Rare Event Handling in Statistical Model Checking

Benoît Barbot, Serge Haddad and Claudine Picaronny

LSV, ENS Cachan, CNRS & INRIA

Barbizon 2012

Plan

Rare Event

Critical systems

- Plane, rocket (failure of the fuel control system)
- Nuclear power plant (failure of all the redundant security systems)
- Security device like an airbag (delayed deployment)
- Telecommunication (overflow)
- Banking system (ruin of an insurance)
- Biology
- etc.

In common

- Consequences of failure are dramatic.
- The probability of failure is very small.

Rare Event

Critical systems

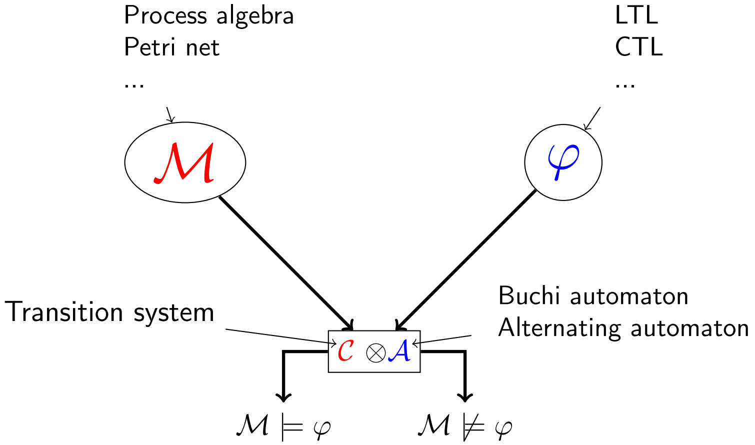
- Plane, rocket (failure of the fuel control system)
- Nuclear power plant (failure of all the redundant security systems)
- Security device like an airbag (delayed deployment)
- Telecommunication (overflow)
- Banking system (ruin of an insurance)
- Biology
- etc.

In common

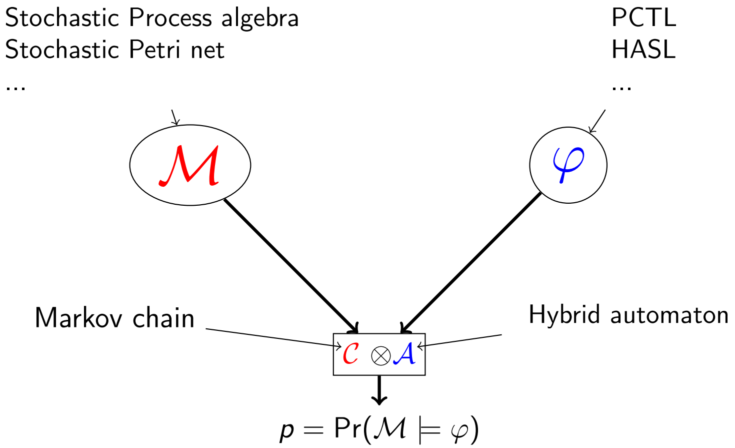
- Consequences of failure are dramatic.
- The probability of failure is very small.

Estimation of this probability is critical.

Model checking



Model checking for stochastic system



Numerical and Statistical Approaches

- Numerical Approach
  - Branching logic (based on CTL)
  - Exact value (but subject to numerical error)
  - Efficiently implemented in many tools
    (Prism, MRMC, GreatSPN)
  - Strong probabilistic hypotheses
  - Memory space
    proportional to the size of the stochastic process

Numerical and Statistical Approaches

- Numerical Approach
  - Branching logic (based on CTL)
  - Exact value (but subject to numerical error)
  - Efficiently implemented in many tools (PRISM, MRMC, GreatSPN)
  - Strong probabilistic hypotheses
  - Memory space proportional to the size of the stochastic process
- Statistical Approach
  - Linear Logic (based on LTL)
  - Confidence interval: probabilistic framing
  - Very small memory space
  - Easy to parallelize
  - Weak probabilistic hypothesis (only an operational semantic)
  - Unsuitable for rare events'probability

Objective: Develop a dedicated method for rare events.

