Probabilistic verification, approximation and metrics for bisimulation

Richard Lassaigne
Equipe de Logique,
CNRS-Université Paris 7

Joint work with Sylvain Peyronnet (LRDE/EPITA).

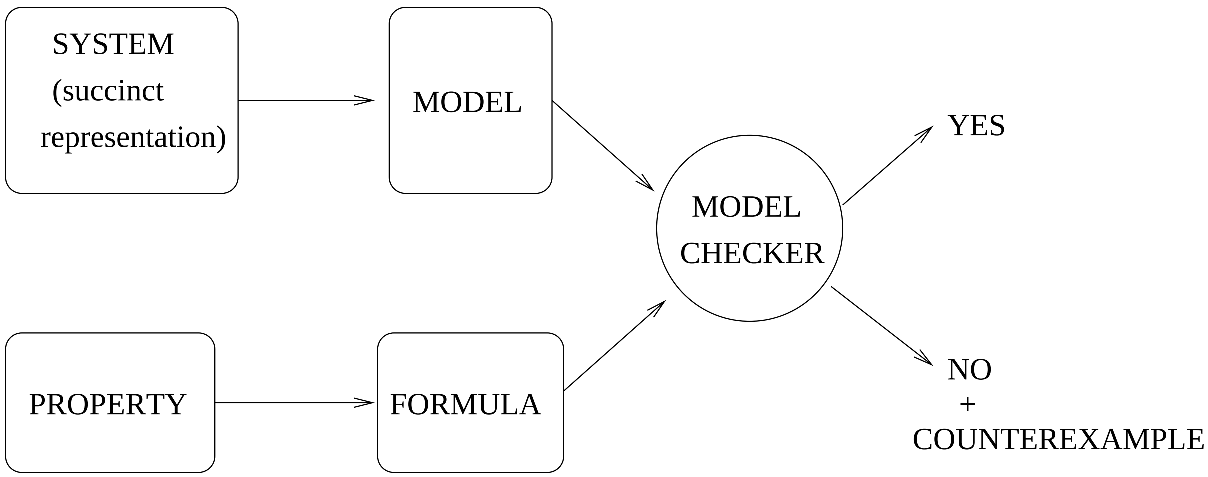
Probabilistic verification

Randomized approximation schemes

Approximate Probabilistic Model Checker

Probabilistic bisimulation

Metrics for labelled Markov Processes



Input :

- Model $\mathcal{M} = (S, R)$ $R \subseteq S^2$ (transition relation)
- Initial state s_0
- Formula φ

Output :

- YES if $(\mathcal{M}, s_0) \models \varphi$
- NO with a counterexample if $(\mathcal{M}, s_0) \not\models \varphi$

Complexity

$O(|M|.|\varphi|)$ (Branching Time Temporal Logic **CTL**)

ou

$O(|M|.2^{|\varphi|})$ (Linear Time Temporal Logic **LTL**)

Problem :

State space explosion phenomenon

(the problem is not the time but the space)

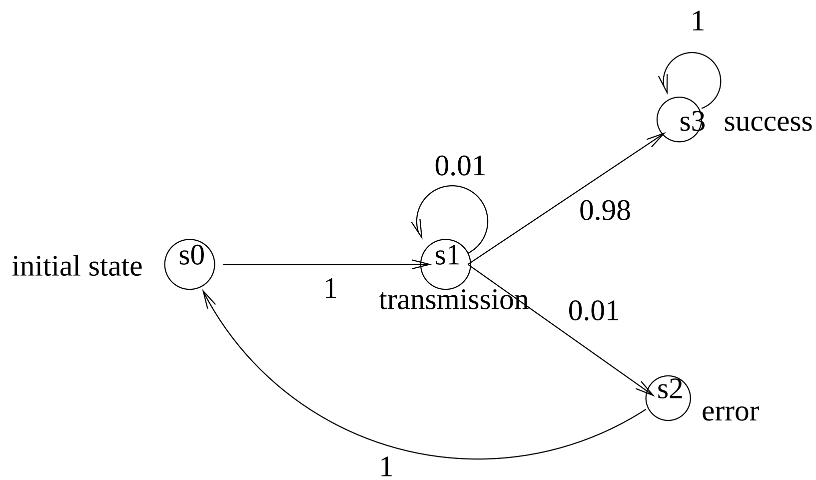
Classical methods :

