + \int_{I^w_v}\mathrm{d}t\Big) \qquad\qquad\quad (\text{induction hypothesis})\\
&\le \lambda\big(T\varepsilon + n\lceil T\rceil\delta\big) \le \lambda(n+1)\lceil T\rceil\varepsilon
\end{aligned}
$$

This ends the proof that $\mathsf{Opt}^{N}_{\mathcal{M}}$ is uniformly continuous in v —as the constant multiplicative factor $\lambda(n+1)\lceil T\rceil$ does not matter— for all $N \in \mathbb{N}$.

Finally, we exploit Lemma 2 to show that these properties of uniform continuity in t and v are inherited by the limit $\mathsf{Opt}_{\mathcal{M}}$. This can be shown using simple triangle inequalities. To establish

$$
\begin{aligned}
&\forall\ell\in L\ \forall G\subseteq L\ \forall\varepsilon>0\ \exists\delta>0\ \forall v,w\in\mathbb{R}^n_{\geq 0}.\ \|v-w\|<\delta\\
&\qquad\qquad\qquad\qquad \Rightarrow \Big|\mathsf{Opt}_{\mathcal{M}}(\ell,v,G,T)-\mathsf{Opt}^{N}_{\mathcal{M}}(\ell,w,G,T)\Big| < 3\varepsilon,
\end{aligned}
$$

we first fix an $N \in \mathbb{N}$ such that $\|\mathsf{Opt}_{\mathcal{M}} - \mathsf{Opt}^{N}_{\mathcal{M}}\| \leq \varepsilon$. Then we spend one ε for $|\mathsf{Opt}^{N}_{\mathcal{M}}(\ell,v,G,T) - \mathsf{Opt}^{N}_{\mathcal{M}}(\ell,w,G,T)|$, because $\mathsf{Opt}^{N}_{\mathcal{M}}$ is uniformly continuous in the valuation, and one ε each for $|\mathsf{Opt}_{\mathcal{M}}(\ell,v,G,T) - \mathsf{Opt}^{N}_{\mathcal{M}}(\ell,v,G,T)|$ and $|\mathsf{Opt}_{\mathcal{M}}(\ell,w,G,T) - \mathsf{Opt}^{N}_{\mathcal{M}}(\ell,w,G,T)|$. □

We can now prove our main theorem using a topological argument that extends the argument from [16] to the more general case of TAMDPs.

Theorem 1. *For every TAMDP $\mathcal{M}$, with initial state $(\ell_0, 0^X)$, reachability objective G and time-bound T, there exists a measurable scheduler σ such that*

$$\mathbb{P}_\sigma(\mathsf{Reach}_{\mathcal{M}}(\ell_0, 0^X, G, T)) = \mathsf{Opt}_{\mathcal{M}}(\ell_0, 0^X, G, T).$$

Proof. As a consequence of the continuity of $\mathsf{Opt}_{\mathcal{M}}(\ell, v+t, G, T-t)$ in v and t, we describe in the following an abstract construction of a *measurable* scheduler σ that chooses, for all $v \in [0,T]^X$ and $t, T' \in [0,T]$, an action e that determines the transition $(\ell, v) \xrightarrow{t,e} (\ell', v')$ that maximises $\mathsf{Opt}_{\mathcal{M}}(\ell', v', G, T'-t)$.

For positions outside of $[0,T]^X \times [0,T] \times [0,T]$, the behaviour of the scheduler does not matter: σ can therefore be fixed to any constant decision for all of these clock valuations and times.

We fix a location ℓ for the rest of the proof, and write Range for $[0,T]^X \times [0,T] \times [0,T]$ the range of triples (v,t,T') we will consider. In order to determine optimal decisions, we start with fixing an arbitrary order $\succ$ on the actions available in ℓ. Next, we let, for each clock valuation $v \in [0,T]^X$, delay $t \in [0,T]$ and remaining time $T' \in [0,T]$ and action e, $\mathsf{val}(v,t,T',e) = \sum_{(\ell,v)\xrightarrow{t,e,p}(\ell',v')} p \cdot \mathsf{Opt}_{\mathcal{M}}(\ell', v', G, T'-t)$, provided e is enabled in $(\ell, v+t)$, otherwise we use $-\infty$. Last we introduce, for all (v,t,T'), an additional order $\sqsupset_{v,t,T'}$ on the actions, determined by $\mathsf{val}(\ell, v+t, T, e)$ and using $\succ$ as a tie-breaker.

