Approximating probabilistic behaviours of biological systems using abstract interpretation

Alessio Coletta, Roberta Gori, Francesca Levi

Department of Computer Science, University of Pisa, Italy

Abstract

In this paper we propose to apply the *Abstract Interpretation* approach [9,10] for approximating the behaviour of biological systems, modeled specifically using the Chemical Ground Form calculus [4], a new stochastic calculus rich enough to model the dynamics of biochemical reactions.

Our analysis computes an *Interval Markov Chains* that safely approximates the *Discrete-Time Markov Chain*, describing the probabilistic behaviour of the system, and reports both *lower* and *upper* bounds for probabilistic temporal properties. Our analysis has several advantages: (i) the method is effective (even for infinite state systems) and allows us to systematically derive an IMC from an abstract labeled transition system; (ii) using intervals for abstracting the multiplicity of reagents allows us to achieve conservative bounds for the concrete probabilities of a set of concrete experiments which differs only for initial concentrations.

Keywords: Stochastic π-calculus, Abstract Interpretation, Verification of Probabilistic Temporal Properties

1 Introduction

Process calculi, originally designed for modeling distributed and mobile systems, are nowadays one of the most popular formalisms for the specification of biological systems. In this new application domain, a great effort has been devoted for adapting the traditional models to characterize the molecular and biochemical aspects of biological systems. On one hand, the proposals, such as BioAmbients [35], Beta-Binders [33], and Brane calculi [3], aim at expressing the concepts of hierarchy, compartment and membrane, which play a key role in the organization of biomolecular systems. On the other hand, there is a new interest in the design of calculi, such as stochastic π-calculus [32,34,36], able to capture the *quantitative* aspect (both time and probability) of real life applications.

The use of process calculi as a specification language offers a range of well established techniques for analysis and verification that can now be applied also to biological systems. The stochastic simulators, for example for π-calculus [36,29,30]), could be used to realize *virtual experiments* on biological systems models. *In silico* experiments could be realized in order to test possible hypotheses and to guide future *in vivo* experimentations.

An orthogonal approach is that based on the verification of temporal properties by means of model checking techniques, recently extended also to the quantitative setting. The tools of [21,23,5] support the validation of probabilistic systems, modeled as a *Discrete-Time Markov Chain* (DTMC) or as a *Markov Decision Process* (MDP), and also of stochastic systems, modeled as *Continous-Time Markov Chain* (CTMC). The study of temporal properties could give to biologists interesting information about the possible behavior of complex biological systems, such as pathways and networks of proteins.

Unfortunately, the practical application of automatic tools to biological systems revealed serious problems. One specific feature of biological processes is that they are composed by a huge number of processes with identical behavior, such as thousand of molecules of the same type. Thus, the state space of the transition system is often very large. Moreover, typically different scenarios have to be analyzed in order to infer interesting information from a biological point of view. For example, the hypotheses have to be tested several times just varying the initial concentrations.

To overcome these limitations we propose in this paper the extension of traditional *abstract model checking* [6,12] techniques to the validation of quantitative temporal properties of biological systems. As a specification language we consider a simple calculus, the *Chemical Ground Form* (CGF)[4], a subset of π-calculus without communication enriched with transition rates that determine the stochastic behavior. The calculus is rich enough for suitably modeling the dynamics of biochemical reactions. Furthermore, we consider the validation of *probabilistic temporal properties*, such as those expressible in the logic PCTL [23]. This means that we consider a probabilistic semantics of CGF formalized as a DTMC. Examples of interesting probabilistic temporal properties could be: which is the probability to reach a state where the concentration of molecule A is greater than n? which is the probability that always (in each state) the concentration of molecule A is greater than the one of molecule B?

Static analysis techniques have been established to be one of the most effective ways for computing safe approximations of the (run-time) behavior of a system, even for infinite systems. In the framework of traditional process calculi for distributed and mobile systems a variety of analyses have been proposed, aimed at proving qualitative properties, such as invariance and even of more general temporal properties [2,16,17,18,19,20,25,27,28,31].

The abstraction of probabilistic systems is much more difficult. Even for handling simple invariance properties correct approximations of the DTMC have to be computed; e.g. the *abstract model* should give conservative bounds on the actual values of the probability on the concrete model. Since the abstraction process introduces uncertainty about the behavior of the system, it is necessary that the probabilistic abstract semantics combines together probabilistic and non-deterministic steps. For these purposes traditional models, such as MDP [13,14,26], or *Interval Markov Chains* (IMC) [37,15,22], where transitions are labeled with intervals of probabilities, can be exploited. Not surprisingly, in both approaches the *lower* and *upper* bounds for the concrete probability can be achieved by considering the *worst-case* and *best-case* scenario w.r.t. all the non-deterministic choices.

The effective application of probabilistic approximations to the validation of process calculi has not yet been investigated. We propose to apply *Abstract Interpretation* [9,10], a general theory of semantics approximations, that allows us to handle infinite systems. More in details, our methodology permits to systematically derive an IMC, approximating the probabilistic behavior of a system, from an abstract *Labeled Transition System* (LTS).

As it is well known, the choice of an adequate concrete semantics is fundamental in order to derive a not too coarse abstraction. This motivates the introduction of a new LTS semantics for CGF, based on a representation of processes as *multisets* (similarly as in [4]). The main difficulty in the definition of a LTS semantics for stochastic calculi is to be able to count the number of distinct reagents combinations described by the same transition. To this aim, we adopt (well-)labeled processes so that labels identifies exactly the basic actions. Then, we design transitions recording information about the labels of the actions that participate to the move, about their rates, and about their number of occurrences.

Such a LTS semantics offers several advantages. By exploiting the labeling of processes, it is easy to extract, for each transition, both the rate and the probability. The rate of a transition can be calculated from the rate of the processes, that participate to the move, and from their multiplicities. Then, the probability of moving from one state to another can be derived in a standard way by computing the constant proportional to the rate.

Moreover, this approach supports a natural abstraction where the information about the multiplicities of reagents, present in each solution, is approximated by adopting the well-know domain of intervals of integers [8]. The abstraction of states, that can be formalized as a Galois connection [10], supports the definition of a corresponding abstract LTS semantics, where multiplicities are replaced by intervals of multiplicities both in states and in transitions. Thus, the abstract transitions record information about the labels of the actions that participate to the move, about their rates, and about their possible number of occurrences (expressed as an interval of multiplicities).

Finally, we show that our abstract LTS semantics is adequate for abstract model checking, by introducing a translation into IMC. The key step of the translation is the computation of intervals of probabilities from intervals of multiplicities. Very accurate intervals of probabilities can be achieved by suitably exploiting the information reported by transition labels.

Our approach is correct, meaning that the derived IMC gives conservative bounds for probabilistic temporal properties. In order to formalize these results we follow a traditional approach (see for example [12,13,14,37,22]) based on the definition of suitable approximation orders. For reasons of space, in this paper we focus on probabilistic reachability properties; the abstraction is however correct for full PCTL.

We present a simple example showing that our analysis is able to compute *lower* and *upper* bounds on the concrete probability of a set of concrete experiments. Thus, the approach gives conservative bounds for an experiment w.r.t. different initial concentrations.

2 Chemical Ground Form

We present the CGF calculus [4], a subset of π-calculus where basic actions are related to *rates*, e.g. values of $\mathbb{R}^+$. These values represent the parameters of the exponential distribution that characterize stochastic calculi [32,29].

The syntax of (labeled) CGF is defined in Table 1. In particular, we consider a set $\mathcal{N}$ (ranged over by $a, b, c, \ldots$) of *names* and a set $\mathcal{L}$ (ranged over by $\lambda, \mu \ldots$) of *labels*. Moreover, we consider a set $\mathcal{X}$ (ranged over by X,Y,....) of variables (representing reagents).

$E ::= 0 \mid X = S, E$	Environment
$S ::= 0 \mid \pi^{\lambda}.P + S$	Molecules
$P ::= 0 \mid X \vert P$	Solutions
$\pi ::= a_r \mid \bar{a}_r \mid \tau_r \quad r \in \mathbb{R}^+$	Basic Actions

Table 1
Syntax of CGF

A CGF is a pair (E, P) where E is a *species environment* and P is a *solution*. The environment E is a (finite) list of reagent definitions $X_i = S_i$ for distinct variables X_i and molecules S_i describing the interaction capabilities. A *molecule* S may do nothing, or may change after a delay or may interact with other reagents. The standard notation of process algebras is adopted. Thus, a delay at rate r is represented by τ_r; the input and output on a channel a at rate r are represented by a_r and $\bar{a}_r$ model, respectively (each channel always has the same rate). A solution P is a parallel composition of variables, that is a finite list of reagents. A solution P evolves according to the definitions of reagents appearing in the environment E. Intuitively, a reagent of P may change after a delay; or two reagents of P may synchronize on a channel a at rate r.

