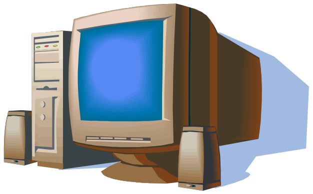
Solving Generalized Semi-Markov Decision Processes using Continuous Phase-Type Distributions

Håkan L. S. Younes Reid G. Simmons
Carnegie Mellon University

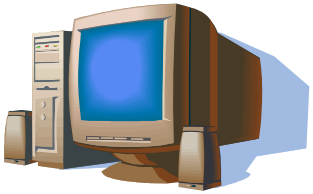
Introduction

- Asynchronous processes are abundant in the real world
  - Telephone system, computer network, etc.
- Discrete-time and semi-Markov models are inappropriate for systems with asynchronous events
- Generalized semi-Markov (decision) processes, GSM(D)Ps, are great for this!
  - Approximate solution using phase-type distributions and your favorite MDP solver

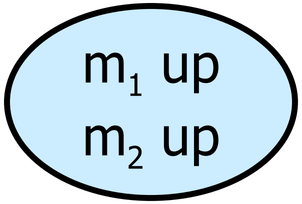
Asynchronous Processes: Example



m_1

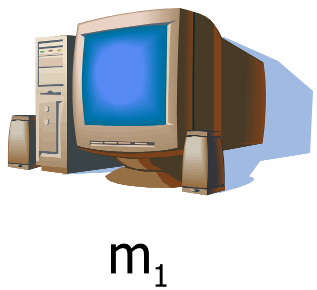


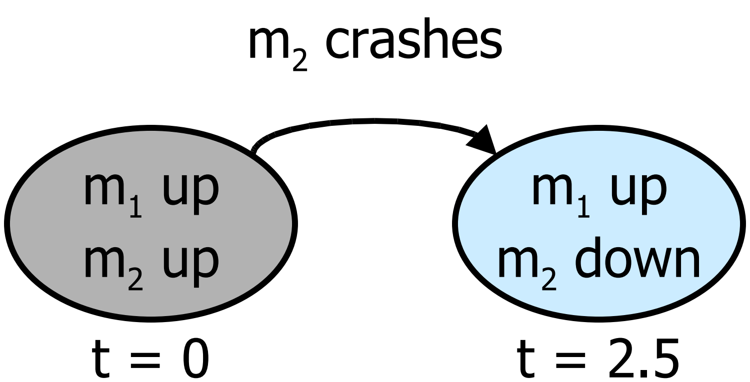
m_2