We now define the following sets for every action e:

- $M_e = \{(v,t,T') \in \mathsf{Range} \mid e$ is maximal w.r.t. $\sqsupset_{v,t,T'}\}$ is the set of triplets (v,t,T') for which e is maximal with respect to the order $\sqsupset_{v,t,T'}$.
- $C_e = \{(v,t,T') \in \mathsf{Range} \mid \forall \delta > 0\ \exists (v',t',T'') \in M_e.\ \|(v,t,T') - (v',t',T'')\| < \delta\}$ is the closure of M_e, and
- $D_e = C_e \smallsetminus \bigcup\limits_{f \succ e} C_f$ is the set of triplets for which action e is $\sqsupset_{v,t,T'}$-better than all other actions and there is no $\succ$-better action with equal quality.

We define σ as the memoryless scheduler σ such that when the last state is (ℓ, v), the delay is t, and the remaining time is T', σ selects action e such that $(v,t,T') \in D_e$. To complete the proof, let us show that σ is (1) optimal and (2) measurable.

To show the first point, we observe that the decision e is optimal in M_e by definition. The fact that e is also optimal in the larger set C_e is a consequence of the continuity of $\mathsf{Opt}_{\mathcal{M}}$ in v and t. $D_e \subseteq C_e$ then implies that e is an optimal decision for all triplets contained in D_e. Note that optimality among the pure decisions entails optimality among mixed ones, as the value of mixed decisions is the convex combination of the values for the respective pure decisions.

To show the second point, we observe that the M_e's partition Range by their definition, because $\sqsupset_{v,t,T'}$ is a total order. Consequently, the C_e's cover Range, and the D_e's again partition Range. The C_e's are closed subsets of Range, and therefore measurable. By their definition, the D_e's inherit this measurability.

Our construction therefore provides us with a measurable scheduler, which is optimal, deterministic, and memoryless. □

4 Extensions

In this section we consider three potential extensions: the extension to games, the extension to time-unbounded reachability, and the strengthening of the results to schedulers with a simple finite structure. We show in Subsection 4.1 that the first of these extensions is possible: our results extend to a generalisation to two-player games. But the results from Theorem 1 do neither extend to time-unbounded reachability (Subsection 4.2), nor can they be strengthened to a simple class of schedulers, whose decisions are only based on polyhedral regions (Subsection 4.3).

4.1 Extension to games

So far, we considered time-bounded reachability objectives for TAMDPs, which can be seen as a stochastic one-player game, that is, a game with a single player interacting with a randomised environment. Let us now discuss how to extend our results to stochastic two-player games by considering timed automata Markov games (TAMGs for short).

Definition 6 (Timed automaton Markov game). *A* timed automaton Markov game *is a tuple $\mathcal{G} = (\mathcal{A}, L_0, L_1, \Lambda)$ where $\mathcal{A} = (L, X, E)$ is a reactive probabilistic timed automaton $L = L_0 \sqcup L_1$ is a partition of the set of locations and $\Lambda : L \to \mathbb{R}_{\geq 0}$ is a rate function.*

Naturally, Player 0 owns states with location in L_0 and Player 1 is responsible for the decisions in states with location in L_1. The semantics of a TAMG is a stochastic two-player game. Informally, from a state (ℓ, v) with $\ell \in L_i$ (for $i \in \{0, 1\}$), the sojourn time in location ℓ follows an exponential distribution with rate $\Lambda(\ell)$ and Player i chooses in the intermediate state $\ell, v + t$ which enabled edge to fire. The resolution of nondeterministic choices by the players is governed by strategies. Similarly to TAMDPs, for which we introduced cylindrical and measurable schedulers, we consider here cylindrical and measurable strategies for each of the players. We write σ (resp. τ) for a measurable strategy of Player 0 (resp. Player 1). Any strategy profile (σ, τ) for $\mathcal{G}$ with σ and τ measurable strategies induces a probability measure $\mathbb{P}_{\sigma,\tau}$ over $\mathsf{Runs}(\mathcal{G}) = \mathsf{Runs}(\mathcal{A})$ (assuming an initial state is fixed).

