DOI 10.1007/s00165-016-0411-1
BCS © 2017
Formal Aspects of Computing (2017) 29: 629–649

Formal Aspects of Computing

Cost vs. time in stochastic games and Markov automata

Hassan Hatefi[1], Ralf Wimmer[2], Bettina Braitling[2], Luis María Ferrer Fioriti[1], Bernd Becker[2], and Holger Hermanns[1]

[1] Chair for Dependable Systems and Software, Saarland University, Saarbrücken, Germany
[2] Chair of Computer Architecture, Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany

Abstract. Costs and rewards are important tools for analysing quantitative aspects of models like energy consumption and costs of maintenance and repair. Under the assumption of transient costs, this paper considers the computation of expected cost-bounded rewards and cost-bounded reachability for Markov automata and Markov games. We provide a fixed point characterization of this class of properties under early schedulers. Additionally, we give a transformation to expected time-bounded rewards and time-bounded reachability, which can be computed by available algorithms. We prove the correctness of the transformation and show its effectiveness on a number of Markov automata case studies.

Keywords: Markov automata, Stochastic games, Expected rewards, Cost bounds, Time bounds.

1. Introduction

Markov automata (MA) [EHZ10] constitute a compositional modelling formalism for concurrent stochastic systems. They generalise discrete-time Markov chains (DTMCs), Markov decision processes (MDPs), probabilistic automata (PA) [Seg95], continuous-time Markov chains (CTMCs), and interactive Markov chains (IMCs) [Her02]. Markov automata form the semantic foundation of, among others, dynamic fault trees [BCS10], stochastic activity networks, and generalised stochastic Petri nets (GSPNs) [EHKZ13]. Compositional modelling for MA [TKvdPS12] is supported by the MAMA tool set [GHH+13, GHH+14], also providing access to effective model analysis via the IMCA tool [GHKN12]. That analysis follows the principles of model checking [BK08]. Concretely speaking, algorithms for model checking time-bounded reachability and continuous stochastic logic (CSL) [HH12] as well as long-run average and expected reachability times [GHH+13, GHH+14] are supported.

Part of this work was done while Ralf Wimmer was a visiting professor at Saarland University, Saarbrücken, Germany. This work was partly supported by the CDZ project CAP (GZ 1023) and by the German Research Foundation (DFG) as part of the Cluster of Excellence BrainLinks/BrainTools (EXC 1086).

Correspondence and offprint requests to: R. Wimmer, Email: wimmer@informatik.uni-freiburg.de

Stochastic games (SGs) [Sha53, NS03] constitute a generalisation of Markov automata. They can be considered as Markov games with a single player and certain restrictions on the allowed nondeterminism [BFH+15]. SGs are turn-based games and allow, e. g., to model interactive stochastic systems in a hostile environment. Here, the user is one of the players, choosing actions to execute. The environment is the other player. The two players can follow independent strategies to reach their goals. For instance, the user can try to maximize the probability to reach a target state within a given time bound, while the environment is hostile and tries to keep the user away from the target [BFK+13].

The main reason we study SGs in this paper comes from their application area. First of all, their ability to model controller-environment games makes them potentially interesting and possibly useful for applications in control domain. One may then utilise the result of this paper to compute the optimal expected cost-bounded rewards for SGs using the solution given in [BFH+15]. The current practical limitation is that, in contrary to MAs, SGs in general are not compositional [GBK16]. Therefore, it is not possible to benefit from compositional model construction to build highly complex models from their smaller building blocks. This is the main reason we do not have any SG model in our experiments. Instead, to motivate using SGs and to demonstrate their modelling capacity we provide the game semantics for the "dynamic power management system", one of our MA case studies which will be discussed in Sect. 4.

Secondly, we have previously exploited the expressive power of SGs in an abstraction refinement framework for MA [BFH+15]. There, an MA is abstracted into a smaller SG such that analyzing the game allows to deduce safe bounds on, e. g., (time-bounded) reachability probabilities [BFH+14] and expected rewards [BFH+15] of the original MA. If the obtained bounds are too far apart, the SG is refined until the requested precision is obtained. It is thereby the fact that any kinds of analysis technique for SGs can in principle be utilised for MAs via the abstraction refinement framework. This was for us promising enough to generalise our results from MAs to SGs.

Apart from timing-related properties, there is an immensely large spectrum of potential applications that ask for the integration of cost-related modelling and analysis. Costs, or dually rewards, are especially convenient to reflect economical implications, power consumption, wear and abrasion, or other quantitative information. Therefore MA have lately been extended to MRA, Markov reward automata. In MRA, states and transitions can be equipped with rewards or costs, which are accumulated as time advances and as transitions are taken. Algorithms for computing the long-run average reward, for the expected cumulative reward until reaching a set of goal states, and for the expected cumulative reward until a certain time bound are known and implemented [GTH+14]. Effective abstraction and refinement strategies for MRA have also been introduced [BFH+15], working on stochastic reward game abstractions of MRA.

Contribution In this paper, we turn our attention to properties that relate multiple dimensions of costs or rewards. In particular, we enable the computation of expected cumulative rewards until exceeding a cost bound, both for Markov reward automata and stochastic reward games. This can, for instance, answer questions of central importance for energy-harvesting battery-powered missions:

Under a given initial budget, what is the maximum probability of the battery running dry, or how many tasks can maximally be expected to be carried out by the battery?

To answer such questions we give a fixed point characterisation of expected cost-bounded rewards and a transformation for stochastic games from cost- to time-bounded rewards. This transformation supports arbitrary non-negative transient costs. If the transformation is applied to a Markov automaton, the result is again a Markov automaton. After the transformation, arbitrary algorithms for expected time-bounded rewards like [GTH+14, BFH+15] can be applied to compute expected cost-bounded rewards.

In order to develop our contribution, we take inspiration from various sources, especially from the domain of continuous-time Markov decision processes (CTMDPs). This encompasses works on necessary and sufficient criteria for optimality with respect to time-bounded rewards [Mil68], and algorithms to compute optimal time-bounded rewards using uniformisation [BS11, BHHK15]. Instantaneous transition rewards have been added to the CTMC setting as well [CKKP05].

Our work is strongly influenced by the study of the duality between time and costs in CTMDPs under time-abstract strategies [BHHK08], built up on the earlier work in the setting of CTMCs [BHHK00]. We extend it in various dimensions: Our technique supports zero-cost states, where previously only strictly positive costs were allowed. We optimise over time-dependent strategies, which are a superclass of time-abstract ones. We extend the setting to expected reward analysis on two-player games with discrete and continuous distributions, which is also an improvement over [Fu14a, Fu14b]. And finally our analysis technique works for any kind of models, not only uniform ones.

A preliminary version of this paper appeared as [HBW+15]. Compared to that version, we have made the paper more self-contained by extending the section on foundations and by adding detailed proofs for all main theorems. In particular, we elaborate on measure-theoretic concepts and measurable strategies in SGs. We have furthermore added results concerning the inclusion of instantaneous costs. This indeed would make the computation of cost-bounded expected rewards NP-hard, while without instantaneous costs, the algorithms run in polynomial time.

Structure of the paper In the following section, we introduce the necessary foundations. Section 3 describes the fixed point characterisation of optimal expected cost-bounded rewards and the transformation from cost to time bounds. We report on experimental results in Sect. 4 and conclude the paper in Sect. 5. Detailed proofs of the main propositions are contained in the appendix of this paper.

2. Foundations

The real numbers are denoted by $\mathbb{R}$, $\mathbb{R}_{\geq 0}$ is the set of non-negative real numbers, and $\mathbb{R}^{\infty}_{\geq 0} := \mathbb{R}_{\geq 0} \cup \{\infty\}$. Accordingly $\mathbb{R}_{>0}$, $\mathbb{R}^{\infty}_{>0}$ etc. are used.

Let V be a finite (or countably infinite) set. A *discrete probability distribution* over V is a function $\mu : V \to [0, 1]$ such that $\sum_{v \in V} \mu(v) = 1$. We denote the set of probability distributions over V by $\mathrm{Distr}(V)$. The support of a distribution $\mu \in \mathrm{Distr}(V)$ is the set $\mathrm{supp}(\mu) = \{v \in V \mid \mu(v) > 0\}$. A distribution $\mu \in \mathrm{Distr}(V)$ such that, for some $v \in V$, $\mu(v) = 1$ is called a Dirac distribution and denoted by Δ_v.

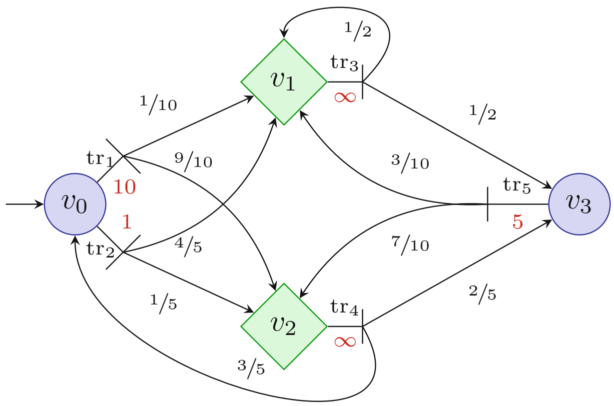
Additionally we need *continuous probability distributions* over $\mathbb{R}_{\geq 0}$. They can either be specified in terms of their probability density function or their cumulative distribution function. A function $p : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a probability density function on $\mathbb{R}_{\geq 0}$ if $\int_0^{\infty} p(x)\,\mathrm{d}x = 1$. Its cumulative distribution function is then given by $P(t) = \int_0^t p(x)\,\mathrm{d}x$. A central role plays the *exponential distribution*. For a parameter $\lambda \in \mathbb{R}_{\geq 0}$, its density is given by $\exp_{\lambda}(x) = \lambda \cdot \mathrm{e}^{-\lambda \cdot x}$ and its cumulative distribution by $\mathrm{Exp}_{\lambda}(x) = 1 - \mathrm{e}^{-\lambda \cdot x}$.

Definition 2.1 (*Stochastic game*) A *stochastic* (*continuous-time two-player*) *game* (SG) is a tuple $\mathcal{G} = \big(V, (V_1, V_2), v_{\mathrm{init}}, T\big)$ such that $V = V_1 \uplus V_2$ is the finite set of states, $v_{\mathrm{init}} \in V$ is the initial state, and $T \subseteq V \times \mathbb{R}^{\infty}_{>0} \times \mathrm{Distr}(V)$ is the transition relation.

V_1 and V_2 are the states of player 1 and player 2, respectively; we also denote them as V_1- and V_2-states. Transitions $(v, \lambda, \mu) \in T$ with rate $\lambda < \infty$ are called *Markovian*, transitions with infinite rate *probabilistic*. We denote the set of Markovian and probabilistic transitions by T_{M} and T_{P}, respectively. We use $T_{\mathrm{M}}(v)$ and $T_{\mathrm{P}}(v)$ to refer to the set of Markovian and probabilistic transitions available at state v. Then, $T(v) = T_{\mathrm{M}}(v) \uplus T_{\mathrm{P}}(v)$ is the set of all available transitions of v. We assume that $T(v) \neq \emptyset$ for all $v \in V$.

The game starts in state v_{init}. If the current state is $v \in V_1$, then it is player 1's turn, otherwise player 2's. The current player chooses a transition $(v, \lambda, \mu) \in T(v)$ for leaving state v. The rate $\theta_{\mathrm{rate}}\big((v, \lambda, \mu)\big) = \lambda \in \mathbb{R}^{\infty}_{\geq 0}$ determines how long we stay at v, whereas $\theta_{\mathrm{distr}}\big((v, \lambda, \mu)\big) = \mu \in \mathrm{Distr}(V)$ gives us the discrete probability distribution which leads to the successor states. If $\lambda = \infty$, the transition is taken instantaneously. Otherwise, λ is taken as the parameter of an exponential distribution that determines the sojourn time in the current state, i. e. the probability that the transition tr is taken within time t is given by $\mathrm{Exp}_{\lambda}(t) = 1 - \mathrm{e}^{-\lambda \cdot t}$. The probability that the transition is taken within $t \geq 0$ time units and leads to state $v' \in V$, is accordingly given by $\mu(v') \cdot \mathrm{Exp}_{\lambda}(t) = \mu(v') \cdot (1 - \mathrm{e}^{-\lambda \cdot t})$. For conciseness, we write λ_{tr} instead of $\theta_{\mathrm{rate}}(\mathrm{tr})$ and μ_{tr} instead of $\theta_{\mathrm{distr}}(\mathrm{tr})$ for $\mathrm{tr} \in T$.

Example 2.1 Figure 1a shows an example of a stochastic game. It consists of two player 1 states (drawn as circles) and two player 2 states (drawn as diamonds). The exit rates of the transitions $\mathrm{tr}_1, \ldots, \mathrm{tr}_5$ are written in red. The game starts in v_0. Player 1 chooses one of the outgoing transitions $\{\mathrm{tr}_1, \mathrm{tr}_2\}$, say tr_1. The probability to stay in v_0 for at most t time units is then given by $1 - \mathrm{e}^{-10 \cdot t}$. When the transition fires, we move to v_1 with probability 0.1 and to v_2 with probability 0.9; say v_1 is the successor state. There it is player 2's turn. As only one outgoing transition is available, namely tr_3, and its exit rate is ∞, v_1 is left immediately, either to v_1, again, or to v_3, both with probability 0.5.



