Privacy-Preserving Policy Synthesis for MDPs

Matthew Hale

Department of Mechanical and Aerospace Engineering
University of Florida

Joint work with Parham Gohari, Bo Wu, and Ufuk Topcu at UT Austin

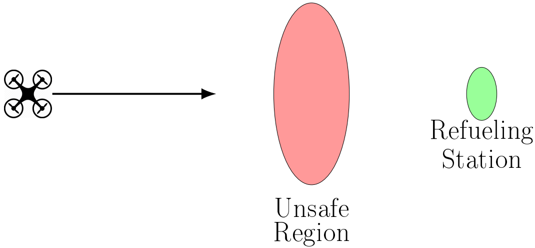
AFOSR Center of Excellence on Assured Autonomy in Contested Environments
April 15th, 2020

Challenge: Adversaries can observe us

- Adversaries can observe us, and actions can reveal intent/knowledge

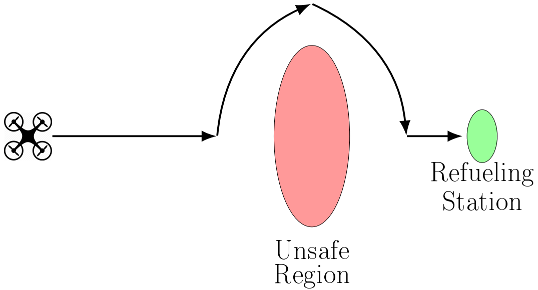
Challenge: Adversaries can observe us

- Adversaries can observe us, and actions can reveal intent/knowledge
  - Direction of travel can reveal a destination



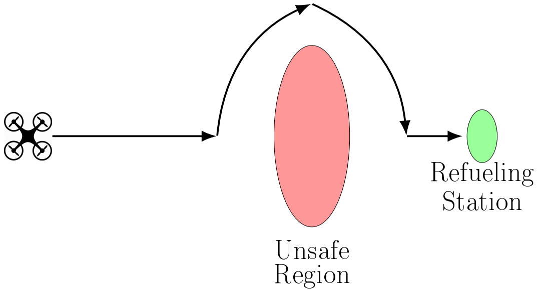
Challenge: Adversaries can observe us

- Adversaries can observe us, and actions can reveal intent/knowledge
  - Direction of travel can reveal a destination
  - Avoiding an area can reveal knowledge of hazards



Challenge: Adversaries can observe us

- Adversaries can observe us, and actions can reveal intent/knowledge
  - Direction of travel can reveal a destination
  - Avoiding an area can reveal knowledge of hazards



Fundamental Problem

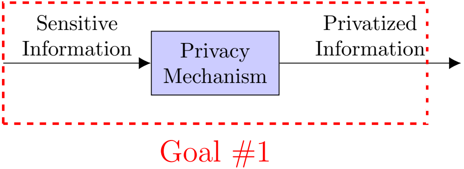
A task must be completed without revealing the information driving it.

Goals of private synthesis

- There are 3 goals in this work:

Goals of private synthesis

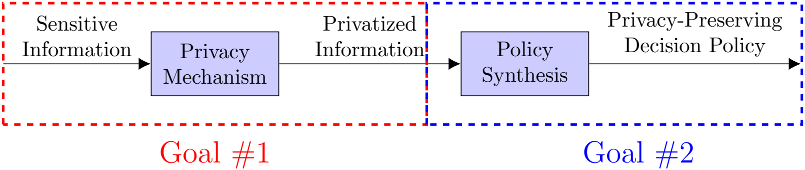
- There are 3 goals in this work:



1. **Provably** protect the information driving a decision

Goals of private synthesis

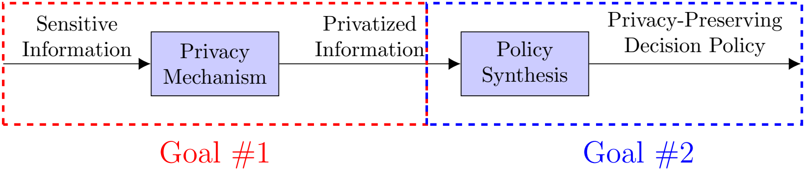
- There are 3 goals in this work:



