Quantitative Automata-based Controller Synthesis for Non-Autonomous Stochastic Hybrid Systems*

Ilya Tkachev
TU Delft
i.tkachev@tudelft.nl

Alexandru Mereacre
University of Oxford
mereacre@cs.ox.ac.uk

Joost-Pieter Katoen
RWTH Aachen
katoen@cs.rwth-aachen.de

Alessandro Abate
TU Delft
a.abate@tudelft.nl

ABSTRACT

This work deals with Markov processes that are defined over an uncountable state space (possibly hybrid) and embedding non-determinism in the shape of a control structure. The contribution looks at the problem of optimization, over the set of allowed controls, of probabilistic specifications defined by automata – in particular, the focus is on deterministic finite-state automata. This problem can be reformulated as an optimization of a probabilistic reachability property over a product process obtained from the model for the specification and the model of the system. Optimizing over automata-based specifications thus leads to maximal or minimal probabilistic reachability properties. For both setups, the contribution shows that these problems can be sufficiently tackled with history-independent Markov policies. This outcome has relevant computational repercussions: in particular, the work develops a discretization procedure leading into standard optimization problems over Markov decision processes. Such procedure is associated with exact error bounds and is experimentally tested on a case study.

Categories and Subject Descriptors

G.3 [**Mathematics of Computing**]: PROBABILITY AND STATISTICS—*Stochastic processes*

General Terms

Theory, verification

*This work is supported by the European Commission MoVeS project FP7-ICT-2009-5 257005, by the European Commission Marie Curie grant MANTRAS 249295, by the European Commission NoE HYCON2 FP7-ICT-2009-5 257462, and by the NWO VENI grant 016.103.020.

HSCC'13, April 8–11, 2013, Philadelphia, Pennsylvania, USA.

Keywords

stochastic hybrid systems, probabilistic reachability, formal verification, stochastic optimal control, approximate abstractions.

1. INTRODUCTION

Stochastic Hybrid Systems (SHS) are a general and widely applicable mathematical framework involving the interaction of discrete, continuous, and probabilistic dynamics. Because of their generality, SHS have been applied over many areas, including telecommunication networks, manufacturing systems, transportation, and biological systems [8, 9].

SHS can be abstractly regarded as general Markov processes defined over an uncountable (in particular, hybrid) state space and embedded with non-determinism in the sense that they are dependent on a control structure [7]. The allowed control policies are functions of the history: in other words, we allow the control inputs to depend not only on the current state, but also on past trajectory realization and on past choices of control inputs. Markov policies are important special instances of these policies and depend solely on the current state.

SHS models are properly tagged by a labeling function that maps points in the state space to elements of a finite labeling set, which we refer to as an alphabet. SHS are structurally useful in the verification of linear time specifications, for example specifications expressed as deterministic finite-state automata (DFA) with labels. The work in [4] has shown that the verification of linear time specifications (in particular, DFA properties) can be performed by solving a probabilistic reachability problem over a new SHS, which is obtained by taking the cross product between the SHS and the DFA. Probabilistic reachability can then be practically assessed by computing its dual, namely probabilistic invariance [5].

This work generalizes the results in [4] to the case of control-dependent SHS and of DFA specifications. Since the verification of DFA specifications boils down to solving a probabilistic reachability problem, we consider this setup both in its finite- and infinite-horizon formulations. As the models under study are control dependent, we look into the possible maximal and minimal probabilistic reachability formulations: these can be studied as limits of dynamic programming recursions or as solutions of integral equations [7]. Since the cost functions expressing the dynamic programming scheme for probabilistic reachability take a multiplicative form [5], the analysis of their properties is in general difficult. This has lead several works in the

literature [18, 10] to try formulating the reach-avoid problem (a generalization of the reachability problem) in terms of additive costs, where the dynamic programming theory is mature [7, 13]. To the best of our knowledge, inspired by the work in [11] the present work provides such a reduction explicitly for the first time. This as a consequence leads to several important results: first of all, we show that Markov policies are sufficient for the optimal probabilistic reachability; additionally, such a technique allows obtaining Bellman recursion and Bellman fixpoint equations for the reachability probability in the most general case, whereas the results in the literature were either focused on Markov policies exclusively [5, 18] or required structural assumptions on the model [10].

Sufficiency of Markov policies has important computational implications, since optimal controls can be synthesized according to the current value of trajectories, rather than based on the entire past history. Further along this computational line, the work introduces a discretization scheme that reduces the original (uncountable) setup to a finite optimization problem over Markov decision processes. Provided some continuity assumptions on the model are valid, the work shows that the discretization scheme can be associated with exact bounds on the introduced error. The obtained error bounds, inspired by [2], are functions of tunable parameters in the model and thus can be made, at the expense of more computations, arbitrarily small.

The article is structured as follows. Section 2 introduces the model syntax and semantics, as well as the class of linear temporal specifications of interest. Section 3 discusses the problem statement (probabilistic reachability) and its alternative formulation via additive cost functions, and derives the main theoretical results in this work: it shows in particular the sufficiency of Markov policies, and elaborates the minimal and maximal optimization problems over finite and infinite horizons. Section 4 puts forward a discretization scheme for the computation of the quantities in Section 3, with an exact quantification of the introduced errors. Finally, Section 5 presents the experimental outcomes over a case study and Section 6 concludes the paper. Due to space constraints, the proofs of the statements are omitted from this manuscript.

2. PRELIMINARIES

2.1 Notations and model definition

In this section we give a brief recap on Borel spaces and related stochastic kernels. It is common in the literature on the theory of controlled discrete-time Markov processes (cdt-MP) to assume that both the state space and the control space are endowed with a certain topological structure [7, 13]. A topological space X is called a Borel space if it is homeomorphic to a Borel subset of a Polish space[1]. Examples of a Borel space are the Euclidean spaces $\mathbb{R}^n$, its Borel subsets endowed with a subspace topology [16], as well as hybrid spaces [5]. Any Borel space X is assumed to be endowed with a Borel σ-algebra, which is denoted by $\mathscr{B}(X)$. We say that a map $f : X \to Y$ is measurable whenever it is Borel measurable.

Given two Borel spaces X, Y the stochastic kernel on X given Y is the map $P : Y \times \mathscr{B}(X) \to [0,1]$ such that $P(\cdot|y)$ is a probability measure on X for any point $y \in Y$, and such that $P(B|\cdot)$ is a measurable function on Y for any set $B \in \mathscr{B}(X)$. Stochastic kernels provide a natural generalization of update laws for deterministic systems, as we further show below.

We adopt the notation from [13] and consider a tuple $\mathfrak{D} = (X, U, \{U(x)\}_{x\in X}, \mathsf{T})$, where X is a Borel space, referred to as the state space of the model, and U is a Borel space to be thought of as the set of controls. Furthermore, $\{U(x)\}_{x\in X}$ is a family of non-empty measurable subsets of U with the property that

$$\mathbb{K} := \{(x,u) : x \in X, u \in U(x)\}$$

is measurable in $X \times U$. Intuitively, $U(x)$ is the set of controls that are feasible at state x. Finally, T is a stochastic kernel on X given $\mathbb{K}$: note that $\mathbb{K}$ is a measurable subset of a Borel space $X \times U$, hence it is itself a Borel space. In order to assure that the set of control policies is not empty we require $\mathbb{K}$ to contain the graph of a measurable function [13, Assumption 2.2.2]. In other words, we assume that there exists a measurable map $\mathsf{k} : X \to U$ such that $\mathsf{k}(x) \in U(x)$ for any $x \in X$.

DEFINITION 1 (cdt-MP). *We call any tuple*

$$\mathfrak{D} = (X, U, \{U(x)\}_{x\in X}, \mathsf{T})$$

that satisfies the assumptions above a cdt-MP.

The following notation is used throughout the paper. We denote the set of positive integers by $\mathbb{N}$ and the set of non-negative integers by $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Furthermore, we denote $\overline{m,n} := \{m, m+1, \dots, n-1, n\}$ for $m, n \in \mathbb{N}_0$ such that $m < n$; $\mathbb{R}$ stands for the set of reals.

For any set X we denote by $X^{\mathbb{N}_0}$ the Cartesian product of a countable number of copies of X, i.e. $X^{\mathbb{N}_0} = \prod_{k=0}^{\infty} X$.

2.2 Model semantics

The semantics of a cdt-MP is characterized by its *paths* or executions, which reflect both the history of previous states of the system and of implemented control actions. Paths (often thought as infinite paths) are used to measure the performance of the system, which is done here via model checking methods. Also, a path up to a time epoch n can be used to derive the control action on the next step: these are finite paths that we also call *histories* as in [13, Section 2.2]. Let us finally mention that for technical reasons, when dealing with uncountable state spaces, it is usual to take into consideration also non-admissible paths, i.e. those containing non-feasible controls.