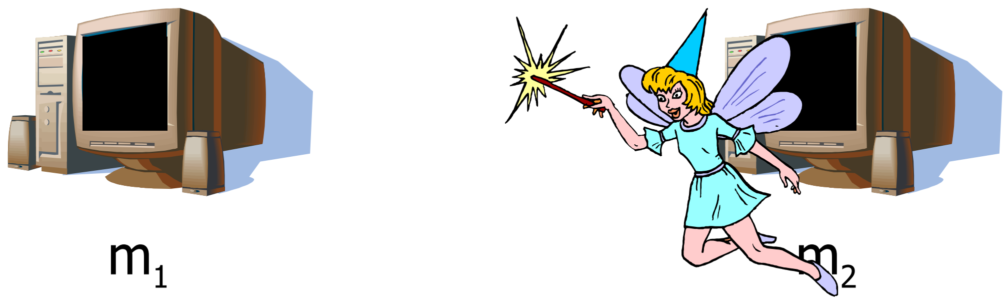
t = 0

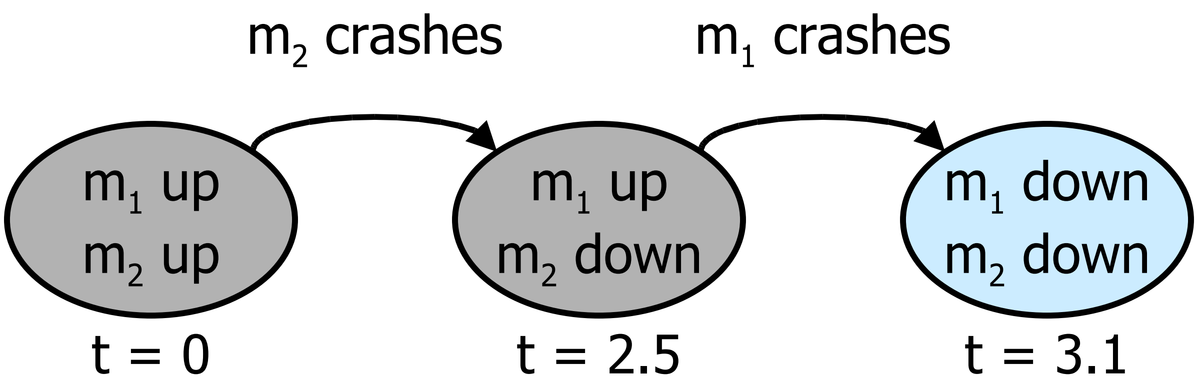
Asynchronous Processes: Example



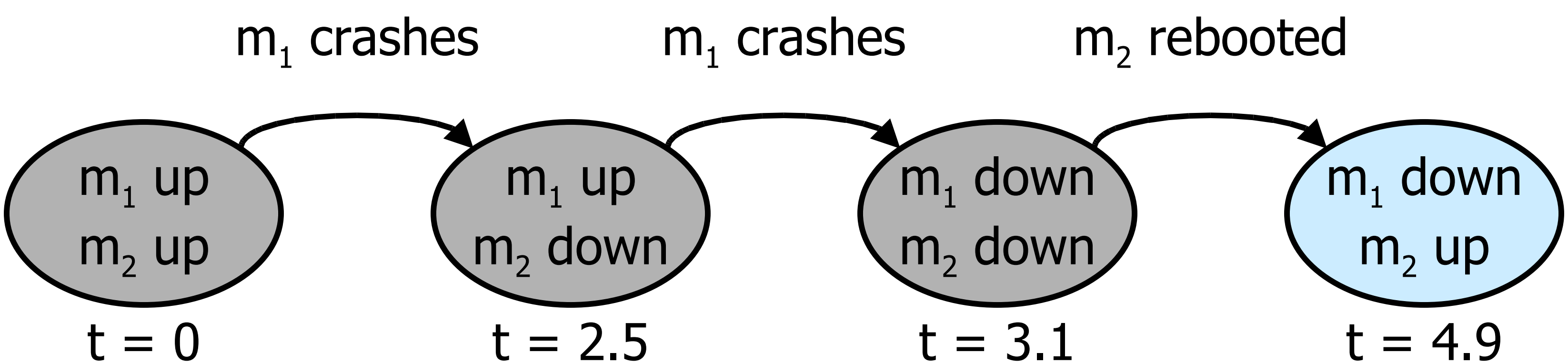


Asynchronous Processes: Example





Asynchronous Processes: Example



A Model of Stochastic Discrete Event Systems

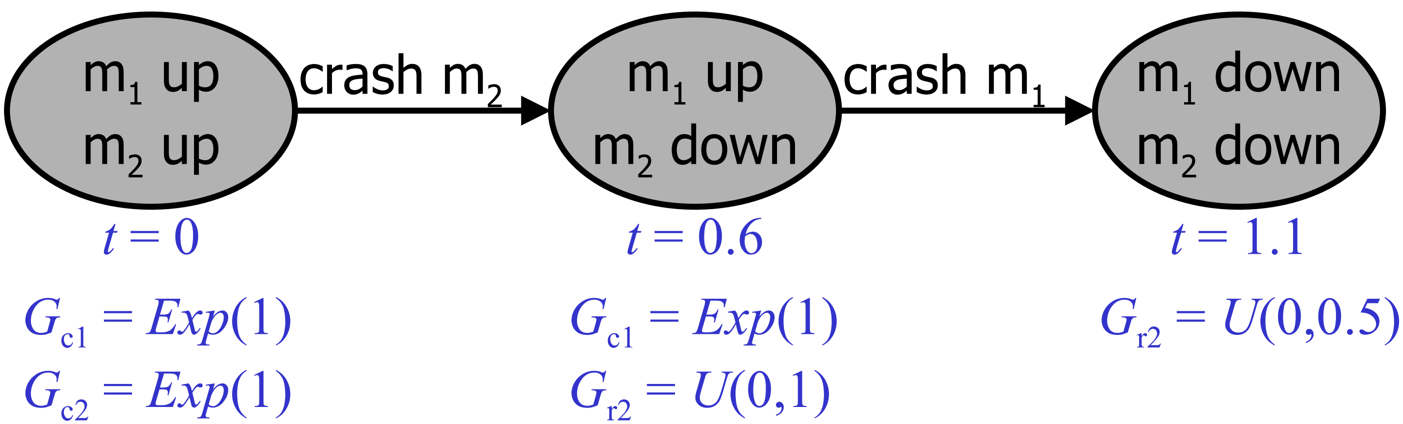
- Generalized semi-Markov process (GSMP) [Matthes 1962]
  - A set of events E
  - A set of states S
- GSMDP
  - Actions $A \subset E$ are controllable events

Events

- With each event e is associated:
  - A condition ϕ_e identifying the set of states in which e is enabled
  - A distribution G_e governing the time e must remain enabled before it triggers
  - A distribution $p_e(s'|s)$ determining the probability that the next state is s' if e triggers in state s

