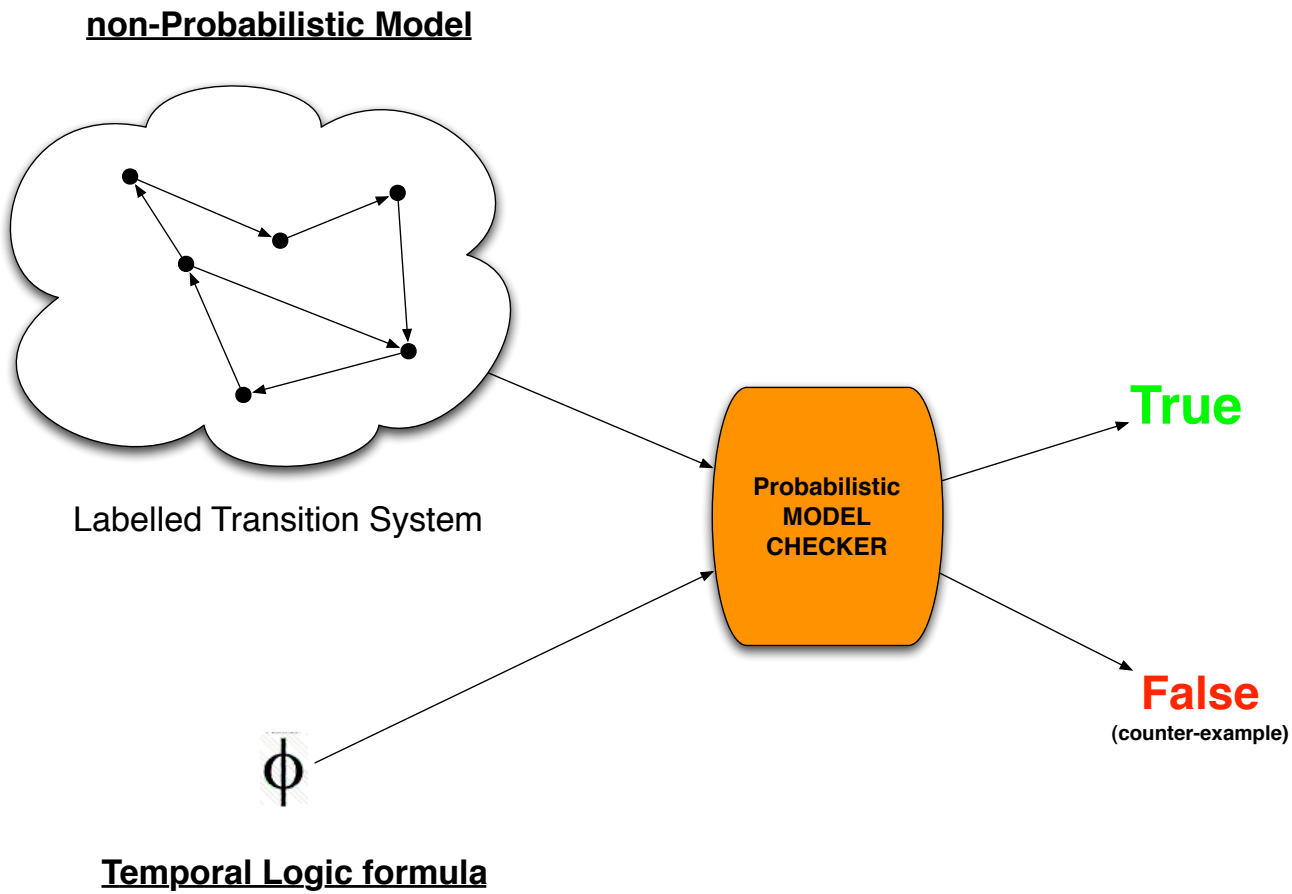
Statistical Model Checking

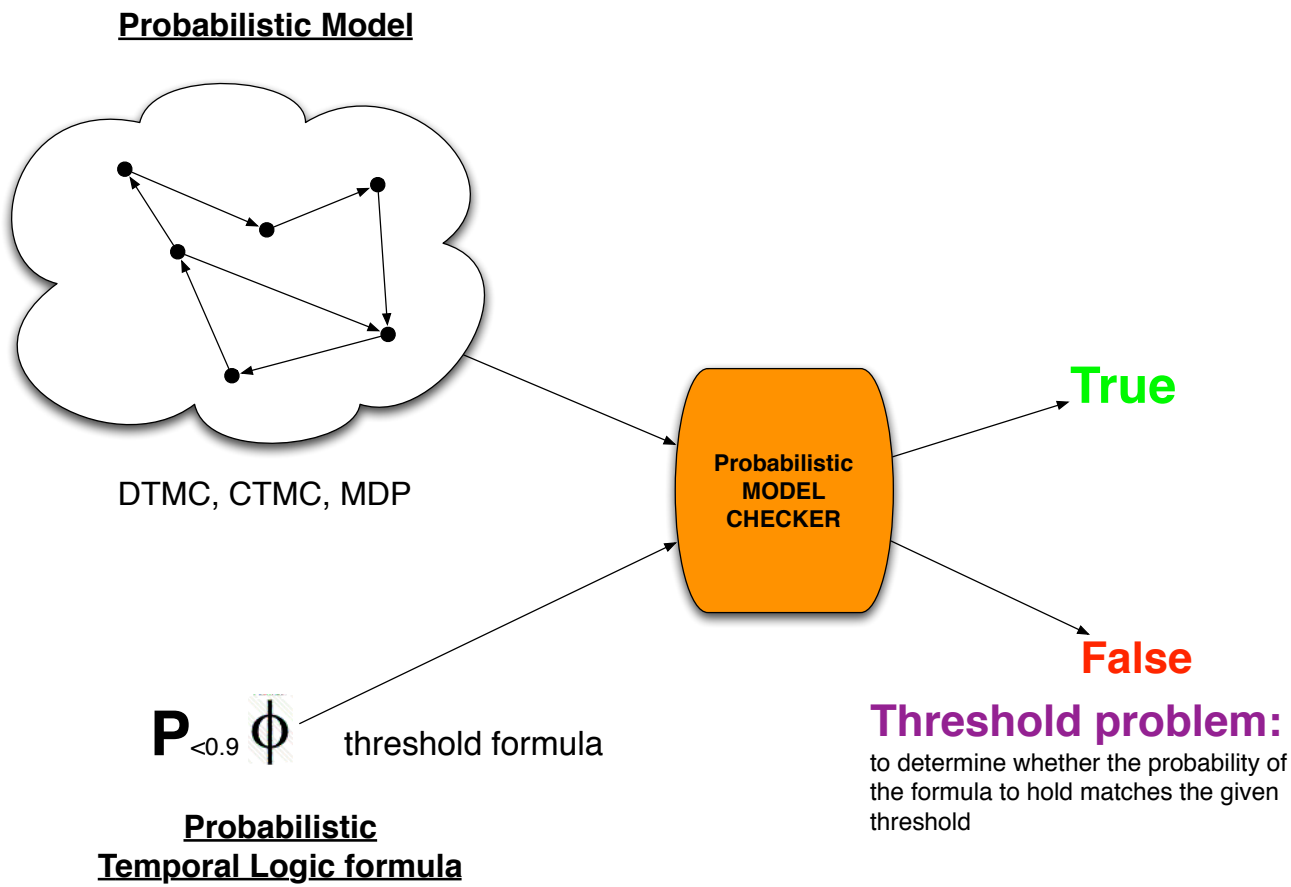
(an overview)

