

Probabilistic Model Checking of Biological Systems with Uncertain Kinetic Rates

Roberto Barbuti¹, Francesca Levi¹, Paolo Milazzo¹, and Guido Scatena²

¹ Dip. di Informatica, Univ. di Pisa, Largo B. Pontecorvo 3, 56127 - Pisa, Italy
{barbuti,levifran,milazzo}@di.unipi.it

² IMT Lucca Inst. for Advance Studies, Piazza San Ponziano 6, 55100 - Lucca, Italy
g.scatena@imtlucca.it

Abstract. We present an abstraction of the probabilistic semantics of Multiset Rewriting to formally express systems of reactions with uncertain kinetic rates. This allows biological systems modelling when the exact rates are not known, but are supposed to lie in some intervals. On these (abstract) models we perform probabilistic model checking obtaining lower and upper bounds for the probabilities of reaching states satisfying given properties. These bounds are under- and over-approximations, respectively, of the probabilities one would obtain by verifying the models with exact kinetic rates belonging to the intervals.

Key words: probabilistic model checking, systems biology, uncertain kinetic rates, abstract interpretation, interval Markov chains.

1 Introduction

When modelling biological systems, the rates of the reactions involved in the evolution of the systems are often not precisely known. Thus, it is necessary to model such systems with some level of approximation. However, approximations should be significant, namely they should preserve, although not precisely, the overall behaviour of the systems.

In this paper we present a formalisation of biological systems based on *Multiset Rewriting* (MSR) [1], and we investigate the use of abstract interpretation [2] on its probabilistic semantics. In particular, we use an *Interval Markov Chain* (IMC) [3,4] to abstract the *Discrete Time Markov Chain* (DTMC) probabilistic semantics of a set of MSR models. The abstraction is able to model the semantics of a biological system for which the exact kinetic rates are not precisely known, but are supposed to lie in some intervals.

We start defining MSR as the formalism used to construct *concrete models*, namely models with exact kinetic rates (Section 2). We give a *Labelled Transition System* (LTS) semantics to MSR and show how to derive, in standard way, a probabilistic semantics from it, in terms of a DTMC. On the DTMC it is possible to perform probabilistic model checking.

In order to deal with uncertainty we define *abstract models* in which the kinetic rates are given as intervals (Section 3). We give an abstract LTS semantics

$$\begin{array}{ccccc}
\mathcal{C} \subset \mathcal{P}(\mathcal{M}) & \xrightarrow{LTS} & \mathcal{LTS} & \xrightarrow{\mathcal{H}} & \mathcal{MC} \\
\downarrow \alpha & & \downarrow \alpha_{\mathcal{LTS}} & & \downarrow \alpha_{\mathcal{MC}} \\
\mathcal{M}^\circ & \xrightarrow{LTS^\circ} & \mathcal{LTS}^\circ & \xrightarrow{\mathcal{H}^\circ} & \mathcal{MC}^\circ
\end{array}$$

Fig. 1: Schematics of the the defined theory; with \circ we indicate abstract structures, with α abstraction functions.

and a method to derive an abstract probabilistic semantics from it, in terms of IMC. On the IMC it is possible to perform probabilistic model checking (that gives *lower* and *upper bounds* for the probability of reaching states satisfying given properties) by following the approach of [5].

We relate the concrete probabilistic semantics with the abstract one by means of concepts of abstract interpretation (see Figure 1). We prove the soundness of the abstract semantics with respect to the concrete one. This implies that the lower and upper bounds obtained by model checking an abstract model are valid for all the models with exact kinetic rates belonging to the specified intervals.

In Section 4, we apply probabilistic model checking to verify reachability properties in an abstract model of tumor growth [6]. We review related work in Section 5 and we conclude with a summary and further research ideas with Section 6.

2 Probabilistic Model Checking of Biological Systems

To model biological systems we adopt *Multiset Rewriting* (MSR) where the rewriting rules are enriched with non negative real kinetic constants. Multisets are states of computation and transitions between states are performed by applying rewriting rules with a probability proportional to their kinetic constants.

Let Σ be a finite set of *species names* with cardinality n . A *multiset* is a function $s : \Sigma \rightarrow \mathbb{N}$ and $\mathcal{S}(\Sigma)$ is the *universe of multisets over Σ* . We assume multiset sum \oplus and difference \ominus , to be defined as follows: given $s', s'' \in \mathcal{S}(\Sigma)$ we have $s' \oplus s''(x) = s'(x) + s''(x)$ and $s' \ominus s''(x) = \max(s'(x) - s''(x), 0)$. In what follows we shall often assume Σ to be given.

A multiset represents the configuration of a biological system model, whereas the description of the possible events is given by rewriting rules. A *rewriting rule* is a pair (l, r) where l and r , called *reactants* and *products*, are multisets. Each rule is associated with a *kinetic constant* that is, roughly, an indication of the likelihood of the represented event.

Definition 1 (Concrete Model). A concrete model M is a triple $(\mathcal{R}, \mathcal{K}, s_0)$:

- $\mathcal{R} = \{R_1, \dots, R_m\}$, with $R_i \in \mathcal{S}(\Sigma) \times \mathcal{S}(\Sigma)$, is a set of rewriting rules;
- $\mathcal{K} = \{k_1, \dots, k_m\}$, with $k_i \in \mathbb{R}_{\geq 0}$, is a set of kinetic constants;
- $s_0 \in \mathcal{S}(\Sigma)$ is the starting state.

We denote the universe of concrete models as \mathcal{M} . When the model $M = (\mathcal{R}, \mathcal{K}, s_0)$ is clear, for $i \in [1, m]$, we use l_i and r_i to denote the multisets of rule R_i and $\mathcal{K}[i]$ for k_i . We use the notation $(l \xrightarrow{k} r)$ for $(l, r) \in \mathcal{R}$ and $k \in \mathcal{K}$. Finally we use $R(M), K(M), S_0(M)$ to denote $\mathcal{R}, \mathcal{K}, s_0$ respectively.

Two concrete models $M_i, i \in \{1, 2\}$, are *isomorphic* ($M_1 \sim M_2$) if and only if $R(M_1) = R(M_2) \wedge S_0(M_1) = S_0(M_2)$.

2.1 Labelled Transition System Semantics

To describe the semantics of a concrete model we adopt a *Labelled Transition System* (LTS) with a *transition* relation of the form $s' \xrightarrow{\eta, \beta} s''$ where η is the number of the applied rule, and $\beta \in \mathbb{R}_{\geq 0}$ is the *transition rate*.