Events: Example

- Network with two machines
  - Crash time: $Exp(1)$
  - Reboot time: $U(0,1)$



Asynchronous events $\Rightarrow$ beyond semi-Markov

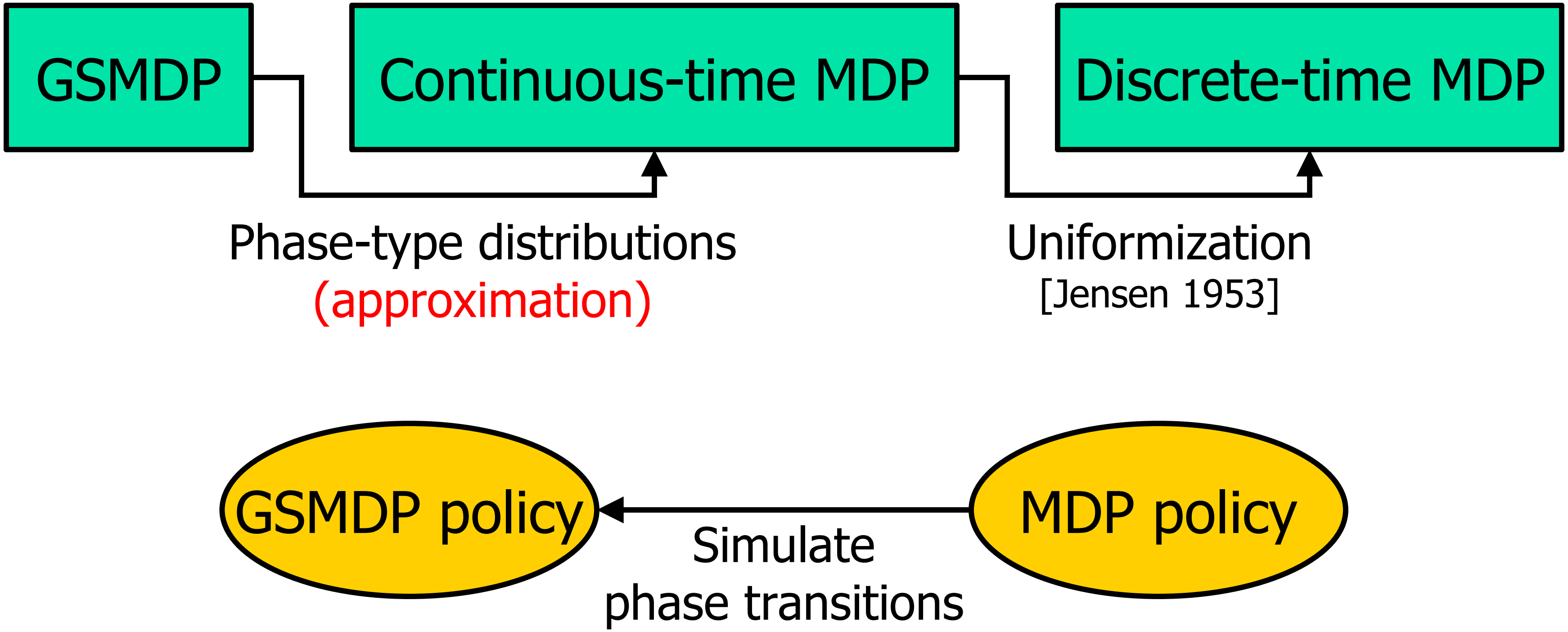
Policies

- Actions as controllable events
  - We can choose to disable an action even if its enabling condition is satisfied
- A policy determines the set of actions to keep enabled at any given time during execution

Rewards and Optimality

- 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
- Infinite-horizon discounted reward
  - Unit reward earned at time t counts as $e^{-\alpha t}$
- Optimal choice may depend on entire execution history

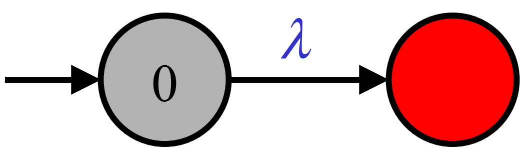
GSMDP Solution Method



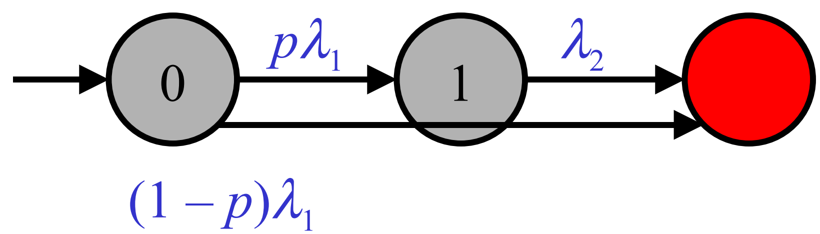
Continuous Phase-Type Distributions [Neuts 1981]