Notice that we assume labeled basic actions in order to identify exactly the actions involved in interactions. For these purposes, however, we have to consider *well-labeled* environments; an environment E is well-labeled if the labels occurring in the definitions of E are all distinct. In the following, we assume that in a CGF (E, P), E is well-labeled and each variable X occurring in E or in P has a corresponding definition in E.

Given an environment E and a label $\lambda \in \mathcal{L}$, we use the notation $E.X.\lambda$ to indicate the process $\pi^\lambda.P$ provided that $X = \ldots + \pi^\lambda.P + \ldots$ is the definition of X occurring in E. Moreover, we use $\mathcal{L}(E.X)$ to denote the set of labels appearing in the definition of X in E.

We introduce an LTS semantics for CGF based on the natural representation of solutions as multisets of reagents.

Definition 2.1 [Multiset] A *multiset* is a function $M : \mathcal{X} \to \mathbb{N}$. We use $\mathcal{M}$ for the set of multisets.

In the following, we call $M(X)$ the multiplicity of a reagent X in the multiset M. We may also use the standard representation for multisets as sets of pair (m, X) where m is the multiplicity of reagent X. Moreover, we may omit the pairs with multiplicity 0.

For multisets we use the standard operations of sum and difference $\oplus$ and $\ominus$, such that $\forall X \in \mathcal{X}$,

$M \oplus N(X) = M(X) + N(X)$

$M \ominus N(X) = M(X) \widehat{-} N(X) \quad$ where $n \widehat{-} m = n - m$ if $n - m \geq 0,\ 0$ otherwise.

For describing the behavior of a multiset we adopt a labeled transition relation of the form

$$M \xrightarrow{\Theta, \Delta, r} M'$$

where $r \in \mathbb{R}^+$ is a rate, $\Theta \in \widehat{\mathcal{L}} = \mathcal{L} \cup (\mathcal{L} \times \mathcal{L})$, $\Delta \in \widehat{Q} = \mathbb{N} \cup (\mathbb{N} \times \mathbb{N})$ such that $arity(\Theta) = arity(\Delta)$. Here, Θ reports the label (the labels) of the basic action (the basic actions) that participate to the move, Δ reports consistent information about the multiplicity (the multiplicities), and r is the related rate.

The transition relation for multisets is defined by the rules Table 2 (we are tacitly assuming to reason with respect to a given environment E). For translating solutions into multisets we exploit a function $[\![]\!] : \mathcal{P} \to \mathcal{M}$, where $\mathcal{P}$ is the set of solutions, defined in the obvious way.

There are two transition rules: one for delay actions, and one for synchronization. Rule (**Delay**) models the execution of a process ${\tau_r}^\lambda.Q$ appearing in the definition of a reagent X. The transition records the label λ together with the multiplicity of X (e.g $M(X)$) as well as the rate of delay r. Rule (**Sync**) models the synchronization between two complementary processes ${a_r}^\lambda.Q_1$ and ${\bar{a_r}}^\mu.Q_2$ appearing in the definition reagents X and Y (that may even coincide). The transition records the labels λ and μ together with the multiplicities of X and Y (e.g $M(X)$ and $M(Y)$) as well as the rate of the channel r.

It is worth noticing that we admit transitions that may report even a zero multiplicity; this choice simplifies the definition of the abstraction.

We recall that our processes are well-labeled, e.g. basic actions have distinct labels. As a consequence, the outgoing transitions for a solution M have distinct labels too. Formally,

$$\frac{E.X.\lambda = {\tau_r}^{\lambda}.Q}{M \xrightarrow{\lambda, M(X), r} (M \ominus X) \oplus [\![Q]\!]} \qquad (\textbf{Delay})$$

$$\frac{E.X.\lambda = {a_r}^{\lambda}.Q_1 \qquad E.Y.\mu = {\bar{a}_r}^{\mu}.Q_2}{M \xrightarrow{(\lambda,\mu),(M(X),M(Y)),r} (M \ominus X \ominus Y) \oplus [\![Q_1]\!] \oplus [\![Q_2]\!]} \qquad (\textbf{Sync})$$

Table 2
Transition relation

let $\mathsf{Next}(E, M)$ be the set of transitions from process M with respect to the environment E. For each $\Theta \in \widehat{\mathcal{L}}$ we may have *at most one* transition $M \xrightarrow{\Theta,\Delta,r} M' \in \mathsf{Next}(E, M)$.

Definition 2.2 [LTS] A *labeled transition system* (LTS) is a tuple $(S, \rightarrow, M_0, E)$ where: (i) $S \subseteq \mathcal{M}$ is the set of states and $M_0 \in S$ is the initial state; (ii) and, $\rightarrow \subseteq S \times \widehat{\mathcal{L}} \times \widehat{Q} \times \mathbb{R}^+ \times S$ is a set of transitions, such that, for each $M \xrightarrow{\Theta,\Delta_1,r_1} M_1$, $M \xrightarrow{\Theta,\Delta_2,r_2} M_2$, we have $\Delta_1 = \Delta_2$, $r_1 = r_2$ and $M_1 = M_2$.

Given an environment E and a multiset $M_0 \in \mathcal{M}$, we use $\mathsf{LTS}((E, M_0)) = (S, \rightarrow, M_0, E)$ for the LTS that has M_0 has initial state, obtained as usual by transitive closure. Hence, the LTS describing the evolution of a CGF (E, P) is $\mathsf{LTS}((E, [\![P]\!]))$.

In the following we use $\mathcal{LTS}$ to denote the set of LTS. Moreover, we use $\mathsf{Ts}(M, M') = \{M \xrightarrow{\Theta,\Delta,r} M'$ for some Θ, Δ and $r\}$ for describing the transitions from a M to M'.

3 Discrete-Time Markov Chains

We present the probabilistic semantics of CGF by means of a translation from LTS into DTMC. We recall the main concepts about the verification for probabilistic reachability properties over DTMC; more details on probabilistic model checking can be found in [23].

Given a finite or countable set of states $S \subseteq \mathcal{M}$ we denote with

$$\mathsf{Distr}(S) = \{\rho \mid \rho\colon S \rightarrow [0,1] \text{ and } \textstyle\sum_{M \in S} \rho(M) = 1\} \quad \mathsf{SDistr}(S) = \{\rho \mid \rho\colon S \rightarrow [0,1]\}$$

the set of (discrete) probability *distributions* and of *pseudo-distributions* on S, respectively.

Definition 3.1 [DTMC] A *Discrete-Time Markov Chain* is a triple $(S, \mathbf{P}, M_0)$ where: (i) $S \subseteq \mathcal{M}$ is a *finite or countable* set of states and $M_0 \in S$ is the initial state; (ii) and, $\mathbf{P}\colon S \rightarrow \mathsf{Distr}(S)$ is the *probability transition function.*

In DTMC state transitions are equipped with probabilities, namely $\mathbf{P}(M)(M')$ reports the probability of moving from state M to state M'. In the following, we restrict the attention to *finitely branching* DTMC, meaning that for each $M \in S$, the set $\{M' \mid \mathbf{P}(M)(M') > 0\}$ is finite. Moreover, we use $\mathcal{MC}$ to denote the set of (finitely branching) DTMC.

In order to derive a DTMC from the LTS semantics, we have to calculate, for each multiset M and M', the probability of moving from M to M' by exploiting the information reported by the labels of transitions. First, we have to extract the rate corresponding to the move from M to M' (namely the rate of the underlying CTMC). Then, we achieve the related probability by considering as usual the constant proportional to the rate of the move.

Therefore, we introduce the concept of *rate of a transition.* For a transition $t = M \xrightarrow{\Theta,\Delta,r} M' \in \mathsf{Next}(E, M)$ we have

$$\mathsf{rate}(t) = \begin{cases} n \cdot r & \Theta = \lambda \wedge \Delta = n, \\ n \cdot (\widehat{m-1}) \cdot r & \Theta = (\lambda, \mu) \wedge \Delta = (n, m) \wedge \lambda, \mu \in \mathcal{L}(E.X), \\ n \cdot m \cdot r & \Theta = (\lambda, \mu) \wedge \Delta = (n, m) \wedge \lambda \in \mathcal{L}(E.X) \wedge \mu \in \mathcal{L}(E.Y) \wedge X \neq Y. \end{cases}$$

For computing $\mathsf{rate}(t)$ we take into account the number of distinct transitions t that may occur in the multiset M. More in details, the rate r of the basic action (actions) related to Θ is multiplied by the number of distinct combinations appearing in M (by exploiting the information recorded by Δ). The resulting rate may be even zero. This is the case, for example, whenever two reagents X and Y interact and one of the two has multiplicity zero; or whenever a reagent X with multiplicity one interacts with X itself.

Then, we introduce functions $\mathbf{R} : S \times S \to \mathbb{R}^{>=0}$ and $\mathbf{E} : S \to \mathbb{R}^{>=0}$, such that

$$\mathbf{R}(M, M') = \textstyle\sum_{t \in \mathsf{Ts}(M,M')} \mathsf{rate}(t) \quad \mathbf{E}(M) = \textstyle\sum_{M' \in S} \mathbf{R}(M, M').$$

Intuitively, $\mathbf{R}(M, M')$ reports the rate corresponding to the move from M to M', while $\mathbf{E}(M)$ is the *exit rate*. As usual in stochastic calculi, the probability of moving from M to M' is computed from $\mathbf{R}(M, M')$ and from the exit rate $\mathbf{E}(M)$.