The application of a rule R_η to a state s' is modelled by the inference rule

$$\boxed{\frac{(l_\eta \xrightarrow{k_\eta} r_\eta) \quad l \subseteq s' \quad \beta = \text{rate}(l_\eta, s', k_\eta) \quad s'' = ((s' \ominus l_\eta) \oplus r_\eta)}{s' \xrightarrow{\eta, \beta} s''}} \quad (1)$$

where $\text{rate}(l_\eta, s', k_\eta) = \text{kin}(l_\eta, s') \times k_\eta$ and $\text{kin}(l_\eta, s') = \prod_{x \in \Sigma} \binom{s'(x)}{l_\eta(x)}$.

To compute $\text{kin}(l_\eta, s')$ we take into account the number of possible distinct applications of the rule R_η to the state s' . Actually, this requires to compute the number of distinct combinations of the reactants l_η into the multiset s' . To compute $\text{rate}(l_\eta, s', k_\eta)$ we multiply the value of $\text{kin}(l_\eta, s')$ by the kinetic constant associated with R_η , namely k_η .

Given a concrete model $M = (\mathcal{R}, \mathcal{K}, s_0) \in \mathcal{M}$, we define the function $LTS : \mathcal{M} \mapsto \mathcal{LTS}$, such that $LTS(M) = (S, \rightarrow, s_0)$ is the LTS, obtained as usual by transitive closure of (1) starting from s_0 . In the following, we use \mathcal{LTS} to denote the universe of LTSs. Moreover, we use $Next(s)$ for the set of transitions from the state s ; in addition, we use $TS(s', s'') = \{s' \xrightarrow{\eta, \beta} s'' \mid \text{for some } \eta, \beta\}$ for describing the set of transitions from s' to s'' . Given a transition $t = s' \xrightarrow{\eta, \beta} s''$, we also use $\text{rate}(t) = \beta$. Note that, $\forall s \in S$, $R_\eta \in \mathcal{R}$, there is at most one transition $s \xrightarrow{\eta, \beta} s' \in Next(s)$ corresponding to R_η .

2.2 Derivation of Probabilistic Semantics

We present the probabilistic semantics of a concrete model by means of a translation from LTS into *Discrete Time Markov Chain* (DTMC).

Given a countable set S we denote with $Distr(S) = \{\rho \mid \rho : S \rightarrow [0, 1] \wedge \sum_{s \in S} \rho(s) = 1\}$ the set of *probability distributions* and with $PDistr(S) = \{\rho \mid \rho : S \rightarrow [0, 1]\}$ the set of *probability pseudo-distributions*. Given a finite set S , a function $P : S \times S \mapsto \mathbb{R}$ and $s \in S$, we denote with $P(s, S) = \sum_{s' \in S} P(s, s')$.

Definition 2 (Discrete Time Markov Chain). A DTMC is a tuple (S, P, s_0) , where: S is the set of states, $s_0 \in S$ is the starting state and $P : S \mapsto Distr(S)$ is the transition probability function.

In the following, we restrict our attention to *finitely branching* DTMC, meaning that for each $s \in S$, the set $\{s' \mid P(s)(s') > 0\}$ is finite. Since our models have m -sized set of rules from each state we have at most m exit transitions. Moreover, we use \mathcal{MC} to denote the universe of (finitely branching) DTMCs.

To derive a DTMC from the LTS, we have to calculate, for each multiset s' and s'' , the probability of moving from s' to s'' , by exploiting transition rates. Thus, we introduce two functions $R : S \times S \mapsto \mathbb{R}_{\geq 0}$ and $E : S \mapsto \mathbb{R}_{\geq 0}$, such that $R(s', s'') = \sum_{t \in TS(s', s'')} \text{rate}(t)$ and $E(s') = \sum_{s'' \in S} R(s', s'')$. Intuitively $R(s', s'')$ gives the rate of the transition from s' to s'' , while $E(s')$ computes the exit rate of the state. The probability of moving from s' to s'' is derived from $R(s', s'')$ and $E(s')$ in standard way.

Definition 3 (Probabilistic Translation Function). *We define $\mathcal{H} : \mathcal{LTS} \rightarrow \mathcal{MC}$ as $\mathcal{H}((S, \rightarrow, s_0)) = (S, P, s_0)$, where $P : S \rightarrow \text{Distr}(S)$ is the probability transition function, s.t. , $\forall s', s'' \neq s' \in S : \text{if } E(s') = 0, \text{ then } P(s')(s'') = 0, \text{ and } P(s')(s') = 1; P(s')(s'') = R(s', s'')/E(s')$ otherwise.*

2.3 Probabilistic Model Checking

In the context of probabilistic model checking [7, 8] we focus our attention on a fragment of the Probabilistic CTL (PCTL) [9] able to express reachability properties. Formally, we have to evaluate the probability of a set of paths.

Let (S, P, s_0) be a DTMC. A *path* π is a non-empty (finite or infinite) ordered succession of states s_0, s_1, \dots of S . We denote the i^{th} state of the path π by $\pi[i]$, starting from 1, and the length of π by $|\pi|$, where $|\pi| = \infty$ if π is infinite. The set of paths over S is denoted by $\text{Paths}(S)$ and its subset of finite paths is denoted as $F\text{Paths}(S)$. The *cylinder* corresponding to a path π is the set of all paths prefixed by π . Formally, for $\pi \in \text{Paths}(s)$, $C(\pi) = \{\pi\pi' \mid \pi' \in \text{Paths}(S)\}$ and $C(s)$ denotes the set of paths starting from the state s .

Definition 4 (Probability of Paths). *Let (S, P, s_0) be a DTMC. Let $\Pi = \bigcup_{\pi \in F\text{Paths}(s)} C(\pi)$ be the set of all cylinder, \mathcal{B} be the smallest σ -algebra containing Π , and $s \in S$ a state. The tuple $(\text{Paths}(S), \mathcal{B}, P_s)$ is a probability space, where P_s is the unique measure satisfying, for all path $s_0 \dots s_n$,*

$$P_s(C(s_0 \dots s_n)) = \begin{cases} 1 & \text{if } s_0 = s \wedge n = 0 \\ P(s_0, s_1) \times \dots \times P(s_{n-1}, s_n) & \text{if } s_0 = s \wedge n > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Our reachability properties are parametric w.r.t. a set AP of propositional symbols (ranged over by $\{A, B, \dots\}$). A symbol $A \in AP$ denotes a set of conditions on multisets that are evaluated by a corresponding notion of satisfaction $\models : \mathcal{S}(\Sigma) \times AP \mapsto \{\text{true}, \text{false}\}$. As usual, given $s \in \mathcal{S}(\Sigma)$ and $A \in AP$, $s \models A$ says that s satisfies A .

