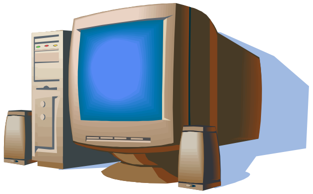
Decision-Theoretic Planning with Asynchronous Events

Håkan L. S. Younes
Carnegie Mellon University

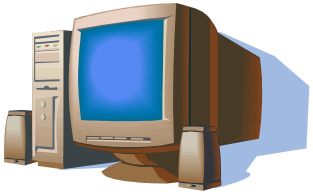
Introduction

- Asynchronous processes are abundant in the real world
- Discrete-time models are inappropriate for systems with asynchronous events
- Generalized semi-Markov (decision) processes are great for this!

Stochastic Processes with Asynchronous Events



m_1

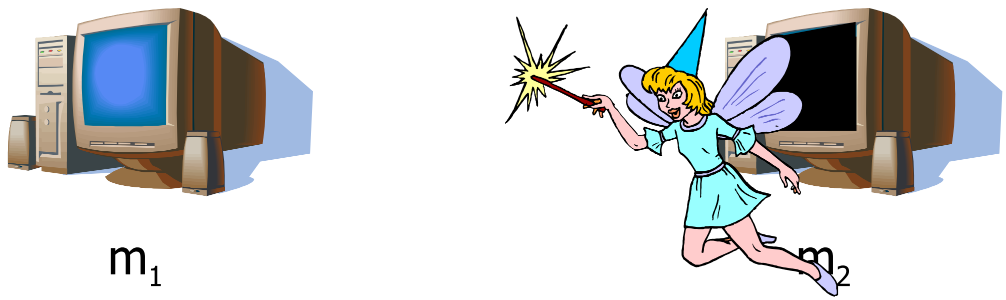


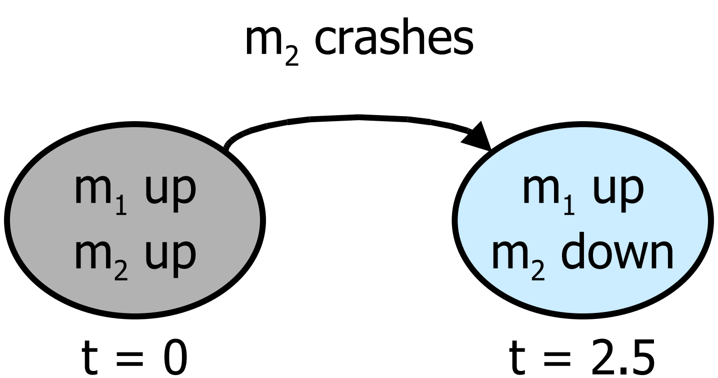
m_2

m_1 up
m_2 up

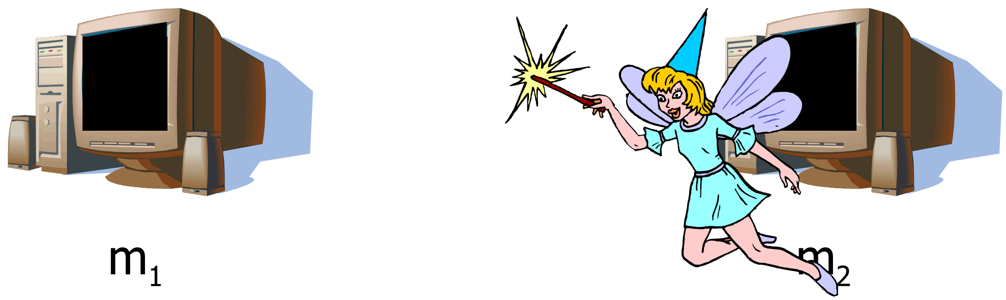
$t = 0$

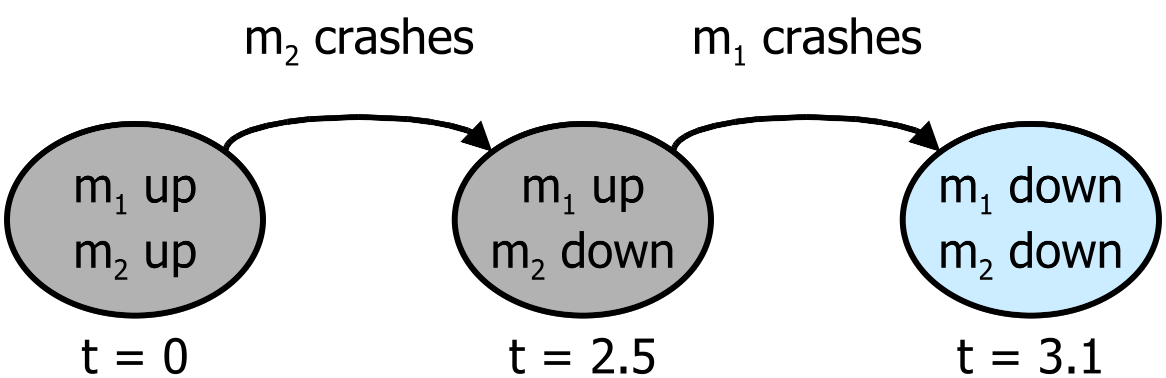
Stochastic Processes with Asynchronous Events