Paolo Ballarini

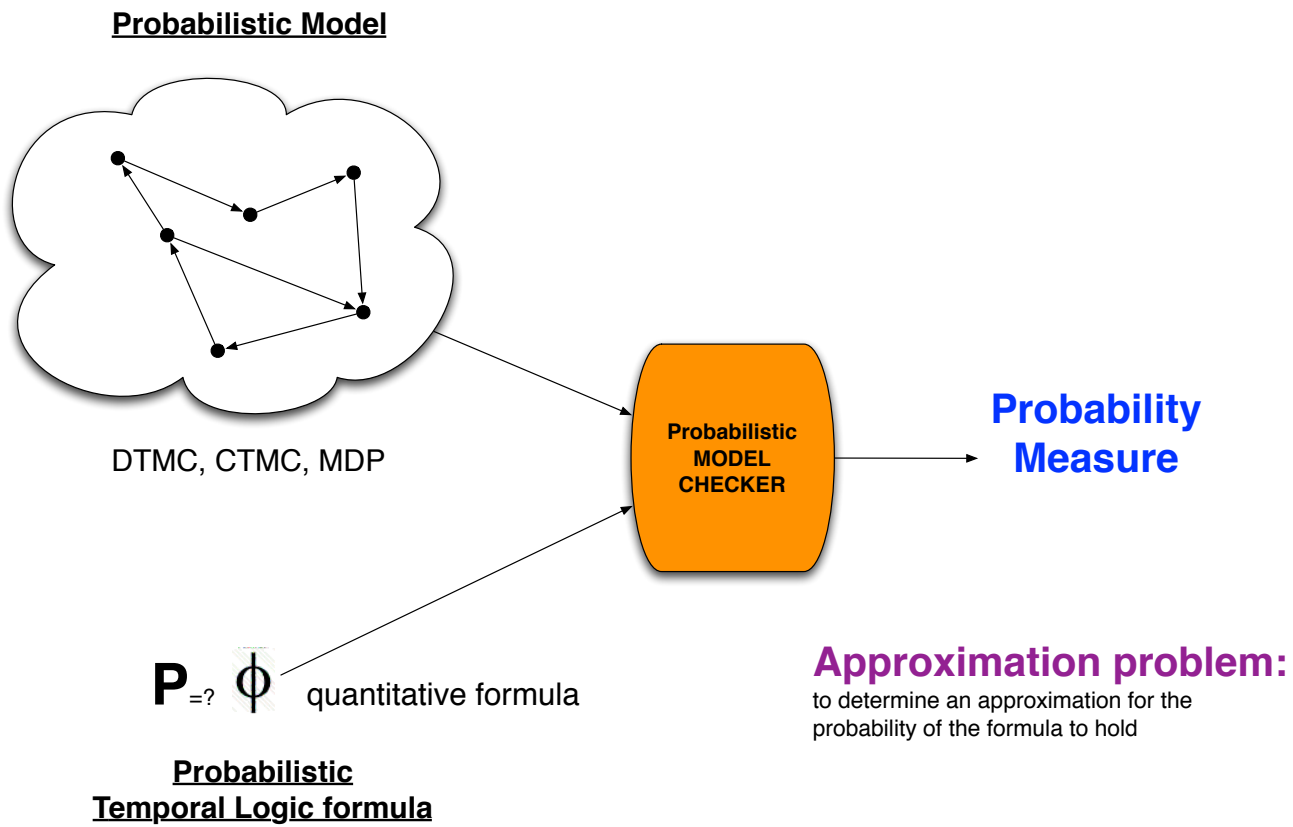
The Model Checking approach



The Probabilistic Model Checking approach



The Probabilistic Model Checking approach



Numerical vs Statistical

- numerical model checking: application of numerical methods for Model's solution (i.e. *transient analysis*), the approximation is in form of truncation error
  - **+** : better when very accurate approximation are required
  - **−** : large memory requirements
  - **−** : many realistic models are numerically untreatable

- statistical model checking: application of discrete-event simulation to sampling model's executions used to produced an approximated output
  - **+** : small memory requirements (no need to store the state-space)
  - **+** : only option for verification of many realistic models
  - **−** : very accurate approximations may be computationally harder

Outline

Temporal Logic for CTMC Models

Statistical Model Checking
- Threshold problem (Hypothesis Testing)
- Approximation problem

Current work, future directions
- Confidence Interval based Model Checking

Outline

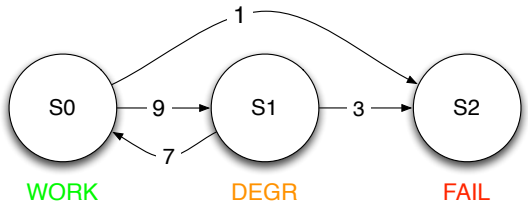
Temporal Logic for CTMC Models

Statistical Model Checking
- Threshold problem (Hypothesis Testing)
- Approximation problem

Current work, future directions
- Confidence Interval based Model Checking

Temporal Queries for CTMC models

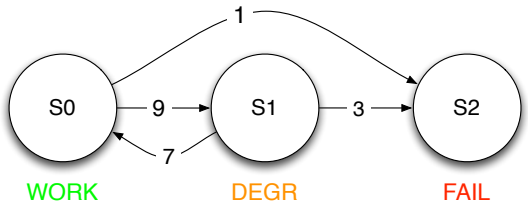
CTMCs are probabilistic and timed models, thus query languages for CTMCs must allow for both *probability measures* and *time bounds*



- *what is the probability that the system will FAIL (within time T)*
  ("unconditional" reachability)
- *what is the probability that the system will go from WORK to FAIL (within time T)*
  ("simple-conditional" reachability)
- *what is the probability that the system will go from WORK to DEGR to FAIL (within time T)*
  ("sequential" reachability)
- *what is the probability that the system will go from WORK to DEGR within time T_1 and then from DEGR to FAIL within time $T2$*
  ("timed-sequential " reachability)