(a)

	tr$_1$	tr$_2$	tr$_3$	tr$_4$	tr$_5$	v_0	v_1	v_2	v_3
c	5	0	0	0	3				
ρ_t	1	3	0	0	1				
ρ_i	4	1	1	5	2				
ρ_f						1	0	2	3

(b)

Fig. 1. An example of a stochastic game with costs and rewards

Markov automata (MA) [EHZ10] are a special type of stochastic games with a single player and without a nondeterministic choice between different Markovian transitions at one state. The reason for this restriction is that Markov automata are designed to be a compositional formalism, i. e. the MA for a system consisting of several components can be constructed from the MA of the individual components.

Definition 2.2 (*Markov automaton*) A *Markov automaton* (MA) is a stochastic game $\mathcal{M} = \big(V, (V, \emptyset), v_{\text{init}}, T\big)$ such that $|\ T_{\text{M}}(v)\ | \leq 1$ holds for all $v \in V$. We simply write $\mathcal{M} = (V, v_{\text{init}}, T)$ for a Markov automaton $\mathcal{M}$.

In this paper we only consider *closed* Markov automata which are not subject to further composition operations. In this case, it is standard for Markov automata to make an *urgency assumption*: Since nothing prevents probabilistic transitions from happening instantaneously and the probability that a Markovian transition is taken without delay is zero, probabilistic transitions take precedence over Markovian transitions. Therefore we assume for MA that Markovian transitions have been removed from all states which also exhibit an outgoing probabilistic transition.

Paths through stochastic games The dynamics of an SG is specified by paths. A *path* $\pi \in (V \times \mathbb{R}_{\geq 0} \times T)^\omega$ is an infinite sequence of states, sojourn times, and transitions. A *history* is such a sequence of finite length, i. e. $h \in (V \times \mathbb{R}_{\geq 0} \times T)^\star \times V$. It ends in a state which we call the terminal state of the history. We usually write $v \xrightarrow{t,\text{tr}}$ instead of $(v, t, \text{tr}) \in (V \times \mathbb{R}_{\geq 0} \times T)$. We use H and Π to denote the set of histories and paths, respectively. Given history h, its length, denoted by $|\ h\ |$, is the of number transitions executed in h. The length is infinite for a path. We write $last(h)$ for the last, i. e. terminal, state of h. Given history $h = v_0 \xrightarrow{t_0,\text{tr}_0} v_1 \xrightarrow{t_1,\text{tr}_1} \cdots v_n$ and $0 \leq i < n$, v_i is the $(i+1)$-th state of π, denoted by $h[i]$; t_i is the time of staying at v_i, denoted by $h\langle i\rangle$; and $\text{trans}(h[i]) = \text{tr}_i$ is the executed transition at v_i. Note that v_i must be left instantaneously, i. e. $h\langle i\rangle = 0$, if $\text{trans}(h[i])$ has an infinite rate. For $0 \leq i \leq j \leq n$, the sub-history $v_i \xrightarrow{t_i,\text{tr}_i} \cdots v_j$ is denoted by $h[i \cdot\cdot j]$. The exact same notations are carried over into paths.

Measurability A collection of paths, also called an *event*, describes a specific behaviour pattern of an SG. Such an event is *measurable* if, intuitively speaking, we can asses how likely it will happen. In mathematical analysis and in particular measure theory, a set of measurable events that is closed under countably many set operations is described by a σ-algebra. Here we briefly explain how we define a σ-algebra over the set of histories and paths of an SG. They are constructed in a modular way using the concept of a *product* σ*-algebra* [ADD99, Def. 2.6.1], here denoted by the $\otimes$ operator. Let 2^V and 2^T be the power sets of V and T, respectively, and $\mathcal{B}(\mathbb{R}_{\geq 0})$ be the Borel σ-algebra over $\mathbb{R}_{\geq 0}$. Then, $\mathcal{F} := 2^V \otimes \mathcal{B}(\mathbb{R}_{\geq 0}) \otimes 2^T$ denotes the σ-algebra over the set of path steps, i. e. $V \times \mathbb{R}_{\geq 0} \times T$. Accordingly, the σ-algebra over the set of n-step histories is referred to as $\mathcal{H}_n := (\bigotimes_{i=1}^n \mathcal{F}) \otimes 2^V$, for $n \in \mathbb{N}$. For the set of infinite paths, the σ-algebra is obtained by applying the cylinder set construction [ADD99, Def. 2.7.1]. Briefly speaking, let B_n be a subset of $\mathcal{H}_n$, then the cylinder of base B_n

is described as $\mathrm{Cyl}(B_n) := \{\pi \in \Pi \mid \pi[0 \cdots n] \in B_n\}$. Cylinder $\mathrm{Cyl}(B_n)$ is *measurable* iff its base is measurable, that is to say iff $B_n \in \mathcal{F}_n$. Then, $\mathcal{F}_\Pi$, the σ-algebra over the set of paths, is the smallest σ-algebra generated by the class of all measurable cylinders, i. e. $\bigcup_{n=0}^{\infty}\{\mathrm{Cyl}(B_n) \mid B_n \in \mathcal{F}_n\}$. Together with the set of infinite paths, it constitutes a measurable space $(\Pi, \mathcal{F}_\Pi)$, on which the probability measure will be defined. For more details see, for instance, [Neu10, Sec. 2.5.4]).

Strategies SGs may exhibit nondeterminism occurring at a state with more than one outgoing transition. In such a state, it is not clear a priori which transition is taken during the execution of the model. In this case, the player who controls the state makes a decision on which transition is going to be taken. The decision is made by a function that is called *strategy* (or policies or schedulers). Each player follows her own strategy in order to accomplish her specific goal. Once the players have fixed their strategies, the SG does not exhibit nondeterminism anymore and its behaviour is purely stochastic. Therefore, a unique probability measure can be defined on the measurable space $(\Pi, \mathcal{F}_\Pi)$.

Strategies can make use of many details from what has been visited thus far to resolve the nondeterminism. It can be, for instance, the states and the transitions that have been observed, or their execution time. The strategies are classified according to the amount of information they employ to make the decision. The most general class exploits the complete history to decide between available transitions at any state. The class is further pruned by considering only the strategies that are *measurable* which is known as *generic measurable strategies* [WJ06, Joh08, Neu10].

Definition 2.3 (*Generic measurable strategy*) A *generic strategy* of player $i = 1, 2$ is a function $\sigma_i : H \to \mathrm{Distr}(T)$ such that for every $h \in H$ it holds that $\mathrm{supp}(\sigma(h)) \subseteq T(last(h))$. Strategy σ is *generic measurable* iff $\{h \mid \sum_{\mathrm{tr}\in T} \sigma(h)(\mathrm{tr}) \in B \wedge \mid h \mid= n\} \in \mathcal{H}_n$ for every $n \in \mathbb{N}$ and $B \in \mathcal{B}([0, 1])$. We use Σ_1 and Σ_2 to denote the set of all generic measurable strategies of player 1 and 2, respectively.

The support restriction of Definition 2.3 indicates that, for every history h, the strategy can only select the outgoing transitions from the last state of h. We indeed apply σ_i to the histories that end in player i's states. Measurability of a strategy intuitively means that it never resolves the nondeterminism for histories such that in any way they induce non-measurable sets. It provides an essential basis for the unique probability measure on $(\Pi, \mathcal{F}_\Pi)$. Although generic measurable strategies are theoretically important, they might be complex and therefore computationally intractable. However, for most of the objectives it is sufficient and even desirable to consider simpler classes of strategies.

We briefly introduce some classes of strategies that are important for our purpose. *Total-time dependent positional deterministic* (TTPD) strategies are in the form $\sigma : V \times \mathbb{R}_{\geq 0} \to T$. They use the total time which has passed since starting in the initial state up to the current state to select an outgoing transition of the current state. Their decision is always deterministic rather than being randomised. TTPD strategies can be easily extended to the more general *total-cost dependent positional deterministic* (TCPD) strategies, where the role of time is taken by costs that have been accumulated since the start of the system. Both classes are important for the properties we consider in this paper [Neu10, Fu14a, Fu14b].

There is yet another dimension for strategy classification that leads to the introduction of *early* vs. *late* strategies [NZ10]. Early strategies make their decision upon entering a state. The decision may not be changed afterwards while residing in the state. On the contrary, late strategies can change their decision at any time while residing in the state. It is well understood that for time-bounded reachability objectives, late strategies are superior to early ones. However, early strategies are the ones that naturally emerge from compositional model construction, e. g. from parallel composition of Markov automata. Hence, we only consider *early* strategies in this paper. We propose a transformation in Sect. 3 that preserves the optimal expected cost-bounded rewards under the class of early strategies. We strongly believe that the transformation preserves the exact same objective under the more general class of late strategies. The claim is posed as a conjecture.

Probability measure Given strategies σ_1, σ_2 for both players and a state $v \in V$, it is possible to define a unique probability measure $\mathrm{Pr}_{v,\sigma_1,\sigma_2}$ on $(\Pi, \mathcal{F}_\Pi)$. For a measurable event $E \in \mathcal{F}_\Pi$, $\mathrm{Pr}_{v,\sigma_1,\sigma_2}(E)$ denotes the probability of observing paths in E, starting from state v, given that player 1 and player 2 play with strategies σ_1 and σ_2, respectively. The main building block is the probability measure on single path steps. The probability measure is then defined in a recursive manner using the product measure theorem [ADD99, Thm. 2.6.2] and the Ionescu–Tulcea extension theorem [ADD99, Thm. 2.7.2]. The construction extends the existing techniques used for MAs and IMCs. The details are highly technical and omitted here; for more information see, e. g. [HH12, Neu10, Joh08].

Zenoness It may happen that an SG contains an end component [BK08, Def. 10.117] consisting of probabilistic transitions only. Such an end component leads to the existence of sets of infinite paths π with finite sojourn times and non-zero probability, i. e. $\lim_{n\to\infty}\sum_{i=0}^{n}\pi\langle i\rangle < \infty$. This phenomenon is known as *Zenoness*. Since such behaviour is unrealistic, we assume that the SGs under consideration are non-Zeno, i. e. that they do not contain any probabilistic end component. Formally, an SG is non-Zeno iff

$$\Pr_{v,\sigma_1,\sigma_2}\left(\left\{\pi \in \Pi : \lim_{n\to\infty}\sum_{i=0}^{n}\pi\langle i\rangle < \infty\right\}\right) = 0$$

holds for all states $v \in V$ and all strategies $\sigma_1 \in \Sigma_1$ and $\sigma_2 \in \Sigma_2$.

For more on strategies and on SGs in general we refer to [Sha53, NS03, BFK$^+$13].

Costs and rewards We now extend stochastic games by costs and rewards to analyse properties like "What is the maximal reward one can earn when the accumulated cost is bounded by b?"

Definition 2.4 (*Cost and reward structures*) Let $\mathcal{G}$ be a stochastic game as in Definition 2.1. A *cost function* $c : T \to \mathbb{R}_{\geq 0}$ assigns a non-negative cost rate to each transition. A *reward structure* ρ is a triple $\rho = (\rho_{\mathrm{t}}, \rho_{\mathrm{i}}, \rho_{\mathrm{f}})$ of functions $\rho_{\mathrm{t}}, \rho_{\mathrm{i}} : T \to \mathbb{R}_{\geq 0}$, and $\rho_{\mathrm{f}} : V \to \mathbb{R}_{\geq 0}$; ρ_{t} is the transient reward rate, ρ_{i} the instantaneous reward, and ρ_{f} the final reward.

For a transition $\mathrm{tr} = (v, \lambda, \mu) \in T$, costs and transient rewards are granted per time unit, i. e. residing in v for t time units before taking transition tr causes a cost of $t \cdot c(\mathrm{tr})$, and a transient reward of $t \cdot \rho_{\mathrm{t}}(\mathrm{tr})$ is granted. In contrast, the instantaneous reward $\rho_{\mathrm{i}}(\mathrm{tr})$ is granted for taking the transition tr. The final reward is granted for the state reached when the maximal cost has been spent. This allows, e. g. to consider cost-bounded reachability probabilities as a special case of expected cost-bounded rewards (for more details, see below).

Cost and reward of paths Given a history $h = v_0 \xrightarrow{t_0,\mathrm{tr}_0} v_1 \xrightarrow{t_1,\mathrm{tr}_1} \cdots v_{n-1} \xrightarrow{t_{n-1},\mathrm{tr}_{n-1}} v_n$, its cost is defined as $\mathrm{cost}(h) := \sum_{i=0}^{n-1} c(\mathrm{tr}_i)\cdot t_i$. The cost can be extended to a path π by $\mathrm{cost}(\pi) := \lim_{n\to\infty}\mathrm{cost}(\pi[0 \cdot\cdot\, n])$. The cumulative reward of a history or a path can be defined in a similar way, i. e. $\mathrm{crew}(h) := \sum_{i=0}^{n-1}\left(\rho_{\mathrm{t}}(\mathrm{tr}_i)\cdot t_i + \rho_{\mathrm{i}}(\mathrm{tr}_i)\right)$ and $\mathrm{crew}(\pi) := \lim_{n\to\infty}\mathrm{crew}(\pi[0 \cdot\cdot\, n])$. Furthermore we define the *cost-bounded reward* of π by

$$cbr^{\mathcal{G}}_{\rho,c}(\pi, b) := \begin{cases} \mathrm{crew}(\pi), & \text{if } \mathrm{cost}(\pi) \leq b,\\ \mathrm{crew}(\pi[0 \cdot\cdot\, n^*]) + \frac{b-\mathrm{cost}(\pi[0\cdot\cdot n^*])}{c(\mathrm{tr}_{n^*})}\cdot\rho_{\mathrm{t}}(\mathrm{tr}_{n^*}) & \\ \qquad + \rho_{\mathrm{f}}(\pi[n^*]), & \text{if } \mathrm{cost}(\pi) > b, \end{cases}$$

where $n^* \in \mathbb{N}$ is the index of the state along path π such that $\mathrm{cost}(\pi[0 \cdot\cdot\, n^*]) \leq b$ and $\mathrm{cost}(\pi[0 \cdot\cdot\, n^* + 1]) > b$. More precisely, the cost exceeds b after residing $\frac{b-\mathrm{cost}(\pi[0\cdot\cdot n^*])}{c(\mathrm{tr}_{n^*})}$ time units in the n^*-th state of the path, and thereby the state is subject to the final reward. Note that such an index exists, provided that $\mathrm{cost}(\pi) > b$.