Rare Event Problem

Illustration

- **Objective:** Estimation of the probability p of an event e with a confidence level of 0.99
- **Hypotheses:**
    1. Computation of 10^9 trajectories
    2. $p \leq 10^{-15}$

Rare Event Problem

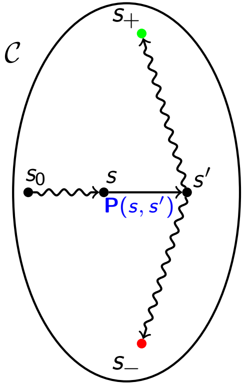
Illustration

- **Objective:** Estimation of the probability p of an event e with a confidence level of 0.99
- **Hypotheses:**
  1. Computation of 10^9 trajectories
  2. $p \leq 10^{-15}$

Possible outcomes

- With probability $\approx 1 - 10^{-6}$, e does not occur in any trajectory
  We obtain as confidence interval: $[0, 7\ 10^{-9}]$
  $\Rightarrow$ Confidence interval too large
- With probability smaller than 10^{-6}, e occurs in one trajectory
  We obtain as confidence interval: $[7\ 10^{-10}, 2\ 10^{-9}]$
  $\Rightarrow$ Value outside the confidence interval
- With a tiny probability, e occurs in more than one trajectory
  $\Rightarrow$ Value outside the confidence interval

Rare Event as a Reachability Problem



A Discrete Time Markov chain $\mathcal{C}$
Two absorbing states s_-, s_+
reached with probability 1

Let $\sigma = s \to s_1 \to s_2 \to \cdots \to s_\pm$
be a random trajectory in $\mathcal{C}$

$$V_s = \begin{cases} 1 \text{ if } \sigma \text{ ends in state } s_+ \\ 0 \text{ if } \sigma \text{ ends in state } s_- \end{cases}$$

Objective:
Estimate $\Pr(\sigma \text{ ends in state } s_+) = \mathbf{E}(V_{s_0})$
when $\mathbf{E}(V_{s_0}) \ll 1$

Difficulty:
$\mathbf{V}(V_{s_0})$ too big to have an accurate estimation

Importance Sampling

Principle: Substitute W_s to V_s with same expectation but reduced variance.

1. Substitute $\mathbf{P}'$ to $\mathbf{P}$ such that $\mathbf{P}(s,s') > 0 \Rightarrow \mathbf{P}'(s,s') > 0 \vee s = s_-$
2. For each trajectory $\sigma = s \to s_1 \to s_2 \cdots s_k \to s_\pm$
   We define

$$W_s = \begin{cases} \frac{\mathbf{P}(s,s_1)}{\mathbf{P}'(s,s_1)} \cdot \frac{\mathbf{P}(s_1,s_2)}{\mathbf{P}'(s_1,s_2)} \cdot \ldots \cdot \frac{\mathbf{P}(s_k,s_+)}{\mathbf{P}'(s_k,s_+)} & \text{if } \sigma \text{ ends in state } s_+ \\ 0 & \text{if } \sigma \text{ ends in state } s_- \end{cases}$$

3. Statistically estimate $\mathbf{E}(W_{s_0})$

Importance Sampling

Principle: Substitute W_s to V_s with same expectation but reduced variance.

1. Substitute $\mathbf{P}'$ to $\mathbf{P}$ such that $\mathbf{P}(s, s') > 0 \Rightarrow \mathbf{P}'(s, s') > 0 \vee s = s_-$
2. For each trajectory $\sigma = s \to s_1 \to s_2 \cdots s_k \to s_\pm$
   We define

$$W_s = \begin{cases} \frac{\mathbf{P}(s,s_1)}{\mathbf{P}'(s,s_1)} \cdot \frac{\mathbf{P}(s_1,s_2)}{\mathbf{P}'(s_1,s_2)} \cdot \ldots \cdot \frac{\mathbf{P}(s_k,s_+)}{\mathbf{P}'(s_k,s_+)} & \text{if } \sigma \text{ ends in state } s_+ \\ 0 & \text{if } \sigma \text{ ends in state } s_- \end{cases}$$

3. Statistically estimate $\mathbf{E}(W_{s_0})$

Expectation is unchanged

$$\forall s \in S,\ \mathbf{E}(W_s) = \mathbf{E}(V_s)$$

Objective: reduction of the variance

$$\mathbf{V}(W_{s_0}) \ll \mathbf{V}(V_{s_0})$$

Optimal Importance Sampling

A non effective result

There exists an importance sampling with variance equal to zero.