Definition 3.2 [Derivation of the DTMC] We define a *probabilistic translation* function $\mathbf{H} : \mathcal{LTS} \to \mathcal{MC}$ such that $\mathbf{H}((S, \to, M_0, E)) = (S, \mathbf{P}, M_0)$, where $\mathbf{P} : S \to \mathsf{Distr}(S)$ is the *probability transition function*, such that for each $M \in S$,

(i) if $\mathbf{E}(M) = 0$, then $\mathbf{P}(M)(M') = 0$, for each $M' \neq M$, and $\mathbf{P}(M, M) = 1$;

(ii) if $\mathbf{E}(M) > 0$, then for each M', $\mathbf{P}(M)(M') = \mathbf{R}(M, M')/\mathbf{E}(M)$.

We are interested in the probability of reaching a state satisfying a given property, starting from the initial state. Formally, we have to evaluate the probability of a set of paths.

Let $(S, \mathbf{P}, M_0)$ be a DTMC. A *path* π is a non-empty (finite of infinite) sequence of states of S. We denote the i-th state in a path π, starting from 0, by $\pi[i]$ and the length of π by $|\pi|$, where $|\pi| = \infty$ if π is infinite. The set of paths over S is denoted by $\mathsf{Paths}(S)$, while the subset of finite paths is denoted by $\mathsf{FPaths}(S)$.

The *cylinder* corresponding to π is the set of all paths prefixed by π. Formally, $C(\pi) = \{\pi\pi' \mid \pi' \in \mathsf{Paths}(S)\}$ and $C(M)$ denotes the set of paths starting from the state M.

Definition 3.3 [Probability of Paths] Let $(S, \mathbf{P}, M_0)$ be a DTMC. Let $\mathcal{C} = \bigcup_{\pi \in \mathsf{FPaths}(S)} C(\pi)$ be the set of all cylinder, $\mathcal{B}$ be the smallest σ-algebra containing $\mathcal{C}$, and $M \in \mathcal{M}$ a state. The tuple $(\mathsf{Paths}(S), \mathcal{B}, \mathbf{P}_M)$ is a probability space, where $\mathbf{P}_M$ is the unique measure satisfying, for all path $M_0 \dots M_n$,

$$\mathbf{P}_M(C(M_0 \dots M_n)) = \begin{cases} 1 & \text{if } M_0 = M \wedge n = 0 \\ \mathbf{P}(M_0, M_1) \cdot \ldots \cdot \mathbf{P}(M_{n-1}, M_n) & \text{if } M_0 = M \wedge n > 0 \\ 0 & \text{otherwise} \end{cases}$$

Our reachability properties are parametric w.r.t. a set AP of propositional symbols (ranged over by A,B) and w.r.t. a corresponding notion of satisfaction for multisets $\mathcal{M}$. As usual we use $M \models A$ to say that M satisfies A, and $M \not\models A$ to say that this is not the case.

Definition 3.4 [Reachability Properties] Let $mc = (S, \mathbf{P}, M_0)$ be a DTMC. The probability of reaching a state satisfying a propositional symbol $A \in AP$, starting from $M \in S$, is defined as $\mathsf{Reach}_{A,mc}(M) = \mathbf{P}_M(\{\pi \in C(M) \mid \pi[i] \models A \text{ for some } i \geq 0\})$.

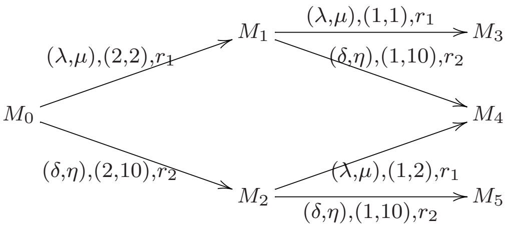

Figure 1. The LTS

Example 3.5 We consider a simple chemical reaction formed by a *complexation* and a *degradation*, i.e. a reaction where two molecules X and Y may bind together to form a multimolecular complex XY but where molecule X may be degraded by molecule W. This situation can be formalized by the following environment,

$$E ::= X = a^{\lambda}_{r_1}.0 + b^{\delta}_{r_2}.0, \quad Y = \bar{a}^{\mu}_{r_1}.XY, \quad W = \bar{b}^{\eta}_{r_2}.W, \quad XY = 0.$$

The reagent X may *either* synchronize with reagent Y along channel a at rate r_1 (and produce XY) *or* it may synchronize with reagent W along channel b at rate r_2 (and produce W).

By examining the evolution of the system for the initial solution M_0 we obtain the LTS (depicted in Fig.1) where[1],

$$M_0 = \{(2, X), (2, Y), (10, W)\} \quad M_1 = \{(1, X), (1, Y), (10, W), (1, XY)\} \quad M_2 = \{(1, X), (2, Y), (10, W)\}$$
$$M_3 = \{(2, XY), (10, W)\} \quad M_4 = \{(1, XY), (1, Y)(10, W)\} \quad M_5 = \{(2, Y), (10, W)\}$$

All the transitions are obtained by rule **Sync**; as an example we comment the case of M_0. Transition $M_0 \xrightarrow{(\lambda,\mu),(2,2),r_1} M_1$ models the *binding*, i.e. the synchronization between X and Y along channel a. The transition records the labels of the basic actions and the multiplicities of reagents X and Y, respectively, and the rate r_1. Similarly transition $M_0 \xrightarrow{(\delta,\eta),(2,10),r_2} M_2$ models the *degradation*, i.e the synchronization between X and W along channel b.

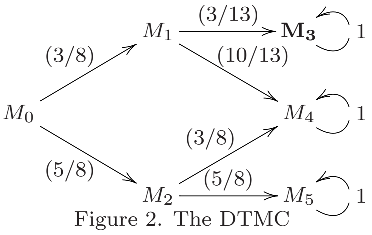

Figure 2. The DTMC

If we assume $r_1 = 3r_2$, showing that the complexion is three times faster than the degradation, we derive the DTMC, depicted in Fig.2[2]. In order to infer relevant information about our biological system, it would be convenient to compute the probability of reaching a state where a given quantity of complexes XY appear. As an example, we consider the probability to reach a state where *at least two* bindings XY are created, i.e., the probability that no degradation will take place and that all the reagents X will bind to reagents Y. Since only state M_3 (depicted in bold) satisfies the previous requirement the probability to reach state M_3 will be $(3/8) \cdot (3/13) = (9/104)$. This shows that even if the rate of the binding is (three times) greater that the one of degradation, the concentration of reagent W makes the degradation more likely to happen than the binding of reagent X and Y. By contrast,

[1] Note that we do not to indicate transitions (and consequently states) where a zero multiplicity may appear.

[2] As usual, we picture $Distr(M)$ by drawing an arrow between M and M' labeled $\rho(M')$ for all M' with $\rho(M') > 0$.

the probability to reach a state where *at least three* complexes XY are created is trivially 0. □

4 Abstraction of the LTS

We define the abstract LTS semantics for CGF. The domain of abstract LTS includes a notion of ordering, expressing precision and correctness of approximations, in the style of [12].

Abstract states. We present the abstract states and we formalize the relation with multisets as a standard Galois connection [10].

In order to approximate the information related to the multiplicities of reagents present in a solution we adopt the well-know domain of intervals of integers [8]. In particular, let

$$\mathcal{I} = \{[m, n] \mid m \in \mathbb{N}, n \in \mathbb{N} \cup \{\infty\} \wedge m \leq n\}.$$

Over intervals we consider the standard union $\cup^\circ$ and the induced order $\leq_I$, defined as follows,
(i) $I \cup^\circ J = [\mathtt{min(a,c)}, \mathtt{max(b,d)}]$ for $I = [a, b], J = [c, d]$; (ii) $I \leq_I J$ iff $I \cup^\circ J = J$.

The abstract states are defined analogously to multisets by replacing multiplicities with intervals of multiplicities.

Definition 4.1 [Abstract states] An *abstract state* is a function $M^\circ : \mathcal{X} \to \mathcal{I}$. We use $\mathcal{M}^\circ$ for the set of abstract states.

Notice that each multiset $M \in \mathcal{M}$ is represented by a corresponding abstract multiset, where each multiplicity, such as n, is replaced by the exact interval $[n, n]$. In the following, we may write M° for the abstract version of a multiset M; analogously, we may use S° for sets of multisets S.

Since an interval represents a set of multiplicities, it is immediate to define the following approximation order over abstract states.

Definition 4.2 [Order on States] Let $M_1^\circ, M_2^\circ \in \mathcal{M}^\circ$, we say that $M_1^\circ \sqsubseteq^\circ M_2^\circ$ iff, for all reagent $X \in \mathcal{X}$, $M_1^\circ(X) \leq_I M_2^\circ(X)$.

