# Coupling and Importance Sampling for Statistical Model Checking

Benoît Barbot, Serge Haddad, and Claudine Picaronny

LSV, ENS Cachan & CNRS & INRIA, Cachan, France
{barbot,haddad,picaronny}@lsv.ens-cachan.fr

**Abstract.** Statistical model-checking is an alternative verification technique applied on stochastic systems whose size is beyond numerical analysis ability. Given a model (most often a Markov chain) and a formula, it provides a confidence interval for the probability that the model satisfies the formula. One of the main limitations of the statistical approach is the computation time explosion triggered by the evaluation of very small probabilities. In order to solve this problem we develop a new approach based on importance sampling and coupling. The corresponding algorithms have been implemented in our tool Cosmos. We present experimentation on several relevant systems, with estimated time reductions reaching a factor of $10^{-120}$.

**Keywords:** statistical model checking, rare events, importance sampling, coupling.

## 1 Introduction

**Quantitative Model Checking.** Model checking [13] is an efficient verification method to check that the behaviour of a system fulfills properties expressed by some temporal logic. It has been successfully implemented in a variety of tools, thanks to it algorithmic simplicity. Although a method initially dedicated to discrete event systems, it has been adapted to performance evaluation in order to check quantitative properties and in particular to estimate probabilities [18].

**Statistical Model-Checking.** Analysis of stochastic systems requires *numerical* or *statistical* techniques. Numerical methods give exact results (up to numerical approximations) but significantly restrict the class of analysable systems (manageable size, Markov properties, etc.). Otherwise, statistical method may be used. By simulating a big sample of trajectories of the system and computing the ratio of these trajectories that satisfy a given property, it produces a probabilistic framing of the expected value. To generate the sample we only need to have an operational stochastic semantic of the system. This usually requires a very small state space compared to the numerical method and allows to deal with huge models [20].

**Rare Events.** The main drawback of the statistical model-checking is its inefficiency in dealing with very small probabilities. The size of the sample of

C. Flanagan and B. König (Eds.): TACAS 2012, LNCS 7214, pp. 331–346, 2012.

simulations required to estimate these small probabilities exceeds achievable capacities. This difficulty is known as the *rare event* problem. Several methods have been developed to cope with this problem whose main one is *importance sampling*. Importance sampling consists in modifying the model and in substituting to the indicator random variable related to the satisfaction of the formula, another variable with same mean and, in the favorable cases, reduced variance. Most of the techniques related to importance sampling are based on heuristics and cannot provide any confidence interval for the estimated probability.

**Our Contribution.** Here we propose a method based on importance sampling to estimate in a reliable way a very small probability[1].

We set up a theoretical framework using coupling theory [21], yielding an efficient importance sampling that guarantees a variance reduction and provides a confidence interval. This is done by performing numerical model checking on a small suitable reduction of the Markov chain associated with the system. The results are then used as parameters required for the importance sampling technique. Such a method deals with huge (possibly infinite) systems which are out of reach of numerical model checking and standard statistical model checking. It can be applied to a large variety of models compared to existing importance sampling methods which are usually put up in an ad-hoc way for particular families of models. Furthermore to the best of our knowledge, this is the first importance sampling method that provides a true (and not an approximate) confidence interval.

We implemented our method in the statistical model-checker Cosmos [4] using the tool Prism for the numerical computation on the reduced model. We tested our tool on several models getting impressive time reductions.

**Organisation.** In section 2, we motivate this work and we give a state of the art related to rare event handling. Then we develop our method in section 3. Afterwards we present and discuss experimentation in section 4. Finally in section 5, we conclude and give some perspectives to this work. Due to lack of place, the proofs can be found in [6].

## 2 Motivation and State of the Art

The temporal logics for probabilistic systems include both the qualitative and quantitative aspects of the systems. For instance, such logics can express (1) boolean assertions like "the probability of failure of a fixed component is below some threshold" and (2) numerical indices like "the mean delivery time of a packet assuming three collisions". The semantics of such formula is based on the probability that a random path fulfills some property (in CSL [2]) or (in a more general setting) on the conditional expectation of a path random variable whose condition is the satisfaction of some property by the random path (in HASL [4]).

[1] We have presented in a previous paper [5] a preliminary approach of this method with stronger assumptions and without using the coupling theory.

Model checking of these logics can be performed in a numerical or in a statistical way. The former approach builds the underlying stochastic process of the model and then computes probabilities or expectations using direct or iterative methods. Such methods have been implemented efficiently in tools like PRISM [17], LiQuor [9] or MRMC [16].

However these methods have two drawbacks. On the one hand, they rely on strong assumptions about the stochastic process that must be a Markov chain (see for instance [2]) or at least a regenerative process (see for instance [1]). On the other hand they suffer from the combinatorial explosion of the size of the stochastic process w.r.t. the size of the model.

Models with huge stochastic process are handled by statistical model checking. The corresponding methods randomly generate a (large) set of execution paths and check whether the paths fulfill the formula. The result is a probabilistic estimation of the satisfaction given by a confidence interval [3]. In principle, it only requires to maintain a current state (and some numerical values in case of a non Markovian process). Furthermore no regenerative assumption is required and it is easier to parallelize the methods. Several tools include statistical model checking: COSMOS [4], GREATSPN [8], PRISM [17], UPPAAL [7], VESTA [23], YMER [25].

Model checking of probabilistic systems is particularly important for events which have disastrous consequences (loss of human life, financial ruin, etc.), but occur with very small probability. Unfortunately statistical model checking of *rare events* triggers a computation time explosion, forbidding its use. To illustrate this point, suppose one wants to estimate an unknown probability $p = 10^{-13}$ and one chooses to generate $10^{10}$ paths (which is already a large number) for such an estimation. With probability larger than 0.999 the result is 0, giving no information on the value of $p$. With probability smaller than 0.001 the result will be greater or equal than $10^{-10}$ which is a very crude estimation.

Thus *acceleration* techniques [22] have been introduced to cope with this problem. The two main families of methods are *splitting* and *importance sampling*.

Splitting methods [19] duplicate or eliminate paths during their generation depending on their intermediate behaviour. When generation is ended, the bias introduced by these operations is taken into account for the estimation of the probability. Splitting methods are by nature heuristics, model dependent and very few theoretical results are known.

