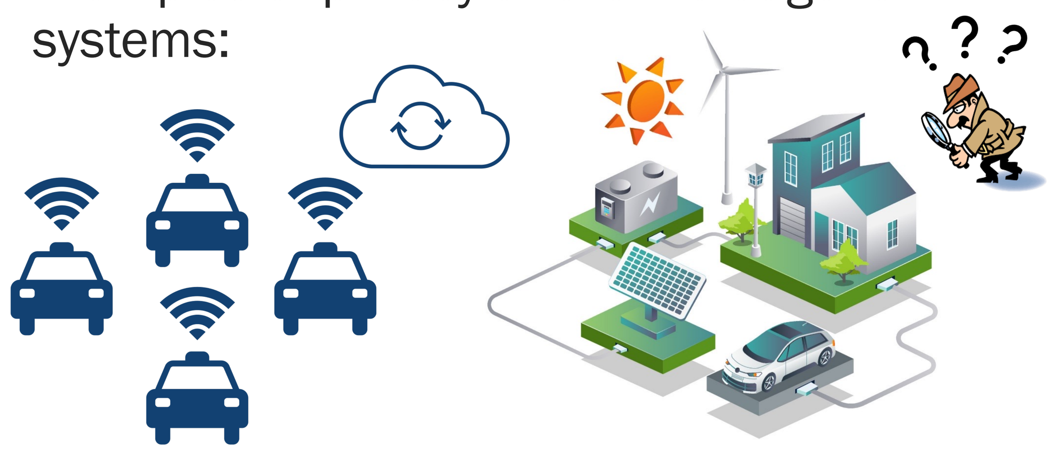
Differential Privacy in Cooperative Multiagent Planning

Bo Chen, Calvin Hawkins, Mustafa Karabag, Cyrus Neary, Matthew Hale, Ufuk Topcu
{bo.chen,calvin.hawkins,matthewhale}@ufl.edu, {karabag,cneary,utopcu}@utexas.edu

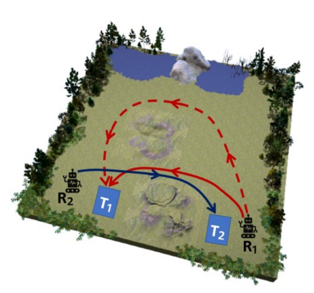
Motivation

- Overall goal: allow agents to interact with a shared environment to achieve a common objective while protecting sensitive information.
- Examples of privacy aware multiagent systems:



The Setting

- Consider a collection of N agents indexed by $i \in \{1, \dots, N\}$.
- Each agent's dynamics are modeled by an MDP: $M^i = (S^i, s_I^i, A^i, T^i)$.
- Cooperative Markov game $\boldsymbol{M} = (\boldsymbol{S}, \boldsymbol{s_I}, \boldsymbol{A}, \boldsymbol{T})$:
  - $\boldsymbol{S} = S^1 \times \cdots \times S^N$
  - $\boldsymbol{s_I} = (s_I^1, \dots, s_I^N)$
  - $\boldsymbol{A} = A^1 \times \cdots \times A^N$
  - $\boldsymbol{T}(\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{y}) = \prod_{i=1}^N T^i(s^i, a^i, y^i)$
- Agent i's policy: $\pi^i : \boldsymbol{S} \to \Delta(A^i)$
  - $a^i \sim \pi^i(s)$
- Team's objective:
  - Target set: $\boldsymbol{S_T} \subseteq \boldsymbol{S}$
  - Avoid set: $\boldsymbol{S_A} \subseteq \boldsymbol{S}$



Privacy Concerns

- At each timestep t, agent i needs the joint state $\boldsymbol{s_t}$ to execute its local policy and generate a local action $a_t^i \sim \pi^i(\boldsymbol{s_t})$
- Thus, agent i needs to share its trajectory $h_t^i = \{s_1^i, s_2^i, \dots, s_t^i\}$ with the rest of the team to achieve the collaborative objective.