The relation between multisets and abstract states is formalized as a Galois connection. The *abstraction function* $\alpha : \mathcal{P}(\mathcal{M}) \to \mathcal{M}^\circ$ reports the *best approximation* for each set of multisets S, given by the l.u.b. (denoted by $\cup^\circ$) of the abstraction of each multiset $M \in S$. Its counterpart is the *concretization function* $\gamma : \mathcal{M}^\circ \to \mathcal{P}(\mathcal{M})$ that reports the set of multisets represented by an abstract state.

Definition 4.3 We define $\alpha : \mathcal{P}(\mathcal{M}) \to \mathcal{M}^\circ$ and $\gamma : \mathcal{M}^\circ \to \mathcal{P}(\mathcal{M})$ such that, for each $S \in \mathcal{P}(\mathcal{M})$ and $M^\circ \in \mathcal{M}^\circ$: (i) $\alpha(S) = \bigcup^\circ_{M \in S} M^\circ$; (ii) $\gamma(M^\circ) = \{M' \mid M'^\circ \sqsubseteq^\circ M^\circ\}$.

Theorem 4.4 *The pair* (α, γ) *is a* Galois connection *between* $(\mathcal{P}(\mathcal{M}), \subseteq)$ *and* $(\mathcal{M}^\circ, \sqsubseteq^\circ)$.

Abstract LTS. We introduce the definition of abstract LTS as well as the notions necessary to state the correctness and precision with respect to the concrete semantics.

The abstract transitions are defined analogously to the concrete case by replacing the information about multiplicities with intervals of multiplicities. Thus, we adopt an abstract transition relation of the form

$$M^\circ \xrightarrow[\circ]{\Theta, \Delta^\circ, r} M_1^\circ$$

where $\Theta \in \widehat{\mathcal{L}}$, $\Delta^\circ \in \widehat{Q}^\circ = \mathcal{I} \cup (\mathcal{I} \times \mathcal{I})$, with *arity*$(\Theta) =$ *arity*(Δ°).

Analogously as in the concrete case the outgoing transitions from an abstract state M° have distinct labels.

Definition 4.5 [Abstract LTS] An *abstract labeled transition system* is a tuple $(S^\circ, \rightarrow_\circ, M_0^\circ, E)$ where: (i) $S^\circ \subseteq \mathcal{M}^\circ$ is a set of abstract states and $M_0^\circ \in S^\circ$ is the initial state; (ii) and, $\rightarrow_\circ \subseteq S^\circ \times \widehat{\mathcal{L}} \times \widehat{Q}^\circ \times \mathbb{R}^+ \times S^\circ$ is a set of abstract transitions, such that for each $M^\circ \xrightarrow[\circ]{\Theta,\Delta_1^\circ,r_1} M_1^\circ$, $M^\circ \xrightarrow[\circ]{\Theta,\Delta_2^\circ,r_2} M_2^\circ$ we have $r_1 = r_2$, $\Delta_1^\circ = \Delta_2^\circ$ and $M_1^\circ = M_2^\circ$.

In the following we use $\mathcal{LTS}^\circ$ to denote the set of abstract LTS. We also assume that the notations defined for LTS are adapted in the obvious way to the abstract case.

We introduce the notion of *best approximation* of an LTS, e.g. a method for deriving the *most precise* abstract LTS that is correct. The most precise information can obviously obtained by replacing, both for states and transitions, the multiplicities with the most precise interval.

Definition 4.6 [Best Abstraction] We define $\alpha_{lts} : \mathcal{LTS} \rightarrow \mathcal{LTS}^\circ$, such that $\alpha_{lts}((S, \rightarrow, M_0, E)) = (S^\circ, \rightarrow^\circ, {M_0}^\circ, E)$ where[3] $\rightarrow^\circ = \{M^\circ \xrightarrow[\circ]{\Theta,\Delta^\circ,r} M_1^\circ \mid M \xrightarrow{\Theta,\Delta,r} M_1 \in \rightarrow\}$.

Notice that α_{lts} does not effectively introduce any approximation. Thus, an approximation order $\sqsubseteq_{lts}^\circ$ is fundamental for expressing the correctness of an abstract LTS with respect to a concrete one. Intuitively, lts° is a correct approximation of lts provided that $\alpha_{lts}(lts) \sqsubseteq_{lts}^\circ lts^\circ$.

For these purposes, we assume to extend the order $\leq_I$ over intervals to pairs of intervals. Given ${\Delta_1}^\circ, {\Delta_2}^\circ \in \widehat{Q}^\circ$ we define ${\Delta_1}^\circ \leq_I {\Delta_2}^\circ$ component-wise.

Definition 4.7 [Order on abstract LTS] Let $lts_i^\circ = (S_i^\circ, \rightarrow_\circ^i, M_{0,i}^\circ, E)$ with $i \in \{1, 2\}$ be two abstract LTS. For $M_1^\circ \in S_1^\circ, M_2^\circ \in S_2^\circ$, we say that $M_1^\circ \preccurlyeq M_2^\circ$ (M_2° simulates M_1°) iff

(i) $M_1^\circ \sqsubseteq^\circ M_2^\circ$,

(ii) for each $t_1^\circ = M_1^\circ \xrightarrow[\circ]{\Theta,\Delta_1^\circ,r} M_{1,1}^\circ \in \rightarrow_\circ^1$ there exists $t_2^\circ = M_2^\circ \xrightarrow[\circ]{\Theta,\Delta_2^\circ,r} M_{2,1}^\circ \in \rightarrow_\circ^2$, such that $\Delta_1^\circ \leq_I \Delta_2^\circ$ and $M_{1,1}^\circ \preccurlyeq M_{2,1}^\circ$.

We say that $lts_1^\circ \sqsubseteq_{lts}^\circ lts_2^\circ$ if $M_{0,1}^\circ \preccurlyeq M_{0,2}^\circ$.

As expected, the definition of the order for abstract LTS is based on a simulation between abstract states. Intuitively, M_2° simulates M_1° whenever: (i) M_2° approximates M_1°; (ii) each transition of M_1°, such as $M_1^\circ \xrightarrow[\circ]{\Theta,\Delta_1^\circ,r} M_{1,1}^\circ$, is matched by a transition $M_2^\circ \xrightarrow[\circ]{\Theta,\Delta_2^\circ,r} M_{2,1}^\circ$. Notice that it must be the case that the transitions are related to the same label Θ and that $\Delta_1^\circ \leq_I \Delta_2^\circ$, showing that the information about multiplicities is properly approximated.

The analysis. We define an effective method to derive an abstract LTS that safely approximates the concrete one. The abstract transition relation for abstract states is defined by the rules Table 3 (as previously, we are tacitly assuming to reason with respect to a given environment E). The rules are adapted from the concrete ones by replacing multiplicities with intervals of multiplicities, and exploit the abstract counterpart of the concrete operation $\oplus$ and $\ominus$. The abstract operations are defined as follows:

$$M^\circ \oplus^\circ N^\circ(X) = M^\circ(X) + N^\circ(X), \qquad I + J = [min(I) + min(J), max(I) + max(J)]$$

$$M^\circ \ominus^\circ N^\circ(X) = M^\circ(X) - N^\circ(X), \qquad I - J = [min(I) \widehat{-} min(J), max(I) \widehat{-} max(J)]$$

Analogously as in the concrete case, we write $\mathsf{LTS}^\circ((E, M_0^\circ)) = (S^\circ, \rightarrow_\circ, M_0^\circ, E)$ for the abstract LTS, obtained for the initial abstract state M_0° by transitive closure. As usual,

[3] We assume that Δ° is the best abstraction of Δ, derived component-wise in the obvious way.

$$\frac{E.X.\lambda = {\tau_r}^{\lambda}.Q}{M^\circ \xrightarrow[\circ]{\lambda, M^\circ(X), r} (M^\circ \ominus^\circ \{([0,0],X)\}) \oplus^\circ [\![Q]\!]^\circ} \qquad (\textbf{Delay-abs})$$

$$\frac{E.X.\lambda = {a_r}^{\lambda}.Q_1 \qquad E.Y.\mu = {\bar{a_r}}^{\mu}.Q_2}{M^\circ \xrightarrow[\circ]{(\lambda,\mu),(M^\circ(X),M^\circ(Y)),r} (M^\circ \ominus^\circ \{([1,1],X),([1,1],Y)\}) \oplus^\circ [\![Q_1]\!]^\circ \oplus^\circ [\![Q_2]\!]^\circ} \qquad (\textbf{Sync-abs})$$

Table 3
Abstract transition relation

$\mathsf{Next}^\circ(E, M^\circ)$ stands for the set of abstract transitions from M° with respect to the environment E.

The following theorem shows that the abstract LTS of an abstract state M° is a correct approximation of the LTS of multiset M, for each M represented by M°. This is the case in particular of the abstract LTS of the translation of the initial solution $[\![P]\!]$.