Importance sampling methods [14] generate paths of a system whose probability distribution of transitions have been changed to increase the probability of the event to occur. A weight is then affected to each path to correct the introduced bias. The goal is to substitute to the Bernoulli random variable corresponding to the occurrence of the rare event, another one with same mean value (the probability of event occurrence) but smaller variance. In Markov chains, an optimal change of distribution exists leading to a zero variance but it requires more information than the searched value! However this optimal importance sampling allows to design efficient heuristics for some classes of models.

The modification of the distribution can be performed at the model level (called *static*) or at the Markov chain level (called *dynamic*). The static

importance sampling requires no additional memory but in general provides a smaller reduction of variance than the dynamic importance sampling. More precisely, it is proved in [11] that asymptotic optimality (a weaker requirement than optimality) cannot be obtained even for very simple classes of models by static importance sampling. In full generality, the dynamic importance sampling [24] requires to maintain a memory whose size is proportional to the size of the Markov chain which is exactly what one wants to avoid. To deal with this problem, in [12] the authors develop the following method: (1) the possible distributions belong to the convex hull of a finite number of distributions, (2) the state space is partitioned and (3) a distribution is selected for each subset of this partition. They prove that for a simple class of models their method is asymptotically optimal. Other empirical approaches turn out to be efficient [15,10].

Summarizing, theoretical results (reduction of variance, asymptotical optimality, etc.) have been obtained for importance sampling. However none of these methods can produce a reliable confidence interval[2] for the mean value since the distribution of the modified random variable is unknown.

## 3 General Approach

### 3.1 Preliminaries

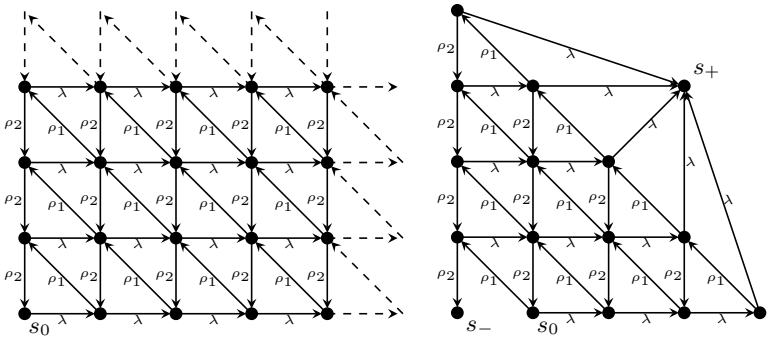
**Definition 1.** *A discrete time Markov chain (DTMC) $\mathcal{C}$ is defined as a set of states $S$, an initial state $s_0$, and a transition probability matrix $\mathbf{P}$ of size $S \times S$. The state of the chain at time $n$ is a random variable $X_n$ defined inductively by* $\Pr(X_0 = s_0) = 1$ *and* $\Pr(X_{n+1} = s' \mid X_n = s, X_{n-1} = s_{n-1}, \ldots, X_0 = s_0) = \Pr(X_{n+1} = s' \mid X_n = s) = \mathbf{P}(s, s')$.

**Example 1.** *The figure 1(a) represents a Markov chain of a tandem queue system. This system contains two queues, the number of clients in the first queue is represented on the horizontal axis and the number of clients in the second one is represented on the vertical axis. In the initial state $s_0$, the two queues are empty. Given some state, a new client comes in the first queue with probability $\lambda$, a client leaves the first queue for the second one with probability $\rho_1$ and a client leaves the second queue and exits with probability $\rho_2$ ($\lambda + \rho_1 + \rho_2 = 1$). An impossible event (due to the emptiness of some queue) corresponds to an event leaving unchanged the state. These loops are not represented in the figure.*

Usually the modeller does not specify its system with a Markov chain. He rather defines a higher level model $\mathcal{M}$ (a queueing network, a stochastic Petri net, etc.), whose operational semantic is a Markov chain $\mathcal{C}$.

In the context of model checking, the states of chain $\mathcal{C}$ are labelled with atomic propositions. The problem we address here is the computation of the probability that a random path starting from state $s_0$ satisfies a formula $aUb$ where $U$ is the *Until* operator and $a, b$ are atomic propositions. Observe that in continuous time

[2] In contrast to the empirical confidence interval based on approximations by the normal distribution.




(a) DTMC for the tandem queues (b) DTMC with absorbing states

**Fig. 1.** DTMC for tandem queues

Markov chains, this probability only depends on its embedded DTMC. Thus our results are also applicable in a continuous time setting. We (implicitly) transform $\mathcal{C}$ by lumping together all the states that satisfy $b$ into an absorbing state $s_+$ (i.e. $\mathbf{P}(s_+, s_+) = 1$) and states that satisfy $\neg a \wedge \neg b$ into an absorbing state $s_-$. We assume that there is no terminal strongly connected component of $\mathcal{C}$ whose every state satisfies $a \wedge \neg b$[3]. Hence in the modified chain, the probability to reach $s_+$ or $s_-$ is equal to 1 and probability of satisfying the formula is the probability to reach $s_+$.

**Example 1.** *The figure 1(b) shows the transformation of the tandem queues were the states have been lumped together w.r.t. the propositions a:* `There is at least one client in some queue` *and b:* `the sum of clients in both queues is equal to` 5. *The initial state $s_0$ is now the state with one client in the first queue (to avoid $s_0 = s_-$). We are looking for the probability to have simultaneously at least five clients between two idle periods.*

The statistical approach consists in generating $K$ paths of the Markov chain which ends in an absorbing state. Let $K_+$ be the number of paths ending in the $s_+$ state. The random variable $K_+$ follows a binomial distribution with parameters $p$ and $K$. Thus the random variable $\frac{K_+}{K}$ has a mean value $p$ and a variance $\frac{p-p^2}{K}$. When $K$ goes to infinity the variance goes to 0. In order to be more precise on the estimation, we introduce the notion of confidence interval.

