Efficient Probabilistic Model Checking of Systems with Ranged Probabilities

Khalil Ghorbal[1,2] Parasara Sridhar Duggirala[1,3]
Franjo Ivančić[1] Vineet Kahlon[1] Aarti Gupta[1]

[1] NEC Laboratories America, Inc.

[2] *now with* Carnegie Mellon University

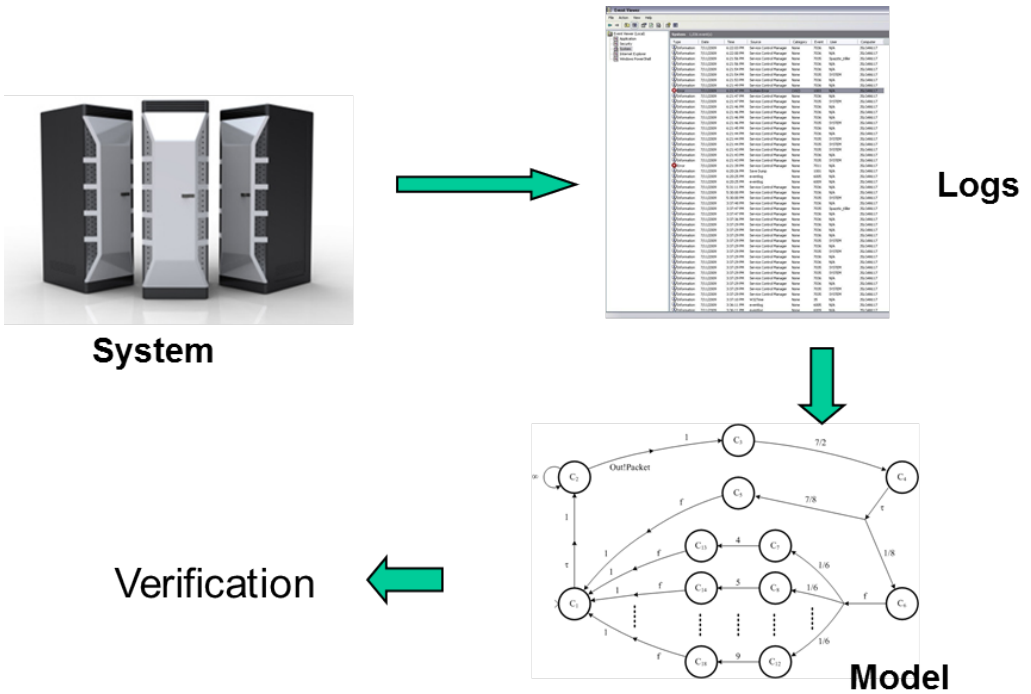
[3] *now with* University of Illinois at Urbana-Champagne

September 18th, 2012
Reachability Problems

Problem Statement

Analyze real-world stochastic systems

- Large systems contain many components (including third-party)
- Full formal system description not available
- *But:* Execution logs are easily generated



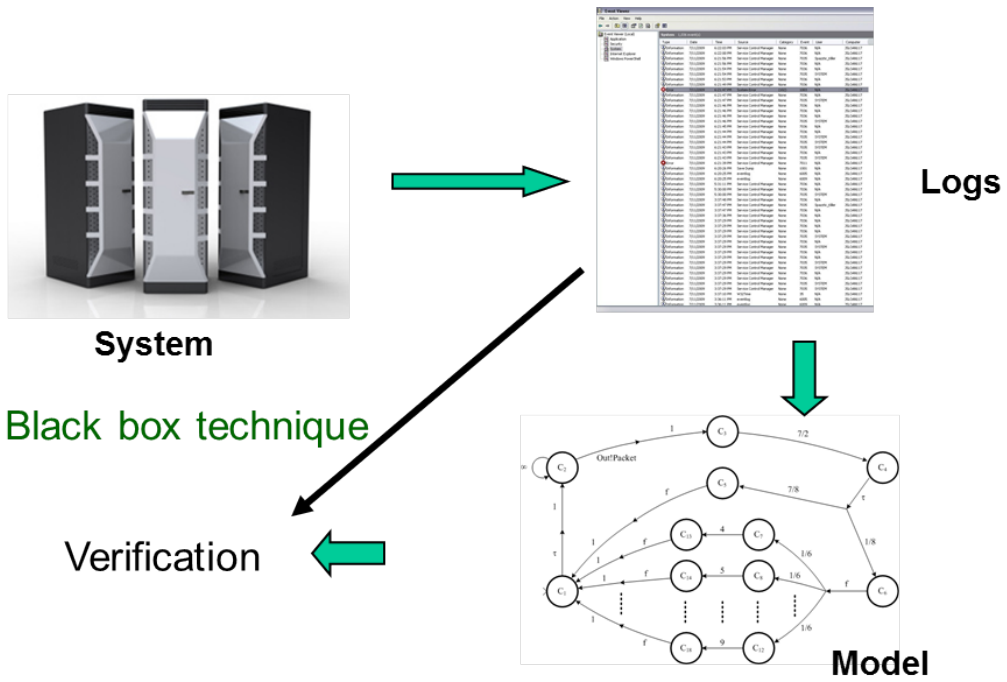
State-of-the-art solution: Black Box Technique

