

---

# Reachability in Continuous-Time Markov Reward Decision Processes

Christel Baier<sup>1</sup>

Boudewijn R. Haverkort<sup>2</sup>

Holger Hermanns<sup>3</sup>

Joost-Pieter Katoen<sup>4</sup>

<sup>1</sup> Faculty of Computer Science  
Technical University of Dresden  
D-01062 Dresden, Germany

<sup>2</sup> Department of Computer Science  
University of Twente  
P.O. Box 217  
7500 AE Enschede, the Netherlands

<sup>3</sup> Department of Computer Science  
Saarland University  
D-66123 Saarbrücken, Germany

<sup>4</sup> Department of Computer Science  
RWTH Aachen University  
D-52056 Aachen, Germany

baier@tcs.inf.tu-dresden.de, brh@cs.utwente.nl,  
hermanns@cs.uni-sb.de, katoen@cs.rwth-aachen.de

---

## Abstract

Continuous-time Markov decision processes (CTMDPs) are widely used for the control of queueing systems, epidemic and manufacturing processes. Various results on optimal schedulers for discounted and average reward optimality criteria in CTMDPs are known, but the typical game-theoretic winning objectives have received scant attention so far. This paper studies various sorts of reachability objectives for CTMDPs. Memoryless schedulers are optimal for simple reachability objectives as it suffices to consider the embedded MDP. Schedulers that may count the number of visits to states are optimal—when restricting to time-abstract schedulers—for timed reachability in uniform CTMDPs. The central result is that for any CTMDP, reward reachability objectives are dual to timed ones. As a corollary,  $\epsilon$ -optimal schedulers for reward reachability objectives in uniform CTMDPs can be obtained in polynomial time using a simple backward greedy algorithm.

## 1 Introduction

Having their roots in economics, Markov decision processes (MDPs, for short) in computer science are used in application areas such as randomised distributed algorithms and security protocols. The discrete probabilities are used to model random phenomena in such algorithms, like flipping a coin or choosing an identity from a fixed range according to a uniform distribution, whereas the nondeterminism in MDPs is used to specify unknown or underspecified behaviour, e.g., concurrency (interleaving) or the unknown malicious behavior of an attacker.

MDPs – also considered as turn-based  $1\frac{1}{2}$ -player stochastic games – consist of decision epochs, states, actions, and transition probabilities. On entering a state, an action,  $\alpha$ , say, is nondeterministically selected and the next state is determined randomly by a probability distribution that depends on  $\alpha$ . Actions may incur a reward, interpreted as gain, or dually, as cost. Schedulers or strategies prescribe which actions to choose in a state. One of the simplest schedulers, the so-called memoryless ones, base their decision solely on the current state and not on the further history. A plethora of results for MDPs are known that mainly focus on finding an optimal scheduler for a certain objective, see e.g. [8]. For, e.g., reachability objectives – find a scheduler, possibly the simplest one, that maximises the probability to reach a set of states – memoryless schedulers suffice and can be determined in polynomial time. For step-bounded reachability objectives, finite memory schedulers are sufficient. These schedulers perform the selection process on the basis of a finite piece of information, typically encoded as a finite-state automaton that runs in parallel to the MDP at hand.

This paper considers turn-based  $1\frac{1}{2}$ -player stochastically timed games, also known as *continuous-time* Markov decision processes (CTMDPs) [8]. They behave as MDPs but in addition their timing behaviour is random. The probability to stay at most  $t$  time units in a state is determined by a negative exponential distribution of which the rate depends on  $\alpha$ . A reward is obtained which is linearly dependent on the time  $t$  spent in state  $s$ , as well as on a factor  $\rho(s, \alpha)$ , the state- and action-dependent reward rate. In contrast to MDPs, CTMDPs have received far less attention; a reason for this might be the increased complexity when moving to continuous time. This paper studies reachability objectives for CTMDPs, in particular time-bounded reachability – what is the optimal policy to reach a set of states within a certain deadline – reward-bounded reachability, and their combination. We survey the results in this field, and show that reward-bounded and time-bounded reachability are interchangeable.

The presented reachability objectives are for instance relevant for job-shop scheduling problems where individual jobs have a random exponential

duration, see e.g., [5]. The problem of finding a schedule for a fixed number of such (preemptable) jobs on a given set of identical machines such that the probability to meet a given deadline is maximised, is, in fact, an instance of timed reachability on CTMDPs. Optimal memoryless strategies exist for minimising the sum of the job completion times, but, as is shown, this is not the case for maximising the probability to reach the deadline. The same applies for maximising the probability to complete all jobs within a fixed cost.

This paper is further structured as follows. Section 2 rehearses the necessary background in the area of Markov decision processes, schedulers, stochastic processes, and reachability objectives. Section 3 then recalls the logic CSRL and discusses its semantics for continuous-time Markov reward decision processes. Section 4 then discusses a number of new results on the duality of the roles of time and reward in such processes. Section 5 concludes the paper.

## 2 Preliminaries

### 2.1 Markov decision processes

Let  $AP$  be a fixed set of atomic propositions.

**Definition 2.1** (CTMDP). A *continuous-time Markov decision process* (CTMDP)  $\mathcal{M}$  is a tuple  $(S, Act, \mathbf{R}, L)$  with  $S$ , a countable set of *states*,  $Act$ , a set of *actions*,  $\mathbf{R} : S \times Act \times S \rightarrow \mathbb{R}_{\geq 0}$ , the rate function such that for each  $s \in S$  there exists a pair  $(\alpha, s') \in Act \times S$  with  $\mathbf{R}(s, \alpha, s') > 0$ , and labeling function  $L : S \rightarrow 2^{AP}$ .

The set of actions that are enabled in state  $s$  is denoted  $Act(s) = \{\alpha \in Act \mid \exists s'. \mathbf{R}(s, \alpha, s') > 0\}$ . The above condition thus requires each state to have at least one outgoing transition. Note that this condition can easily be fulfilled by adding self-loops.

The operational behavior of a CTMDP is as follows. On entering state  $s$ , an action  $\alpha$ , say, in  $Act(s)$  is nondeterministically selected. Given that action  $\alpha$  has been chosen, the probability that the transition  $s \xrightarrow{\alpha} s'$  can be triggered within the next  $t$  time units is  $1 - e^{-\mathbf{R}(s, \alpha, s') \cdot t}$ . The delay of transition  $s \xrightarrow{\alpha} s'$  is thus governed by a negative exponential distribution with rate  $\mathbf{R}(s, \alpha, s')$ . If multiple outgoing transitions exist for the chosen action, they compete according to their exponentially distributed delays. For  $B \subseteq S$ , let  $\mathbf{R}(s, \alpha, B)$  denote the total rate from state  $s$  to some state in  $B$ , i.e.,

$$\mathbf{R}(s, \alpha, B) = \sum_{s' \in B} \mathbf{R}(s, \alpha, s').$$

Let  $\underline{E}(s, \alpha) = \mathbf{R}(s, \alpha, S)$  denote the exit rate of state  $s$  under action  $\alpha$ . If  $\underline{E}(s, \alpha) > 0$ , the probability to move from  $s$  to  $s'$  via action  $\alpha$  within  $t$

time units, i.e., the probability that  $s \xrightarrow{\alpha} s'$  wins the competition among all outgoing  $\alpha$ -transitions of  $s$  is:

$$\frac{\mathbf{R}(s, \alpha, s')}{\underline{E}(s, \alpha)} \cdot \left(1 - e^{-\underline{E}(s, \alpha) \cdot t}\right),$$

where the first factor describes the discrete probability to take transition  $s \xrightarrow{\alpha} s'$  and the second factor reflects the sojourn time in state  $s$  given that  $s$  is left via action  $\alpha$ . Note that the sojourn time is distributed negative exponentially with rate equal to the sum of the rates of the outgoing  $\alpha$ -transitions of state  $s$ . This is conform the minimum property of exponential distributions.