**Definition 2.** *Let $X_1, \ldots, X_n$ be independent random variables following a common distribution including a parameter $\theta$. Let $0 < \gamma < 1$ be a confidence level. Then a confidence interval for $\theta$ with level at least $\gamma$ is given by two random variables $l(X_1, \ldots, X_n)$ and $u(X_1, \ldots, X_n)$ such that for all $\theta$:*

$$\Pr\left(l(X_1, \ldots, X_n) \leq \theta \leq u(X_1, \ldots, X_n)\right) \geq \gamma$$

[3] There is currently no satisfactory solution for the statistical model checking of the unbounded until for chains that do not fulfill this assumption.

For standard parametrized distributions like the normal or the Bernoulli ones, it is possible to compute confidence intervals [3]. Thus, given a number of paths $K$ and a confidence level $1 - \varepsilon$, the method produces a confidence interval. As discussed before when $p \ll 1$, the number of paths required for a small confidence interval is too large to be simulated.

The importance sampling method uses a modified transition matrix $\mathbf{P'}$ during the generation of paths. $\mathbf{P'}$ must satisfy:

$$\mathbf{P}(s, s') > 0 \Rightarrow \mathbf{P'}(s, s') > 0 \vee s' = s_- \tag{1}$$

which means that this modification cannot remove transitions that have not $s_-$ as target, but can add new transitions. The method maintains a correction factor called $L$ initialized to 1; this factor represents the *likelihood* of the path. When a path crosses a transition $s \to s'$ with $s' \neq s_-$, $L$ is updated by $L \leftarrow L\frac{\mathbf{P}(s,s')}{\mathbf{P'}(s,s')}$. When a path reaches $s_-$, $L$ is set to zero. If $\mathbf{P'} = \mathbf{P}$ (i.e. no modification of the chain), the value of $L$ when the path reaches $s^+$ (resp. $s^-$) is 1 (resp. 0).

Let $V_s$ (resp. $W_s$ ) be the random variable associated with the final value of $L$ for a path starting in $s$ in the original model (resp. in the modified one). By definition, $\mathbf{E}(V_{s_0}) = p$. The following proposition establishes the correctness of the method.

**Proposition 1.** $\mathbf{E}(W_{s_0}) = p$.

A good choice of $\mathbf{P'}$ should reduce the variance of $W_{s_0}$ w.r.t. to variance of $V_{s_0}$. The following proposition shows that there exists a matrix $\mathbf{P'}$ which leads to a null variance. We denote the probability to reach $s_+$ starting from $s$ by $\mu(s)$.

**Proposition 2.** *Let* $\mathbf{P'}$ *be defined by*

- $\forall s$ *such that* $\mu(s) \neq 0$, $\mathbf{P'}(s, s') = \frac{\mu(s')}{\mu(s)}\mathbf{P}(s, s')$
- $\forall s$ *such that* $\mu(s) = 0$, $\mathbf{P'}(s, s') = \mathbf{P}(s, s')$

*Then for all* $s$, *we have* $\mathbf{V}(W_s) = 0$.

This result has a priori no practical application since it requires the knowledge of $\mu$ for all states, whereas we only want to estimate $\mu(s_0)$!

The coupling method [21] is a classical method for comparing two stochastic processes, applied in different contexts (establishing ergodicity of a chain, stochastic ordering, bounds, etc.). In the sequel we will develop a new application for coupling. A coupling between two Markov chains is a chain whose space is a subset of the product of the two spaces which satisfies: (1) the projection of the product chain on any of its components behaves like the original corresponding chain, (2) an additional constraint which depends on the property to be proved (here related to the absorbing states).

**Definition 3.** *Let* $\mathcal{C} = (S, \mathbf{P})$ *and* $\mathcal{C}' = (S', \mathbf{P'})$ *be two Markov chains with* $s_+$ *and* $s_-$ *two absorbing states of* $\mathcal{C}$ *and* $s'_+$ *and* $s'_-$ *two absorbing states of* $\mathcal{C}'$. *A coupling between* $\mathcal{C}$ *and* $\mathcal{C}'$ *is a DTMC* $\mathcal{C}^{\otimes} = (S^{\otimes}, \mathbf{P}^{\otimes})$ *such that :*

- $S^{\otimes} \subseteq S \times S'$
- $\forall s \neq s_1 \in S\ ,\ \forall (s,s') \in S^{\otimes},\ \mathbf{P}\ (s\ ,s_1) = \sum_{s'_1 \in S'} \mathbf{P}^{\otimes}((s,s'),(s_1,s'_1))$ *and* $\forall s' \neq s'_1 \in S',\ \forall (s,s') \in S^{\otimes},\ \mathbf{P}'(s',s'_1) = \sum_{s_1 \in S} \mathbf{P}^{\otimes}((s,s'),(s_1,s'_1))$
- $\forall (s,s') \in S^{\otimes}\ s' = s'_+ \Rightarrow s = s_+$

*The set $S^{\otimes}$ defines a coupling relation between the two chains.*

The following proposition allows to compare probabilities without any numerical computation. As before, $\mu(s)$ (resp. $\mu'(s')$) denotes the probability to reach the state $s_+$ (resp. $s'_+$) in $\mathcal{C}$ (resp. in $\mathcal{C}'$) starting from $s$ (resp. from $s'$).

**Proposition 3.** *Let $\mathcal{C}^{\otimes}$ be a coupling between $\mathcal{C}$ and $\mathcal{C}'$. Then, for all $(s,s') \in S^{\otimes}$, we have:*

$$\mu(s) \geq \mu'(s')$$

**Example 1.** *Let us illustrate coupling for the Markov chain represented in figure 2 and called $\mathcal{C}^{\bullet}$. This chain is obtained from the tandem queues by lumping together states which have the same number of clients and at least $R$ clients in the second queue (in the figure $R = 2$). Its set of state is $S^{\bullet} = [0..N] \times [0..R]$. Here there is a coupling of this chain with itself defined by $S^{\otimes} = \{((n_1,n_2),(n'_1,n'_2)) \mid n_1 + n_2 \geq n'_1 + n'_2 \wedge n_1 \geq n'_1\}$.*

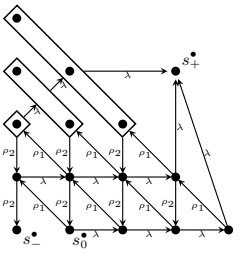



**Fig. 2.** Reduced DTMC

***Lemma 1.*** *$S^{\otimes}$ is a coupling relation.*

*Thus:* $\forall ((n_1,n_2),(n'_1,n'_2)) \in S^{\otimes},\ \mu^{\bullet}(n_1,n_2) \geq \mu^{\bullet}(n'_1,n'_2)$

### 3.2 An Importance Sampling Method with Variance Reduction and Confidence Interval

The proposed method combines statistical model checking on the original chain preceded by numerical model checking on a reduced chain whose formal definition is given below.