DEFINITION 2 (HISTORY). *Given a* cdt-MP $\mathfrak{D}$ *and a number* $n \in N_0$*, an* n-history *is a finite sequence*

$$h_n = (x_0, u_0, \dots, x_{n-1}, u_{n-1}, x_n), \tag{1}$$

where $x_i \in X$ *are state coordinates and* $u_i \in U$ *are control coordinates of the history. An* n*-history* h_n *is called* admissible *if* $u_i \in U(x_i)$*,* $i \in \overline{0, n-1}$*. The space of all* n*-histories is denoted by* $\bar{H}_n$*, and its subspace of admissible* n*-histories is denoted by* H_n*:*

$$\bar{H}_n = (X \times U)^n \times X, \quad H_n = \mathbb{K}^n \times X.$$

Further, we denote projections by $h_n[i] := x_i$ *and* $h_n(i) := u_i$*.*

DEFINITION 3 (PATH). *An infinite* path *of a* cdt-MP $\mathfrak{D}$ *is*

$$\omega = (x_0, u_0, x_1, u_1, \dots), \tag{2}$$

where $x_i \in X$ *and* $u_i \in U$ *for all* $i \in \mathbb{N}_0$*. As above, let us introduce projections* $\omega[i] := x_i$ *and* $\omega(i) := u_i$*.*

The space of all infinite paths $\Omega = (X \times U)^{\mathbb{N}_0}$ *together with its product* σ*-algebra* $\mathscr{F}$ *is called a* canonical sample space *for*

[1] A Polish space is a topological space which is separable and completely metrizable.

a cdt-MP $\mathfrak{D}$ *[13, Section 2.2]. An infinite path* $\omega \in \Omega$ *is called* admissible *if* $u_n \in U(x_n)$ *for any* $n \in \mathbb{N}_0$. *The space of all infinite admissible paths is denoted by* $H_\infty = \mathbb{K}^\infty$ *and is a subspace of* Ω.

Given a path ω or a history h_n, we assume below that x_i and u_i are their state and control coordinates respectively, unless otherwise stated. In order to emphasize the control structure, for the path ω as in (2) we also write

$$\omega = x_0 \xrightarrow{u_0} x_1 \xrightarrow{u_1} x_2 \xrightarrow{u_2} \cdots .$$

Clearly, for any infinite path $\omega \in \Omega$ its n-prefix (ending in a state) ω_n is an n-history. However, for notational reasons we aim on denoting finite histories through h_n and paths through ω exclusively, since they usually serve for different purposes. We are now ready to introduce the notion of control policy.

DEFINITION 4 (POLICY). *A* policy *is a sequence* $\pi = (\pi_n)_{n\in\mathbb{N}_0}$ *of universally measurable stochastic kernels* π_n *[7, Chapter 8.1], each defined on the control set* U *given* H_n *and such that*

$$\pi_n(U(x_n)|h_n) = 1 \tag{3}$$

for all $h_n \in H_n$, $n \in \mathbb{N}_0$. *The set of all policies is denoted by* Π.

Equation (2) shows that all policies are "admissible" by definition. More precisely, the class of non-admissible policies is of 0 probability: given a policy $\pi \in \Pi$ and an admissible n-history $h_n \in H_n$, the distribution of the next control action u_n given by $\pi(\cdot|h_n)$ is supported on $U(x_n)$. Among the class of all possible policies, we are especially interested in those with a simple structure in that they depend only on the current state, rather than on the whole history.

DEFINITION 5 (MARKOV POLICY). *A policy* $\pi \in \Pi$ *is called a* Markov *policy if for any* $n \in \mathbb{N}_0$ *it holds that* $\pi_n(\cdot|h_n) = \pi_n(\cdot|x_n)$, *i.e.* π_n *depends on the history* h_n *only through the current state* x_n. *The class of all Markov policies is denoted by* $\Pi_M \subset \Pi$.

Although a Markov policy is not history-dependent, it is not necessary *stationary*, i.e. it may depend on the time variable. This fact highlights the difference between Markov policies and memoryless schedulers as defined in [6, Section 10.6]. More precisely, a policy $\pi \in \Pi$ is a memoryless scheduler if and only if it is Markov and stationary, i.e. $\pi_n = \pi_0$ for all $n \in \mathbb{N}_0$.

A central measurability result is that given a cdt-MP $\mathfrak{D}$ and a policy $\pi \in \Pi$, it is always possible to construct a suitable probability measure over the set of paths. Due to technical reasons, such a measure is constructed on the space $(\Omega, \mathscr{F})$ which also contains non-admissible paths, but it is supported on the space H_∞ of admissible paths. More precisely, let a cdt-MP $\mathfrak{D}$, a policy π and a probability measure α on X be given – the latter is referred to be the initial probability distribution of the cdt-MP. By the theorem of Ionescu Tulcea [13], there exists a unique probability measure P^π_α on the canonical sample space $(\Omega, \mathscr{F})$ supported on H_∞, i.e. $\mathsf{P}^\pi_\alpha(H_\infty) = 1$ and such that

$$\begin{aligned}
\mathsf{P}^\pi_\alpha(x_0 \in B) &= \alpha(B),\\
\mathsf{P}^\pi_\alpha(u_n \in C|h_n) &= \pi_n(C|h_n),\\
\mathsf{P}^\pi_\alpha(x_{n+1} \in B|h_n, u_n) &= \mathsf{T}(B|x_n, u_n)
\end{aligned}$$

for all $B \in \mathscr{B}(X)$, $C \in \mathscr{B}(U)$ and $h_n \in H_n$, $n \in \mathbb{N}_0$. In the case when the initial distribution is supported on a single point, i.e. $\alpha(\{x\}) = 1$, we write P^π_x in place of P^π_α. Note that the last equation means that T is a transition kernel for the cdt-MP: whenever the current state is known and the control action is chosen, T gives a distribution for the next state.

2.3 Controlled Stochastic Hybrid Systems

Controlled discrete-time Stochastic Hybrid Systems (cdt-SHS) are a particular subclass of cdt-MP with a more explicit structure that distinguishes between continuous and discrete states of the system. The class of cdt-SHS provides rich modeling power and is applicable to various areas [9, 5]. It has been introduced in [2] and further studied in [5, 19]. Here we introduce it directly through the discussed framework: a cdt-SHS is a cdt-MP $\mathfrak{D} = (X, U, \{U(x)\}_{x\in X}, \mathsf{T})$ whose state space is $X = \bigcup_{q\in Q}\{q\} \times D_q$, where Q is a finite set of discrete *modes* and the measurable sets $D_q \in \mathscr{B}(\mathbb{R}^{n(q)})$ are the continuous components of the state space.

The control space U is often taken to be some Borel space. Furthermore, the transition kernel of cdt-SHS is often defined through its hybrid components [5] as:

$$\mathsf{T}(\{q'\}\times \mathrm{d}c'|(q,c),u) = \begin{cases} \mathsf{T}_q(q'|(q,c),u)\mathsf{T}_x(\mathrm{d}c'|(q,c),u), & q' = q,\\ \mathsf{T}_q(q'|(q,c),u)\mathsf{T}_r(\mathrm{d}c'|(q,c,q'),u), & q' \neq q, \end{cases}$$

for any $c \in D_q$ and $q \in Q$ and where T_q is a discrete probability law, whereas T_r, T_x are continuous (reset and transition) kernels. The semantical meaning of the conditional distributions T_q,T_r and T_x is given in [5]. The hybrid structure of the kernel further allows considering the control space U to be a product of the controls affecting the continuous dynamics and of those affecting the discrete dynamics as $U = U_q \times U_c$, where U_q, U_c are some Borel spaces [5].

Markov Decision Processes (MDP) are a subclass of cdt-SHS characterized by finite state and control spaces. Formally, U is finite and each continuous component D_q of a MDP is a singleton, which can be identified with the discrete component q itself. The theory of MDP is mature and allows for explicit solutions of many synthesis problems [6, Section 10.6].

2.4 Deterministic Finite State Automata

We are interested in linear temporal properties of trajectories of a given cdt-MP. For this purpose, in this section we introduce a model known as Deterministic Finite-state Automaton (DFA).

DEFINITION 6 (DFA). *A DFA is a tuple* $\mathscr{A} = (Q, q_0, \Sigma, F, \mathrm{t})$, *where* Q *is a finite set of locations,* $q_0 \in Q$ *is the initial location,* Σ *is a finite set,* $F \subseteq Q$ *is a set of accept locations, and* $\mathrm{t} : Q \times \Sigma \to Q$ *is a transition function.*