The objective of Player 0 is to maximise the probability to reach a set of goal locations $G \subseteq L$ within time T. The optimum is thus defined as:

$$\mathsf{Opt}_{\mathcal{G}}(\ell, v, G, T) = \sup_{\sigma} \inf_{\tau} \mathbb{P}_{\sigma,\tau}(\mathsf{Reach}_{\mathcal{G}}(\ell, v, G, T)).$$

As announced earlier, the result established for TAMDPs carries over to TAMGs:

Theorem 2. *For every TAMG $\mathcal{G}$ with initial state $(\ell_0, 0^X)$, reachability objective G for Player 0 and time-bound T, there exists a measurable strategy profile (σ, τ) such that*

$$\mathbb{P}_{\sigma,\tau}(\mathsf{Reach}_{\mathcal{G}}(\ell_0, 0^X, G, T)) = \mathsf{Opt}_{\mathcal{G}}(\ell_0, 0^X, G, T) =$$
$$\sup_{\sigma'} \mathbb{P}_{\sigma',\tau}(\mathsf{Reach}_{\mathcal{G}}(\ell_0, 0^X, G, T)) = \inf_{\tau'} \mathbb{P}_{\sigma,\tau'}(\mathsf{Reach}_{\mathcal{G}}(\ell_0, 0^X, G, T)).$$

In order to extend the proof, we proceed in two steps. The first step is the extension of the lemmata from Section 3. This extension is simple: following the same structure, it suffices to replace, in the lemmata and their proofs, the max by min for all locations of Player 1.

Consequently, we obtain a set of equations that describe the value of the time-bounded reachability probability. We can then proceed with fixing optimal measurable strategies for σ only and τ only, respectively, such that their decisions is locally optimal. Note that the proof of Theorem 1 treats the different locations independently, so that a restriction to a subset of locations does not affect the proof at all. The proof is also not affected by swapping max for min for the locations owned by Player 1.

4.2 Time-unbounded reachability

We considered optimisation problem for time-bounded reachability, and justify now, a posteriori, why the time-bound is crucial for Theorem 1. Indeed, we show that optimal scheduling policies may not exist for *time-unbounded* reachability objectives. The situation thus ressembles the framework of stochastic real-time games [9] for which it was shown that optimal strategies do not always exist, using a similar example. To exemplify this, we consider the TAMDP $\mathcal{M}$ depicted on Figure 1 with constant transition rate $\Lambda = 1$ and the objective to reach the goal region G. We argue that this control objective does not admit an optimal scheduling policy.

It is easy to see that the chances of reaching G from ℓ_1 are 0 if the value of the clock x is greater or equal to 1, and $e^{-\varepsilon} - e^{-1}$ for a clock value $\varepsilon \in [0, 1]$. This implies an upper bound on the time-unbounded reachability of $1 - e^{-1}$. This value can easily be approximated by choosing a scheduling policy that guarantees a time-unbounded reachability $> 1 - e^{-1} - \varepsilon$ by progressing to ℓ_1 iff the clock value of x is smaller than ε. (Almost surely such a value is eventually taken.)

While this determines the *value* of time-unbounded reachability, it does not provide a scheduling policy that realises this value. If we consider a scheduling policy that, for any $\varepsilon \in]0, 1]$, provides a positive probability p_ε to progress to ℓ_1 with a clock valuation $\geq \varepsilon$, then the likelihood of reaching G is bounded by $1 - e^{-1} - p_\varepsilon(1 - e^{-\varepsilon})$. At the same time, this chance being 0 for all $\varepsilon > 0$ implies that we almost surely never progress to ℓ_1. (Progressing with clock valuation 0 can only happen on a 0 set.)

Consequently, no optimal scheduling policy exists.

4.3 Simple schedulers

Beyond the existence of optimal schedulers, the simplicity of schedulers is also a concern. In the proof of Theorem 1 we show the existence of an optimal scheduler which is measurable, deterministic and memoryless. It is an interesting question whether the optimum can be reached by even simpler schedulers. For CTMDPs, e.g., timed positional schedulers (whose decisions only depend on the location and the time that remains) and even cylindrical schedulers (that have only finitely many intervals of constant decisions) are sufficient [16]. Timed positional schedulers are clearly not sufficient for TAMDPs. (Consider a scheduler that makes the decision of whether to progress to ℓ_1 in the example from Figure 1, with 0 to 0.5 time-units left and the clock valuation of x (a) less than 0.5 and (b) greater than 1. For (a), the optimal decision is clearly to progress, for (b), the optimal decision is clearly to stay in ℓ_0 and reset the clock. A scheduler that does not distinguish these cases cannot be optimal.) Still the question remains if considering simple regions of clock valuations suffices. A natural generalisation of finitely many intervals would be finitely many *polyhedra* that are distinguished by a scheduler.