Example 2.2 Consider again the stochastic game in Fig. 1a. We extend it by the cost function and reward structure shown in Fig. 1b. Now consider the path $\pi = v_0 \xrightarrow{3,\mathrm{tr}_1} v_1 \xrightarrow{0,\mathrm{tr}_3} v_3 \xrightarrow{2,\mathrm{tr}_5} v_2 \xrightarrow{0,\mathrm{tr}_4} v_0 \to \cdots$ and assume the cost bound $b = 20$. The cost incurring in v_0 before taking tr_1 is $5\cdot 3 = 15$. Since tr_3 is probabilistic, no cost incurs in v_1. In v_3 we have costs $3\cdot 2 = 6$. Therefore the cost bound is reached while staying in v_3, after $1/3\cdot(20-15) = 5/3$ time units. We then have $n^* = 2$. Since v_3 is the state in which the cost bound is reached, we additionally get its final reward $\rho_{\mathrm{f}}(v_3) = 3$. The cost-bounded reward for this path is accordingly $cbr^{\mathcal{G}}_{\rho,c}(\pi, 20) = (3\cdot 1+4)+(0\cdot 0+1)+(5/3\cdot 1)+3 = 12\,2/3$.

Given strategies $\sigma_1 \in \Sigma_1$ and $\sigma_2 \in \Sigma_2$ we can define the *expected cost-bounded reward* (ECR) as the expectation of cbr:

$$\mathbb{E}cbr^{\sigma_1,\sigma_2}_{\mathcal{G},\rho,c}(v, b) := \int_{\pi\in\Pi} cbr^{\mathcal{G}}_{\rho,c}(\pi, b)\,\Pr_{v,\sigma_1,\sigma_2}(\mathrm{d}\pi).$$

The two players can independently try to maximise or minimise the reward earned until the cost bound is reached. Hence, for $\mathrm{opt}_1, \mathrm{opt}_2 \in \{\inf, \sup\}$ we define the *optimal* expected cost-bounded reward by

$$\mathbb{E}cbr^{\mathrm{opt}_1,\mathrm{opt}_2}_{\mathcal{G},\rho,c}(v, b) := \operatorname*{opt_1}_{\sigma_1\in\Sigma_1} \operatorname*{opt_2}_{\sigma_2\in\Sigma_2} \mathbb{E}cbr^{\sigma_1,\sigma_2}_{\mathcal{G},\rho,c}(v, b).$$

Two important classes of properties can be considered as special cases of expected cost-bounded rewards:

For *time-bounded rewards*, denoted by random variable tbr, the time is limited during which reward is collected. This corresponds to using the constant **1**-function as cost. We therefore define $\mathbb{E}\mathrm{tbr}^{\sigma_1,\sigma_2}_{\mathcal{G},\rho}(v, b) := \mathbb{E}cbr^{\sigma_1,\sigma_2}_{\mathcal{G},\rho,\mathbf{1}}(v, b)$.

The second class encompasses *cost-bounded reachability probabilities*, i. e. questions like "What is the maximal probability to reach a set $V_{\mathrm{goal}} \subseteq V$ of states with cost $\leq b$?". We first make the states in V_{goal} absorbing and add a Markovian self-loop $\mathrm{tr}_v = (v, \lambda, \{v \mapsto 1\})$ with arbitrary finite rate $0 < \lambda < \infty$ to each state $v \in V_{\mathrm{goal}}$ and define the final reward by $\rho_{\mathrm{f}}(v) = 1$ if $v \in V_{\mathrm{goal}}$, and $\rho_{\mathrm{f}}(v) = 0$ otherwise. The transient and instantaneous rewards are constantly 0. Then the expected reward until cost b is reached corresponds to the probability of reaching V_{goal} with costs $\leq b$.

Algorithms to compute optimal expected time-bounded rewards are available both for Markov automata [GTH+14] and stochastic games [BFH+15]. To the best of our knowledge, up to now there are no algorithms available to compute the optimal expected cost-bounded rewards for MA and SG.

Instantaneous Costs Similar to the definition of a reward structure, we could also define instantaneous costs which occur when taking a transition. We do not consider instantaneous costs in this paper for two reasons:

First, they would render the transformation in Sect. 3 impossible, since there is no instantaneous time. In principle, adapting the analysis algorithm for time-bounded rewards [GTH+14, BFH+15] to cost bounds should be possible. That algorithm is based on discretising the time interval, yielding a discrete-time probabilistic game. However, analysing cost-bounded properties for discrete-time models is expensive, even more so as we would have to support non-integer costs [AHK03].

The second reason is that instantaneous costs increase the complexity of the problem. Without instantaneous costs, computing cost-bounded expected rewards up to a user-defined precision $\varepsilon > 0$ can be done in polynomial time. We extend for the moment the cost structure given in Definition 2.4 to encompass instantaneous as well as transient costs. That is to say, each transition imposes a non-negative instantaneous cost via function $c_{\mathrm{i}} : T \to \mathbb{R}_{\geq 0}$. Accordingly, the cost of histories and paths and the cost-bounded reward, computed by cost and $cbr^{\mathcal{G}}_{\rho,c}$, resp., are extended in a straightforward way. We now study the problem of the optimal expected cost-bounded rewards under the new cost structure. The following theorem shows the complexity of the problem.

Theorem 2.1 (Complexity of instantaneous costs) *Computation of the optimal cost-bounded reward of an SG under presence of instantaneous cost is NP-hard.*

Proof We provide a reduction from the knapsack problem. The goal is to select a subset from n items, each with value x_i and weight w_i ($i = 1, \ldots, n$), such that the weight of the items in the subset is at most equal to a given bound W and their value is maximal. To solve the problem, we define a stochastic game $\mathcal{G} = (V, (V_1, \emptyset), v_1, T)$ (which is actually an MA) such that $V = V_1 = \{v_1, \ldots, v_{n+1}\}$ and $T = \{\mathrm{tr}^{(0)}_i, \mathrm{tr}^{(1)}_i \mid i = 1, \ldots, n\} \cup \{\mathrm{tr}_{n+1}\}$. States $v_1, \ldots, v_n$ correspond to the respective items, whereas v_{n+1} is just a terminal state. It is equipped with a Markovian self-loop $\mathrm{tr}_{n+1} = (v_{n+1}, 1, \Delta_{v_{n+1}})$ in order to ensure deadlock freedom and to avoid zenoness. Recall that $\Delta_{v_{n+1}}$ denotes the Dirac distribution to v_{n+1}.

Executing transitions $\mathrm{tr}^{(0)}_i$ and $\mathrm{tr}^{(1)}_i$ is equivalent to ignoring and picking item i, respectively. For $i = 1, \ldots, n$, each v_i has a pair of those transitions, i. e. $\mathrm{tr}^{(0)}_i = \mathrm{tr}^{(1)}_i = (v_i, \infty, \Delta_{v_{i+1}})$. The transitions however have different cost and reward, namely $c_{\mathrm{i}}(\mathrm{tr}^{(1)}_i) = w_i$ and $\rho_{\mathrm{i}}(\mathrm{tr}^{(1)}_i) = x_i$. All other rewards and costs are zero. It is not hard to see that knapsack problem can be solved via computing $\mathbb{E}cbr^{\sup,\sup}_{\mathcal{G},\rho,c}(v_1, W)$. □

Theorem 2.1 shows that instantaneous cost consumption substantially increases the hardness of the problem. It adds one level of combinatorial complexity (from polynomial to NP-hard) into the problem. This is the main reason why we do not consider it in this paper. From now on we assume as before that the cost consumption is transient.

3. Transformation of stochastic games

In this section, we first give a fixed point characterisation of expected cost-bounded rewards for stochastic games and prove its correctness in the Appendix. Similar to time-bounded properties [BFH+15], this fixed point characterisation is not amenable to an efficient solution. Therefore we transform the stochastic game so that the optimal expected cost-bounded reward coincides with the optimal expected time-bounded reward in the transformed game. This allows us to apply arbitrary algorithms like [GTH+14, BFH+15] for expected time-bounded rewards to compute optimal expected cost-bounded rewards.

Theorem 3.1 (Fixed point characterisation) *Let $\mathcal{G}$ be a stochastic game with cost function c and reward structure $\rho = (\rho_{\mathrm{t}}, \rho_{\mathrm{i}}, \rho_{\mathrm{f}})$. Let $b \in \mathbb{R}_{\geq 0}$ be a cost bound,* $\mathrm{opt}_1, \mathrm{opt}_2 \in \{\inf, \sup\}$, *and* $\mathrm{opt}_{[v]} = \mathrm{opt}_i$ *if* $v \in V_i$. *Then,* $\mathbb{E}cbr^{\mathrm{opt}_1,\mathrm{opt}_2}_{\mathcal{G},\rho,c}(v, b)$ *is the least fixed point of the higher-order operator* $\Omega^{\mathrm{opt}_1}_{\mathrm{opt}_2} : (V \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}) \to (V \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0})$, *such that*

$$\Omega^{\mathrm{opt}_1}_{\mathrm{opt}_2}(F)(v, b) = \underset{\mathrm{tr} \in T(v)}{\mathrm{opt}_{[v]}} \begin{cases} \displaystyle\int_0^{b/c(\mathrm{tr})} \lambda_{\mathrm{tr}} \cdot \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \cdot \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot F\big(v', b - c(\mathrm{tr}) \cdot t\big)\,\mathrm{d}t & \\ \quad + \left(\frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr})\right) \cdot \left(1 - \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}} \cdot b}{c(\mathrm{tr})}}\right) + \rho_{\mathrm{f}}(v) \cdot \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}} \cdot b}{c(\mathrm{tr})}}, & \\ & \textit{if } \mathrm{tr} \in T_{\mathrm{M}}(v) \wedge c(\mathrm{tr}) > 0, \\ \frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr}) + \displaystyle\sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot F(v', b), & \textit{if } \mathrm{tr} \in T_{\mathrm{M}}(v) \wedge c(\mathrm{tr}) = 0, \\ \rho_{\mathrm{i}}(\mathrm{tr}) + \displaystyle\sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot F(v', b), & \textit{if } \mathrm{tr} \in T_{\mathrm{P}}(v). \end{cases}$$

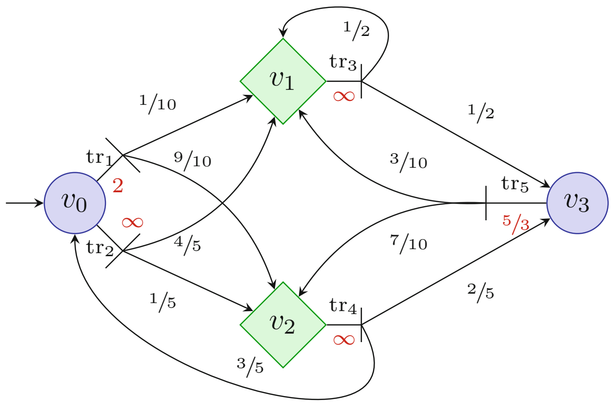
The existence of the least fixed point is guaranteed through *Tarski's fixed point theorem* [Tar55]. It states that for a complete lattice $\langle A, \leq \rangle$, a monotone function $f : A \to A$, and the set $F := \{x \in A \mid f(x) = x\}$ of all fixed points of f in A, it holds that $F \neq \emptyset$ and that $\langle F, \leq \rangle$ is a complete lattice as well. In our case, we have that $A := (V \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0})$ and function $f := \Omega^{\mathrm{opt}_1}_{\mathrm{opt}_2}$. The partial order "$\leq$" is given as follows: For all $f, g : (V \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}) \to (V \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0})$ it holds that $f \leq g \Leftrightarrow \forall\, v \in V, b \in \mathbb{R}_{\geq 0} : f(v, b) \leq g(v, b)$.

The remaining proof of Theorem 3.1 can be found in Appendix B.

The fixed point characterisation of expected cost-bounded rewards yields a system of integral equations, which are typically hard to solve. Instead, the following transformation turns cost-bounded rewards into time-bounded rewards. For the latter, not only a fixed point characterisation is available [BFH+15], but also a more efficient algorithm, based on discretisation [GTH+14, BFH+15].

