Playing Stochastically in Weighted Timed Games to Emulate Memory

Benjamin Monmege ✉ 🆔
Aix Marseille Univ, Université de Toulon, CNRS, LIS, France

Julie Parreaux ✉
Aix Marseille Univ, Université de Toulon, CNRS, LIS, France

Pierre-Alain Reynier ✉
Aix Marseille Univ, Université de Toulon, CNRS, LIS, France

—— Abstract ——————————————————————————————————

Weighted timed games are two-player zero-sum games played in a timed automaton equipped with integer weights. We consider optimal reachability objectives, in which one of the players, that we call Min, wants to reach a target location while minimising the cumulated weight. While knowing if Min has a strategy to guarantee a value lower than a given threshold is known to be undecidable (with two or more clocks), several conditions, one of them being the divergence, have been given to recover decidability. In such weighted timed games (like in untimed weighted games in the presence of negative weights), Min may need finite memory to play (close to) optimally. This is thus tempting to try to emulate this finite memory with other strategic capabilities. In this work, we allow the players to use stochastic decisions, both in the choice of transitions and of timing delays. We give for the first time a definition of the expected value in weighted timed games, overcoming several theoretical challenges. We then show that, in divergent weighted timed games, the stochastic value is indeed equal to the classical (deterministic) value, thus proving that Min can guarantee the same value while only using stochastic choices, and no memory.

1 Introduction

Real-time aspects are often inherent in the behaviour of critical software systems. *Timed automata* [2] extend finite-state automata with timing constraints, providing an automata-theoretic framework to model and verify real-time systems. While this has lead to the development of mature verification tools, the design of programs verifying some real-time specifications remains a notoriously difficult problem. One way to avoid the need to a posteriori debugging is to automatise the process as much as possible. To do so, the situation is modelled into a *timed game*, played by a *controller* and an antagonistic *environment*: they act, in a turn-based fashion, over a timed automaton. A simple, yet realistic, objective for the controller is to reach a target location. We are thus looking for a *strategy* of the controller, that is a recipe dictating how to play so that the target is reached no matter how the environment plays. Reachability timed games are decidable [4], and EXPTIME-complete [19].

For many applications, this qualitative setting is often too coarse to model faithfully the system. This motivated a shift to a quantitative setting, based on weighted extensions of the models considered so far. Weighted extensions of these timed automata and games have thus been considered in order to measure the quality of the winning strategy for the controller [11, 1]: when the controller has several winning strategies, the quantitative version of the game helps choosing a good one with respect to some metrics. More precisely, the controller, which we now call player Min, wants to reach the target while *minimising* the cumulated weight. The model we consider, called *weighted timed game* (WTG for short), is defined as follows: the game takes place over a weighted (or priced) timed automaton [5, 3], where locations are split among the two players, transitions are equipped with weights, and locations with rates of weights (the cost is then proportional to the time spent in this location, with the rate as proportional coefficient). In this setting, the possibility to use negative weights on transitions and locations is crucial when one wants to model energy or other resources that can grow or decrease during the execution of the system under study.

While solving the optimal reachability problem on weighted timed automata has been shown to be PSPACE-complete [8] (i.e. the same complexity as the non-weighted version), WTGs are known to be undecidable [13]. Many restrictions have then been considered in order to regain decidability, the first and most interesting one being the class of strictly non-Zeno cost with only non-negative weights (in transitions and locations) [11]: this hypothesis requires that every execution of the timed automaton that follows a cycle of the region automaton has a weight far from 0 (in interval $[1, +\infty)$, for instance). This setting has been extended in the presence of negative weights in transitions and locations [16]: in the so-called *divergent WTGs*, each execution that follows a cycle of the region automaton has a weight in $(-\infty, -1] \cup [1, +\infty)$. A triply-exponential-time algorithm allows one to compute the values and almost-optimal strategies, while deciding the divergence of a WTG is PSPACE-complete.

When studying optimal reachability objectives with both positive and negative weights, it is known that strategies of player Min require memory to play optimally (see [15] for the case of finite games). More precisely, the memory needed is pseudo-polynomial (i.e. polynomial if constants are encoded in unary). For WTGs, the memory needed even becomes exponential. An important challenge is thus to find ways to avoid using such complex strategies, e.g. by proposing alternative classes of strategies that are more easily amenable to implementation.

Strategies considered so far are deterministic. Though the game has no stochastic edges, it is possible to allow players to use stochastic strategies. This approach has been recently studied in the setting of finite games [20], where it is shown that memory may indeed be emulated using randomness in finite reachability games with integer weights. More precisely, the minimal value Min can achieve using memoryless stochastic strategies is the same as the value achievable using deterministic strategies. In the present work, we lift the results obtained in [20] for finite games to the timed setting.

A first important challenge is to analyse how to play stochastically in WTGs. To our knowledge, this has not been studied before. Starting from a notion of stochastic behaviours in a timed automaton considered in [7] (for the one-player setting), we propose a new class of stochastic strategies. Compared with [7], our class is larger in the sense that we allow Dirac distributions for delays, which subsumes the setting of deterministic strategies. However, in order to ensure that strategies yield a well-defined probability distribution on sets of executions, we need measurability properties stronger than the one considered in [7] (we actually provide an example showing that their hypothesis was not strong enough).

Then, we turn our attention towards the expected cumulated weight of the set of plays conforming to a pair of stochastic strategies. We first prove that under the previous measurability hypotheses, this expectation is well-defined when restricting to the set of plays
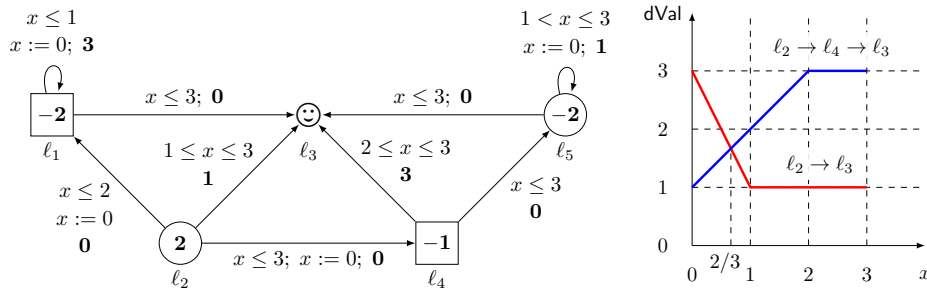
Figure 1 On the left, a weighted timed game. Locations belonging to Min (resp. Max) are depicted by circles (resp. squares). The target location is ℓ_3. Location ℓ_1 (resp. ℓ_5) has (deterministic) value $+\infty$ (resp. $-\infty$). As a consequence, the value in ℓ_4 is determined by the edge to ℓ_3, and depicted in blue on the right. In location ℓ_2, the value associated with the transition to ℓ_3 is depicted in red, and the deterministic value in ℓ_2 is obtained as the minimum of these two curves.

following a finite sequence of transitions. In order to have the convergence of the global expectation, we identify another property of strategies of Min, which intuitively ensures that the set of target locations is reached quickly enough. This allows us to define a notion of stochastic value (resp. memoryless stochastic value) of the game, i.e. the best value Min can achieve using stochastic strategies (resp. memoryless stochastic strategies), when Max uses stochastic strategies (resp. memoryless stochastic strategies) too.

In a second step, we aim at adapting the proof techniques of [20] from finite to infinite games. It is well-known that the classical region abstraction of timed automata is not suited to analyse WTGs (there are cases in which one has to split regions). In order to obtain positive results, we focus on the class of divergent WTGs. We prove that the notion of optimal deterministic switching strategy, which was central in the approach of [20], can be adapted to divergent WTGs. Our main result is then to show that for these games, the two versions of stochastic values are equal to the deterministic value. In other terms, we show that Min can emulate memory using randomisation, and vice versa. Moreover, combining memory and randomisation does not increase Min's capabilities. Due to the lack of space, detailed proofs of all results can be found in the long version [21].

2 Weighted timed games

We let $\mathcal{C}$ be a finite set of variables called clocks. A valuation is a mapping $\nu \colon \mathcal{C} \to \mathbb{R}_{\geq 0}$. For a valuation ν, a delay $t \in \mathbb{R}_{\geq 0}$ and a subset $Y \subseteq \mathcal{C}$ of clocks, we define the valuation $\nu + t$ as $(\nu + t)(x) = \nu(x) + t$, for all $x \in \mathcal{C}$, and the valuation $\nu[Y := 0]$ as $(\nu[Y := 0])(x) = 0$ if $x \in Y$, and $(\nu[Y := 0])(x) = \nu(x)$ otherwise. The valuation $\mathbf{0}$ assigns 0 to every clock. A (non-diagonal) guard on clocks of $\mathcal{C}$ is a conjunction of atomic constraints of the form $x \bowtie c$, where $\bowtie \in \{\leq, <, =, >, \geq\}$ and $c \in \mathbb{N}$. A valuation $\nu \colon \mathcal{C} \to \mathbb{R}_{\geq 0}$ satisfies an atomic constraint $x \bowtie c$ if $\nu(x) \bowtie c$. The satisfaction relation is extended to all guards g naturally, and denoted by $\nu \models g$. We let $\mathsf{Guards}(\mathcal{C})$ denote the set of guards over $\mathcal{C}$.