We call the set Σ an *alphabet* and its elements $\sigma \in \Sigma$ *letters*. We denote by $\mathfrak{S} = \Sigma^{\mathbb{N}_0}$ (by $\mathfrak{S}_{<\infty}$) the collection of all infinite (finite) words over Σ. A finite word $w = (w[0], \ldots, w[n]) \in \mathfrak{S}_{<\infty}$ is accepted by a DFA $\mathscr{A}$ if there exists a finite *run* $z = (z[0], \ldots, z[n+1]) \in Q^{n+2}$ such that $z[0] = q_0$, $z[i+1] = \mathrm{t}(z[i], w[i])$ for all $0 \le i \le n$ and $z[n+1] \in F$. Although the verification of DFA is classically based on finite words, it is here more convenient to work with infinite words and define the corresponding accepting languages over infinite words. We say that an infinite word $w \in \mathfrak{S}$ is accepted by a DFA $\mathscr{A}$ if there exists a finite prefix of w accepted by $\mathscr{A}$ as a finite word. This is equivalent to the following statement: an infinite word $w \in \mathfrak{S}$ is accepted by $\mathscr{A}$ if and only if there exists an infinite run $z \in Q^{\mathbb{N}_0}$ such that $z[0] = q_0$, $z[i+1] = \mathrm{t}(z[i], w[i])$ for all $i \in \mathbb{N}_0$ and there exists $j \in \mathbb{N}_0$ such that $z[j] \in F$. Note that $w[i]$ (resp. $z[i]$) denotes the i-th letter (resp. state of the automaton) on w (resp. z). The accepted language of $\mathscr{A}$, denoted $\mathfrak{L}(\mathscr{A})$, is the set of all words accepted by $\mathscr{A}$. We are further interested in other kinds of accepting conditions over the DFA $\mathscr{A}$: we say that the word w is n-accepted by the DFA $\mathscr{A}$ if there exists a

run $z \in Q^{\mathbb{N}_0}$ such that $z[0] = q_0$, $z[i+1] = \mathsf{t}(z[i], w[i])$ for all $i \in \mathbb{N}_0$ and $z[j] \in F$ for some $j \leq n$. We denote the set of all n-accepted words by $\mathfrak{L}_n(\mathscr{A})$. It is clear that $\mathfrak{L}_\infty(\mathscr{A}) = \mathfrak{L}(\mathscr{A})$ and furthermore that $\mathfrak{L}_n(\mathscr{A}) \subseteq \mathfrak{L}_{n+1}(\mathscr{A})$, where $n \in \mathbb{N}_0$ is arbitrary.

We use the DFA $\mathscr{A}$ to specify properties of the cdt-MP as follows. Let $\mathsf{L} : X \to \Sigma$ be a measurable function which we call the *labeling function* for a cdt-MP $\mathfrak{D}$. To each state $x \in X$ it assigns the letter $\mathsf{L}(x) \in \Sigma$. In the same fashion, each path $\omega = (x_0, u_0, x_1, u_1, x_2, u_2, \dots) \in \Omega$ induces the word $w \in \mathfrak{S}$ given by $w = (\mathsf{L}(x_0), \mathsf{L}(x_1), \mathsf{L}(x_2), \dots)$. We define the function L_Ω which maps paths onto words, i.e., $\mathsf{L}_\Omega(\omega) = w$, as above. Using this function we can introduce the satisfaction relation between paths of $\mathfrak{D}$ and the DFA specification as follows:

$$\omega \models \mathscr{A} \quad \Leftrightarrow \quad \mathsf{L}_\Omega(\omega) \in \mathfrak{L}(\mathscr{A}). \tag{4}$$

As a result, given a policy $\pi \in \Pi$ we can define the probability that a path of $\mathfrak{D}$ satisfies $\mathscr{A}$, i.e. $\mathsf{P}^\pi_\alpha(\omega \models \mathscr{A})$. It can be shown that under the assumption made on the measurability of $\mathsf{L} : X \to \Sigma$ it holds that $\{\omega \in \Omega : \omega \models \mathscr{A}\} \in \mathscr{F}$ and hence such probability is well-defined. For any $n \in N_0$ we also define

$$\omega \models_n \mathscr{A} \quad \Leftrightarrow \quad \mathsf{L}_\Omega(\omega) \in \mathfrak{L}_n(\mathscr{A}). \tag{5}$$

By now it should be clear why we prefer dealing with infinite words: the reason for this is that runs of the cdt-MP are always infinite, i.e. the cdt-MP "never stops". At the same time, the input to the DFA $\mathscr{A}$ is a word $\mathsf{L}_\Omega(\omega)$ which is infinite as well.

The work in [4] studied model-checking of automata specifications against autonomous (i.e. uncontrolled) discrete-time stochastic models over uncountable state spaces. It was shown that the computation of the probability of satisfying a DFA can be restated in terms of a *probabilistic reachability* problem over the product between the original model and the DFA. In this paper we extend this result to the case of cdt-MP. As a result, we need to introduce the probabilistic reachability problem in general terms, and discuss its solution.

3. PROBABILISTIC REACHABILITY

3.1 Problem formulation

In this section we consider a basic and important problem where the probability of reaching a goal set is to be optimized. Let us consider a cdt-MP $\mathfrak{D} = (X, U, \{U(x)\}, \mathsf{T})$ and some goal set $G \in \mathscr{B}(X)$. For any $n \in \mathbb{N}_0$ we define

$$\Diamond^{\leq n} G = \{\omega \in \Omega : x_k \in G \text{ for some } 0 \leq k \leq n\} \tag{6}$$

and further $\Diamond^{\leq \infty} G = \bigcup_{n=0}^\infty \Diamond^{\leq n} G$. We call the event in $\Diamond^{\leq n} G$ an *n-bounded* reachability if $n < \infty$, and and an *unbounded* reachability if $n = \infty$. Since it holds that $\Diamond^{\leq n} G = \bigcup_{k=0}^n \{x_k \in G\}$ for any $n \in \bar{\mathbb{N}}_0 = \mathbb{N}_0 \cup \{\infty\}$, we obtain that $G \in \mathscr{B}(X)$ implies that $\Diamond^{\leq n} G \in \mathscr{F}$ for all $n \in \bar{\mathbb{N}}_0$. Due to this reason, the quantities $\mathsf{P}^\pi_\alpha(\Diamond^{\leq n} G)$ are well-defined for any initial distribution α and any control policy $\pi \in \Pi$. We further refer to these quantities as *reachability probabilities*.

We assume that the target set $G \in \mathscr{B}(X)$ is given and fixed. We are interested in the problem of reachability probability optimization, being either a maximization or a minimization over all possible policies $\pi \in \Pi$. For this purpose we restrict our attention to initial distributions supported at single points and define the following *value functions*: $V^\pi_n(x) := \mathsf{P}^\pi_x(\Diamond^{\leq n} G)$ for all $n \in \bar{\mathbb{N}}_0$. For the problem of maximal reachability, the corresponding value functions are given by

$$V^*_n(x) := \sup_{\pi \in \Pi} V^\pi_n(x) \tag{7}$$

and for the problem of minimal reachability by

$$V_{*,n}(x) := \inf_{\pi \in \Pi} V^\pi_n(x). \tag{8}$$

A policy $\pi^* \in \Pi$ is called optimal for the problem (7) if it satisfies $V^{\pi^*}_n(x) = V^*_n(x)$. Similarly, a policy $\pi_* \in \Pi$ is called optimal for the problem (8) if $V^{\pi^*}_n(x) = V_{*,n}(x)$

The reachability problem introduced above relates to other important problems in the analysis of probabilistic systems. First of all, it is a dual to the probabilistic safety (or invariance) problem, which was studied for cdt-SHS in [5]. We can define such problem as follows: for some $S \in \mathscr{B}(X)$ and any $n \in \mathbb{N}_0$,

$$\square^{\leq n} S = \{\omega \in \Omega : x_k \in S \text{ for all } 0 \leq k \leq n\} \tag{9}$$

and $\square^{\leq \infty} S = \bigcap_{n=0}^\infty \square^{\leq n} S$. The duality between the safety and the reachability is given by the identity $(\square^{\leq n} S)^c = \Diamond^{\leq n} S^c$, which holds for any $n \in \bar{\mathbb{N}}_0$. As a result, we obtain that the maximal and minimal safety problems can be successfully reformulated in terms of optimal reachability value functions, i.e. if $S = G^c$

$$\begin{aligned} \sup_{\pi \in \Pi} \mathsf{P}^\pi_x(\square^{\leq n} S) &= 1 - V_{*,n}(x), \\ \inf_{\pi \in \Pi} \mathsf{P}^\pi_x(\square^{\leq n} S) &= 1 - V^*_n(x), \end{aligned} \tag{10}$$

for any $n \in \bar{\mathbb{N}}_0$. To find the maximal safety one has to look for the minimal reachability over the complement of the goal set.

Another problem related to probabilistic reachability is known as probabilistic reach-avoid [18]. Let us introduce this problem for two given sets $S, G \in \mathscr{B}(X)$ as follows. For $n \in \mathbb{N}_0$ we define

$$S \mathsf{U}^{\leq n} G = \left\{ \omega \in \Omega : \begin{array}{c} x_k \in G \text{ for some } 0 \leq k \leq n \text{ and} \\ x_j \in S \text{ for all } 0 \leq j < k \end{array} \right\}$$

and $S \mathsf{U}^{\leq \infty} G = \bigcup_{n=0}^\infty S \mathsf{U}^{\leq n} G$. The measurability of the defined events is clear. The set S is referred to as the safe set (or the set of legal states) and as above the set G is referred to as the goal set. We further introduce the corresponding reach-avoid value function for $x \in X, n \in \bar{\mathbb{N}}_0$, as

$$W^*_n(x) := \sup_{\pi \in \Pi} \mathsf{P}^\pi_x(S \mathsf{U}^{\leq n} G), \quad W_{*,n}(x) := \inf_{\pi \in \Pi} \mathsf{P}^\pi_x(S \mathsf{U}^{\leq n} G). \tag{11}$$

It is well-known that the reach-avoid problem is more general than the reachability one, since $\Diamond^{\leq n} G = X \mathsf{U}^{\leq n} G$ for any $G \in \mathscr{B}(X)$ and any $n \in \bar{\mathbb{N}}_0$. However, the converse statement also holds true, as we are going to show now: the idea is to state the reach-avoid problem over the cdt-MP $\mathfrak{D}$ as a reachability problem over a new cdt-MP $\check{\mathfrak{D}}$, where the avoid set $A := (S \cup G)^c$ is identified with an auxiliary single state with a loop.