Let $\mu(s) = \mathbf{E}(V_s)$
Let $\mathbf{P}'(s,t) = \frac{\mu(t)}{\mu(s)} \cdot \mathbf{P}(s,t)$

$$W_s = \frac{\mathbf{P}(s,s_1)}{\mathbf{P}'(s,s_1)} \cdot \frac{\mathbf{P}(s_1,s_2)}{\mathbf{P}'(s_1,s_2)} \cdots \frac{\mathbf{P}(s_k,s_+)}{\mathbf{P}'(s_k,s_+)} = \frac{\mu(s)}{\mu(s_1)} \cdot \frac{\mu(s_1)}{\mu(s_2)} \cdots \frac{\mu(s_k)}{1} = \mu(s)$$

Problem: Need to know μ which is what one wants to compute.

An help to design good importance sampling.

State of the art

Asymptotically optimal importance sampling
(P.Dupuis, A.D. Sezer, H. Wang 2007)

Reduced to an optimization problem (Cross Entropy Method)
(E. Clarke, P. Zuliani 2011)
(C. Jegourel, A. Legay, S. Sedwards 2012)

Use of heuristic
(P.E Heegaard, W. Sandmann 2007)

Case by case analysis
(Rubino, Tuffin 2009)

State of the art

Asymptotically optimal importance sampling
(P.Dupuis, A.D. Sezer, H. Wang 2007)

Reduced to an optimization problem (Cross Entropy Method)
(E. Clarke, P. Zuliani 2011)
(C. Jegourel, A. Legay, S. Sedwards 2012)

Use of heuristic
(P.E Heegaard, W. Sandmann 2007)

Case by case analysis
(Rubino, Tuffin 2009)

Problems

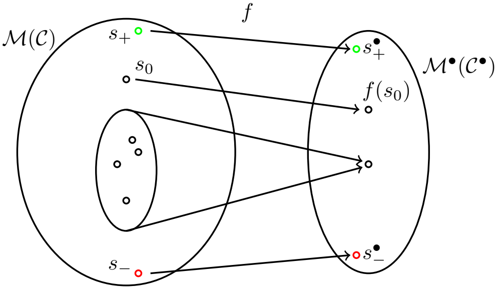
- None of these methods is fully automatic.
- None of these methods produces a true confidence interval.

Principle of efficient importance sampling

Design a reduced model $\mathcal{M}^\bullet$ of $\mathcal{M}$ and an abstraction function $f : S \to S^\bullet$.

Numerically compute $\mu^\bullet$.

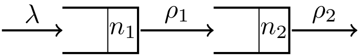
Substitute $\mu^\bullet$ to μ in the optimal importance sampling.



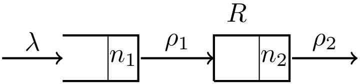
Exemple

Rare event: The are at least N clients between two idle periods.

From a tandem queues to a bounded capacity tandem queues ($R \ll N$).



$\mathcal{M}$



$\mathcal{M}^\bullet$

The clients in excess are moved back to the first queue.

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

How to guarantee variance reduction?

Goal: a modified Benoulli law for W_{s_0}

- $V_{s_0} \sim \mathcal{B}\text{ernoulli}(\{0, 1\}, \mu(s_0))$
- $W_{s_0} \sim \mathcal{B}\text{ernoulli}(\{0, \mu^{\bullet}(f(s_0))\}, \frac{\mu(s_0)}{\mu^{\bullet}(f(s_0))})$

How to guarantee variance reduction?

Goal: a modified Benoulli law for W_{s_0}

- $V_{s_0} \sim \mathcal{B}\text{ernoulli}(\{0,1\}, \mu(s_0))$
- $W_{s_0} \sim \mathcal{B}\text{ernoulli}(\{0, \mu^\bullet(f(s_0))\}, \frac{\mu(s_0)}{\mu^\bullet(f(s_0))})$

Theorem (necessary and sufficient condition)

$$\forall s \in S,\ \mu^\bullet(f(s)) \geq \sum_{s' \in S} \mathbf{P}(s,s') \cdot \mu^\bullet(f(s'))$$

Is a necessary and sufficient condition for W_{s_0} to follow a Bernoulli law.

How to guarantee variance reduction?

