**Resource-Bounded Goal Obfuscation**

Anagha Kulkarni*, PhD Student, Arizona State University
Matthew Klenk, Shantanu Rane, Hamed Soroush, Palo Alto Research Center

- In an adversarial environment, agents should not reveal their objectives.
- A goal obfuscated plan produces a sequence of observations from which an adversary is unable to derive the agent’s true objective.

### Application areas

<table>
<thead>
<tr>
<th>Application areas</th>
<th>Information to be concealed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mission Planning</td>
<td>Point of attack</td>
</tr>
<tr>
<td>Corporate Strategy</td>
<td>Potential partners during new product launch</td>
</tr>
<tr>
<td>Component Design</td>
<td>Intended use of designed components</td>
</tr>
</tbody>
</table>

### Problem Setting

- We consider two agents: an actor, and an observer.
- The actor performs deterministic actions and has full observability.
- The observer can have partial/full observability of actor’s actions.
- The observer knows that the actor has \( n \) candidate goals but in unaware of actor’s true goal.
- The solution involves choosing \( v \) (\( \leq n \)) candidate goals and generating a goal obfuscated plan that is equally consistent with all \( v \) candidate goals.

### Example

1. **Actor** selects \( v \)-1 candidate goals using landmark similarity to obfuscate the true goal.

2. Search the state space to obtain equidistant states. Equidistant states have equal distance to all the \( v \) goals.

3. From each equidistant state, perform bounded length belief space search until a solution is found.

4. Observation sequence obtained for the given \( k \) goals: \(< O_2, O_1, O_1, O_1 >\)

### A goal obfuscation planning problem is a tuple, \( P_{GO} = (\mathcal{D}, \mathcal{I}, \mathcal{G}, \Omega, \mathcal{O}, \mathcal{R}) \), where:

- \( \mathcal{G} = \{ G_A \cup G_1 \cup G_2 \ldots \cup G_{n-1} \} \) is a set of \( n \) candidate goal conditions, and \( G_A \) is the true goal of the actor.
- \( \Omega = \{ o_i | i = 1, \ldots, m \} \) is a finite set of observations symbols corresponding to the domain.
- \( \mathcal{O} : (\mathcal{A} \times \mathcal{S}) \rightarrow \Omega \) is the observation function which allows either partial or full observability mapping the pair: action taken and state reached to observation symbols.
- \( \mathcal{R} \) is the cost budget of the actor.

### To compute a secure goal obfuscated plan:

**Step 1:** Choose \( v \) goals with high landmark similarity

**Step 2:** Compute a set of equidistant states

**Step 3:** Compute a bounded length belief plan by exploring equidistant states

**Step 4:** Return optimal plan to equidistant state + bounded length belief plan

### Cryptography Assumptions

- Adversary knows the agent's algorithm
- Independence of inputs
- Delivery of observations is fair and in-order
- Semi-Honest Adversary

### Empirical Evaluation

We evaluated our approach using three IPC domains

**Metric 1:** The impact of different observation models on the extent of obfuscation.

**Metric 2:** The trade-off between additional cost and extent of obfuscation possible.

**Metric 3:** The comparison between run time and plan costs for goal obfuscation versus optimal planning.