**Definition 4.** *Let $\mathcal{C}$ be a DTMC, a DTMC $\mathcal{C}^{\bullet}$ is called a* reduction *of $\mathcal{C}$ by a function $f$ that maps $S$ to $S^{\bullet}$, the state space of $\mathcal{C}^{\bullet}$, if, denoting $s^{\bullet}_- = f(s_-)$ and $s^{\bullet}_+ = f(s_+)$, the following assertions are satisfied:*

- *$f^{-1}(s^{\bullet}_-) = \{s_-\}$ and $f^{-1}(s^{\bullet}_+) = \{s_+\}$.*
- *$s^{\bullet}_-$ and $s^{\bullet}_+$ are absorbing states reached with probability 1.*
- *Let $s^{\bullet} \in S^{\bullet}$ and denote by $\mu^{\bullet}(s^{\bullet})$, the probability to reach $s^{\bullet}_+$ starting from $s^{\bullet}$. Then for all $s \in S$, we have $\mu^{\bullet}(f(s)) = 0 \Rightarrow \mu(s) = 0$.*

The two first assertions entail that the reduced chain has two absorbing states reached with probability 1 which are images of the absorbing states of the original chain. The last assertion requires that when from the image of some state $s$, one cannot reach $s^{\bullet}_+$, then one cannot reach $s_+$ from $s$. These (weak) assumptions

ensure that the mapping $f$ preserves the basic features of the original chain. Two states $s$ and $s'$ are *equivalent* if $f(s) = f(s')$, in other words $f^{-1}$ define equivalence classes for this reduction.

**Example 1.** *In the example of tandem queues, the reduced chain $\mathcal{C}^\bullet$ is obtained from the original chain by applying the following function to the state space.*

$$f(n_1, n_2) = \begin{cases} (n_1, n_2) & \text{if } n_2 \leq R \\ (n_1 + n_2 - R, R) & \text{otherwise} \end{cases}$$

*The intuition behind this reduction is to block clients in the first queue when there are $R$ clients in the second one, thus increasing the probability of a global overflow.* Given some reduced chain $\mathcal{C}^\bullet$, our goal is to replace the random variable (r.v.) $V_{s_0}$ which takes value in $\{0, 1\}$ by a r.v. $W_{s_0}$ which takes value in $\{0, \mu^\bullet(f(s_0))\}$. This requires that $\mu(s_0) \leq \mu^\bullet(f(s_0))$. By applying an homogeneity principle, we get the stronger requirement $\forall s \in S, \mu(s) \leq \mu^\bullet(f(s))$. In fact, the appropriate requirement which implies the previous one (see later proposition 4) is expressed by the next definition.

**Definition 5.** *Let $\mathcal{C}$ be a DTMC and $\mathcal{C}^\bullet$ a reduction of $\mathcal{C}$ by $f$. $\mathcal{C}^\bullet$ is a* reduction with guaranteed variance *if for all $s \in S$ such that $\mu^\bullet(f(s)) > 0$ we have :*

$$\sum_{s' \in S} \mu^\bullet(f(s')) \cdot \mathbf{P}(s, s') \leq \mu^\bullet(f(s)) \tag{2}$$

Given $s \in S$, let $h(s)$ be defined by $h(s) = \sum_{s' \in S} \frac{\mu^\bullet(f(s'))}{\mu^\bullet(f(s))} \mathbf{P}(s, s')$. We can now construct an efficient important sampling based on a reduced chain with guaranteed variance.

**Definition 6.** *Let $\mathcal{C}$ be a DTMC and $\mathcal{C}^\bullet$ be a reduction of $\mathcal{C}$ by $f$ with guaranteed variance. Then $\mathbf{P}'$ is transition matrix on $S$ defined by:*
*Let $s$ be a state of $S$,*

- *if $\mu^\bullet(f(s)) = 0$ then for all $s' \in S$, $\mathbf{P}'(s, s') = \mathbf{P}(s, s')$*
- *if $\mu^\bullet(f(s)) > 0$ then for all $s' \in S \setminus \{s_-\}$,*
  $\mathbf{P}'(s, s') = \frac{\mu^\bullet(f(s'))}{\mu^\bullet(f(s))} \mathbf{P}(s, s')$ *and* $\mathbf{P}'(s, s_-) = 1 - h(s)$.

The following proposition justifies the definition of $\mathbf{P}'$.

**Proposition 4.** *Let $\mathcal{C}$ be a DTMC and $\mathcal{C}^\bullet$ be a reduction with guaranteed variance. The importance sampling based on matrix $\mathbf{P}'$ of definition 6 has the following properties:*

- *For all $s$ such that $\mu(s) > 0$,*
  *$W_s$ is a random variable which has value in $\{0, \mu^\bullet(f(s))\}$.*
- *$\mu(s) \leq \mu^\bullet(f(s))$ and $\mathbf{V}(W_s) = \mu(s)\mu^\bullet(f(s)) - \mu^2(s)$.*
- *One can compute a confidence interval for this importance sampling.*

Since $\mu(s_0) \ll 1$, $\mathbf{V}(V_{s_0}) \approx \mu(s_0)$. If $\mu(s_0) \ll \mu^\bullet(f(s_0))$, we obtain $\mathbf{V}(W_{s_0}) \approx \mu(s_0)\mu^\bullet(f(s_0))$, so the variance is reduced by a factor $\mu^\bullet(f(s_0))$. In the case where $\mu(s_0)$ and $\mu^\bullet(f(s_0))$ have same magnitude order, the reduction of variance is even bigger.

Unfortunately, Equation (2) requires to compute the function $\mu^\bullet$ in order to check that $\mathcal{C}^\bullet$ is a reduction with guaranteed variance. We are looking for a structural requirement that does not involve the computation of $\mu^\bullet$.