- Symbolic representation (OBDD)
- SAT-based methods (Bounded moded checking)
- Abstraction

Probabilistic Transition Systems

Input :

- Model $\mathcal{M} = (S, \pi, L)$ and initial state s_0
- $\pi : S^2 \longrightarrow [0,1]$ Probability function
- $L : S \longrightarrow 2^{AP}$ (state labelling)
- Formula ψ (**LTL**)



Output : $Prob_{\Omega}[\psi]$

where (for example) $\psi \equiv transmission \mathbf{U} success$

(Ω probabilistic space of execution paths starting at s_0)

Probability space (and measure) :

Finite paths $\rho = (s_0, s_1, \ldots, s_n)$:

$Prob(\{\sigma/\sigma$ is a path and $(s_0, s_1, \ldots, s_n)$ is a prefix of $\sigma\}) =$

$$\prod_{i=1}^{n} P(s_{i-1}, s_i)$$

Measure extended to the Borel family of sets generated by the sets $\{\sigma/\rho$ is a prefix of $\sigma\}$ where ρ is a finite path.

The set of paths $\{\sigma/\sigma(0) = s$ and $\mathcal{M}, \sigma \models \psi\}$ is measurable (Vardi).

Complexity : (Coucourbetis and Yannakakis) [CY95]

Qualitative verification (i.e. prob=1 ?)

Same **complexity** as **LTL model checking**

$$O(|M|.2^{|\psi|})$$

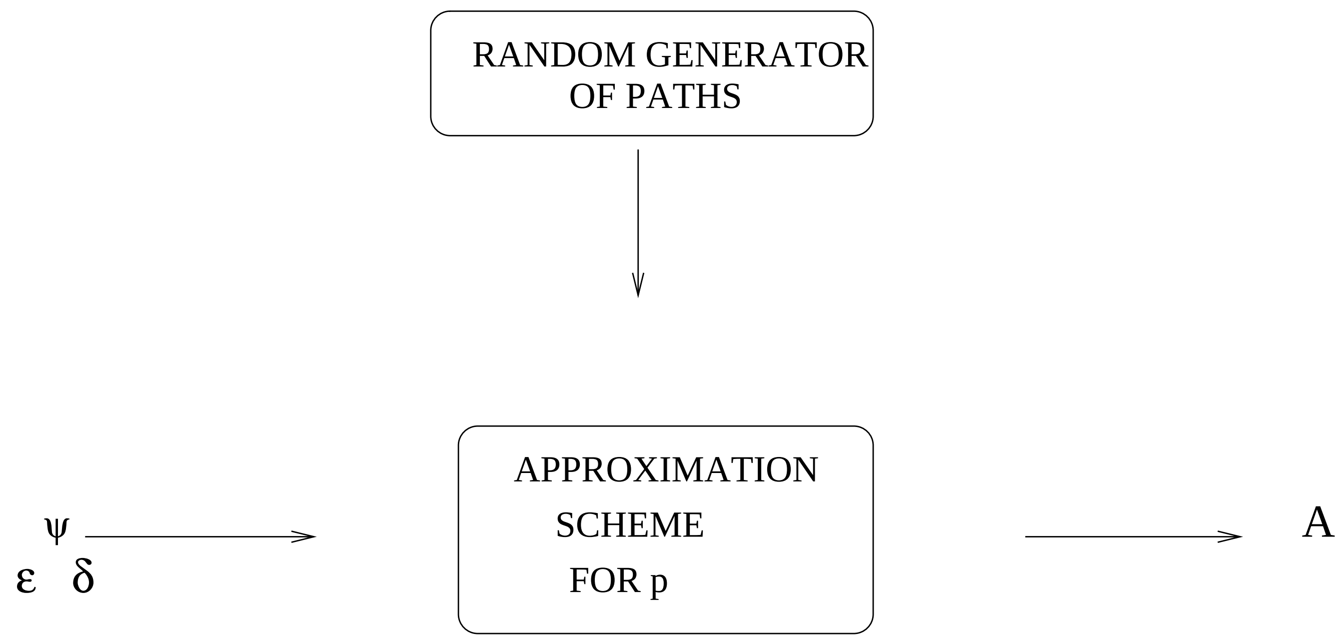
Quantitative verification (i.e. prob= ?)