Black box techniques

- No system model
- Qualitative and quantitative properties

Learning Models

- Many applications need models (for example: anomaly detection)
- Bootstrapping to learn stochastic models



Can we use *approximate* learned models for sound analysis?

Motivation

Analyze real-world stochastic systems

- Follow model based approach
- Analysis based on the (finite) set of execution logs generated at runtime (usually available for debugging purposes)
- Try to bridge the gap between the model and the system under analysis
- Need to provide a way of capturing confidence about the learned model

Overview

- Phase I: Learning: set of logs $\mapsto$ Stochastic Model (*Interval-Valued* Discrete-Time Markov Models)
- Phase II: Model Checking (sound quantitative analysis ... of the model !)

Why Interval Discrete Time Markov Chains (IDTMC) ?

- Finite set of logs leads to approximate transition probabilities $\pm$ error due to the learning technique.
- To quantify the confidence in the model we use interval transition probabilities where the width of interval is related to the confidence parameters of the learning technique.

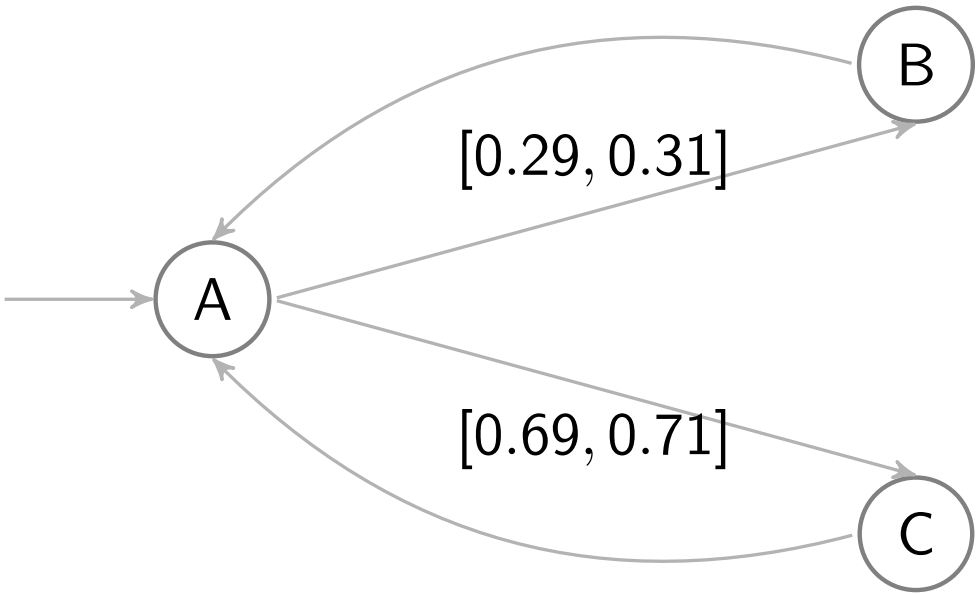


Figure : Small IDTMC Example

Outline

Definitions

DTMC

A DTMC is a 4-tuple: $\mathbf{M} \stackrel{\text{def}}{=} (S, s_0, P, \ell)$:

- S is a finite set of states,
- $s_0 \in S$ the initial state,
- P a transition probability matrix,
- $\ell : S \rightarrow 2^{AP}$ is a labelling function, $\ell(s_i)$ gives the set of atomic propositions $a \in AP$ that are valid in s,
- AP denotes a finite set of atomic propositions.