Definition 5 (Reachability Probability). Let $mc = (S, P, s_0)$ be a DTMC. The probability of reaching a state satisfying $A \in AP$, starting from $s \in S$, is $Reach_{A,mc}(s) = P_s(\{\pi \in C(s) \mid \pi[i] \models A \text{ for some } i \geq 0\})$.

We use $Reach(A)$ to denote $Reach_{A,mc}(s_0)$ where $mc = \mathcal{H}(LTS(M))$, for a model M clear from the context.

Example 1. We consider a chemical reactions system where molecules X and Y may bind to form complex XY and molecule X may be degraded by molecule W . With $\Sigma = \{X, Y, W, XY\}$, the system is modelled by $s_0 = \{(X, 2), (Y, 2), (W, 10)\}$, $\mathcal{R} = \{(R_1 = \{X, Y\} \xrightarrow{k_1} \{XY\}), (R_2 = \{X, W\} \xrightarrow{k_2} \{W\})\}$ and $\mathcal{K} = \{k_1 = 3, k_2 = 1\}$. Notice that we assume that the complexation is three times faster than the degradation. Figure 1 shows the derived $LTS(M)$ and $\mathcal{H}(LTS(M))$ where

$$S = \left\{ \begin{array}{ll} s_0 = \{(X, 2), (Y, 2), (W, 10), (XY, 0)\} & s_1 = \{(X, 1), (Y, 1), (W, 10), (XY, 1)\} \\ s_2 = \{(X, 1), (Y, 2), (W, 10), (XY, 0)\} & s_3 = \{(X, 0), (Y, 0), (W, 10), (XY, 2)\} \\ s_4 = \{(X, 0), (Y, 1), (W, 10), (XY, 1)\} & s_5 = \{(X, 0), (Y, 2), (W, 10), (XY, 0)\} \end{array} \right\}.$$

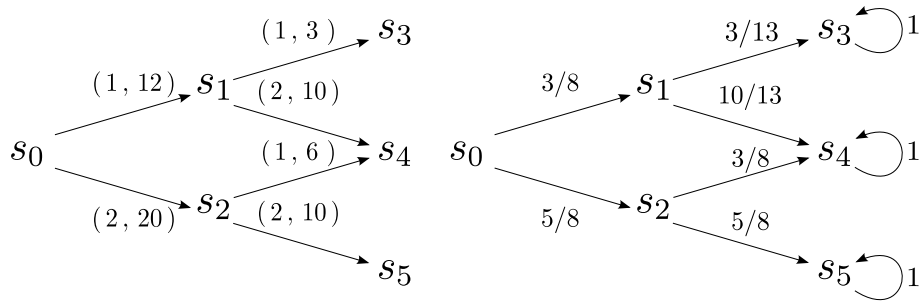


Fig. 2: $LTS(M)$, and $\mathcal{H}(LTS(M))$.

The probability of obtaining at least two complexes XY is given by the probability to reach s_3 . Therefore we obtain $3/8 \times 3/13 = 9/104$. This shows that, even if the rate of the complexation is (three times) greater than the one of the degradation, the concentration of reagent W makes the degradation more likely to happen than the binding of reagent X and Y .

3 Abstract Modelling and Model Checking

In order to approximate the information related to the kinetic rates of the reaction rules we adopt the *domain of intervals of (non negative) reals* \mathbb{I} (the real valued version of intervals of integers $[2, 10, 11]$).

Definition 6 (Intervals). $\mathbb{I} = \{[m, n] \mid m \in \mathbb{R}_{\geq 0}, n \in \mathbb{R}_{\geq 0} \cup \{\infty\} \wedge m \leq n\}$.

Over intervals of reals \mathbb{I} we use the operations and the order defined as follows.

$$\begin{aligned} \forall \mathcal{I}, \mathcal{J} \in \mathbb{I}, \mathcal{I} = [a, b], \mathcal{J} = [c, d] : \\ \mathcal{I} \times^{\mathbb{I}} \mathcal{J} = [a \times c, b \times d], & \quad \mathcal{I} \cup^{\mathbb{I}} \mathcal{J} = [\min(a, c), \max(b, d)], \\ \mathcal{I} +^{\mathbb{I}} \mathcal{J} = [a + c, b + d], & \quad \mathcal{I} \sqsubseteq_{\mathbb{I}} \mathcal{J} \text{ iff } (\mathcal{I} \cup_{\mathbb{I}} \mathcal{J} = \mathcal{J}). \end{aligned}$$

We consider both $\cup_{\mathbb{I}}$ and $\sqsubseteq_{\mathbb{I}}$ extended component-wise to m -sized vectors of intervals, and we use the same symbols. For $x \in \mathbb{R}_{\geq 0}$ we use $x^{\bullet} = [x, x] \in \mathbb{I}$ for its best abstraction as interval, and we consider \bullet extended to vector of reals.

In abstract models each *reaction rule* does not have associated a precise kinetic constant ($\in \mathbb{R}$) but instead an interval of reals ($\in \mathbb{I}$).

Definition 7 (Abstract Model). *An abstract model M is a triple $(\mathcal{R}, \mathcal{K}^{\circ}, s_0)$ with \mathcal{R} and s_0 as in the concrete case, while $\mathcal{K}^{\circ} = \{k_1^{\circ}, \dots, k_m^{\circ}\}$, $k_i^{\circ} \in \mathbb{I}$, is a set of interval values.*

We denote the universe of abstract models as \mathcal{M}° . We assume the notation used for concrete models extended in the oblivious way to concrete models. The order $\sqsubseteq_{\mathbb{I}}$ over intervals introduces a corresponding order $\sqsubseteq_{\mathcal{M}^{\circ}}$ over abstract models.