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

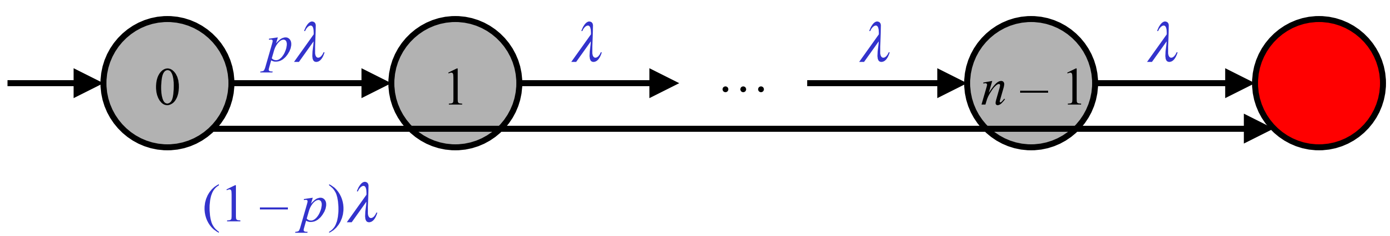
Exponential



Two-phase Coxian



n-phase generalized Erlang



Approximating GSMDP with Continuous-time MDP

- Approximate each distribution G_e with a continuous phase-type distribution
  - Phases become part of state description
  - Phases represent discretization into random-length intervals of the time events have been enabled

Policy Execution

- The policy we obtain is a mapping from modified state space to actions
- To execute a policy we need to simulate phase transitions
- Times when action choice may change:
  - Triggering of actual event or action
  - Simulated phase transition

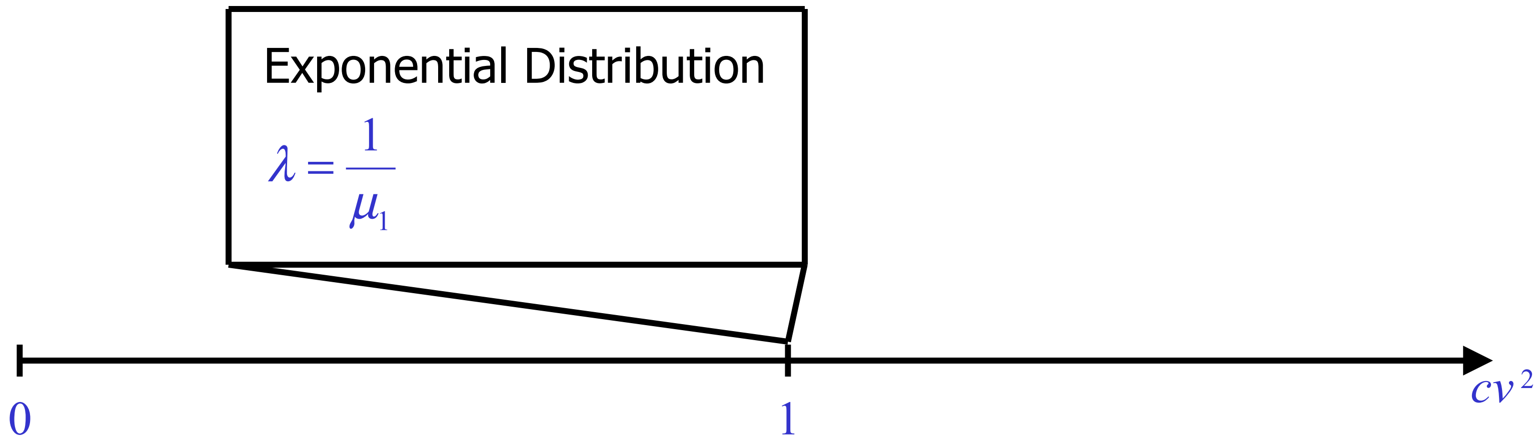
Method of Moments

- Approximate general distribution G with phase-type distribution PH by matching the first k moments
  - Mean (first moment): μ_1
  - Variance: $\sigma^2 = \mu_2 - \mu_1^2$
  - The ith moment: $\mu_i = \mathrm{E}[X^i]$
  - Coefficient of variation: $cv = \sigma / \mu_1$

Matching One Moment

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

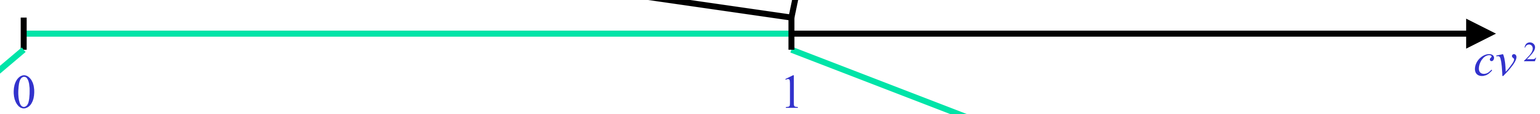
Matching Two Moments



Matching Two Moments

Exponential Distribution

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



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 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