Theorem 4.8 *Let E be an environment and $M^\circ \in \mathcal{M}^\circ$ be an abstract state. For each multiset $M' \in \gamma(M^\circ)$, we have $\alpha_{lts}(\mathsf{LTS}((E, M'))) \sqsubseteq_{lts} \mathsf{LTS}^\circ((E, M^\circ))$.*

For a sake of simplicity in this paper we proposed an approximation which admits infinite abstract LTS. Note, however, that further approximations able to deal with infinite systems can be easily derived by means of widening operators [11]. In this context, for example, it is sufficient to replace the abstract operator $\oplus^\circ$ by its parametric version $\oplus^\circ_k$, which, given $k \in \mathbb{N}$, is defined as follows

$$M^\circ \oplus^\circ_k N^\circ(X) = \begin{cases} [min(M^\circ(X)) + min(N^\circ(X)), \infty] & \text{if } max(M^\circ(X)) + max(N^\circ(X)) > k, \\ M^\circ(X) + N^\circ(X) & \text{otherwise.} \end{cases}$$

It's easy to see that replacing $\oplus^\circ$ with $\oplus^\circ_k$ for any $k < \infty$ always lead to a finite and correct abstract LTS.

5 Abstract Markov Chains

We adopt the model of *Interval Markov Chains* (IMC) proposed in [15,22] in order to abstract DTMC. The correctness and precision of the abstraction are formalised, similarly as for LTS, by introducing a notion of best abstraction and an order over IMC. These notions are adapted from those proposed in [13,14,15] to our abstract interpretation framework.

Definition 5.1 An *Interval Discrete-Time Markov Chain* (IMC) is a tuple $(S^\circ, \mathbf{P}^-, \mathbf{P}^+, M_0^\circ)$ where: (i) $S^\circ \subseteq \mathcal{M}^\circ$ is a *finite or countable* set of abstract states and ${M_0}^\circ \in S^\circ$ is the *initial state*; (ii) and, $\mathbf{P}^-, \mathbf{P}^+ : S^\circ \to \mathsf{SDistr}(S^\circ)$ are the *lower* and *upper* probability functions, such that for all ${M_1}^\circ, {M_2}^\circ \in S^\circ$, $\mathbf{P}^-({M_1}^\circ)({M_2}^\circ) \leq \mathbf{P}^+({M_1}^\circ)({M_2}^\circ)$.

In the following we use $\mathcal{MC}^\circ$ to denote the set of IMC. The IMC model combines together non-deterministic and probabilistic steps similarly as in *Markov Decision Process* (MDP). Here, $\mathbf{P}^-({M_1}^\circ)({M_2}^\circ)$ and $\mathbf{P}^+({M_1}^\circ)({M_2}^\circ)$ define *intervals of probabilities*, that represent *lower* and *upper* bounds for the transition probabilities of moving from ${M_1}^\circ$ to ${M_2}^\circ$. Thus, for each abstract state there is a choice for the distribution yielding the probability to reach any other state. As usual, the non-determinism is resolved by a *scheduler* that chooses an *admissible* distribution for each step.

Definition 5.2 Let $mc^\circ = (S^\circ, \mathbf{P}^-, \mathbf{P}^+, M_0^\circ)$ be an IMC and let $M^\circ \in S^\circ$. We say that a distribution $\rho \in \mathsf{Distr}(S^\circ)$ is *admissible* for M° iff, for each $M'^\circ \in S^\circ$, $\mathbf{P}^-(M^\circ)(M'^\circ) \leq \rho(M'^\circ) \leq \mathbf{P}^+(M^\circ)(M'^\circ)$. We use $\mathsf{ADistr}_{mc^\circ}(M^\circ)$ for the admissible distributions for M°.

The notion of path for IMC is analogous to that presented for DTMC in Section 3. We therefore use the same notation.

Definition 5.3 [Sheduler] Let $mc^\circ = (S^\circ, \mathbf{P}^-, \mathbf{P}^+, M_0^\circ)$ be an IMC. A *sheduler* is a function $\Pi \colon \mathsf{FPaths}(S^\circ) \to \mathsf{Distr}(S^\circ)$ such that $\Pi(\pi) \in \mathsf{ADistr}_{mc^\circ}(\pi_{\text{last}})$ for each path $\pi \in \mathsf{FPaths}(S^\circ)$. We use $\mathsf{Adv}(mc^\circ)$ to denote the set of schedulers.

Given a scheduler $\Pi \in \mathsf{Adv}(mc^\circ)$ the probability space over paths can be defined analogously as for DTMC (see Definition 3.3). Thus, $\mathbf{P}^{\Pi}_{M^\circ}$ stands for the probability starting from the abstract state M° w.r.t. the scheduler Π.

In order to define the abstract validation of probabilistic reachability properties, we introduce a *may* and *must* notion of satisfaction for abstract states and propositional symbols. We say that an abstract state M° *must* satisfy A, $M^\circ \models^\forall A$, iff for each $M \in \gamma(M^\circ)$ we have $M \models A$. Analogously, an abstract state M° *may* satisfy A, $M^\circ \models^\exists A$, iff there exists $M \in \gamma(M^\circ)$ such that $M \models A$.

Notice that the notion of satisfaction is preserved by state approximation. Suppose that ${M_1}^\circ \sqsubseteq^\circ {M_2}^\circ$. If ${M_1}^\circ \models^\exists A$, then also ${M_2}^\circ \models^\exists A$; conversely, if ${M_2}^\circ \models^\forall A$, then also ${M_1}^\circ \models^\forall A$.

For reachability properties we derive both *under* and *over* approximations of the probability of reachability properties. Not surprisingly, it is enough to take the *minimum* and *maximum* probabilities w.r.t. all the schedulers.

Definition 5.4 [Reachability Properties] Let $mc^\circ = (S^\circ, \mathbf{P}^-, \mathbf{P}^+, M_0^\circ)$ be an IMC. The *lower* and *upper bound* of the probability of reaching a state satisfying a propositional symbol $A \in AP$, starting from $M^\circ \in S^\circ$, are defined as follows

$$\begin{aligned}
\mathsf{Reach}^-_{A,mc^\circ}(M^\circ) &= \inf_{\Pi \in \mathsf{Adv}(mc^\circ)} \mathbf{P}^{\Pi}_{M^\circ}(\{\pi^\circ \in C(M^\circ) \mid \pi^\circ[i] \models^\forall A \text{ for some } i \geq 0\}) \\
\mathsf{Reach}^+_{A,mc^\circ}(M^\circ) &= \sup_{\Pi \in \mathsf{Adv}(mc^\circ)} \mathbf{P}^{\Pi}_{M^\circ}(\{\pi^\circ \in C(M^\circ) \mid \pi^\circ[i] \models^\exists A \text{ for some } i \geq 0\})
\end{aligned}$$

Abstraction of Markov Chains. We define a function $\alpha_{MC} : \mathcal{MC} \to \mathcal{MC}^\circ$ that gives the *best abstraction* of a DTMC. Since there is no effective abstraction of states, similarly as for LTS, the derived probabilities are exact.

Definition 5.5 [Best Abstraction] We define $\alpha_{MC} : \mathcal{MC} \to \mathcal{MC}^\circ$ as follows

$$\alpha_{MC}((S, \mathbf{P}, M_0)) = (S^\circ, {\mathbf{P}_\alpha}^-, {\mathbf{P}_\alpha}^+, {M_0}^\circ)$$

where ${\mathbf{P}_\alpha}^-({M_1}^\circ, {M_2}^\circ) = {\mathbf{P}_\alpha}^+({M_1}^\circ, {M_2}^\circ) = \mathbf{P}(M_1)(M_2)$, for each $M_1, M_2 \in S$.

Effective approximations of a DTMC mc can be introduced by considering an IMC mc° such that $\alpha_{MC}(mc) \sqsubseteq^\circ_{mc} mc^\circ$, where $\sqsubseteq^\circ_{mc}$ is defined as follows.

Definition 5.6 [Order on IMC] Let $mc_i^\circ = ({S_i}^\circ, \mathbf{P}_i^-, \mathbf{P}_i^+, M_{0,i}^\circ)$ be two IMC for $i \in \{1, 2\}$. Given two abstract states ${M_i}^\circ \in {S_i}^\circ$ for $i \in \{1, 2\}$, we say that ${M_1}^\circ \preccurlyeq_{mc} {M_2}^\circ$ (${M_2}^\circ$ simulates ${M_1}^\circ$) iff

(i) ${M_1}^\circ \sqsubseteq^\circ {M_2}^\circ$;

(ii) for each distribution $\rho_1 \in \mathsf{ADistr}({M_1}^\circ)$ there exists a function $H \colon {S_1}^\circ \to {S_2}^\circ$ and a distribution $\rho_2 \in \mathsf{ADistr}({M_2}^\circ)$ such that,

   (a) for each $M^\circ \in {S_2}^\circ$, $\rho_2(M^\circ) = \sum_{M'^\circ \in H^{-1}(M^\circ)} \rho_1(M'^\circ)$.