Definition 8 (Order on Abstract Models). *Given $M_i^{\circ}, i \in \{1, 2\}$: $M_1^{\circ} \sqsubseteq_{\mathcal{M}^{\circ}} M_2^{\circ}$ iff $M_1^{\circ} \sim M_2^{\circ} \wedge K(M_1^{\circ}) \sqsubseteq_{\mathbb{I}} K(M_2^{\circ})$.*

3.1 Abstraction and Concretization

To formalise the relation between concrete and abstract models we introduce the concepts of *abstraction function* and *concretization function* [2].

Let $\mathcal{C} = \{X \in \mathcal{P}(\mathcal{M}) \mid \forall M_i, M_j \in X, M_i \sim M_j\}$ the *domain of isomorphic concrete models*. Given $X \in \mathcal{C}$ we denote with $R(X)$ and $S_0(X)$ the shared rule set and the shared starting state respectively.

Definition 9 (Order on Set of Isomorphic Concrete Models).

Given two set of isomorphic concrete models $X_i \in \mathcal{C}$, $i \in \{1, 2\}$: $X_1 \sqsubseteq_{\mathcal{C}} X_2$ iff $\overline{\mathcal{K}}_1 \sqsubseteq_{\mathbb{I}} \overline{\mathcal{K}}_2$ where $\overline{\mathcal{K}}_i = \cup_{M \in X_i}^{\mathbb{I}} (K(M))^{\bullet}$.

Definition 10 (Abstraction and Concretization Functions).

We define $\alpha : \mathcal{C} \mapsto \mathcal{M}^{\circ}$ and $\gamma : \mathcal{M}^{\circ} \mapsto \mathcal{C}$ s.t. $\forall X \in \mathcal{C}, \forall M^{\circ} \in \mathcal{M}^{\circ}$:

- $\alpha(X) = (R(X), \overline{\mathcal{K}}^{\circ}, S_0(X))$ where $\overline{\mathcal{K}}^{\circ} \equiv \cup_{M \in X}^{\mathbb{I}} (K(M))^{\bullet}$;
- $\gamma(M^{\circ}) = \{M \mid \alpha(M) \sqsubseteq_{\mathcal{M}^{\circ}} M^{\circ}\}$.

Theorem 1.

The pair (α, γ) is a Galois connection between $(\mathcal{C}, \sqsubseteq_{\mathcal{C}})$ and $(\mathcal{M}^{\circ}, \sqsubseteq_{\mathcal{M}^{\circ}})$.

This formalisation shows that an abstract model M° represents a (infinite) set of concrete models with the same set of rules (same multiset of reactants and products) with kinetic rates in the specified interval.

3.2 Abstract LTS Semantics

We introduce the LTS semantics associated with abstract models, adopting an *abstract transition* relation $s' \xrightarrow{\eta, \beta^\circ} s''$, where η is as in the concrete case, while $\beta^\circ \in \mathbb{I}$. The application of a rule R_η to a state s' is modelled by the rule

$$\boxed{\frac{(l_\eta \xrightarrow{k_\eta^\circ} r_\eta) \quad l_\eta \subseteq s' \quad \beta^\circ = \text{rate}^\circ(l_\eta, s', k_\eta^\circ) \quad s'' = ((s' \ominus l_\eta) \oplus r_\eta)}{s' \xrightarrow{\eta, \beta^\circ} s''}} \quad (2)$$

where $\text{rate}^\circ(l_\eta, s', k_\eta^\circ) = \text{kin}(l_\eta, s') \times^{\mathbb{I}} k_\eta^\circ$.

We define the function $LTS^\circ : M^\circ \mapsto LTS^\circ$ such that $LTS^\circ((\mathcal{R}, \mathcal{K}^\circ, s_0)) = (S, \rightarrow_\circ, s_0)$ is obtained by transitive closure of (2) starting from s_0 . As in the concrete case the outgoing transitions from a state have distinct labels. In the following we use \mathcal{LTS}° to denote the universe of abstract LTSs and we assume the notation defined for LTSs adapted in the obvious way to the abstract case.

To relate an LTS to its abstract counterpart we introduce the concept of *best abstraction* of an LTS. The *most precise* abstract LTS can be obviously obtained by replacing the rate β of each transition with $\beta^\bullet = [\beta, \beta]$.

Definition 11 (Best Abstraction of LTS). We define $\alpha_{LTS} : \mathcal{LTS} \mapsto \mathcal{LTS}^\circ$ s.t. $\alpha_{LTS}((S, \rightarrow, s_0)) = ((S, \rightarrow_\alpha, s_0))$ with $\rightarrow_\alpha = \{s' \xrightarrow{\eta, \beta^\bullet} s'' \mid s' \xrightarrow{\eta, \beta} s'' \in \rightarrow\}$.

Notice that α_{LTS} does not effectively introduce any approximation. For expressing the correctness of an abstract LTS with respect to a concrete one, we need an approximation order $\sqsubseteq_{\mathcal{LTS}^\circ}$. In this way, we can say that $lts^\circ \in \mathcal{LTS}^\circ$ is a *sound approximation* of $lts \in \mathcal{LTS}$ provided that $\alpha_{\mathcal{LTS}^\circ}(lts) \sqsubseteq_{\mathcal{LTS}^\circ} lts^\circ$.

Definition 12 (Abstract LTS Order). Let $lts_i^\circ = (S_i, \rightarrow_\circ^i, s_{0,i}), i \in \{1, 2\}$, two abstract LTS. For $s_1 \in S_1$ and $s_2 \in S_2$, $s_1 \preceq_{\mathcal{LTS}^\circ} s_2$ (s_2 simulates s_1) iff

1. $s_1 = s_2$;
2. $\forall t_1^\circ = (s_1 \xrightarrow{\eta, \beta_1^\circ} s') \in \rightarrow_\circ^1, \exists t_2^\circ = (s_2 \xrightarrow{\eta, \beta_2^\circ} s') \in \rightarrow_\circ^2$ such that $\beta_1^\circ \sqsubseteq_{\mathbb{I}} \beta_2^\circ$.

We say that $lts_1^\circ \sqsubseteq_{\mathcal{LTS}^\circ} lts_2^\circ$ iff $s_{0,1} \preceq_{\mathcal{LTS}^\circ} s_{0,2}$.

The definition of order for abstract LTS is based on a notion of simulation between states. Intuitively, a state s' simulates another state s'' if they represents the same system configuration and if each outgoing transition from s' is matched by a transition from s'' to the same arrival configuration, with a coarser transition rate interval.

The following theorem states that the abstract LTS of an abstract model is a correct approximation of the LTS, of all the corresponding concrete models.