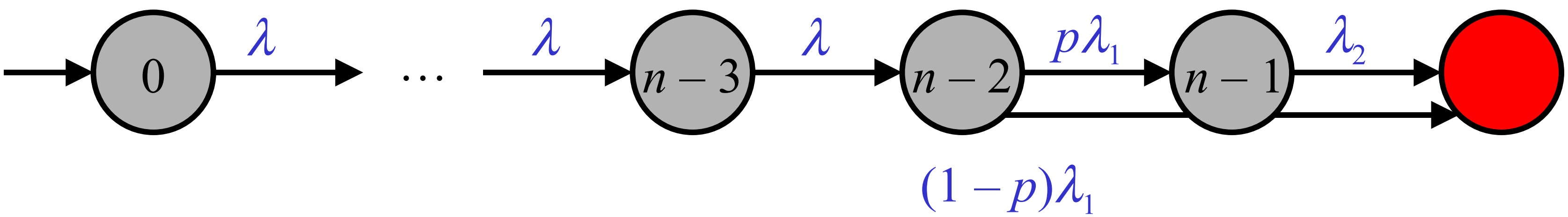
Generalized Erlang Distribution

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

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

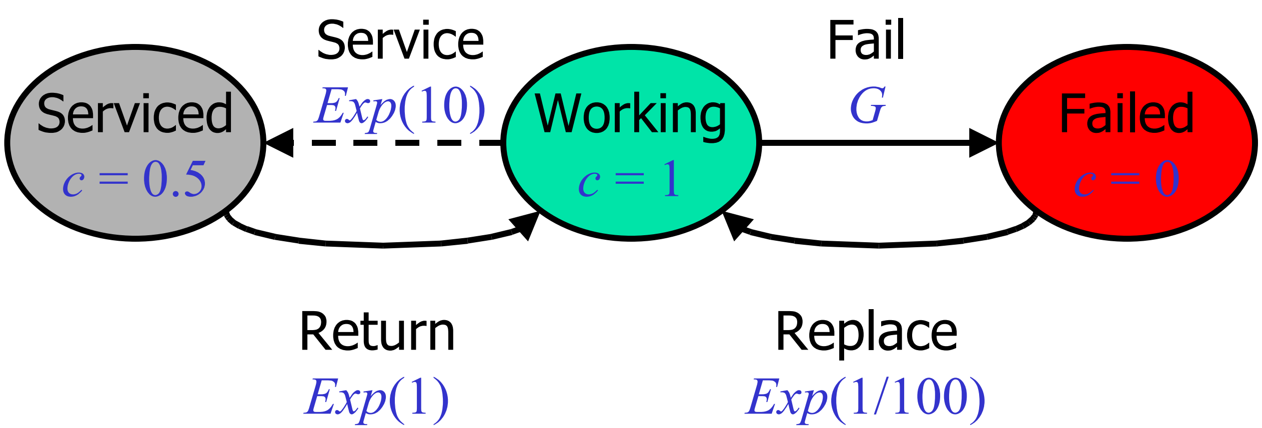
Matching Three Moments

- Combination of Erlang and two-phase Coxian [Osogami & Harchol-Balter, TOOLS'03]