(b) for each ${M'}^\circ \in {S_1}^\circ$, if $H({M'}^\circ) = M^\circ$ then ${M'}^\circ \preccurlyeq_{mc} M^\circ$.

Moreover, we say that ${mc_1}^\circ \sqsubseteq^\circ_{mc} {mc_2}^\circ$ iff $M^\circ_{0,1} \preccurlyeq_{mc} M^\circ_{0,2}$.

The order uses a sort of probabilistic simulation similarly as in [15,15,14]. Intuitively, ${M_2}^\circ$ simulates ${M_1}^\circ$ whenever: (i) ${M_2}^\circ$ approximates ${M_1}^\circ$: (ii) each distribution of ${M_1}^\circ$ is matched by a corresponding distribution of ${M_2}^\circ$, where the probabilities of the target states are eventually summed up.

The simulation $\preccurlyeq_{mc}$ provides sufficient conditions for the preservation of extremum probabilities, as stated by the following theorem.

Theorem 5.7 (Soundness of the order) *Let* $mc^\circ_i = ({S_i}^\circ, \mathbf{P}^-_i, \mathbf{P}^+_i, M^\circ_{0,i})$ *be two IMC and let* ${M_i}^\circ \in {S_i}^\circ$ *be two abstract states, for* $i \in \{1,2\}$*. If* ${M_1}^\circ \preccurlyeq_{mc} {M_2}^\circ$*, then for each propositional symbol* $A \in AP$*, we have*

$$\mathsf{Reach}^-_{A,mc^\circ_2}({M_2}^\circ) \leq \mathsf{Reach}^-_{A,mc^\circ_1}({M_1}^\circ) \leq \mathsf{Reach}^+_{A,mc^\circ_1}({M_1}^\circ) \leq \mathsf{Reach}^+_{A,mc^\circ_2}({M_2}^\circ)$$

6 Derivation of IMC

We define the abstract counterpart of the probabilistic translation function $\mathbf{H} : \mathcal{LTS} \to \mathcal{MC}$. Moreover, we discuss the soundness of our approach.

Our abstract LTS reports on transitions information about the label of the process (the labels of the processes) that participate to the move, the interval (the intervals) representing a possible range for its (their) multiplicities, and the rate of the basic action. Therefore, it should be well understood that the abstract rate associated to each transition is an *interval of rates*. From this kind of information, both *lower* and *upper* bounds for the probabilities of moving from an abstract state to another could be calculated following the guidelines of the derivation of the DTMC from the concrete LTS.

It is convenient, however, to maintain the calculation of the intervals of rates symbolic in order not to loose relational information on quantities of different occurrences of the same reagent. This means that the interval of rates assigned to each abstract transition will be represented by a symbolic expression on reagent variables. More in details, we adopt expressions such as (e, c) where: (i) $e \in \mathcal{Z}$ is a *symbolic expression* over the variables of $\mathcal{X}$; (ii) $c \in \mathcal{C}$ is a set of *membership constraints* of the form $X \in I$. We require that each expression (e, c) is *well-formed* meaning that, for each variable X occurring in e, there exists one and only one constraint $X \in I$ occurring in c.

Hence, we define the *abstract rate of a transition* as follows. Given a transition $t^\circ = M^\circ \xrightarrow[\circ]{\Theta,\Delta^\circ,r} M^\circ_1 \in \mathsf{Next}^\circ(E, M^\circ)$ we have

$$\mathsf{rate}^\circ(t^\circ) = \begin{cases} (X \cdot r, \{X \in I\}) & \Theta = \lambda, \lambda \in \mathcal{L}(E.X) \wedge \Delta^\circ = I, \\ (X \cdot (X\widehat{-}1) \cdot r, \{X \in I\}) & \Theta = (\lambda, \mu) \wedge \Delta^\circ = (I, I) \wedge \lambda, \mu \in \mathcal{L}(E.X), \\ (X \cdot Y \cdot r, \{X \in I_1, Y \in I_2\}) & \Theta = (\lambda, \mu) \text{ and } \Delta^\circ = (I_1, I_2) \wedge \lambda \in \mathcal{L}(E.X), \mu \in \mathcal{L}(E.Y) \wedge X \neq Y. \end{cases}$$

Moreover, we introduce the functions $\mathbf{R}^\circ : S^\circ \times S^\circ \to \mathcal{Z} \times \mathcal{C}$, and $\mathbf{E}^\circ : S^\circ \to \mathcal{Z} \times \mathcal{C}$, analogously as in the concrete case,

$$\mathbf{R}^\circ(M^\circ, {M'}^\circ) = \textstyle\sum^\circ_{t^\circ \in \mathsf{Ts}(M^\circ, {M'}^\circ)} \mathsf{rate}^\circ(t^\circ), \quad \mathbf{E}^\circ(M^\circ) = \textstyle\sum^\circ_{{M'}^\circ \in S^\circ} \mathbf{R}^\circ(M^\circ, {M'}^\circ),$$

$$(e_1, c_1) op^\circ (e_2, c_2) = (e_1 \; op \; e_2, \textstyle\bigcup_{X \in \mathcal{X}} \{X \in \bigcup^\circ_{(X \in I) \in c_i, i \in \{1,2\}} I\}) \text{ for } op \in \{+, /\}.$$

Intuitively, $\mathbf{R}^\circ(M^\circ, {M'}^\circ)$ reports the interval of rates corresponding to the move from M° to ${M'}^\circ$, while $\mathbf{E}^\circ(M^\circ)$ is the *abstract exit rate*. Both *lower* and *upper* bounds of the probability of moving from M° to M'° can be determined by $\mathbf{R}^\circ(M^\circ, M'^\circ)$ and by $\mathbf{E}^\circ(M^\circ)$. For these purposes we need to consider the *worst* case and *best* case scenario, respectively.

Definition 6.1 [Derivation of the IMC] We define an *abstract probabilistic translation* function $\mathbf{H}^\circ : \mathcal{LTS}^\circ \to \mathcal{MC}^\circ$ such that $\mathbf{H}^\circ((S^\circ, \to^\circ, {M_0}^\circ, E)) = (S^\circ, \mathbf{P}^-, \mathbf{P}^+, M_0^\circ)$, where $\mathbf{P}^-, \mathbf{P}^+ \colon S^\circ \to \mathsf{SDistr}(S^\circ)$ are the *lower* and *upper* probability functions, such that for all ${M_1}^\circ \in S^\circ$

(i) if $max(\mathbf{E}^\circ(M_1^\circ)) = 0$, then $\mathbf{P}^+(M_1^\circ)(M_2^\circ) = \mathbf{P}^-(M_1^\circ)(M_2^\circ) = 0$, for each $M_1^\circ \neq M_2^\circ$, and $\mathbf{P}^+(M_1^\circ)(M_1^\circ) = \mathbf{P}^-(M_1^\circ)(M_1^\circ) = 1$;

(ii) if $max(\mathbf{E}^\circ(M_1^\circ)) > 0$ then
   (a) if $min(\mathbf{E}^\circ(M_1^\circ)) = 0$ then $\mathbf{P}^+(M_1^\circ)(M_1^\circ) = 1$ and $\mathbf{P}^-(M_1^\circ)(M_1^\circ) = 0$,
   (b) for each M_2°, if $min(\mathbf{R}^\circ(M_1^\circ, M_2^\circ)) = 0$ then $\mathbf{P}^-(M_1^\circ)(M_2^\circ) = 0$ else $\mathbf{P}^-(M_1^\circ)(M_2^\circ) = min(\mathbf{R}^\circ(M_1^\circ, M_2^\circ)/^\circ\mathbf{E}^\circ(M_1^\circ))$,
   (c) for each M_2°, if $max(\mathbf{R}^\circ(M_1^\circ, M_2^\circ)) = 0$ then $\mathbf{P}^+(M_1^\circ)(M_2^\circ) = 0$ else $\mathbf{P}^+(M_1^\circ)(M_2^\circ) = max(\mathbf{R}^\circ(M_1^\circ, M_2^\circ)/^\circ\mathbf{E}^\circ(M_1^\circ))$.

Intuitively, the lower and upper bound probabilities for the move from M° to M'° are computed by minimizing and maximizing the solution of $\mathbf{R}^\circ(M^\circ, M'^\circ)/^\circ\mathbf{E}^\circ(M^\circ)$, respectively. This reasoning has to be properly combined with the special cases when $max(\mathbf{E}^\circ(M^\circ)) = 0$ or $min(\mathbf{E}^\circ(M^\circ)) = 0$. When $max(\mathbf{E}^\circ(M^\circ)) = 0$ all the states represented by M° are stable, while when $min(\mathbf{E}^\circ(M^\circ)) = 0$ a state represented by M° is stable.

Note that in order to find the maximum and minimum of a symbolic, constrained expression $(e, c) \in \mathcal{Z} \times \mathcal{C}$, when it's not trivial, it's sufficient to evaluate the expression e for the *stationary points* (that can be found by differentiate e and by setting the result equal to 0) and for the boundaries of the intervals in c constraining variables of e.