1. **Provably** protect the information driving a decision
2. Synthesize an altered, privacy-preserving decision policy

Goals of private synthesis

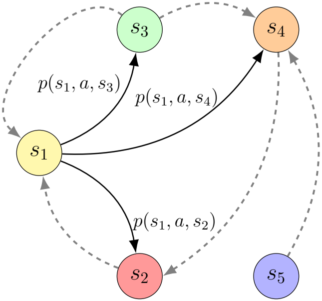
- There are 3 goals in this work:



1. **Provably** protect the information driving a decision
2. Synthesize an altered, privacy-preserving decision policy
3. Quantify the "cost of privacy," formalize tradeoffs between privacy and performance

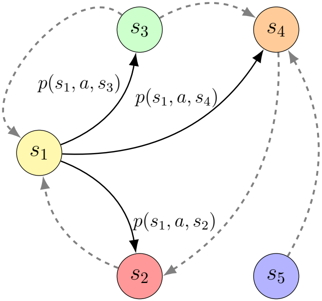
Problem Setup

- We consider MDP models:



Problem Setup

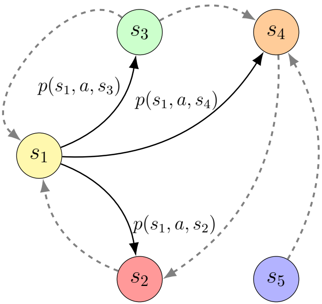
- We consider MDP models:



- We want to take actions that don't reveal transition probabilities

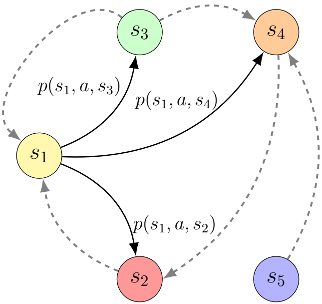
Problem Setup

- We consider MDP models:



- We want to take actions that don't reveal transition probabilities
- In state s, taking action a transitions to state s' with prob. $p(s, a, s')$

- We consider MDP models:



- We want to take actions that don't reveal transition probabilities
- In state s, taking action a transitions to state s' with prob. $p(s, a, s')$
- For all s, we have $p(s, a, s') \geq 0$ and $\sum_{s'} p(s, a, s') = 1$

How should we provide privacy?

How should we provide privacy?

Differential Privacy (DP)

DP is a privacy framework with several key features:

- It offers a formal definition of "privacy"

How should we provide privacy?

Differential Privacy (DP)

DP is a privacy framework with several key features:

- It offers a formal definition of “privacy”
- It is immune to post-processing
  - x private $\Rightarrow f(x)$ private for all f

How should we provide privacy?

Differential Privacy (DP)

DP is a privacy framework with several key features:

- It offers a formal definition of "privacy"
- It is immune to post-processing
  - x private $\Rightarrow f(x)$ private for all f
- It is robust to side information

How should we provide privacy?

Differential Privacy (DP)

DP is a privacy framework with several key features:

- It offers a formal definition of "privacy"
- It is immune to post-processing
  - x private $\Rightarrow f(x)$ private for all f
- It is robust to side information

- Used by: Apple Google Uber

Differential Privacy (DP)

DP is a privacy framework with several key features:

- It offers a formal definition of "privacy"
- It is immune to post-processing
  - x private $\Rightarrow f(x)$ private for all f
- It is robust to side information

- Used by: Apple Google Uber

UBER

DP Idea

Make probability vectors look "similar"