A CTMC (a *continuous-time Markov chain*) is a CTMDP in which for each state  $s$ ,  $\text{Act}(s)$  is a singleton. In this case, the selection of actions is purely deterministic, and  $\mathbf{R}$  can be projected on an  $(S \times S)$  matrix, known as the transition rate matrix.

**Definition 2.2** (MDP). A (discrete-time) *Markov decision process* (MDP)  $\mathcal{M}$  is a tuple  $(S, \text{Act}, \mathbf{P}, L)$  with  $S$ ,  $\text{Act}$ , and  $L$  as before and  $\mathbf{P} : S \times \text{Act} \times S \rightarrow [0, 1]$ , a probability function such that for each pair  $(s, \alpha)$ :

$$\sum_{s' \in S} \mathbf{P}(s, \alpha, s') \in \{0, 1\}.$$

A DTMC (a *discrete-time Markov chain*) is an MDP in which for each state  $s$ ,  $\text{Act}(s)$  is a singleton. In this case,  $\mathbf{P}$  can be projected on an  $(S \times S)$  matrix, known as the transition probability matrix of a DTMC.

**Definition 2.3** (Embedded MDP of a CTMDP). For CTMDP  $\mathcal{M} = (S, \text{Act}, \mathbf{R}, L)$ , the discrete probability of selecting transition  $s \xrightarrow{\alpha} s'$  is determined by the *embedded MDP*, denoted  $\text{emb}(\mathcal{M}) = (S, \text{Act}, \mathbf{P}, L)$ , with:

$$\mathbf{P}(s, \alpha, s') = \begin{cases} \frac{\mathbf{R}(s, \alpha, s')}{\underline{E}(s, \alpha)}, & \text{if } \underline{E}(s, \alpha) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

$\mathbf{P}(s, \alpha, s')$  is the time-abstract probability for the  $\alpha$ -transition from  $s$  to  $s'$  on selecting action  $\alpha$ . For  $B \subseteq S$  let  $\mathbf{P}(s, \alpha, B) = \sum_{s' \in B} \mathbf{P}(s, \alpha, s')$ .

**Definition 2.4** (Path in a CTMDP). An infinite path in a CTMDP  $\mathcal{M} = (S, \text{Act}, \mathbf{R}, L)$  is a sequence  $s_0, \alpha_0, t_0, s_1, \alpha_1, t_1, s_2, \alpha_2, t_2, \dots$  in  $(S \times \text{Act} \times \mathbb{R}_{>0})^\omega$ , written as:

$$s_0 \xrightarrow{\alpha_0, t_0} s_1 \xrightarrow{\alpha_1, t_1} s_2 \xrightarrow{\alpha_2, t_2} \dots$$

Any finite prefix of  $\sigma$  that ends in a state is a finite path in  $\mathcal{M}$ . Let  $\text{Paths}(\mathcal{M})$  denote the set of infinite paths in  $\mathcal{M}$ .

Let  $\sigma = s_0 \xrightarrow{\alpha_0, t_0} s_1 \xrightarrow{\alpha_1, t_1} s_2 \xrightarrow{\alpha_2, t_2} \dots \in Paths(\mathcal{M})$ . The time-abstract path of  $\sigma$  is  $s_0 \xrightarrow{\alpha_0} s_1 \xrightarrow{\alpha_1} s_2 \xrightarrow{\alpha_2} \dots$ , the corresponding action-abstract path is:  $s_0 \xrightarrow{t_0} s_1 \xrightarrow{t_1} s_2 \xrightarrow{t_2} \dots$ , and the time- and action-abstract path is the state sequence  $s_0, s_1, s_2, \dots$ . Let  $first(\sigma)$  denote the first state of  $\sigma$ . For finite path  $\sigma$ ,  $last(\sigma)$  denotes the last state of  $\sigma$ , and  $\sigma \rightarrow s$  denotes the finite time- and action-abstract path  $\sigma$  followed by state  $s$ . For  $i \in \mathbb{N}$ , let  $\sigma[i] = s_i$  denote the  $(i+1)$ -st state of  $\sigma$ .  $\sigma@t$  denotes the state occupied at time instant  $t \in \mathbb{R}_{\geq 0}$ , i.e.,  $\sigma@t = \sigma[k]$  where  $k$  is the smallest index such that  $\sum_{i=0}^k t_i > t$ .

**Definition 2.5** (CMRDP). A *continuous-time Markov reward decision process* (CMRDP) is a pair  $(\mathcal{M}, \rho)$  with  $\mathcal{M}$  a CTMDP with state space  $S$  and  $\rho : S \times Act \rightarrow \mathbb{R}_{\geq 0}$  a reward function.

CMRDPs are often called CTMDPs in the literature [8]. The state reward function  $\rho$  assigns to each state  $s \in S$  and action  $\alpha \in Act$  a reward rate  $\rho(s, \alpha)$ . Under the condition that action  $\alpha$  has been selected in state  $s$ , a reward  $\rho(s, \alpha) \cdot t$  is acquired after residing  $t$  time units in state  $s$ . Recall that  $t$  is governed by an exponential distribution with rate  $E(s, \alpha)$ , i.e.,  $t$  randomly depends on action  $\alpha$ . A path through a CMRDP is a path through its underlying CTMDP. For timed path  $\sigma = s_0 \xrightarrow{\alpha_0, t_0} s_1 \xrightarrow{\alpha_1, t_1} s_2 \xrightarrow{\alpha_2, t_2} \dots$  and  $t = \sum_{i=0}^{k-1} t_i + t'$  with  $t' \leq t_k$  let:

$$y(\sigma, t) = \sum_{i=0}^{k-1} t_i \cdot \rho(s_i, \alpha_i) + t' \cdot \rho(s_k, \alpha_k)$$

the accumulated reward along  $\sigma$  up to time  $t$ . An MRM (*Markov reward model*) is a CTMC equipped with a reward function. As an MRM is action-deterministic,  $\rho$  may be viewed as a function of the type  $S \rightarrow \mathbb{R}_{\geq 0}$ .

## 2.2 Schedulers

CTMDPs incorporate nondeterministic decisions, not present in CTMCs. Nondeterminism in a CTMDP is resolved by a *scheduler*. In the literature, schedulers are sometimes also referred to as adversaries, policies, or strategies. For deciding which of the next nondeterministic actions to take, a scheduler may “have access” to the current state only or to the path from the initial to the current state (either with or without timing information). Schedulers may select the next action either (i) *deterministically*, i.e., depending on the available information, the next action is chosen in a deterministic way, or (ii) in a *randomized* fashion, i.e., depending on the available information the next action is chosen probabilistically. Accordingly, the following classes of schedulers  $D$  are distinguished [8], where  $Distr(Act)$  denotes the collection of all probability distributions on  $Act$ :

- stationary Markovian deterministic (*SMD*),  $D : S \rightarrow Act$  such that  $D(s) \in Act(s)$
- stationary Markovian randomized (*SMR*),  $D : S \rightarrow Distr(Act)$  such that  $D(s)(\alpha) > 0$  implies  $\alpha \in Act(s)$
- Markovian deterministic (*MD*, also called *step-dependent schedulers*),  $D : S \times \mathbb{N} \rightarrow Act$  such that  $D(s, n) \in Act(s)$
- Markovian randomized (*MR*),  $D : S \times \mathbb{N} \rightarrow Distr(Act)$  such that  $D(s, n)(\alpha) > 0$  implies  $\alpha \in Act(s)$
- (time-abstract) history-dependent, deterministic (*HD*),  $D : (S \times Act)^* \times S \rightarrow Act$  such that

$$D(\underbrace{s_0 \xrightarrow{\alpha_0} s_1 \xrightarrow{\alpha_1} \dots \xrightarrow{\alpha_{n-1}}}_{\text{time-abstract history}}, s_n) \in Act(s_n)$$

- (time-abstract) history-dependent, randomized (*HR*),  $D : (S \times Act)^* \times S \rightarrow Distr(Act)$  such that  $D(s_0 \xrightarrow{\alpha_0} s_1 \xrightarrow{\alpha_1} \dots \xrightarrow{\alpha_{n-1}}, s_n)(\alpha) > 0$  implies  $\alpha \in Act(s_n)$ .