Definition 3.1 (*Cost-to-time transformation*) Let $\mathcal{G} = (V, (V_1, V_2), v_{\mathrm{init}}, T)$ be a stochastic game with cost function $c : T \to \mathbb{R}_{\geq 0}$ and reward structure $\rho = (\rho_{\mathrm{t}}, \rho_{\mathrm{i}}, \rho_{\mathrm{f}})$. We define the *cost-transformed game* $\mathcal{G}^c = \big(V, (V_1, V_2), v_{\mathrm{init}}, T^c\big)$ with

$$\begin{aligned} T^c = &\big\{\mathrm{tr} \in T \;\big|\; \lambda_{\mathrm{tr}} = \infty\big\} \\ &\cup \big\{(v, \infty, \mu) \;\big|\; \exists\, \lambda \in \mathbb{R}_{\geq 0} : \mathrm{tr} = (v, \lambda, \mu) \in T \wedge c(\mathrm{tr}) = 0\big\} \\ &\cup \big\{(v, \lambda/c(\mathrm{tr}), \mu) \;\big|\; \mathrm{tr} = (v, \lambda, \mu) \in T \wedge c(\mathrm{tr}) \neq 0\big\} \end{aligned}$$



(a)

	tr_1	tr_2	tr_3	tr_4	tr_5	v_0	v_1	v_2	v_3
ρ_{t}^c	$1/5$	0	0	0	$1/3$				
ρ_{i}^c	4	4	1	5	2				
ρ_{f}^c						1	0	2	3

(b)

Fig. 2. Figure 1 after transformation

and reward structure $\rho^c = (\rho_{\mathrm{t}}^c, \rho_{\mathrm{i}}^c, \rho_{\mathrm{f}}^c)$ such that $\rho_{\mathrm{f}}^c = \rho_{\mathrm{f}}$,

$$\rho_{\mathrm{t}}^c(\mathrm{tr}) = \begin{cases} \rho_{\mathrm{t}}(\mathrm{tr})/c(\mathrm{tr}), & \text{if } c(\mathrm{tr}) \neq 0, \\ 0, & \text{if } c(\mathrm{tr}) = 0, \text{ and} \end{cases}$$

$$\rho_{\mathrm{i}}^c(\mathrm{tr}) = \begin{cases} \rho_{\mathrm{i}}(\mathrm{tr}) + \rho_{\mathrm{t}}(\mathrm{tr})/\lambda_{\mathrm{tr}}, & \text{if } c(\mathrm{tr}) = 0 \wedge \lambda_{\mathrm{tr}} < \infty, \\ \rho_{\mathrm{i}}(\mathrm{tr}), & \text{otherwise.} \end{cases}$$

The motivation behind this transformation is as follows: Since we want to transform the cost bound b into a time bound we have to divide b through the cost gained per time unit. This is done by dividing the rate λ of a Markovian transition $\mathrm{tr} \in T_{\mathrm{M}}$ through its cost $c(\mathrm{tr})$. The same has to be done with the transient reward $\rho_{\mathrm{t}}(\mathrm{tr})$. If tr has no cost, i. e. $c(\mathrm{tr}) = 0$, the transition is transformed into a probabilistic transition. The expected transient reward $\rho_{\mathrm{t}}(\mathrm{tr})/\lambda_{\mathrm{tr}}$ has to be added to the instantaneous reward of the transition in this case.

The transformation does not change the structure or size of the SG, and the transformed system is an SG as well. Additionally, Markov automata are closed under this transformation, i. e. if the original SG is actually an MA, so is the transformed system. To see this, we only have to consider Markovian transitions, since probabilistic transitions are not affected by the transformation. We have to distinguish between two scenarios: (1) Markovian transitions with positive costs are transformed into Markovian transitions with new rates and new transient rewards. (2) Markovian transitions without costs are transformed into probabilistic transitions with infinite rate and new instantaneous reward. It is easy to see that the transformed transitions still adhere to Definition 2.1.

Example 3.1 Consider again the stochastic game in Fig. 1a with the costs and rewards in Fig. 1b. We assume a cost bound of $b = 20$. Then the rewards of the five transitions after transformation are shown in Fig. 2b. Transitions tr_3 and tr_4 remain unchanged as they are probabilistic. The Markovian transitions tr_1, tr_2, and tr_5 are modified as follows. The expected residence time before taking tr_1 is scaled such that it matches the expected cost in the original game, i. e. the new exit rate becomes $\lambda_{\mathrm{tr}_1}/c(\mathrm{tr}_1) = 10/5 = 2$. The transient reward rate is adjusted accordingly and becomes $\rho_{\mathrm{t}}(\mathrm{tr}_1)/c(\mathrm{tr}_1) = 1/5$. The instantaneous reward does not change. The transition tr_5 is modified in the same way. As the cost of tr_2 is zero, tr_2 becomes probabilistic and the expected reward $\rho_{\mathrm{t}}(\mathrm{tr}_2)/\lambda_{\mathrm{tr}_2}$ earned in v_1 until tr_2 being taken is added to the instantaneous reward of tr_2. The stochastic game after the transformation is shown in Fig. 2a.

Theorem 3.2 (Measure preservation) *Let $\mathcal{G}$ be a stochastic game with reward structure ρ, cost function c, cost bound $b \in \mathbb{R}_{\geq 0}$, $v \in V$, and* $\mathrm{opt}_1, \mathrm{opt}_2 \in \{\inf, \sup\}$. *Then we have*

$$\mathbb{E}cbr_{\mathcal{G},\rho,c}^{\mathrm{opt}_1,\mathrm{opt}_2}(v, b) = \mathbb{E}\mathrm{tbr}_{\mathcal{G}^c,\rho^c}^{\mathrm{opt}_1,\mathrm{opt}_2}(v, b).$$

Proof Here we sketch the proof of the theorem. It is done by showing that the original and the transformed games have indeed the same fixed point characterisation for the respective objectives. For this, on the one hand, we construct the fixed point characterisation of the transformed game using Theorem 3.1 by assigning the constant cost of one to all Markovian transitions. On the other hand, we reinterpret the representation of the fixed point characterisation of the original model by a series of sound variable substitutions, partly inspired by the transformation. At the end we conclude that both of the fixed point characterisations are the same, and thereby their least fixed points are exactly equal. For more details, see the complete proof in Appendix C. □

We strongly believe that the result of Theorem 3.2 is valid under the class of *late* strategies. The claim is posed in the following conjecture.

Conjecture 3.1 Theorem 3.2 holds also when the optimal measures are computed over the class of *late* (instead of early) strategies.

Here we provide the intuition behind the conjecture. The challenge exists in the transformation of Markovian transitions with zero cost consumption into probabilistic transitions. It must be the case that any strategy in the original game that chooses those transitions for finite amount of time is sub-optimal. Otherwise the strategy cannot be simulated in the transformed game, since their corresponding transitions are probabilistic, and thus the measure preservation fails. However, the sub-optimality of such a strategy can be justified by the fact that those transitions are cost preserving. Therefore, holding them for finite amount of time cannot be optimal as they can earn reward for free (with no cost consumption). The optimal case might then happen by either ignoring those transitions or holding them until firing, both can be imitated by the corresponding probabilistic transitions.

Zero-cost transitions[1] in the original game can introduce Zenoness in the transformed game. That happens if a set of such transitions constitutes an end component in the transformed game. This will be problematic for the analysis, in particular if the end component contains positive rewards. Therefore the strategy that keeps the control of the game inside the end component delivers infinite expected rewards, since staying there gains reward without any cost. Nevertheless the analysis may ignore such a strategy in some cases, for instance in the analysis of MA against minimal ECR. By any means and for simplicity we exclude such models from our analysis technique.

4. Case studies and experimental results

In this section we report on experimental results of computing the optimal ECR on several case studies. The case studies are all modelled as MA. However, we explain how to give a game semantics to one of them, namely the dynamic power management system. The game interpretation is not specific to this model and can in principle be applied to other systems.

(1) The *Dynamic Power Management System* (DPMS) [QQP01] describes the following scenario: A service requester generates tasks which are stored within a queue until they are handled by a processor. This processor (P) can either be "busy" with processing a job, "idle" while the queue is empty, in a "standby" mode, or in a "sleep" mode. In the latter two modes P is inactive and cannot handle tasks. The change between "busy" and "idle" occurs automatically, depending on whether there are tasks in the queue or not. If P has been "idle" for some time, it is switched into "standby" or "sleep" by a power manager. The power manager is also responsible for switching from these two modes back to "idle". P consumes the least power in "standby" and "sleep" (0.35 W and 0.13 W, respectively), whereas it consumes more power while "idle" (0.95 W) and the most if it is "busy" (2.15 W) [QQP01, SBGM00]. We model the DPMS as an MA with the costs representing the power consumption of P. The reward corresponds to the number of served tasks. For our experiments we varied the number of different task types (T) and the size of the queue (Q). We explore the expected cost-bounded reward. The model instances are denoted as "DPMS-T-Q".

It is possible to give a game interpretation to the DPMS case. For this, we can distinguish two kinds of nondeterminism in this model. The nondeterminism that is resolved by the power manager is assumed to be *controllable*, meaning that the system itself can control it, e. g. the power manager can switch the system into "idle" or "standby". The service requester however exhibits a different kind of nondeterminism, by generating tasks of different types. This nondeterminism is considered *non-controllable* since the system has no control on the type of incoming tasks.

[1] Note that the cost of probabilistic transitions is implicitly zero as the delay until taking such transitions is zero.

Table 1. Expected reward in the dynamic power management system

Name	#States	Budget = 10		Budget = 20		Budget = 50	
		Min	Max	Min	Max	Min	Max
DPMS-2-5	508	0.759	0.859	1.557	1.924	3.910	5.150
DPMS-2-10	1588	0.759	0.859	1.557	1.924	3.910	5.150
DPMS-2-20	5548	0.759	0.859	1.557	1.924	3.910	5.150
DPMS-3-5	5190	0.785	0.883	1.617	1.930	4.129	5.088
DPMS-3-10	29,530	0.785	0.883	1.617	1.930	4.129	5.088
DPMS-3-20	195,810	0.785	0.883	1.617	1.930	4.129	5.088
DPMS-4-5	47,528	0.784	0.877	1.617	1.889	4.143	4.936
DPMS-4-10	492,478	0.784	0.877	1.617	1.889	4.143	4.936

Table 2. Expected reward of the queueing system

Name	#States	Budget = 1		Budget = 5		Budget = 10	
		Min	Max	Min	Max	Min	Max
QS-2-2	2314	0.249	0.857	1.294	4.078	2.634	7.975
QS-2-3	10,778	0.249	0.857	1.294	4.078	2.634	7.975
QS-2-4	46,234	0.249	0.857	1.294	4.078	2.634	7.975
QS-2-5	191,258	0.249	0.857	1.294	4.078	2.634	7.975
QS-2-6	777,754	0.249	0.857	1.294	4.078	2.634	7.975
QS-3-2	12,205	0.125	0.857	0.649	4.078	1.332	7.972
QS-3-3	117,532	0.125	0.857	0.649	4.078	1.332	7.972
QS-3-4	1,080,865	0.125	0.857	0.649	4.078	1.332	7.972
QS-4-2	42,616	0.125	1.287	0.649	6.127	1.333	12.075
QS-4-3	708,088	0.125	1.287	0.649	6.127	1.333	12.075
QS-6-2	266,974	0.084	1.713	0.433	8.187	0.892	16.201

Hence a more refined model for this system is a stochastic game. The two kinds of nondeterminism can be considered as either co-operating with or competing against each other. In the competitive semantics one can investigate, e. g., situations where the power manager tries to increase the number of processed task within a certain energy budget while the service requester aims to decrease it. The MA model considered here implicitely uses a co-operative semantics.

(2) The *Queueing System* (QS) [HH12] stores requests of T different types into two queues of size Q each. A server is attached to each queue, which fetches requests from its corresponding queue, and then processes them. One of the servers might insert, with probability 0.1, the already served request into the other queue to be reprocessed by the other server. Power is consumed by both servers when they are processing. We compute the minimum and the maximum number of processed requests under different energy budgets. The model instances are denoted as "QS-T-Q".

(3) The *Polling System* (PS) [GHH+13, TvdPS13] consists of S stations and one server. Each station comes with a queue of size Q, and buffers incoming jobs of T different types. The jobs are then polled and processed by the server. There is a probability of 0.1 for a job to be processed while erroneously remaining in the queue. Each job brings an instantaneous reward when it is completely processed by the server. Whenever processing, the server consumes energy. The model is subject to two kinds of analysis: First we compute the minimum and the maximum probability of encountering the error under some energy budget. The second analysis is on the computation of the minimum and the maximum expected energy-bounded reward of the model. The instances of the polling system are denoted as "PS-S-T-Q".

(4) The *Stochastic Job Scheduling* benchmark (SJS) [BDF81] originally stems from economy. In this setting, a number of jobs with different service rates are distributed between processors. Each processor consumes resources, e. g. energy which has to be paid for. The costs in our model represent these expenses. The goal is to have all jobs processed within a certain cost budget. In our experiments we explore the reachability of this goal with homogeneous costs ("all processors have the same costs") and heterogeneous costs ("all processors have different costs"), while varying the number of jobs (M) and the number of processors (N). Since the system degenerates to a CTMC if the service rates are homogeneous, we do not consider this case. The model instances are denoted as "SJS-N-M".