**Proposition 5.** *Let $\mathcal{C}$ be a DTMC, $\mathcal{C}^\bullet$ be a reduction of $\mathcal{C}$ by $f$. Assume there exists a family of functions $(g_s)_{s\in S}$, $g_s : \{t \mid \mathbf{P}(s,t) > 0\} \to S^\bullet$ such that:*

1. $\forall s \in S,\ \forall t^\bullet \in S^\bullet,\ \mathbf{P}^\bullet(f(s), t^\bullet) = \sum_{s' \mid g(s') = t^\bullet} \mathbf{P}(s, s')$
2. $\forall s, t \in S$ *such that* $\mathbf{P}(s,t) > 0,\ \mu^\bullet(f(t)) \leq \mu^\bullet(g_s(t))$

*Then $\mathcal{C}^\bullet$ is a reduction of $\mathcal{C}$ with guaranteed variance.*

The family of functions $(g_s)$ assigns to each transition of $\mathcal{C}$ starting from $s$ a transition of $\mathcal{C}^\bullet$ starting from $f(s)$. The first condition can be checked by straightforward examination of the probability transition matrices. The second condition still involves the mapping $\mu^\bullet$ but here there are only comparisons between its values. Thanks to proposition 3, it can be proved by exhibiting a coupling of $\mathcal{C}$ with itself.

We are now in position to describe the whole method for a model $\mathcal{M}$ with associated DTMC $\mathcal{C}$.

1. Specify a model $\mathcal{M}^\bullet$ with associated DTMC $\mathcal{C}^\bullet$, a function $f$ and a family of functions $(g_s)_{s\in S}$. The specification of this family is done at the level of models $\mathcal{M}$ and $\mathcal{M}^\bullet$ as shown in the next example and in section 4.
2. Prove using a coupling on $\mathcal{C}^\bullet$ that proposition 5 holds. Again the proof is performed at the level of models.
3. Compute function $\mu^\bullet$ with a numerical model checker applied on $\mathcal{M}^\bullet$.
4. Compute $\mu(s_0)$ with a statistical model checker applied on $\mathcal{M}$ using the importance sampling of definition 6.

The last two steps are done by tools. The second step is currently done by hand (see [6]) but could be handled by theorem provers. The only manual step is the specification of $\mathcal{M}^\bullet$ which requires to study $\mathcal{M}$ and the formula to be checked (see section 4).

**Example 1.** *To apply the method on the example it remains to specify the family of functions $(g_s)_{s\in S}$.*

$$
\begin{aligned}
g_{(n_1,n_2)}(n_1, n_2) &= f(n_1, n_2) \\
g_{(n_1,n_2)}(n_1 + 1, n_2) &= f(n_1 + 1, n_2) \\
g_{(n_1,n_2)}(n_1 - 1, n_2 + 1) &= f(n_1 - 1, n_2 + 1) \\
g_{(n_1,n_2)}(n_1, n_2 - 1) &= \begin{cases} (n_1, n_2 - 1) & \text{if } n_2 \leq R \\ (n_1 + n_2 - R, R - 1) & \text{otherwise} \end{cases}
\end{aligned}
$$

*The condition 2 always trivially holds except for the last case with $n_2 > R$. We have to check that $\mu^\bullet(n_1 + n_2 - 1 - R, R) \leq \mu^\bullet(n_1 + n_2 - R, R - 1)$. As $(n_1 + n_2 - R, R - 1), (n_1 + n_2 - 1 - R, R))$ belongs to the coupling relation the inequality holds.*

### 3.3 Generalisation

We generalize the method but with no guarantee about the variance reduction.

**Definition 7.** *Let $\mathcal{C}$ be a DTMC and $\mathcal{C}^\bullet$ a reduction $\mathcal{C}$ of by $f$. We define a transition matrix $\mathbf{P}'$ on $S$ by the following rules. Let $s \in S$:*

- *if $\mu^\bullet(f(s)) = 0$ then for all $s' \in S$, $\mathbf{P}'(s,s') = \mathbf{P}(s,s')$*
- *if $\mu^\bullet(f(s)) > 0$ and $h(s) \leq 1$ then for all $s' \in S \setminus \{s_-\}$,*
  $\mathbf{P}'(s,s') = \frac{\mu^\bullet(f(s'))}{\mu^\bullet(f(s))}\mathbf{P}(s,s')$ *and* $\mathbf{P}'(s,s_-) = 1 - h(s)$
- *if $h(s) > 1$, then for all $s' \in S$, $\mathbf{P}'(s,s') = \frac{\mu^\bullet(f(s'))}{h(s)\mu^\bullet(f(s))}\mathbf{P}(s,s')$*

When Equation 2 does not hold for state $s$, we have to "normalize" the matrix row $\mathbf{P}'(s,-)$. The next proposition characterises the range of the random variable $W_s$ for this importance sampling. Thus the precision of the estimator highly depends on the shape of the (unknown) distribution of $W_s$ beyond $\mu^\bullet(f(s))$.

**Proposition 6.** *Let $\mathcal{C}$ be a DTMC and $\mathcal{C}^\bullet$ his reduction. The importance sampling of the definition 7 has the following property: for all $s$ such that $\mu(s) > 0$, $W_s$ is a random variable which takes its values in $\{0\} \cup [\mu^\bullet(f(s)), \infty[$.*

## 4 Experimentation

### 4.1 Implementation

**Tools.** Our experiments [4] have been performed on a modified version of COSMOS (downloadable at `http://www.lsv.ens-cachan.fr/Software/Cosmos`). COSMOS is a statistical model checker whose input model is a stochastic Petri net with general distributions and formulas are expressed by the Hybrid Automata Stochastic Logic [4]. The numerical model checking of the reduced model have been performed by PRISM whereas we have also used the statistical model checker PRISM 4.0 for comparisons with our method.

**Adaptation of COSMOS.** Since COSMOS takes as input a stochastic Petri net with a continuous time semantic, we have adapted our method to work with continuous time Markov chains. As discussed before, for formulas that we consider, this does not present serious difficulty.