Theorem 2. $\forall M^\circ \in \mathcal{M}^\circ, \forall M \in \gamma(M^\circ) : \alpha_{\mathcal{LTS}}(LTS(M)) \sqsubseteq_{\mathcal{LTS}^\circ} LTS^\circ(M^\circ)$.

3.3 Abstract Probabilistic Semantics

We use the *Interval Discrete-Time Markov Chain* [3, 4] to define the probabilistic semantics of an abstract model.

Definition 13 (IMC). *An IMC is a tuple (S, P^-, P^+, s) , where: $S \subseteq \mathcal{S}(\Sigma)$ and $s \in S$ are a countable set of states and the initial state; $P^-, P^+ : S \rightarrow PDistr(S)$ are the lower and upper probability transition function s.t. $\forall s', s'' \in S, P^-(s')(s'') \leq P^+(s')(s'')$ and $P^-(s, S) \leq 1 \leq P^+(s, S)$.*

Here, $P(s')(s'')$ and $P^+(s')(s'')$ define intervals of probabilities, that represent *lower and upper bounds for the transition probabilities* of moving from s' to s'' . In the following we use \mathcal{MC}° to denote the universe of IMCs.

On a IMC, for any state s , there is a choice for an *admissible distribution* yielding the probabilities to reach successor states. A distribution is admissible for an IMC $mc^\circ = (S, P^-, P^+, s_0)$ and a state $s \in S$, iff, $\forall s' \in S : P^-(s)(s') \leq \rho(s') \leq P^+(s)(s')$. We use $ADistr_{mc^\circ}(s)$ for the admissible distributions for s and mc° .

The notion of path for IMC is analogous to that presented for DTMC and we use therefore the same notation.

Definition 14 (Scheduler). *Let $mc^\circ = (S, P^-, P^+, s_0)$ be an IMC. A scheduler is a function $\mathcal{S} : FPaths(S) \mapsto ADistr_{mc^\circ}(\pi_{last})$ for each path $\pi \in FPaths(S)$. We use $Adv(mc^\circ)$ for the set of schedulers on mc° .*

Given a scheduler $\mathcal{S} \in Adv(mc^\circ)$ the probability space over paths can be defined analogously as for DTMC (see Definition 4). Thus, $P_s^\mathcal{S}$ stands for the probability on an IMC starting from the state s w.r.t. the scheduler \mathcal{S} .

To relate a DTMC to its abstract counterpart IMC we introduce the concept of *best abstraction* of a DTMC. As for LTS, the derived probabilities are exact.

Definition 15 (Best Abstraction of DTMC).

Let $\alpha_{\mathcal{MC}} : \mathcal{MC} \mapsto \mathcal{MC}^\circ$ s.t. $\alpha_{\mathcal{MC}}((S, P, s_0)) = ((S, P_\alpha^-, P_\alpha^+, s_0))$ where, $\forall s', s'' \in S, P_\alpha^-(s', s'') = P_\alpha^+(s', s'') = P(s', s'')$.

In the style of [12, 13], we introduce an approximation order $\sqsubseteq_{\mathcal{MC}^\circ}$.

Definition 16 (Order on IMC). *Let $mc_i^\circ = (S_i, P_i^-, P_i^+, s_{0,i}), i \in \{1, 2\}$, two IMC. Given two states $s_i \in S_i, i \in \{1, 2\}$, $s_1 \preceq_{\mathcal{MC}^\circ} s_2$ (s_2 simulates s_1) iff*

(i) $s_1 = s_2$ and (ii) $ADistr_{mc_1^\circ}(s_1) \subseteq ADistr_{mc_2^\circ}(s_2)$.

We say that $mc_1^\circ \preceq_{\mathcal{MC}^\circ} mc_2^\circ$ iff $s_{0,1} \preceq_{\mathcal{MC}^\circ} s_{0,2}$.

3.4 Derivation of Abstract Markov Chain Semantics

We define the abstract probabilistic translation function $\mathcal{H}^\circ : \mathcal{LTS}^\circ \rightarrow \mathcal{MC}^\circ$. The abstract LTS reports on transitions the number of the rule which is applied and the interval representing a possible range for its rate. From this kind of information, both lower and upper bounds for the probabilities of moving from

a state to another can be calculated. Following the guidelines of the derivation of the DTMC from the concrete LTS, we introduce $R^\circ : S \times S \mapsto \mathbb{I}$, and $E^\circ : S \mapsto \mathbb{I}$ s.t. $R^\circ(s', s'') = \sum_{t \in TS(s', s'')} rate^\circ(t)$ and $E^\circ(s') = \sum_{s'' \in S} R^\circ(s', s'')$.

Intuitively, $R^\circ(s')(s'')$ reports the interval of rates corresponding to the move from s' to s'' , while $E^\circ(s')$ is the abstract exit rate. Both lower and upper bounds of the probability of moving from s' to s'' can be determined by $R^\circ(s')(s'')$ and by $E^\circ(s')$. For these purposes we need to consider the *worst case* and *best case* scenario, respectively. That is, the transition to be maximised (minimised) takes as rate value its upper (lower) bound and all the others take their lower (upper) bound. This reasoning has to be properly combined with the special cases when $\max(E^\circ(s')) = 0$ (the state s' is stable) or $\min(E^\circ(s')) = 0$ (the state s' is stable for some values of kinetic constant of some rules).

Definition 17 (Abstract Probabilistic Translation Function). *We define $\mathcal{H}^\circ : \mathcal{LTS}^\circ \rightarrow \mathcal{MC}^\circ$ such that $\mathcal{H}^\circ((S, \rightarrow^\circ, s_0)) = (S, P^-, P^+, s_0)$, where $P^-, P^+ : S \rightarrow PDistr(S)$ are computed, for each $s' \in S$, as follows:*

- if $\max(E^\circ(s')) = 0$, then $P^+(s')(s'') = P^-(s')(s'') = 0$, for each $s' \neq s''$ and $P^+(s')(s') = P^-(s')(s') = 1$;
- if $\max(E^\circ(s')) > 0$, then
 - (a) if $\min(E^\circ(s')) = 0$, then $P^+(s')(s') = 1$ and $P^-(s')(s') = 0$
 - (b) for each s'' , if $\min(R^\circ(s', s'')) = 0$, then $P^-(s')(s'') = 0$ else $P^-(s')(s'') = \min(R^\circ(s', s'')) / \max(E^\circ(s')) - \max(R^\circ(s', s'')) + \min(R^\circ(s', s''))$
 - (c) for each s'' , if $\max(R^\circ(s', s'')) = 0$, then $P^+(s')(s'') = 0$ else $P^+(s')(s'') = \max(R^\circ(s', s'')) / \min(E^\circ(s')) - \min(R^\circ(s', s'')) + \max(R^\circ(s', s''))$.