More precisely, given a cdt-MP $\mathfrak{D}$ and two sets $S, G \in \mathscr{B}(X)$ we define $\check{\mathfrak{D}} = (\check{X}, U, \{\check{U}_x\}_{x \in \check{X}}, \check{\mathsf{T}})$ where $\check{X} = S \cup G \cup \{\psi\}$ for some auxiliary state $\psi \notin X$. We further define $\check{U}(x) = U(x)$ for $x \in S \cup G$ and $\check{U}(\psi) = \check{u}$ where $\check{u}$ is an element of U. Finally,

$$\check{\mathsf{T}}(B|x,u) = \begin{cases} \mathsf{T}(B|x,u), & \text{if } B \in \mathscr{B}(S \cup G),\ x \in S \cup G \\ \mathsf{T}(A|x,u), & \text{if } B = \{\psi\},\ x \in S \cup G \\ 1, & \text{if } B = \{\psi\},\ x = \psi. \end{cases}$$

Let us now define the corresponding optimal reachability value functions for the goal set G over the cdt-MP $\check{\mathfrak{D}}$, which we denote by $\check{V}^*_n$ for the maximal reachability and $\check{V}_{*,n}$ for the minimal one. The following result relates the reach-avoid value functions over $\mathfrak{D}$ to the reachability ones over $\check{\mathfrak{D}}$.

PROPOSITION 1. *For any $x \in A$ and any $n \in \bar{\mathbb{N}}_0$ it holds that $W_{*,n}(x) = W_n^*(x) = 0$. Furthermore, for any $x \in S \cup G$ and $n \in \mathbb{N}_0$*

$$W_n^*(x) = \check{V}_n^*(x), \quad W_{*,n}(x) = \check{V}_{*,n}(x).$$

It follows from Proposition 1 that results obtained for the optimization over the reachability probabilities can be directly applied to the case of reach-avoid. Due to this reason, we focus on the former problem and later show explicitly how to extend the obtained results from the reachability to the reach-avoid.

3.2 Formulation with an additive cost

The work in [5] considered the optimization of reachability probabilities over the class of cdt-MP where the cost functional took a multiplicative form. Focusing exclusively on Markov policies allowed obtaining DP recursions for value functions. In this work, however, we are interested in a wider class of policies, so the results in [5] are not directly applicable. In particular, an interesting question is the following: is it sufficient to consider only Markov policies in the optimization procedure? In order to answer this question, as well as to derive DP recursions, we are going to reformulate the original optimization problem via an additive cost functional, for which the theory of DP is rather rich [7, 13]. This approach is inspired by the one in [11].

Given a goal set $G \in \mathscr{B}(X)$ we consider a new cdt-MP

$$\hat{\mathfrak{D}} := (\hat{X}, U, \{\hat{U}(x,y)\}_{(x,y)\in\hat{X}}, \hat{\mathsf{T}})$$

with an augmented state space $\hat{X} = X \times Y$, where $Y = \{0,1\}$. The states are of the form (x,y) with coordinates being $x \in X$, $y \in Y$. The control space U is the same and we further define $\hat{U}(x,y) := U(x)$. The dynamics of $\hat{\mathfrak{D}}$ are given as follows:

$$\begin{cases} x_{n+1} & \sim \mathsf{T}(\cdot|x_n, u_n) \\ y_{n+1} & = 1_{G^c}(x_n) \cdot y_n, \end{cases}$$

hence the corresponding transition kernel $\hat{\mathsf{T}}$ is given by

$$\hat{\mathsf{T}}\left(B \times \{y'\}|x,y,u\right) := \begin{cases} y \cdot 1_{G^c}(x)\mathsf{T}(B|x,u), & \text{if } y' = 1, \\ \left(1 - y \cdot 1_{G^c}(x)\right)\mathsf{T}(B|x,u), & \text{if } y' = 0. \end{cases}$$

We construct a space of policies $\hat{\Pi}$ and for each $\hat{\pi} \in \hat{\Pi}$ and initial distribution $\hat{\alpha}$ on $\hat{X}$, a probability space $(\hat{\Omega}, \hat{\mathscr{F}}, \hat{\mathsf{P}}_{\hat{\alpha}}^{\hat{\pi}})$ with the expectation $\hat{\mathsf{E}}_{\hat{\alpha}}^{\hat{\pi}}$. We denote by $\hat{\Pi}_M \subset \hat{\Pi}$ the corresponding class of Markov policies for $\hat{\mathfrak{D}}$.

The additive cost structure consists of a cost $c : \hat{X} \to \{0,1\}$ given by $c(x,y) := y \cdot 1_G(x)$ and a functional

$$J_n^{\hat{\pi}}(x,y) := \hat{\mathsf{E}}_{(x,y)}^{\hat{\pi}}\left[\sum_{k=0}^{n} c(x_k, y_k)\right].$$

In order to relate it to the original formulation defined over the cdt-MP $\mathfrak{D}$, we first have to establish an explicit relationship between classes of strategies Π and $\hat{\Pi}$. Clearly, we can treat Π as a subset of $\hat{\Pi}$ as any policy $\pi \in \Pi$ for the cdt-MP $\mathfrak{D}$ serves also as a policy for the cdt-MP $\hat{\mathfrak{D}}$. We let $\iota : \Pi \to \hat{\Pi}$ be the *inclusion* map. On the other hand, we define the *projection* map $\theta : \hat{\Pi} \to \Pi$ by

$$\theta_i(\pi)(\mathrm{d}u_i|x_0, u_0, \ldots, x_i) := \hat{\pi}_i(\mathrm{d}u_i|x_0, y_0 = 1, u_0, \ldots, x_i, y_i = 1)$$

The following result relates the two optimization problems.

THEOREM 1. *For any $n \in \bar{\mathbb{N}}_0$, $\pi \in \Pi$ and $\hat{\pi} \in \hat{\Pi}$, it holds that*

$$J_n^{\hat{\pi}}(x,1) = V_n^{\theta(\hat{\pi})}(x), \quad V_n^{\pi}(x) = J_n^{\iota(\pi)}(x,1). \tag{12}$$

Theorem 1 has several important corollaries. First of all, it can be used to prove that Markov policies are sufficient for the original optimal reachability problem over a bounded time horizon, i.e. in case when $n < \infty$. At the same time, the optimal policy may depend on time and thus is not necessary stationary. Note that for a MDP, a special case of cdt-MP, this fact has been already known [6, Section 10.6]. More precisely:

COROLLARY 1. *For any $n \in \mathbb{N}_0$ and $\pi \in \Pi$ there exists $\pi' \in \Pi_M$ such that $V_n^{\pi} = V_n^{\pi'}$, and as a consequence $V_n^*(x) := \sup_{\pi'\in\Pi_M} V_n^{\pi'}(x)$ and $V_{*,n}(x) := \inf_{\pi'\in\Pi_M} V_n^{\pi'}(x)$.*

Moreover, it follows from Theorem 1 that one can do equivalently optimization of the additive cost functionals J to solve the original optimal reachability problem. Let us further define $J_n^*(x,y) := \sup_{\hat{\pi}\in\hat{\Pi}} J_n^{\hat{\pi}}(x,y)$ and $J_{*,n}(x,y) := \inf_{\hat{\pi}\in\hat{\Pi}} J_n^{\hat{\pi}}(x,y)$.

COROLLARY 2. *For any $n \in \bar{\mathbb{N}}_0$, $x \in X$, the following equalities hold true: $J_n^*(x,0) = J_{*,n}(x,0) = 0$ and*

$$V_n^*(x) = J_n^*(x,1), \quad V_{*,n}(x) = J_{*,n}(x,1). \tag{13}$$

Finally, we can exploit DP recursions for the additive cost functionals J_n^* and $J_{*,n}$ to study DP recursion for the optimal reachability value functions. Let us introduce the following operators

$$\mathfrak{I}^* f(x) := 1_G(x) + 1_{G^c}(x) \sup_{u\in U(x)} \int_X f(y)\mathsf{T}(\mathrm{d}y|x,u),$$

$$\mathfrak{I}_* f(x) := 1_G(x) + 1_{G^c}(x) \inf_{u\in U(x)} \int_X f(y)\mathsf{T}(\mathrm{d}y|x,u),$$

which act on the space of bounded universally measurable functions. These operators can be used to compute optimal value functions recursively, as the following result states.