The following theorems state the soundness of our approach.

Theorem 6.2 *Let $lts_i^\circ = ({S_i}^\circ, \to_i{}^\circ, {M_{0,i}}^\circ)$ be two abstract LTS. If $lts_1^\circ \sqsubseteq_{lts}^\circ lts_2^\circ$, then also $\mathbf{H}^\circ(lts_1^\circ) \sqsubseteq_{mc}^\circ \mathbf{H}^\circ(lts_2^\circ)$.*

Theorem 6.3 *Let E be an environment and $M_0 \in \mathcal{M}$ be a multiset. We have*

$$\alpha_{MC}(\mathbf{H}(\mathsf{LTS}((E, M_0)))) \sqsubseteq_{mc}^\circ \mathbf{H}^\circ(\alpha_{lts}(\mathsf{LTS}((E, M_0)))).$$

Example 6.4 Consider again the chemical reaction described by the environment E of Example 3.5. We show the ability of our analysis for predicting the probabilistic behavior of the reaction, described by E, w.r.t. different initial concentrations. The idea is to model simultaneously the behavior of several experiments, described by the abstract initial state. As an example we can consider the initial abstract state $M^\circ = \{([2, 3], X)$,$([2, 5], Y)$, $([1, 10], W)\}$.

The obtained abstract LTS is depicted in Fig.3 where [4],

$M_1^\circ = \{([1, 2], X), ([1, 4], Y), ([1, 10], W), ([1, 1], XY)\}$ $\quad M_2 = \{([1, 2], X), ([2, 5], Y), ([1, 10], W)\}$
$M_3 = \{([0, 1], X), ([0, 3], Y)([1, 10], W), ([2, 2], XY)\}$ $\quad M_4 = \{([0, 1], X), ([1, 4], Y), ([1, 10], W), ([1, 1], XY)\}$
$M_5 = \{([0, 1], X), ([2, 5], Y), ([1, 10], W)\}$ $\quad M_6 = \{([0, 2], Y)([1, 10], W), ([3, 3], XY)\}$
$M_7 = \{([0, 3], Y)([1, 10], W), ([2, 2], XY)\}$ $\quad M_8 = \{([1, 4], Y)([1, 10], W), ([1, 1], XY)\}$ $\quad M_9 = \{([2, 5], Y)([1, 10], W)\}$

It is convenient to consider again the probabilistic reachability properties discussed in the previous example. Thus, we assume $r_1 = 3r_2$ and we derive the IMC of Fig.4.

We compute the minimum and maximum probabilities (denoted by $\mathbf{P}^-(M^\circ)$ and $\mathbf{P}^+(M^\circ)$, respectively) to reach, from M°, a state where *at least two* binding XY appear. We summarize the most relevant steps of the reasoning that permits to compute $\mathbf{P}^-(M^\circ)$ and $\mathbf{P}^+(M^\circ)$; for a complete discussion on probabilistic model checking we refer the reader to [23,37,15].

[4] We adopt the same notation of Example 3.5 for states

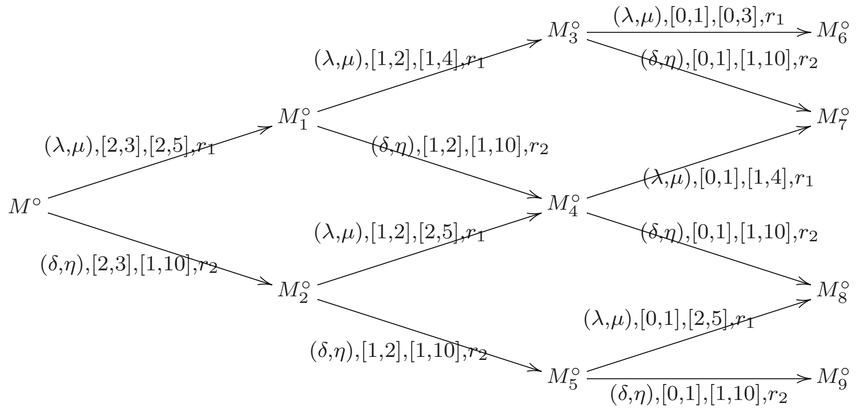
M°
(λ,μ),[2,3],[2,5],r_1
(δ,η),[2,3],[1,10],r_2
M_1°
M_2°
(λ,μ),[1,2],[1,4],r_1
(δ,η),[1,2],[1,10],r_2
(λ,μ),[1,2],[2,5],r_1
(δ,η),[1,2],[1,10],r_2
M_3°
M_4°
M_5°
(λ,μ),[0,1],[0,3],r_1
(δ,η),[0,1],[1,10],r_2
(λ,μ),[0,1],[1,4],r_1
(δ,η),[0,1],[1,10],r_2
(λ,μ),[0,1],[2,5],r_1
(δ,η),[0,1],[1,10],r_2
M_6°
M_7°
M_8°
M_9°

Figure 3. The LTS

The states that contain at least two occurrences of XY are $\mathbf{M_3}$, $\mathbf{M_7}$ and $\mathbf{M_6}$. As a consequence, we have $\mathbf{P}^-(M_3^\circ) = \mathbf{P}^-(M_6^\circ) = \mathbf{P}^-(M_7^\circ) = 1$, and $\mathbf{P}^+(M_3^\circ) = \mathbf{P}^+(M_6^\circ) = \mathbf{P}^+(M_7^\circ) = 1$. Moreover, we have also $\mathbf{P}^-(M_8^\circ) = \mathbf{P}^-(M_9^\circ) = \mathbf{P}^-(M_5^\circ) = 0$ and $\mathbf{P}^+(M_8^\circ) = \mathbf{P}^+(M_9^\circ) = \mathbf{P}^+(M_5^\circ) = 0$.

Let us consider the lower bound. The most important observation is that $\mathbf{P}^-(M_4^\circ) = 0$ since there is a self-loop that allows state M_4° not to reach state M_7°. Thus, we have also $\mathbf{P}^-(M_2^\circ) = 0$. Moreover, for M_1° we have to consider the admissible distributions ρ that minimize $\rho(M_1^\circ)(M_3{}^\circ) \cdot \mathbf{P}^-(M_3^\circ) + \rho(M_1^\circ)(M_4{}^\circ) \cdot \mathbf{P}^-(M_4^\circ) = \rho(M_1^\circ)(M_3{}^\circ)$. Thus, we derive $\mathbf{P}^-(M_1^\circ) = (3/13)$, and analogously $\mathbf{P}^-(M^\circ) = (18/48) \cdot \mathbf{P}^-(M_1^\circ) + (30/48) \cdot \mathbf{P}^-(M_2^\circ) = (18/48) \cdot (3/13)$.

For the upper bound we obtain in a similar way $\mathbf{P}^+(M^\circ) = (45/48) \cdot \mathbf{P}^+(M_1^\circ) + (3/48) \cdot \mathbf{P}^+(M_2^\circ) = (45/48) \cdot ((12/13) + (1/13) \cdot (12/13)) + (3/48) \cdot ((15/16) \cdot (12/13))$. The most relevant difference is that in this case $\mathbf{P}^+(M_4^\circ) = (12/13)$ by maximizing the probability of moving from M_4° to M_7°.

Finally, we consider the probability of reaching a state where *at least three* bindings XY are created. In this case, only state M_6° satisfies the requirement, and we obtain $\mathbf{P}^-(M^\circ) = 0$ and $\mathbf{P}^+(M^\circ) = (45/48) \cdot (12/13) \cdot (9/10)$.

It is worth noticing that the result of our analysis is very accurate. In fact, for the reachability properties previously considered both lower and upper bound correspond to the concrete probability of one of the experiments represented by the abstract initial state M°. For example, the lower bound of the probability that we reach a state where *at least two*

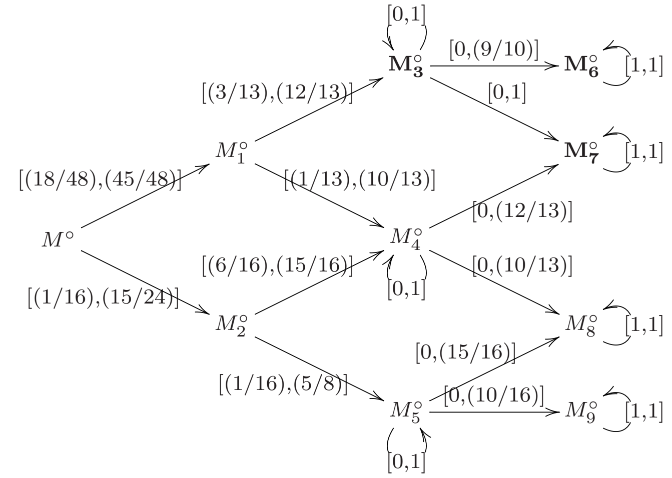

Figure 4. The DTMC

binding XY are created is $(18/48) \cdot (3/13)$, e.g. the probability of the "worst case' ' concrete experiment we have illustrated in Example 3.5. This result could not be achieved without the relational information about the number of occurrences of reagents, that we profitably exploit for computing the intervals of probabilities. □

7 Conclusions and Related Works

In this paper we have proposed a methodology, based on abstract interpretation, for probabilistic abstract model checking of biological systems, modeled in the CGF calculus. We believe that the approach could be extended in a simple way to the full calculus with communication, by considering the language of [30].