The following theorems states the soundness of \mathcal{H}° w.r.t. the approximation order $\sqsubseteq_{\mathcal{MC}^\circ}$, and that $\alpha_{\mathcal{MC}} \circ \mathcal{H} = \mathcal{H}^\circ \circ \alpha_{\mathcal{LTS}}$.

Theorem 3. *Let $lts_i^\circ = (S_i, \rightarrow_i^\circ, s_{0,i})$, $i \in \{1, 2\}$, two abstract LTS. If $lts_1^\circ \sqsubseteq_{\mathcal{LTS}^\circ} lts_2^\circ$ then $\mathcal{H}^\circ(lts_1^\circ) \sqsubseteq_{\mathcal{MC}^\circ} \mathcal{H}^\circ(lts_2^\circ)$.*

Theorem 4. *Let $M \in \mathcal{M}$, $\alpha_{\mathcal{MC}}(\mathcal{H}(LTS(M))) = \mathcal{H}^\circ(\alpha_{\mathcal{LTS}}(LTS(M)))$.*

3.5 Probabilistic Model Checking of Interval Markov Chains

By realizing probabilistic model checking on an abstract model we compute *lower* and *upper bounds* for the concrete reachability probability of all the abstracted models. On IMCs the computation of reachability probabilities considers the minimum and maximum probabilities w.r.t. all the schedulers, giving *under* and *over* approximations (for details see [5]).

Definition 18 (Reachability Probability). *Let $mc^\circ = (S, P^-, P^+, s_0)$ be an IMC. The lower and upper bound of the probability of reaching a state satisfying a propositional symbol $A \in AP$, starting from $s \in S$, are defined as follows:*

$$Reach_{A, mc^\circ}^-(s) = \inf_{\mathfrak{s} \in Adv(mc^\circ)} P_s^{\mathfrak{s}}(\{\pi \in C(s) \mid \pi[i] \models A \text{ for some } i \geq 0\});$$

$$Reach_{A, mc^\circ}^+(s) = \sup_{\mathfrak{s} \in Adv(mc^\circ)} P_s^{\mathfrak{s}}(\{\pi \in C(s) \mid \pi[i] \models A \text{ for some } i \geq 0\}).$$

Theorem 5. Let $mc_i^\circ = (S_i, P_i^-, P_i^+, s_{0,i}), i \in \{1, 2\}$, two IMC and $s_i \in S_i$, $i \in \{1, 2\}$, two states. If $s_1 \preceq_{MC^\circ} s_2$ then $\forall A \in AP$:

$$Reach_{A,mc_2^\circ}^-(s_2) \leq Reach_{A,mc_1^\circ}^-(s_1) \leq Reach_{A,mc_1^\circ}^+(s_1) \leq Reach_{A,mc_2^\circ}^+(s_2).$$

The Theorems 3, 4 and 5 show that the IMC, derived from the abstract LTS, gives *conservative bounds* for probabilistic reachability properties.

Example 2. We consider the model of Example 1 but, in this case, we assume that we are not sure about the kinetic rate of each rule, but we only estimate the interval in which they lie in. For instance, we consider $M^\circ = (\mathcal{R}, \mathcal{K}^\circ, s_0)$ where \mathcal{R} and s_0 are the same of Example 1, while $\mathcal{K}^\circ = \{k_1^\circ = [1, 5], k_2^\circ = [1, 5]\}$.

Figure 2 shows the derived $LTS^\circ(M^\circ)$ and $\mathcal{H}^\circ(LTS^\circ(M^\circ))$, where the state space S is the same of Example 1. By computing the probability of obtaining at least two complexes XY , we obtain $[4/104, 1/2] \times^{\mathbb{I}} [1/51, 1/3] = [1/1326, 1/6]$. This result shows that, even if the rates of the reactions are not precise, we have the same behaviour of Example 1. The concentration of reagent W makes the degradation more likely to happen than the binding of reagent X and Y .

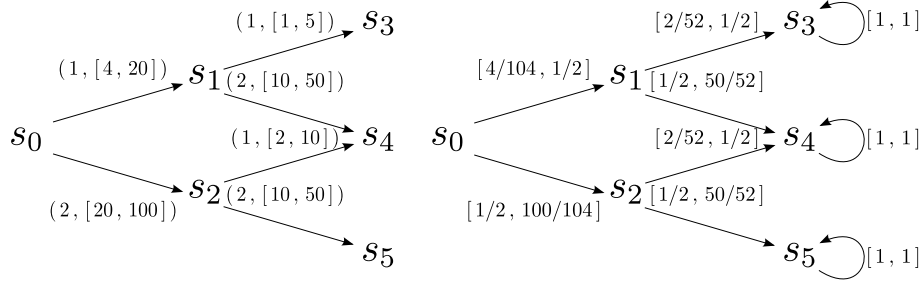


Fig. 3: $LTS^\circ(M^\circ)$ and $\mathcal{H}^\circ(LTS^\circ(M^\circ))$.

4 Case Study

We briefly present the application of the proposed approach to a model of tumor growth proposed by Villasana and Radunskaya and studied with Delay Differential Equations (DDEs) in [6].

Tumor growth is based on cell divisions (or *mitosis*). The cell cycle is the process between two mitosis, and it consists of four phases: the G_1 phase (a resting phase or gap period) called pre-synthetic phase, the S phase where the replication of DNA occurs, the G_2 gap period, called the post-synthetic phase, and the mitosis phase M in which the cells segregate the duplicated sets of chromosomes between daughter cells. The three phases G_1 , S, and G_2 constitute the pre-mitotic phase, also called *interphase*.