▶ **Definition 1.** *A weighted timed game (WTG) is a tuple* $\mathcal{G} = \langle L_{\mathsf{Min}}, L_{\mathsf{Max}}, L_T, \Delta, \mathsf{wt} \rangle$ *where* L_{Min}, L_{Max}, L_T *are finite disjoint subsets of* Min *locations,* Max *locations, and target locations, respectively (we let* $L = L_{\mathsf{Min}} \uplus L_{\mathsf{Max}} \uplus L_T$*),* $\Delta \subseteq L \times \mathsf{Guards}(\mathcal{C}) \times 2^{\mathcal{C}} \times L$ *is a finite set of transitions,* $\mathsf{wt} \colon \Delta \uplus L \to \mathbb{Z}$ *is the weight function.*

Without loss of generality, we suppose the absence of deadlocks except on target locations, i.e. for each location $\ell \in L \backslash L_T$ and valuation ν, there exists $(\ell, g, Y, \ell') \in \Delta$ such that $\nu \models g$, and no transitions start in L_T. The semantics of a WTG $\mathcal{G}$ is defined in terms of a game played on an infinite transition system whose vertices are configurations of the WTG. A configuration is a pair (ℓ, ν) with a location and a valuation of the clocks. Configurations are split into players according to the location. A configuration is final if its location is a target location of L_T. The alphabet of the transition system is given by $\Delta \times \mathbb{R}_{\geq 0}$: a pair (δ, t) encodes the delay t that a player wants to spend in the current location, before firing transition δ. For every delay $t \in \mathbb{R}_{\geq 0}$, transition $\delta = (\ell, g, Y, \ell') \in \Delta$ and valuation ν, there is an edge $(\ell, \nu) \xrightarrow{\delta, t} (\ell', \nu')$ if $\nu + t \models g$ and $\nu' = (\nu + t)[Y := 0]$. The weight of such an edge e is given by $t \times \mathsf{wt}(\ell) + \mathsf{wt}(\delta)$. An example is depicted on Figure 1.

A *finite play* is a finite sequence of consecutive edges $\rho = (\ell_0, \nu_0) \xrightarrow{\delta_0, t_0} (\ell_1, \nu_1) \xrightarrow{\delta_1, t_1} \cdots (\ell_k, \nu_k)$. We sometimes denote such a play $(\ell_0, \nu_0) \xrightarrow{(\delta_0, t_0) \cdots (\delta_{k-1}, t_{k-1})}$, since intermediate locations and valuations are uniquely defined by the initial configuration and the sequence of transitions and delays. We denote by $|\rho|$ the length k of ρ. The concatenation of two finite plays ρ_1 and ρ_2, such that ρ_1 ends in the same configuration as ρ_2 starts, is denoted by $\rho_1 \rho_2$. We denote by $I(\rho, \delta)$ the interval of delays t such that the play ρ can be extended with the edge $\xrightarrow{\delta, t}$. We let FPlays be the set of all finite plays, whereas $\mathsf{FPlays_{Min}}$ (resp. $\mathsf{FPlays_{Max}}$) denote the finite plays that end in a configuration of Min (resp. Max). A *play* is then a maximal sequence of consecutive edges (it is either infinite or it reaches L_T).

We call *path* a finite or infinite sequence π of transitions of $\mathcal{G}$. Each play ρ of $\mathcal{G}$ is associated with a unique path π (by projecting away everything but the transitions): we say that ρ *follows* the path π. A *target path* is a finite path ending in the target set L_T. We denote by TPaths the set of target paths. We let TPaths_ρ (resp. TPaths_ρ^n) the subset of target paths that start from the last location of the finite play ρ (resp. containing n transitions). A path is said to be *maximal* if it is infinite or if it is a target path.

A *deterministic strategy* for Min (resp. Max) is a mapping $\sigma \colon \mathsf{FPlays_{Min}} \to \Delta \times \mathbb{R}_{\geq 0}$ (resp. $\tau \colon \mathsf{FPlays_{Max}} \to \Delta \times \mathbb{R}_{\geq 0}$) such that for all finite plays $\rho \in \mathsf{FPlays_{Min}}$ (resp. $\rho \in \mathsf{FPlays_{Max}}$) ending in non-target configuration (ℓ, ν), there exists an edge $(\ell, \nu) \xrightarrow{\sigma(\rho)} (\ell', \nu')$. We let $\mathsf{dStrat_{Min}}$ and $\mathsf{dStrat_{Max}}$ denote the set of deterministic strategies in $\mathcal{G}$ for players Min and Max, respectively. A play or finite play $\rho = (\ell_0, \nu_0) \xrightarrow{\delta_0, t_0} (\ell_1, \nu_1) \xrightarrow{\delta_1, t_1} \cdots$ conforms to a deterministic strategy σ of Min (resp. Max) if for all k such that (ℓ_k, ν_k) belongs to Min (resp. Max), we have that $(\delta_k, t_k) = \sigma((\ell_0, \nu_0) \xrightarrow{\delta_0, t_0} \cdots (\ell_k, \nu_k))$. For all deterministic strategies σ and τ of players Min and Max, respectively, and for all configurations (ℓ_0, ν_0), we let $\mathsf{Play}((\ell_0, \nu_0), \sigma, \tau)$ be the outcome of σ and τ, defined as the unique maximal play conforming to σ and τ and starting in (ℓ_0, ν_0).

The objective of Min is to reach a target configuration, while minimising the cumulated weight up to the target. Hence, we associate to every finite play $\rho = (\ell_0, \nu_0) \xrightarrow{\delta_0, t_0} (\ell_1, \nu_1) \xrightarrow{\delta_1, t_1} \cdots (\ell_k, \nu_k)$ its cumulated weight, taking into account both discrete and continuous costs: $\mathsf{wt}_\Sigma(\rho) = \sum_{i=0}^{k-1} [t_i \times \mathsf{wt}(\ell_i) + \mathsf{wt}(\delta_i)]$. Then, the weight of a maximal play ρ, denoted by $\mathsf{wt}(\rho)$, is defined by $+\infty$ if ρ is infinite (does not reach L_T), and $\mathsf{wt}_\Sigma(\rho)$ if it ends in (ℓ_T, ν) with $\ell_T \in L_T$.

A deterministic strategy $\sigma \in \mathsf{dStrat_{Min}}$ guarantees a certain value, against all possible strategies of the opponent: for all locations ℓ and valuations ν, we let $\mathsf{dVal}_{\ell, \nu}^\sigma = \sup_{\tau \in \mathsf{dStrat_{Max}}, \mathcal{G}} \mathsf{wt}(\mathsf{Play}((\ell, \nu), \sigma, \tau))$. Then, for all locations ℓ and valuations ν, we let $\mathsf{dVal}_{\ell, \nu}$ be the *deterministic value* of $\mathcal{G}$ in (ℓ, ν), defined as $\mathsf{dVal}_{\ell, \nu} = \inf_{\sigma \in \mathsf{dStrat_{Min}}} \mathsf{dVal}_{\ell, \nu}^\sigma$. We say that a deterministic strategy σ of Min is *ε-optimal wrt the deterministic value* if $\mathsf{dVal}_{\ell, \nu}^\sigma \leq \mathsf{dVal}_{\ell, \nu} + \varepsilon$ for all (ℓ, ν). It is said optimal if this holds for $\varepsilon = 0$.

As usual in related work [1, 11, 12, 16], we assume that all clocks are *bounded* by a constant $M \in \mathbb{N}$, i.e. every transition of the WTG is equipped with a guard g such that $\nu \models g$ implies $\nu(x) \leq M$ for all clocks $x \in \mathcal{C}$. We denote by $w_{\max}^L$ (resp. $w_{\max}^\Delta$, $w_{\max}^e$) the maximal weight in absolute values of locations (resp. of transitions, edges) of $\mathcal{G}$, i.e. $w_{\max}^L = \max_{\ell \in L} |\mathsf{wt}(\ell)|$ (resp. $w_{\max}^\Delta = \max_{\delta \in \Delta} |\mathsf{wt}(\delta)|$, $w_{\max}^e = M w_{\max}^L + w_{\max}^\Delta$).

In the following, we rely on the crucial notion of regions, as introduced in the seminal work on timed automata [2]. A game $\mathcal{G}$ can be populated with the region information, without loss of generality, as described formally in [16], e.g. The *region automaton, or region game*, $\mathcal{R}(\mathcal{G})$ is thus the WTG with locations $S = L \times \mathsf{Reg}(\mathcal{C}, M)$ and all transitions $((\ell, r), g'', Y, (\ell', r'))$ with $(\ell, g, Y, \ell') \in \Delta$ such that the model of guard g'' (i.e. all valuations ν such that $\nu \models g''$) is a region r'', time successor of r such that r'' satisfies the guard g, and r' is the region obtained from r'' by resetting all clocks of Y. Distribution of locations to players, final locations, and weights are inherited from $\mathcal{G}$. We call *region path* a finite or infinite sequence of transitions in this automaton, and we again denote by π such paths. A play ρ in $\mathcal{G}$ is projected on a region path π, with a similar definition as the projection on paths: we again say that ρ *follows* the region path π. It is important to notice that, even if π is a *cycle* (i.e. starts and ends in the same location of the region game), there may exist plays following it in $\mathcal{G}$ that are not cycles, due to the fact that regions are sets of valuations.

As shown in previous work [11, 16], knowing whether $\mathsf{dVal}_{\ell,\nu} = +\infty$ for a certain configuration is a purely qualitative problem that can be decided easily by using the region game: indeed, $\mathsf{dVal}_{\ell,\nu} = +\infty$ if and only if Min has no strategies that guarantee reaching the target L_T. This is thus a reachability objective, where weights are useless. Moreover, Max has a strategy that guarantees that no plays reach the target L_T from any configuration (ℓ, ν) such that $\mathsf{dVal}_{\ell,\nu} = +\infty$. In this situation, considering stochastic choices is not interesting. **We thus rule out this case by supposing in the following that no configurations of $\mathcal{G}$ have a value $+\infty$**: such configurations can be removed in the region game by strengthening the guard on transitions.