The component p_{ij} of the square matrix P denotes the transition probability between state s_i and state s_j:

$$P[X_t = s_j \mid X_{t-1} = s_i]$$

.

Example

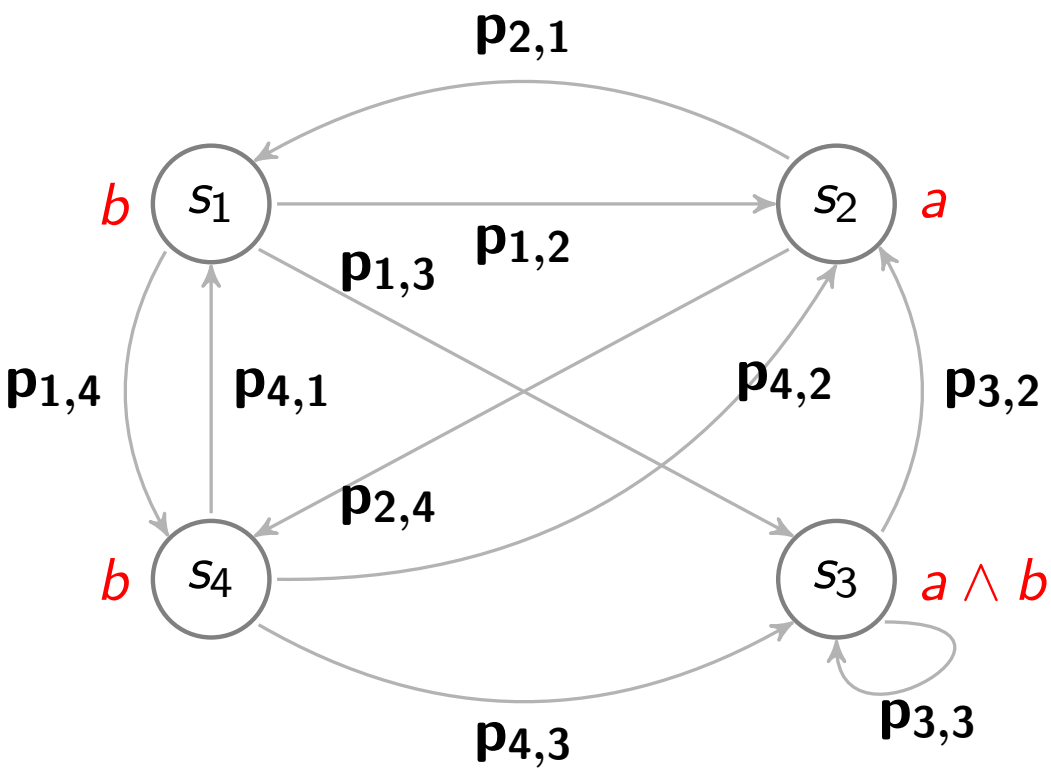


Figure : DTMC representation

Probabilistic Computation Tree Logic (PCTL)

$$\phi ::= \mathit{true} \mid a \mid \neg\phi \mid \phi \wedge \phi \mid P_{\bowtie\gamma}[\psi]$$
$$\psi ::= \mathcal{X}\phi \mid \phi\, \mathcal{U}^{\leq k} \phi$$

- $a \in AP$
- $\bowtie \in \{<, \leq, >, \geq\}$
- $\gamma \in [0, 1]$ a threshold probability
- $k \in \mathbb{N} \cup \{+\infty\}$ (bounded and standard until)

Semantics of the P operator

Let $Prob_M(s, \psi)$ denote the probability that a random path σ in M starting from s ($\sigma[0] = s$) satisfies ψ, i.e. $\sigma \models \psi$.

$$s \models P_{\bowtie \gamma}[\psi] \iff Prob_M(s, \psi) \bowtie \gamma$$

for an IDTMC:

$$\mathbf{M}, s \models \phi \iff \forall M \in \mathbf{M} : M, s \models \phi \ .$$