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

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

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

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

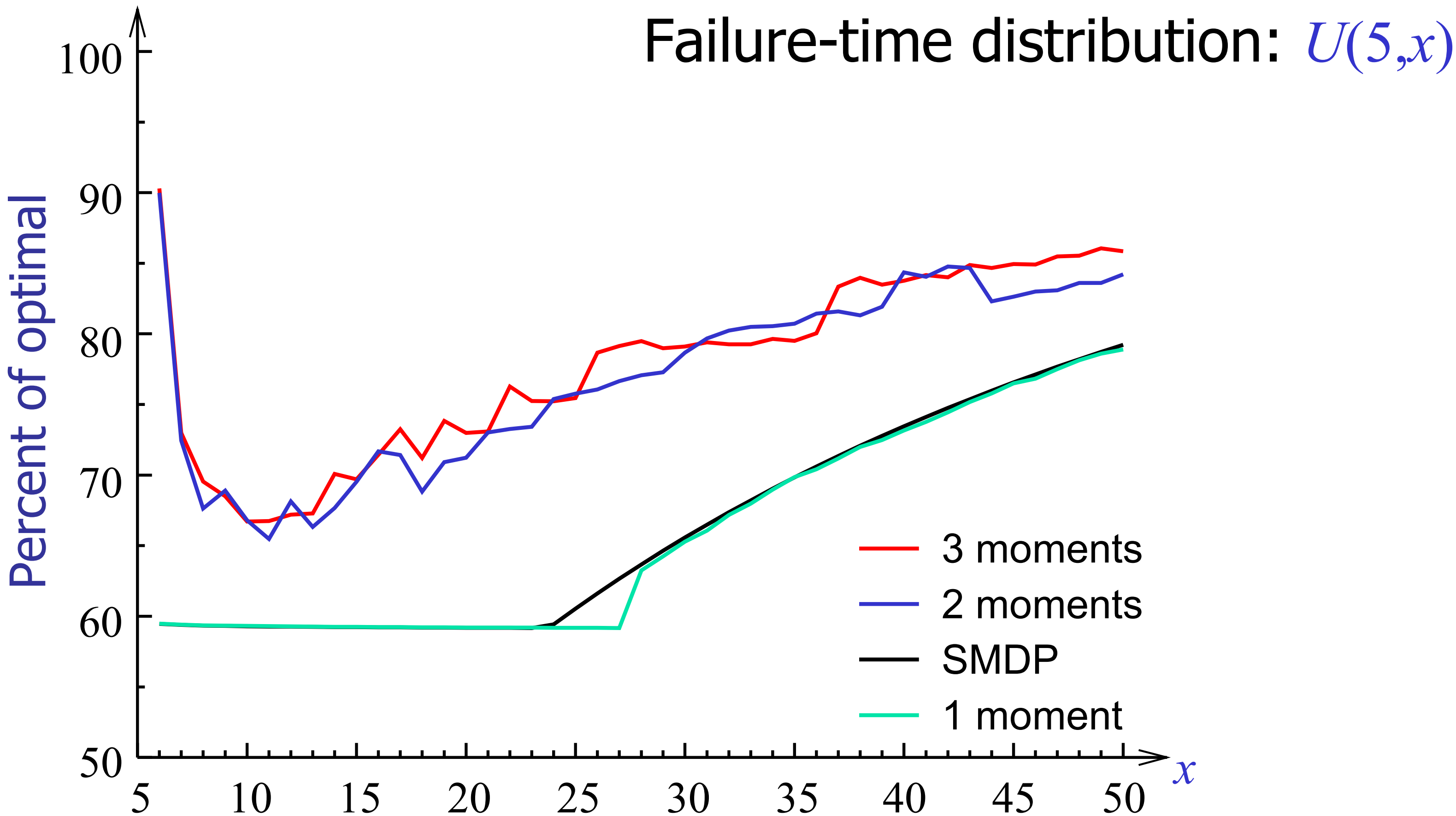
$$F_X(t) = \int_0^t f_X(x)\,dx$$

Y is the time to failure in "Working" state