Temporal Queries for CTMC models

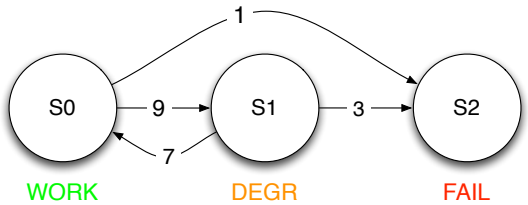
CTMCs are probabilistic and timed models, thus query languages for CTMCs must allow for both *probability measures* and *time bounds*



- *what is the probability that the system will FAIL (within time T)*
  ("unconditional" reachability)
- *what is the probability that the system will go from WORK to FAIL (within time T)*
  ("simple-conditional" reachability)
- *what is the probability that the system will go from WORK to DEGR to FAIL (within time T)*
  ("sequential" reachability)
- *what is the probability that the system will go from WORK to DEGR within time T_1 and then from DEGR to FAIL within time $T2$*
  ("timed-sequential " reachability)

Temporal Queries for CTMC models

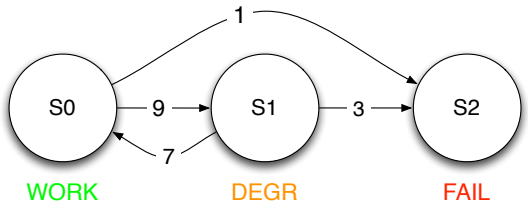
CTMCs are probabilistic and timed models, thus query languages for CTMCs must allow for both *probability measures* and *time bounds*



- *what is the probability that the system will FAIL (within time T)*
  ("unconditional" reachability)
- *what is the probability that the system will go from WORK to FAIL (within time T)*
  ("simple-conditional" reachability)
- *what is the probability that the system will go from WORK to DEGR to FAIL (within time T)*
  ("sequential" reachability)
- *what is the probability that the system will go from WORK to DEGR within time T_1 and then from DEGR to FAIL within time $T2$*
  ("timed-sequential " reachability)

Temporal Queries for CTMC models

CTMCs are probabilistic and timed models, thus query languages for CTMCs must allow for both *probability measures* and *time bounds*



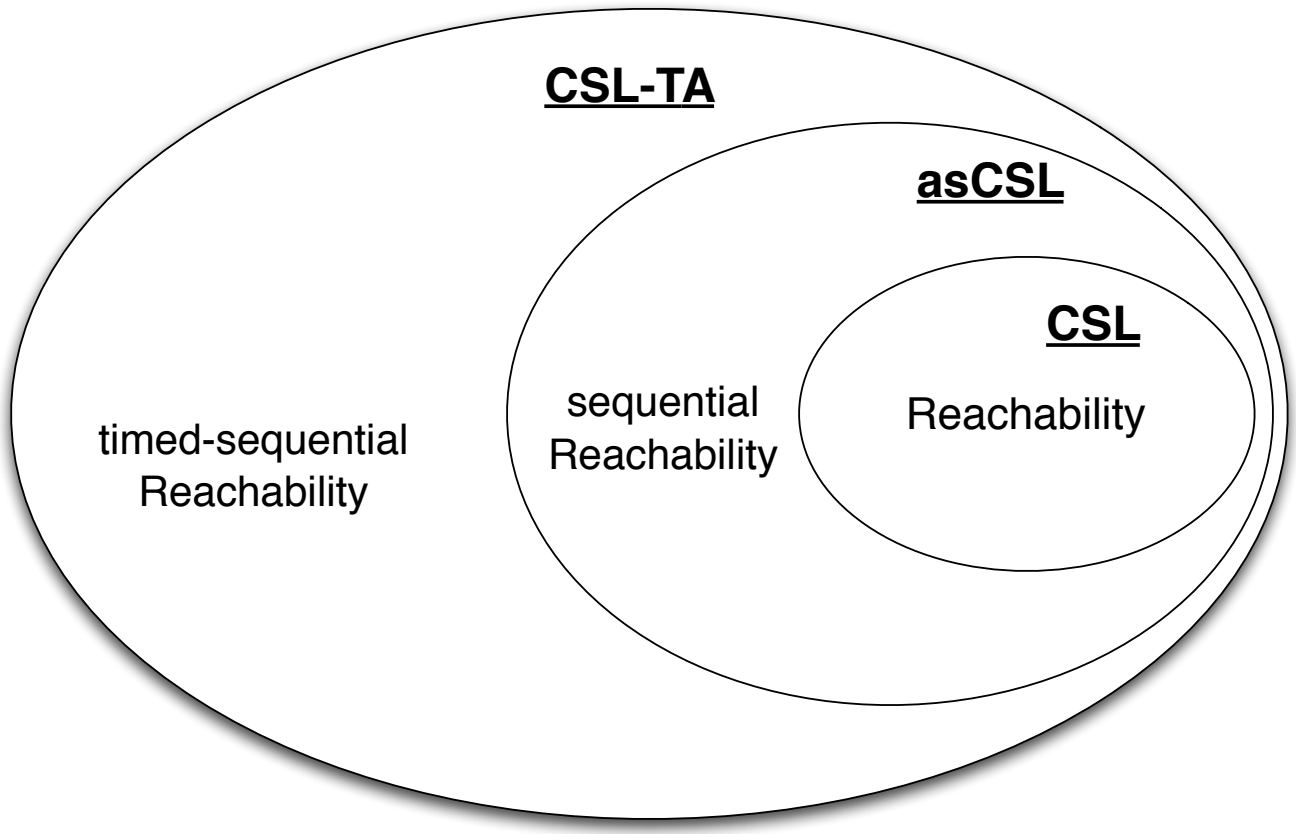
- *what is the probability that the system will FAIL (within time T)*
  ("unconditional" reachability)
- *what is the probability that the system will go from WORK to FAIL (within time T)*
  ("simple-conditional" reachability)
- *what is the probability that the system will go from WORK to DEGR to FAIL (within time T)*
  ("sequential" reachability)
- *what is the probability that the system will go from WORK to DEGR within time T_1 and then from DEGR to FAIL within time $T2$*
  ("timed-sequential " reachability)