3   Playing stochastically in WTGs

Our first contribution consists in allowing both players to use stochastic choices in their strategies. From a game theory point of view, this seems natural. From a controller synthesis point of view, we claim that the question is natural too, especially because player Min may require exponential memory to play optimally in WTGs. This is already the case even without clocks (such games are then sometimes called *shortest-path games*) where it has been shown in [20] that the memory required by Min could be traded for stochastic choices instead (and vice versa). We aim at extending this result in the context of weighted timed games. Before doing so, we must introduce stochastic strategies in the context of weighted timed games, which has never been explored until now, as far as we are aware of. We will however strongly rely on a recent line of works aiming at studying stochastic timed automata [7, 9, 6, 10], thus extending the results in the context of two-player games (instead of model-checking) and with weights, which indeed represents the main challenge in order to give a meaning to the expected payoff.

Naturally, deterministic strategies for Min are extended to more general stochastic strategies as mappings $\eta \colon \mathsf{FPlays}_{\mathsf{Min}} \to \mathsf{Dist}(\Delta \times \mathbb{R}_{\geq 0})$ where each finite play is associated to a probability distribution over the set of pairs of transition and delay. Here, we let $\mathsf{Dist}(S)$ the set of all probability distributions over a set S (equipped with an underlying σ-algebra). Since Δ is a finite set, this is equivalent to letting first Min choose a transition

via $\eta_\Delta \colon \mathsf{FPlays}_{\mathsf{Min}} \to \mathsf{Dist}(\Delta)$, and then, knowing the chosen transition, choose a delay via $\eta_{\mathbb{R}^+} \colon \mathsf{FPlays}_{\mathsf{Min}} \times \Delta \to \mathsf{Dist}(\mathbb{R}_{\geq 0})$, the support of the distribution $\eta_{\mathbb{R}^+}(\rho, \delta)$ being included in the interval $I(\rho, \delta)$ of valid delays. We can then recombine η_Δ and $\eta_{\mathbb{R}^+}$ to obtain the distribution $\eta(\rho)$. Similar definitions hold for Max whose general strategies are denoted by θ.

Notice that deterministic strategies are a special case of strategies, where the distributions are chosen to be Dirac distributions. Another useful restriction over strategies is the non-use of memory: a strategy η is *memoryless* if for all finite plays ρ, ρ' ending in the same configuration, we have that $\eta(\rho) = \eta(\rho')$. A similar definition holds for Max.

Probability measure on plays. We fix two strategies η and θ for both players, and an initial configuration (ℓ_0, ν_0). Our goal is to define a probability measure on plays. To do so, and following the contribution of [7] for stochastic timed automata, the set of plays of a WTG $\mathcal{G}$ starting from (ℓ_0, ν_0) and conforming to η and θ can naturally be equipped with a structure of σ-algebra whose generators are all subsets of plays that start with a finite prefix following the same finite path π (remember that paths are sequences of transitions, with no information on the delayed time) with some Borel-measurable constraints on the delays taken along π. The a priori idea is thus to define a probability measure $\mathbb{P}^{\eta, \theta}_{\ell_0, \nu_0}$ on such generators which extends uniquely as a probability measure over the whole σ-algebra, by Carathéodory's extension theorem.

Consider thus a finite path π, starting in location ℓ, and a play ρ ending in the same location ℓ. We define the probability $\mathbb{P}^{\eta, \theta}_\rho(\pi)$ taking into account all possible plays that start with ρ and continue according to π (we leave the Borel-measurable constraints on the delays for now, but discuss them later). It is defined by induction on the length of π by $\mathbb{P}^{\eta, \theta}_\rho(\varepsilon) = 1$, and for all transitions $\delta = (\ell, g, Y, \ell') \in \Delta$, $\mathbb{P}^{\eta, \theta}_\rho(\delta\pi) = \int_{I(\rho, \delta)} \eta_\Delta(\rho)(\delta) \times \mathbb{P}^{\eta, \theta}_{\rho \xrightarrow{\delta, t}}(\pi) \, \mathrm{d}\eta_{\mathbb{R}^+}(\rho, \delta)(t)$. This definition is very similar to the one in [7] except that we choose to decouple the distribution on pairs of $\Delta \times \mathbb{R}_{\geq 0}$ by first selecting a transition and then delay, whereas authors of [7] consider independent choices, the one on transitions being described by some weights on transitions (depending on the current region).

For modelling purposes, authors of [7] enforce that probability distributions on delays do not forbid any delays of the interval $I(\rho, \delta)$ of possible delays, thus ruling out singular distributions like Dirac ones that would consider taking a single possible delay (like deterministic strategies do). More formally, they require $\eta_{\mathbb{R}^+}(\rho, \delta)$ to be absolutely continuous (i.e. equivalent to the Lebesgue measure) on interval $I(\rho, \delta)$. We claim that even with this assumption, the previous definition of the probability may not be well-founded, as demonstrated by the example given in [21, Appendix A]. From this example, we see the importance to moreover enforce that the distributions $\eta_\Delta(\rho)$ and $\eta_{\mathbb{R}^+}(\rho, \delta)$ are "measurable wrt the sequence of delays along the play ρ". This is easy to define for the transition part. For delays, since we want deterministic strategies to be a subset of stochastic strategies, we must be able to choose delays by using Dirac distributions, and by extension discrete distributions (that are not absolutely continuous, as [7] requires). This results in the following hypothesis:

▶ **Hypothesis 1.** *A strategy η satisfies this hypothesis if*
1. *for all transitions $\delta_0, \ldots, \delta_k, \delta$, the mapping*

$$(t_0, \ldots, t_{k-1}) \mapsto \eta_\Delta\big((\ell_0, \nu_0) \xrightarrow{(\delta_0, t_0)\cdots(\delta_{k-1}, t_{k-1})}\big)(\delta)$$

is measurable; and

2. *for all plays ρ and transition δ, the probability distribution $\eta_{\mathbb{R}^+}(\rho, \delta)$ (of the random variable t) is described by a cumulative distribution function (CDF) that is the sum of an absolutely continuous function $G(\rho, \delta)$ and Heaviside functions[1] $t \mapsto \sum_i \alpha_i(\rho, \delta) H(t - a_i(\rho, \delta))$. Moreover, for all transitions $\delta_0, \ldots, \delta_{k-1}, \delta$, the mappings $(t_0, \ldots, t_{k-1}, t) \mapsto G(\rho, \delta)(t)$, $(t_0, \ldots, t_{k-1}) \mapsto \alpha_i(\rho, \delta)$, and $(t_0, \ldots, t_{k-1}) \mapsto a_i(\rho, \delta)$ must be measurable (where we use notation ρ to denote the play $(\ell_0, \nu_0) \xrightarrow{(\delta_0, t_0) \cdots (\delta_{k-1}, t_{k-1})}$).*

This hypothesis allows us to obtain :

▶ **Lemma 2.** *If η and θ are strategies satisfying Hypothesis 1, the probabilities $\mathbb{P}_\rho^{\eta,\theta}(\pi)$ of following a path π after the play ρ are well defined. It can be extended into a probability distribution over maximal paths π starting in the last location of ρ.*

Apart from the well-definition that is new, the rest of the proof is very close to the one of [7]. The probability measure easily extends to unions of maximal paths: in particular, $\mathbb{P}_{\ell_0,\nu_0}^{\eta,\theta}(\mathsf{TPaths}_{\ell_0,\nu_0})$ is set as the sum $\sum_{\pi \in \mathsf{TPaths}_{\ell_0,\nu_0}} \mathbb{P}_{\ell_0,\nu_0}^{\eta,\theta}(\pi)$ of probabilities of all paths reaching L_T from ℓ_0. Authors of [7] go one step further, by using Carathéodory's theorem to extend the probability measure on paths $(\mathbb{P}_\rho^{\eta,\theta}(\pi))$ to a measure on plays $(\mathbb{P}_\rho^{\eta,\theta})$, whose σ-algebra is generated by maximal plays with Borel-measurable constraints on the delays. We do not formally need this further extension and will only use such extension to give an intuitive introduction of the expected payoff below. In the following, we let $\mathsf{Strat}_{\mathsf{Min}}$ and $\mathsf{Strat}_{\mathsf{Max}}$ be the sets of (stochastic) strategies satisfying Hypothesis 1, for both players. We let $\mathsf{mStrat}_{\mathsf{Min}}$ and $\mathsf{mStrat}_{\mathsf{Max}}$ be the respective subsets of memoryless strategies.

Expected payoff of plays.   As explained before, by Carathéodory's theorem, the set of plays can be equipped with a probability distribution, and we are interested in the expectation of the random variable $\mathsf{wt}(\rho)$ (where ρ conforms with two fixed strategies η and θ). This only makes sense if the probability to reach a target location is equal to 1, since otherwise, the expected weight will intuitively be $+\infty$ (there is a non-zero probability to not reach the target location, the weight of all such plays being $+\infty$). We thus now require that $\mathbb{P}_{\ell_0,\nu_0}^{\eta,\theta}(\mathsf{TPaths}_{\ell_0,\nu_0}) = 1$ (i.e. the probability to follow an infinite path is 0). We will see afterwards that this is not a sufficient condition to ensure that the expected weight is finite.

We would like to define the expected weight to reach the target as (we write $\mathbb{E}_{\ell_0,\nu_0}^{\eta,\theta}$ instead of $\mathbb{E}_{\ell_0,\nu_0}^{\eta,\theta}(\mathsf{wt})$, since we only consider the expectation of the weight wt): $\mathbb{E}_{\ell_0,\nu_0}^{\eta,\theta} = \int_\rho \mathsf{wt}(\rho) \, \mathrm{d}\mathbb{P}_{\ell_0,\nu_0}^{\eta,\theta}(\rho)$ where the integral is over all plays ρ that start in (ℓ_0, ν_0) and reach the target L_T (such restriction is again justified by the fact that the probability mass of all other plays is 0). This is problematic a priori (and we will see below an example where this indeed would be a problem) since the cumulated weight is not known to be a measurable function of the play, wrt the measure $\mathbb{P}_{\ell_0,\nu_0}^{\eta,\theta}$.