All these schedulers are time-abstract and cannot base their decisions on the sojourn times. Timed (measurable) schedulers [9, 7] are not considered in this paper. Finally, let  $X$  denote the class of all  $X$ -schedulers over a fixed CTMDP  $\mathcal{M}$ .<sup>1</sup>

Note that for any HD-scheduler, the actions can be dropped from the history, i.e., HD-schedulers may be considered as functions  $D : S^+ \rightarrow Act$ , as for any sequence  $s_0, s_1, \dots, s_n$  the relevant actions  $\alpha_i$  are given by  $\alpha_i = D(s_0, s_1, \dots, s_i)$ , and, hence, the scheduled action sequence can be constructed from prefixes of the path at hand. Any state-action sequence  $s_0 \xrightarrow{\alpha_0} s_1 \xrightarrow{\alpha_1} \dots \xrightarrow{\alpha_{n-1}} s_n$  where  $\alpha_i \neq D(s_0, s_1, \dots, s_i)$  for some  $i$ , does not describe a path fragment that can be obtained from  $D$ .

The scheduler-types form a hierarchy, e.g., any SMD-scheduler can be viewed as an MD-scheduler (by ignoring parameter  $n$ ) which, in turn, can be viewed as an HD-scheduler (by ignoring everything from the history except its length). A similar hierarchy exists between SMR, MR, and HR schedulers. Moreover, deterministic schedulers can be regarded as trivial versions of their corresponding randomized counterparts that assign probability one to the actions selected.

<sup>1</sup> Strictly speaking, we should write  $X(\mathcal{M})$  but  $\mathcal{M}$  is omitted as it should be clear from the context.

### 2.3 Induced stochastic process

Given a scheduler  $D$  (of arbitrary type listed above) and a starting state,  $D$  induces a stochastic process on a CTMDP  $\mathcal{M}$ . For deterministic schedulers (HD, MD, and SMD), the induced process is a CTMC, referred to as  $\mathcal{M}_D$  in the sequel. For MD- and HD-schedulers, though, the state space of  $\mathcal{M}_D$  will in general be infinitely large (but countable).

**Definition 2.6** (Induced CTMC of a CTMDP). Let  $\mathcal{M} = (S, Act, \mathbf{R}, L)$  be a CTMDP and  $D : S^+ \rightarrow Act$  an HD-scheduler on  $\mathcal{M}$ . The CTMC  $\mathcal{M}_D = (S^+, \mathbf{R}_D, L')$  with:

$$\mathbf{R}_D(\sigma, \sigma') = \begin{cases} \mathbf{R}(\text{last}(\sigma), D(\sigma), s), & \text{if } \sigma' = \sigma \rightarrow s, \\ 0, & \text{otherwise,} \end{cases}$$

and  $L'(\sigma) = L(\text{last}(\sigma))$ .

The embedded DTMC  $\text{emb}(\mathcal{M}_D)$  is a tuple  $(S^+, \mathbf{P}_D, L)$  where:

$$\mathbf{P}_D(\sigma, \sigma') = \begin{cases} \frac{\mathbf{R}_D(\sigma, \sigma')}{E_D(\sigma)}, & \text{if } E_D(\sigma) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Here,  $E_D(\sigma) = \mathbf{R}_D(\sigma, S^+)$ , i.e., the exit rate of  $\sigma$  in  $\mathcal{M}_D$ . States in CTMC  $\mathcal{M}_D$  can be seen as state sequences  $s_0 \rightarrow s_1 \rightarrow \dots \rightarrow s_{n-1} \rightarrow s_n$  corresponding to time- and action-abstract path fragments in the CTMDP  $\mathcal{M}$ . State  $s_n$  stands for the current state in the CTMDP whereas states  $s_0$  through  $s_{n-1}$  describe the history. Intuitively, the stochastic process induced by an HD-scheduler  $D$  on the CTMDP  $\mathcal{M}$  results from unfolding  $\mathcal{M}$  into an (infinite) tree while resolving the nondeterministic choices according to  $D$ . For SMD-schedulers, the induced CTMC is guaranteed to be finite. More precisely, for SMD-scheduler  $D$ ,  $\mathcal{M}_D$  can be viewed as a CTMC with the original state space  $S$ , as all sequences that end in  $s$ , say, are lumping equivalent [6].

In contrast to a CTMDP (or MDP), a CTMC (or DTMC) is a fully determined stochastic process. For a given initial state  $s_0$  in CTMC  $\mathcal{M}$ , a unique probability measure  $Pr_{s_0}$  on  $\text{Paths}(s_0)$  exists, where  $\text{Paths}(s_0)$  denotes the set of timed paths that start in  $s_0$ . Timed paths through a CTMC are defined as for CTMDPs, but by nature are action-abstract. The inductive construction of the probability measure below follows [2], the fact that we allow countable-state Markov chains does not alter the construction. Let  $\mathbf{P}$  be the probability matrix of the embedded DTMC of  $\mathcal{M}$  and let  $\text{Cyl}(s_0 \xrightarrow{I_0} \dots \xrightarrow{I_{k-1}} s_k)$  denote the cylinder set consisting of all timed paths  $\sigma$  that start in state  $s_0$  such that  $s_i$  ( $i \leq k$ ) is the  $(i+1)$ -th state on  $\sigma$  and the time spent in  $s_i$  lies in the non-empty interval  $I_i$  ( $i < k$ ) in  $\mathbb{R}_{\geq 0}$ .

The cylinder sets induce the probability measure  $Pr_{s_0}$  on the timed paths through  $\mathcal{M}$ , defined by induction on  $k$  by  $Pr_{s_0}(Cyl(s_0)) = 1$ , and, for  $k > 0$ :

$$Pr_{s_0}(Cyl(s_0 \xrightarrow{I_0} \dots \xrightarrow{I_{k-1}} s_k \xrightarrow{I'} s')) = Pr_{s_0}(Cyl(s_0 \xrightarrow{I_0} \dots \xrightarrow{I_{k-1}} s_k)) \cdot \mathbf{P}(s_k, s') \cdot (e^{-E(s_k) \cdot a} - e^{-E(s_k) \cdot b}),$$

where  $a = \inf I'$  and  $b = \sup I'$ .

#### 2.4 Reachability objectives

For CMRDP  $\mathcal{M}$  with state space  $S$  and  $B \subseteq S$ , we consider the maximum (or, dually, minimum) probability to reach  $B$  under a given class of schedulers. Let  $\diamond B$  denote the event to eventually reach some state in  $B$ ,  $\diamond^{\leq t} B$  the same event with the extra condition that  $B$  is reached within  $t$  time units, and  $\diamond_{\leq r} B$  the event that  $B$  is eventually reached within accumulated reward  $r$ . The event  $\diamond_{\leq r}^{\leq t} B$  asserts that  $B$  is reached within  $t$  time units and accumulated reward at most  $r$ . Note that the accumulated reward gained depends on the sojourn times in states, hence the bounds  $t$  and  $r$  are not independent. It is not difficult to assess that these events are measurable for the time-abstract schedulers considered here. A detailed proof of the measurability of  $\diamond^{\leq t} B$  for measurable timed schedulers (a richer class of schedulers) can be found in [7]. The probability for such an event  $\varphi$  to hold in state  $s$  of  $\mathcal{M}$  is denoted  $Pr(s \models \varphi)$ , i.e.,

$$Pr(s \models \varphi) = Pr_s\{\sigma \in Paths(\mathcal{M}) \mid \sigma \models \varphi\}.$$

The maximal probability to reach a state in  $B$  under a *HR*-scheduler is given by:

$$Pr_{HR}^{\max}(s \models \diamond B) = \sup_{D \in HR} Pr(s \models \diamond B).$$

In a similar way,  $Pr_{HR}^{\min}(s \models \diamond B) = \inf_{D \in HR} Pr(s \models \diamond B)$ .