Differential Privacy

- Goal: make "similar" pieces of data appear approximately indistinguishable.
- Adjacency encodes when two trajectories are "similar":
  - Fix an adjacency parameter $k \in \mathbb{N}^+$ and length $T \in \mathbb{N}^+$
  - Fix two trajectories $v, w \in (S^i)^T$, these two trajectories are adjacent if $d(v, w) \le k$, where $d()$ denotes the Hamming distance.
- A mechanism $\mathcal{M} : (S^i)^T \times \Omega \to (S^i)^T$ is ϵ-differentially private if for any adjacent v & w

$$P[\mathcal{M}(v) \in S] \le e^{\epsilon} P[\mathcal{M}(w) \in S]$$

$h_t^i = s_1^i, \dots, s_T^i$ → Privacy Mechanism tuned by ϵ → $\tilde{h}_t^i = \tilde{s}_1^i, \dots, \tilde{s}_T^i$

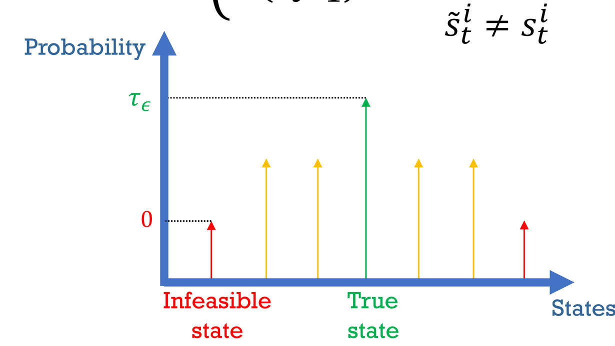
Problem Statements

1. Design an online privacy mechanism to provide ϵ-differential privacy to $h_t^i = \{s_1^i, s_2^i, \dots, s_t^i\}$ in real time
2. Define an algorithm for the decentralized execution of local policies $\{\pi^i\}_{i=1}^N$ under private communications
3. Given local policies $\{\pi^i\}_{i=1}^N$, bound the probability of success under private communications, $\boldsymbol{v^{pr}}$
4. Synthesize policies that achieve high performance under private communications

The Privacy Mechanism

- At time t, agent i is at state s_t^i and generates a private state $\tilde{s}_t^i \sim \mu_\epsilon^i(\cdot \mid \tilde{s}_{t-1}^i, s_t^i)$

$$\mu_\epsilon^i(\tilde{s}_t^i \mid \tilde{s}_{t-1}^i, s_t^i) = \begin{cases} 0 & \text{If } \tilde{s}_t^i \text{ is not feasible from } \tilde{s}_{t-1}^i \\ \tau_\epsilon(\tilde{s}_{t-1}^i) & \text{If } \tilde{s}_t^i \text{ is feasible from } \tilde{s}_{t-1}^i \text{ and } \tilde{s}_t^i = s_t^i \\ \frac{1-\tau_\epsilon(\tilde{s}_{t-1}^i)}{\rho(\tilde{s}_{t-1}^i)-1} & \text{If } \tilde{s}_t^i \text{ is feasible from } \tilde{s}_{t-1}^i \text{ and } \tilde{s}_t^i \neq s_t^i \end{cases}$$



- The probability of true transition, τ_ϵ, will be tuned to meet ϵ-differential privacy.

Algorithm 1: Online Mechanism Construction

Input: Probability of true transition τ_ϵ
Output: μ_ϵ^i
for $(s_t^i, \tilde{s}_{t-1}^i, \tilde{s}_t^i) \in \mathcal{S}^i \times \mathcal{S}^i \times \mathcal{S}^i$ **do**
  if $s_t^i = \tilde{s}_t^i$ *and* $\beta(\tilde{s}_t^i, \tilde{s}_{t-1}^i) = 1$ **then**
    $\mu_\epsilon^i(\tilde{s}_t^i \mid s_t^i, \tilde{s}_{t-1}^i) = \tau_\epsilon(\tilde{s}_{t-1}^i)$.
  else if $s_t^i \neq \tilde{s}_t^i$ *and* $\beta(\tilde{s}_t^i, \tilde{s}_{t-1}^i) = 1$ **then**
    $\mu_\epsilon^i(\tilde{s}_t^i \mid s_t^i, \tilde{s}_{t-1}^i) = \frac{1-\tau_\epsilon(\tilde{s}_{t-1}^i)\beta(s_t^i, \tilde{s}_{t-1}^i)}{\rho(\tilde{s}_{t-1}^i)-\beta(s_t^i, \tilde{s}_{t-1}^i)}$.
  else
    $\mu_\epsilon^i(\tilde{s}_t^i \mid s_t^i, \tilde{s}_{t-1}^i) = 0$.