To overcome this challenge, we follow a different approach, consisting in mimicking the construction of the probability before: first define the expected payoff of all plays following a given path, and then sum over all possible paths.

▶ **Definition 3.** *We define the expected weight $\mathbb{E}_\rho^{\eta,\theta}(\pi)$ of plays that can extend ρ (the weight of ρ is thus not counted in the expectation) and that follow the path π. It is defined by induction on the length of π by $\mathbb{E}_\rho^{\eta,\theta}(\varepsilon) = 0$ and for all transitions $\delta = (\ell, g, Y, \ell')$:*

$$\mathbb{E}_\rho^{\eta,\theta}(\delta\pi) = \int_{I(\rho,\delta)} \eta_\Delta(\rho)(\delta) \Big[\big(t\,\mathsf{wt}(\ell) + \mathsf{wt}(\delta) \big) \mathbb{P}_{\rho \xrightarrow{\delta,t}}^{\eta,\theta}(\pi) + \mathbb{E}_{\rho \xrightarrow{\delta,t}}^{\eta,\theta}(\pi) \Big] \, \mathrm{d}\eta_{\mathbb{R}^+}(\rho,\delta)(t)$$

We then define the expected weight $\mathbb{E}_\rho^{\eta,\theta} = \sum_{\pi \in \mathsf{TPaths}_\rho} \mathbb{E}_\rho^{\eta,\theta}(\pi)$, when this sum converges.

---

[1] We let H denote the mapping from $\mathbb{R}$ to $[0, 1]$ such that $H(t) = 0$ if $t < 0$ and $H(t) = 1$ otherwise. Recall that it is the CDF of the Dirac distribution choosing $t = 0$.

Hypothesis 1 is sufficient to show the well-definition of all expectations $\mathbb{E}_\rho^{\eta,\theta}(\pi)$:

▶ **Lemma 4.** *If $\eta \in \mathsf{Strat}_{\mathsf{Min}}$ and $\theta \in \mathsf{Strat}_{\mathsf{Max}}$, $\mathbb{E}_\rho^{\eta,\theta}(\pi)$ is well-defined for all ρ and π.*

However, the infinite sum in $\mathbb{E}_\rho^{\eta,\theta}$ can be problematic. We thus need a stronger hypothesis to ensure its convergence. We adopt here an asymmetrical point of view, relying only on hypothesis on the strategy η of Min. Our choice is grounded in our controller synthesis view, Min being the controller desiring to reach a target location with minimum expected payoff, while Max is an uncontrollable environment.

▶ **Definition 5.** *A strategy $\eta \in \mathsf{Strat}_{\mathsf{Min}}$ of Min is said* proper *if for all finite plays ρ and strategies $\theta \in \mathsf{Strat}_{\mathsf{Max}}$, $\mathbb{P}_\rho^{\eta,\theta}(\mathsf{TPaths}_\rho) = 1$ and the infinite sum $\sum_{\pi \in \mathsf{TPaths}_\rho} \mathbb{E}_\rho^{\eta,\theta}(\pi)$ converges.*

We let $\mathsf{Strat}_{\mathsf{Min}}^{\mathsf{p}}$ be the set of proper strategies of Min, $\mathsf{mStrat}_{\mathsf{Min}}^{\mathsf{p}}$ the subset of memoryless proper strategies. Notice that a deterministic strategy of Min is proper as soon as it guarantees to reach the target set of locations (remember that we have ruled out configurations with a deterministic value $\mathsf{dVal}(\ell,\nu) = +\infty$ where Min cannot deterministically guarantee to reach the target L_T): this shows that proper strategies exist (even without using memory). For stochastic strategies, we have seen above that reaching the target set of locations with probability 1 is a necessary but not sufficient condition to be proper. Not only Max must reach the target almost surely, but he must do it *quickly enough* so that the expectation converges. We now give a sufficient condition for a strategy to be proper, that we will use in the rest of this article.

▶ **Hypothesis 2.** *A strategy $\eta \in \mathsf{Strat}_{\mathsf{Min}}$ of Min satisfies this hypothesis if there exist $m \in \mathbb{N}$ and $\alpha \in (0,1]$ such that for all finite plays ρ and strategies $\theta \in \mathsf{Strat}_{\mathsf{Max}}$, $\mathbb{P}_\rho^{\eta,\theta}(\bigcup_{n \leq m} \mathsf{TPaths}_\rho^n) \geq \alpha$.*

This hypothesis is indeed a sufficient condition for a strategy to be proper:

▶ **Lemma 6.** *All strategies of Min satisfying Hypothesis 2 are proper.*

Sketch of proof. The idea is to decompose $\mathbb{E}_\rho^{\eta,\theta}$ for all ρ as $\mathbb{E}_\rho^{\eta,\theta} = \sum_{n=0}^\infty \sum_{\pi \in \mathsf{TPaths}_\rho^n} \mathbb{E}_\rho^{\eta,\theta}(\pi)$. Since TPaths_ρ^n is a finite set, only the first sum must be shown to be converging. It is done by noticing that the weight of plays of length n grows linearly wrt n, while the probability $\sum_{\pi \in \mathsf{TPaths}_\rho^n} \mathbb{P}_\rho^{\eta,\theta}(\pi)$ decreases exponentially wrt n (thanks to Hypothesis 2). More precisely, we show for all $n \in \mathbb{N}$ and all $\rho \in \mathsf{FPlays}$, that
1. for all $\pi \in \mathsf{TPaths}_\rho^n$, $|\mathbb{E}_\rho^{\eta,\theta}(\pi)| \leq \mathbb{P}_\rho^{\eta,\theta}(\pi)\, n\, w_{\max}^e$; and
2. $\sum_{\pi \in \mathsf{TPaths}_\rho^n} \mathbb{P}_\rho^{\eta,\theta}(\pi) \leq (1-\alpha)^{\lfloor n/m \rfloor}$.
This allows us to show that $\mathbb{P}_\rho^{\eta,\theta}(\mathsf{TPaths}_\rho) = 1$ and that the sum $\mathbb{E}_\rho^{\eta,\theta} = \sum_{\pi \in \mathsf{TPaths}_\rho} \mathbb{E}_\rho^{\eta,\theta}(\pi)$ converges. ◀

Now that we have associated an expected payoff to each convenient pair of strategies, we are able to mimic the classical definition of *value* to stochastic strategies. Let ℓ be a location and ν be a valuation. For all $\eta \in \mathsf{Strat}_{\mathsf{Min}}^{\mathsf{p}}$ and $\theta \in \mathsf{Strat}_{\mathsf{Max}}$, we let $\mathsf{Val}_{\ell,\nu}^\eta = \sup_{\theta \in \mathsf{Strat}_{\mathsf{Max}}} \mathbb{E}_{\ell,\nu}^{\eta,\theta}$. Then, we let $\mathsf{Val}_{\ell,\nu}$ be the value of $\mathcal{G}$ in (ℓ,ν), defined as the best expected payoff Min can hope for: $\mathsf{Val}_{\ell,\nu} = \inf_{\eta \in \mathsf{Strat}_{\mathsf{Min}}^{\mathsf{p}}} \mathsf{Val}_{\ell,\nu}^\eta$. Both definitions can be generalised by replacing configurations (ℓ,ν) by finite plays ρ: we let Val_ρ^η and Val_ρ be the generalised versions. We also define the *memoryless values* mVal^η and mVal, where all strategies are taken memoryless.

Our main contribution, presented in details in Section 5, is to compare the memoryless (stochastic) value, the deterministic value and the stochastic value, showing their equality for a fragment of WTGs. Along the way, we will need the following result showing that when Min plays with a proper strategy, Max always has a *best response* strategy that is deterministic:

▶ **Lemma 7.** *Let $\eta \in \mathsf{Strat}^{\mathsf{P}}_{\mathsf{Min}}$ and $\varepsilon > 0$. There exists a deterministic strategy $\tau \in \mathsf{dStrat}_{\mathsf{Max}}$ such that for all finite plays ρ, $\mathbb{E}^{\eta,\tau}_{\rho} \geq \mathsf{Val}^{\eta}_{\rho} - \varepsilon$. If $\eta \in \mathsf{mStrat}^{\mathsf{P}}_{\mathsf{Min}}$ is memoryless, then there exists a deterministic strategy $\tau \in \mathsf{dStrat}_{\mathsf{Max}}$ such that for all finite plays ρ, $\mathbb{E}^{\eta,\tau}_{\rho} \geq \mathsf{mVal}^{\eta}_{\rho} - \varepsilon$.*

4 Divergent weighted timed games

As we have already seen in the introduction, interesting fragments of WTGs have been designed, in order to regain decidability of the problem of determining whether the value of a WTG is below a certain threshold. One such fragment is obtained by enforcing a semantical property of divergence (originally called *strictly non-Zeno cost* when only dealing with non-negative weights [11]): it asks that every play following a cycle in the region automaton has weight far from 0. We will consider this restriction in the following, since it allows for a large class of decidable WTGs, with no limitations on the number of clocks. Formally, a cyclic region path π of $\mathcal{R}(\mathcal{G})$ is said to be a positive cycle (resp. a negative cycle) if every finite play ρ following π satisfies $\mathsf{wt}_{\Sigma}(\rho) \geq 1$ (resp. $\mathsf{wt}_{\Sigma}(\rho) \leq -1$).