COROLLARY 3. *For any $n \in \mathbb{N}_0$, the functions V_n^*, $V_{*,n}$ are universally measurable. Moreover, $V_0^* = V_{*,0} = 1_G$ and for any $n \in \mathbb{N}_0$*

$$V_{n+1}^* = \mathfrak{I}^* V_n^*, \quad V_{*,n+1} = \mathfrak{I}_* V_{*,n}. \tag{14}$$

3.3 Infinite time horizon

In the previous section we have successfully restated the original reachability problem as a classical additive-cost problem, which made it possible to derive several important results. In particular, Corollary 3 allows one to compute finite-horizon optimal reachability functions recursively, which can be done using numerical methods based on approximate abstractions: we present them later in Section 4. With focus on the infinite time horizon, it is expected that the solution can be obtained as a limit of solutions of finite-horizon optimization problems, which is a fixpoint of an appropriate operator: either $\mathfrak{I}^*$ for the maximal value function, or $\mathfrak{I}_*$ for the minimal one. However, it is known from the literature that this is not necessarily true [7]. In particular, in our case the fixpoint characterization $V_\infty^* = \mathfrak{I}^* V_\infty^*$ holds in general, whereas additional assumptions are needed to show that $V_{*,\infty} = \mathfrak{I}_* V_{*,\infty}$.

We start with the following result, which describes the properties of reachability probabilities with a fixed policy.

LEMMA 1. *For any $n \in \mathbb{N}_0$, $\pi \in \Pi$ and an initial distribution α,*

$$\mathsf{P}_\alpha^\pi(\Diamond^{\le n} G) \le \mathsf{P}_\alpha^\pi(\Diamond^{\le n+1} G) \le \mathsf{P}_\alpha^\pi(\Diamond^{<\infty} G). \tag{15}$$

Moreover, it holds that

$$\mathsf{P}_\alpha^\pi(\Diamond^{\infty} G) = \lim_n \mathsf{P}_\alpha^\pi(\Diamond^{\le n} G) = \sup_n \mathsf{P}_\alpha^\pi(\Diamond^{\le n} G). \tag{16}$$

The monotonicity result above is crucial for the proof of the fixpoint characterization for the maximal reachability over an infinite time horizon. When considering the limit of functions V_n^* as $n \to \infty$ the key step is to swap the order of $\lim_{n\to\infty}$ and the supremum over the control actions, which comes from $\mathfrak{I}^*$ – this can be done as the limit of an increasing sequence is a supremum itself. This leads us to the following:

LEMMA 2. *For any $x \in X$, $(V_n(x))_{n\in\mathbb{N}_0}$ is a non-decreasing sequence of real numbers and there exists a point-wise limit*

$$V^\star(x) := \lim_n V_n(x) = \sup_n V_n(x), \tag{17}$$

which is the least non-negative fixpoint of $\mathfrak{I}$, i.e. $V^\star = \mathfrak{I}V^\star$ and if there is another fixpoint $f \in \mathscr{B}(X)$ such that $f \geq 0$ then $f \geq V^\star$.

THEOREM 2. *The maximal reachability value function V_∞^* is the least non-negative fixpoint of the operator $\mathfrak{I}^*$.*

REMARK 1. *To our knowledge, this is the first result on the fixpoint characterization of the maximal reachability value function for* cdt-MP. *Although [10, Theorem (2.10)] provides a fixpoint characterization for a (more general) reach-avoid problem, one of the assumptions required there is that under any Markov policy, for any initial condition the set of legal states is left by the path of the* cdt-MP *with a probability* 1 *in some finite time. This clearly leads to the fact that $V_\infty^* \equiv 1$. Thus, with focus on the reachability problem [10, Theorem (2.10)] can be applied only for the case when the solution is known to be constant.*

Let us now consider the minimal reachability problem: in this case the $\lim_{n\to\infty}$ has to be swapped with the infimum that comes from $\mathfrak{I}_*$, which cannot been done in general. We then tailor a technique in [13, Section 4] to our case. In order to establish the main result we need the following assumption:

ASSUMPTION 1. *The kernel T is strongly continuous: $\mathsf{T}(A|\cdot)$ is a continuous function on $\mathbb{K}$ for any $A \in \mathscr{B}(X)$ [13, Appendix C].*

THEOREM 3. *Under Assumption 1, the minimal reachability function $V_{*,\infty}$ is the least non-negative fixpoint of $\mathfrak{I}_*$.*

Let us discuss some properties of the optimal infinite-horizon reachability value functions and of the fixpoint equations

$$f = \mathfrak{I}^* f, \tag{18}$$

$$f = \mathfrak{I}_* f. \tag{19}$$

First of all, note that $V^*(x) = V_*(x) = 1$ for all $x \in G$, so in case $V^* \equiv 1$ or $V_* \equiv 1$ we say that the corresponding optimization problem has a trivial solution. This case refers to the reachability of the goal state G in some finite time with probability 1 starting from any initial condition. However, by substitution we find that the function $f \equiv 1$ solves both equations (18) and (19), regardless of the shape of the transition kernel T. Due to this reason, we are able to formulate the following result.

PROPOSITION 2. *Equations (18) and (19) have unique solutions if and only if the corresponding optimization problems have trivial solutions.*

Some examples when the solutions of the optimization problems are not trivial can be constructed using appropriate notions of absorbing sets over the cdt-MP.

DEFINITION 7 (STRONGLY AND WEAKLY ABSORBING SETS). *Given the* cdt-MP *$\mathfrak{D} = (X, U, \{U(x)\}_{x\in X}, \mathsf{T})$, the set $A \in \mathscr{B}(X)$ is called* strongly absorbing *if $\mathsf{T}(A|x,u) = 1$ for all $u \in U(x)$ and $x \in A$.*

The set B is called weakly absorbing *if there exists a kernel μ on U given B such that $\mu(U(x)|x) = 1$ for all $x \in B$ and such that*

$$\int_{U(x)} \mathsf{T}(B|x,u)\mu(\mathrm{d}u|x) = 1, \quad \forall x \in B.$$

Let us briefly comment on the definition above. First of all, every strongly absorbing set A is a weakly absorbing set since the required kernel μ as per Definition 7 in such case can be chosen to be a deterministic one, obtained by the restriction of the map k (defined in Section 2.1) to the set A, i.e.

$$\mu(C|x) = 1_C(\mathsf{k}(x))$$

for any $x \in A$ and $C \in \mathscr{B}(U)$. Furthermore, in the autonomous case when the control set U is identified with a singleton, the notion of weak and strong sets coincide with that of an absorbing set [15]. Intuitively, a strongly absorbing set remains absorbing under any action whereas for a weakly absorbing set there exists a control which makes such set absorbing.

PROPOSITION 3. *If $A \subseteq G^c$ is a strongly (weakly) absorbing set, then $V^*(x) = 0$ ($V_*(x) = 0$) for all $x \in A$.*

The latter result in particular shows that the presence of absorbing subsets on the complement of the goal state violates the uniqueness of the fixpoint equations, and leads to a non-trivial solution of the problem. This fact is already known for the case of autonomous systems [20, 21]. There, it has been shown that under some structural assumptions on the model and on the goal set, the presence of absorbing subsets is not only a sufficient condition for the lack of uniqueness, but is also necessary. In particular, let us further mention that in the second part of [4, Theorem 3], where the infinite-horizon (autonomous) reachability value function is considered, the result holds only under the assumption that the solution of the fixpoint equation is unique – in which case as we have shown the solution is trivial and equal to a constant function 1.

Finally, let us apply the derived results to the probabilistic reach-avoid problem using Proposition 1. We keep G as the goal set, and define $S \in \mathscr{B}(X)$ to be the set of legal states. As above, let W_n^* and $W_{*,n}$ be the optimal reach-avoid value functions as in (11) and define the following operators

$$\mathfrak{R}^* f(x) := 1_G(x) + 1_A(x) \sup_{u\in U(x)} \int_X f(y)\mathsf{T}(\mathrm{d}y|x,u),$$

$$\mathfrak{R}_* f(x) := 1_G(x) + 1_A(x) \inf_{u\in U(x)} \int_X f(y)\mathsf{T}(\mathrm{d}y|x,u),$$

where as above $A = (S \cup G)^c$ is the set to be avoided. The next result follows immediately from those we have obtained for reachability and from Proposition 1.

THEOREM 4. *It holds that $W_{*,0} = W_0^* = 1_G$ and for any $n \in \mathbb{N}_0$*

$$W_{n+1}^* = \mathfrak{R}^* W_n^*, \quad W_{*,n} = \mathfrak{R}_* W_{*,n}.$$

In particular, Markov policies are sufficient for optimization:

$$W_n^*(x) := \sup_{\pi\in\Pi_M} \mathsf{P}_x^\pi(S\mathsf{U}^{\leq n}G), \quad W_{*,n}(x) := \inf_{\pi\in\Pi_M} \mathsf{P}_x^\pi(S\mathsf{U}^{\leq n}G).$$

Moreover, function W_∞^* is the least non-negative fixpoint of the operator $\mathfrak{R}^*$ and, under Assumption 1, function $W_{*,\infty}$ is the least non-negative fixpoint of the operator $\mathfrak{R}_*$.

Theorem 4 generalizes several results in the literature. First, it extends the work in [18] by considering all possible policies, rather than focusing on Markov policies exclusively. Second, it extends results obtained in [10] over the infinite-time horizon by considering both optimization problems, rather than the maximization one only. Furthermore, we are able to relax the assumptions in [10, Theorem (2.10)] and show that the fixpoint characterization for W^*_∞ holds in a more general case.