Implementing Local Policies with Private Communications

- Agents treat the information they receive from the rest of the network as the truth and store this information in $\hat{s}_{t,i}$

Algorithm 2: Privatized Policy Execution

Input for every agent i**:** Local policy π^i
Set $\tilde{s}_0^i = s_I^i$ for all $i \in [N]$.
for $t = 0, 1, \dots$ **every agent** i **does in parallel**
  Set $\hat{\boldsymbol{s}}_{t,i} = (\tilde{\boldsymbol{s}}_{t,i}^{Succ}, s_t^i)$.
  Sample an action $a_t^i \sim \pi^i(\hat{\boldsymbol{s}}_{t,i})$.
  Execute a_t^i and transition to $s_{t+1}^i \sim \mathcal{T}^i(s_t^i, a_t^i)$.
  Share $\tilde{s}_{t+1}^i \sim \mu_\epsilon^i(\cdot \mid s_{t+1}^i, \tilde{s}_t^i)$ with agents in $Pred(i)$.

Theoretical results

Theorem 1 (Privacy)

Fix an adjacency parameter $k \in \mathbb{N}^+$, and a privacy parameter $\epsilon > 0$. The online mechanism is ϵ-differentially private if $\tau_\epsilon(\tilde{s}_{t-1}^i)$ satisfies

$$\tau_\epsilon(\tilde{s}_{t-1}^i) = \frac{1}{\left(\rho(\tilde{s}_{t-1}^i) - 1\right)\exp\left(-\frac{\epsilon}{k+1}\right)}$$

- Given a joint policy $\boldsymbol{\pi} = \{\pi_i\}_{i=1}^N$:
  - We represent the total correlation as
    $C_{\boldsymbol{\pi}} = \sum_{i=1}^N H(S_0^i A_0^i \dots S_\eta^i) - H(\boldsymbol{S_0 A_0} \dots \boldsymbol{S_\eta})$
  - Where $H(Y) = -\sum_{y \in \mathcal{Y}} \Pr(Y = y) \log(\Pr(Y = y))$
  - Denote the expected trajectory length under true communication as l^{tr}
  - Denote the probability of success under truthful communication as $\boldsymbol{v^{tr}}$, and denote the probability of success under private communication as $\boldsymbol{v^{pr}}$

Theorem 2 (Performance)

Given N agents implementing the policies $\boldsymbol{\pi} = \{\pi^i\}_{i=1}^N$ with private communications according to Algorithm 2, then

$$\boldsymbol{v^{pr}} \ge \boldsymbol{v^{tr}} - \sqrt{1 - e^{-C_{\boldsymbol{\pi}}}\left((\rho_m - 1)e^{-\frac{\epsilon}{k}} + 1\right)^{Nl^{tr}}}$$

Policy Synthesis

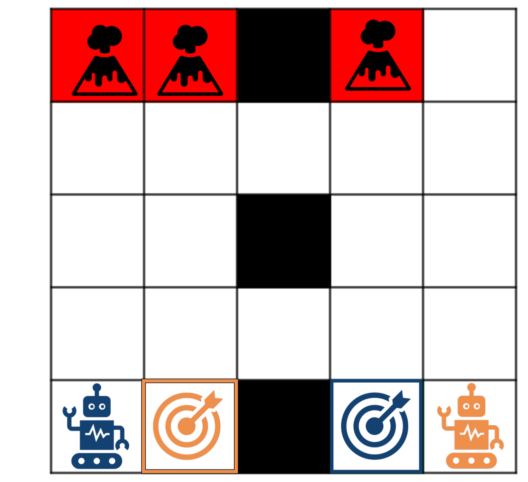
- Here we provide a method for the synthesis of policies $\boldsymbol{\pi} = \{\pi_i\}_{i=1}^N$ that remain performant under private communications.
- Goal: Maximize the probability of success under private communications
- Use Theorem 2 and solve