The following result follows immediately from the fact that for event  $\diamond B$  it suffices to consider the embedded MDP of a given CTMDP, and the fact that memoryless schedulers for finite MDPs exist that maximize the reachability probability for  $B$ . Such memoryless schedulers are obtained in polynomial time by solving a linear optimization problem. A similar result holds for minimal probabilities and for events of the form  $\diamond^{\leq n} B$ , i.e., the event that  $B$  is reached within  $n \in \mathbb{N}$  steps (i.e., transitions). Note that the event  $\diamond^{\leq t} B$  requires a state in  $B$  to be reached within  $t$  time units (using an arbitrary number of transitions), while  $\diamond^{\leq n} B$  requires  $B$  to be reached in  $n$  discrete steps, regardless of the time spent to reach  $B$ .

**Lemma 2.7** (Optimal SMD schedulers for reachability). Let  $\mathcal{M}$  be a finite CTMDP with state space  $S$  and  $B \subseteq S$ . There exists an SMD scheduler  $D$

such that for any  $s \in S$ :

$$Pr^D(s \models \diamond B) = Pr_{HR}^{\max}(s \models \diamond B).$$

## 2.5 Time- and cost-bounded reachability

Consider the following class of CTMDPs:

**Definition 2.8** (Uniform CTMDP). A CTMDP  $(S, Act, \mathbf{R}, L)$  is *uniform* if for some  $E > 0$  it holds  $E(s, \alpha) = E$  for any state  $s \in S$  and  $\alpha \in Act(s)$ .

Stated in words, in a uniform CTMDP the exit rates for all states and all enabled actions are equal. It follows from [3]:

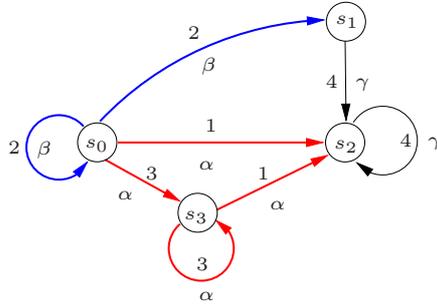
**Theorem 2.9** (Optimal MD schedulers for timed reachability). Let  $\mathcal{M}$  be a finite *uniform* CTMDP with state space  $S$ ,  $t \in \mathbb{R}_{\geq 0}$  and  $B \subseteq S$ . There exists an MD scheduler  $D$  such that for any  $s \in S$ :

$$Pr^D(s \models \diamond^{\leq t} B) = Pr_{HR}^{\max}(s \models \diamond^{\leq t} B).$$

An  $\epsilon$ -approximation of such scheduler, i.e., a scheduler that obtains  $Pr^D(s \models \diamond^{\leq t} B)$  up to an accuracy of  $\epsilon$ , can be obtained in polynomial time by a greedy backward reachability algorithm as presented in [3]. A similar result can be obtained for minimal time-bounded reachability probabilities by selecting a transition with smallest, rather than largest, probability in the greedy algorithm.

The following example shows that memoryless schedulers for maximal time-bounded reachability probabilities may not exist.

**Example 2.10** (Optimal SMD schedulers may not exist). Consider the following uniform CTMDP:



Action labels and rates are indicated at each edge. Let  $B = \{s_2\}$ , and consider the SMD-schedulers,  $D_\alpha$ , selecting action  $\alpha$  in state  $s_0$ , and  $D_\beta$ , selecting action  $\beta$ . Comparing them with  $D_{\beta\alpha}$ , i.e., the scheduler that after selecting  $\beta$  once switches to selecting  $\alpha$  in state  $s_0$ , we find that for a certain

range of time bounds  $t$ ,  $D_{\beta\alpha}$  outperforms both  $D_\beta$  and  $D_\alpha$ . Intuitively, the probability of stuttering in state  $s_0$  (by choosing  $\beta$  initially) may influence the remaining time to reach  $B$  to an extent that it becomes profitable to continue choosing  $\alpha$ . For  $t = 0.5$ , for instance,  $\Pr_{D_{\beta\alpha}}(s_0, \diamond^{\leq 0.5} B) = 0.4152$ , whereas for  $D_\alpha$  and  $D_\beta$  these probabilities are 0.3935 and 0.3996, respectively.

The following result is of importance later and is partially based on a result in [3]. Informally, it states that maximal (and minimal) probabilities for timed reachabilities in CTMDPs under deterministic and randomised HD schedulers coincide. As this result holds for arbitrary CTMDPs, there is no need to restrict to uniform ones here.

**Theorem 2.11** (Maximal probabilities are invariant under randomization). For CMRDP  $\mathcal{M}$  with state space  $S$ ,  $s \in S$  and  $B \subseteq S$ , it holds for any  $r, t \in \mathbb{R}_{\geq 0} \cup \{\infty\}$ :

$$\begin{aligned} \sup_{D \in HD} \Pr^D(s \models \diamond^{\leq t} B) &= \sup_{D \in HR} \Pr^D(s \models \diamond^{\leq t} B) \\ \sup_{D \in HD} \Pr^D(s \models \diamond_{\leq r} B) &= \sup_{D \in HR} \Pr^D(s \models \diamond_{\leq r} B) \\ \sup_{D \in HD} \Pr^D(s \models \diamond_{\leq r}^{\leq t} B) &= \sup_{D \in HR} \Pr^D(s \models \diamond_{\leq r}^{\leq t} B). \end{aligned}$$

Analogous results hold for minimal probabilities for the events  $\diamond^{\leq t} B$ ,  $\diamond_{\leq r} B$ , and  $\diamond_{\leq r}^{\leq t} B$ .

*Proof.* For any HD-scheduler  $D$  for the CTMDP  $\mathcal{M}$  it holds:

$$\Pr^D(s \models \diamond^{\leq t} B) = \lim_{n \rightarrow \infty} \Pr^D(s \models \diamond^{\leq t, \leq n} B)$$

where the superscript  $\leq n$  denotes that  $B$  has to be reached within at most  $n$  transitions. Similarly, we have:

$$\Pr^D(s \models \diamond_{\leq r} B) = \lim_{n \rightarrow \infty} \Pr^D(s \models \diamond_{\leq r}^{\leq n} B).$$

By induction on  $n$ , it can be shown (cf. [3, Theorem 7]) that there is a finite family  $(D_i)_{i \in J_n}$  (with  $J_n$  an index set) of HD-schedulers such that the measure  $\Pr_{D'}$  induced by an HR-scheduler  $D'$  for the cylinder sets induced by path fragments consisting of  $n$  transitions is a convex combination of the measures  $\Pr_{D_i}$ ,  $i \in J_n$ . Q.E.D.