The simplest model proposed in [6] considers two populations of tumor cells: the population of tumor cells during cell cycle interphase, and the population of

tumor cells during mitosis. Such a model can be expressed as the following set of reactions:

$$\mathcal{R} = \{R_1 : T_I \xrightarrow{a_1} T_M, \quad R_2 : T_M \xrightarrow{a_4} 2T_I, \quad R_3 : T_I \xrightarrow{d_2}, \quad R_4 : T_M \xrightarrow{d_3}\}$$

where T_I and T_M are tumor cells in interphase and in mitosis, respectively. Reaction R_1 represents the passage of a tumor cell from the interphase to the mitosis phase, reaction R_2 represents the mitosis, and reactions R_3 and R_4 represent tumor cell death.

Let d be the rate at which mitotic cells disappear, namely $d = d_3 + a_4$. Figure 4 shows the results of the analytical study of the DDEs model, by setting the parameters a_4 and d_2 to 0.5 and 0.3, respectively, and by varying a_1 and d . There are two regions. The region in which the tumor grows is R-I, while in R-II both kinds of tumor cells disappear. A concrete probabilistic model of

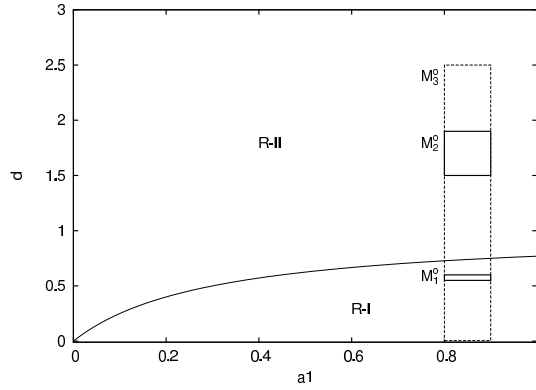


Fig. 4: The regions which describe the different behaviours of the DDEs model by varying parameters a_1 and d .

tumor growth could be trivially obtained from reactions \mathcal{R} . We have constructed three abstract models of tumor growth M_1° , M_2° and M_3° by replacing rates in the reactions with intervals. Actually, in all the three models we have replaced a_1 with $[0.8, 0.9]$, a_4 with 0.5^\bullet , d_2 with 0.3^\bullet . As regards d_3 , we have replaced it with $[0.05, 0.1]$, $[1, 1.4]$ and $[0.005, 2]$ in M_1° , M_2° and M_3° , respectively. This corresponds to consider a region in R-I, a region in R-II and a region across the line separating R-I and R-II (see Figure 4). Moreover, we have considered an initial population consisting of 10 tumor cells in interphase and 10 tumor cells in mitosis.

Formally, $M_i^\circ = (\mathcal{R}, \mathcal{K}_i^\circ, s_0)$, with $i \in \{1, 2, 3\}$, where $s_0 = \{(T_I, 10), (T_M, 10)\}$, $\mathcal{K}_i^\circ = [[0.8, 0.9]; 0, 5^\bullet; 0, 3^\bullet; d_3^i]$, where $d_3^1 = [0.05, 0.1]$, $d_3^2 = [1, 1.4]$, $d_3^3 = [0.005, 2]$.

In order to perform model checking on the abstract model we have developed a translator [14] of abstract MSR models into equivalent MDP models (by

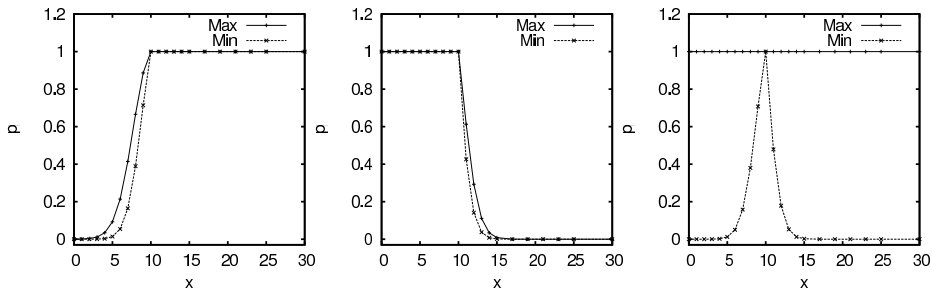


Fig. 5: Model checking results of $Reach(T_M = x)$ in, from left to right, M_1^o , M_2^o , M_3^o .

following the extreme distribution approach of [5]) that invokes PRISM [15] for the verification of the properties of interest on the MDP model. Moreover, to obtain a finite MDP, we have limited the number of states of the model to 10^4 by applying standard abstraction techniques.

In Figure 5 we show the results of model checking of property $Reach(T_M = x)$ in M_1^o , M_2^o and M_3^o by varying x . In M_1^o both the minimum and the maximum probabilities tend to zero for small values of x while they are both equal to 1 for values greater or equal to 10 (the initial value of T_M). In M_2^o it holds the opposite. In M_3^o we have that both probabilities are equal to 1 when x is 10, but they tend to the interval $[0, 1]$, namely to complete uncertainty, both for greater and smaller values of x .

The obtained results agree with the analytic ones. In fact, the results on M_1^o suggest tumor growth, those on M_2^o suggest tumor decay while those on M_3^o leave uncertainty.

Our approach is more precise with respect to analytic studies as it looks at all possible behaviours of the modeled system, rather than a single average behaviour. Moreover, a discrete probabilistic semantics is considered, instead of a continuous deterministic one.

5 Related Work

The abstraction of DTMC probabilistic semantics in terms of IMC is presented in [5, 16–19]. In the context of formal studies of biological systems, in [20, 21] abstract interpretation techniques are used to coarse-grain a system model and to perform static analyses, respectively.

Most of the above techniques differ substantially from our application. In particular, their goal is to address the state-explosion problem, e.g. to obtain a smaller abstract model by collapsing sets of concrete states into abstract states. An abstract model is thus derived from a single concrete model. Instead, we use abstraction to formalize uncertainty, representing with an abstract model an infinite set of concrete models.

Different kinds of abstraction of probabilistic semantics are proposed in [22, 23], where abstract interpretation is applied to probabilistic programs and concurrent constraint programs, respectively.

Many approaches to uncertain parameter tuning and parameters synthesis are present in the literature. The problem is examined in [24, 25] and [26] for deterministic, continuous-state and discrete-state respectively, semantics. In [27] the problem is approached by statistical model checking, while both, simulation traces based, parameter tuning and model revision are considered in [28].

The closest approach to ours is presented in [12], where a systematic technique for abstracting a set of DTMC, each representing a concrete experiment, is proposed. Their abstraction approximates the information about the multiplicity of reagents present in a solution by means of intervals of integers.