Temporal Logic Taxonomy

CSL: Continuous Stochastic Logic

asCSL: action-state CSL

CSL-TA: CSL Timed Automata

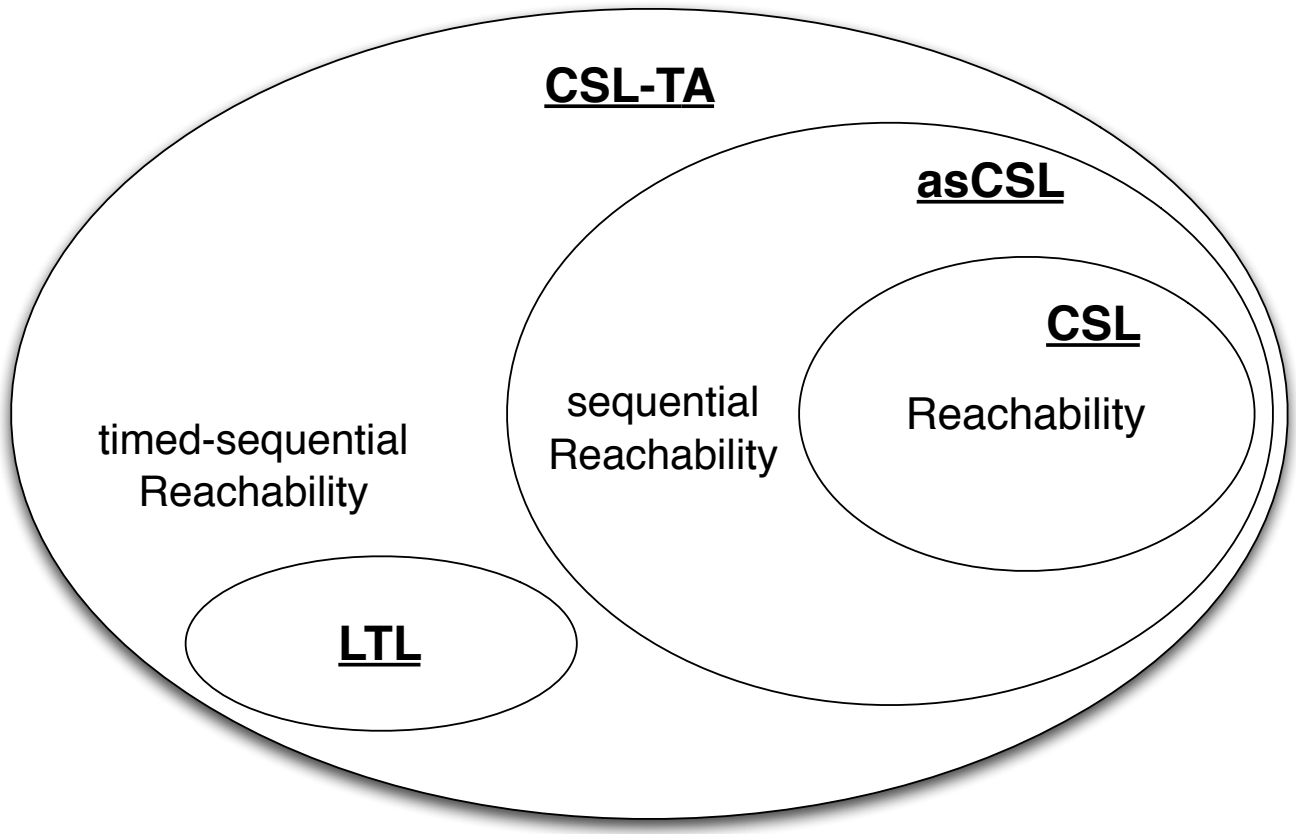


Temporal Logic Taxonomy

CSL: Continuous Stochastic Logic

asCSL: action-state CSL

CSL-TA: CSL Timed Automata

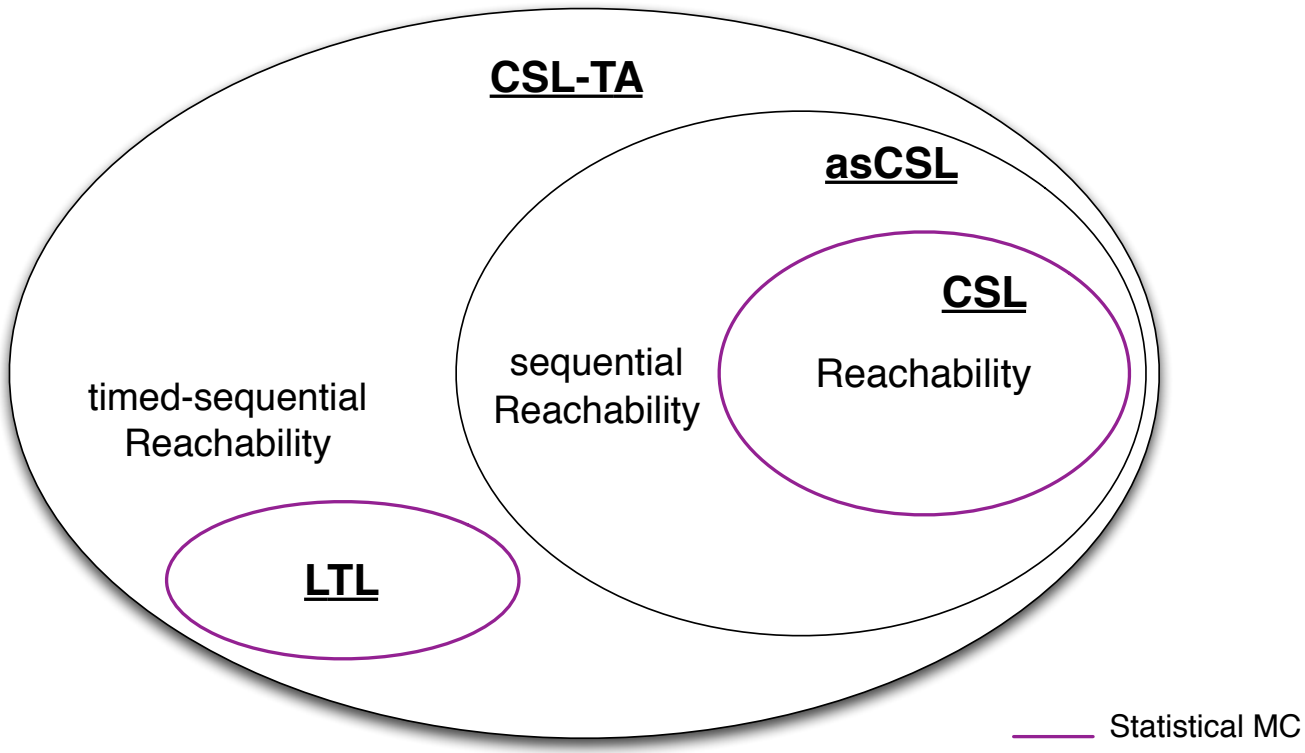


Temporal Logic Taxonomy

CSL: Continuous Stochastic Logic

asCSL: action-state CSL

CSL-TA: CSL Timed Automata



CSL (Continuous Stochastic Logic)