The Foreman's Dilemma: SMDP Solution

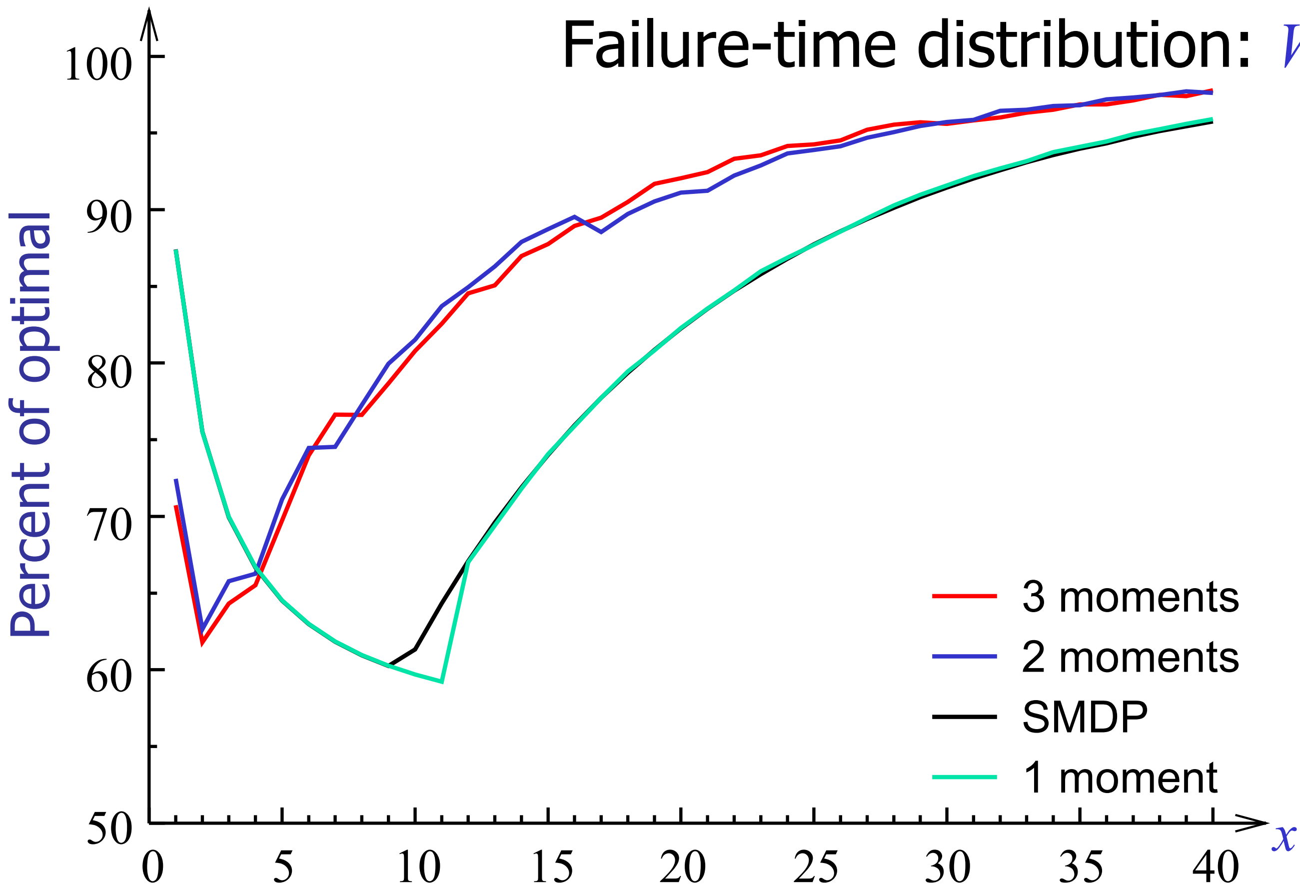
- Same formulas, but restricted choice:
  - Action is immediately enabled ($t_0 = 0$)
  - Action is never enabled ($t_0 = \infty$)

The Foreman's Dilemma: Performance



The Foreman's Dilemma: Performance

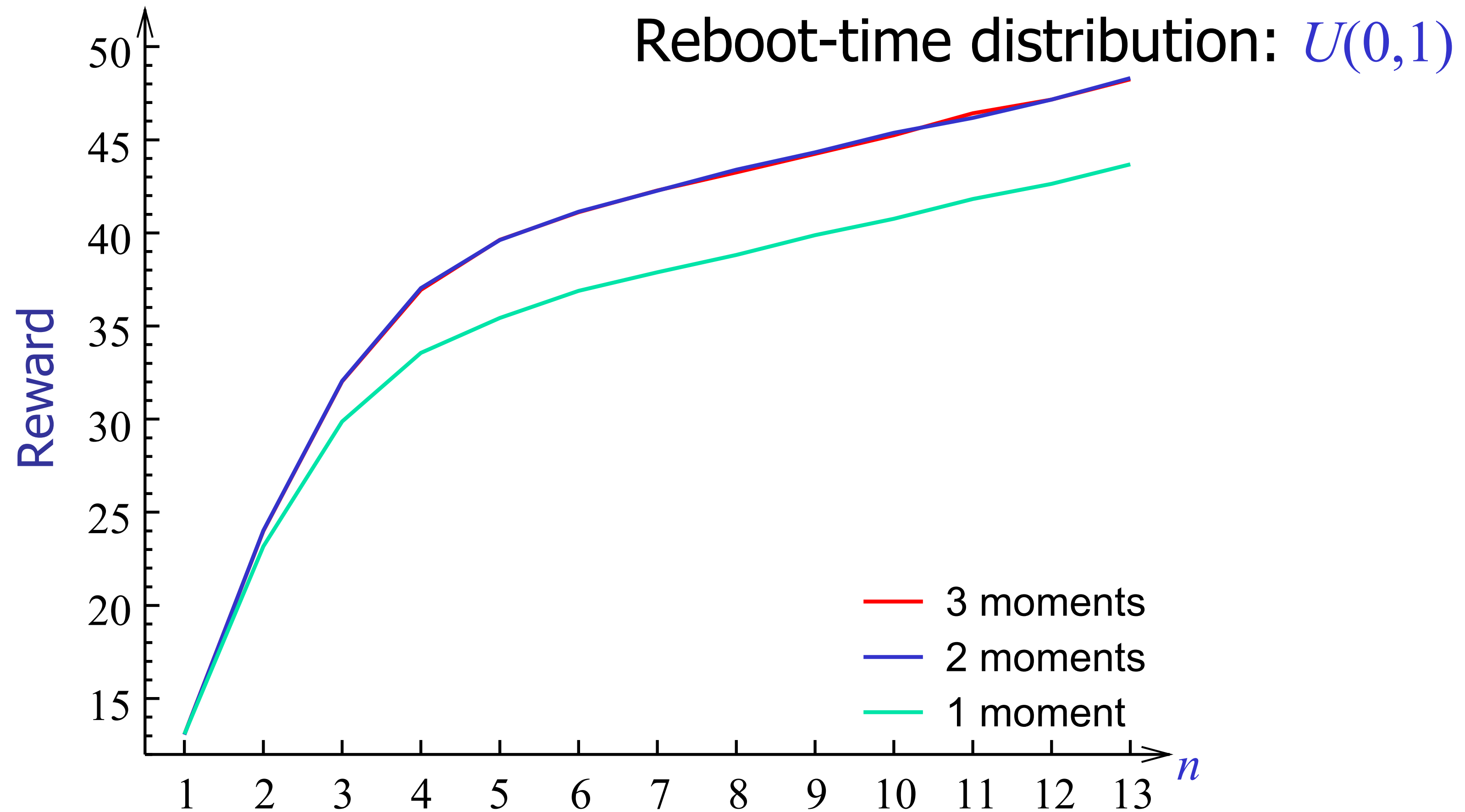
Failure-time distribution: $W(1.6x, 4.5)$