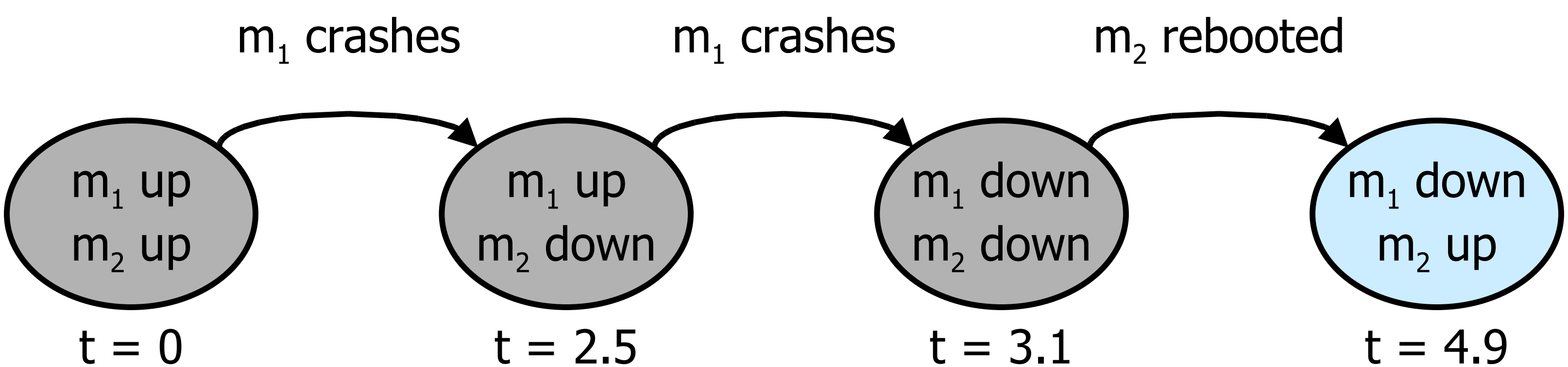


Stochastic Processes with Asynchronous Events





Stochastic Processes with Asynchronous Events



A Model of Stochastic Discrete Event Systems

- Generalized semi-Markov process (GSMP) [Matthes 1962]
  - A set of events E
  - A set of states S

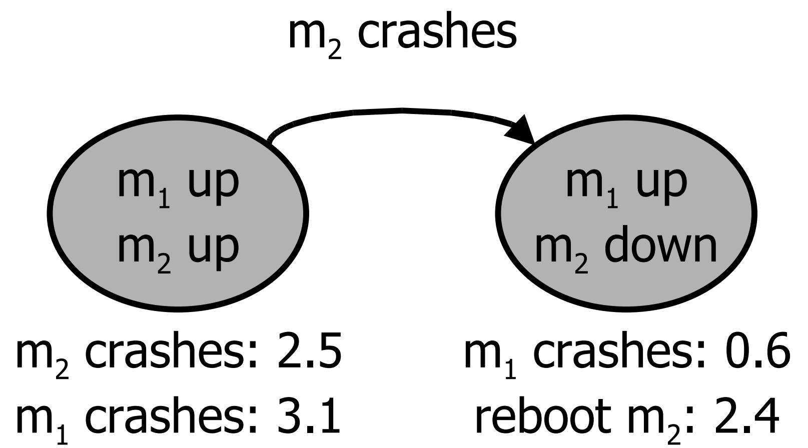
Events

- In a state s, events $E_s \subset E$ are enabled
- With each event e is associated:
  - A distribution G_e governing the time e must remain enabled before it triggers
  - A next-state probability distribution $p_e(s'|s)$

Semantics of GSMP Model

- Associate a real-valued clock t_e with e
- For each $e \in E_s$ sample t_e from G_e
- Let $e^* = \mathrm{argmin}_{e \in E_s} t_e$, $t^* = \min_{e \in E_s} t_e$
  - Sample s' from $p_{e^*}(s'|s)$
  - For each $e \in E_{s'}$
    - $t_e' = t_e - t^*$ if $e \in E_s \setminus \{e^*\}$
    - sample t_e' from G_e otherwise
- Repeat with $s = s'$ and $t_e = t_e'$

Semantics: Example



Notes on Semantics

- Events that remain enabled across state transitions without triggering are not rescheduled
  - Asynchronous events!
  - Differs from semi-Markov process in this respect
  - Continuous-time Markov chain if all G_e are exponential distributions

General State-Space Markov Chain (GSSMC)

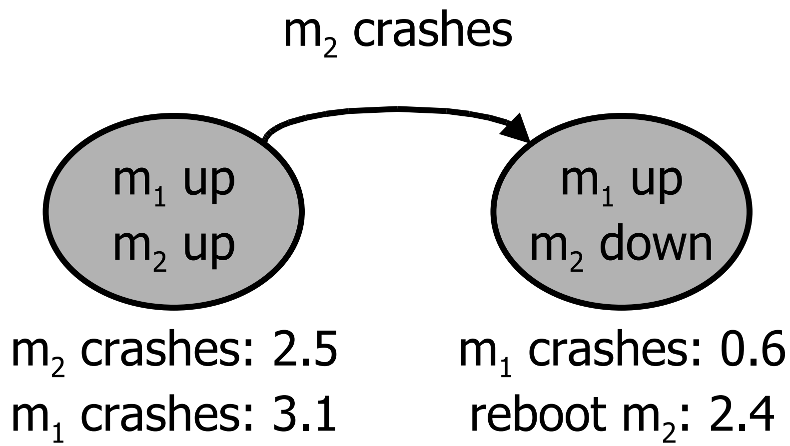
- Model is Markovian if we include the clocks in the state space
  - Extended state space X
  - Next-state distribution $f(x'|x)$ well-defined
- Clock values are not know to observer
  - Time events have been enabled *is* know