Goal: a modified Benoulli law for W_{s_0}

- $V_{s_0} \sim \mathcal{B}\text{ernoulli}(\{0,1\}, \mu(s_0))$
- $W_{s_0} \sim \mathcal{B}\text{ernoulli}(\{0, \mu^\bullet(f(s_0))\}, \frac{\mu(s_0)}{\mu^\bullet(f(s_0))})$

Theorem (necessary and sufficient condition)

$$\forall s \in S,\ \mu^\bullet(f(s)) \geq \sum_{s' \in S} \mathbf{P}(s,s') \cdot \mu^\bullet(f(s'))$$

Is a necessary and sufficient condition for W_{s_0} to follow a Bernoulli law.

Intuition: $\forall s \in S,\ \mu(s) = \sum_{s' \in S} \mathbf{P}(s,s') \cdot \mu(s)$

How to guarantee variance reduction?

Goal: a modified Benoulli law for W_{s_0}

- $V_{s_0} \sim \mathcal{B}\text{ernoulli}(\{0,1\}, \mu(s_0))$
- $W_{s_0} \sim \mathcal{B}\text{ernoulli}(\{0, \mu^\bullet(f(s_0))\}, \frac{\mu(s_0)}{\mu^\bullet(f(s_0))})$

Theorem (necessary and sufficient condition)

$$\forall s \in S,\ \mu^\bullet(f(s)) \geq \sum_{s' \in S} \mathbf{P}(s,s') \cdot \mu^\bullet(f(s'))$$

Is a necessary and sufficient condition for W_{s_0} to follow a Bernoulli law.

Intuition: $\forall s \in S,\ \mu(s) = \sum_{s' \in S} \mathbf{P}(s,s') \cdot \mu(s)$

Results

- Variance reduction is at least $\mu^\bullet(f(s_0))$.
- A true confidence interval can be computed.

How to check the property in a structural way?

Theorem

Assume there exists a family of functions $(g_s)_{s \in S}$*,*
$g_s : \{t \mid \mathbf{P}(s,t) > 0\} \rightarrow S^\bullet$ *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* $\mathbf{P}(s,t) > 0,\ \mu^\bullet(f(t)) \leq \mu^\bullet(g_s(t))$

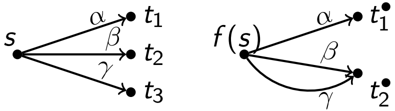
Then $\mathcal{C}^\bullet$ *is a reduction of* $\mathcal{C}$ *with guaranteed variance.*

Interest

- Condition 1 checked by examination of $\mathcal{M}$ and $\mathcal{M}^\bullet$.
- Condition 2 only involves comparison of items of $\mu^\bullet$.

Illustration of the local conditions

1. $\forall s \in S,\ \forall t^\bullet \in S^\bullet,\ \mathbf{P}^\bullet(f(s), t^\bullet) = \sum_{s' | g(s') = t^\bullet} \mathbf{P}(s, s')$

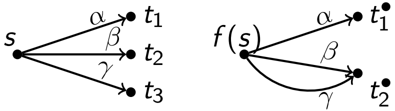


2. $\forall s, t \in S$ such that $\mathbf{P}(s, t) > 0,\ \mu^\bullet(f(t)) \leq \mu^\bullet(g_s(t))$

$$\mu^\bullet(t_1^\bullet) \leq \mu^\bullet(f(t_1))$$
$$\mu^\bullet(t_2^\bullet) \leq \mu^\bullet(f(t_2))$$
$$\mu^\bullet(t_2^\bullet) \leq \mu^\bullet(f(t_3))$$

Illustration of the local conditions

1. $\forall s \in S,\ \forall t^\bullet \in S^\bullet,\ \mathbf{P}^\bullet(f(s), t^\bullet) = \sum_{s' | g(s') = t^\bullet} \mathbf{P}(s, s')$



2. $\forall s, t \in S$ such that $\mathbf{P}(s,t) > 0,\ \mu^\bullet(f(t)) \leq \mu^\bullet(g_s(t))$

$$\mu^\bullet(t_1^\bullet) \leq \mu^\bullet(f(t_1))$$
$$\mu^\bullet(t_2^\bullet) \leq \mu^\bullet(f(t_2))$$
$$\mu^\bullet(t_2^\bullet) \leq \mu^\bullet(f(t_3))$$

A coupling theorem

Let $\mathcal{S}^\otimes$ be a coupling relation of $\mathcal{C}^\bullet$ whith itself by respect to $s_-^\bullet$ and $s_+^\bullet$, Then for all $(s, s') \in S^\otimes$, we have $\mu^\bullet(s) \geq \mu^\bullet(s')$.