3.4 Automata model checking

In section 2.4 we have introduced the following problem: given a cdt-MP $\mathfrak{D}$, a policy $\pi \in \Pi$, an initial distribution α and a DFA $\mathscr{A}$, find the probability $\mathsf{P}^\pi_\alpha(\omega \models \mathscr{A})$. We are further interested in maximizing and minimizing such a probability over all possible policies $\pi \in \Pi$. For this purpose, we are going to reduce this general problem over $\mathfrak{D}$ to the reachability problem studied above over another cdt-MP $\mathfrak{D} \otimes \mathscr{A}$, which we refer to as a product of the cdt-MP $\mathfrak{D}$ and the automaton $\mathscr{A}$. This product is defined as follows:

DEFINITION 8 (PRODUCT BETWEEN cdt-MP AND DFA). *Given a cdt-MP $\mathfrak{D} = (X, U, \{U(x)\}_{x\in X}, \mathsf{T})$, a finite alphabet Σ, a labeling function $\mathsf{L} : X \to \Sigma$, and a DFA $\mathscr{A} = (Q, q_0, \Sigma, F, \mathsf{t})$, we define the product between $\mathfrak{D}$ and $\mathscr{A}$ to be another cdt-MP denoted as $\mathfrak{D} \otimes \mathscr{A} = (\bar{X}, U, \{U(x)\}_{x\in\bar{X}}, \bar{\mathsf{T}})$. Here $\bar{X} = X \times Q$ and*

$$\bar{\mathsf{T}}(A \times \{q'\}|x, q, u) = 1_{\mathsf{t}(q,\mathsf{L}(x))}(q') \cdot \mathsf{T}(A|x, u).$$

We want to show that $\mathsf{P}^\pi_x(\omega \models \mathscr{A})$ can be related to the reachability probability over the cdt-MP $\mathfrak{D} \otimes \mathscr{A}$ with a goal state $G := X \times F$, as it was shown to be the case when the state space is finite [12, Proposition 1]. In order to reformulate the automaton verification as a reachability problem, we are going to follow a procedure similar to the one in Section 3.2, where the reachability problem has been reformulated by an additive cost. We again construct a space of policies $\bar{\Pi}$ for $\mathfrak{D} \otimes \mathscr{A}$, and for each $\bar{\pi} \in \bar{\Pi}$ and initial distribution $\bar{\alpha}$ on $\bar{X}$, a probability space $(\bar{\Omega}, \bar{\mathscr{F}}, \bar{\mathsf{P}}^{\bar{\pi}}_{\bar{\alpha}})$. As in Section 3.2 we need to relate policies in Π to those in $\bar{\Pi}$: again, we can use the fact that $\Pi \subset \bar{\Pi}$ hence any policy $\pi \in \Pi$ over the original cdt-MP $\mathfrak{D}$ is also a policy over $\mathfrak{D} \otimes \mathscr{A}$. By $\iota : \Pi \to \bar{\Pi}$ we can hence denote the corresponding inclusion map. For the other direction, to any policy $\bar{\pi} \in \bar{\Pi}$ we can assign a policy $\theta(\bar{\pi}) \in \Pi$ as follows

$$\theta_i(\pi)(\mathrm{d}u_i|x_0, u_0, \ldots, x_i) = \bar{\pi}_i(\mathrm{d}u_i|x_0, z_0, u_0, \ldots, x_i, z_i),$$

where $z_0 = q_0$ is the initial state of $\mathscr{A}$ and $z_{j+1} = \mathsf{t}(z_j, \mathsf{L}(x_j))$ is defined recursively for $j \in \overline{0, i-1}$. The following technical lemma is necessary to state the main result.

LEMMA 3. *For any $n \in \bar{\mathbb{N}}_0$, $x \in X$ and $\pi \in \Pi$ it holds that*

$$\mathsf{P}^\pi_x(\omega \models_n \mathscr{A}) = \bar{\mathsf{P}}^{\iota(\pi)}_{(x,q_0)}(\Diamond^{\leq n+1} G),$$

where $\infty + 1 := \infty$, and for any $\bar{\pi} \in \bar{\Pi}$ it holds that

$$\bar{\mathsf{P}}^{\bar{\pi}}_{(x,q_0)}(\Diamond^{\leq n+1} G) = \mathsf{P}^{\theta(\bar{\pi})}_x(\omega \models_n \mathscr{A}).$$

THEOREM 5. *For any $n \in \bar{\mathbb{N}}_0$ and $x \in X$ it holds that*

$$\begin{aligned} \sup_{\pi\in\Pi} \mathsf{P}^\pi_x(\omega \models_n \mathscr{A}) &= \bar{V}^*_{n+1}(x), \\ \inf_{\pi\in\Pi} \mathsf{P}^\pi_x(\omega \models_n \mathscr{A}) &= \bar{V}_{*,n+1}(x), \end{aligned} \tag{20}$$

where $\infty + 1 := \infty$ and $\bar{V}^*_n$, $\bar{V}_{*,n}$ are the optimal reachability functions that are defined over the cdt-MP $\mathfrak{D} \otimes \mathscr{A}$.

Notice that in the above theorem we need to consider only Markov policies as was shown in the previous section. However, such policies are Markov with respect to $\mathfrak{D} \otimes \mathscr{A}$ and are not necessarily Markov with respect to $\mathfrak{D}$. In terms of $\mathfrak{D}$ this means that the policy is dependent on the history through the state of the automaton. The actual computation of the optimal reachability value functions can be implemented by discretizing the set of states as well as the set of controls, as elaborated in Section 4.

4. APPROXIMATE ABSTRACTIONS

Since in general there is no hope that the iterations in (14) yield value functions in an explicit form, we introduce an abstraction procedure that leads to numerical methods for the computation of such functions. Moreover, we provide an explicit upper bound on the error caused by the abstraction. We present the results with focus on the maximal reachability problem, with the understanding that a similar procedure applies to the minimal reachability one too.

Let us consider some cdt-MP $\mathfrak{D} = (X, U, \{U(x)\}_{x\in X}, \mathsf{T})$. Since X and U are Borel spaces they are metrizable topological spaces. Let ρ_X and ρ_U be some metrics on X and U respectively, which are consistent with the given topologies of the underlying spaces. We first introduce some technical considerations that are important for the abstraction over the control space. Let $U = \bigcup_{j=1}^M U_j$ be a measurable partition of U, and let $u_j \in U_j$ for $1 \leq j \leq M$ be arbitrary representative points. Define $\Delta := \max_{1\leq j\leq M} \mathrm{diam}_U(U_j)$ where the diameter of a subset of U for any $C \subseteq U$ is given by

$$\mathrm{diam}_U(C) = \sup_{u',u''\in C} \rho_U(u', u'').$$

LEMMA 4. *Let $g, \hat{g} : U \to \mathbb{R}$ be two functions. Define two optimization problems: $g^* := \sup_{u\in U} g(u)$ and $\hat{g}^* := \max_{1\leq j\leq M} \hat{g}(u_j)$. If g is Lipschitz continuous, i.e. if there exists $K > 0$ such that*

$$|g(u') - g(u'')| \leq K \cdot \rho_U(u', u'') \tag{21}$$

for all $u', u'' \in U$, then it holds that $|g^* - \hat{g}^*| \leq K \cdot \Delta + \|g - \hat{g}\|_U$.

Let us further proceed with the abstraction procedure and choose $G \in \mathscr{B}(X)$ to be a target set. Clearly, the solution of any reachability problem on G is trivial, so we only need to solve these problems on G^c. For this purpose we define $G^c = \bigcup_{i=1}^N G_i$ to be a measurable partition of the set G^c, and we choose representative points $x_i \in G_i$ for $1 \leq i \leq N$ in an arbitrary way. We further denote $\delta_i := \mathrm{diam}_X(G_i)$.

We abstract the original cdt-MP $\mathfrak{D}$ as an MDP denoted by $\tilde{\mathfrak{D}} = (\tilde{X}, \tilde{U}, \{\tilde{U}(x)\}_{x\in\tilde{X}}, \tilde{\mathsf{T}})$. The finite state and control spaces are $\tilde{X} := \{i\}_{i=1}^N \cup \{\phi\}$ and $\tilde{U} := \{j\}_{j=1}^M$, where $\phi \notin X$ is a "sink" state that corresponds to the target set G of the original cdt-MP. To complete the definition of $\tilde{\mathfrak{D}}$ we have to specify the map $\tilde{U}$ and the kernel $\tilde{\mathsf{T}}$. We choose $\tilde{U}(i) = \tilde{U}$ for any $i \in \tilde{X}$ and

$$\begin{cases} \tilde{\mathsf{T}}(k|i, j) := \mathsf{T}(G_k|x_i, u_j) & \text{for all } 1 \leq j \leq M,\ 1 \leq i, k \leq N \\ \tilde{\mathsf{T}}(\phi|i, j) := \mathsf{T}(G|x_i, u_j) & \text{for all } 1 \leq j \leq M,\ 1 \leq i \leq N \\ \tilde{\mathsf{T}}(\phi|\phi, j) = 1 & \text{for all } 1 \leq j \leq M. \end{cases}$$

For $\tilde{\mathfrak{D}}$, let us define $\tilde{V}^*_n$ to be the n-horizon maximal reachability value function of the target state ϕ. Clearly, $\tilde{V}^*_0 = 1_{\{\phi\}}$ and it follows from Corollary 3 that

$$\tilde{V}^*_{n+1}(i) = 1_{\{\phi\}}(i) + 1_{\{\phi\}^c}(i) \max_{1\leq j\leq M} \sum_{k\in\tilde{X}} \tilde{V}^*_n(k)\, \tilde{\mathsf{T}}(k|i, j).$$

Such value functions can be computed with numerically efficient methods [14]. Clearly, we expect that such functions serve as a good approximation for functions V_n^* defined for the original cdt-MP $\mathfrak{D}$. However, so far they are defined over a different state space, which makes it hard to compare them. Due to this reason, we choose $\hat{V}_n^* : X \to \mathbb{R}$ to be a piece-wise constant interpolation of $\tilde{V}_n^*$, defined in the following manner:

$$\hat{V}_n^*(x) = 1_G(x) + \sum_{i=1}^{N} 1_{G_i}(x)\tilde{V}_n^*(i).$$

Clearly, $\hat{V}_n^*(x) = 1$ for all $x \in G$. Moreover, for all other states x the following result holds true.