▶ **Definition 8** ([16]). *A WTG is divergent if every cyclic region path is positive or negative.*

In [16], it is shown that this definition is equivalent to requiring that for all strongly connected components (SCC) S of the graph of $\mathcal{R}(\mathcal{G})$, either every cycle π inside S is positive (we say that the SCC is positive), or every cycle π inside S is negative (we say that the SCC is negative). The best computability result in this setting is:

▶ **Theorem 9** ([16]). *The deterministic value of a divergent WTG can be computed in triply-exponential-time.*

We explain how to recover from Theorem 9 the needed shape of ε-optimal strategies, since this is one of the new technical ingredient we need afterwards.

Switching strategies for Min. Theorem 9 is obtained in [16] by using a *value iteration algorithm* (originally described in [1] for acyclic timed automata). If V represents a value function, i.e. a mapping $L \times \mathbb{R}^{\mathcal{C}}_{\geq 0} \to \mathbb{R}_{\infty}$, we denote by $V_{\ell,\nu}$ the image $V(\ell, \nu)$, for better readability. One step of the game is summarised in the following operator $\mathcal{F}$ mapping each value function V to the value function defined for all $(\ell, \nu) \in L \times \mathbb{R}^{\mathcal{C}}_{\geq 0}$ by $\mathcal{F}(V)_{\ell,\nu} = 0$ if $\ell \in L_T$, $\mathcal{F}(V)_{\ell,\nu} = \sup_{(\ell,\nu) \xrightarrow{\delta,t} (\ell',\nu')} \left[t \times \mathsf{wt}(\ell) + \mathsf{wt}(\delta) + V_{\ell',\nu'} \right]$ if $\ell \in L_{\mathsf{Max}}$, and $\mathcal{F}(V)_{\ell,\nu} = \inf_{(\ell,\nu) \xrightarrow{\delta,t} (\ell',\nu')} \left[t \times \mathsf{wt}(\ell) + \mathsf{wt}(\delta) + V_{\ell',\nu'} \right]$ if $\ell \in L_{\mathsf{Min}}$, where $(\ell, \nu) \xrightarrow{\delta,t} (\ell', \nu')$ ranges over valid edges in $\mathcal{G}$. Then, starting from V^0 mapping every configuration to $+\infty$, except for the targets mapped to 0, we let $V^i = \mathcal{F}(V^{i-1})$ for all $i > 0$. The value function V^i is intuitively what Min can guarantee when forced to reach the target in at most i steps.

The value computation of Theorem 9 is then obtained in two steps. First, configurations (ℓ, ν) of value $\mathsf{dVal}_{\ell,\nu} = -\infty$ are found by using a decomposition of the region game $\mathcal{R}(\mathcal{G})$ into strongly-connected components (SCC). Indeed, in divergent WTGs, configurations of value $-\infty$ are all the ones from which Min has a strategy to visit infinitely many times configurations of a single location (ℓ, r) of $\mathcal{R}(\mathcal{G})$ contained in a negative SCC. This is thus a Büchi objective on the region game, that can easily be solved with some attractor computations. Notice that if a configuration (ℓ, ν) has value $-\infty$, this implies that all configurations (ℓ, ν') with ν' in the same region as ν have value $-\infty$. As we explained at the end of Section 2 for the values $+\infty$, we can then remove configurations of value $-\infty$ by strengthening the guards on transitions, while letting unchanged other finite values.

Then, the (finite) value dVal is obtained as an iterate V^H of the previous operator, with H polynomial in the size of the region game and the maximal weights of $\mathcal{G}$. This means that playing for only a bounded number of steps is equivalent to the original game. In particular, at horizon H, we have that $\mathcal{F}(V^H) = V^{H+1} = \mathsf{dVal}$ so that dVal is a fixpoint of $\mathcal{F}$. As a side effect, this allows one to decompose the clock space $\mathbb{R}^{\mathcal{C}}_{\geq 0}$ into a finite number α of *cells* (a refinement of the classical regions) such that dVal is affine on each cell.

Based on this, we can construct good strategies for Min that have a special form, the so-called *switching strategies* (introduced in [15] in the untimed setting, further extended in the timed setting with only one-clock in [14]).

▶ **Definition 10.** *A switching strategy σ is described by two deterministic memoryless strategies σ^1 and σ^2, as well as a switching threshold K. The strategy σ then consists in playing strategy σ^1 until either we reach a target location, or the finite play has length at least K, in which case we switch to strategy σ^2.*

Our new contribution is as follows:

▶ **Theorem 11.** *In a divergent WTG, for all $\varepsilon > 0$ and $N \in \mathbb{N}$, there exists a switching strategy σ for Min, for which the two components σ^1 and σ^2 satisfy Hypothesis 1, such that for all configurations (ℓ, ν), $\mathsf{dVal}^\sigma_{\ell,\nu} \leq \max(-N, \mathsf{dVal}_{\ell,\nu}) + \varepsilon$.*

In particular, if all configurations have a finite deterministic value, **there exists an ε-optimal switching strategy wrt the deterministic value**. In the presence of a configuration with a deterministic value $-\infty$, we build from Theorem 11 a family of switching strategies (indexed by the parameter N) whose value tends to $-\infty$.

The proof of Theorem 11 requires to build both strategies σ^1 and σ^2, as well as a switching threshold K. The second strategy σ^2 only consists in reaching the target and is thus obtained as a deterministic memoryless strategy from a classical attractor computation in the region game $\mathcal{R}(\mathcal{G})$. It is easy to choose σ^2 smooth enough so that it fulfils Hypothesis 1. In contrast, the first strategy σ^1 requires more care. We build it so that it fulfils two properties, that we summarise in saying that σ^1 is *fake-ε-optimal* wrt the deterministic value:

1. each finite play conforming to σ^1 from (ℓ, ν) and reaching the target has a cumulated weight at most $\mathsf{dVal}_{\ell,\nu} + |\rho|\,\varepsilon$ (in particular, if $\mathsf{dVal}_{\ell,\nu} = -\infty$, no such plays should exist);

2. each finite play conforming to σ^1 following a *long enough* cycle in the region game $\mathcal{R}(\mathcal{G})$ has a cumulated weight at most -1.

Here, "fake" means that σ^1 is not obliged to guarantee reaching the target, but if it does so, it must do it with a cumulated weight close to $\mathsf{dVal}_{\ell,\nu}$, the error factor depending linearly on the size of the play. The second property ensures that playing long enough σ^1 without reaching the target results in diminishing the cumulated weight. Then, if the switch happens at horizon K big enough, $(K = (w^e_{\max}|\mathcal{R}(\mathcal{G})|(|L|\alpha + 2) + N)(|\mathcal{R}(\mathcal{G})|(|L|\alpha + 1) + 1)$ suffices for instance), Min is sure that the cumulated weight so far is low enough so that the rest of the play to reach a target location (following σ^2 only) will not make the weight increase too much. In the absence of values $-\infty$ in dVal, the first property allows one to obtain a $K\varepsilon$-optimal strategy even in the case where the switch does not occur (because we reach the target prematurely). The construction of a fake-ε/K-optimal strategy σ^1 (the linear dependency on the length of the play in the first property of fake-optimality is thus taken care by a division by K here) relies on the fact that $\mathcal{F}(\mathsf{dVal}) = \mathsf{dVal}$ to play almost-optimally at horizon 1. More formally:

▬ For all configurations of value $-\infty$, σ^1 is built as a winning strategy for the Büchi objective "visit infinitely often configurations of a location (ℓ, r) of $\mathcal{R}(\mathcal{G})$ contained in a negative SCC". By definition, all cyclic paths following σ^1 will be inside a negative

SCC, and thus be of cumulated weight at most -1, by divergence of the WTG. Moreover, no plays conforming to σ^1 from such a configuration of value $-\infty$ will reach a target location, since the chosen negative SCC is a trap controlled by Min. It is easy to choose σ^1 smooth enough so that it fulfils Hypothesis 1.

▬ For the remaining configurations of finite value, we rely upon operator $\mathcal{F}$, letting σ^1 choose a decision that minimises the value at horizon 1. However, because of the guards on clocks, infimum/supremum operators in $\mathcal{F}$ are not necessarily minima/maxima, and we thus need to allow for a small error at each step of the strategy: this is the main difference with the untimed setting, which by the way explains why our definition of switching strategy needed to be adapted. We will use the $\mathrm{arginf}^\varepsilon$ operator defined for all mappings $f\colon A \to \mathbb{R}$ and $B \subseteq A$ by $\mathrm{arginf}_B^\varepsilon f = \{a \in B \mid f(a) \leq \inf_B f + \varepsilon\}$. Then, for all configurations $(\ell, \nu) \in L_{\mathsf{Min}} \times \mathbb{R}_{\geq 0}^{\mathcal{C}}$, we choose $\sigma^1(\ell, \nu)$ as a pair (δ, t) in

$$\mathrm{arginf}^{\varepsilon/K}_{(\ell,\nu)\xrightarrow{\delta,t}(\ell',\nu')}(t\,\mathsf{wt}(\ell) + \mathsf{wt}(\delta) + \mathsf{dVal}_{\ell',\nu'})$$

This set is non empty since dVal is a fixpoint of operator $\mathcal{F}$ in this case. Moreover, knowing that the mapping dVal_ℓ is piecewise affine by the results shown in [16], it is possible to choose σ^1 so that it fulfils the measurability (even piecewise continuity) conditions of Hypothesis 1. More precisely, we can consider it to take the same kind of decision for all configurations of a same cell: same transition, and either no delay or a delay jumping to the same border of cell.

The strategy σ^1 thus built makes a small error wrt the optimal at each step. But once again strongly relying on the divergence of the WTG, we can nevertheless show that σ^1 is fake-ε/K-optimal wrt the deterministic value.