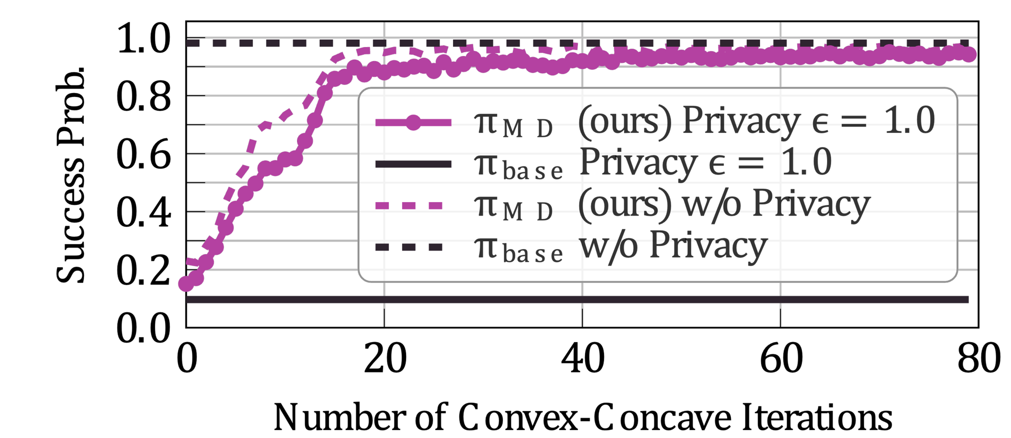
$$\sup_{\boldsymbol{\pi}} \boldsymbol{v^{tr}} - \delta l^{tr} - \beta C_{\boldsymbol{\pi}}$$

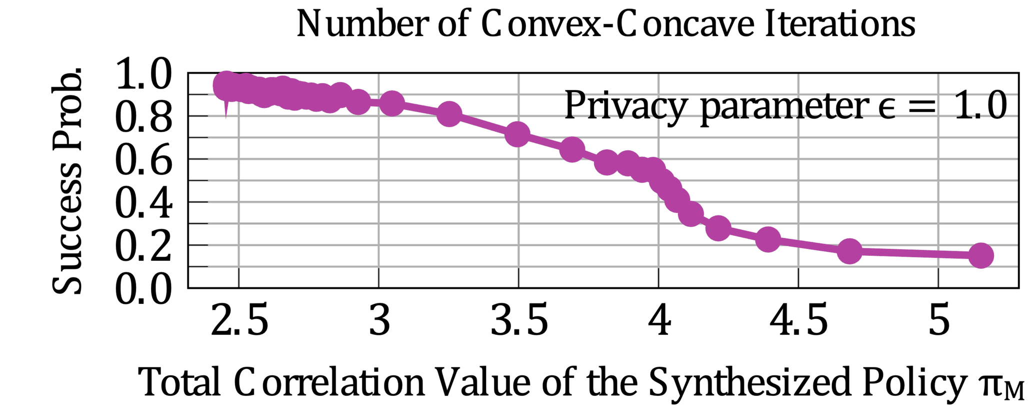
- We represent this optimization problem using occupancy measures, where the occupancy measure x_{s^i, a^i} denotes the expected times action a^i is taken at state s^i
- The objective function contains concave and convex functions of the occupancy measures, thus we solve with the convex-concave procedure
- See "Planning Not to Talk: Multiagent Systems that are Robust to Communication Loss" Karabag et al. 2022 for more details