The importance sampling increases the computation time of simulation. First we have to compute and store in an hash table the probability vector $\mu^\bullet$ of the reduced model in polynomial time w.r.t. the reduced Markov chain $\mathcal{C}^\bullet$. Then after a transition of the path we must compute $\mathbf{P}'(s,-)$ where $s$ is the current state whose computation time is linear w.r.t. the number of events of $\mathcal{M}$.

[4] All the experiments have been performed on a computer with a 2.6Ghz processor and 48Go of memory without parallelism.

### 4.2 Example 1: Global Overflow in Tandem Queues

This example is a classical benchmark for importance sampling. We compare our results with those of [12] which provides an efficient solution (see section 2). We take the same parameters with $\lambda = 0.1$, $\rho_1 = \rho_2 = 0.45$, $N = 50$ and we also simulate 20000 paths. We set the parameter $R$ to 4. The probability (computed by a numerical model checker) is $3.8\ 10^{-31}$. The width of confidence interval produced by [12] is $6.4\ 10^{-32}$ whereas ours is six times smaller ($9.63\ 10^{-33}$).

We also compare our method to both numerical and standard statistical model checking done by PRISM. We change the parameter of the model to $\lambda = 0.32$, $\rho_1 = \rho_2 = 0.34$ in order to evaluate the methods for large values of $N$. Results are depicted in table 1. We set the value of $R$ to the minimal value such that $\frac{\mu(s_0)}{\mu^\bullet(f(s_0))} < 1.5$. We found that $R$ and $N$ satisfy $R \approx 36.3 \log(N) - 126$. The narrowness of the obtained confidence interval confirms the validity of this choice. The reduced model has $\Theta(n \log(n))$ states whereas the initial one has $\Theta(n^2)$ states.

The standard statistical model checker fails to find a relevant confidence interval (i.e. the width of interval is half the value of the estimation) when $N \geq 100$ while the numerical model checker does not end when $N \geq 5000$. Our method can handle greater values of this parameter. We can approximate the number of paths required by standard statistical model checking and deduce the estimated corresponding computation time which is $10^{120}$ greater than ours!

**Table 1.** Experimental results for example 1

| N | Size of | Prism num | | Prism stat | | | Cosmos | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mathcal{C}$ | T (s) | $\mu(s_0)$ | T (s) | $\mu(s_0)$ | Conf. Int. width | R | T $\mathcal{C}^\bullet$ (s) | T (s) | $\mu(s_0)$ | Conf. Int. width |
| 50 | 2601 | 0.3 | 0.0929 | 1.45 | 0.091 | 0.016 | 4 | 0.03 | 7 | 0.090 | 0.017 |
| 100 | 10 201 | 1.6 | 0.01177 | 2.7 | 0.015 | 0.007 | 30 | 1 | 36 | 0.01156 | 8.6E-4 |
| 500 | 251 001 | 126 | 2.06E-12 | 2.3 | 0 | # | 87 | 23 | 145 | 2.075E-12 | 1.72E-13 |
| 1000 | 1E6 | 860 | 2.87E-25 | No path reaches the rare event | | | 111 | 113 | 263 | 2.906E-25 | 2.52E-26 |
| 5000 | 25E6 | >12h | # | # | | | 150 | 3061 | 1099 | 7.10E-130 | 1.21E-130 |

### 4.3 Example 2 : Parallel Random Walk

The Petri net depicted in figure 3 models a parallel random walk of $N$ walkers. A walk is done between position 1 and position $L$ starting in position $L/2$ and ends up in the extremal positions. At every round, some random walker can randomly go in either direction. However when walkers $i$ and $i+1$ are in the same position, walker $i$ can only go toward 1. We represent on this figure the walker $i$ and his interactions with walker $i+1$. Transition $A_{i,j}$ (resp. $R_{i,j-1}$) corresponds to a step toward $L$ (resp. 1).

This model is a paradigm of failure tolerant systems in which each walker represents a process which finishes its job when it reaches position 1. Failures can occur and move the process away of its goal. When position $L$ is reached the job is aborted. We want to evaluate the probability that a majority of players have reached position $L$.

This model has $L^N$ states. In order to get a reduced model, we remove all synchronisation between walkers. Behaviours of all walkers are now independent and thus a state of the reduced system is now defined by the number of walkers in each position. The size of the reduced system is $\binom{N+L-1}{L-1}$.

Proposition 5 holds for this reduced model. Intuitively, removing synchronisation between walkers increases the probability to reach position $L$.

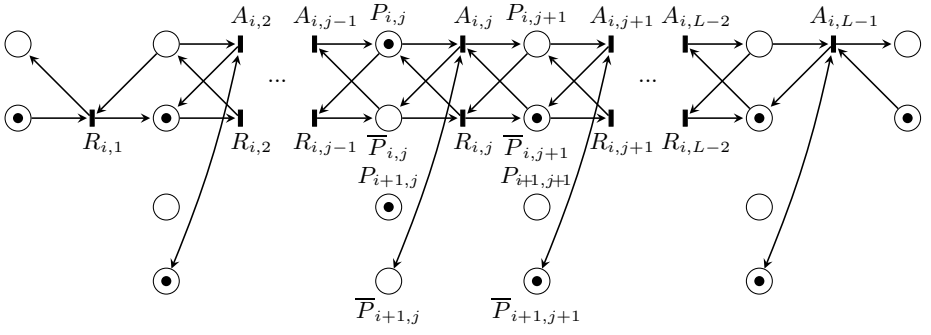



**Fig. 3.** The Petri net for example 2

Table 2 shows the experimental result with the following parameters $p = 0.3$ , $q = 0.7$, $L = 15$. We stop the simulation when the confidence interval width reaches one tenth of the estimated value. Our method handles huge models (with size up to 8 $10^{12}$) with very small probabilities (8 $10^{-18}$) whereas the standard statistical model checking and numerical model checkings fail due to either the low probability or the size of the system.