- Verifying PCTL properties over IDTMCs is known to be an **NP-hard problem.**

Model Checking over a DTMC

$\mathcal{X}$ property: $\psi = \mathcal{X}\phi$

$$Prob_M(s_i, \mathcal{X}\phi) = \sum_{s_j \models \phi} p_{ij}$$

$\mathcal{U}$ property: $\psi = \phi_1 \, \mathcal{U}^{\leq k} \phi_2$

- $S_{yes} \stackrel{\text{def}}{=} \{s_i \mid s_i \models \phi_2\}$,
- $S_{no} \stackrel{\text{def}}{=} \{s_i \mid s_i \not\models \phi_1 \wedge s_i \not\models \phi_2\}$,
- $S_{maybe} \stackrel{\text{def}}{=} S \setminus (S_{yes} \cup S_{no})$.

- If $s_i \in S_{yes}$, then $Prob_M(s_i, \psi) = 1$.
- If $s_i \in S_{no}$, then $Prob_M(s_i, \psi) = 0$.

Model Checking over a DTMC (Cont'd)

Let $v_k[i] \stackrel{\text{def}}{=} Prob_M(s_i, \psi, k)$, then

$$\begin{aligned} v_k[i] &= \sum_{j=1}^{n} p_{ij} v_{k-1}[j] \\ &= \sum_{j \in I_{maybe}} p_{ij} v_{k-1}[j] + \underbrace{\sum_{j \notin I_{maybe}} p_{ij} v_{k-1}[j]}_{b_i} \quad . \end{aligned}$$

$v_{k-1}[j]$ are known for $j \notin I_{maybe}$ (either 0 or 1).

$$v_k = P' v_{k-1} + b,$$

The square matrix P' is extracted from P such that: for all i such that $s_i \in S_{yes} \cup S_{no}$, we delete the ith row and the ith column.

Example

- $\mathbf{M} = (S, s_1, \mathbf{P}, \ell)$
- $S = \{s_1, s_2, s_3, s_4\}$
- $AP = \{a, b\}$
- s_1 is initial state
- $\ell(s_1) = \{b\}$, $\ell(s_2) = \{a\}$, $\ell(s_3) = \{a \wedge b\}$, $\ell(s_4) = \{b\}$

$$P = \begin{bmatrix} 0 & 0.5 & 0.1 & 0.4 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0.8 & 0.2 & 0 \\ 0.5 & 0.3 & 0.2 & 0 \end{bmatrix}$$

Example (Cont'd)

- $P_{\leq\gamma}[b\, \mathcal{U}^{\leq 2}(a \wedge b)]$
- $S_{yes} = \{s_3\}$, $S_{no} = \{s_2\}$ and $S_{maybe} = \{s_1, s_4\}$

$$P' = \begin{bmatrix} 0 & 0.4 \\ 0.5 & 0 \end{bmatrix} \text{ and } b = (0.1, 0.2)^t$$

$$\begin{bmatrix} Prob_M(s_1, \psi) \\ Prob_M(s_4, \psi) \end{bmatrix} = \begin{bmatrix} 0.18 \\ 0.25 \end{bmatrix}$$

Extension to IDTMCs

Sample probability transition relation for IDTMC

$$P = \begin{bmatrix} 0 & [0.49, 0.51] & [0.09, 0.11] & [0.39, 0.41] \\ [0.49, 0.51] & 0 & 0 & [0.49, 0.51] \\ 0 & [0.79, 0.81] & [0.19, 0.21] & 0 \\ [0.49, 0.51] & [0.29, 0.31] & [0.19, 0.21] & 0 \end{bmatrix}$$

Analysis using Interval Arithmetic

$$\mathbf{v_k} = \mathbf{P'v_{k-1}} + \mathbf{b}$$

Successive computation inherits from the loss of precision due to interval arithmetic

To overcome this loss of precision, in the bounded case, we use **affine arithmetic**

Affine Forms

Interval Analysis Problem: Compute $x - x$

$[a, b] - [a, b] = [a - b, b - a] \supset [0, 0]$

In AA, the interval $[a, b]$ is represented using the affine expression:

$$\frac{a+b}{2} + \frac{b-a}{2}\epsilon_1,$$

$\epsilon_1 \in [-1, 1]$ is introduced to capture the uncertainty.

$$\hat{a} \stackrel{\text{def}}{=} \alpha_0^a + \alpha_1^a \epsilon_1 + \cdots + \alpha_l^a \epsilon_l = \alpha_0^a + \sum_{i=1}^{l} \alpha_i^a \epsilon_i,$$

- $\alpha_0^a, \ldots, \alpha_l^a$ are real coefficients (error weights).
- $\epsilon_1, \ldots, \epsilon_l$ are symbolic error variables.