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 (single agent)

System Administration: Performance



System Administration: Performance

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

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 and SMDPs using existing MDP techniques
- Phase does matter!

Future Work

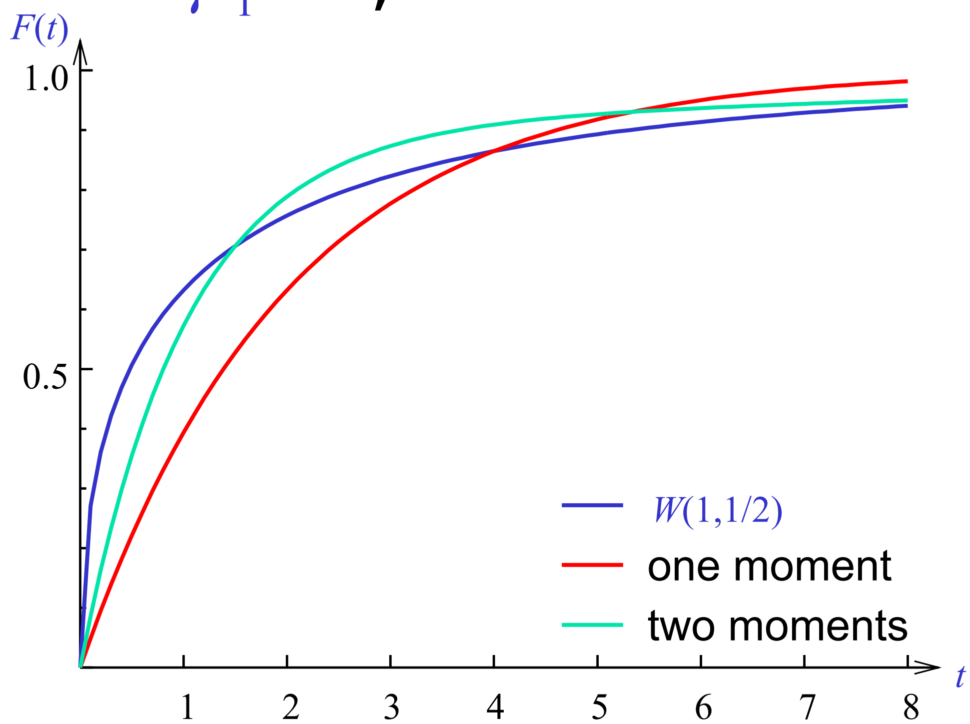
- Discrete phase-type distributions
  - Handles deterministic distributions
  - Avoids uniformization step
- Other optimization criteria
  - Finite horizon, etc.
- Computational complexity of optimal GSMDP planning

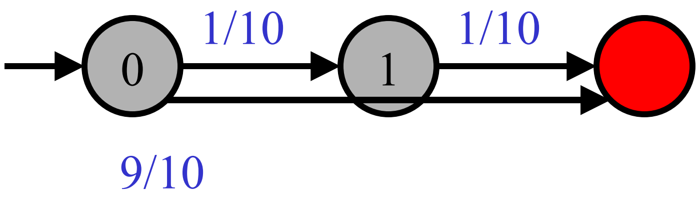
Tempastic-DTP

- A tool for GSMDP planning:
  http://www.cs.cmu.edu/~lorens/tempastic-dtp.html

Matching Moments: Example 1

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





Matching Moments: Example 2

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