**Table 2.** Experimental results for example 2

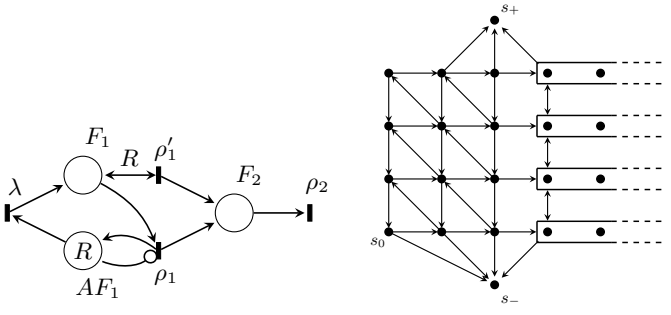
| N | size of | Prism num | | Prism stat | | | Cosmos | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mathcal{C}$ | T (s) | $\mu(s_0)$ | T (s) | $\mu(s_0)$ | Conf. Int. | Nb Traj. | T $\mathcal{C}^\bullet$ (s) | T (s) | $\mu(s_0)$ | Conf. Int. |
| 1 | 15 | $\approx 0$ | 0.00113 | 12 | 1.15E-3 | 1E-4 | 1 | $\approx$0 | $\approx$0 | 0.00113 | 0 |
| 5 | 7.5E5 | 6 | 1.88E-9 | 21 | 0 | # | 18000 | 0.5 | 13 | 1.94E-9 | 1.89E-10 |
| 6 | 1.1E7 | 127 | 1.14E-12 | No path reaches the rare event | | | 53000 | 1 | 57 | 1.17E-12 | 1.17E-13 |
| 7 | 1.7E8 | 2248 | 2.93E-12 | # | | | 50000 | 2.8 | 186 | 2.92E-12 | 2.89E-13 |
| 8 | 2.0E9 | Out of memory | | # | | | 145000 | 7.9 | 1719 | 1.86E-15 | 1.86E-16 |
| 9 | 3.8E10 | # | | # | | | 128000 | 24 | 3800 | 4.7E-15 | 4.75E-16 |
| 10 | 5.7E11 | # | | # | | | 371000 | 71 | 26000 | 3.12E-18 | 3.11E-19 |
| 11 | 8.0E12 | # | | # | | | 321000 | 228 | 67000 | 7.90E-18 | 7.89E-19 |

### 4.4 Example 3: Local Overflow in Tandem Queues

We consider the tandem queues system with a different property to check: `The second queue contains` $N$ `clients` ($n_2 = N$) `before the second queue is empty`($n_2 = 0$). The state space is $S = \mathbb{N} \times [0..N]$ with initial state $(0, 1)$. Contrary to the first example $\mathcal{C}$ is now infinite but $\mathcal{C}^\bullet$ must be finite in order to apply the numerical model checker.

The reduced model behaves as the original one until the first queue contains $R$ clients. Then the model behaves as if there is an infinite number of clients in the first queue. The corresponding Petri net is depicted in figure 4(a). The corresponding reduction function $f$ (whose effect on the original chain is represented in figure 4(b)) is defined by:

$$f(n_1, n_2) = \begin{cases} (n_1, n_2) & \text{if } n_1 \leq R \\ (R, n_2) & \text{otherwise} \end{cases}$$




(a) Reduced model (b) Reduction chain

**Fig. 4.** Petri net for example 3($R = 3, N = 5$)

**Table 3.** Experimental results for example 3

| N | R | T (s) | Size of | $\mu^\bullet(f(s_0))$ | Cosmos | | | | Prism stat | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\mathcal{C}^\bullet$ | $\mathcal{C}^\bullet$ | | $\mu(s_0)$ | Conf. Int. | T (s) | Nb Traj. | T (s) | $\mu(s_0)$ | Conf. Int. |
| 25 | 12 | $\approx 0$ | 338 | 1.16E-5 | 1.48E-6 | 2.83E-7 | 2 | 5000 | 33 | 1.1E-6 | 1.6E-6 |
| 50 | 29 | $\approx 0$ | 1530 | 2.98E-10 | 3.81E-11 | 7.19E-12 | 13 | 5000 | No path reaches the rare event | | |
| 100 | 66 | 1.44 | 6767 | 1.87E-19 | 4.22E-20 | 7.34E-21 | 17 | 3000 | # | | |
| 500 | 370 | 1770 | 185871 | 1.03E-90 | 6.63E-91 | 8.05E-32 | 37 | 2000 | # | | |
| 1000 | 740 | 24670 | 741741 | 3.24E-177 | 3.95E-179 | 4.00E-179 | 180 | 3000 | # | | |

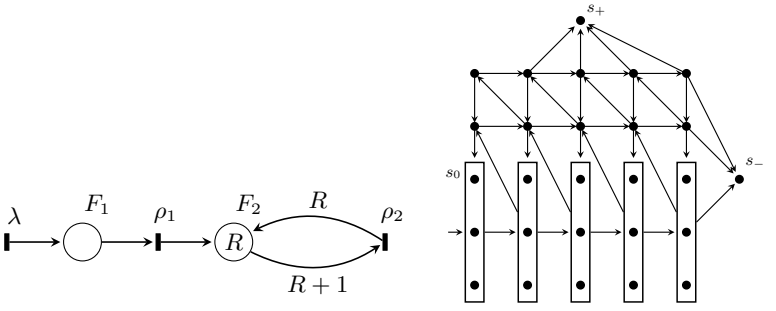
We found by running experiments on small values of $N$ and $R$ that for $R \geq 0.74 \times N$ we have $\mu(s_0) \geq \mu^\bullet(f(s_0))/10$. This example shows that we can apply our method on an infinite model subject to the specification of a finite reduced model. Observe that computation time reductions w.r.t. standard statistical model checking are still impressive.

### 4.5 Example 4: Bottleneck in Tandem Queues

We consider the tandem queues system with a different property to check: `The second queue is full` ($n_2 = N$) `before the first one` ($n_1 = N$). The reduced model is obtained by considering that the second queue is full when it contains $N - R$ clients or in an equivalent way that the second queue always contains at least $R$ clients. The corresponding Petri net is depicted in figure 5(a).

The reduction function (whose effect on the original chain is represented in figure 5(b)) is defined by:

$$f(n_1, n_2) = \begin{cases} (n_1, R) & \text{if } n_2 \leq R \\ (n_1, n_2) & \text{otherwise} \end{cases}$$




(a) Reduced model (b) Reduction function

**Fig. 5.** Petri net for the tandem queues ($R = 2, N = 5$)

However, the experimental results are not satisfactory since $\mu(s_0) \ll \mu^\bullet(f(s_0))$ when $R$ is small compared to $N$. This shows that designing a reduced model with relevant computation time reduction is sometimes tricky (and remains to be done for this example).