Affine Arithmetic

- $\hat{a}$ and $\hat{b}$ are two affine forms
- λ, ζ be two finite real numbers

Linear Operations

$$\hat{a} \pm \hat{b} \stackrel{\text{def}}{=} (\alpha_0^a \pm \alpha_0^b) + \sum_{i=1}^{l} (\alpha_i^a \pm \alpha_i^b)\epsilon_i$$

$$\lambda\hat{a} \stackrel{\text{def}}{=} \lambda\alpha_0^a + \sum_{i=1}^{l} (\lambda\alpha_i^a)\epsilon_i$$

$$\hat{a} + \zeta \stackrel{\text{def}}{=} (\alpha_0^a + \zeta) + \sum_{i=1}^{l} \alpha_i^a \epsilon_i$$

Model Checking IDTMC

Main idea

Split **P** into a central matrix P_c, and an interval matrix **E**, which encodes the uncertainty of the model: $\mathbf{P} = P_c + \mathbf{E}$

- Matrix P_c is stochastic (all rows sum up to 1) *in our case*
- The matrix **E** is represented using AA error terms

Thus, the equation for DTMC analysis $v_k = P' v_{k-1} + b$ becomes:

$$v_k(\epsilon) = (P'_c + E'(\epsilon))v_{k-1}(\epsilon) + (b + b(\epsilon))$$

The updated components of $v_k(\epsilon)$ are non-linear (polynomial) functions of the perturbations $(\epsilon_{ij})_{1 \leq i,j \leq n}$.

Combining AA and IA

Overapproximation

Split non-linear component computation of $v_k(\epsilon)$ into three parts:

- a constant value c_k
- $l_k(\epsilon)$ is the linear part of $v_k(\epsilon)$ using AA
- $\square_k$ is an IA-overapproximation of $v_k(\epsilon) - (c_k + l_k(\epsilon))$

$v_k(\epsilon) \in \tilde{\mathbf{P}}_k \stackrel{\text{def}}{=} c_k + l_k(\epsilon) + \square_k$

$$
\begin{aligned}
c_k &= P_c' c_{k-1} + b \\
l_k(\epsilon) &= P_c' l_{k-1}(\epsilon) + E'(\epsilon) c_{k-1} + b(\epsilon) \\
\square_k &= P_c' \square_{k-1} + \mathbf{E}'(\square_{k-1} + \mathbf{I}_{k-1})
\end{aligned}
$$

We still need to compute $\square_k$: that is evaluate $\mathbf{I}_{k-1}$.
$\mathbf{I}_{k-1}$ contains component-wise wrapping interval bounds for $l_{k-1}(\epsilon)$.

Computing $\mathbf{I}_{k-1}$

- For each component of the n-dimensional interval-vector $\mathbf{I}_{k-1}$:

$$\begin{aligned} \max / \min \quad & \sum_{1 \le i,j \le n} \alpha_{ij}\epsilon_{ij} \\ \text{s.t.} \quad & -e_{ij} \le \epsilon_{ij} \le e_{ij}, 1 \le i,j \le n \\ & \sum_{j=1}^{n} p'_{c\,ij} + \epsilon_{ij} = 1, 1 \le i \le n \end{aligned} \tag{LP}$$

- the feasible region is not empty for a *normalized* IDTMC
- any off-the-shelf LP solver can be used

Specificity of learned IDTMC

- The matrix P_c is stochastic (rows sum up to 1),
- which makes $\sum_{j=1}^{n} \epsilon_{ij} = 0, 1 \leq i \leq n$

$\Longrightarrow$ It turns out that under these assumptions, we need to only sort affine error weights to compute $\mathbf{l}_{k-1}$ (see next slide)

- In fact: it can be done in linear time by reduction to *weighted median problem* (see paper)

Saturation

Lemma

Given a linear programming problem of the form of (LP), there exists a feasible maximizing solution that leaves at most one variable non-saturated. All other variables are positively or negatively saturated.