Methodology with guaranteed variance reduction

1. Specify a reduced model $\mathcal{M}^\bullet$ with associated Markov chain $\mathcal{C}^\bullet$ and a function f.
2. Establish using analysis of $\mathcal{C}$ and $\mathcal{C}^\bullet$ and using a coupling $\mathcal{C}^\bullet$ that the reduction guarantees the variance reduction.
3. Compute numerically $\mu^\bullet$.
4. Compute statistically $\mu(s_0)$ using the importance sampling induced by $\mu^\bullet$.

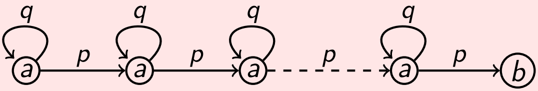
Handling Time Bounded Reachability

Time bounded reachability is strongly related to reactivity.

Difficulties

Observation 1

The rarity of an event can be triggered by the time bound.



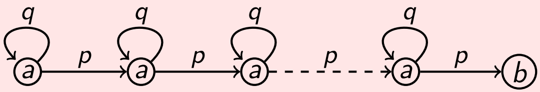
Handling Time Bounded Reachability

Time bounded reachability is strongly related to reactivity.

Difficulties

Observation 1

The rarity of an event can be triggered by the time bound.

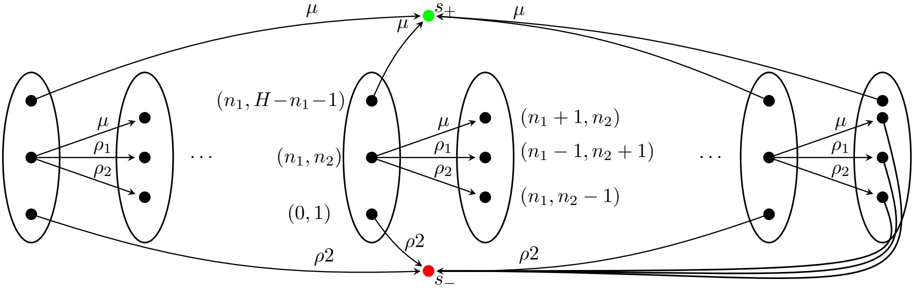


Observation 2

For finite horizon discrete and continuous time Markov chains behave differently.

From bounded reachability to unbounded reachability

$$S_u = S_{\bar{a}\bar{b}} \times [1, u] \cup \{s_-, s_+\}$$



Requires a stronger coupling theorem.

Principle of the method

Apply guaranteed importance sampling to $\mathcal{C}_u$

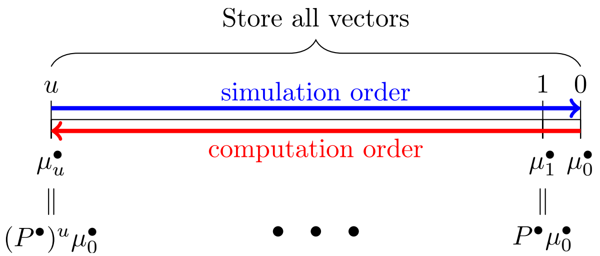
Let $\mu_v^\bullet$ be the time bounded reachability probability with horizon v.

$\mu_v^\bullet$ can be computed using equalities $\begin{cases} \mu_v^\bullet = \mathbf{P}^\bullet \cdot \mu_{v-1}^\bullet \\ \mu_0^\bullet(s_+) = 1 \\ \mu_0^\bullet(s) = 0 \ \forall s \neq s_+ \end{cases}$

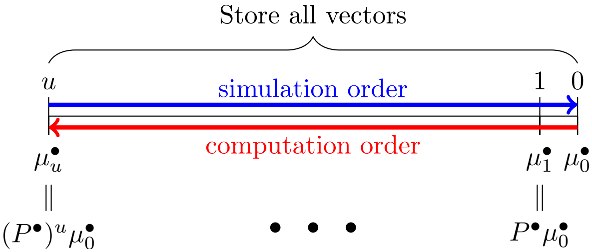
Problem

- $\mu_v^\bullet$ is computed by increasing values of v.
- During the simulation $\mu_v^\bullet$ are used by decreasing values of v.

