Multiagent Planning under Uncertainty with Stochastic Communication Delays

Matthijs Spaan§ Frans Oliehoek* Nikos Vlassis#

§Institute for Systems and Robotics, IST, Lisbon, Portugal
*University of Amsterdam, The Netherlands
#Technical University of Crete, Chania, Greece

Lisbon Workshop on New Challenges for Cooperative Robotics
October 24, 2008
Introduction

Focus:
- Cooperative multiagent planning under uncertainty.
- Formalized as a decentralized partially observable Markov decision process (Dec-POMDP).
- Very hard to solve.

Communication:
- When available can be leveraged.
- Sharing observations $\rightarrow$ synchronized knowledge.
- But requires instantaneous, perfect communication.

Contributions:
- We consider stochastic delays: towards more realistic communication models for Dec-POMDPs.
Outline

• POMDPs vs. Dec-POMDPs.
• Review of communication models:
  ► Instantaneous, perfect communication.
  ► Delayed communication (fixed delay).
• Stochastic communication delays:
  ► Model.
  ► Compactly representing the value function.
  ► Value iteration method.
  ► Experimental results.
• Conclusions.
POMDP

time

$st$

agent

$ot$

$a$

$st+1$

$t + 1$
Dec-POMDP (without communication)

agent i

agent j

$t$

$t + 1$

$s^t$

$s^{t+1}$

$O_i$

$O_j$

$a_i$

$a_j$
POMDPs vs. Dec-POMDPs

Frameworks for acting optimally given:

- limited sensing,
- stochastic environments.

Dec-POMDPs:

- Decentralized execution.
- Usually centralized planning.
- No common state estimate, no joint belief.
- Optimal policies based on observation history mapping.
Instantaneous communication

agent i

agent j

\(s^t\)

\(s^{t+1}\)

\(o_i\)

\(a_i\)

\(o_j\)

\(a_j\)

decision interval

decision interval

\(\text{time}\)
Instantaneous communication

- Instantaneous, perfect communication reduces a Dec-POMDP to a POMDP (Pynadath and Tambe, 2002).
- Observation sharing → each agent knows belief $b^t$ at time $t$.
- Piecewise-linear and convex (PWLC) value function.
- But this synchronization step takes time.

```
... is common knowledge...
```
Delayed communication

agent i  

agent j  

\( s_t \)  

\( s_{t+1} \)  

do time  

\( o_i \)  

\( o_j \)  

\( a_i \)  

\( a_j \)  

\{ decision interval \}
Fixed delay communication

- Each agent knows the last commonly known $b^{t-k}$, $k > 0$.
- If $k = 1$, fixed delay communication can be modeled as a Bayesian game (Oliehoek et al., 2007).
- If $k = 1$, the optimal value function is PWLC, and equal to the $Q_{BG}$ value function (Oliehoek et al., 2008).
- Otherwise the value function is not separable (Varaiya and Walrand, 1978), it is not a function over the joint belief space.
- But what if the delay can vary?
We formalize the probability that synchronization succeeds within a particular stage (details in Spaan et al. (2008)):

\[ p^{0\text{TD}}(s) \] synchronization is instantaneous
\[ p^{1\text{TD}}(s) \] synchronization succeeds within 1 stage
\[ p^{2\text{TD}}(s) \] synchronization succeeds within 2 stages
\[ \vdots \]

Can depend on state, and \( \sum_i p^{i\text{TD}}(s) = 1 \).

E.g., robots close \( \rightarrow \) more reliable communication.

Unifying framework:

\[ p^{0\text{TD}}(s) = 1, \forall s: \text{instantaneous communication.} \]
\[ p^{1\text{TD}}(s) = 1, \forall s: \text{one-step delay communication.} \]
Stochastic communication delays

- Optimal value function:
  \[ Q^*_\text{SD} = R + p^{0\text{TD}} F_{0\text{TD}} + p^{1\text{TD}} F_{1\text{TD}} + p^{2\text{TD}} F_{2\text{TD}} + \ldots, \]
  where \( F_{i\text{TD}} \) is the expected future reward given a delay of \( i \) stages.

- We assume during planning that delay is at most 1 step:
  \[ p^D = p^{1\text{TD}} + p^{2\text{TD}} + \cdots = 1 - p^{0\text{TD}}, \]
  and define an approximate value function as
  \[ \tilde{Q}^*_\text{SD} = R + p^{0\text{TD}} F_{0\text{TD}} + p^D F_{1\text{TD}}. \]

- We prove that \( \tilde{Q}^*_\text{SD} \) is PWLC over the joint belief space.
Algorithm

- We show how to perform finite-horizon value iteration for the stochastically delayed communication setting.
- Point-based approximate POMDP techniques transfer directly to our $\tilde{Q}_*^{SD}$ value function.
- For delays of more than 1 one time step, we propose an online algorithm similar to Dec-COMM (Roth et al., 2005).
Experimental results

Uniform delay between 0 and 1:

Dec-Tiger

GridSmall
The policies consider potential future communication capabilities.

The CW route is cheaper, and with uniform $p^{0TD}$ it is chosen.

We set $p^{0TD}(s) = 0, \forall s \in CW$ and to 1 everywhere else.

Now the agents choose the CCW route, leading to higher payoff.

---

<table>
<thead>
<tr>
<th></th>
<th>CCW</th>
<th>CCW</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CW</td>
<td></td>
<td>CCW</td>
</tr>
<tr>
<td>CW</td>
<td>CW</td>
<td>S</td>
</tr>
</tbody>
</table>
# Experimental results

## Arbitrary delays:

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p^0_{TD}$</td>
<td>0.8</td>
<td>0.8</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>$p^1_{TD}$</td>
<td>0.2</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>$p^2_{TD}$</td>
<td></td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>$p^3_{TD}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1</td>
</tr>
</tbody>
</table>

![Graph](#)
Experimental results

Policies computed for a specific value of $p^{0_{TD}}$, but evaluated under different values of $p^{0_{TD}}$.

Meet in corner (large)

![Plot showing the relationship between $V(\theta^0)$ and $p^{0_{TD}}$ for different policies.]

- $\pi_{p^{0_{TD}}=0}$
- $\pi_{p^{0_{TD}}=0.5}$
- $\pi_{p^{0_{TD}}=1.0}$
Conclusions:

- We discussed different communication assumptions made in the Dec-POMDP literature.
- We introduced a model for stochastically delayed communication:
  - More realistically models unreliable communication.
  - We showed how we can approximate its optimal value function, and how to compute it efficiently.
- Encouraging experimental results.

Future work:

- Not only delayed, but also noisy communication.
- Alternative methods for longer delays.
- Infinite-horizon value iteration.
References