It is then sufficient to determine:

- the non-saturated index, say k
- the set $\oplus$ of positively saturated variables
- the set $\ominus$ of negatively saturated variables

The value of ϵ_k is then determined by

$$\epsilon_k = -\sum_{i \in \ominus \cup \oplus} \epsilon_i = \sum_{i \in \ominus} \epsilon_i - \sum_{i \in \oplus} \epsilon_i \ .$$

(unbounded) Until properties - DTMC

Fixpoint formulation:

$$v = P'v + b$$

Proposition

Let A be a square matrix of dimension $n \times n$ such that

- $\forall i, j, 1 \leq i, j \leq n, a_{ij} \in [0, 1]$
- $\forall i, 1 \leq i \leq n, 0 < \sum_{j=1}^{n} a_{ij} \leq 1$
- $\exists i, 1 \leq i \leq n, \sum_{j=1}^{n} a_{ij} < 1$

Let I_n denote the identity matrix of dimension n. Then the matrix $A - I_n$ is invertible.

$\Longrightarrow$ Therefore $v = (I - P')^{-1} b$

(unbounded) Until properties - IDTMC

Fixpoint formulation

$$c = P'_c c + b$$
$$l(\epsilon) = P'_c l(\epsilon) + E'(\epsilon)c + b(\epsilon)$$
$$\square = P'_c \square + \mathbf{E}'(\square + \mathbf{l})$$

As for DTMCs, we derive c and $l(\epsilon)$ as follows:

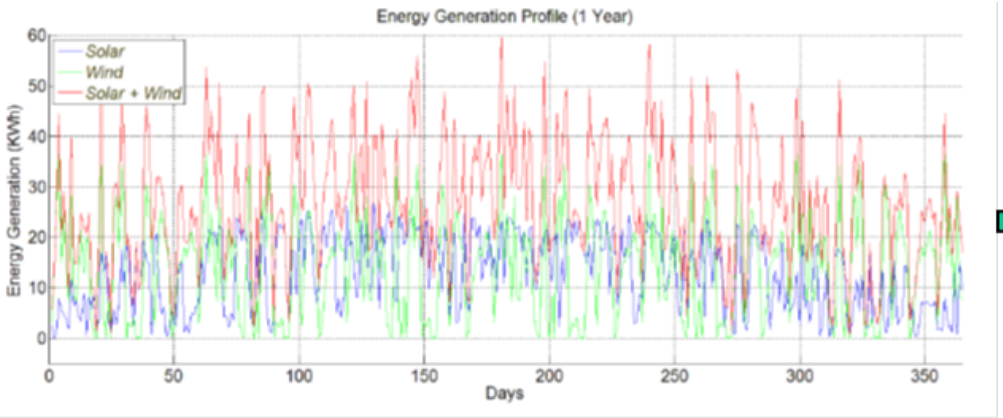
$$c = (I - P'_c)^{-1} b$$
$$l(\epsilon) = (I - P'_c)^{-1}(E'(\epsilon)c + b(\epsilon))$$

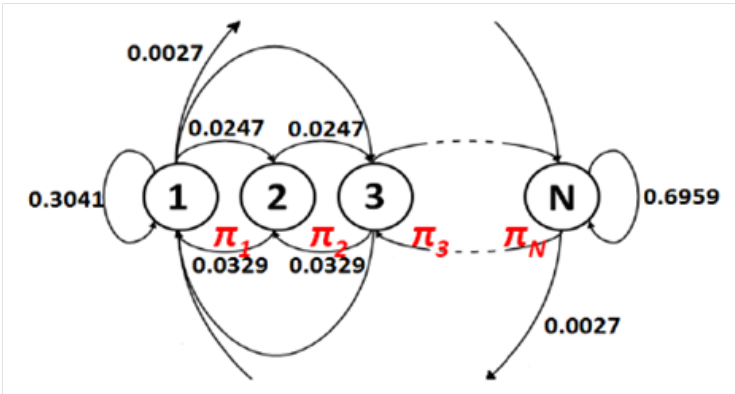
and compute an overapproximation of $\square$

$$(I - P'_c - \mathbf{E'})\square = \mathbf{E'l}$$

Smart Grid Management System

- Data collected for renewable energy sources (wind, solar)
- Fluctuations in demand and supply modeled as Markov chain





- Instead: We learned IDTMC and performed analysis

Smart Grid Management System (cont.)