Table 3. Results for the polling system

		Reachability		Reward	
Name	#States	Min	Max	Min	Max
PS-2-2-2	455	0.743	0.773	3.128	3.219
PS-2-2-3	2055	0.483	0.551	3.980	4.117
PS-2-2-4	8421	0.998	0.999	1.045	1.080
PS-2-3-2	2392	0.995	0.996	1.209	1.253
PS-2-3-3	22,480	0.973	0.983	1.730	1.848
PS-2-3-4	137,445	0.990	0.994	1.489	1.583
PS-3-2-2	3577	0.888	0.917	2.549	2.685
PS-3-2-3	34,425	0.665	0.760	3.493	3.732
PS-3-3-2	35,659	1.000	1.000	0.918	0.965
PS-3-4-2	300,793	0.402	0.543	4.180	4.412
PS-4-2-2	27,783	0.955	0.973	2.166	2.307
PS-4-3-2	570,375	0.793	0.879	3.116	3.403
PS-5-2-2	213,689	0.983	0.992	1.908	2.039

Table 4. Reachability in the stochastic job scheduling benchmark

		Homogeneous costs		Heterogeneous costs	
Name	#States	Min	Max	Min	Max
SJS-2-2	34	0.713	0.713	0.699	0.799
SJS-2-4	464	0.241	0.241	0.186	0.243
SJS-2-6	4144	0.041	0.041	0.021	0.029
SJS-2-8	29,344	0.004	0.004	0.001	0.002
SJS-4-2	104	0.713	0.713	0.542	0.995
SJS-4-4	3168	0.241	0.241	0.120	0.610
SJS-4-6	71,644	0.041	0.041	0.013	0.130
SJS-4-8	1,032,272	0.004	0.004	0.001	0.012
SJS-6-2	214	0.713	0.713	0.424	1.000
SJS-6-4	13,924	0.241	0.241	0.059	0.945
SJS-6-6	685,774	0.041	0.041	0.005	0.374
SJS-8-2	364	0.713	0.713	0.337	1.000
SJS-8-4	41,552	0.241	0.241	0.033	0.999
SJS-10-2	554	0.713	0.713	0.274	1.000
SJS-10-4	98,436	0.241	0.241	0.019	1.000

We used SCOOP [TKvdPS12] to create the model files. The transformation from cost to time was done with a python script; the computation time for this was negligible. We then employed the tool IMCA [GHKN12, GHH+13, GTH+14] to determine the minimum and maximum expected cost-bounded reward or the minimum and maximum cost-bounded reachability probabilities of the models. It would be possible to use any other analyser for MA, e. g. MeGARA, the prototype from [BFH+15].

All experiments were run on an Intel Xeon quad-core processor with 3.3 GHz per core and 64 GB of memory. We set a time limit of 12 h. The memory consumption was negligible; all experiments needed less than 300 MB.

We will not give detailed time measurements due to space restrictions, nevertheless we want to briefly discuss the computation times. The shortest computations took only fractions of a second, e. g. the computation of the minimum reachability for SJS-2-4 with cost budget 5 took 0.06 s, whereas the longer computations needed several hours, e. g. for DPMS-4-10 the computation of the minimum reachability with cost budget 50 took almost 11 h, which was the longest computation time of all our experiments. The computation time is also influenced by the size of the cost budget. For example, for cost budget 10 the computation of the minimum reachability for DPMS-4-10 took less than 6 min. This is due to the fact that IMCA uses discretisation [GHH+13, GHH+14, GTH+14] to compute the values; for a larger bound more discretisation steps are needed. There is also an interesting connection between the costs within the system, its maximum rate, and the computation time: The size of a discretisation step depends on the maximum rate of the transformed system. The higher the maximum rate is, the smaller the discretisation step must be chosen in order to satisfy the given accuracy level. For the computation of cost-bounded rewards, this means that the computation time is strongly influenced by the value of $\max\left\{ {}^{\lambda_{\mathrm{tr}}}/_{c(\mathrm{tr})} \,\middle|\, \mathrm{tr} \in T_{\mathrm{M}} : c(\mathrm{tr}) > 0 \right\}$. For details on the discretisation, see [GTH+14, BFH+15].

Tables 1, 2, 3 and 4 show the results of our experiments. The first two columns of each table contain the name of the respective model instance and its number of states.

In case of DPMS (Table 1) and QS (Table 2) we explore the minimum and maximum expected reward under different cost budgets. For DPMS we used cost budgets of 10, 20, and 50, whereas for QS we used cost budgets of 1, 5, and 10 (see the respective blocks in Tables 1, 2). It holds for both DPMS and QS that the expected reward grows with the budget, as does the difference between minimum and maximum reward, as to be expected. Another interesting fact is that the size of the queues in the models – while having a big influence on the size of the system – has practically no impact on the expected reward. It is completely determined by the number of different task types. This observation can be explained as follows: For the processing unit of DPMS (or of QS) it is not important how many jobs exactly can be stored in the queue(s), as long as there *are* jobs in the queue(s).

For PS (Table 3) we studied both minimum and maximum reachability and minimum and maximum expected reward (see the respective blocks in the table) under a cost budget of 5. If we increase the queue size, the minimum and maximum probability for encountering the error decreases, while the expected minimum and maximum reward increases. At the same time we can observe that the reachability increases with the number of stations, e. g. for PS-2-2-2, containing two stations, the maximum probability is 0.773, whereas for PS-5-2-2, containing five stations, it is 0.992. This makes sense, since the error is caused by the stations and the probability to encounter the error therefore increases with having more stations.

For SJS (Table 4) we also used a cost budget of 5. Here we studied the minimum and maximum reachability while assuming homogeneous or heterogeneous costs for the different processors of the system (see the respective blocks in Table 4). For homogeneous costs we can observe a similar effect as for DPMS and PS: The number of processors influences the number of states in the system, but has a negligible impact on the reachability. The latter is completely determined by the number of jobs. What's more, the minimum and the maximum reachability are the same in this case. These effects vanish if we assume heterogeneous costs. In this case, the distance between minimum and maximum reachability increases, especially the maximum reachability becomes higher. These observations make sense: in case of a homogeneous system it does not matter, which processor handles which job. However, in a heterogeneous system there is a choice between more and less expensive processors which can handle the jobs, which in turn leads to a higher (lower) maximum (minimum) reachability.

5. Conclusion

We studied the computation of Markov automata and stochastic games against cost-bounded reward objectives. In this regard, we provided a fixed point characterisation for the optimal expected cost-bounded reward. Moreover, we proposed an efficient measure-preserving transformation from cost-bounded to time-bounded objectives. For the latter, an analysis technique based on discretisation with strict error bound exists. Our experiments demonstrate the effectiveness of the approach.

In the future, we want to prove Theorem 3.2 for the larger class of late schedulers. We plan to improve the efficiency of the analysis, e. g. via uniformization-based techniques and abstraction refinement on very large games and automata.

A. Fixed point characterisation for expected time-bounded reward

We recall that the optimal expected time-bounded reward (ETR) is the special case of the optimal expected cost-bounded reward (ECR) when the costs of all transitions are one. We recap the fixed point characterisation of the optimal ETR for stochastic games as described in [BFH+15, Lemma 1]. The characterisation is slightly extended for reward structures with final rewards and adapted to our notations.

Theorem A.1 (Fixed point characterisation for the optimal ETR) *Let $\mathcal{G}$ be a stochastic game with reward structure $\rho = (\rho_t, \rho_i, \rho_f)$. Let $b \in \mathbb{R}_{\geq 0}$ be a time bound and* $\mathrm{opt}_i \in \{\inf, \sup\}$*, and* $\mathrm{opt}_{[v]} = \mathrm{opt}_i$ *if $v \in V_i$ for $i \in \{1, 2\}$. Then,* $\mathbb{E}\mathrm{tbr}_{\mathcal{G},\rho}^{\mathrm{opt}_1,\mathrm{opt}_2}(v, b)$ *is the least fixed point of the higher-order operator* $\Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1} : (V \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}) \to (V \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0})$*,* *such that*

$$\Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1}(F)(v,b)=\underset{\mathrm{tr}\in T(v)}{\mathrm{opt}_{[v]}}\begin{cases}\int_0^b \lambda_{\mathrm{tr}}\cdot \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t}\cdot \sum_{v'\in V}\mu_{\mathrm{tr}}(v')\cdot F(v',b-t)\,\mathrm{d}t & \\ \quad +\left(\frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}}+\rho_{\mathrm{i}}(\mathrm{tr})\right)\cdot\left(1-\mathrm{e}^{-\lambda_{\mathrm{tr}}\,b}\right)+\rho_{\mathrm{f}}(v)\mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot b}, & \textit{if } \mathrm{tr}\in T_{\mathrm{M}}(v),\\ \rho_{\mathrm{i}}(\mathrm{tr})+\sum_{v'\in V}\mu_{\mathrm{tr}}(v')\cdot F(v',b), & \textit{if } \mathrm{tr}\in T_{\mathrm{P}}(v).\end{cases}$$

Proof The proof is the direct result of Theorem 3.1 by choosing a constant cost of one for each transition. □

B. Proof of Theorem 3.1

We recall Theorem 3.1:

Theorem 3.1 (Fixed point characterisation) *Let $\mathcal{G}$ be a stochastic game with cost function c and reward structure $\rho=(\rho_{\mathrm{t}},\rho_{\mathrm{i}},\rho_{\mathrm{f}})$. Let $b\in\mathbb{R}_{\geq 0}$ be a cost bound,* $\mathrm{opt}_1,\mathrm{opt}_2\in\{\inf,\sup\}$, *and* $\mathrm{opt}_{[v]}=\mathrm{opt}_i$ *if $v\in V_i$. Then, $\mathbb{E}cbr_{\mathcal{G},\rho,c}^{\mathrm{opt}_1,\mathrm{opt}_2}(v,b)$ is the least fixed point of the higher-order operator $\Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1}:(V\times\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0})\to(V\times\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0})$, such that*

$$\Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1}(F)(v,b)=\underset{\mathrm{tr}\in T(v)}{\mathrm{opt}_{[v]}}\begin{cases}\int_0^{b/c(\mathrm{tr})} \lambda_{\mathrm{tr}}\cdot \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t}\cdot \sum_{v'\in V}\mu_{\mathrm{tr}}(v')\cdot F\left(v',b-c(\mathrm{tr})\cdot t\right)\mathrm{d}t & \\ \quad +\left(\frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}}+\rho_{\mathrm{i}}(\mathrm{tr})\right)\cdot\left(1-\mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}}\right)+\rho_{\mathrm{f}}(v)\cdot\mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}}, & \\ & \text{if } \mathrm{tr}\in T_{\mathrm{M}}(v)\wedge c(\mathrm{tr})>0,\\ \frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}}+\rho_{\mathrm{i}}(\mathrm{tr})+\sum_{v'\in V}\mu_{\mathrm{tr}}(v')\cdot F(v',b), & \text{if } \mathrm{tr}\in T_{\mathrm{M}}(v)\wedge c(\mathrm{tr})=0,\\ \rho_{\mathrm{i}}(\mathrm{tr})+\sum_{v'\in V}\mu_{\mathrm{tr}}(v')\cdot F(v',b), & \text{if } \mathrm{tr}\in \mathrm{T_P(v)}.\end{cases}$$

Proof The proof is done in two steps. First we show that $\mathbb{E}cbr_{\mathcal{G},\rho,c}^{\mathrm{opt}_1,\mathrm{opt}_2}$ is a fixed point of the operator Ω described in the theorem. Then we show that it is indeed the least fixed point. Since $\mathcal{G}$, ρ and c are clear from the context, we will drop the respective subscripts from now on and write $\mathbb{E}cbr^{\mathrm{opt}_1,\mathrm{opt}_2}$ only.

We recall the definition of the optimal expected cost-bounded reward. Given a stochastic game $\mathcal{G}$, a cost bound $b\geq 0$, and $v\in V$, the maximum time-bounded expected reward is defined as:

$$\mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}(v,b):=\underset{\sigma_1\in\Sigma_1}{\mathrm{opt}_1}\underset{\sigma_2\in\Sigma_2}{\mathrm{opt}_2}\int_{\pi\in\Pi(v)} cbr(\pi,b)\,\mathrm{d}\mathrm{Pr}_{v,\sigma_1,\sigma_2}(\pi). \tag{1}$$

Let v be a state and further assume w. l. o. g. that $v\in V_1$. Since the initial state is v, all the paths starting not from v have measure zero and are out of the computation. Therefore the optimal early strategy at state v tries to pick the best combination of transitions available at v. Obviously it selects the transition that optimises the objective rather than any convex combination of available transitions since the latter gives anyway an inferior value than or at most the same value as the optimal one. Consequently we can only optimise over the deterministic strategies, that select one of the available transitions of v with probability one.

