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

\[ \text{m}_1 \quad \text{m}_2 \]

\[ \text{m}_1 \text{ up} \quad \text{m}_2 \text{ up} \]

\[ t = 0 \]
Asynchronous Processes: Example

$m_1$

$m_2$

$m_2$ crashes

$m_1$ up  $m_2$ up

$t = 0$

$m_1$ up  $m_2$ down

$t = 2.5$
Asynchronous Processes: Example

$m_1$  

$m_2$ crashes  

$m_1$ crashes

$m_1$ up  

$m_2$ up  

t = 0

$m_1$ up  

$m_2$ down  

t = 2.5

$m_1$ down  

$m_2$ down  

t = 3.1
Asynchronous Processes: Example

$m_1$ crashes  
$m_1$ up  
$m_2$ up  
$t = 0$

$m_1$ up  
$m_1$ crashes  
$m_2$ down  
$t = 2.5$

$m_1$ down  
$m_2$ down  
$m_2$ reboots  
$t = 3.1$

$m_1$ down  
$m_2$ up  
$t = 4.9$
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 \( A \subseteq E \) are controllable events
Events

- With each event $e$ is associated:
  - A condition $\phi_e$ 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: $U(0,1)$

\[ G_{c1} = Exp(1) \]
\[ G_{c2} = Exp(1) \]
\[ G_{r2} = U(0,0.5) \]

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

- GSMDP
- Continuous-time MDP
- Discrete-time MDP

Phase-type distributions (approximation)

- GSMDP policy
- MDP policy

Simulate phase transitions

Uniformization [Jensen 1953]
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): $\mu_1$
  - Variance: $\sigma^2 = \mu_2 - \mu_1^2$
  - The $i$th moment: $\mu_i = E[X^i]$
  - Coefficient of variation: $cv = \sigma / \mu_1$
Matching One Moment

- Exponential distribution: $\lambda = 1/\mu_1$
Matching Two Moments

Exponential Distribution

\[ \lambda = \frac{1}{\mu_1} \]
Matching Two Moments

**Exponential Distribution**

\[ \lambda = \frac{1}{\mu_1} \]

**Generalized Erlang Distribution**

\[
n = \left\lceil \frac{1}{cv^2} \right\rceil \\
p = 1 - \frac{2n \cdot cv^2 + n - 2 - \sqrt{n^2 + 4 - 4n \cdot cv^2}}{2(n-1)(cv^2 + 1)} \\
\lambda = \frac{1 - p + np}{\mu_1}
\]
Matching Two Moments

Exponential Distribution
\[ \lambda = \frac{1}{\mu_1} \]

Generalized Erlang Distribution
\[ n = \left[ \frac{1}{cv^2} \right] \quad p = 1 - \frac{2n \cdot cv^2 + n - 2 - \sqrt{n^2 + 4 - 4n \cdot cv^2}}{2(n-1)(cv^2 + 1)} \]
\[ \lambda = \frac{1 - p + np}{\mu_1} \]

Two-Phase Coxian Distribution
\[ p = \frac{1}{2 \cdot cv^2} \quad \lambda_1 = \frac{2}{\mu_2} \quad \lambda_2 = \frac{1}{\mu_1 \cdot cv^2} \]
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?

![Diagram showing the states and actions: Serviced (c = 0.5), Working (c = 1), Failed (c = 0). The transitions include Service (Exp(10)), Fail (G), Return (Exp(1)), and Replace (Exp(1/100)).]
The Foreman’s Dilemma: Optimal Solution

- Find $t_0$ that maximizes $v_0$

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

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

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

$Y$ 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

Failure-time distribution: $U(5, x)$
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

Reboot-time distribution: $U(0,1)$
## System Administration: Performance

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

\[
2^n \quad (n+1)2^n \quad (1.5n+1)2^n
\]
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: $W(1,1/2)$
  - $\mu_1 = 2$, $cv^2 = 5$

Graph showing the cumulative distribution function $F(t)$ for different scenarios:
- $W(1,1/2)$
- One moment
- Two moments
Matching Moments: Example 2

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