Observation Model

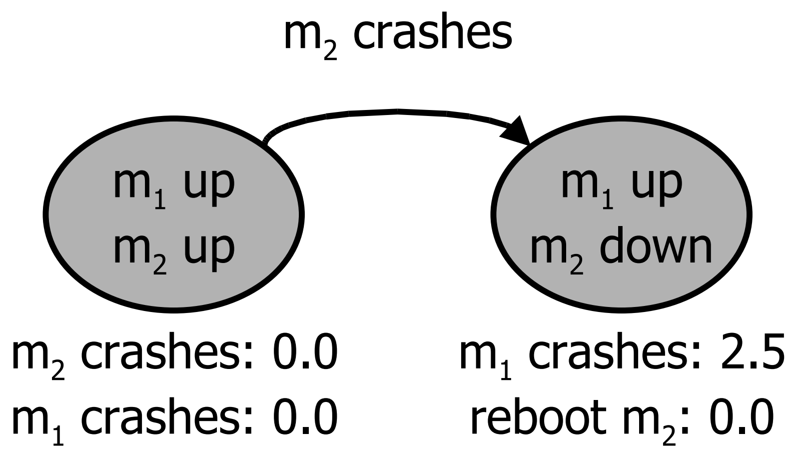
- An observation o is a state s and a real value u_e for each event representing the time e has currently been enabled
  - $f(x|o)$ is well-defined

Observations: Example

Actual model:



Observed model:



Actions and Policies (GSM**D**Ps)

- Identify a set $A \subset E$ of controllable events (actions)
- A policy is a mapping from observations to sets of actions
  - Action choice can change at any time in a state s

Rewards and Discounts

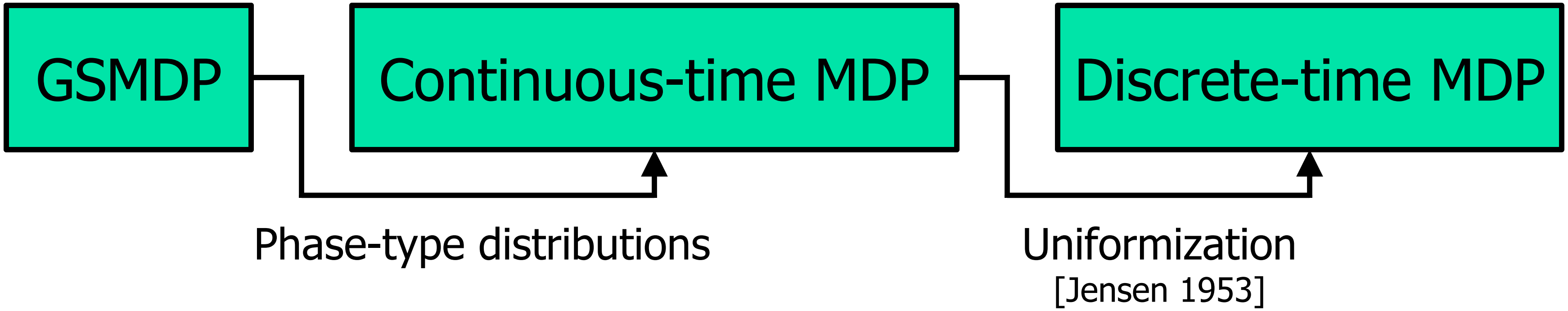
- Lump sum reward $k(s,e,s')$ associated with transition from s to s' caused by e
- Continuous reward rate $r(s,a)$ associated with a being enabled in s
- Discount factor α
  - Unit reward earned at time t counts as $e^{-\alpha t}$

Value Function for GSMDPs

$$V_\pi(o) = \int_X f(x \mid o)\left(\int_0^{t^*} e^{-\alpha t} c(s,\pi(o))dt + e^{-\alpha t^*}\int_X f(x' \mid x,o)\big(k(s,e^*,s') + V_\pi(obs(x,o,s'))\big)dx'\right)dx$$

$$= \int_X f(x \mid o)\left(\frac{1}{\alpha}\left(1-e^{-\alpha t^*}\right)c(s,\pi(o)) + e^{-\alpha t^*}\sum_{s' \in S} p_{e^*}(s' \mid s)\big(k(s,e^*,s') + V_\pi(obs(x,o,s'))\big)\right)dx$$

GSMDP Solution Method [Younes & Simmons 2004]



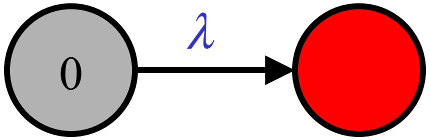
Continuous Phase-Type Distributions [Neuts 1981]

- Time to absorption in a continuous-time Markov chain with n transient states

$$\Pr[X \le t] = 1 - \vec{\alpha} e^{\mathbf{Q}t} \vec{e}$$

$$\vec{e} = \left.\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}\right\} n \text{ rows}$$