The results for the events  $\diamond B$  and  $\diamond^{\leq t} B$  in finite CTMDP  $\mathcal{M}$  can be generalized towards constrained reachability properties  $C \cup B$  and  $C \cup^{\leq t} B$ , respectively, where  $C \subseteq S$ . This works as follows. First, all states in  $S \setminus (C \cup B)$  and in  $B$  are made absorbing, i.e., their enabled actions are replaced by a single action,  $\alpha_s$ , say, with  $\mathbf{R}(s, \alpha_s, s) > 0$ . The remaining

states are unaffected. Paths that visit some state in  $S \setminus (C \cup B)$  contribute probability zero to the event  $C \cup B$  while the continuation of paths that have reached  $B$  is of no importance to the probability of this event. For the resulting CTMDP  $\mathcal{M}'$  it follows:

$$\begin{aligned} Pr_{\mathcal{M},X}^{\max}(s \models C \cup^{\leq n} B) &= Pr_{\mathcal{M}',X}^{\max}(s \models \diamond^{\leq n} B), \\ Pr_{\mathcal{M},X}^{\max}(s \models C \cup B) &= Pr_{\mathcal{M}',X}^{\max}(s \models \diamond B), \\ Pr_{\mathcal{M},X}^{\max}(s \models C \cup^{\leq t} B) &= Pr_{\mathcal{M}',X}^{\max}(s \models \diamond^{\leq t} B), \end{aligned}$$

where the subscript of  $Pr$  indicates the CTMDP of interest. Similar results hold for  $Pr^{\min}$ .

For the event  $C \cup_{\leq r} B$  in CMRDP  $\mathcal{M}$ , the states in  $S \setminus C \cup B$  are made absorbing (as before) and the reward of states in  $B$  is set to zero. The latter ensures that the accumulation of reward halts as soon as  $B$  is reached. Then it follows:

$$Pr_{\mathcal{M},X}^{\max}(s \models C \cup_{\leq r} B) = Pr_{\mathcal{M}^*,X}^{\max}(s \models \diamond_{\leq r} B)$$

and similar for  $Pr^{\min}$  where  $\mathcal{M}^*$  is the resulting CMRDP after the transformations indicated above.

### 3 Continuous Stochastic Reward Logic

CSRL is a branching-time temporal logic, based on the Computation Tree Logic (CTL). A CSRL formula asserts conditions on a state of a CMRDP. Besides the standard propositional logic operators, CSRL incorporates the probabilistic operator  $\mathbb{P}_J(\varphi)$  where  $\varphi$  is a path-formula and  $J$  is an interval of  $[0, 1]$ . The path-formula  $\varphi$  imposes a condition on the set of paths, whereas  $J$  indicates a lower bound and/or upper bound on the probability. The intuitive meaning of the formula  $\mathbb{P}_J(\varphi)$  in state  $s$  is: the probability for the set of paths satisfying  $\varphi$  and starting in  $s$  meets the bounds given by  $J$ . The probabilistic operator can be considered as the quantitative counterpart to the CTL-path quantifiers  $\exists$  and  $\forall$ .

The path formulae  $\varphi$  are defined as for CTL, except that a bounded until operator is additionally incorporated. The intuitive meaning of the path formula  $\Phi \cup_K^I \Psi$  for intervals  $I, K \subseteq \mathbb{R}_{\geq 0}$  is that a  $\Psi$ -state should be reached within  $t \in I$  time units via a  $\Phi$ -path with total cost  $r \in K$ .

**Definition 3.1** (Syntax of CSRL). CSRL *state-formulae* over the set  $AP$  of atomic propositions are formed according to the following grammar:

$$\Phi ::= \text{true} \mid a \mid \Phi_1 \wedge \Phi_2 \mid \neg \Phi \mid \mathbb{P}_J(\varphi),$$

where  $a \in AP$ ,  $\varphi$  is a path-formula and  $J \subseteq [0, 1]$  is an interval with rational bounds. CSRL *path-formulae* are formed according to:

$$\varphi ::= \bigcirc_K^I \Phi \mid \Phi_1 \cup_K^I \Phi_2,$$

where  $\Phi$ ,  $\Phi_1$  and  $\Phi_2$  are state-formulae, and  $I, K \subseteq \mathbb{R}_{\geq 0} \cup \{\infty\}$ .

Other boolean connectives such as  $\vee$  and  $\rightarrow$  are derived in the obvious way. The reachability event considered before is obtained by  $\diamond_K^I \Phi = \text{true} \cup_K^I \Phi$ . The always-operator  $\square$  can be obtained by the duality of always/eventually and lower/upper probability bounds, e.g.,

$$\mathbb{P}_{\geq p}(\square_K^I \Phi) = \mathbb{P}_{\leq 1-p}(\diamond_K^I \neg\Phi) \text{ and } \mathbb{P}_{[p,q]}(\square_K^I \Phi) = \mathbb{P}_{[1-q,1-p]}(\diamond_K^I \neg\Phi).$$

Special cases occur for the trivial time-bound  $I = [0, \infty)$  and the trivial reward-bound  $K = [0, \infty)$ :

$$\bigcirc \Phi = \bigcirc_{[0,\infty)}^{[0,\infty)} \Phi \text{ and } \Phi \cup \Psi = \Phi \cup_{[0,\infty)}^{[0,\infty)} \Psi.$$

The semantics of CSRL is defined over the class of *HR*-schedulers.

**Definition 3.2** (Semantics of CSRL). Let  $a \in AP$ ,  $\mathcal{M} = (S, Act, \mathbf{R}, L)$  a CMRDP,  $s \in S$ ,  $\Phi, \Psi$  CSRL state-formulae, and  $\varphi$  a CSRL path-formula. The satisfaction relation  $\models$  for state-formulae is defined by:

$$\begin{aligned} s \models a & \quad \text{iff} \quad a \in L(s) \\ s \models \neg\Phi & \quad \text{iff} \quad s \not\models \Phi \\ s \models \Phi \wedge \Psi & \quad \text{iff} \quad s \models \Phi \text{ and } s \models \Psi \\ s \models \mathbb{P}_J(\varphi) & \quad \text{iff} \quad \text{for any scheduler } D \in HR : Pr^D(s \models \varphi) \in J. \end{aligned}$$

For path  $\sigma = s_0 \xrightarrow{\alpha_0, t_0} s_1 \xrightarrow{\alpha_1, t_1} s_2 \xrightarrow{\alpha_2, t_2} \dots$  in  $\mathcal{M}$ :

$$\begin{aligned} \sigma \models \bigcirc_K^I \Phi & \quad \text{iff} \quad \sigma[1] \models \Phi, t_0 \in I \text{ and } y(\sigma, t_0) \in K \\ \sigma \models \Phi \cup_K^I \Psi & \quad \text{iff} \quad \exists t \in I. (\sigma @ t \models \Psi \wedge (\forall t' < t. \sigma @ t' \models \Phi) \wedge y(\sigma, t) \in K). \end{aligned}$$

The semantics for the propositional fragment of CSRL is standard. The probability operator  $\mathbb{P}_J(\cdot)$  imposes probability bounds for all (time-abstract) schedulers. Accordingly,  $s \models \mathbb{P}_{\leq p}(\varphi)$  if and only if  $Pr_{HR}^{\max}(s \models \varphi) \leq p$ , and similarly,  $s \models \mathbb{P}_{\geq p}(\varphi)$  if and only if  $Pr_{HR}^{\min}(s \models \varphi) \geq p$ . The well-definedness of the semantics of  $\mathbb{P}_J(\varphi)$  follows from the fact that for any CSRL path-formula  $\varphi$ , the set  $\{\sigma \in Paths(s) \mid \sigma \models \varphi\}$  is measurable. This follows from a standard measure space construction over the infinite paths in the stochastic process induced by an *HD*-scheduler over the CMRDP  $\mathcal{M}$ . In fact, the measurability of these events can also be guaranteed for measurable timed schedulers, cf. [7].

Recall that  $\sigma @ t$  denotes the current state along  $\sigma$  at time instant  $t$ , and  $y(\sigma, t)$  denotes the accumulated reward along the prefix of  $\sigma$  up to time  $t$ . The intuition behind  $y(\sigma, t)$  depends on the formula under consideration and

the interpretation of the rewards in the CMRDP  $\mathcal{M}$  at hand. For instance, for  $\varphi = \diamond good$  and path  $\sigma$  that satisfies  $\varphi$ , the accumulated reward  $y(\sigma, t)$  can be interpreted as the cost to reach a *good* state within  $t$  time units. For  $\varphi = \diamond bad$ , it may, e.g., be interpreted as the energy used before reaching a *bad* state within  $t$  time units.