Memoryless strategies for Max.   WTGs are known to be determined [14], i.e. the deterministic value is also equal to $\mathsf{dVal}_{\ell,\nu} = \sup_{\tau \in \mathsf{dStrat}_{\mathsf{Max}}} \inf_{\sigma \in \mathsf{dStrat}_{\mathsf{Min}}} \mathsf{wt}(\mathsf{Play}((\ell,\nu),\sigma,\tau))$. In this setting, we can turn our study to the point of view of Max, looking for good strategies for this other player. A deterministic strategy τ of Max has an associated value: $\mathsf{dVal}^\tau_{\ell,\nu} = \inf_{\sigma \in \mathsf{dStrat}_{\mathsf{Min}}} \mathsf{wt}(\mathsf{Play}((\ell,\nu),\sigma,\tau))$. It is ε-optimal wrt the deterministic value if $\mathsf{dVal}^\tau_{\ell,\nu} \geq \mathsf{dVal}_{\ell,\nu} - \varepsilon$ for all (ℓ,ν).

As Max does not wish to go to the target, we show that no switch is necessary to play ε-optimally: memoryless strategies are sufficient to guarantee a value as close as wanted to the deterministic value. For a configuration with a value equal to $-\infty$, all the deterministic strategies for Max are equivalent where they are all equally bad. Without loss of generality, we can therefore suppose that there are no configurations in $\mathcal{G}$ with a value equal to $-\infty$. Then, it is shown in [16] that remaining values are bounded in absolute value by $w_{\max}^e|\mathcal{R}(\mathcal{G})|$, since *optimal plays* have no cycles. We use that fact to build a memoryless deterministic strategy τ analogous to strategy σ^1 before:

▶ **Theorem 12.** *In a divergent WTG, there exists a memoryless ε-optimal strategy for player* Max *wrt the deterministic value (that moreover satisfies Hypothesis 1).*

5   Emulate memory with randomness, and vice versa

The main contribution of this article, apart from defining a notion of expected value in weighted timed games, is to relate the different notions of values. In divergent WTGs, memory can thus be fully emulated with stochastic choices, and combining memory and stochastic choices does not bring more power to players, which we summarise by:

▶ **Theorem 13.** *In all divergent WTGs, for all (ℓ, ν), $\mathsf{dVal}_{\ell,\nu} = \mathsf{Val}_{\ell,\nu} = \mathsf{mVal}_{\ell,\nu}$.*

The proof of this result is decomposed into several inequalities on these values. One is easier, and holds for all WTGs: the stochastic value is at most equal to the deterministic value, using the inclusion of deterministic strategies into stochastic ones, and Lemma 7.

▶ **Lemma 14.** *In all WTGs $\mathcal{G}$, for all configurations (ℓ, ν), $\mathsf{Val}_{\ell,\nu} \leq \mathsf{dVal}_{\ell,\nu}$.*

We show in the rest of this section the inequalities comparing the deterministic value with the other values: first we show that memoryless stochastic strategies can emulate deterministic ones ($\mathsf{mVal}_{\ell,\nu} \leq \mathsf{dVal}_{\ell,\nu}$); then we show that deterministic strategies can emulate stochastic ones ($\mathsf{dVal}_{\ell,\nu} \leq \mathsf{Val}_{\ell,\nu}$ and $\mathsf{dVal}_{\ell,\nu} \leq \mathsf{mVal}_{\ell,\nu}$).

Simulating deterministic strategies with memoryless strategies.   We focus here on showing that, for all configurations (ℓ, ν), $\mathsf{mVal}_{\ell,\nu} \leq \mathsf{dVal}_{\ell,\nu}$. We build a memoryless strategy of Min at least as good as a deterministic strategy. By Theorem 11, we can start from a switching strategy for Min. For $N \in \mathbb{N}$ and $\varepsilon > 0$, we thus consider a switching strategy $\sigma = (\sigma^1, \sigma^2, K)$ of value $\mathsf{dVal}^\sigma_{\ell,\nu} \leq \max(-N, \mathsf{dVal}_{\ell,\nu}) + \varepsilon$, and simulate it with a memoryless strategy for Min, denoted η^p, with a probability parameter $p \in (0, 1)$. This new strategy is a probabilistic superposition of the two memoryless deterministic strategies σ^1 and σ^2.

Formally, we define $\eta^p(\ell, \nu)$, with $\ell \in L_{\mathsf{Min}}$, depending on the sign of the SCC containing the location (ℓ, r), with r the region of ν, of the region game $\mathcal{R}(\mathcal{G})$.

In a positive SCC, Min always chooses to play σ^1, thus looking for a negative cycle in the next SCCs (in topological order) if any. Formally, letting $(\delta_1, t_1) = \sigma^1(\ell, \nu)$, we define $\eta^p_\Delta(\ell, \nu) = \mathsf{Dirac}_{\delta_1}$ and $\eta^p_{\mathbb{R}^+}((\ell, \nu), \delta_1) = \mathsf{Dirac}_{t_1}$.

In a negative SCC, we let $\eta^p(\ell, \nu)$ be the distribution picking $\sigma^1(\ell, \nu)$ with probability p, and $\sigma^2(\ell, \nu)$ with probability $1 - p$. Formally, η^p_Δ chooses the transition given by $\sigma^1(\ell, \nu)$ with probability p and of σ^2, with probability $1 - p$, except if those transitions are equal, in which case η^p_Δ chooses it with probability 1. If transitions chosen by σ^1 and σ^2 are distinct, $\eta^p_{\mathbb{R}^+}$ is a Dirac distribution over the corresponding delay. Otherwise, $\eta^p_{\mathbb{R}^+}$ chooses with probability p the delay given by σ^1, and with probability $1 - p$ the one given by σ^2.

Theorem 11 ensuring that strategies σ^1 and σ^2 satisfy Hypothesis 1, the superposition η^p also satisfies these hypotheses. Moreover, we use the sufficient condition in Hypothesis 2 to show that η^p is also proper:

▶ **Lemma 15.** *For all $p \in (0, 1)$, the strategy η^p satisfies Hypothesis 2.*

To show the expected result, we prove that $\mathsf{mVal}^{\eta^p}_{\ell,\nu} \leq \max(-N, \mathsf{dVal}_{\ell,\nu}) + 3\varepsilon$ for all (ℓ, ν), for p *close enough to* 1: we conclude that $\mathsf{mVal}_{\ell,\nu} \leq \mathsf{dVal}_{\ell,\nu}$ by taking the limit when N tends to $+\infty$ and ε tends to 0. We get that inequality by showing the following result, paired with the fact that $\mathsf{dVal}^\sigma_{\ell,\nu} \leq \max(-N, \mathsf{dVal}_{\ell,\nu}) + \varepsilon$.

▶ **Proposition 16.** *For all configurations (ℓ, ν) and $\varepsilon > 0$ small enough, there exists $\tilde{p} \in (0, 1)$ so that for all $p \in [\tilde{p}, 1)$, $\mathsf{mVal}^{\eta^p}_{\ell,\nu} \leq \mathsf{dVal}^\sigma_{\ell,\nu} + 2\varepsilon$.*

Proof. Lemma 7 allows us to limit ourselves to deterministic strategies for Max. For all deterministic strategies τ of Max, we compute a lower bound on p independent of τ such that $\mathbb{E}^{\eta^p,\tau}_{\ell,\nu} \leq \mathsf{dVal}^\sigma_{\ell,\nu} + 3\varepsilon/2$. By Lemma 7 (with $\varepsilon/2$), we obtain the desired $\mathsf{mVal}^{\eta^p}_{\ell,\nu} \leq \mathsf{dVal}^\sigma_{\ell,\nu} + 2\varepsilon$.

The case where the whole region game only contains positive SCCs is easy, since then η^p chooses the transition and delay given by σ^1 with probability 1. By divergence, $\mathcal{G}$ then contains no negative cycles. A play conforming to η^p is also conforming to the deterministic
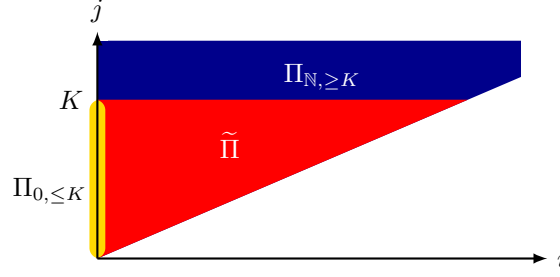
Figure 2 Partition of paths TPaths.

strategy σ^1, so it must be acyclic. In particular, there exists only one play ρ conforming to η^p and τ. This one is also conforming to σ and thus reaches the target with a cumulated weight $\mathsf{wt}_\Sigma(\rho) = \mathbb{E}_{\ell,\nu}^{\eta^p,\tau} \leq \mathsf{dVal}_{\ell,\nu}^\sigma$ as expected.

Now, suppose that the region graph contains at least a negative SCC. Thus, we let $c > 0$ be the maximal size of an elementary cycle of the region game (that visits a pair (ℓ, r) at most once) and $w^- > 0$ be the opposite of the maximal cumulated weight of an elementary negative cycle in $\mathcal{R}(\mathcal{G})$ (necessarily bounded by $w_{\max}^e |\mathcal{R}(\mathcal{G})|$).

We partition the set $\mathsf{FPlays}_{\ell,\nu}^{\eta^p,\tau}$ into subsets $\Pi_{i,j}$ according to the number i of choices of probability $1 - p$ along the play (the probability as described previously with the product of the probabilities given by η_Δ^p and $\eta_{\mathbb{R}^+}^p$), and their length j (we always have $i \leq j$). The partition is depicted in Figure 2:

- $\Pi_{\mathbb{N}, \geq K}$, depicted in blue, contains all plays with a length greater than K (the switching threshold)
- $\Pi_{0, \leq K}$, depicted in yellow, contains all plays without any probability $1 - p$, with a length at most K;
- $\widetilde{\Pi}$, depicted in red, contains the rest of the plays.