Table : IA versus AA+LP

	# Days	IA	AA+LP
P_1	7	$[0.55, 1]$	$[0.83, 0.98]$
P_2	7	$[0.35, 1]$	$[0.70, 0.80]$

P_1: What is the probability that within k days, the power grid will switch from high supply mode to low supply mode: $P[\frac{1}{2}\delta_M \leq \delta \leq \delta_M \; \mathcal{U}^{\leq k} 0 \leq \delta \leq \frac{1}{2}\delta_M]$.

P_2: What is the probability that within k days, the power grid will switch from low supply mode to low demand mode: $P[0 \leq \delta \leq \frac{1}{2}\delta_M \; \mathcal{U}^{\leq k} \frac{1}{2}\delta_m \leq \delta \leq 0]$.

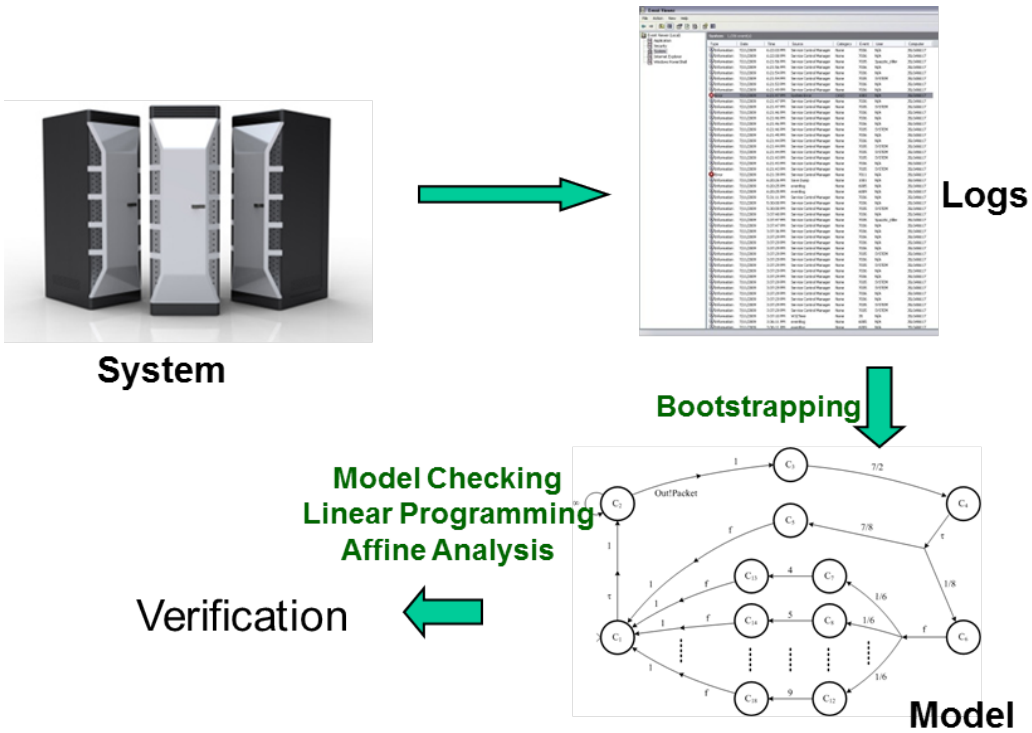
Conclusion and Future Work

Conclusion

- Efficient computation of simple reachability properties over IDTMC.
- Exact propagation of first order error terms.

Future work

- The propagation of first order error terms allow witness generation.
- Extension to nested and multiple P operators.



Thank you for your attention!

Questions???

HSCC 2013 (part of *CPSWeek 2013*)

- **Submission deadline: October 15th, 2012 (strict!)**
- `http://2013.hscc-conference.org`