- CSL: a branching time logic

$$\boxed{\phi := a \mid \neg\phi \mid \ \phi \wedge \phi \mid \mathcal{S}_{\sim p}(\phi) \mid \mathcal{P}_{\sim p}(\varphi)}$$

$$\boxed{\varphi := X^I \ \phi \mid \phi \, U^I \phi}$$

  - $\sim \in \{<, \leq, \geq, >, =, \neq\}$, $I \subseteq \mathbb{R}_{\geq 0}$ (time-bound)

- characteristic of CSL: each temporal operator must be preceded by a probability bound $\sim p$

$$\phi \equiv \mathcal{P}_{\leq 0.9}(WORK \ U \ [\mathcal{P}_{<0.5}(DEGR \, U \ FAIL)])$$

Outline

Statistical MC: two formulations

1. **Threshold problem:** does the probability measure of query Q meets the bound $\sim p$?
   - CSL
     - Younes, Simmons: YMER (sequential Acceptance sampling)
     - San et al: VESTA (simple hypothesis testing)
     - Katoen, Zapreev: MRMC (through Estimation)
   - LTL
     - Clarke et al.: (hypoth. Testing + Bayesian)

2. **Estimation problem:** what is the probability measure of query Q?
   - LTL
     - Peyronnet *et al.*: APMC (Chernoff-Hoeffding bound)
   - CSL
     - Parker *et al.*: PRISM (implementation of Peyronnet *et al* method)
     - Katoen, Zapreev: MRMC (?)
   - LTL
     - Ballarini et al: OCEAN (Wilson Confidence Interval)

Statistical MC: two formulations

1. **Threshold problem:** does the probability measure of query Q meets the bound $\sim p$?
   - CSL
     - Younes, Simmons: YMER (sequential Acceptance sampling)
     - San et al: VESTA (simple hypothesis testing)
     - Katoen, Zapreev: MRMC (through Estimation)
   - LTL
     - Clarke et al.: (hypoth. Testing + Bayesian)

2. **Estimation problem:** what is the probability measure of query Q?
   - LTL
     - Peyronnet *et al.*: APMC (Chernoff-Hoeffding bound)
   - CSL
     - Parker *et al.*: PRISM (implementation of Peyronnet *et al* method)
     - Katoen, Zapreev: MRMC (?)
   - LTL
     - Ballarini et al: OCEAN (Wilson Confidence Interval)

Statistical MC: two formulations

1. **Threshold problem:** does the probability measure of query Q meets the bound $\sim p$?
   - CSL
     - Younes, Simmons: YMER (sequential Acceptance sampling)
     - San et al: VESTA (simple hypothesis testing)
     - Katoen, Zapreev: MRMC (through Estimation)
   - LTL
     - Clarke et al.: (hypoth. Testing + Bayesian)

2. **Estimation problem:** what is the probability measure of query Q?
   - LTL
     - Peyronnet *et al.*: APMC (Chernoff-Hoeffding bound)
   - CSL
     - Parker *et al.*: PRISM (implementation of Peyronnet *et al* method)
     - Katoen, Zapreev: MRMC (?)
   - LTL
     - Ballarini et al: OCEAN (Wilson Confidence Interval)

Statistical MC: basic idea

- **Inputs:**
  - a model M and property ϕ
  - a bound for the desired level of approximation:
    - **hypothesis testing :** error probability bounds $< \alpha, \beta >$ + width of indifference region δ
    - **error bounded estimation:** the desired error ϵ + the confidence level δ
    - **confidence interval estimation:** a confidence level α + and width of the confidence interval ϵ
- generate N (finite) *sample trajectories* σ_i through discrete-event stochastic simulation
- to each trajectory σ_i corresponds a Bernoulli variable X_i:

$$\begin{cases} X_i = 1 \longleftrightarrow (\sigma_i \models \phi) \\ X_i = 0 \longleftrightarrow (\sigma_i \not\models \phi) \end{cases}$$

- **outcome:**
  - an approximated answer to $M \models \phi$ with a bounded error probability.
  - an approximated estimate of $Pr(\phi, M)$

Hypothesis-testing based Statistical MC

- **the problem:** decide whether $Pr(\phi, M) \sim \theta$, for example:

$$\mathcal{P}_{\leq 0.05}(true\ U^{[0,3]}\ FAIL)$$

("there's at most $\theta = 0.05$% chance that system will FAIL within time 3")

- **principle:** there's no need to get an accurate estimate of $Pr(\phi, M)$ in order to decide whether $Pr(\phi, M) \sim \theta$: a "rough" estimate (resulting from small samples) may be enough
- **hypothesis testing formulation:** the decision on $Pr(\phi, M) \sim \theta$ can be formulated in terms of an Hypothesis Testing problem, with

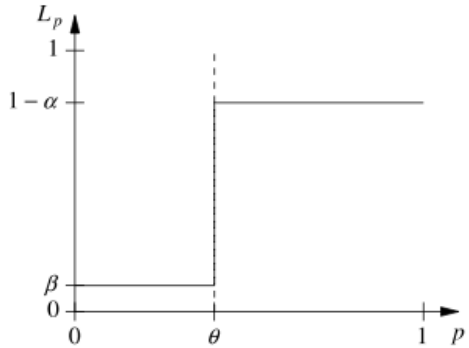
$$H : p \geq \theta$$
$$K : p < \theta$$

where $p = Pr(\phi, M)$ is the actual probability.

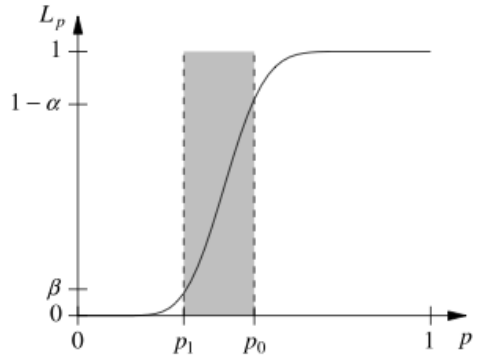
α :probability of accepting K when H holds (false negative)
β :probability of accepting H when K holds (false positive)

$\langle \alpha, \beta \rangle$ is referred to as the strength of the test

Acceptance Sampling [Younes, Simmons]



(a) Prob. of accepting H (ideal)



(b) Prob. of accepting H (with indifference region)

- **a)** problems: requires exhaustive sample
- **b)** an indifference region is introduced and hypothesis testing is reformulated:

$$H_0 : p \geq p_0 \qquad p_0 = \theta + \delta$$
$$H_1 : p < p_1 \qquad p_1 = \theta - \delta$$

Acceptance Sampling (1): fixed-size samples

The number of successes $Y = \sum_i X_i$ is Binomial

$$Pr(Y \leq c) = F(c, n, p) = \sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i}$$

- **fixed-size samples**
  - H_0 is accepted if $\sum_i X_i > c$ (c is a constant)
  - the acceptance problem corresponds to an optimization problem: *"find n and c such that;"*

$$F(c, n, p_0) \leq \alpha \qquad 1 - F(c, n, p_1) \leq \beta$$

given the strength $\langle \alpha, \beta \rangle$ and the indifference region $IR = [p_1, p_0]$.

approximated sample size is:

$$\tilde{n} = \frac{\Phi^{-1}(\alpha)\sqrt{p_0(1-p_0)} + \Phi^{-1}(\beta)\sqrt{p_1(1-p_1)}}{(p_0 - p_1)^2}$$

Note that: for a fixed width of IR the sample size is largest if IR is centered in $p = 1/2$ and decreases when moving towards the extremes 0 and 1.