Differential Privacy is a Statistical Guarantee

Fundamental Inequality of Differential Privacy

For probability vectors p and q, we generate private forms $\tilde{p}$, $\tilde{q}$ to satisfy

$$\mathbb{P}(\tilde{p}) \leq e^{\epsilon}\mathbb{P}(\tilde{q}) + \delta$$

Differential Privacy is a Statistical Guarantee

Fundamental Inequality of Differential Privacy

For probability vectors p and q, we generate private forms $\tilde{p}$, $\tilde{q}$ to satisfy

$$\mathbb{P}(\tilde{p}) \leq e^{\epsilon}\mathbb{P}(\tilde{q}) + \delta$$

Differential Privacy is a Statistical Guarantee

Fundamental Inequality of Differential Privacy

For probability vectors p and q, we generate private forms $\tilde{p}$, $\tilde{q}$ to satisfy

$$\mathbb{P}(\tilde{p}) \leq e^{\epsilon}\mathbb{P}(\tilde{q}) + \delta$$

- For us this will take the form

$$\xrightarrow[p]{\text{Vector}} \boxed{\text{Dirichlet}(kp)} \xrightarrow[\tilde{p}]{\text{Private Vector}}$$

where

$$\text{Dirichlet}(kp) = \frac{\Gamma\left(\sum_{i=1}^{n} kp_i\right)}{\prod_{i=1}^{n} \Gamma(kp_i)} \prod_{i=1}^{n} x_i^{kp_i}$$

Differential Privacy is a Statistical Guarantee

Fundamental Inequality of Differential Privacy

For probability vectors p and q, we generate private forms $\tilde{p}$, $\tilde{q}$ to satisfy

$$\mathbb{P}(\tilde{p}) \leq e^{\epsilon}\mathbb{P}(\tilde{q}) + \delta$$

- For us this will take the form

$$\xrightarrow[p]{\text{Vector}} \boxed{\text{Dirichlet}(kp)} \xrightarrow[\tilde{p}]{\text{Private Vector}}$$

where

$$\text{Dirichlet}(kp) = \frac{\Gamma\left(\sum_{i=1}^{n} kp_i\right)}{\prod_{i=1}^{n}\Gamma(kp_i)} \prod_{i=1}^{n} x_i^{kp_i}$$

Privacy Theorem (ACC 2020 Paper; Gohari, Wu, Hale, and Topcu)

The Dirichlet mechanism provides $\big(\epsilon(k), \delta(k)\big)$-differential privacy.

- Example: $k = 24$ gives $(1.18, 0.05)$-DP

- Objective is to maximize the accumulated reward

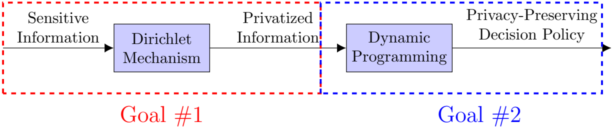
$$\sum_{t=1}^{T} \gamma^t R(s_t, a_t)$$

Private Synthesis

- Objective is to maximize the accumulated reward

$$\sum_{t=1}^{T} \gamma^t R(s_t, a_t)$$

- We privatize transition probabilities, then synthesize a decision policy

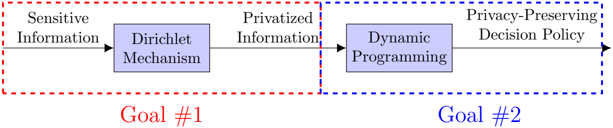


Private Synthesis

- Objective is to maximize the accumulated reward

$$\sum_{t=1}^{T} \gamma^t R(s_t, a_t)$$

- We privatize transition probabilities, then synthesize a decision policy



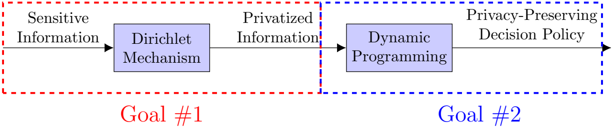
- Synthesis is just post-processing, so its output protects p as well

Private Synthesis

- Objective is to maximize the accumulated reward

$$\sum_{t=1}^{T} \gamma^t R(s_t, a_t)$$

- We privatize transition probabilities, then synthesize a decision policy



- Synthesis is just post-processing, so its output protects p as well
- Our actions protect transition probabilities!