LEMMA 5. *For any $1 \le i \le N$ and any $x \in G_i$ it holds that*

$$\hat{V}_{n+1}^*(x) = \max_{1\le j\le M} \int_X \hat{V}_n^*(y)\mathsf{T}(\mathrm{d}y|x_i,u_j).$$

To ensure that the difference between V_n^* and $\hat{V}_n^*$ can be made as small as needed by tuning parameters of the partition Δ and δ_i, we need the following assumption.

ASSUMPTION 2. *The action space $U(x)$ does not depend on x, i.e. it holds that $U(x) = U$ for all $x \in X$. In addition, T is an integral kernel, i.e. there exists a σ-finite measure μ on X and a jointly measurable function $t : X \times X \times U \to R$ such that*

$$\mathsf{T}(B|x,u) = \int_B t(y|x,u)\mu(\mathrm{d}y), \tag{22}$$

for any $B \in \mathscr{B}(X), x \in X, u \in U$. Moreover, there exist:

1. *measurable functions $\lambda_i : X \to \mathbb{R}$ for $1 \le i \le N$, such that for all $x', x'' \in G_i$ and for all $y \in X, u \in U$:*

$$|t(y|x',u) - t(y|x'',u)| \le \lambda_i(y),$$

where $\Lambda_i := \int_X \lambda_i(y)\mu(\mathrm{d}y) < \infty$;

2. *measurable functions $\kappa_i : X \to \mathbb{R}$ for $1 \le i \le N$, such that for all $u', u'' \in U$ and for all $y \in X, x \in G_i$:*

$$|t(y|x,u') - t(y|x,u'')| \le \kappa_i(y),$$

where $K_i := \int_X \kappa_i(y)\mu(\mathrm{d}y) < \infty$.

The following result provides upper-bounds on the difference between the value functions of the original model and those computed via the MDP abstraction.

THEOREM 6. *Let V_n^* and $\hat{V}_n^*$ be the functions defined above. Introduce $r := \sup_{x\in G^c, u\in U} \mathsf{T}(G^c|x,u)$ and let Assumption 2 hold true. If $r < 1$, then for all $n \ge 1$ it holds that*

$$\|V_n^* - \tilde{V}_n^*\| \le \frac{1-r^n}{1-r} \cdot \max_{1\le i\le N} \left(\Lambda_i\delta_i + K_i\Delta\right). \tag{23}$$

If $r = 1$, then for all $n \ge 0$ it holds that

$$\|V_n^* - \tilde{V}_n^*\| \le n \cdot \max_{1\le i\le N} \left(\Lambda_i\delta_i + K_i\Delta\right). \tag{24}$$

Since the case of autonomous SHS can be regarded as a cdt-SHS with a control space being a singleton, $U = \{u\}$, it is worth commenting on the relation between Theorem 6 and approximation techniques known from the literature on autonomous SHS [3, 17]. First of all, in such a case Assumption 2 is a slight generalization of assumptions on the uniform Lipschitz continuity of kernels in [3] and on the local Lipschitz continuity in [17]. In particular, the application of Theorem 6 under Assumption 2 to the autonomous case implies [17, Theorem 6] as a special case, where the functions λ_i are assumed to be piecewise constant. In turn, [17, Theorem 6] is further known to be a generalization of [3, Theorem 1].

Let us mention that the structure of the bounds allows applying the adaptive gridding procedure developed for the autonomous SHS in [17] over the state space discretization, by computing the local errors Λ_i and choosing the discretization size δ_i accordingly. However, the error introduced by the partition of the control space depends on the global discretization parameter $\max_{1\le i\le N} K_i\Delta$, so the adaptive gridding of the control space may not improve results versus those obtained by the uniform gridding of the control space.

5. CASE STUDY: ENERGY CONTROL

Inspired by [1], we consider a resource allocation problem for an energy network comprised of two subnetworks $i = 1, 2$. The energy provider for each subnetwork is given the choice of generating energy at capacity either by a polluting device (say, a coal plant), or alternatively via local renewables: accordingly, the (normalized) decision variables are $0 \le u^i_{(\cdot)} \le 1, i = 1, 2$, denoting the production of polluting power (u^i_p) and of renewable power (u^i_r) within the i^{th} subnetwork. There is a constraint on the maximal power generation, both globally (over the total generated polluting power) and locally (whatever is not generated via coal can be obtained from the renewables u^i_r), so that

$$u^1_p + u^2_p \le 2T, \qquad u^i_p + u^i_r = T,\ i = 1, 2. \tag{25}$$

Each of the two subnetworks sets forth a time-varying demand $D_i(k)$, which is not exactly known due to the intrinsic variability of power demand.

The state variables of the model are the time-varying energy levels of each subnetwork, which we denote by E_i for $i = 1, 2$. They are driven by the following dynamics:

$$E_i(k+1) = E_i(k) + P(k)u^i_p(k) + R_i(k)u^i_r(k) - D_i(k), \tag{26}$$

where P is the actual power generated by the coal plant (polluting device) and R_i is the generated power by local renewables. Furthermore, P and R_i are independent random variables, identically distributed in time, and such that $\mathsf{E}[P] \doteq \mu_P > \mathsf{E}[R_i] \doteq \mu_{R_i}$, whereas $\mathsf{Var}[P] \doteq \sigma_P^2 < \mathsf{Var}[R_i] \doteq \sigma_{R_i}^2$. The above relation among parameters is suggested by assuming that the coal plant is a stronger and more reliable (though less desirable) source of power. Variables R_1 and R_2 can be correlated due to the spatial adjacency of the two subnetworks and its effect on the production based on renewables. This, along with the presence of P and of the constraints in (25), couples the two dynamics. We select the demand variables D_i to be independent and identically distributed, and independent from P and R_i, $i = 1, 2$.

We consider two scenarios expressed via specifications, and proceed optimizing over. Let us introduce two additional constraints on the energy levels for the subnetworks, namely:

$$E_1 + E_2 \le S, \tag{27}$$

$$E_i \ge M_i, i = 1, 2, \tag{28}$$

where the threshold S denotes a (constant) limit due to storage, while the second inequality refers to (constant) minimal energy

requirements. We are looking at the following problem:

$$\sup_{\pi} \mathsf{P}_x^{\pi} \left\{ \Box^{\leq N} (27) \wedge (28) \right\}, \tag{29}$$

which corresponds to the following safety specification: the energy levels have to be above the thresholds and the sum of them must not exceed the storage capacity. Another problem which is interesting to us is whether the energy levels can simultaneously exceed some given value F without ever falling below the zero level beforehand. This problem can be expressed as:

$$E_i \geq F_i, i = 1,2, \tag{30}$$

$$E_i \geq 0, i = 1,2, \tag{31}$$

and the related synthesis problem is

$$\sup_{\pi} \mathsf{P}_x^{\pi} \left\{ (31) \mathsf{U}^{\leq N} (30) \right\}, \tag{32}$$

The specification in (29) denotes maximal probabilistic invariance (equivalently, minimal reachability), whereas (32) is a maximal reach-avoid property, which can be reformulated as a maximal probabilistic reachability problem as discussed in Sec. 3.

In the implementation we have used the following model parameters: $T = 1$, $P = 2$ (constant polluting power production), whereas $R_i \sim \mathcal{N}(0.5, 2)$ and $D_i \sim \mathcal{N}(1, 2)$. None of the variables P, R and D depend on the step size k. We have selected the parameters M_i for (29) to be smaller than F_i for (32): for the first synthesis problem we have picked $M_i = 5$ and $S = 30$, whereas for the second one we have picked $F_i = 25$. We have chosen a discretization step $\delta_s = 1$ for the state variables and $\delta_c = 0.05$ for the control variables. All the experiments have been run on a 2.83GHz 4 Core(TM)2 Quad CPU with 3.7Gb of memory. The total running time for the experiments has amounted to 14.37 min.

Fig. 1(a) displays the maximal probabilistic invariance obtained (at the initial time step) for the synthesis problem in (29). It can be seen that the probability decreases close to the boundary of the region defined by the conditions in (30) and (31). This is as expected, since there is a higher probability close to the boundary to falsify the invariance property. The optimal control action at the initial time step for the synthesis problem (29) is plotted in Fig. 1(b). Note that this action suggests using some polluting power generation (rather than relying on the more uncertain renewable power) whenever the energy levels are too low and close to the thresholds M_i. Conversely, whenever the energy levels are high enough, the policy suggests relying on renewables only.