6 Conclusion

In the paper we consider models of biological systems defined by Multiset Rewriting where rewriting rules, corresponding to reactions, are enriched by real valued kinetic constraints. Our framework allows probabilistic systems with uncertain kinetics to be exhaustively model checked without any artificial assumption, obtaining conservative probabilistic bounds as result.

The computational complexity of the proposed approach is exponential in the number of uncertain parameters. The cause of this is the translation of IMC to MDP that requires for each state the computation of all the extreme distributions that grow exponentially with the number of uncertain parameters.

We plan to investigate the application of parametric DTMC [29, 30] to perform parameters tuning, and the extension of our approach to CTMC, by using the theory presented in [31] where uniform CTMC [32] is used.

References

1. Cervesato, I., Durgin, N.A., Lincoln, P., Mitchell, J.C., Scedrov, A.: A meta-notation for protocol analysis. In: CSFW, pp. 55–69 (1999)
2. Cousot, P., Cousot, R.: Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints. In: POPL, pp. 238–252 ACM Press (1977)
3. Jonsson, B., Larsen, K.G.: Specification and refinement of probabilistic processes. In: LICS, pp. 266–277. IEEE (1991)
4. Kozine, I., Utkin, L.V.: Interval-valued finite markov chains. *Reliable Computing* 8 (2), pp. 97–113 (2002)
5. Fecher, H., Leucker, M., Wolf, V.: *Don't Know* in probabilistic systems. In: SPIN. LNCS, 3925, pp. 71–88. Springer (2006)
6. Villasana, M., Radunskaya, A.: A delay differential equation model for tumor growth. *J. of Math. Biol.* 47, pp. 270–294 (2003)
7. Kwiatkowska, M.Z.: Model checking for probability and time: from theory to practice. In: LICS, pp. 351–360. IEEE (2003)
8. Kwiatkowska, M.Z., Norman, G., Parker, D.: Prism: Probabilistic symbolic model checker. In: TOOLS. LNCS, 2324, pp. 200–204. Springer (2002)
9. Hansson, H., Jonsson, B.: A logic for reasoning about time and reliability. *Formal Asp. Comput.* 6 (5), pp. 512–535 (1994)

10. Kearfott, R.B.: Interval computations: Introduction, uses, and resources. *Euromath Bulletin* 2, pp. 95–112. European Mathematical Trust (1996)
11. Weichselberger, K.: The theory of interval-probability as a unifying concept for uncertainty. In: *ISIPTA*, pp. 387–396. (1999)
12. Coletta, A., Gori, R., Levi, F.: Approximating probabilistic behaviors of biological systems using abstract interpretation. *ENTCS*, 229, pp. 165–182. Elsevier (2009)
13. D’Argenio, P.R., Jeannet, B., Jensen, H.E., Larsen, K.G.: Reachability analysis of probabilistic systems by successive refinements. In: *PAPM-PROBMIV*. LNCS, 2165, pp. 39–56. Springer (2001)
14. *AMSR2PRISM*, <http://www.di.unipi.it/~milazzo/biosims/>
15. *PRISM model checker*, <http://www.prismmodelchecker.org>
16. Huth, M.: On finite-state approximants for probabilistic computation tree logic. *Theor. Comput. Sci.* 246, pp. 113–134. Elsevier (2005)
17. Sen, K., Viswanathan, M., Agha, G.: Model-checking markov chains in the presence of uncertainties. In: *TACAS*. LNCS, 3920, pp. 394–410. Springer (2006)
18. Škulj, D.: Finite Discrete Time Markov Chains with Interval Probabilities. In: *SMPS*, pp. 299–306. Springer (2006)
19. Blanc, J.P.C., Hertog, D.D.: On Markov Chains with Uncertain Data. *CentER Discussion Paper Series* 50, Tilburg Univ., Center for Economic Research (2008)
20. Danos, V., Feret, J., Fontana, W., Krivine, J.: Abstract interpretation of cellular signalling networks. In: *VMCAI*. LNCS, 4905, pp. 83–97. Springer (2008)
21. Fages, F., Soliman, S.: Formal Cell Biology in Biocham. In: *SFM*. LNCS, 5016, pp. 54–80. Springer (2008)
22. Monniaux, D.: Abstract interpretation of programs as Markov decision processes. *Sci. Comput. Program.* 58, pp. 179–205. Springer-Verlag (2005)
23. Di Pierro, A., Wiklicky, H.: Concurrent constraint programming: towards probabilistic abstract interpretation. In: *PPDP*. pp. 127–138. ACM (2000)
24. Wilkinson, S. J., Benson, N., Kell, D. B.: Proximate parameter tuning for biochemical networks with uncertain kinetic parameters. *Mol. bioSys.* 4, pp. 74–97. RSC Publishing (2008)
25. Batt, G., Belta, C., Weiss, R.: Model Checking Genetic Regulatory Networks with Parameter Uncertainty. In: *HSCC*. LNCS, 4416, pp. 61–75. Springer (2007)
26. Manca, V.: The Metabolic Algorithm for P Systems: Principles and Applications. *Theor. Comput. Sci.* 404, pp. 142–155. Elsevier (2008)
27. Donaldson, R., Gilbert, D.: A Model Checking Approach to the Parameter Estimation of Biochemical Pathways. In: *CMSB*. LNCS, 5307, pp. 269–287. Springer (2008)
28. Fages, F., Soliman, S.: Model Revision from Temporal Logic Properties in Systems Biology. In: *PILP*. LNCS, 4911, pp. 287–304. Springer (2008)
29. Lanotte, R., Maggiolo-Schettini, A., Troina, A.: Parametric probabilistic transition systems for system design and analysis. *Form. Asp. Comput.* 19, pp. 93–109. Springer-Verlag (2007)
30. Han, T., Katoen, J.P., Mereacre, A.: Approximate Parameter Synthesis for Probabilistic Time-Bounded Reachability. In: *RTSS*, pp. 173–182. IEEE (2008)
31. Katoen, J.P., Klink, D., Leucker, M., Wolf, V.: Three-Valued Abstraction for Continuous-Time Markov Chains. In: *CAV*. LNCS, 4590, pp. 311–324. Springer (2007)
32. Baier, C., Hermanns, H., Katoen, J.P., Haverkort, B.R.: Efficient computation of time-bounded reachability probabilities in uniform continuous-time Markov decision processes. *Theor. Comput. Sci.* 345, pp. 2–26. Elsevier (2005)