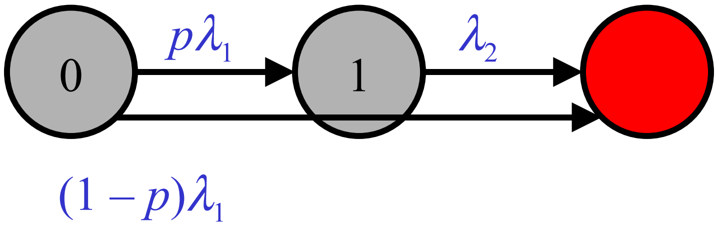
Exponential Distribution



$$\vec{\alpha} = [1] \qquad \mathbf{Q} = [-\lambda]$$

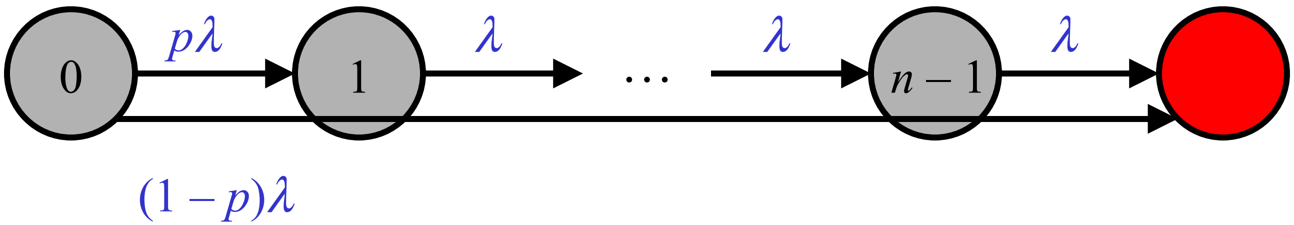
$$\Pr[X \leq t] = 1 - e^{-\lambda t}$$

Two-Phase Coxian Distribution



$$\vec{\alpha} = \begin{bmatrix} 1 & 0 \end{bmatrix} \qquad \mathbf{Q} = \begin{bmatrix} -\lambda_1 & p\lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}$$

Generalized Erlang Distribution



$$\vec{\alpha} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \qquad \mathbf{Q} = \begin{bmatrix} -\lambda & p\lambda & 0 & \cdots & 0 \\ 0 & -\lambda & \lambda & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\lambda & \lambda \\ 0 & \cdots & 0 & 0 & -\lambda \end{bmatrix}$$

Method of Moments

- Approximate general distribution G with phase-type distribution PH by matching the first n moments

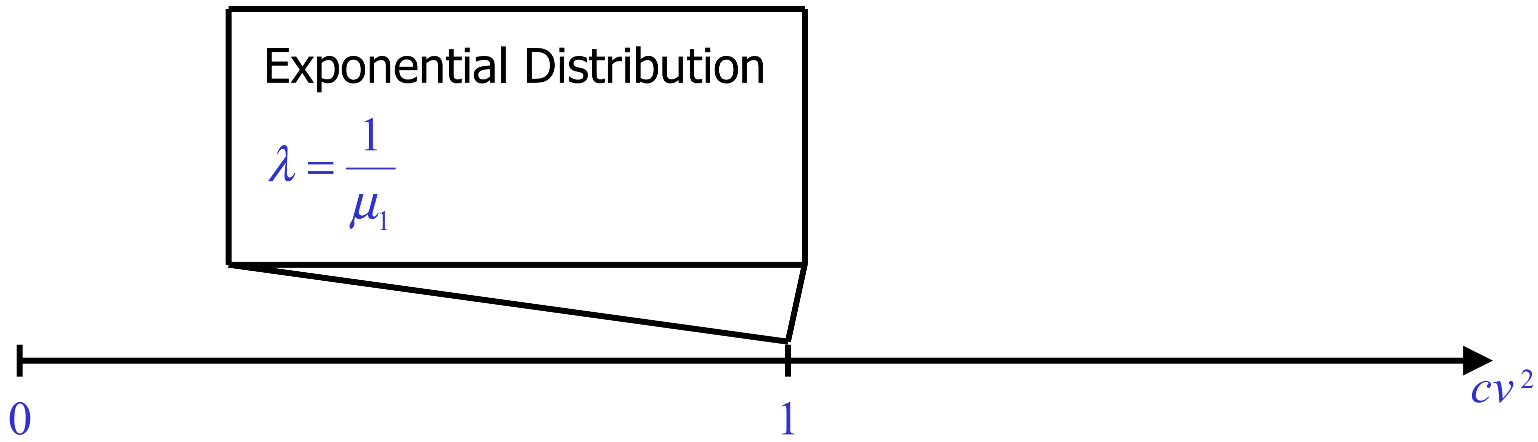
Moments of a Distribution

- The *i*th moment: $\mu_i = \mathrm{E}[X^i]$
  - Mean: μ_1
  - Variance: $\sigma^2 = \mu_2 - \mu_1^{\ 2}$
  - Coefficient of variation: $cv = \sigma / \mu_1$

Matching One Moment

- Exponential distribution: $\lambda = 1/\mu_1$

Matching Two Moments



Matching Two Moments

Exponential Distribution

$$\lambda = \frac{1}{\mu_1}$$

0 — 1 — cv^2

Generalized Erlang Distribution

$$n = \left\lceil \frac{1}{cv^2} \right\rceil \quad p = 1 - \frac{2n \cdot cv^2 + n - 2 - \sqrt{n^2 + 4 - 4n \cdot cv^2}}{2(n-1)(cv^2+1)}$$

$$\lambda = \frac{1 - p + np}{\mu_1}$$

Matching Two Moments

Exponential Distribution

$$\lambda = \frac{1}{\mu_1}$$

Two-Phase Coxian Distribution

$$p = \frac{1}{2 \cdot cv^2} \qquad \lambda_1 = \frac{2}{\mu_2} \qquad \lambda_2 = \frac{1}{\mu_1 \cdot cv^2}$$