Therefore we have

$$\mathbb{E}cbr^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v, b) = \underset{\mathrm{tr}\in T(v)}{\mathrm{opt}_1} \underbrace{\underset{\sigma_1\in\Sigma_1^{\mathrm{tr}}}{\mathrm{opt}_1} \underset{\sigma_2\in\Sigma_2}{\mathrm{opt}_2} \overbrace{\int_\pi cbr(\pi, b)\,\mathrm{dPr}_{v,\sigma_1,\sigma_2}(\pi)}^{\mathcal{R}_{\sigma_1,\sigma_2}(v,\mathrm{tr},b)}}_{\overset{\star}{\mathcal{R}}{}^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v,\mathrm{tr},b)}, \tag{2}$$

where Σ_1^{tr} is the set of strategies that select transition $\mathrm{tr} \in T(v)$ for path v, i. e. a path containing only state v. Now we consider different cases of transition tr:

(a) $\mathrm{tr} \in T_{\mathrm{M}}(v)$, $c(\mathrm{tr}) > 0$ and $b = 0$: For this case the final reward is applied, then

$$\overset{\star}{\mathcal{R}}{}^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v, \mathrm{tr}, b) = \rho_{\mathrm{f}}(v). \tag{3}$$

(b) $\mathrm{tr} \in T_{\mathrm{M}}(v)$, $c(\mathrm{tr}) > 0$ and $b > 0$: We split each path starting from v at the point it takes transition tr, and write it as $\pi = v \xrightarrow{t,\mathrm{tr}} \pi'$. We can therefore split the reward of such paths accordingly:

$$cbr(\pi, b) = \begin{cases} \rho_{\mathrm{t}}(\mathrm{tr}) \cdot t + \rho_{\mathrm{i}}(\mathrm{tr}) + cbr(\pi', b - c(\mathrm{tr}) \cdot t), & t \le \frac{b}{c(\mathrm{tr})}, \\ \rho_{\mathrm{t}}(\mathrm{tr}) \cdot \frac{b}{c(\mathrm{tr})} + \rho_{\mathrm{f}}(v), & t > \frac{b}{c(\mathrm{tr})}. \end{cases} \tag{4}$$

The probability measure of such paths can be split in a similar way. To do that we first need to construct a new strategy from a given arbitrary strategy σ_i, $i = \{1, 2\}$. Its aim is to mimic the decision of σ_i on path π when it takes the suffix of π after the splitting point t. It is required to guarantee that the splitting of the probability measure is sound. Formally strategy $\sigma_i^{t,\mathrm{tr}}$ resolves nondeterminism for the suffix π' of path π after splitting point t as σ_i does it for π, i. e. for $\pi = v \xrightarrow{t,\mathrm{tr}} \pi'$, $\sigma_i^{t,\mathrm{tr}}(\pi') = \sigma_i(\pi)$. Whenever clear from the context, we drop tr from $\sigma_i^{t,\mathrm{tr}}$ and write $\sigma_i^t(\pi')$. We can subsequently split the probability measure of path π:

$$\mathrm{dPr}_{v,\sigma_1,\sigma_2}(\pi) = \lambda_{\mathrm{tr}} \cdot \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t}\,\mathrm{d}t \cdot \sum_{v'\in V} \mu_{\mathrm{tr}}(v')\,\mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi'). \tag{5}$$

We proceed with the simplification of $\mathcal{R}_{\sigma_1,\sigma_2}(v, \mathrm{tr}, b)$ according to Eqs. (4) and (5):

$$\begin{aligned}
\mathcal{R}_{\sigma_1,\sigma_2}&(v, \mathrm{tr}, b) \\
=&\int_\pi cbr(\pi, b)\,\mathrm{dPr}_{v,\sigma_1,\sigma_2}(\pi) \\
=&\int_0^{b/c(\mathrm{tr})}\int_{\pi'} \Big(\big(\rho_{\mathrm{t}}(\mathrm{tr})\cdot t + \rho_{\mathrm{i}}(\mathrm{tr}) + cbr(\pi', b - c(\mathrm{tr})\cdot t)\big)\,\lambda_{\mathrm{tr}}\,\mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t}\cdot \sum_{v'\in V}\mu_{\mathrm{tr}}(v')\,\mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\Big)\,\mathrm{d}t \\
&+ \int_{b/c(\mathrm{tr})}^{\infty}\int_{\pi'} \Big(\big(\rho_{\mathrm{t}}(\mathrm{tr})\cdot \tfrac{b}{c(\mathrm{tr})} + \rho_{\mathrm{f}}(v)\big)\,\lambda_{\mathrm{tr}}\,\mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t}\cdot \sum_{v'\in V}\mu_{\mathrm{tr}}(v')\,\mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\Big)\,\mathrm{d}t \\
=&\int_0^{b/c(\mathrm{tr})}\int_{\pi'} \Big(\rho_{\mathrm{t}}(\mathrm{tr})\cdot t + \rho_{\mathrm{i}}(\mathrm{tr})\Big)\,\lambda_{\mathrm{tr}}\,\mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t}\sum_{v'\in V}\mu_{\mathrm{tr}}(v')\,\mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\,\mathrm{d}t \\
&+ \int_0^{b/c(\mathrm{tr})}\int_{\pi'} cbr(\pi', b - c(\mathrm{tr})\cdot t)\,\lambda_{\mathrm{tr}}\,\mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t}\sum_{v'\in V}\mu_{\mathrm{tr}}(v')\,\mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\,\mathrm{d}t \\
&+ \int_{b/c(\mathrm{tr})}^{\infty}\int_{\pi'} \Big(\rho_{\mathrm{t}}(\mathrm{tr})\cdot \tfrac{b}{c(\mathrm{tr})} + \rho_{\mathrm{f}}(v)\Big)\,\lambda_{\mathrm{tr}}\,\mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t}\sum_{v'\in V}\mu_{\mathrm{tr}}(v')\,\mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\,\mathrm{d}t
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^{b/c(\mathrm{tr})} \Big(\rho_t(\mathrm{tr})\cdot t + \rho_i(\mathrm{tr})\Big)\, \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t} \sum_{v'\in V} \mu_{\mathrm{tr}}(v') \int_{\pi'} \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\,\mathrm{d}t \\
&\quad + \int_0^{b/c(\mathrm{tr})} \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t} \int_{\pi'} cbr(\pi', b - c(\mathrm{tr})\cdot t) \sum_{v'\in V} \mu_{\mathrm{tr}}(v')\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\,\mathrm{d}t \\
&\quad + \int_{b/c(\mathrm{tr})}^{\infty} \Big(\rho_t(\mathrm{tr})\cdot \tfrac{b}{c(\mathrm{tr})} + \rho_f(v)\Big)\, \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t} \sum_{v'\in V} \mu_{\mathrm{tr}}(v') \int_{\pi'} \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\,\mathrm{d}t \\
&\overset{(*)}{=} \int_0^{b/c(\mathrm{tr})} \Big(\rho_t(\mathrm{tr})\cdot t + \rho_i(\mathrm{tr})\Big)\, \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t}\,\mathrm{d}t + \int_{b/c(\mathrm{tr})}^{\infty} \Big(\rho_t(\mathrm{tr})\cdot \tfrac{b}{c(\mathrm{tr})} + \rho_f(v)\Big)\, \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t}\,\mathrm{d}t \\
&\quad + \int_0^{b/c(\mathrm{tr})} \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t} \int_{\pi'} cbr(\pi', b - c(\mathrm{tr})\cdot t) \sum_{v'\in V} \mu_{\mathrm{tr}}(v')\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\,\mathrm{d}t \\
&= \left(\frac{\rho_t(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_i(\mathrm{tr})\right)\left(1 - \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}}\right) + \rho_f(v)\cdot \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}} \\
&\quad + \int_0^{b/c(\mathrm{tr})} \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t} \int_{\pi'} cbr(\pi', b - c(\mathrm{tr})\cdot t) \sum_{v'\in V} \mu_{\mathrm{tr}}(v')\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\,\mathrm{d}t,
\end{aligned} \tag{6}
$$

where $(*)$ holds since $\int_{\pi'} \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi') = 1$ and $\sum_{v'\in V} \mu_{\mathrm{tr}}(v') = 1$.

With Eq. (6) one can determine $\overset{\star}{\mathcal{R}}{}^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v, \mathrm{tr}, b)$:

$$
\begin{aligned}
&\overset{\star}{\mathcal{R}}{}^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v, \mathrm{tr}, b) \\
&\quad = \underset{\sigma_1\in\Sigma_1^{\mathrm{tr}}}{\mathrm{opt}_1}\ \underset{\sigma_2\in\Sigma_2}{\mathrm{opt}_2}\ \mathcal{R}_{\sigma_1,\sigma_2}(v, \mathrm{tr}, b) && \text{(* by Eq. (2) *)} \\
&\quad = \left(\frac{\rho_t(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_i(\mathrm{tr})\right)\left(1 - \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}}\right) + \rho_f(v)\cdot \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}} \\
&\qquad + \underset{\sigma_1\in\Sigma_1^{\mathrm{tr}}}{\mathrm{opt}_1}\ \underset{\sigma_2\in\Sigma_2}{\mathrm{opt}_2} \int_0^{b/c(\mathrm{tr})} \Big(\lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t} \int_{\pi'} \big(cbr(\pi', b - c(\mathrm{tr})\cdot t)\cdot \sum_{v'\in V} \mu_{\mathrm{tr}}(v')\big)\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\Big)\,\mathrm{d}t && \text{(* by Eq. (6) *)} \\
&\quad = \left(\frac{\rho_t(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_i(\mathrm{tr})\right)\left(1 - \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}}\right) + \rho_f(v)\cdot \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}} \\
&\qquad + \underset{\sigma_1\in\Sigma_1^{\mathrm{tr}}}{\mathrm{opt}_1}\ \underset{\sigma_2\in\Sigma_2}{\mathrm{opt}_2} \int_0^{b/c(\mathrm{tr})} \Big(\lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t} \sum_{v'\in V} \mu_{\mathrm{tr}}(v')\cdot \int_{\pi'} cbr(\pi', b - c(\mathrm{tr})\cdot t)\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\Big)\,\mathrm{d}t \\
&\quad \overset{(*)}{=} \left(\frac{\rho_t(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_i(\mathrm{tr})\right)\left(1 - \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}}\right) + \rho_f(v)\cdot \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}} \\
&\qquad + \int_0^{b/c(\mathrm{tr})} \Big(\lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t} \sum_{v'\in V} \mu_{\mathrm{tr}}(v')\cdot \underset{\sigma_1\in\Sigma_1^{\mathrm{tr}}}{\mathrm{opt}_1}\ \underset{\sigma_2\in\Sigma_2}{\mathrm{opt}_2} \int_{\pi'} cbr(\pi', b - c(\mathrm{tr})\cdot t)\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\Big)\,\mathrm{d}t \\
&\quad \overset{(\dagger)}{=} \left(\frac{\rho_t(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_i(\mathrm{tr})\right)\left(1 - \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}}\right) + \rho_f(v)\cdot \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}} \\
&\qquad + \int_0^{b/c(\mathrm{tr})} \Big(\lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t} \sum_{v'\in V} \mu_{\mathrm{tr}}(v')\cdot \underset{\sigma_1\in\Sigma_1}{\mathrm{opt}_1}\ \underset{\sigma_2\in\Sigma_2}{\mathrm{opt}_2} \int_{\pi'} cbr(\pi', b - c(\mathrm{tr})\cdot t)\, \mathrm{dPr}_{v',\sigma_1,\sigma_2}(\pi')\Big)\,\mathrm{d}t \\
&\quad \overset{(\ddagger)}{=} \left(\frac{\rho_t(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_i(\mathrm{tr})\right)\left(1 - \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}}\right) + \rho_f(v)\cdot \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\cdot b}{c(\mathrm{tr})}} \\
&\qquad + \int_0^{b/c(\mathrm{tr})} \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}}\cdot t} \sum_{v'\in V} \mu_{\mathrm{tr}}(v')\cdot \mathbb{E}cbr^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v', b - c(\mathrm{tr})\cdot t)\,\mathrm{d}t,
\end{aligned} \tag{7}
$$

where the role of the integral and opt operators can be changed in $(*)$ as the integral is over variable t and both σ_1^t and σ_2^t are function of t. Moreover, (†) holds since σ_1^t and σ_2^t are now optimising paths from v', leaving no chance for tr to be selected. Hence they function like general strategies. Lastly (‡) follows from Eq. (1).

(c) $\mathrm{tr} \in T_{\mathrm{M}}(v)$, $c(\mathrm{tr}) = 0$: Here we again split the paths as it is done for case (b). A path π starting from v is then represented by $\pi = v \xrightarrow{t,\mathrm{tr}} \pi'$. The splitting of the probability measure follows Eq. (5) accordingly. However the reward splitting in this case is different since staying at state v has no cost. Therefore we can write:

$$cbr(\pi, b) = \rho_{\mathrm{t}}(\mathrm{tr}) \cdot t + \rho_{\mathrm{i}}(\mathrm{tr}) + cbr(\pi', b). \tag{8}$$

We proceed with the simplification of $\mathcal{R}_{\sigma_1,\sigma_2}(v, \mathrm{tr}, b)$ according to Eqs. (5) and (8).

$$\begin{aligned}
\mathcal{R}_{\sigma_1,\sigma_2} & (v, \mathrm{tr}, b) \\
&= \int_{\pi} cbr(\pi, b)\, \mathrm{dPr}_{v,\sigma_1,\sigma_2}(\pi) \\
&= \int_0^{\infty} \int_{\pi'} \Big(\rho_{\mathrm{t}}(\mathrm{tr}) \cdot t + \rho_{\mathrm{i}}(\mathrm{tr}) + cbr(\pi', b)\Big)\, \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \sum_{v' \in V} \mu_{\mathrm{tr}}(v')\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\, \mathrm{d}t \\
&= \int_0^{\infty} \int_{\pi'} \Big(\rho_{\mathrm{t}}(\mathrm{tr}) \cdot t + \rho_{\mathrm{i}}(\mathrm{tr})\Big)\, \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \sum_{v' \in V} \mu_{\mathrm{tr}}(v')\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\, \mathrm{d}t \\
&\quad + \int_0^{\infty} \int_{\pi'} cbr(\pi', b)\, \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \sum_{v' \in V} \mu_{\mathrm{tr}}(v')\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\, \mathrm{d}t \\
&= \int_0^{\infty} (\rho_{\mathrm{t}}(\mathrm{tr}) \cdot t + \rho_{\mathrm{i}}(\mathrm{tr}))\, \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \int_{\pi'} \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\, \mathrm{d}t \\
&\quad + \int_0^{\infty} \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \int_{\pi'} cbr(\pi', b)\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\, \mathrm{d}t \\
&\overset{(*)}{=} \int_0^{\infty} \Big(\rho_{\mathrm{t}}(\mathrm{tr}) \cdot t + \rho_{\mathrm{i}}(\mathrm{tr})\Big)\, \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t}\, \mathrm{d}t \\
&\quad + \int_0^{\infty} \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \int_{\pi'} cbr(\pi', b)\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\, \mathrm{d}t \\
&= \frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr}) + \int_0^{\infty} \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \int_{\pi'} cbr(\pi', b)\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\, \mathrm{d}t,
\end{aligned} \tag{9}$$

where $(*)$ follows from $\int_{\pi'} \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi') = 1$ and $\sum_{v' \in V} \mu_{\mathrm{tr}}(v') = 1$.