Fig. 2(a) depicts the maximal probabilistic reach-avoid (at the initial time step) for the synthesis problem in (32). As expected, the obtained probability is higher for energy levels close to F_i. The optimal control action for the synthesis problem (32) is given in Fig. 2(b) (here plotted at the final step): since the goal is essentially to maximize the energy levels, it can be seen that the obtained policy selects the reliable coal plant for maximal energy generation (except when already in the goal sets).

The computational results have been obtained by discretizing the state space and computing the discrete value functions, and can be further improved by using an adaptive gridding procedure as in [17]. On the other hand, rather than computing value functions, the obtained MDP abstractions for safety and reach-avoid can be engaged in the verification of the properties of interest using model-checking software [14].

6. CONCLUSIONS AND NEXT STEPS

The contributions of this work are twofold. On the theoretical side, the article has zoomed in on issues related to reachability and invariance optimization over both finite and infinite horizons for non-autonomous stochastic hybrid systems, showing the sufficiency of memoryless optimal policies and tackling the problem of verification of specifications expressed as deterministic, finite-state automata. On the applicative side, a new computational scheme with explicit error quantification has been introduced and applied over a controller synthesis case study from the area of power systems. Both the theoretical and the computational outcomes nicely tailor back to known special models (autonomous, Markovian) from the literature.

The authors are interested in considering models with more complicated control structures (both discrete and continuous), and in looking at verification problems over ω-regular properties expressed as Büchi automata: the latter require non-trivial measure-theoretical results dealing with infinite-horizon problems that go beyond the scope of this work.

7. ACKNOWLEDGMENTS

The authors would like to thank Sadegh E. Z. Soudjani for the discussions on the bounds on the discretization procedure.

8. REFERENCES

[1] MoVeS website. http://www.movesproject.eu.

[2] A. Abate, S. Amin, M. Prandini, J. Lygeros, and S. Sastry. Computational approaches to reachability analysis of stochastic hybrid systems. In *Hybrid Systems: Computation and Control*, pages 4–17. Springer Verlag, 2007.

[3] A. Abate, J.-P. Katoen, J. Lygeros, and M. Prandini. Approximate model checking of stochastic hybrid systems. *European Journal of Control*, 16(6):1–18, 2010.

[4] A. Abate, J.-P. Katoen, and A. Mereacre. Quantitative automata model checking of autonomous stochastic hybrid systems. In *Proceedings of the 14th ACM international conference on Hybrid Systems: Computation and Control*, pages 83–92, Chicago, IL, April 2011.

[5] A. Abate, M. Prandini, J. Lygeros, and S. Sastry. Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems. *Automatica*, 44(11):2724–2734, 2008.

[6] C. Baier and J.-P. Katoen. *Principles of Model Checking*. MIT Press, 2008.

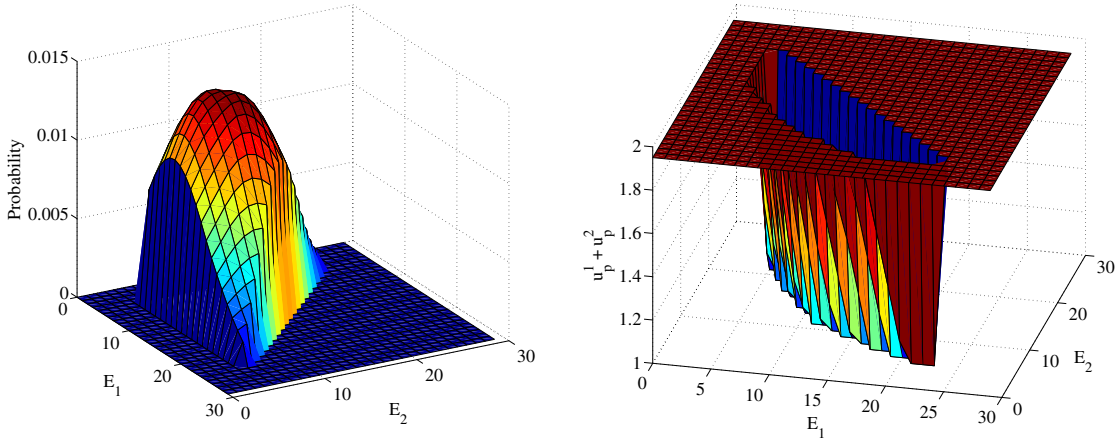
[7] D. Bertsekas and S. Shreve. *Stochastic Optimal Control: The Discrete Time Case*, volume 139. Academic Press, 1978.

[8] H. Blom and J. Lygeros (Eds.). *Stochastic Hybrid Systems: Theory and Safety Critical Applications*. Number 337 in Lecture Notes in Control and Information Sciences. Springer Verlag, Berlin Heidelberg, 2006.

[9] C. Cassandras and J. E. Lygeros. *Stochastic hybrid systems*, volume 24. CRC Press, 2007.

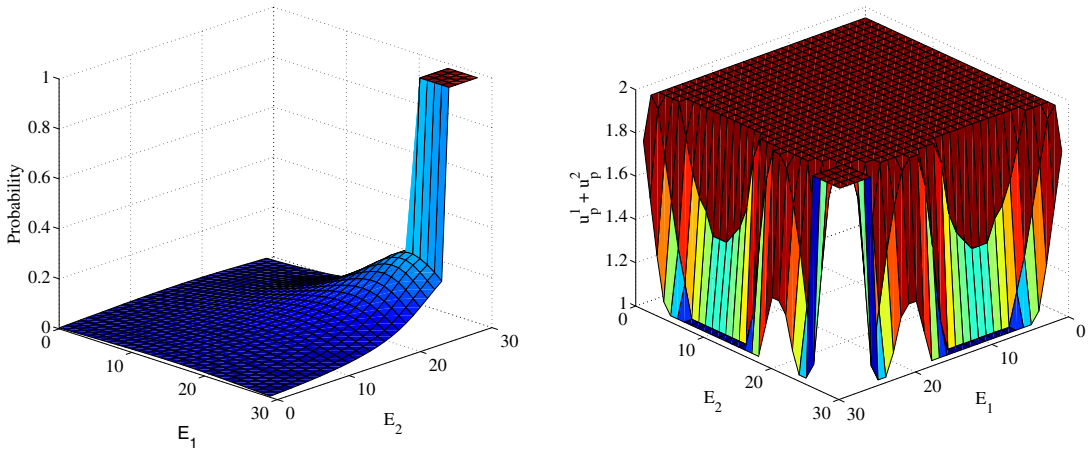
[10] D. Chatterjee, E. Cinquemani, and J. Lygeros. Maximizing the probability of attaining a target prior to extinction. *Nonlinear Analysis: Hybrid Systems*, 5(2):367–381, 2011.

[11] J. Ding, A. Abate, and C. Tomlin. Optimal control of partially observable discrete time stochastic hybrid systems for safety specifications. *Proceedings of the 32nd American Control Conference*, 2013.



(a) Maximal probabilistic invariance (b) Optimal control action at the initial step

Figure 1: Synthesis problem formulated in (29).



(a) Maximal probabilistic reach-avoid (b) Optimal control action at the final step

Figure 2: Synthesis problem formulated in (32).

[12] V. Forejt, M. Kwiatkowska, G. Norman, and D. Parker. Automated verification techniques for probabilistic systems. *Formal Methods for Eternal Networked Software Systems*, pages 53–113, 2011.

[13] O. Hernández-Lerma and J. B. Lasserre. *Discrete-time Markov control processes*, volume 30 of *Applications of Mathematics (New York)*. Springer Verlag, New York, 1996.

[14] A. Hinton, M. Kwiatkowska, G. Norman, and D. Parker. PRISM: A tool for automatic verification of probabilistic systems. In H. Hermanns and J. Palsberg, editors, *Tools and Algorithms for the Construction and Analysis of Systems*, volume 3920 of *Lecture Notes in Computer Science*, pages 441–444. Springer Verlag, 2006.

[15] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Springer Verlag, 1993.

[16] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[17] S. Soudjani and A. Abate. Adaptive gridding for abstraction and verification of stochastic hybrid systems. In *Quantitative Evaluation of Systems (QEST)*, pages 59–68, Aachen, DE, 2011.

[18] S. Summers and J. Lygeros. A Probabilistic Reach-Avoid Problem for Controlled Discrete Time Stochastic Hybrid Systems. In *IFAC Conference on Analysis and Design of Hybrid Systems, ADHS*, Zaragoza, Spain, September 2009.

[19] S. Summers and J. Lygeros. Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem. *Automatica*, 46(12):1951–1961, 2010.

[20] I. Tkachev and A. Abate. On infinite-horizon probabilistic properties and stochastic bisimulation functions. In *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, pages 526–531, Orlando, FL, December 2011.

[21] I. Tkachev and A. Abate. Regularization of Bellman Equations for Infinite-Horizon Probabilistic Properties. In *Proceedings of the 15th International Conference on Hybrid Systems: Computation and Control*, pages 227–236, Beijing, PRC, April 2012.