0 1 cv^2

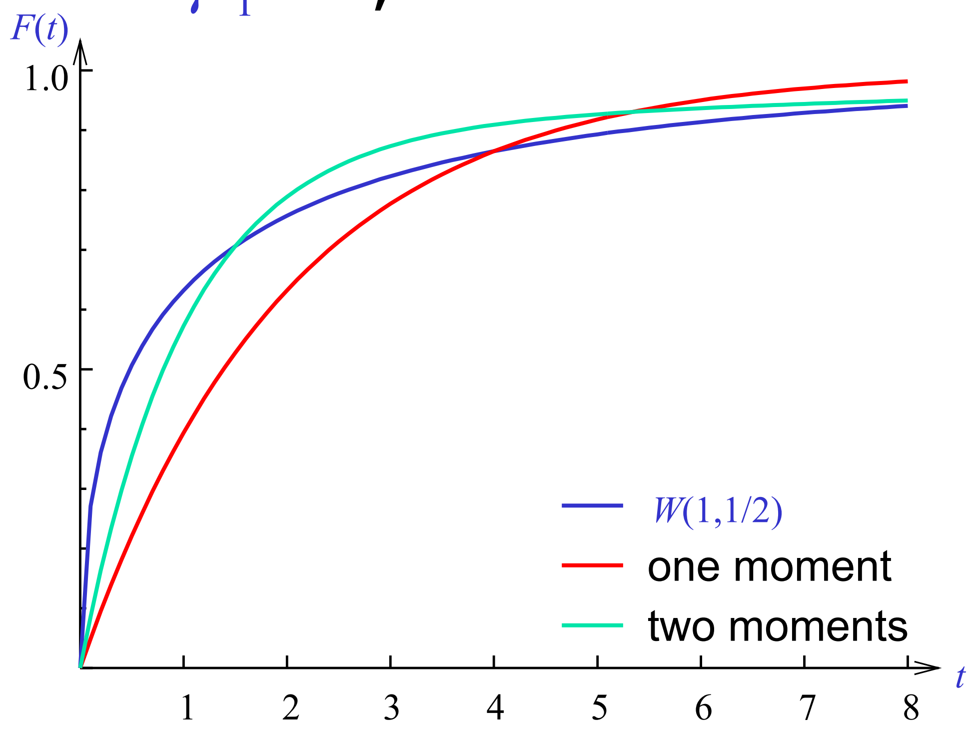
Generalized Erlang Distribution

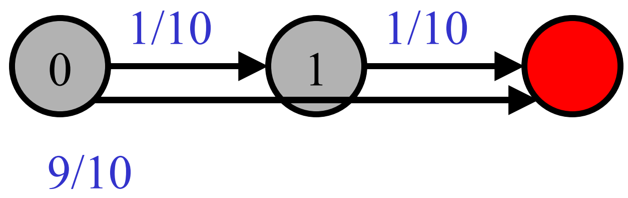
$$n = \left\lceil \frac{1}{cv^2} \right\rceil \qquad p = 1 - \frac{2n \cdot cv^2 + n - 2 - \sqrt{n^2 + 4 - 4n \cdot cv^2}}{2(n-1)(cv^2+1)}$$

$$\lambda = \frac{1 - p + np}{\mu_1}$$

Matching Moments: Example 1

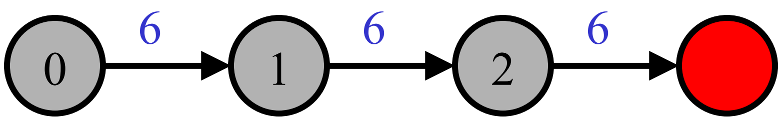
- Weibull distribution: $W(1,1/2)$
  - $\mu_1 = 2$, $cv^2 = 5$

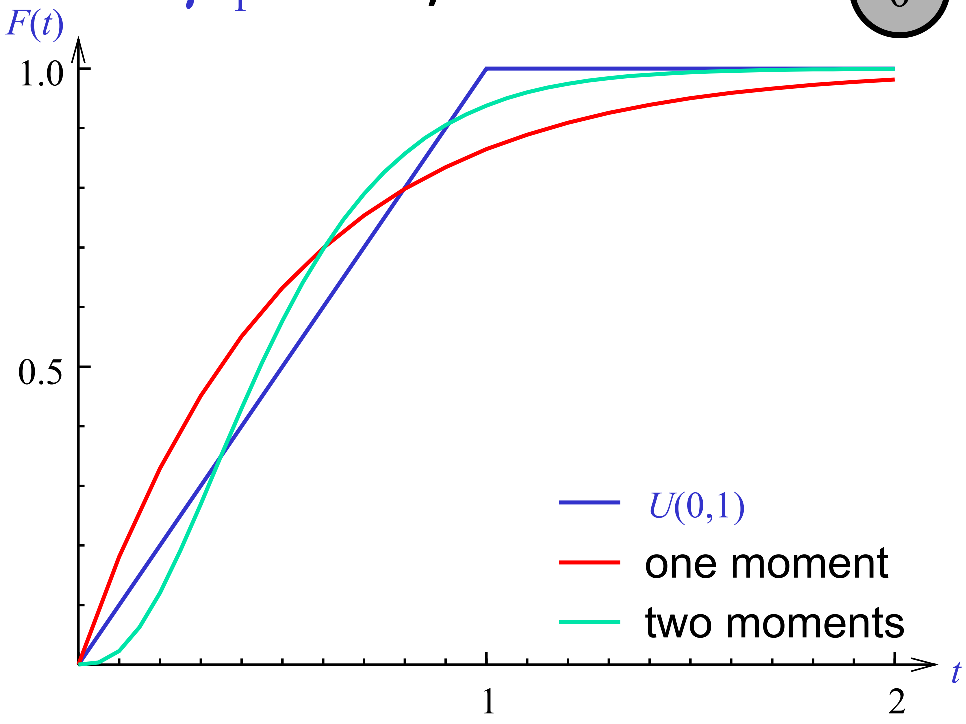




Matching Moments: Example 2

- Uniform distribution: $U(0,1)$
  - $\mu_1 = 1/2$, $cv^2 = 1/3$





Matching More Moments

- Closed-form solution for matching three moments of positive distributions
  [Osogami & Harchol-Balter 2003]
  - Combination of Erlang distribution and two-phase Coxian distribution