## 4 Duality of Time and Reward

The main aim of this section is to show the duality of rewards and the elapse of time in a CMRDP. The proof strategy is as follows. We first consider the action-deterministic case, i.e., MRMs, and show that – in spirit of the observations in the late nineteen seventies by Beaudry [4] – the progress of time can be regarded as the earning of reward and vice versa in the case of non-zero rewards. The key to the proof of this result is a least fixed-point characterization of  $Pr(C \cup_K^I B)$  in MRMs. This result is then lifted to CMRDPs under *HD*-schedulers. By Theorem 2.11, the duality result also applies to *HR*-schedulers.

Consider first CMRDPs for which  $Act(s)$  is a singleton for each state  $s$ , i.e., MRMs. For time-bounded until-formula  $\varphi$  and MRM  $\mathcal{M}$ ,  $Pr^{\mathcal{M}}(s \models \varphi)$  is characterized by a fixed-point equation. This is similar to CTL where appropriate fixed-point characterizations constitute the key towards model checking until-formulas. It suffices to consider time bounds specified by closed intervals since:

$$Pr(s, \Phi \cup_K^I \Psi) = Pr(s, \Phi \cup_{cl(K)}^{cl(I)} \Psi),$$

where  $cl(I)$  denotes the closure of interval  $I$ . A similar result holds for the next-step operator. The result follows from the fact that the probability measure of a basic cylinder set does not change when some of the intervals are replaced by their closure. In the sequel, we assume that intervals  $I$  and  $K$  are compact.

In the sequel, let  $I \ominus x$  denote  $\{t-x \mid t \in I \wedge t \geq x\}$  and  $\mathbf{T}(s, s', x)$  denotes the density of moving from state  $s$  to  $s'$  in  $x$  time units, i.e.,

$$\mathbf{T}(s, s', x) = \mathbf{P}(s, s') \cdot \underline{E}(s) \cdot e^{-\underline{E}(s) \cdot x} = \mathbf{R}(s, s') \cdot e^{-\underline{E}(s) \cdot x}.$$

Here,  $\underline{E}(s) \cdot e^{-\underline{E}(s) \cdot x}$  is the probability density function of the residence time in state  $s$  at instant  $x$ . Let  $Int$  denote the set of all (nonempty) intervals in  $\mathbb{R}_{\geq 0}$ . Let  $L = \{x \in I \mid \rho(s) \cdot x \in K\}$  for closed intervals  $I$  and  $K$ . As we consider MRMs, note that  $\rho$  can be viewed as function  $S \rightarrow \mathbb{R}_{\geq 0}$ . (Strictly speaking,  $L$  is a function depending on  $s$ . As  $s$  is clear from the context, we omit it and write  $L$  instead of  $L(s)$ .) Stated in words,  $L$  is the subset of  $I$  such that the accumulated reward (in  $s$ ) lies in  $K$ .

**Theorem 4.1.** Let  $s \in S$ , interval  $I, K \subseteq \mathbb{R}_{\geq 0}$  and  $\Phi, \Psi$  be CSRL state-formulas. The function  $(s, I, K) \mapsto Pr(s, \Phi \cup_K^I \Psi)$  is the least fixed point of the (monotonic) higher-order operator

$$\Omega : (S \times Int^2 \rightarrow [0, 1]) \rightarrow (S \times Int^2 \rightarrow [0, 1]),$$

where  $\Omega(F)(s, I, K)$  is defined as:

$$\left\{ \begin{array}{ll} 1, & \text{if } s \models \neg\Phi \wedge \Psi \text{ and} \\ & \text{inf } I = \text{inf } K = 0, \\ \int_0^{\sup L} \sum_{s' \in S} \mathbf{T}(s, s', x) \cdot F(s', I \ominus x, K \ominus \rho(s) \cdot x) \, dx, & \text{if } s \models \Phi \wedge \neg\Psi, \\ e^{-\underline{E}(s) \cdot \text{inf } L} + \\ \int_0^{\text{inf } L} \sum_{s' \in S} \mathbf{T}(s, s', x) \cdot F(s', I \ominus x, K \ominus \rho(s) \cdot x) \, dx, & \text{if } s \models \Phi \wedge \Psi, \\ 0, & \text{otherwise.} \end{array} \right.$$

*Proof.* Along the same lines as the proof of Theorem 1 in [2]. Q.E.D.

The above characterisation is justified as follows. If  $s$  satisfies  $\Phi$  and  $\neg\Psi$  (second case), the probability of reaching a  $\Psi$ -state from  $s$  at time  $t \in I$  by earning a reward  $r \in K$  equals the probability of reaching some direct successor  $s'$  of  $s$  within  $x$  time units ( $x \leq \sup I$  and  $\rho(s) \cdot x \leq \sup K$ , that is,  $x \leq \sup L$ ), multiplied by the probability of reaching a  $\Psi$ -state from  $s'$  in the remaining time  $t-x$  while earning a reward of at most  $r - \rho(s) \cdot x$ . If  $s$  satisfies  $\Phi \wedge \Psi$  (third case), the path-formula  $\varphi$  is satisfied if no outgoing transition of  $s$  is taken for at least  $\text{inf } L$  time units<sup>2</sup> (first summand).

Alternatively, state  $s$  should be left before  $\text{inf } L$  in which case the probability is defined in a similar way as for the case  $s \models \Phi \wedge \neg\Psi$  (second summand). Note that  $\text{inf } L = 0$  is possible (if e.g.,  $\text{inf } K = \text{inf } I = 0$ ). In this case,  $s \models \Phi \wedge \Psi$  yields that any path starting in  $s$  satisfies  $\varphi = \Phi \cup_K^I \Psi$  and  $Pr(s, \varphi) = 1$ .

**Definition 4.2** (Dual CMRDP). The *dual* of CMRDP  $\mathcal{M} = (S, Act, \mathbf{R}, L, \rho)$  with  $\rho(s, \alpha) > 0$  for all  $s \in S$  and  $\alpha \in Act$  is the CMRDP  $\mathcal{M}^* = (S, Act, \mathbf{R}^*, L, \rho^*)$  where for  $s, s' \in S$  and  $\alpha \in Act$ :

$$\mathbf{R}^*(s, \alpha, s') = \frac{\mathbf{R}(s, \alpha, s')}{\rho(s, \alpha)} \quad \text{and} \quad \rho^*(s, \alpha) = \frac{1}{\rho(s, \alpha)}.$$

<sup>2</sup> By convention,  $\text{inf } \emptyset = \infty$ .

Intuitively, the transformation of  $\mathcal{M}$  into  $\mathcal{M}^*$  stretches the residence time in state  $s$  under action  $\alpha$  with a factor that is proportional to the reciprocal of reward  $\rho(s, \alpha)$  if  $0 < \rho(s, \alpha) < 1$ . The reward function is changed similarly. Thus, for pairs  $(s, \alpha)$  with  $\rho(s, \alpha) < 1$  the sojourn time in  $s$  is extended, whereas if  $\rho(s, \alpha) > 1$  they are accelerated. For fixed action  $\alpha$ , the residence of  $t$  time units in state  $s$  in  $\mathcal{M}^*$  may be interpreted as the earning of  $t$  reward in  $s$  in  $\mathcal{M}$ , or reversely, earning a reward  $r$  in state  $s$  in  $\mathcal{M}$  corresponds to a residence of  $r$  time units in  $s$  in  $\mathcal{M}^*$ .