Space consumption problem



Space consumption problem



Notations:

- m is the number of states of $\mathcal{C}^\bullet$.
- d is the maximal number of outgoing transitions of a state of $\mathcal{C}^\bullet$.

Complexity

- Time complexity: $\Theta(mdu)$
- Space complexity: $\Theta(mu)$

Comparison

Three algorithms

- The naive method
- Static and dynamic storage for $\mu_v^\bullet$
- Fully dynamic storagefor $\mu_v^\bullet$

Complexity	Algo 1	Algo 2	Algo 3
Space	$\Theta(mu)$	$\Theta(m\sqrt{u})$	$\Theta(m \log u)$
Time for the precomputation	$\Theta(mdu)$	$\Theta(mdu)$	$\Theta(mdu)$
Additional time for the simulation	0	$\Theta(mdu)$	$\Theta(mdu \log(u))$

Bounded reachability in CTMC

Uniformization

- Every CTMC is equivalent to a *uniform* CTMC, i.e. where all sojourn time is state are equal.
- Transient behavior of a uniform CTMC can be efficiently computed from the transient behavior of the associated DTMC.

Application to rare event handling

- Estimation of the time bounded reachability probabilities in the DTMC.
- Computation of the time bounded reachability probabilities in the CTMC via the uniformization formula..
- Elaborated tuning for the confidence interval.

Adaptation of Cosmos

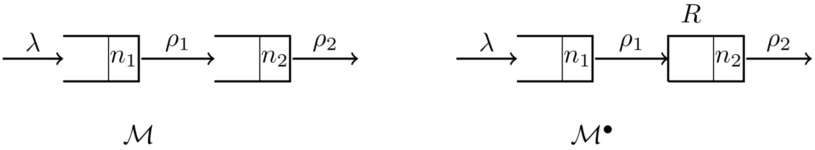
Modifications related to rare event

- Implementation of the importance sampling.
- Numerical computation of the transient behaviors.
- Implementation of the three algorithms.
- Implementation of the uniformization method.

General purpose improvements

- Parallelization of the simulation.
- Integration of Cosmos into the platform CosyVerif

An example



- Parameters: $\lambda = 0.1$, $\rho_1 = \rho_2 = 0.45$,
- Formula: They are at least N clients between two idle periods.
- Generation of 20000 trajectories
- Numerical result: $\mu(s_0) = 3.80122 \cdot 10^{-31}$

Example of the tandem ($N = 50$)

We perform experimentation with different values of R.

R	size of $\mathcal{C}$	size of $\mathcal{C}^\bullet$	$\mu^\bullet(s_0)$	$\mu(s_0)$ estimated	Confidence interval	T (s) simulation
2	2500	100	1.24904E-28	3.96541E-31	2.25E-31	21.47
3	2500	150	2.28771E-30	3.78565E-31	2.76E-32	39.48
4	2500	200	6.55440E-31	3.80168E-31	9.63E-33	57.32
5	2500	250	5.10457E-31	3.79642E-31	4.18E-33	64.81
6	2500	300	3.97544E-31	3.80229E-31	1.86E-33	67.18
7	2500	350	3.97544E-31	3.79973E-31	8.90E-34	68.56

$\mathcal{C}^\bullet$ is much smaller than $\mathcal{C}$.

Example of the tandem ($N = 50$)

We perform experimentation with different values of R.

R	size of $\mathcal{C}$	size of $\mathcal{C}^\bullet$	$\mu^\bullet(s_0)$	$\mu(s_0)$ estimated	Confidence interval	T (s) simulation
2	2500	100	1.24904E-28	3.96541E-31	2.25E-31	21.47
3	2500	150	2.28771E-30	3.78565E-31	2.76E-32	39.48
4	2500	200	6.55440E-31	3.80168E-31	9.63E-33	57.32
5	2500	250	5.10457E-31	3.79642E-31	4.18E-33	64.81
6	2500	300	3.97544E-31	3.80229E-31	1.86E-33	67.18
7	2500	350	3.97544E-31	3.79973E-31	8.90E-34	68.56

The estimated value is always close to the true value of $\mu(s_0)$.