Approximating GSMDP with Continuous-time MDP

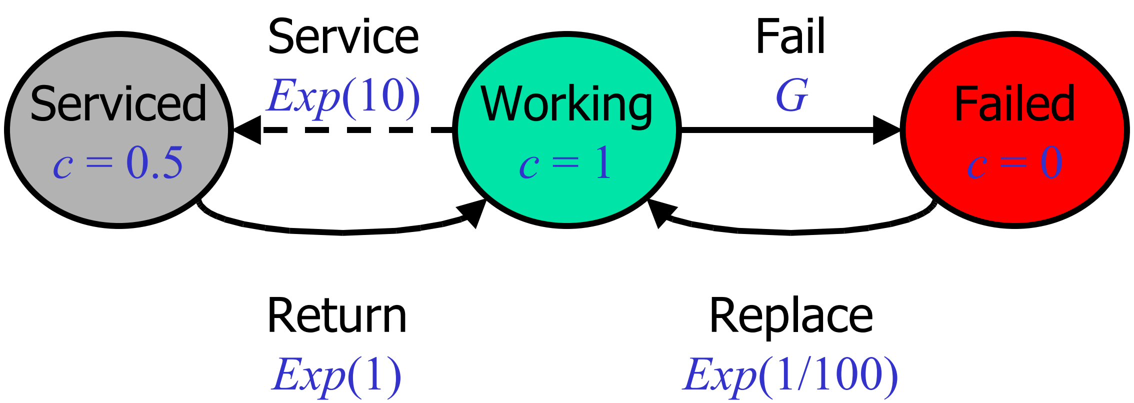
- Each event with a non-exponential distribution is approximated by a set of events with exponential distributions
  - Phases become part of state description

Policy Execution

- Phases represent discretization into random-length intervals of the time events have been enabled
- Phases are not part of real model
  - Simulate phase transition during execution

The Foreman's Dilemma

- When to enable "Service" action in "Working" state?



The Foreman's Dilemma: Optimal Solution

- Find t_0 that maximizes v_0

$$v_0 = \int_0^\infty f_X(t)\left(1 - F_Y(t)\right)\left(\left(\frac{1}{\alpha}\left(1 - e^{-\alpha t}\right) + e^{-\alpha t} v_1\right)\right) + f_Y(t)\left(1 - F_X(t)\right)\left(\frac{1}{\alpha}\left(1 - e^{-\alpha t}\right) + e^{-\alpha t} v_2\right) dt$$

$$v_1 = \frac{1}{1 + 100\alpha} v_0 \qquad v_2 = \frac{1}{1 + \alpha}\left(\frac{1}{2} + v_0\right)$$

$$f_X(t) = \begin{cases} 0 & t < t_0 \\ 10e^{-10(t - t_0)} & t \geq t_0 \end{cases}$$