## 5 Conclusion

We proposed a method of statistical model checking which computes a reduced confidence interval for the probability of a rare event. Our method is based on importance sampling techniques. Other methods usually rely on heuristics and fail to provide a confidence interval. We have developed a theoretical framework ensuring the reduction of the variance and providing a confidence interval. This framework requires a structural analysis of the model but no numerical computation thanks to coupling theory. Our method is implemented in the statistical model checker Cosmos and we have done experiments with impressive results.

We plan to go further in four directions. First we want to deal with more complex infinite systems. Secondly we want to handle "bounded until" formulas requiring to deal with non Markovian systems. We also would mechanize the proofs of coupling since they consist to check parametrized inequalities. Finally we are looking for a class of models which structurally fulfill the required assumptions.

## References

1. Amparore, E.G., Donatelli, S.: Model checking CSL$^{\text{TA}}$ with deterministic and stochastic Petri nets. In: DSN, pp. 605–614 (2010)

2. Baier, C., Haverkort, B.R., Hermanns, H., Katoen, J.-P.: Model-checking algorithms for continuous-time Markov chains. IEEE Trans. Software Eng. 29(6), 524–541 (2003)
3. Bain, L.J., Engelhardt, M.: Introduction to Probability and Mathematical Statistics, 2nd edn. Duxbury Classic Series (1991)
4. Ballarini, P., Djafri, H., Duflot, M., Haddad, S., Pekergin, N.: HASL: An expressive language for statistical verification of stochastic models. In: VALUETOOLS 2011, Cachan, France (May 2011) (to appear)
5. Barbot, B., Haddad, S., Picaronny, C.: Échantillonnage préférentiel pour le model checking statistique. In: MSR 2011. Journal Européen des Systèmes Automatisés, vol. 45, pp. 237–252 (2011)
6. Barbot, B., Haddad, S., Picaronny, C.: Coupling and importance sampling for statistical model checking. Research Report LSV-12-01, Laboratoire Spécification et Vérification. ENS Cachan, France (January 2012)
7. Bengtsson, J., Larsen, K.G., Larsson, F., Pettersson, P., Yi, W.: UPPAAL - a tool suite for automatic verification of real-time systems. In: Hybrid Systems, pp. 232–243 (1995)
8. Chiola, G., Franceschinis, G., Gaeta, R., Ribaudo, M.: GreatSPN 1.7: Graphical editor and analyzer for timed and stochastic Petri nets. Perform. Eval. 24(1-2), 47–68 (1995)
9. Ciesinski, F., Baier, C.: LiQuor: A tool for qualitative and quantitative linear time analysis of reactive systems. In: QEST 2006, pp. 131–132 (2006)
10. Clarke, E.M., Zuliani, P.: Statistical Model Checking for Cyber-Physical Systems. In: Bultan, T., Hsiung, P.-A. (eds.) ATVA 2011. LNCS, vol. 6996, pp. 1–12. Springer, Heidelberg (2011)
11. de Boer, P.-T.: Analysis of state-independent importance-sampling measures for the two-node tandem queue. ACM Trans. Model. Comput. Simul. 16(3), 225–250 (2006)
12. Dupuis, P., Sezer, A.D., Wang, H.: Dynamic importance sampling for queueing networks. Annals of Applied Probability 17, 1306–1346 (2007)
13. Emerson, E.A., Clarke, E.M.: Characterizing Correctness Properties of Parallel Programs Using Fixpoints. In: de Bakker, J.W., van Leeuwen, J. (eds.) ICALP 1980. LNCS, vol. 85, pp. 169–181. Springer, Heidelberg (1980)
14. Glynn, P.W., Iglehart, D.L.: Importance sampling for stochastic simulations. Management Science (1989)
15. Heegaard, P.E., Sandmann, W.: Ant-based approach for determining the change of measure in importance sampling. In: Winter Simulation Conference, pp. 412–420 (2007)
16. Katoen, J.-P., Zapreev, I.S., Hahn, E.M., Hermanns, H., Jansen, D.N.: The ins and outs of the probabilistic model checker MRMC. In: International Conference on Quantitative Evaluation of Systems, pp. 167–176 (2009)
17. Kwiatkowska, M., Norman, G., Parker, D.: PRISM: Probabilistic Symbolic Model Checker. In: Field, T., Harrison, P.G., Bradley, J., Harder, U. (eds.) TOOLS 2002. LNCS, vol. 2324, pp. 113–140. Springer, Heidelberg (2002)
18. Kwiatkowska, M., Norman, G., Parker, D.: Stochastic Model Checking. In: Bernardo, M., Hillston, J. (eds.) SFM 2007. LNCS, vol. 4486, pp. 220–270. Springer, Heidelberg (2007)
19. L'Ecuyer, P., Demers, V., Tuffin, B.: Splitting for rare-event simulation. In: Winter Simulation Conference, pp. 137–148 (2006)

20. Legay, A., Delahaye, B., Bensalem, S.: Statistical Model Checking: An Overview. In: Barringer, H., Falcone, Y., Finkbeiner, B., Havelund, K., Lee, I., Pace, G., Roşu, G., Sokolsky, O., Tillmann, N. (eds.) RV 2010. LNCS, vol. 6418, pp. 122–135. Springer, Heidelberg (2010)
21. Lindvall, T.: Lectures on the coupling method. Dover (2002)
22. Rubino, G., Tuffin, B.: Rare Event Simulation using Monte Carlo Methods. Wiley (2009)
23. Sen, K., Viswanathan, M., Agha, G.: VESTA: A statistical model-checker and analyzer for probabilistic systems. In: QEST, pp. 251–252 (2005)
24. Srinivasan, R.: Importance sampling – Applications in communications and detection. Springer, Berlin (2002)
25. Younes, H.L.S.: Ymer: A Statistical Model Checker. In: Etessami, K., Rajamani, S.K. (eds.) CAV 2005. LNCS, vol. 3576, pp. 429–433. Springer, Heidelberg (2005)