Definition 7. *Let $\mathcal{M}$ be a TAMDP over $\mathcal{A}$ a timed automaton with N clocks. A scheduler σ for $\mathcal{M}$ is* polyhedral *if there exists a finite partition $\mathcal{P}$ of $(\mathbb{R}_{\geq 0})^{N+1}$ into polyhedra $P_1 \cdots P_k$ such that, for every pair of runs $\rho = (\ell_0, v_0) \xrightarrow{t_0,e_0,p_0} (\ell_1, v_1) \cdots (\ell_n, v_n)$ and $\rho' = (\ell_0, v_0') \xrightarrow{t_0',e_0',p_0'} (\ell_1', v_1') \cdots (\ell_m', v_m')$ and every pair of delays $(t_n, t_m') \in \mathbb{R}_{\geq 0}$, as soon as $\ell_n = \ell_m'$ and $(v_n + t_n, \sum_{i \leq n} t_i)$ and $(v_m' + t_m', \sum_{i \leq m} t_i')$ belong to the same polyhedron P_j, then $\sigma(\rho, t_n) = \sigma(\rho', t_m')$.*

Note that polyhedral schedulers are in particular memoryless (and timed positional in the special case of CTMDPs). Polyhedral schedulers are natural in the context of timed automata since they extend region-based schedulers that are for example sufficient for timed games [2].

Proposition 1. *In TAMDPs, the optimal time-bounded reachability probability may not be taken by any polyhedral scheduler.*

Proof. To prove that polyhedral schedulers are not sufficient to obtain optimal control in TAMDPs, we consider again the example of Figure 1, where the rate is constant $\Lambda = 1$, the goal location is G and the time-bound is set to 1.

The only non-trivial decision the scheduler has to make in that example is in location ℓ_0, where it has to choose between looping back to ℓ_0 (action *loop* in the sequel) or moving right to ℓ_1 (action *progress* in the sequel). We are interested in determining a partition (D_l, D_p) of $(\mathbb{R}_{\geq 0_+})^2$ representing sets of valuation for x and remaining time t such that *loop* is optimal in D_l and *progress* is optimal in D_p. We focus on the sub-region $[0,1]^2$ to show that neither D_l nor D_p can be composed of finite unions of polyhedra, and consequently optimal schedulers cannot be polyhedral for this example. For this sub-region, we start with the following observations:

1. If $t \leq 1 - x$, then it is always advisable to select *progress*. The time-bounded reachability probability in this case is $1 - e^{-t}$.
2. If $t \geq 1 - x$ and the selected action is to *progress*, then the time-bounded reachability probability is $1 - e^{x-1}$.
3. If the selected action is to *loop*, then the time-bounded reachability probability is $1 - (t+1)e^{-t}$. (When looping, x is reset. After the reset of x, the guard of the edge from ℓ_1 to G is always satisfied in the remaining $t \leq 1$

time-units. The chance of reaching G is thus the probability of taking two or more steps in the remaining t time-units, and the number of such steps is Poisson distributed with parameter t.)

Consequently, we loop in $[0,1]^2$ iff $(t+1)e^{-t} \leq e^{x-1}$ (modulo 0 sets). Obviously, this set is not representable by a finite union of polyhedra.

The construction of the partition (D_l, D_p) illustrates the proof of Theorem 1, where a measurable optimal scheduler is defined. The partition (D_p, D_l) intersected with $[0,1]^2$ is depicted on Figure 2. The area D_p, below the curve, represents pairs (t, x) of remaining time and clock valuation, for which progressing to ℓ_1 is the optimal decision. □

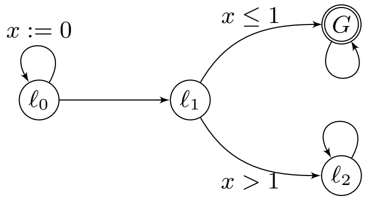

Fig. 1. A simple TAMDP example

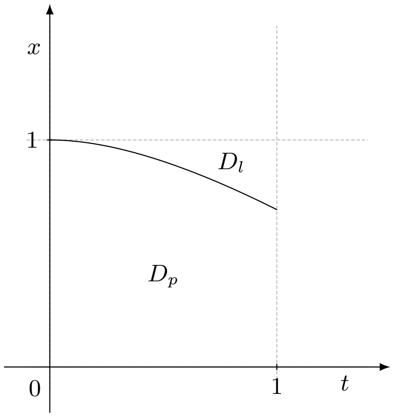

Fig. 2. Illustration of partition $(D_e)_{e \in E}$.