The exit rates in  $\mathcal{M}^*$  are given by  $\underline{E}^*(s, \alpha) = \underline{E}(s, \alpha)/\rho(s, \alpha)$ . It follows that  $(\mathcal{M}^*)^* = \mathcal{M}$  and that  $\mathcal{M}$  and  $\mathcal{M}^*$  have the same time-abstract transition probabilities as  $\underline{E}^*(s, \alpha) = 0$  iff  $\underline{E}(s, \alpha) = 0$  and for  $\underline{E}^*(s, \alpha) > 0$ :

$$\mathbf{P}^*(s, \alpha, s') = \frac{\mathbf{R}^*(s, \alpha, s')}{\underline{E}^*(s, \alpha)} = \frac{\mathbf{R}(s, \alpha, s')/\rho(s, \alpha)}{\underline{E}(s, \alpha)/\rho(s, \alpha)} = \frac{\mathbf{R}(s, \alpha, s')}{\underline{E}(s, \alpha)} = \mathbf{P}(s, \alpha, s').$$

Note that a time-abstract scheduler on CMRDP  $\mathcal{M}$  is also a time-abstract scheduler on  $\mathcal{M}^*$  and vice versa, as such schedulers can only base their decisions on time-abstract histories, and the set of time-abstract histories for  $\mathcal{M}$  and  $\mathcal{M}^*$  coincide. Finally, observe that uniformity is *not* maintained by  $*$ :  $\mathcal{M}^*$  is in general not uniform for uniform  $\mathcal{M}$ .

**Definition 4.3** (Dual formula). For state formula  $\Phi$ ,  $\Phi^*$  is the state formula obtained from  $\Phi$  by swapping the time- and reward-bound in each subformula of the form  $\bigcirc_K^I$  or  $\bigcup_K^I$ .

For state-formula  $\Phi$ , let  $\text{Sat}(\Phi) = \{s \in S \mid s \models \Phi\}$ .

**Theorem 4.4** (Duality for MRMs). For MRM  $\mathcal{M} = (S, \mathbf{R}, L, \rho)$  with  $\rho(s) > 0$  for all  $s \in S$  and CSRL state-formula  $\Phi$ :

$$\text{Sat}^{\mathcal{M}}(\Phi) = \text{Sat}^{\mathcal{M}^*}(\Phi^*).$$

*Proof.* By induction on the structure of  $\Phi$ . Let MRM  $\mathcal{M} = (S, \mathbf{R}, L, \rho)$  with  $\rho(s) > 0$  for all  $s \in S$ . We show that for each  $s \in S$  and sets of states  $B, C \subseteq S$ :

$$\text{Pr}^{\mathcal{M}}(s \models C \bigcup_K^I B) = \text{Pr}^{\mathcal{M}^*}(s \models C \bigcup_I^K B).$$

The proof for a similar result for the next-step operator is obtained in an analogous, though simpler way. For the sake of simplicity, let  $I = [0, t]$  and  $K = [0, r]$  with  $r, t \in \mathbb{R}_{\geq 0}$ . The general case can be obtained in a similar way. Let  $s \in C \setminus B$ . From Theorem 4.1 it follows:

$$\text{Pr}^{\mathcal{M}^*}(s \models C \bigcup_I^K B) = \int_{L^*} \sum_{s' \in S} \mathbf{T}^*(s, s', x) \cdot \text{Pr}^{\mathcal{M}^*}(s', C \bigcup_{I \ominus \rho^*(s) \cdot x}^{K \ominus x} B) dx$$

for  $L^* = \{x \in [0, t] \mid \rho^*(s) \cdot x \in [0, r]\}$ , i.e.,  $L^* = [0, \min(t, \frac{r}{\rho^*(s)})]$ . By the definition of  $\mathcal{M}^*$  and  $\mathbf{T}^*(s, s', x) = \mathbf{R}^*(s, s') \cdot e^{-\underline{E}^*(s) \cdot x}$ , the right-hand side equals:

$$\int_{L^*} \sum_{s' \in S} \frac{\mathbf{R}(s, s')}{\rho(s)} \cdot e^{-\frac{\underline{E}(s)}{\rho(s)} \cdot x} \cdot Pr^{\mathcal{M}^*}(s', C \cup_{I \ominus \frac{x}{\rho(s)}}^K B) dx.$$

By substitution  $y = \frac{x}{\rho(s)}$  this integral reduces to:

$$\int_L \sum_{s' \in S} \mathbf{R}(s, s') \cdot e^{-\underline{E}(s) \cdot y} \cdot Pr^{\mathcal{M}^*}(s', C \cup_{I \ominus y}^{K \ominus \rho(s) \cdot y} B) dy,$$

where  $L = [0, \min(\frac{t}{\rho(s)}, r)]$ . Thus, the values  $Pr^{\mathcal{M}^*}(s, C \cup_I^K B)$  yield a solution to the equation system in Theorem 4.1 for  $Pr^{\mathcal{M}}(s, C \cup_K^I B)$ . In fact, these values yield the *least* solution. The formal argument for this latter observation uses the fact that  $\mathcal{M}$  and  $\mathcal{M}^*$  have the same underlying digraph, and hence,  $Pr^{\mathcal{M}}(s, C \cup_K^I B) = 0$  iff  $Pr^{\mathcal{M}^*}(s, C \cup_I^K B) = 0$  iff there is no path starting in  $s$  where  $C \cup B$  holds. In fact, the equation system restricted to  $\{s \in S \mid Pr^{\mathcal{M}}(s, C \cup_K^I B) > 0\}$  has a unique solution. The values  $Pr^{\mathcal{M}^*}(s, C \cup_I^K B)$  and  $Pr^{\mathcal{M}}(s, C \cup_K^I B)$  are least solutions of the same equation system, and are thus equal. Hence, we obtain:

$$\int_L \sum_{s' \in S} \mathbf{T}(s, s', y) \cdot Pr^{\mathcal{M}}(s', C \cup_{K \ominus \rho(s) \cdot y}^{I \ominus y} B) dy$$

which equals  $Pr^{\mathcal{M}}(s \models C \cup_K^I B)$  for  $s \in C \setminus B$ . Q.E.D.

If  $\mathcal{M}$  contains states equipped with a zero reward, the duality result does not hold, as the reverse of earning a zero reward in  $\mathcal{M}$  when considering  $\Phi$  should correspond to a residence of 0 time units in  $\mathcal{M}^*$  for  $\Phi^*$ , which – as the advance of time in a state cannot be halted – is in general not possible. However, the result of Theorem 4.4 applies to some restricted, though still practical, cases, viz. if (i) for each sub-formula of  $\Phi$  of the form  $\bigcirc_K^I \Phi'$  we have  $K = [0, \infty)$ , and (ii) for each sub-formula of the form  $\Phi \cup_K^I \Psi$  either  $K = [0, \infty)$  or  $Sat^{\mathcal{M}}(\Phi) \subseteq \{s \in S \mid \rho(s) > 0\}$ . The intuition is that either the reward constraint (i.e., time constraint) is trivial in  $\Phi$  (in  $\Phi^*$ ), or that zero-rewarded states are not involved in checking the reward constraint. In such cases, let  $\mathcal{M}^*$  be defined by  $\mathbf{R}^*(s, s') = \mathbf{R}(s, s')$  and  $\rho^*(s) = 0$  in case  $\rho(s) = 0$  and  $\mathbf{R}^*$  and  $\rho^*$  be defined as before otherwise.