Several abstraction methods for probabilistic systems, modeled as DTMC or MDP, have been recently investigated. The proposals of [13,14,15] present similar approaches, based on MDP and IMC, respectively. The abstract model is built over a partition of the concrete state space by computing the abstract probabilities from the concrete probabilities (this is a sort of best abstraction of the concrete DTMC). As a consequence, these approaches can handle finite state systems only. Huth [22] proposes a more general approach based on IMC where the abstraction of states is formalized using a sort of abstract interpretation. Even if the framework admits infinite state systems, no effective methods for deriving an abstract model for a given language is investigated. The technique of [24] extends the approaches of [13,14], using games, in order to more accurately abstract MDP. De Alfaro [1] proposes an original method for the abstraction of finite state MDP, based on regions. Monnieux [26] proposes an approximation method, based on abstract interpretation, for the validation of trace properties of probabilistic and non-deterministic transition systems. Techniques of *backward* and *forward* analysis are successfully applied.

Our approach differs from most of the previous proposals in that we have introduced an effective method (even for infinite state systems) to compute an abstraction of the probabilistic semantics, relying on the abstract LTS. In our opinion, such an abstraction is particularly adequate for achieving correct predictions (*lower* and *upper* bounds on the concrete probability) of the possible behavior of a set of experiments. As it is outlined in Example 6.4, the set of experiments we want to analyze is chosen by considering a suitable initial abstract multiset.

In our framework, based on abstract interpretation, new analyses could be designed by introducing new abstract LTS semantics. We are currently investigating a parametric version of our framework where the partitioning of intervals, and thus of abstract states, could be realized in a coarser or finer way. This could give the possibility to find a trade-off between precision and complexity and also to address different applications. In this setting, it would be interesting to study refinement techniques, guided by the formula in the style of [7]. Moreover, we intend to investigate whether the domain of intervals could be replaced with more precise numerical domains able to model also relational information, such as the domain of convex polyhedra.

References

[1] L. de Alfaro and Pritam Roy. *Magnifying-Lens Abstraction for Markov decision Process.* Proc. of CAV '07, LNCS 4590, 325–338, 2007.

[2] C. Bodei, P.Degano, F.Nielson and H.Riis Nielson. *Static Analysis for the Pi-Calculus with Applications to Security.* Information and Computation, 168: 68-92, 2001.

[3] L. Cardelli. *Brane Calculi.* Proc. of CMSB '04, LNCS 3082, 257–278, 2004.

[4] L. Cardelli. *On Process Rate Semantics.* To appear in Theoretical Computer Science, 2008.

[5] N. Chabrier, Marc Chiaverini, Vincent Danos and F. Fages. *Modeling and Querying Biomolecular Interaction Networks.* Theoretical Computer Science 325(1), 25-44, 2004.

[6] M. Clarke, O. Grumberg and E. Long. *Model Checking and Abstraction.* TOPLAS, 16(5), 1512-1542, 1994.

[7] M. Clarke, O. Grumberg, S. Jha, Y. Lu and H. Veith. *Counter-example Guided Abstraction Refinement.* Proc. of CAV '00, LNCS 1855, 154–169, 2000.

[8] P. Cousot and R. Cousot. *Static Determination of Dynamic Properties of Programs.* Proc. of POPL'76 , 106–130, 1976.

[9] P. Cousot and R. Cousot. *Abstract Interpretation: A Unified Lattice Model for Static Analysis of Programs by Construction or Approximation of Fixpoints.* Proc. of POPL'77, 238–252, 1977.

[10] P. Cousot and R. Cousot. *Systematic Design of Program Analysis Frameworks.* Proc. of POPL'79 , 269–282, 1979.

[11] P. Cousot and R. Cousot. *Comparing the Galois Connection and Widening/Narrowing Approaches to Abstract Interpretation.* Proc. of PLILP'92, LNCS 631, 269–295, 1992.

[12] D. Dams, R. Gerth and O. Grumberg. *Abstract Interpretation of Reactive Systems.* TOPLAS, 19(2), 253-291, 1997.

[13] P. D'Argenio, B. Jeannet, H. Jensen and K. Larsen. *Reachability Analysis of Probabilistic Systems by Successive Refinements.* Proc. of PAPM-PROMIV'01, LNCS 2165, 39–56, 2001.

[14] P. D'Argenio, B. Jeannet, H. Jensen and K. Larsen. *Reduction and Refinement Strategies for Probabilistic Analysis.* Proc. of PAPM-PROMIV'02, LNCS 2399, 57–76, 2002.

[15] H. Fecher, M. Leucker and V. Wolf. *Don't Know in Probabilistic Systems.* Proc. of SPIN'06, LNCS 3925, 71–88, 2006.

[16] J. Feret. *Occurrence Counting Analysis for the pi-calculus.* ENTCS 39 (2), 2001.

[17] J. Feret. *Abstract Interpretation-Based Static Analysis of Mobile Ambients.* Proc. of SAS'01, LNCS 2126, 412-430, Springer Verlag, 2001.

[18] R.Gori and F. Levi. *A new occurrence Counting analysis for BioAmbients.* Proc. of APLAS '05, LNCS 3780, 381–400, 2005.

[19] R.Gori and F. Levi. *An Analysis for proving Temporal Properties of Biological Systems.* Proc. of APLAS '06, LNCS 4279, 234–252, 2006.

[20] R. R. Hansen and J. G. Jensen and F. Nielson and H. R.Nielson. *Abstract Interpretation of Mobile Ambients.* Proc. of SAS'99, LNCS 1694, 135-148, Springer-Verlag, 1999.

[21] A. Hinton, M. Kwiatkowska, G. Norma and D. Parker. *PRISM: a tool for automatic verification of probabilistic systems.* Proc. of TACAS'06, LNCS 3920, 441-444, Springer-Verlag, 2006.

[22] M. Huth. *On finite-state approximants for probabilistic computation tree logic.* Theoretical Computer Science, 346(1), 113–134, 2005.

[23] M. Kwiatkowska. *Model checking for probability and time: from theory to practice.* Proc. of LICS' 03, 351–360, 2003.

[24] M. Kwiatkowska, G. Norman and D. Parker. *Game-based Abstraction for Markov Decision Processes.* Proc. of QEST'06, 157–166, 2006.

[25] F. Levi and S. Maffeis. *On Abstract Interpretation of Mobile Ambients.* Information and Computation 188, 179–240, 2004.

[26] D. Monnieaux. *Abstract interpretation of programs as Markov Decision Processes.* Science of Computer Programming, 58(1-2), 179–205, 2005.

[27] F. Nielson, H.R. Nielson, R.R. Hansen. *Validating firewalls using flow logics.* Theoretical Computer Science, 283(2), 381-418, 2002.

[28] F. Nielson, H.R. Nielson and H. Pilegaard. *Spatial Analysis of BioAmbients.* Proc. of SAS'04, LNCS 3148, pp. 69–83, Springer-Verlag, 2004.

[29] A. Phillips and L. Cardelli. *A Correct Abstract Machine for the Stochastic Pi-calculus.* Proc. of BioCONCUR '04, ENTCS, 2004.

[30] A. Phillips and L. Cardelli. *Efficient, Correct Simulation of Biological Processes in the Stochastic Pi-calculus.* Proc. of CMSB '07, LNCS 4695, 184–199, 2007.

[31] H. Pilegaard, F. Nielson and H.R. Nielson. *Static Analysis of a Model of the LDL Degradation Pathway.* Proc. of CMSB'05, 2005.

[32] C.Priami. *Stochastic π-calculus.* The Computer Journal, 38, 578–589,1995.

[33] C.Priami and P. Quaglia. *Beta binders for biological interactions.* Proc. of CMSB'04, LNCS 3082,20–33,2005.

[34] C. Priami, A. Regev, W. Silverman and E. Shapiro. *Application of a stochastic name-passing calculus to representation and simulation of molecular processes.* Information Processing Letters, 80 (1), 25–31, 2001.

[35] A. Regev, E. M. Panina, W. Silverman, L. Cardelli and E. Shapiro. *BioAmbients: an Abstraction for Biological Compartments.* Theoretical Computer Science, 325, 141–167, 2004.

[36] A. Regev, W. Silverman and E. Shapiro. *Representation and Simulation of Biochemical Processes using the pi-calculus process algebra.* Proc. of the Pacific Symposium on Biocomputing 2001, 6, 459–470, 2001.

[37] K. Sen, M. Viswanathan and G. Agha. *Model Checking Markov Chains in the Presence of Uncernainties.* Proc. of TACAS'06, LNCS 3920, 394-410, 2006.