With Eq. (9) one can determine $\overset{\star\ \mathrm{opt}_1}{\mathcal{R}}_{\mathrm{opt}_2}(v, \mathrm{tr}, b)$:

$$\begin{aligned}
\overset{\star\ \mathrm{opt}_1}{\mathcal{R}}_{\mathrm{opt}_2}(v, \mathrm{tr}, b) &= \underset{\sigma_1 \in \Sigma_1^{\mathrm{tr}}}{\mathrm{opt}_1}\ \underset{\sigma_2 \in \Sigma_2}{\mathrm{opt}_2}\ \mathcal{R}_{\sigma_1,\sigma_2}(v, \mathrm{tr}, b) && \text{(* by Eq. (2) *)} \\
&= \underset{\sigma_1 \in \Sigma_1^{\mathrm{tr}}}{\mathrm{opt}_1}\ \underset{\sigma_2 \in \Sigma_2}{\mathrm{opt}_2} \int_0^{\infty} \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \int_{\pi'} cbr(\pi', b)\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\, \mathrm{d}t + \frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr}) && \text{(* by Eq. (9) *)} \\
&\overset{(*)}{=} \int_0^{\infty} \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot \underset{\sigma_1 \in \Sigma_1^{\mathrm{tr}}}{\mathrm{opt}_1}\ \underset{\sigma_2 \in \Sigma_2}{\mathrm{opt}_2} \int_{\pi'} cbr(\pi', b)\, \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi')\, \mathrm{d}t + \frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr}) \\
&\overset{(\dagger)}{=} \int_0^{\infty} \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot \underset{\sigma_1 \in \Sigma_1}{\mathrm{opt}_1}\ \underset{\sigma_2 \in \Sigma_2}{\mathrm{opt}_2} \int_{\pi'} cbr(\pi', b)\, \mathrm{dPr}_{v',\sigma_1,\sigma_2}(\pi')\, \mathrm{d}t + \frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr})
\end{aligned}$$

$$= \int_0^\infty \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot \mathbb{E}cbr^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v', b)\, \mathrm{d}t + \frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr}) \qquad (\text{* by Eq. (1) *})$$

$$= \int_0^\infty \lambda_{\mathrm{tr}}\, \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t}\, \mathrm{d}t \cdot \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot \mathbb{E}cbr^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v', b) + \frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr})$$

$$= \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot \mathbb{E}cbr^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v', b) + \frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr}), \tag{10}$$

where, similar to case (b), the role of the integral and opt operators can be changed in $(*)$ as the integral is over the variable t and both σ_1^t and σ_2^t are functions of t. Moreover, (†) holds, since σ_1^t and σ_2^t are now optimising paths from v', leaving no chance for tr to be selected. Hence they function like general strategies.

(d) $\mathrm{tr} \in T_{\mathrm{P}}(v)$: We split each path starting from v at the point where transition tr is taken. In contrast to the cases (b) and (c), the transition is taken at time zero as it is probabilistic and therefore immediate. Hence, no cost is imposed, and also no transient reward is gained by tr. We can thus write $\pi = v \xrightarrow{0,\mathrm{tr}} \pi'$ and then:

$$cbr(\pi, b) = \rho_{\mathrm{i}}(\mathrm{tr}) + cbr(\pi', b). \tag{11}$$

The probability measure can be split in the same way. Similar to cases (b) and (c) we first need to construct a new strategy from a given arbitrary schedule σ_i, $i = \{1, 2\}$. Its aim is to mimic the decision of σ_i on path π when it takes the suffix of π after tr is taken. Formally strategy σ_i^{tr} resolves the nondeterminism for the suffix π' of path π after the splitting point as σ_i does it for π, i. e. for $\pi = v \xrightarrow{0,\mathrm{tr}} \pi'$, $\sigma_i^{\mathrm{tr}}(\pi') = \sigma_i(\pi)$. We can subsequently split the probability measure of path π:

$$\mathrm{dPr}_{v,\sigma_1,\sigma_2}(\pi) = \sum_{v' \in V} \mu_{\mathrm{tr}}(v')\, \mathrm{dPr}_{v',\sigma_1^{\mathrm{tr}},\sigma_2^{\mathrm{tr}}}(\pi'). \tag{12}$$

We proceed with the simplification of $\mathcal{R}_{\sigma_a,\sigma_2}(v, \mathrm{tr}, b)$ according to Eqs. (11) and (12):

$$\begin{aligned}
\mathcal{R}_{\sigma_1,\sigma_2}(v, \mathrm{tr}, b) &= \int_\pi cbr(\pi, b)\, \mathrm{dPr}_{v,\sigma_1,\sigma_2}(\pi) \\
&= \int_{\pi'} \Big(\rho_{\mathrm{i}}(\mathrm{tr}) + cbr(\pi', b)\Big) \sum_{v' \in V} \mu_{\mathrm{tr}}(v')\, \mathrm{dPr}_{v',\sigma_1^{\mathrm{tr}},\sigma_2^{\mathrm{tr}}}(\pi') \\
&= \rho_{\mathrm{i}}(\mathrm{tr}) \cdot \int_{\pi'} \sum_{v' \in V} \mu_{\mathrm{tr}}(v')\, \mathrm{dPr}_{v',\sigma_1^{\mathrm{tr}},\sigma_2^{\mathrm{tr}}}(\pi') + \int_{\pi'} cbr(\pi', b) \sum_{v' \in V} \mu_{\mathrm{tr}}(v')\, \mathrm{dPr}_{v',\sigma_1^{\mathrm{tr}},\sigma_2^{\mathrm{tr}}}(\pi') \\
&\overset{(*)}{=} \rho_{\mathrm{i}}(\mathrm{tr}) + \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot \int_{\pi'} cbr(\pi', b)\, \mathrm{dPr}_{v',\sigma_1^{\mathrm{tr}},\sigma_2^{\mathrm{tr}}}(\pi'),
\end{aligned} \tag{13}$$

where $(*)$ follows from $\int_{\pi'} \mathrm{dPr}_{v',\sigma_1^t,\sigma_2^t}(\pi') = 1$ and $\sum_{v' \in V} \mu_{\mathrm{tr}}(v') = 1$.

With Eq. (13) one can determine $\overset{\star}{\mathcal{R}}{}^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v, \mathrm{tr}, b)$:

$$\begin{aligned}
\overset{\star}{\mathcal{R}}{}^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v, \mathrm{tr}, b) &= \underset{\sigma_1 \in \Sigma_1^{\mathrm{tr}}}{\mathrm{opt}_1}\ \underset{\sigma_2 \in \Sigma_2}{\mathrm{opt}_2}\ \mathcal{R}_{\sigma_1,\sigma_2}(v, \mathrm{tr}, b) && (\text{* by Eq. (2) *}) \\
&= \rho_{\mathrm{i}}(\mathrm{tr}) + \underset{\sigma_1 \in \Sigma_1^{\mathrm{tr}}}{\mathrm{opt}_1}\ \underset{\sigma_2 \in \Sigma_2}{\mathrm{opt}_2} \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot \int_{\pi'} cbr(\pi', b)\, \mathrm{dPr}_{v',\sigma_1^{\mathrm{tr}},\sigma_2^{\mathrm{tr}}}(\pi') && (\text{* by Eq. (13) *}) \\
&= \rho_{\mathrm{i}}(\mathrm{tr}) + \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot \underset{\sigma_1 \in \Sigma_1^{\mathrm{tr}}}{\mathrm{opt}_1}\ \underset{\sigma_2 \in \Sigma_2}{\mathrm{opt}_2} \int_{\pi'} cbr(\pi', b)\, \mathrm{dPr}_{v',\sigma_1^{\mathrm{tr}},\sigma_2^{\mathrm{tr}}}(\pi') \\
&\overset{(*)}{=} \rho_{\mathrm{i}}(\mathrm{tr}) + \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot \underset{\sigma_1 \in \Sigma_1}{\mathrm{opt}_1}\ \underset{\sigma_2 \in \Sigma_2}{\mathrm{opt}_2} \int_{\pi'} cbr(\pi', b)\, \mathrm{dPr}_{v',\sigma_1,\sigma_2}(\pi') \\
&\overset{(\dagger)}{=} \rho_{\mathrm{i}}(\mathrm{tr}) + \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot \mathbb{E}cbr^{\mathrm{opt}_1}_{\mathrm{opt}_2}(v', b),
\end{aligned} \tag{14}$$

where $(*)$ holds since σ_1^{tr} and σ_2^{tr} are now optimising paths from v', leaving no chance for tr to be selected. Hence they function like general strategies. Furthermore, (†) follows from Eq. (1).

Plugging Eqs. (3), (7), (10) and (14) into Eq. (2) directly indicates that $\mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}$ is a fixed point of operator $\Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1}$. The next part of the proof shows that $\mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}$ is indeed the least fixed point of $\Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1}$.

Let F be any fixed point of operator $\Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1}$, we show that $\mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}(v, b) \leq F(v, b)$ for all $v \in V$ and $b \geq 0$. We show by $\Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1}[n]$ $(n > 0)$ the n-level recursive composition of operator $\Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1}$ and write $F[n] = \Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1}[n](F[0])$, where $F[0]$ is the starting bottom function. For the optimal ECR, $\mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}[n](v, b)$ intuitively refers to the optimal expected cost-bounded reward from v gained by taking up to n transitions. Its bottom function is thus defined as

$$\mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}[0](v, b) = \begin{cases} \rho_{\mathrm{f}}(v), & \text{if } \mathrm{tr} \in T_{\mathrm{M}}(v) \wedge c(\mathrm{tr}) > 0 \wedge b = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

Now we consider an arbitrary fixed point of operator $\Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1}$. Obviously it holds that $\mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}[0](v, b) \leq F(v, b)$ for all $v \in V$ and $b \in \mathbb{R}_{\geq 0}$. It is not hard to prove, by induction on n, that $\forall\, n.\ \mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}[n](v, b) \leq F(v, b)$. It is enough to inductively show for each case of the fixed point characterisation, discussed in the first part of the proof, that the inequality holds. Finally it holds for $v \in V$ and $b \in \mathbb{R}_{\geq 0}$ that

$$\mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}(v, b) = \lim_{n\to\infty} \mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}[n](v, b) \leq F(v, b).$$

The proof that sequence $\{\mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}[n](v, b)\}_{n\in\mathbb{N}}$ converges to $\mathbb{E}cbr_{\mathrm{opt}_2}^{\mathrm{opt}_1}(v, b)$ can be done via a slight adaptation of [Neu10, Theorem 5.1]. □

C. Proof of Theorem 3.2

We recall the theorem here:

Theorem 3.2 (Measure preservation) *Let $\mathcal{G}$ be a stochastic game with reward structure ρ, cost function c, cost bound $b \in \mathbb{R}_{\geq 0}$, $v \in V$, and* $\mathrm{opt}_1, \mathrm{opt}_2 \in \{\inf, \sup\}$. *Then we have*

$$\mathbb{E}cbr_{\mathcal{G},\rho,c}^{\mathrm{opt}_1,\mathrm{opt}_2}(v, b) = \mathbb{E}\mathrm{tbr}_{\mathcal{G}^c,\rho^c}^{\mathrm{opt}_1,\mathrm{opt}_2}(v, b).$$

Proof The proof is done by showing that the original and the transformed games have indeed the same fixed point characterisation for the optimal ECR. We start with the fixed point characterisation of the original game, described in Theorem 3.1, and consider an arbitrary state v and its transitions case by case.