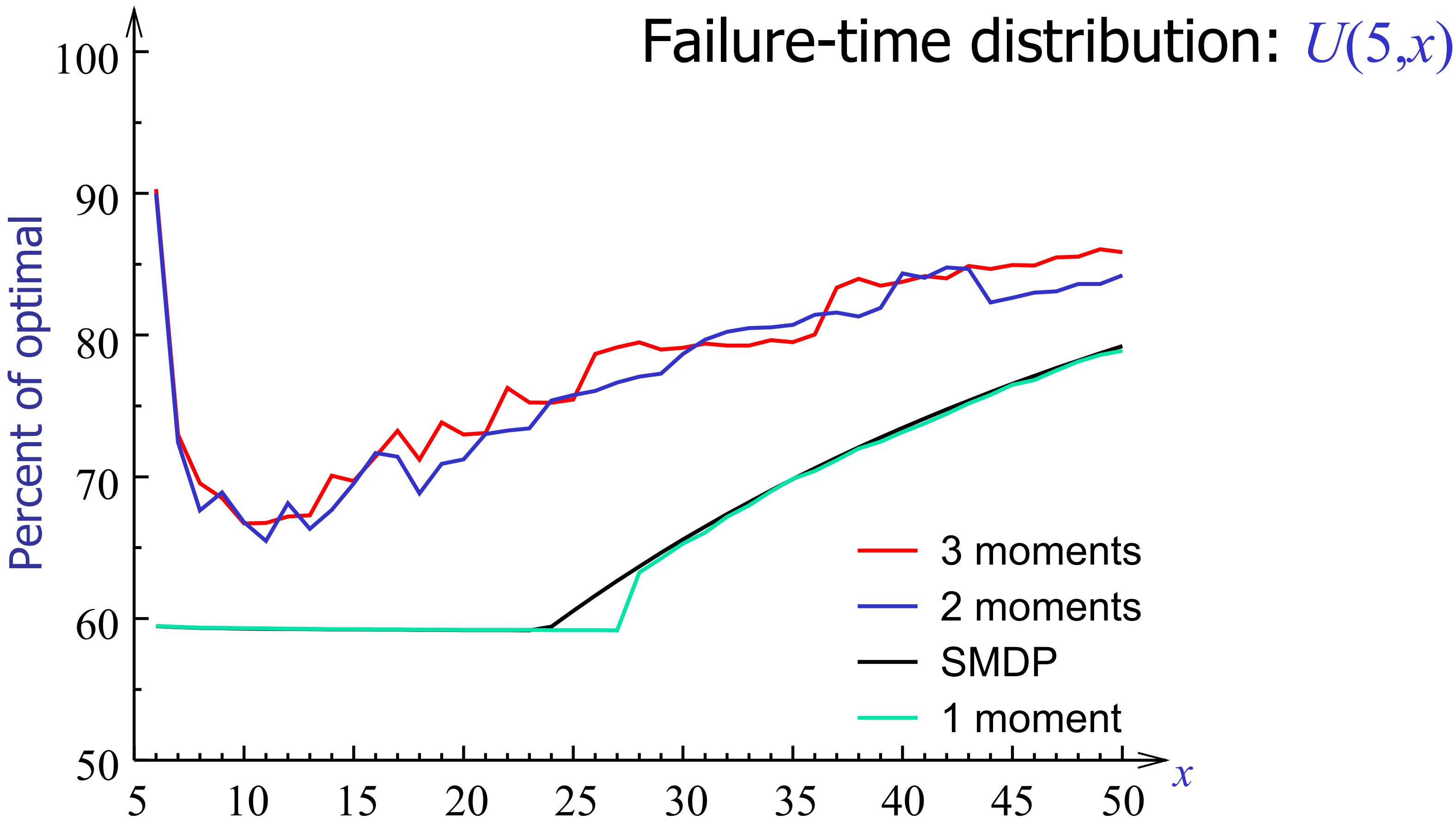
$$F_X(t) = \int_0^t f_X(x)\,dx$$

Y is the time to failure in "Working" state

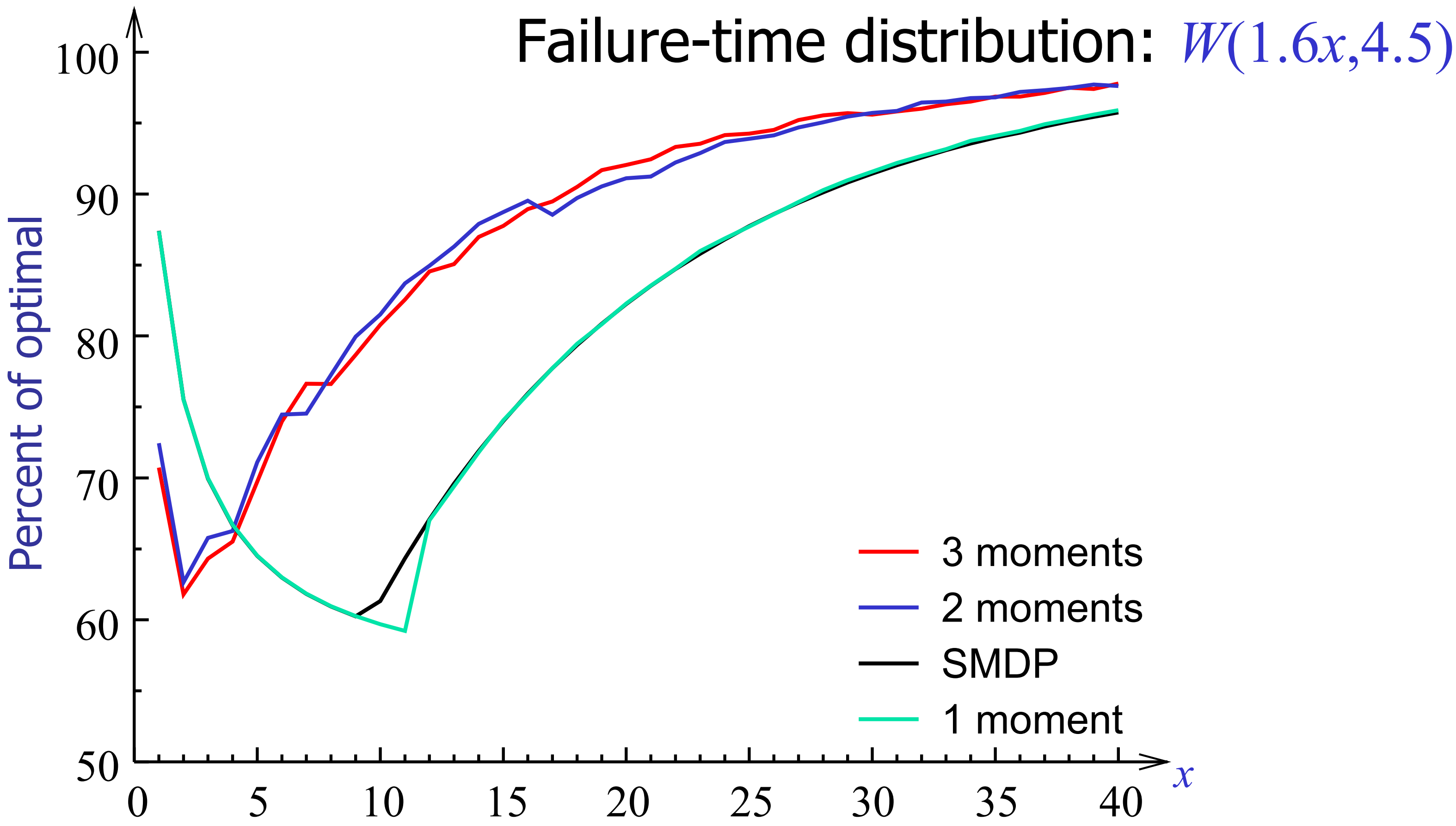
The Foreman’s Dilemma: SMDP Solution

- Same formulas, but restricted choice:
  - Action is immediately enabled ($t_0 = 0$)
  - Action is never enabled ($t_0 = \infty$)