We can use the particular shape of the memoryless strategy η^p for Min, and the fact that we fixed a deterministic strategy τ for Max, to decompose the expectation $\mathbb{E}_\rho^{\eta^p,\tau}$ on the partition. Indeed, notice that the set of plays conforming to η^p and τ, from a particular configuration (ℓ, ν), is countable. Moreover, we can associate a probability to each *play* (instead of a probability to a *path*). For a finite play $\rho = (\ell_0, \nu_0) \xrightarrow{\delta_0, t_0} \cdots (\ell_{k-1}, \nu_{k-1})$ conforming to η^p and τ, we let

$$\mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\rho) = \prod_{i=0}^{k-1} p_i$$

where, for all $i \in \{0, \ldots, k-1\}$

$$p_i = \begin{cases} 1 & \text{if } \ell_i \in L_{\mathsf{Max}} \\ \eta_\Delta^p(\ell_i, \nu_i)(\delta_i) \times \eta_{\mathbb{R}^+}^p((\ell_i, \nu_i), \delta_i)(t_i) & \text{if } \ell_i \in L_{\mathsf{Min}} \end{cases}$$

This definition allows us to recover the probability of a path π using the probability of all plays following π. Then, we obtain easily

$$\mathbb{E}_{\ell,\nu}^{\eta^p,\tau} = \underbrace{\sum_{\rho \in \Pi_{0,\leq K}} \mathsf{wt}(\rho) \mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\rho)}_{\gamma_{0,\leq K}} + \underbrace{\sum_{\rho \in \Pi_{\mathbb{N},\geq K}} \mathsf{wt}(\rho) \mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\rho)}_{\gamma_{\mathbb{N},\geq K}} + \underbrace{\sum_{\rho \in \widetilde{\Pi}} \mathsf{wt}(\rho) \mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\rho)}_{\widetilde{\gamma}} \tag{1}$$

We now compute and bound the three expectations $\gamma_{0,\leq K}$, $\gamma_{\mathbb{N},\geq K}$ and $\widetilde{\gamma}$. In the following, for a set Π of plays, we let $\mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\Pi) = \sum_{\rho \in \Pi} \mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\rho)$.

Red zone is such that $\widetilde{\gamma} \leq \varepsilon/4$. All plays in $\widetilde{\Pi}$ have a length at most K: so the cumulated weight of all such play is at most $K w_{\max}^e$. So, we have

$$\widetilde{\gamma} = \sum_{\rho \in \widetilde{\Pi}} \mathsf{wt}(\rho) \mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\rho) \leq \sum_{\rho \in \widetilde{\Pi}} K w_{\max}^e \mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\rho) = K w_{\max}^e \mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\widetilde{\Pi})$$

But, all plays $\rho \in \bigcup_{j \leq K} \Pi_{i,j}$ with $i \leq K$ take i transitions of probability $1 - p$. In particular, by bounding all other probabilities by 1, and since there are at most 2^K plays in $\bigcup_{j \leq K} \Pi_{i,j}$, we obtain (using that $1 - (1-p)^K \leq 1$)

$$\mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\widetilde{\Pi}) \leq 2^K \sum_{i=1}^{K-1} (1-p)^i = 2^K \frac{(1-p)(1-(1-p)^K)}{p} \leq 2^K \frac{1-p}{p} \longrightarrow_{p \to 1} 0 \tag{2}$$

If we suppose that

$$p \geq \frac{1}{1 + \varepsilon/(4 \times 2^K K w_{\max}^e)}$$

we obtain as desired

$$\widetilde{\gamma} \leq 2^K K w_{\max}^e \frac{1-p}{p} \leq \frac{\varepsilon}{4}$$

Yellow and blue zones are such that $\gamma_{0,\leq K} + \gamma_{\mathbb{N},\geq K} \leq \mathsf{dVal}_{\ell,\nu}^\sigma + 5\varepsilon/4$. All plays in $\Pi_{0,\leq K}$ reach the target without taking any probability $1 - p$ from η_Δ^p, so they are conforming to σ^1. In the case where $\mathsf{dVal}_{\ell,\nu} = -\infty$, $\Pi_{0,\leq K} = \emptyset$ and $\gamma_{0,\leq K} = 0$, since no play conforming to σ^1 from (ℓ, ν) reaches the target. In this case, Min can stay in a cycle with a negative cumulated weight as long as he wants. Now, if $\mathsf{dVal}_{\ell,\nu}$ is finite, Theorem 11 (see [21, Lemma 19]) allows us to show that the cumulated weight of a play in $\Pi_{0,\leq K}$ is at most $\mathsf{dVal}_{\ell,\nu} + K\varepsilon/K = \mathsf{dVal}_{\ell,\nu} + \varepsilon$, as $\mathsf{dVal}_{\ell,\nu} = \inf_{\sigma \in \mathsf{dStrat_{Min}}} \mathsf{dVal}_{\ell,\nu}^\sigma \leq \mathsf{dVal}_{\ell,\nu}^\sigma$. Therefore, in both cases, we can write

$$\gamma_{0,\leq K} \leq \left(\mathsf{dVal}_{\ell,\nu}^\sigma + \varepsilon\right) \mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\Pi_{0,\leq K})$$

Let ρ be a play in $\Pi_{i,j}$ with $0 \leq i$ and $j \geq K$. Since η^p only allows cycles in negative SCCs, all region cycles in ρ have a cumulated weight at most -1. By definition of K and the proof of Theorem 11, $\mathsf{wt}(\rho) \leq \mathsf{dVal}_{\ell,\nu}^\sigma \leq \mathsf{dVal}_{\ell,\nu}^\sigma + \varepsilon$.

By summing up the contribution of yellow and blue zones, we get

$$\gamma_{0,\leq K} + \gamma_{\mathbb{N},\geq K} \leq \left(\mathsf{dVal}_{\ell,\nu}^\sigma + \varepsilon\right) \mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\Pi_{0,\leq K} \cup \Pi_{\mathbb{N},\geq K}) \tag{3}$$

We distinguish two cases.

- If $\mathsf{dVal}_{\ell,\nu}^\sigma \geq -5\varepsilon/4$, having $\mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\Pi_{0,\leq K} \cup \Pi_{\mathbb{N},\geq K}) \leq 1$, we get

$$\gamma_{0,\leq K} + \gamma_{\mathbb{N},\geq K} \leq \left(\mathsf{dVal}_{\ell,\nu}^\sigma + \frac{5\varepsilon}{4}\right) \mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\Pi_{0,\leq K} \cup \Pi_{\mathbb{N},\geq K}) \leq \mathsf{dVal}_{\ell,\nu} + \frac{5\varepsilon}{4}$$

- If $\mathsf{dVal}_{\ell,\nu}^\sigma < -5\varepsilon/4$, then by (2),

$$\mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\Pi_{0,\leq K} \cup \Pi_{\mathbb{N},\geq K}) = 1 - \mathbb{P}_{\ell,\nu}^{\eta^p,\tau}(\widetilde{\Pi}) \geq 1 - 2^K \frac{1-p}{p} \longrightarrow_{p \to 1} 1$$

If we suppose that

$$p \geq \frac{2^K}{2^K + 1 - \frac{\mathsf{dVal}^\sigma_{\ell,\nu} + 5\varepsilon/4}{\mathsf{dVal}^\sigma_{\ell,\nu} + \varepsilon}} \in (0,1)$$

then

$$\mathbb{P}^{\eta^p,\tau}_{\ell,\nu}(\Pi_{0,\leq K} \cup \Pi_{\mathbb{N},\geq K}) \geq \frac{\mathsf{dVal}^\sigma_{\ell,\nu} + 5\varepsilon/4}{\mathsf{dVal}^\sigma_{\ell,\nu} + \varepsilon}$$

and, by negativity of $\mathsf{dVal}^\sigma_{\ell,\nu} + \varepsilon$, we can rewrite (3) as

$$\gamma_{0,\leq K} + \gamma_{\mathbb{N},\geq K} \leq \left(\mathsf{dVal}^\sigma_{\ell,\nu} + \varepsilon\right) \frac{\mathsf{dVal}^\sigma_{\ell,\nu} + 5\varepsilon/4}{\mathsf{dVal}^\sigma_{\ell,\nu} + \varepsilon} = \mathsf{dVal}^\sigma_{\ell,\nu} + \frac{5\varepsilon}{4}$$

In all cases, we have $\gamma_{0,\leq K} + \gamma_{\mathbb{N},\geq K} \leq \mathsf{dVal}^\sigma_{\ell,\nu} + 5\varepsilon/4$.