Acceptance Sampling (2): sequential test

- **Wald sequential test:** a fixed size sample n may be a waist
  - if after $m < n$ observations $Y > c$ then we can accept H_0
  - if after $m < n$ observations Y cannot exceed n then we can accept H_1

$$\boxed{f_m = \prod_{i=1}^{m} \frac{Pr[X_i = x_i | p = p_1]}{Pr[X_i = x_i | p = p_0]} = \frac{p_1^{d_m}(1-p_1)^{m-d_m}}{p_0^{d_m}(1-p_0)^{m-d_m}}}$$

$d_m = \sum_{i=1}^{m} x_i$

- H_0 is accepted if $f_m \leq B$ with $B = \frac{\beta}{(1-\alpha)}$
- H_1 is accepted if $f_m \geq A$ with $A = \frac{1-\beta}{\alpha}$

CSL Model Checking through acceptance sampling

so far: *given a simple formula ϕ we set the error bounds α and β and the method estimates whether ϕ holds with errors α and β*

$$\mathcal{P}_{\leq}0.05(WORK\ U^{[0,3]}\ FAIL)$$

problem: *how do we set the error bounds for "complex formulae"?*

- $\mathcal{P}_{\leq}0.05(WORK\ U^{[0,3]}\ FAIL) \wedge \mathcal{P}_{\geq}0.6(WORK\ U^{[1,2]}\ DEGR)$ (conjunction)
- $\mathcal{P}_{\leq}0.5(WORK\ U^{[0,3]}\ \mathcal{P}_{\geq}0.1(DEGR\ U^{[1,2]}\ FAIL))$ (nested UNTIL)

- **non-nested path formulae:**
  - **conjunction:** to verify $\bigwedge_{i=1}^{n} \Phi_i$ with strenght $\langle \alpha, \beta \rangle$ it is sufficient to verify each conjunct with strength $\langle \alpha/n, \beta/n \rangle$.
- **nested path formulae:**
  - the strength of the test for a property with nested prob. operators is a function on the strength of the test of its nested prob. operators
  - in order to verify $\mathcal{P}_{\sim\theta}(\varphi)$ with strength $\langle \alpha, \beta \rangle$ set the IR to $p_0 = (\theta + \delta(\theta)(1 - \alpha')$ and $p_1 = 1 - (1 - (\theta - \delta(\theta)))(1 - \beta')$, where $\langle \alpha', \beta' \rangle$ is the strength of the test for φ.

Statistical CSL threshold problem: summary

- **Approach:** statistical solution to the *"threshold problem"* for (subset of) CSL
- **Featured properties:**
  - time-bounded CSL (both simple and nested)
- **Non-Featured properties:**
  - unbounded CSL (a tentative-solution has been proposed by Sen *et al.* [2005])
  - steady-state formulae (on-going work by Pekergin, El Raib [2009])
- **Tools:**
  - YMER (Younes, Simmons)
  - VESTA (Sen *et al.*)

Statistical MC as an Approximation Problem

Problem: *given a formula ϕ and probabilistic model M, calculate an approximation of* $p = Pr(\phi, M)$ based on N sampled trajectories of M.

- the point estimator is $\hat{p} = \frac{\sum_i X_i}{N}$

there are 2 approaches:

1. **error bounded estimation (Chernoff-Hoeffding bound):**
   - ϵ: error
   - δ: confidence-level

$$Prob\Big((p - \epsilon) \leq \hat{p} \leq (p + \epsilon)\Big) \geq 1 - \delta$$

the probability that estimate $\hat{p}$ is farther than ϵ from actual value is less than δ

2. **confidence-interval estimation:**
   - α: confidence-level
   - ϵ: interval-width

$$Prob\Big(u(\hat{p}) \leq p \leq v(\hat{p})\Big) \geq 1 - \alpha$$

$v(\hat{p}) - u(\hat{p}) = \epsilon$: functions of the sample

the probability that the actual value is farther than δ from the estimate $\hat{p}$ is less than α

Approximate Probabilistic Model Checking

- APMC tool [Peyronnet *et al.*]: approximate LTL model checking tool
- it calculates approximated probability measure for:
  - general time-bounded LTL
  - time-unbounded monotone LTL (i.e. formulae of the *positive fragment* of LTL, i.e. negation only applied to atomic-propositions)
- it supports two *error bounded estimation* schemes:
  - Randomized approximation with additive error
    - fixed sample-size: $\boxed{N = \ln(\frac{2}{\delta})/2\epsilon^2}$
  - Randomized approximation with multiplicative error
    - complex 3-steps estimation procedure

Outline

Confidence Interval based Model Checking

- OCEAN [Ballarini *et al.*]: a prototype tool for confidence-interval estimation of the probability of time-bounded LTL formulae vs CTMC biological models expressed in SBML

- it is based on the *Wilson-score interval* (an alternative to the most popular *Wald interval*) which guarantees a better coverage-probability than Wald's

confidence interval

$$[L,U] = \frac{\hat{p} + \frac{1}{2N} z^2_{1-\alpha/2} \mp z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{N} + \frac{z^2_{1-\alpha/2}}{4N^2}}}{1 + \frac{1}{N} z^2_{1-\alpha/2}}$$

sample size

$$N \geq z^2_{1-\alpha/2} \frac{\hat{p}(1-\hat{p}) - 2\epsilon^2 + \sqrt{\hat{p}^2(1-\hat{p})^2 + 4\epsilon^2(\hat{p}-0.5)^2}}{2\epsilon^2}$$