The Foreman's Dilemma: Performance



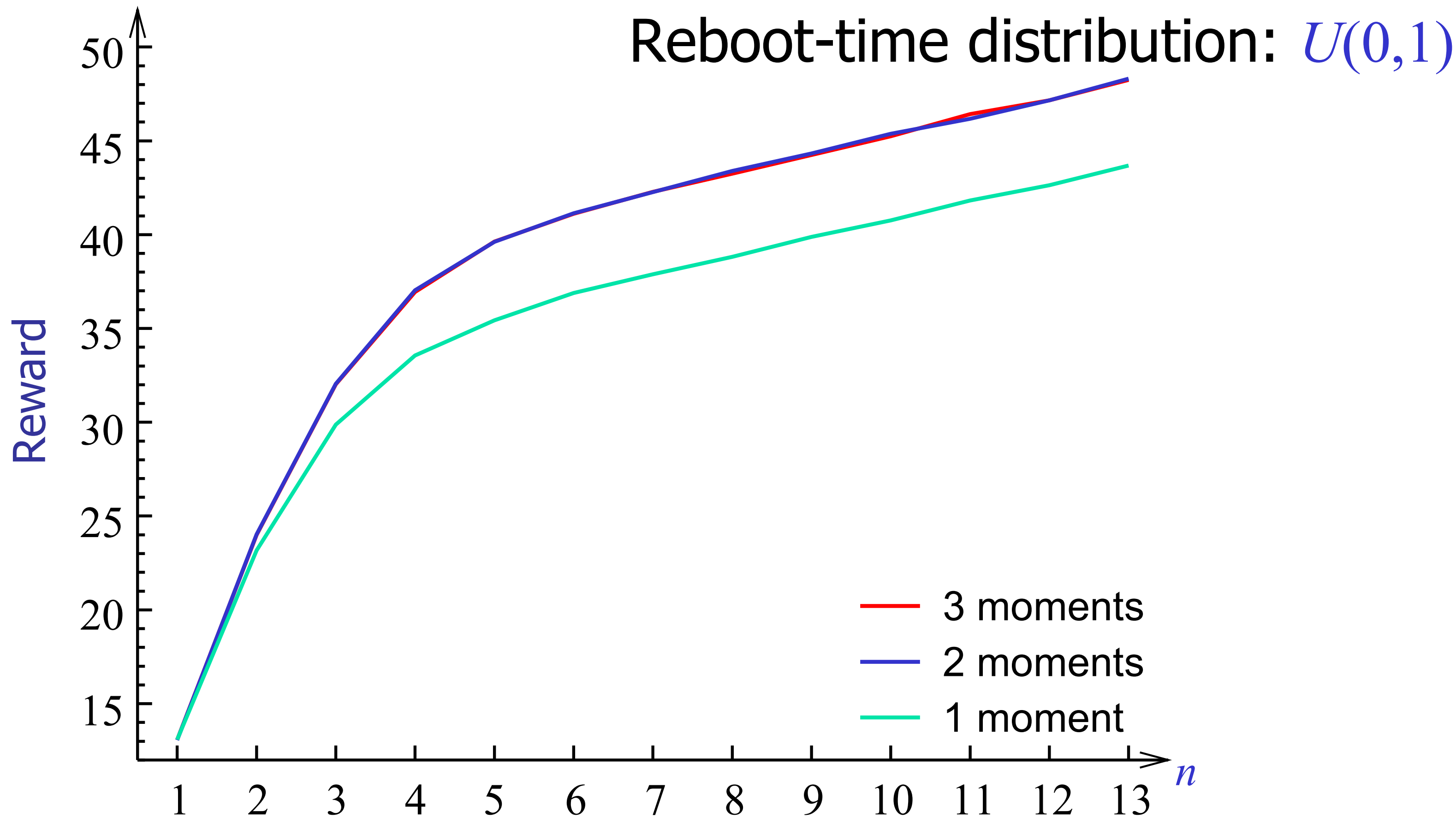
The Foreman's Dilemma: Performance



System Administration

- Network of n machines
- Reward rate $c(s) = k$ in states where k machines are up
- One crash event and one reboot action per machine
  - At most one action enabled at any time

System Administration: Performance



System Administration: Performance

	1 moment			2 moments			3 moments		
size	states	q	time (s)	states	q	time (s)	states	q	time (s)
4	16	5	0.36	32	9	3.57	112	21.0	10.30
5	32	6	0.82	80	10	7.72	272	22.0	22.33
6	64	7	1.89	192	11	16.24	640	23.0	40.98
7	128	8	3.65	448	12	28.04	1472	24.0	69.06
8	256	9	6.98	1024	13	48.11	3328	25.0	114.63
9	512	10	16.04	2304	14	80.27	7424	26.0	176.93
10	1024	11	33.58	5120	15	136.4	16384	27.0	291.70
11	2048	12	66.00	24576	16	264.17	35840	28.0	481.10
12	4096	13	111.96	53248	17	646.97	77824	29.0	1051.33
13	8192	14	210.03	114688	18	2588.95	167936	30.0	3238.16
	2^n			$(n+1)2^n$			$(1.5n+1)2^n$		

The Role of Phases

- Foreman’s dilemma:
  - Phases permit delay in enabling of actions
- System administration:
  - Phases allow us to take into account the time an action has already been enabled

Summary

- Generalized semi-Markov (decision) processes allow asynchronous events
- Phase-type distributions can be used to approximate a GSMDP with an MDP
  - Allows us to approximately solve GSMDPs using existing MDP techniques
- Phase does matter!

Future Work

- Discrete phase-type distributions
  - Handles deterministic distributions
  - Avoids uniformization step
- Value function approximation
  - Take advantage of GSMDP structure
- Other optimization criteria
  - Finite horizon, etc.

Tempastic-DTP

- A tool for GSMDP planning:

  http://www.cs.cmu.edu/~lorens/tempastic-dtp.html