$$O(|M|^3.2^{|\psi|})$$

Method : Computing $Prob_\Omega[\psi]$

- Transforming step by step the formula and the Markov chain $\mathcal{M}$
- Eliminating one by one the temporal connectives
- Preserving the satisfaction probability
- Solving system of linear equations of size $|M|$.

We want to approximate a probability p.



$$Pr[(p-\varepsilon) \leq A \leq (p+\varepsilon)] \geq 1-\delta$$

ε : *error parameter (additive approximation)*

δ : *confidence parameter (probabilistic algorithm)*

Can we efficiently approximate $Prob_{\Omega}(\psi)$?

FPRAS : (Karp, Luby and Madras)
Fully polynomial randomized approximation scheme with time complexity $poly(|\psi|, (1/\varepsilon), \log(1/\delta))$

General case : (Lassaigne and Peyronnet)

There is **no** probabilistic approximation algorithm with polynomial time complexity for computing $Prob_{\Omega}(\psi)$ $(\psi \in LTL)$ unless $BPP = NP$.

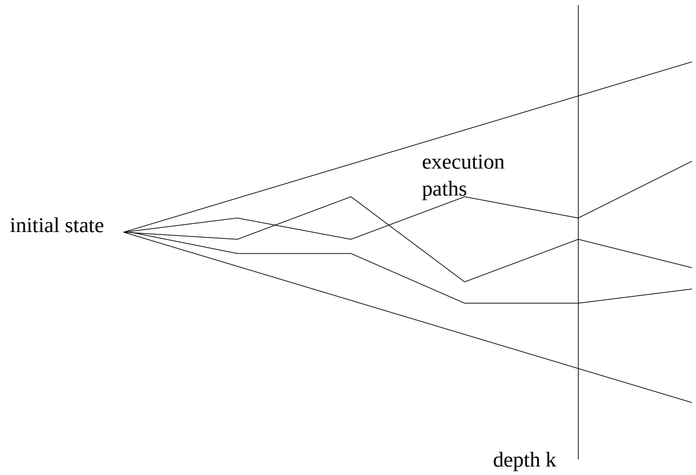
BPP **:** Complexity class of problems decidable by a Monte-Carlo randomized algorithm.

Sketch of the proof

- The problem of counting the number of paths of length $\leq |M|$, whose infinite extensions satisfy ψ reduces to $\#SAT$
- Computing the probability reduces to counting this number of paths
- $\#SAT$ is a $\#P$-complete problem
- So, if there was a FPRAS for computing $Prob_{\Omega}(\psi)$, then we could randomly approximate $\#SAT$
- If the previous statement holds, then $BPP = NP$

We consider $Prob_k(\phi)$ with :

- the probability space is the space over paths of length $\leq k$



- ψ express a monotone property

$$\lim_{k \to \infty} Prob_k(\phi) = Prob_\Omega(\phi)$$

Generic approximation algorithm $\mathcal{GAA}$

input : $\phi, diagram, \varepsilon, \delta$

Let $A := 0$

Let $N := \log(\frac{2}{\delta})/2\varepsilon^2$

For i from 1 to N do

1. Generate a random path σ of depth k
2. If ϕ is true on σ then $A := A + 1$

Return (A/N)

Algorithm based on Monte-Carlo estimation and Chernoff-Hoeffding bound

Diagram : succinct representation of the system (for example in Reactive Modules)