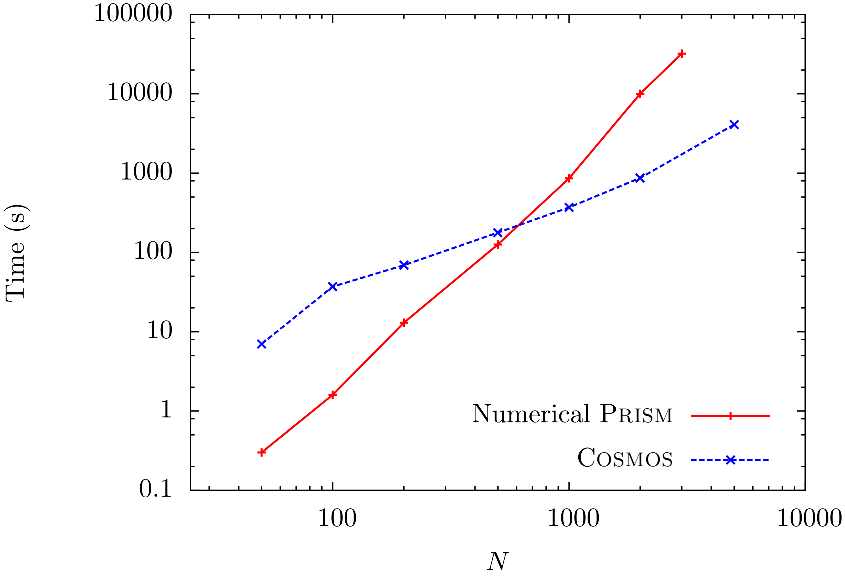
Example of the tandem ($N = 50$)

We perform experimentation with different values of R.

R	size of $\mathcal{C}$	size of $\mathcal{C}^\bullet$	$\mu^\bullet(s_0)$	$\mu(s_0)$ estimated	Confidence interval	T (s) simulation
2	2500	100	1.24904E-28	3.96541E-31	2.25E-31	21.47
3	2500	150	2.28771E-30	3.78565E-31	2.76E-32	39.48
4	2500	200	6.55440E-31	3.80168E-31	9.63E-33	57.32
5	2500	250	5.10457E-31	3.79642E-31	4.18E-33	64.81
6	2500	300	3.97544E-31	3.80229E-31	1.86E-33	67.18
7	2500	350	3.97544E-31	3.79973E-31	8.90E-34	68.56

The confidence interval is tight even for small R.

Example of the tandem with large values of N



Other examples

- Tandem (the second queue is full before the system is empty)
  - Infinite system (the first queue is unbounded)
  - Finite reduced system
- Tandem (the second queue is full before the first one)
  - Theoretical guarantee
  - Experimentally the acceleration is sufficient.
- Parallel ruin
  - Concurrent system
  - The reduced system is build by removing synchronization between process
- Dining philosopher problem
  - Extension of the method but no theoretical guarantee.
  - The distribution of W_{s_0} is heavy tailed.

Conclusion and Perspectives

- Contributions
  - Design of an importance sampling method
    with variance reduction and true confidence interval
  - Integration in a tool
  - Several conclusive case studies
- Perspectives
  - Handling more general infinite systems
  - Search of Petri net classes
    with automatic computation of the reduced model.
  - Automated or assisted proofs of coupling

Publications

- B. Barbot, S. Haddad and C. Picaronny.
  Échantillonnage préférentiel pour le model checking statistique.
  In MSR'11, Journal Européen des Systèmes Automatisés 45(1-3), pages 237-252. Hermès, 2011.
- B. Barbot, S. Haddad and C. Picaronny.
  Coupling and Importance Sampling for Statistical Model Checking.
  In TACAS'12, LNCS 7214, pages 331-346. Springer, 2012.
- B. Barbot, S. Haddad and C. Picaronny.
  Importance Sampling for Model Checking of Time-Bounded Until.
  Research Report LSV-12-04, February 2012. 14 pages.
- B. Barbot, S. Haddad and C. Picaronny.
  Importance Sampling for Model Checking of Continuous-Time Markov Chains.
  Research Report LSV-12-08, May 2012. 15 pages.
- Submission to "Discrete Event Dynamic Systems"