5 Conclusion

We have introduced the model of timed automata Markov games that synthesises stochastic timed automata and continuous time Markov games: TAMGs enhance stochastic timed automata with two conscious players and add timing constraints to the firing of actions in continuous time Markov games. We have proven the existence of measurable strategies that optimise the probability of time-bounded reachability properties. Different to CTMGs, optimal strategies are not necessarily simple: they cannot be represented by finite families of polyhedral regions. Also, in contrast to the positive result for time-bounded reachability, we have shown that optimal scheduling policies for time-unbounded reachability do not always exist.

References

1. R. Alur and D. L. Dill. A Theory of Timed Automata. *Theoretical Computer Science*, 126(2):183–235, 1994.
2. E. Asarin, O. Maler, A. Pnueli, and J. Sifakis. Controller Synthesis for Timed Automata. In *Proc. of SCC'98*, pages 469–474. Elsevier, 1998.
3. C. Baier, N. Bertrand, P. Bouyer, Th. Brihaye, and M. Größer. Probabilistic and Topological Semantics for Timed Automata. In *Proc. of FSTTCS'07*, LNCS 4855, pages 179–191. Springer, 2007.
4. C. Baier, N. Bertrand, P. Bouyer, Th. Brihaye, and M. Größer. Almost-Sure Model Checking of Infinite Paths in One-Clock Timed Automata. In *Proc. of LICS'08*, pages 217–226. IEEE, 2008.
5. 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. *Theoretical Computer Science*, 345(1):2–26, 2005.
6. P. Bouyer, Th. Brihaye, M. Jurdziński, and Q. Menet. Almost-Sure Model-Checking of Reactive Timed Automata. In *Proc. of QEST'12*. IEEE, To appear.
7. P. Bouyer and V. Forejt. Reachability in Stochastic Timed Games. In *Proc. of ICALP'09*, LNCS 5556, pages 103–114. Springer, 2009.
8. T. Brázdil, V. Forejt, J. Krcál, J. Kretínský, and A. Kucera. Continuous-Time Stochastic Games with Time-Bounded Reachability. In *Proc. of FSTTCS'09*, LIPIcs 4, pages 61–72. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2009.
9. T. Brázdil, J. Krcál, J. Kretínský, A. Kucera, and V. Rehák. Stochastic Real-Time Games with Qualitative Timed Automata Objectives. In *Proc. of CONCUR'10*, LNCS 6269, pages 207–221. Springer, 2010.
10. T. Chen, T. Han, J.-P. Katoen, and A. Mereacre. Model Checking of Continuous-Time Markov Chains Against Timed Automata Specifications. *Logical Methods in Computer Science*, 7(1:12):1–34, 2011.
11. T. Chen, T. Han, J.-P. Katoen, and A. Mereacre. Observing Continuous-Time MDPs by 1-clock Timed Automata. In *Proc. of RP'11*, LNCS 6945, pages 2–25. Springer, 2011.
12. T. Chen, T. Han, J.-P. Katoen, and A. Mereacre. Reachability Probabilities in Markovian Timed Automata. In *Proc. of CDC-ECC'11*, pages 7075–7080. IEEE, 2011.
13. J. Fearnley, M. N. Rabe, S. Schewe, and L. Zhang. Efficient Approximation of Optimal Control for Continuous-Time Markov Games. In *Proc. of FSTTCS'11*, LIPIcs 13, pages 399–410. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2011.
14. M. Z. Kwiatkowska, G. Norman, R. Segala, and J. Sproston. Automatic Verification of Real-Time Systems with Discrete Probability Distributions. *Theoretical Computer Science*, 282(1):101–150, 2002.
15. M. R. Neuhäußer and L. Zhang. Time-Bounded Reachability Probabilities in Continuous-Time Markov Decision Processes. In *Proc. of QEST'10*, pages 209–218. IEEE, 2010.
16. M. N. Rabe and S. Schewe. Finite Optimal Control for Time-Bounded Reachability in CTMDPs and Continuous-Time Markov Games. *Acta Informatica*, 48(5-6):291–315, 2011.
17. N. Wolovick and S. Johr. A Characterization of Meaningful Schedulers for Continuous-Time Markov Decision Processes. In *Proc. of FORMATS'06*, LNCS 4202, pages 352–367. Springer, 2006.
18. L. Zhang and M. R. Neuhäußer. Model Checking Interactive Markov Chains. In *Proc. of TACAS'10*, LNCS 6015, pages 53–68. Springer, 2010.