The cost of privacy

- Set **Cost of privacy** = (Reward without DP) - (Reward with DP)

- Set **Cost of privacy** = (Reward without DP) - (Reward with DP)

Theorem: Cost of Privacy

- Privatize all transition probabilities with the Dirichlet mechanism. Then:

$$\text{Cost of privacy} \leq w_0 - v_0,$$

where, for all $t \in \{0, \dots, T\}$,

$$v_t = \sum_{a \in \mathcal{A}_s} \pi(a \mid s) \left(R(s,a) + \gamma \min_{p \in \hat{\mathcal{P}}} p(s,a,s') v_{t+1}(s') \right)$$

$$w_t = \sum_{a \in \mathcal{A}_s} \pi(a \mid s) \left(R(s,a) + \gamma \max_{p \in \hat{\mathcal{P}}} p(s,a,s') w_{t+1}(s') \right)$$

The cost of privacy

- Set **Cost of privacy** = (Reward without DP) - (Reward with DP)

Theorem: Cost of Privacy

- Privatize all transition probabilities with the Dirichlet mechanism. Then:

$$\text{Cost of privacy} \leq w_0 - v_0,$$

where, for all $t \in \{0, \dots, T\}$,

$$v_t = \sum_{a \in \mathcal{A}_s} \pi(a \mid s) \left(R(s,a) + \gamma \min_{p \in \hat{\mathcal{P}}} p(s,a,s') v_{t+1}(s') \right)$$

$$w_t = \sum_{a \in \mathcal{A}_s} \pi(a \mid s) \left(R(s,a) + \gamma \max_{p \in \hat{\mathcal{P}}} p(s,a,s') w_{t+1}(s') \right)$$

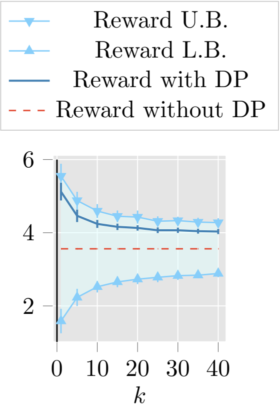
- This is computable in $O\left(T|S|^{4.5}|A_s|\right)$ time

- Implement privacy and synthesize a policy for a 30-state MDP

Simulation Results

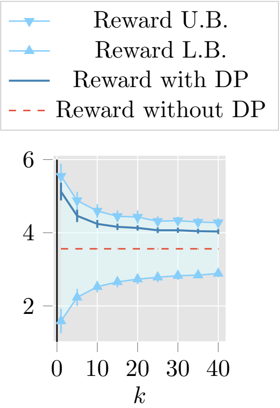
- Implement privacy and synthesize a policy for a 30-state MDP
- Total time required is 4.88s on a desktop computer

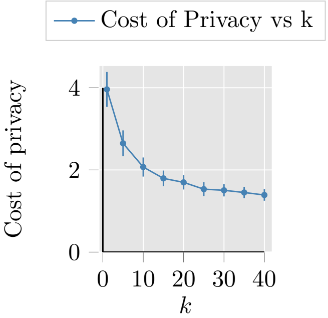
- Implement privacy and synthesize a policy for a 30-state MDP
- Total time required is 4.88s on a desktop computer



Simulation Results

- Implement privacy and synthesize a policy for a 30-state MDP
- Total time required is 4.88s on a desktop computer





Next Steps

- Incorporating temporal logic specifications:
  - What are the tradeoffs in privacy, safety, and performance?
  - What is the complexity of computing a safe, private policy and bounds on the cost of safety & privacy?

Next Steps

- Incorporating temporal logic specifications:
  - What are the tradeoffs in privacy, safety, and performance?
  - What is the complexity of computing a safe, private policy and bounds on the cost of safety & privacy?
- What are the effects of privatizing other characteristics of MDPs?

Thank you