Method : Estimation (Monte-Carlo) + Chernoff-Hoeffding bound

X Bernoulli $(0, 1)$ random variable with success probability p

- Do N independent Bernoulli trials $X_1, X_2, \ldots, X_N$
- Estimate p by $\mu = \sum_{i=1}^{N} X_i/N$ with error ε
- Sample size N is such that the error probability $< \delta$

Chernoff-Hoeffding bound :

$$Pr[\mu < p - \varepsilon] + Pr[\mu > p + \varepsilon] < 2e^{-2N\varepsilon^2}$$

If $N \geq \ln(\frac{2}{\delta})/2\varepsilon^2$, then

$$Pr[p - \varepsilon \leq \mu \leq p + \varepsilon)] \geq 1 - \delta$$

Theorem :

$\mathcal{GAA}$ is a FPRAS for $Prob_k(\psi)$

Methodology : To approximate $Prob_\Omega[\psi]$

- Choose $k \approx log|M| \cdot ln(1/\varepsilon)$
- Iterate approximation of $Prob_k[\psi]$

Remark :

- Length of needed paths can be the diameter of the system
- Convergence time may be long, but space is saved...

Improvement :

Optimal Approximation Algorithm (Dagum, Karp, Luby and Madras) with multiplicative error.

APMC : Approximate Probabilistic Model Checker

- Freely available GPL software
- Developped at LRDE/EPITA, Paris VII and Paris XI Universities
- Use **randomized approximation algorithm**
- **Distributed** computation
- Integrated in the probabilistic model checker **PRISM**
- Main advantage : **space complexity** eliminated...

Action-Labelled Markov Chain :

$\mathcal{M} = (S, s_0, \{\mu_a \ / \ a \in \mathcal{A}\})$ where $\mathcal{A}$ is the set of actions

- S set of states, s_0 initial state
- $\mu_a : S^2 \longrightarrow [0,1]$ s. t. $(\forall s \in S) \sum_{t \in S} \mu_a(s,t) \leq 1$

If $X \subseteq S$ we note : $\mu_a(s,X) = \sum_{t \in X} \mu_a(s,t)$

Bisimulation between 2 processes $\mathcal{M}, \mathcal{M}'$:

Equivalence relation R on $S \uplus S'$ s. t.

$\forall s, s' \ sRs' \implies \forall C \ R-$equivalence class, $\mu_a(s,C) = \mu_a(s,C')$

Bisimilar states s, s' : there is a bisimulation relation R s. t. sRs'

Bisimilar processes $\mathcal{M}, \mathcal{M}'$: initial states are bisimilar

Problems :

- Probabilistic bisimulation is **too exact**

2 states are bisimilar only if the probabilities of outgoing transitions match exactly

- We would like a notion of **approximation** between processes

- We need **pseudo-metrics**

A pseudo-metric is a function d that associates a real number to each pair of processes, s. t.
$d(\mathcal{M}, \mathcal{M}') = 0$ iff $\mathcal{M}, \mathcal{M}'$ are bisimilar
$d(\mathcal{M}, \mathcal{M}') = d(\mathcal{M}', \mathcal{M})$
$d(\mathcal{M}, \mathcal{M}'') \leq d(\mathcal{M}, \mathcal{M}') + d(\mathcal{M}', \mathcal{M}'')$

- Algorithm to **compute** such a metric ?

Real-valued Logic for Labelled Markov Processes :

$\mathcal{F}_c$: family of functional expressions, indexed by $c \in]0,1]$

$f ::= \ \mathbf{1} \mid \mathbf{1} - f \mid \ < a > f \mid sup(f,g) \mid f \dot{-} q \qquad q \in \mathbb{Q}$
$\mathcal{F}_c^+$: without $\mathbf{1} - f$

Interpretation in $\mathcal{M} = (S, s_0, \Sigma, \{\mu_a \ / \ a \in \mathcal{A}\})$: $f_{\mathcal{M}} : \ S \longrightarrow [0,1]$