**Corollary 4.5** (Optimal MD schedulers for cost reachability). Let  $\mathcal{M}$  be a finite *uniform* CMRDP with state space  $S$ ,  $r \in \mathbb{R}_{\geq 0}$  and  $B \subseteq S$ . There exists an MD scheduler  $D$  such that for any  $s \in S$ :

$$Pr^D(s \models \diamond_{\leq r} B) = Pr_{HR}^{\max}(s \models \diamond_{\leq r} B).$$

*Proof.* Let  $\mathcal{M}$  be a uniform CMRDP. By Theorem 2.9 it follows:

$$\sup_{D \in HD} Pr_{\mathcal{M}}^D(s \models \diamond^{\leq t} B) = \sup_{D \in MD} Pr_{\mathcal{M}}^D(s \models \diamond^{\leq t} B).$$

Observe that there is a one-to-one relationship between schedulers of  $\mathcal{M}$  and of its dual  $\mathcal{M}^*$  as  $\mathcal{M}$  and  $\mathcal{M}^*$  have the same time-abstract scheduler for any class  $X$  as defined before. Moreover, for  $HD$ -scheduler  $D$ , the dual of MRM  $\mathcal{M}_D$  is identical to the induced MRM of the dual of  $\mathcal{M}$ , i.e.,  $(\mathcal{M}_D)^* = (\mathcal{M}^*)_D$ . Thus:

$$\sup_{D \in HD} Pr_{\mathcal{M}}^D(s \models \diamond^{\leq t} B) = \sup_{D^* \in HD} Pr_{\mathcal{M}^*}^{D^*}(s \models \diamond^{\leq t} B).$$

Applying Theorem 4.4 to  $\mathcal{M}^*$  yields:

$$\sup_{D \in HD} Pr_{\mathcal{M}}^D(s \models \diamond^{\leq t} B) = \sup_{D^* \in HD} Pr_{\mathcal{M}^*}^{D^*}(s \models \diamond_{\leq r} B),$$

and by an analogous argument for  $MD$ -schedulers:

$$\sup_{D \in MD} Pr_{\mathcal{M}}^D(s \models \diamond^{\leq t} B) = \sup_{D^* \in MD} Pr_{\mathcal{M}^*}^{D^*}(s \models \diamond_{\leq r} B).$$

Thus:

$$\sup_{D \in HD} Pr_{\mathcal{M}^*}^D(s \models \diamond_{\leq r} B) = \sup_{D \in MD} Pr_{\mathcal{M}^*}^D(s \models \diamond_{\leq r} B).$$

In addition, Theorem 2.11 asserts:

$$\sup_{D \in HD} Pr_{\mathcal{M}}^D(s \models \diamond_{\leq r} B) = \sup_{D \in HR} Pr_{\mathcal{M}}^D(s \models \diamond_{\leq r} B)$$

and hence  $\sup_{D^* \in MD} Pr_{\mathcal{M}^*}^{D^*}(s \models \diamond_{\leq r} B)$  coincides with the suprema for the probability to reach  $B$  within reward bound  $r$  under all  $HD$ -,  $HR$ - and  $MD$ -schedulers. As  $MR$ -schedulers are between  $HR$ - and  $MD$ -schedulers, the stated result follows. Q.E.D.

Unfortunately, this result does not imply that the algorithm in [3] applied on  $\mathcal{M}^*$  yields the optimal result for the event  $\diamond_{\leq r} B$ , as  $\mathcal{M}^*$  is not guaranteed to be uniform whereas the algorithm ensures optimality only for uniform CTMDPs.

We conclude this note by a duality result for arbitrary CMRDPs.

**Corollary 4.6** (Duality for CMRDPs). For CMRDP  $\mathcal{M} = (S, Act, \mathbf{R}, L, \rho)$  with  $\rho(s, \alpha) > 0$  for all  $s \in S$  and  $\alpha \in Act$ , and CSRL state-formula  $\Phi$ :

$$Sat^{\mathcal{M}}(\Phi) = Sat^{\mathcal{M}^*}(\Phi^*).$$

*Proof.* By induction on the structure of  $\Phi$ . Let CMRDP  $\mathcal{M} = (S, Act, \mathbf{R}, L, \rho)$  with  $\rho(s, \alpha) > 0$  for all  $s \in S$  and  $\alpha \in Act$ . Consider  $\Phi = \mathbb{P}_{\leq p}(C U_K^I B)$ . The proof for bounds of the form  $\geq p$ , and for the next-step operator are similar. From the semantics of CSRL it follows:

$$s \models_{\mathcal{M}} \mathbb{P}_{\leq p}(C U_K^I B) \text{ iff } \sup_{D \in HR} Pr_{\mathcal{M}}^D(s \models C U_K^I B) \leq p.$$

In a similar way as stated in the third item of Theorem 2.11 it follows:

$$\sup_{D \in HR} Pr_{\mathcal{M}}^D(s \models C U_K^I B) = \sup_{D \in HD} Pr_{\mathcal{M}}^D(s \models C U_K^I B).$$

$\mathcal{M}$  and  $\mathcal{M}^*$  have the same time-abstract  $HD$ -schedulers and  $(\mathcal{M}_D)^* = \mathcal{M}_D^*$ . Theorem 4.4 yields:

$$\sup_{D \in HD} Pr_{\mathcal{M}}^D(s \models C U_K^I B) = \sup_{D^* \in HD} Pr_{\mathcal{M}^*}^{D^*}(s \models C U_I^K B).$$

As  $HD$ - and  $HR$ -schedulers are indistinguishable for events of the form  $C U_K^I B$  (the proof of this fact is analogous to that of Theorem 2.11), it follows:

$$\sup_{D^* \in HD} Pr_{\mathcal{M}^*}^{D^*}(s \models C U_I^K B) = \sup_{D^* \in HR} Pr_{\mathcal{M}^*}^{D^*}(s \models C U_I^K B).$$

Thus:

$$s \models_{\mathcal{M}} \mathbb{P}_{\leq p}(C U_K^I B) \text{ iff } s \models_{\mathcal{M}^*} \mathbb{P}_{\leq p}(C U_I^K B).$$

Q.E.D.

## 5 Epilogue

In this paper we have brought together results on the use of the logic CSRL and time and reward duality for MRMs [1], with recent results on reachability in CTMDPs [3]. This leads to a duality result for CMRDPs, as well as to the existence of optimal  $MD$  schedulers for cost reachability in uniform CMRDPs.

### Acknowledgement

We thank Gethin Norman (Oxford) for his comments on an earlier version of this paper. This work has been partially funded by the bilateral NWO-DFG Project Validation of Stochastic Systems 2 (VOSS2).

## References

- [1] C. Baier, B.R. Haverkort, H. Hermanns and J.-P. Katoen. On the logical characterisation of performability properties. In *Automata, Languages, and Programming (ICALP)*, LNCS 1853: 780–792, Springer, 2000.

- [2] C. Baier, B.R. Haverkort, H. Hermanns and J.-P. Katoen. Model-checking algorithms for continuous-time Markov chains. *IEEE Trans. on Softw. Eng.*, **29**(6): 524–541, 2003.
- [3] C. Baier, H. Hermanns, J.-P. Katoen, B.R. Haverkort. Efficient computation of time-bounded reachability probabilities in uniform continuous-time Markov decision processes. *Th. Comp. Sc.* **345**(1): 2–26, 2005.
- [4] M.D. Beaudry. Performance-related reliability measures for computing systems. *IEEE Trans. on Comp. Sys.*, **27**(6): 540–547, 1978.
- [5] J. L. Bruno, P. J. Downey, and G. N. Frederickson. Sequencing tasks with exponential service times to minimize the expected flow time or makespan. *J. of the ACM*, 28(1): 100-113, 1981.
- [6] P. Buchholz. Exact and ordinary lumpability in finite Markov chains. *J. of Applied Probability*, **31**: 59–75, 1994.
- [7] M. Neuhäüßer and J.-P. Katoen. Bisimulation and logical preservation for continuous-time Markov decision processes. In *Concurrency Theory (CONCUR)*, LNCS, Springer, 2007.
- [8] M.L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 1994.
- [9] N. Wolovick and S. Johr. A characterization of meaningful schedulers for continuous-time Markov decision processes. In *Formal Modeling and Analysis of Timed Systems (FORMATS)*, LNCS 4202: 352–367, 2006.