We first look at the case when $\mathrm{tr} \in T_{\mathrm{M}}(v)$, $c(\mathrm{tr}) > 0$, and $b > 0$. We change then the variables in this case according to the transformation given in Definition 3.1. It is done by introducing a transition tr^c of the transformed model with the following quantities: $\lambda_{\mathrm{tr}^c} := \frac{\lambda_{\mathrm{tr}}}{c(\mathrm{tr})}$ and $\rho_{\mathrm{t}}^c(\mathrm{tr}^c) := \frac{\rho_{\mathrm{t}}(\mathrm{tr})}{c(\mathrm{tr})}$. Other quantities of tr^c are inherited by tr as suggested by the transformation, i. e. $\mu_{\mathrm{tr}^c} := \mu_{\mathrm{tr}}$ and $\rho_{\mathrm{i}}^c(\mathrm{tr}^c) := \rho_{\mathrm{i}}(\mathrm{tr})$. Furthermore we change the integration in this case by substitution of t with $\tau := c(\mathrm{tr}) \cdot t$, and thereby $\mathrm{d}\tau = c(\mathrm{tr})\,\mathrm{d}t$. With this we rewrite the case:

$$\begin{aligned} &\int_0^{b/c(\mathrm{tr})} \lambda_{\mathrm{tr}} \cdot \mathrm{e}^{-\lambda_{\mathrm{tr}} \cdot t} \cdot \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot F(v', b - c(\mathrm{tr}) \cdot t)\,\mathrm{d}t + \left(\frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr})\right) \cdot \left(1 - \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\, b}{c(\mathrm{tr})}}\right) \\ &= \int_0^{b} \lambda_{\mathrm{tr}} \cdot \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}}{c(\mathrm{tr})} \cdot \tau} \cdot \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot F(v', b - \tau) \frac{1}{c(\mathrm{tr})}\,\mathrm{d}\tau + \left(\frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr})\right) \cdot \left(1 - \mathrm{e}^{-\frac{\lambda_{\mathrm{tr}}\, b}{c(\mathrm{tr})}}\right) \\ &= \int_0^{b} \lambda_{\mathrm{tr}^c} \cdot \mathrm{e}^{-\lambda_{\mathrm{tr}^c} \cdot \tau} \cdot \sum_{v' \in V} \mu_{\mathrm{tr}^c}(v') \cdot F(v', b - \tau)\,\mathrm{d}\tau + \left(\frac{\rho_{\mathrm{t}}^c(\mathrm{tr}^c)}{\lambda_{\mathrm{tr}^c}} + \rho_{\mathrm{i}}^c(\mathrm{tr})\right) \cdot \left(1 - \mathrm{e}^{-\lambda_{\mathrm{tr}^c} \cdot b}\right). \end{aligned} \tag{16}$$

Now we consider the case that $\mathrm{tr} \in T_{\mathrm{M}}(v)$ and $c(\mathrm{tr}) = 0$ and apply the transformation therein. We therefore turn it into a probabilistic transition tr^c of the transformed game by $\lambda_{\mathrm{tr}^c} := \infty$, $\mu_{\mathrm{tr}^c} := \mu(\mathrm{tr})$ and $\rho_{\mathrm{i}}^c(\mathrm{tr}^c) := \frac{\rho_{\mathrm{t}}(\mathrm{tr})}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr})$. With this we get:

$$\frac{\rho_{\mathrm{t}}(v)}{\lambda_{\mathrm{tr}}} + \rho_{\mathrm{i}}(\mathrm{tr}) + \sum_{v' \in V} \mu_{\mathrm{tr}}(v') \cdot F(v', b) = \rho_{\mathrm{i}}^c(\mathrm{tr}^c) + \sum_{v' \in V} \mu_{\mathrm{tr}^c}(v') \cdot F(v', b). \tag{17}$$

There are still two cases left: (1) The case when $\mathrm{tr} \in T_{\mathrm{P}}(v)$, and (2) the complement of all other cases. For both of them we just take the transition as it is and do not change anything. Putting this with Eqs. (16) and (17) all together gives:

$$\Omega_{\mathrm{opt}_2}^{\mathrm{opt}_1}(F)(v, b) = \underset{\mathrm{tr} \in T(v)}{\mathrm{opt}_{[v]}} \begin{cases} \int_0^b \lambda_{\mathrm{tr}^c} \cdot \mathrm{e}^{-\lambda_{\mathrm{tr}^c} \cdot \tau} \cdot \sum_{v' \in V} \mu_{\mathrm{tr}^c}(v') \cdot F(v', b - \tau)\, \mathrm{d}\tau & \\ \quad + \left(\frac{\rho_{\mathrm{t}}^c(\mathrm{tr}^c)}{\lambda_{\mathrm{tr}^c}} + \rho_{\mathrm{i}}^c(\mathrm{tr}) \right) \cdot \left(1 - \mathrm{e}^{-\lambda_{\mathrm{tr}^c} \cdot b} \right), & \text{if } \mathrm{tr} \in T_{\mathrm{M}}(v) \wedge b > 0, \\ \rho_{\mathrm{i}}^c(\mathrm{tr}^c) + \sum_{v' \in V} \mu_{\mathrm{tr}^c}(v') \cdot F(v', b), & \text{if } \mathrm{tr}^c \in T_{\mathrm{P}}(v), \\ \rho_{\mathrm{f}}(v), & \text{otherwise.} \end{cases} \tag{18}$$

Note that two cases, when the transition is Markovian with zero cost, and when it is probabilistic, are now merged, since we turn Markovian transitions with zero cost into probabilistic transitions. Moreover, after the transformation being Markovian implies having nonzero cost; as a consequence, cost checking is omitted from the cases.

It is not hard to see that the fixed point characterisation offered by Eq. (18) coincides with that of the transformed game for expected time-bounded reward, similar to the one in [BFH+15, Lemma 1], which is adapted in our notations and recalled in Appendix A. It means that even though the operators characterise different objectives and their representations are different, they can be derived from each other and therefore their least fixed point is the same. □

References

[ADD99] Ash RB, Doléans-Dade CA (1999) Probability & measure theory, 2nd edn. Academic Press, New York

[AHK03] Andova S, Hermanns H, Katoen JP (2003) Discrete-time rewards model-checked. In: Int'l conf. on formal modeling and analysis of timed systems (FORMATS). Lecture notes in computer science. vol 2791. Springer, Berlin, pp 88–104

[BCS10] Boudali H, Crouzen P, Stoelinga M (2010) A rigorous, compositional, and extensible framework for dynamic fault tree analysis. IEEE Tran Depend Sec Comput 7(2):128–143

[BDF81] Bruno JL, Downey PJ, Frederickson GN (1981) Sequencing tasks with exponential service times to minimize the expected flow time or makespan. J ACM 28(1):100–113

[BFH+14] Braitling B, María Ferrer Fioriti L, Hatefi H, Wimmer R, Becker B, Hermanns H (2014) MeGARA: menu-based game abstraction refinement for Markov automata. In: Bertrand N, Bertolussi L (eds) Int'l workshop on quantitative aspects of programming languages and systems (QAPL). In: Electronic proceedings in theoretical computer science, vol 154. Open Publishing Association, Grenoble, pp 48–63

[BFH+15] Braitling B, María Ferrer FL, Hatefi H, Wimmer R, Hermanns H, Becker B (2015) Abstraction-based computation of reward measures for Markov automata. In: Int'l conf. on verification, model checking, and abstract interpretation (VMCAI). Lecture Notes in Computer Science, vol 8931. Springer, Berlin, pp 172–189

[BFK+13] Brázdil T, Forejt V, Krcál J, Kretínský J, Kucera A (2013) Continuous-time stochastic games with time-bounded reachability. Inf Comput 224:46–70

[BHHK00] Baier C, Haverkort BR, Hermanns H, Katoen JP (2000) On the logical characterisation of performability properties. In: Int'l colloquium on automata, languages, and programming (ICALP). In: Lecture notes in computer science, vol 1853. Springer, Berlin, pp 780–792

[BHHK08] Baier C, Haverkort BR, Hermanns H, Katoen JP (2008) Reachability in continuous-time Markov reward decision processes. In: Logic and automata: history and perspectives. Honor of Wolfgang Thomas. Texts in logic and games, vol 2. Amsterdam University Press, Amsterdam, pp 53–72

[BHHK15] Butkova Y, Hatefi H, Hermanns H, Krcál J (2015) Optimal continuous time Markov decisions. In: Finkbeiner B, Pu G, Zhang L (eds) Int'l symp. on automated technology for verification and analysis (ATVA). Lecture notes in computer science, vol 9364. Springer, Shanghai, pp 166–182

[BK08] Baier C, Katoen J-P (2008) Principles of model checking. The MIT Press, Massachusetts

[BS11] Buchholz P, Schulz I (2011) Numerical analysis of continuous time Markov decision processes over finite horizons. Comput Oper Res 38(3):651–659

[CKKP05] Cloth L, Katoen J-P, Khattri M, Pulungan R (2005) Model checking Markov reward models with impulse rewards. In: Int'l conf. on dependable systems and networks (DSN). IEEE Computer Society, New York, pp 722–731

[EHKZ13] Eisentraut C, Hermanns H, Katoen J-P, Zhang L (2013) A semantics for every GSPN. In: Proc. of petri nets. Lecture notes in computer science, vol 7927. Springer, Berlin, pp 90–109
[EHZ10] Eisentraut C, Hermanns H, Zhang L (2010) On probabilistic automata in continuous time. In: Annual IEEE symp. on logic in computer science (LICS). IEEE Computer Society, New York, pp 342–351
[Fu14a] Fu H (2014) Maximal cost-bounded reachability probability on continuous-time Markov decision processes. In: Int'l conf. on foundations of software science and computation structures (FoSSaCS). Lecture notes in computer science, vol 8412. Springer, Berlin, pp 73–87
[Fu14b] Fu H (2014) Verifying probabilistic systems: new algorithms and complexity results. PhD thesis, RWTH Aachen University
[GBK16] Gburek D, Baier C, Klüppelholz S (2016) Composition of stochastic transition systems based on spans and couplings. In: 43rd international colloquium on automata, languages, and programming, ICALP 2016, July 11–15, 2016, Rome, Italy, pp 102:1–102:15
[GHH+13] Guck D, Hatefi H, Hermanns H, Katoen J.-P. Timmer M (2013) Modelling, reduction and analysis of Markov automata. In: Int'l conf. on quantitative evaluation of systems (QEST). Lecture notes in computer science, vol 8054. Springer, Berlin, pp 55–71
[GHH+14] Guck D, Hatefi H, Hermanns H, Katoen JP, Timmer M (2014) Analysis of timed and long-run objectives for Markov automata. Logical Methods Comput Sci 10(3)
[GHKN12] Guck D, Han T, Katoen JP, Neuhäußer MR (2012) Quantitative timed analysis of interactive Markov chains. In: NASA formal methods symposium (NFM). Lecture notes in computer science, vol 7226. Springer, Berlin, pp 8–23
[GTH+14] Guck D, Timmer M, Hatefi H, Ruijters E, Stoelinga M (2014) Modelling and analysis of Markov reward automata. In: Int'l symp. on automated technology for verification and analysis (ATVA). Lecture notes in computer science, vol 8837. Springer, Berlin, pp 168–184
[HBW+15] Hatefi H, Braitling B, Wimmer R, María Ferrer Fioriti L, Hermanns H, Becker B (2015) Cost vs. time in stochastic games and Markov automata. In: Li X, Liu Z, Yi W (eds) int'l symp. on dependable software engineering: theory, tools and applications (SETTA). Lecture notes in computer science, vol 9409. Springer, Nanjing, pp 19–34
[Her02] Hermanns H (2002) Interactive Markov chains: the quest for quantified quality. In: Lecture notes in computer science, vol 2428. Springer, Berlin
[HH12] Hatefi H, Hermanns H (2012) Model checking algorithms for Markov automata. ECEASST 53
[Joh08] Johr S (2008) Model checking compositional Markov systems. PhD thesis, Saarland University, Saarbrücken
[Mil68] Miller BL (1968) Finite state continuous time Markov decision processes with a finite planning horizon. SIAM J Control 6(2):266–280
[Neu10] Neuhäußer MR (2010) Model checking nondeterministic and randomly timed systems. PhD thesis, RWTH Aachen University and University of Twente
[NS03] Neyman A, Sorin S (eds) (2003) Stochastic games and applications. Springer, Berlin
[NZ10] Neuhäußer MR, Zhang L (2010) Time-bounded reachability probabilities in continuous-time Markov decision processes. In: Int'l conf. on quantitative evaluation of systems (QEST). IEEE Computer Society, New York, pp 209–218
[QQP01] Qiu Q, Qu Q, Pedram M (2001) Stochastic modeling of a power-managed system-construction and optimization. IEEE Trans CAD Integrat Circ Syst 20(10):1200–1217
[SBGM00] Simunic T, Benini L, Glynn P.W, De Micheli G (2000) Dynamic power management for portable systems. In: Annual int'l conf. on mobile computing and networking (MOBICOM), Boston, MA, USA, August, pp 11–19
[Seg95] Segala R (1995) A compositional trace-based semantics for probabilistic automata. In: Int'l conf. on concurrency theory (CONCUR). Lecture notes in computer science, vol 962. Springer, Berlin, pp 234–248
[Sha53] Shapley LS (1953) Stochastic games. Proc Natl Acad Sci USA 39(10):1095
[Tar55] Tarski A (1955) A lattice-theoretical fixpoint theorem and its applications. Pac J Math 5(2):285–309
[TKvdPS12] Timmer M, Katoen J-P van de Pol J, Stoelinga M (2012) Efficient modelling and generation of Markov automata. In: Int'l conf. on concurrency theory (CONCUR). Lecture notes in computer science, vol 7454. Springer, Berlin, pp 364–379
[TvdPS13] Timmer M, van de Pol J, Stoelinga M (2013) Confluence reduction for Markov automata. In: Int'l conf. on formal modeling and analysis of timed systems (FORMATS). Lecture notes in computer science, vol 8053. Springer, Berlin, pp 243–257
[WJ06] Wolovick N, Johr S (2006) A characterization of meaningful schedulers for continuous-time Markov decision processes. In: Asarin E, Bouyer P (eds) Int'l conf. on formal modeling and analysis of timed systems (FORMATS). Lecture notes in computer science, vol 4202. Springer, Paris, pp 352–367

Received 25 March 2016
Accepted in revised form 22 November 2016 by Cliff Jones, Xuandong Li, and Zhiming Liu
Published online 9 January 2017