$\mathbf{1}_{\mathcal{M}}(s) = 1$
$(\mathbf{1} - f)_{\mathcal{M}}(s) = 1 - f_{\mathcal{M}}(s)$
$sup(f,g)_{\mathcal{M}}(s) = max(f_{\mathcal{M}}(s), g_{\mathcal{M}}(s))$
$(< a > f)_{\mathcal{M}}(s) = c. \int_S f_{\mathcal{M}}(t) \mu_a(s, dt)$
$(f \dot{-} q)_{\mathcal{M}}(s) = max(f_{\mathcal{M}}(s) - q, 0)$

Theorem : (Desharnais, Gupta, Jagadeesan and Panangaden)
For any labelled Markov processes $\mathcal{M}, \mathcal{M}'$, for all $c \in]0,1]$, $s \in S$ and $s' \in S'$ are bisimilar iff $(\forall f \in \mathcal{F}_c^+) \ f_{\mathcal{M}}(s) = f_{\mathcal{M}'}(s)$

Each collection $\mathcal{F}_c$ of functional expressions induces a **pseudo-metric** d_c :

$$d_c(\mathcal{M}, \mathcal{M}') = sup\{|f_{\mathcal{M}}(s_0) - f_{\mathcal{M}'}(s'_0)| \ / \ f \in \mathcal{F}_c\}$$

Theorem : (van Breugel and Worrell)
There exist pseudo-metrics d_n $(n \geq 0)$ s. t.

$$(\forall s_1, s_2 \in S) \ |d_n(s_1, s_2) - d_c(s_1, s_2)| \leq 2.c^n$$

Corollary :
To approximate d_c with error parameter ε, it is sufficient to compute $d_{\lceil log_c(\varepsilon/2)\rceil}$

Algorithm :

d_n can be computed as the solution of a **linear programming** problem

Input : (S, π) probabilistic transition system

- $S = \{s_0, \ldots, s_{N-1}\}$ $s_N = \mathbf{0}$ (refusal state)
- $(\pi(s_i, s_j))_{0 \leq i \leq N, 0 \leq j \leq N}$ probability matrix
- $(d_{n-1}(s_i, s_j))_{0 \leq i \leq N, 0 \leq j \leq N}$

Problem : Maximize $\sum_{0 \leq k \leq N} (\pi(s_i, s_k) - \pi(s_j, s_k)).y_k$ with

- $y_k - y_l \leq c.d_{n-1}(s_k, s_l)$ $(0 \leq k \leq N, 0 \leq l \leq N, k \neq l)$
- $y_k - y_N \leq 1$ and $y_N - y_k \leq 1$
- $y_k \geq 0$ $(0 \leq k \leq N)$

Conclusion

- Efficacité de l'approximation probabiliste (élimination de la complexité en espace)
- Vérification de propriétés monotones (accessibilité) et anti-monotones (sureté)
- Extension de la méthode à d'autres classes de propriétés (vivacité) ?
- Extension aux chaînes de Markov en temps continu et à la logique CSL (*Continuous Stochastic Logic*)
- Approximation de pseudo-métriques pour la bisimulation probabiliste

- [CY95] C. Courcoubetis et M. Yannakakis. *The complexity of probabilistic verification*. Journal of the ACM, 24(4) :857-907, 1995.
- [HLMP04] T. Hérault, R. Lassaigne, F. Magniette et S. Peyronnet. *Approximate Probabilistic Model Checking*. Int. Conf. on Verification, Model Checking and Abstraction, LNCS n° 2937.
- [KLM89] R. Karp, M. Luby et N. Madras. *Monte-Carlo Approximation Algorithms for Enumeration Problems.* Journal of Algorithms 10, 429-448, 1989.
- [KNP02] M. Kwiatkowska, G. Norman et D. Parker. *Probabilistic symbolic model checking with PRISM : A hybrid approach.* Proc. of 8th Int. Conf. TACAS, LNCS n° 2280, p.52-66, 2002.
- [LR03] R. Lassaigne et M. de Rougemont : *Logic and Complexity.* Springer-Verlag, 350 p. (nov. 2003).