The Iterative Wilson Procedure

WILSON PROCEDURE

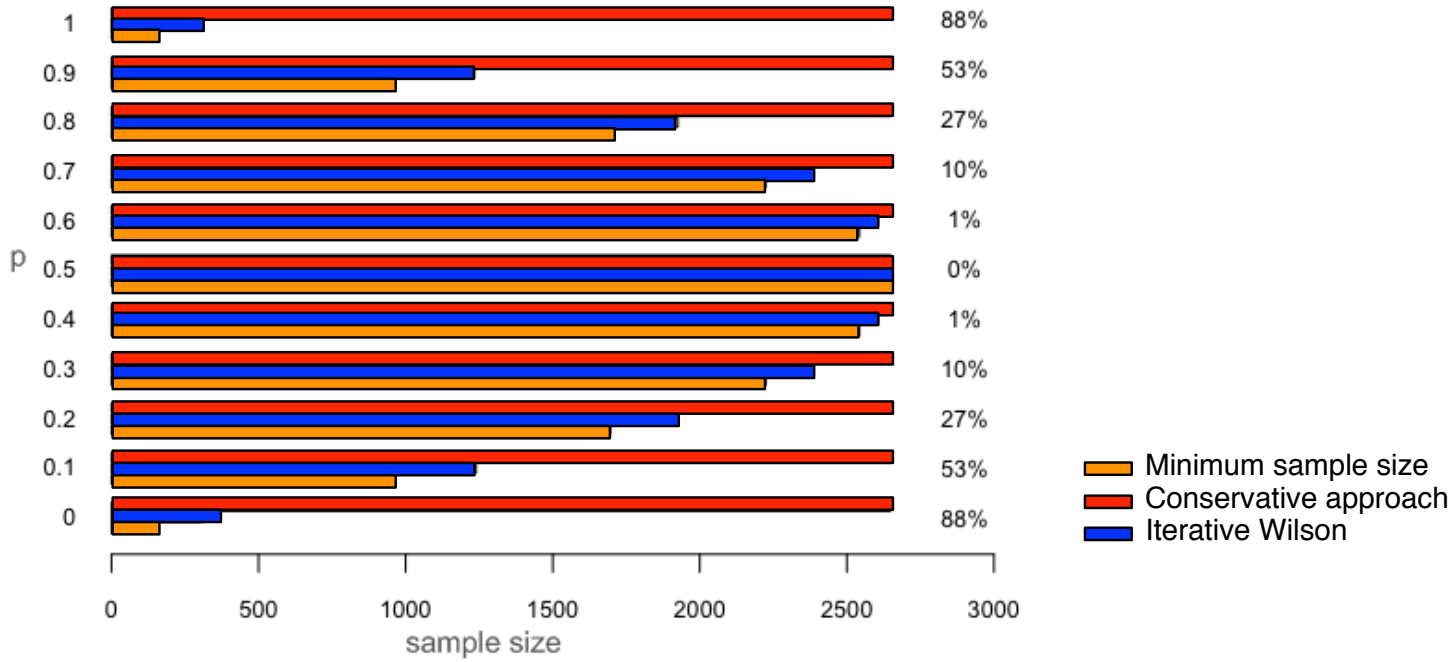
1 $N_{tot} = 0$
2 $N = \text{Wilson\_sample}(\hat{p}_0, \epsilon, \alpha)$
3 Perform N experiments ($\sigma \models \phi$)
4 $N_{tot} = N_{tot} + N; \hat{p} = \frac{|YES|}{N_{tot}}$
5 **if** $\hat{p} \leq 0.5$
6 $\quad p' = \hat{p} + \epsilon$
7 **else** $p' = \hat{p} - \epsilon$
8 $N' = \text{Wilson\_sample}(\text{p}', \epsilon, \alpha); N_{new} = N' - N_{tot}$
9 **if** $N_{new} > 0$
10 $\quad N = N_{new}$; **goto** 3
11 **else return** $\hat{p}$ and $\text{Wilson\_interval}(\hat{p}, N_{tot}, \alpha)$

---

property: the number of iterations depends on the chosen initial estimate $\hat{p}_0$

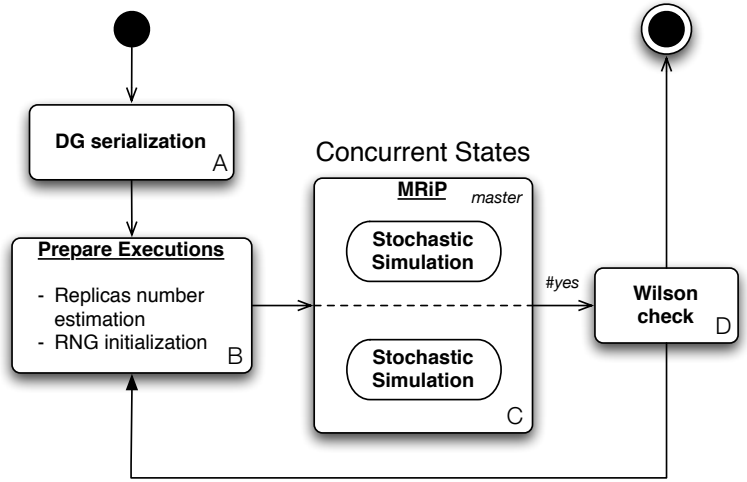
- **conservative approach** $(\hat{p}_0 = 0.5)$**:** worst case
- **iterative approach** $(\hat{p}_0 = 1$ **or** $\hat{p}_0 = 0)$**:** better case
- **Minimum sample size approach** $(\hat{p}_0 =$**actual** $p)$**:** best case

Performance of Wilson interval method

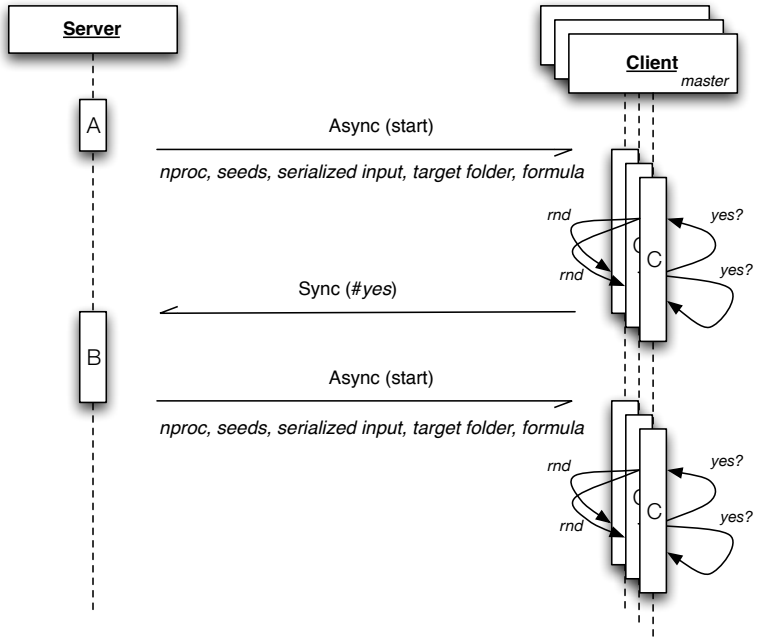


result: the iterative approach outperforms the (most popular) conservative approach particularly with *e*xtreme (actual) probability (up to 88% fewer iterations required)