Gathering all constraints on p. We gather all the lower bounds over p that we need in the proof:

$$\tilde{p} = \begin{cases} \max\left(\frac{1}{1+\varepsilon/(4\times 2^K K w^e_{\max})}, \frac{1}{2}\right) & \text{if } \mathsf{dVal}^\sigma_{\ell,\nu} \geq -5\varepsilon/4 \\[2ex] \max\left(\frac{1}{1+\varepsilon/(4\times 2^K K w^e_{\max})}, \frac{1}{2}, \frac{2^K}{2^K+1-\frac{\mathsf{dVal}^\sigma_{\ell,\nu}+5\varepsilon/4}{\mathsf{dVal}^\sigma_{\ell,\nu}+\varepsilon}}\right) & \text{otherwise} \end{cases}$$

Then, for $p \in (\tilde{p}, 1)$, we have $\mathbb{E}^{\eta^p,\tau}_{\ell,\nu} \leq \mathsf{dVal}^\sigma_{\ell,\nu} + 3\varepsilon/2$. Since $\tilde{p}$ does not depend on τ, we conclude that for $p \in (\tilde{p}, 1)$, we have $\mathsf{mVal}^{\eta^p}_{\ell,\nu} \leq \mathsf{dVal}^\sigma_{\ell,\nu} + 2\varepsilon$. ◀

Simulating (stochastic) strategies with deterministic strategies. Finally, we show that, for all configurations (ℓ, ν), $\mathsf{dVal}_{\ell,\nu} \leq \mathsf{Val}_{\ell,\nu}$ and $\mathsf{dVal}_{\ell,\nu} \leq \mathsf{mVal}_{\ell,\nu}$. To do so, we consider a proper strategy $\eta \in \mathsf{Strat}^{\mathsf{P}}_{\mathsf{Min}}$ of Min and an initial configuration (ℓ, ν). We build a deterministic strategy σ such that $\mathsf{dVal}^\sigma_{\ell,\nu} \leq \mathsf{Val}^\eta_{\ell,\nu} + \varepsilon$. Thus, in the case where $\mathsf{Val}_{\ell,\nu} \neq -\infty$, we can consider η to be an ε-optimal strategy so that $\mathsf{dVal}^\sigma_{\ell,\nu} \leq \mathsf{Val}_{\ell,\nu} + 2\varepsilon$ which allows us to conclude. In the case where $\mathsf{Val}_{\ell,\nu} = -\infty$, then, for all $N \in \mathbb{N}$, we can consider η to be such that $\mathsf{Val}^\eta_{\ell,\nu} \leq -N$. Then, $\mathsf{dVal}^\sigma_{\ell,\nu} \leq -N + \varepsilon$, which implies that $\mathsf{dVal}_{\ell,\nu} = -\infty$, by taking the limit when N tends to $+\infty$.

The deterministic strategy σ uses the same kind of *memory* as η (in particular, it will be memoryless if η is memoryless). However, we want this strategy to be relatively simple to define, independent of the memory of η. Intuitively, we want to build a switching strategy (as in Section 4) on a game induced by the memory of η, i.e. a deterministic strategy σ^1 that uses the memory capabilities of η, a memoryless deterministic strategy σ^2 obtained by an attractor in the region game, and a threshold K. Strategy σ then consists in playing σ^1 for at most K steps, before switching to strategy σ^2. The construction of σ^1 is done in a similar way as in the deterministic case, Min always choosing the best possible candidate according to the choices of η, thus trying to minimise the immediate reward obtained in one turn. Under this condition, we verify that σ^1 satisfies some properties similar to the fake-ε-optimality encountered in Section 4. Then, by mimicking the techniques of Theorem 11, we obtain that the switching strategy η obtained from η^1 satisfies the desired inequality $\mathsf{dVal}^\sigma_{\ell,\nu} \leq \mathsf{Val}^\eta_{\ell,\nu} + \varepsilon$.

6   Conclusion

We have introduced stochastic strategies for WTGs, showing that, in divergent games, Min can use randomisation to emulate memory, and vice versa. We aim at extending our study to more general WTGs. As a first step, we may consider the class of almost-divergent WTGs (adding the possibility for an execution following a region cycle to have weight *exactly 0*), used in [12, 17] to obtain an approximation schema of the optimal value. We wonder if similar ε-optimal switching strategies may exist also in this context, one of the crucial argument in order to extend our emulation result. Another question concerns the implementability of the randomised strategies: even if they use no memory, they still need to know the precise current clock valuation. In (non-weighted) timed games, previous work [18] aimed at removing this need for precision, by using stochastic strategies where the delays are chosen with probability distributions that do not require exact knowledge of the clocks measurements. In our setting, we aim at further studying the implementability of the randomised strategies of Min in WTGs, e.g. by requiring them to be robust against small imprecisions.

References

1   Rajeev Alur, Mikhail Bernadsky, and P. Madhusudan. Optimal reachability for weighted timed games. In *Proceedings of the 31st International Colloquium on Automata, Languages and Programming (ICALP'04)*, volume 3142 of *LNCS*, pages 122–133. Springer, 2004.

2   Rajeev Alur and David L. Dill. A theory of timed automata. *Theoretical Computer Science*, 126(2):183–235, 1994.

3   Rajeev Alur, Salvatore La Torre, and George J. Pappas. Optimal paths in weighted timed automata. *Theoretical Computer Science*, 318(3):297–322, 2004.

4   Eugene Asarin and Oded Maler. As soon as possible: Time optimal control for timed automata. In *Hybrid Systems: Computation and Control*, volume 1569 of *LNCS*, pages 19–30. Springer, 1999.

5   Gerd Behrmann, Ansgar Fehnker, Thomas Hune, Kim Larsen, Paul Pettersson, Judi Romijn, and Frits Vaandrager. Minimum-cost reachability for priced time automata. In *International Workshop on Hybrid Systems: Computation and Control*, pages 147–161. Springer, 2001.

6   Nathalie Bertrand, Patricia Bouyer, Thomas Brihaye, and Pierre P. Carlier. When are stochastic transition systems tameable? *Journal of Logical and Algebraic Methods in Programming*, 99:41–96, 2018.

7   Nathalie Bertrand, Patricia Bouyer, Thomas Brihaye, Quentin Menet, Christel Baier, Marcus Größer, and Marcin Jurdzinski. Stochastic timed automata. *Log. Methods Comput. Sci.*, 10(4), 2014.

8   Patricia Bouyer, Thomas Brihaye, Véronique Bruyère, and Jean-François Raskin. On the optimal reachability problem of weighted timed automata. *Formal Methods in System Design*, 31(2):135–175, 2007.

9   Patricia Bouyer, Thomas Brihaye, Pierre Carlier, and Quentin Menet. Compositional design of stochastic timed automata. In Alexander S. Kulikov and Gerhard J. Woeginger, editors, *Computer Science - Theory and Applications - 11th International Computer Science Symposium in Russia, CSR 2016, St. Petersburg, Russia, June 9-13, 2016, Proceedings*, volume 9691 of *Lecture Notes in Computer Science*, pages 117–130. Springer, 2016.

10   Patricia Bouyer, Thomas Brihaye, Mickael Randour, Cédric Rivière, and Pierre Vandenhove. Decisiveness of stochastic systems and its application to hybrid models. In Davide Bresolin and Jean-François Raskin, editors, *Proceedings of the 11th International Symposium on Games, Automata, Logics, and Formal Verification (GandALF'20)*, volume 326 of *Electronic Proceedings in Theoretical Computer Science*, 2020.

11 Patricia Bouyer, Franck Cassez, Emmanuel Fleury, and Kim G. Larsen. Optimal strategies in priced timed game automata. In Kamal Lodaya and Meena Mahajan, editors, *FSTTCS 2004: Foundations of Software Technology and Theoretical Computer Science*, pages 148–160, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

12 Patricia Bouyer, Samy Jaziri, and Nicolas Markey. On the value problem in weighted timed games. In *Proceedings of the 26th International Conference on Concurrency Theory (CONCUR'15)*, volume 42 of *Leibniz International Proceedings in Informatics*, pages 311–324. Leibniz-Zentrum für Informatik, 2015.

13 Thomas Brihaye, Véronique Bruyère, and Jean-François Raskin. On optimal timed strategies. In Paul Pettersson and Wang Yi, editors, *Formal Modeling and Analysis of Timed Systems*, pages 49–64, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

14 Thomas Brihaye, Gilles Geeraerts, Axel Haddad, Engel Lefaucheux, and Benjamin Monmege. Simple priced timed games are not that simple. In *Proceedings of the 35th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'15)*, volume 45 of *LIPIcs*, pages 278–292. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2015.

15 Thomas Brihaye, Gilles Geeraerts, Axel Haddad, and Benjamin Monmege. Pseudopolynomial iterative algorithm to solve total-payoff games and min-cost reachability games. *Acta Informatica*, 54(1):85–125, 2017.

16 Damien Busatto-Gaston, Benjamin Monmege, and Pierre-Alain Reynier. Optimal reachability in divergent weighted timed games. In *Foundations of Software Science and Computation Structures - 20th International Conference, FOSSACS 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22-29, 2017, Proceedings*, pages 162–178, 2017.

17 Damien Busatto-Gaston, Benjamin Monmege, and Pierre-Alain Reynier. Symbolic approximation of weighted timed games. In Sumit Ganguly and Paritosh Pandya, editors, *Proceedings of the 38th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'18)*, volume 122 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 28:1–28:16. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2018.

18 Krishnendu Chatterjee, Thomas A. Henzinger, and Vinayak S. Prabhu. Trading infinite memory for uniform randomness in timed games. In *Hybrid Systems: Computation and Control, 11th International Workshop, HSCC 2008, St. Louis, MO, USA, April 22-24, 2008. Proceedings*, pages 87–100, 2008.

19 Marcin Jurdziński and Ashutosh Trivedi. Reachability-time games on timed automata. In *Proceedings of the 34th International Colloquium on Automata, Languages and Programming (ICALP'07)*, volume 4596 of *LNCS*, pages 838–849. Springer, 2007.

20 Benjamin Monmege, Julie Parreaux, and Pierre-Alain Reynier. Reaching your goal optimally by playing at random with no memory. In Igor Konnov and Laura Kovács, editors, *Proceedings of the 31st International Conference on Concurrency Theory (CONCUR 2020)*, volume 171 of *LIPIcs*, pages 26:1–26:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

21 Benjamin Monmege, Julie Parreaux, and Pierre-Alain Reynier. Playing stochastically in weighted timed games to emulate memory. Technical Report 2105.00984, arXiv, 2021. `arXiv:2105.00984`.