Software Prototype: parallel on-the-fly verification

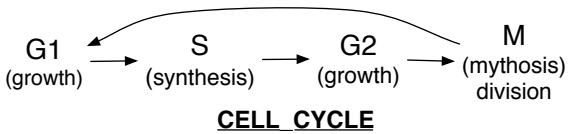


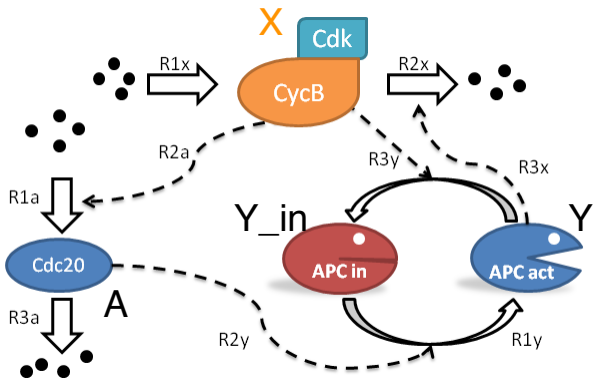
(c) Wilson-driven computational loop



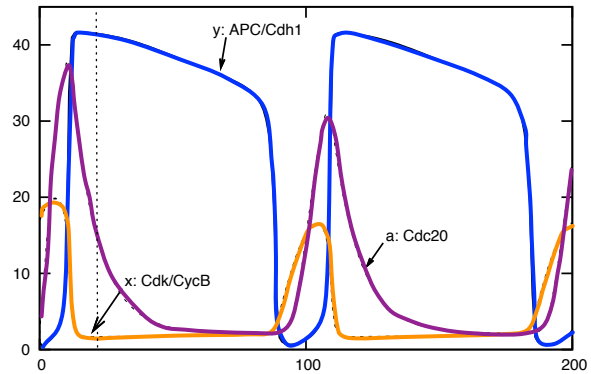
(d) Inter-process and client-server communications

Case study: the Cell-Cycle regulatory network





CELL-CYCLE
control network



simulation of X, Y, and A species

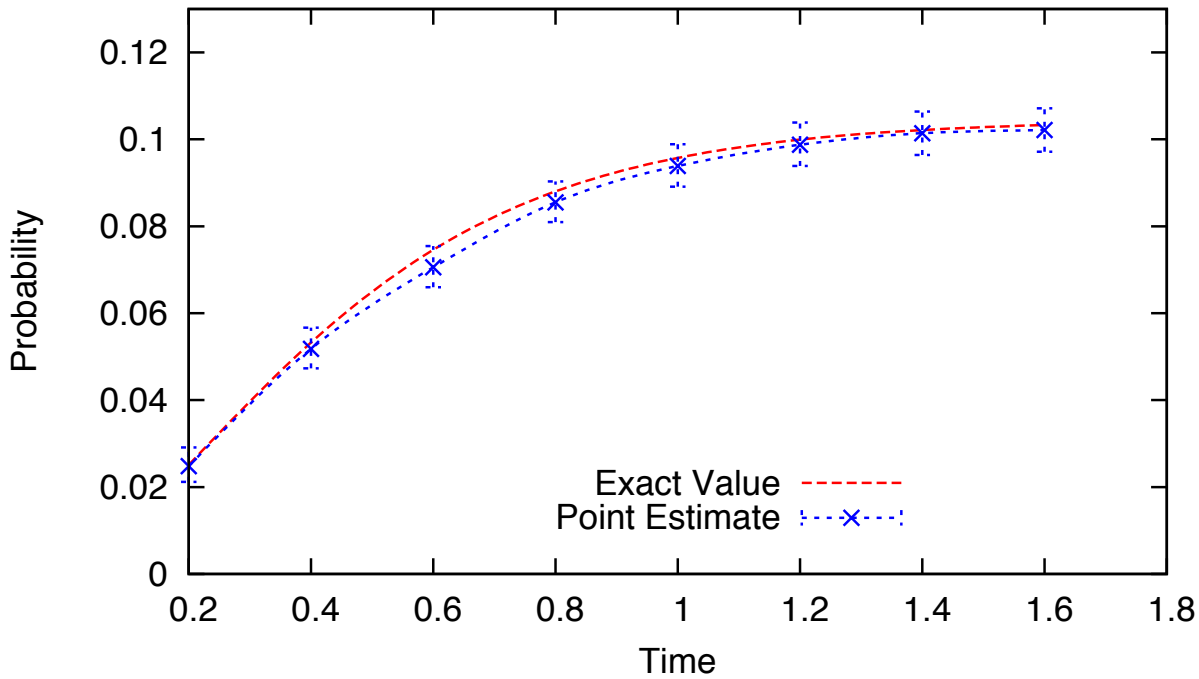
S-G2-M
transition

Cdk/CycB (X)	APC/Cdh1 (Y, Y_{in})	Cdc20 (A)
$\emptyset \xrightarrow{k_1 \alpha} X$	$Y_{in} \xrightarrow{k_3^*} Y$	$\emptyset \xrightarrow{k_5' \alpha} A$
$X \xrightarrow{k_2'} \emptyset$	$Y_{in} + a \xrightarrow{k_3'''} Y + A$	$X \xrightarrow{k_5^*} X + A$
$X + Y \xrightarrow{k_2'' \alpha} Y$	$X + Y \xrightarrow{k_4^*} X + Y_{in}$	$A \xrightarrow{k_6} \emptyset$

Statistical Verification of Cell-Cycle model

- Relevant property: "the influence of A (Cdc20) on $Y(APC)$ during the $S/G2/M$ transition"

$$\phi \equiv (A \leq 4)\ U^{\leq t}\ (Y \geq 5)$$



Threshold Problem vs Approximation Problem

point: what statistical MC approach is better, Threshold Problem (Hypothesis Testing) or Approximation Problem (Error bounded estimation/Confidence Interval)?

- Hypothesis Testing MC is generally cheaper to run (smaller sample size)
- however Hypothesis Testing MC does not provide a probability measure

it depends on the application

- Hypothesis Testing is preferable if strict Safety/Dependability constraints are known
  - e.g. *"the prob. of a breakdown must not exceeds* 10^{-2} *"*
- Approximation based MC is preferable for applications which require a lot of SENSITIVITY ANALYSIS
  - e.g. *"to check the effect of a model's parameter (e.g. a rate of a transition) has on a given property* ϕ*"*

  this is the case for example in Biological Modeling

Future Work, Open Issues

- CSL-TA statistical model checking (starting now)
  - Approximation problem
  - Threshold problem ?
- Error bound approximation vs Confidence Interval approximation
  - how do they compare?
- Time-unbounded properties: verification through finite length experiments
- Combination of Numerical and Statistical methods as a way to improve performances