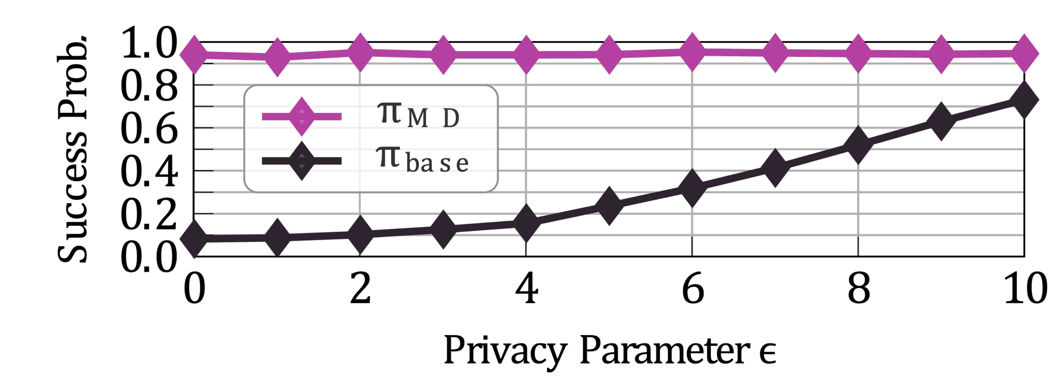
Navigation Example

- Two agents, one safe corridor and one risky corridor
- Each agent slips with probability 0.05

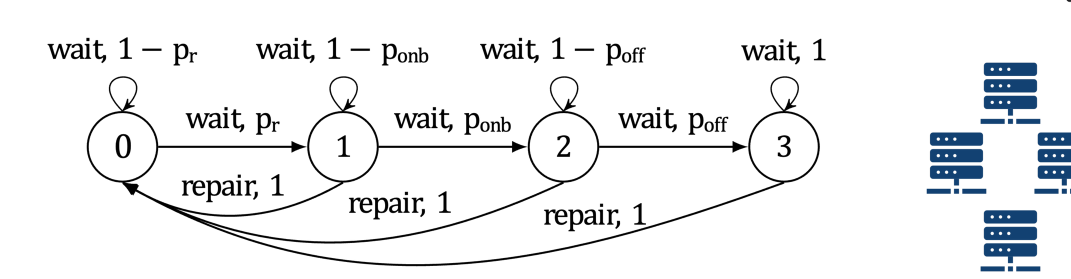




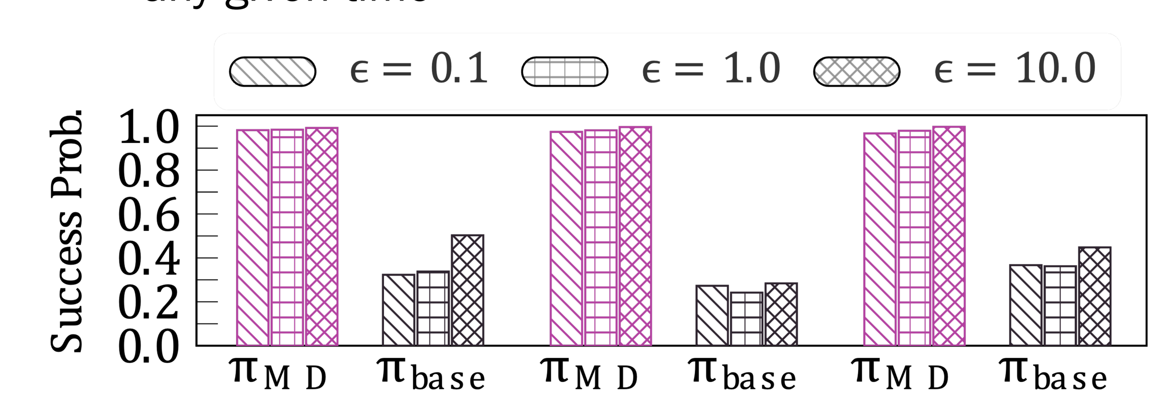




SysAdmin Example



- Collection of 4 servers, each server has four local states:
  - State 0: In repair
  - State 1: Nominal
  - State 2: Needs repairs
  - State 3: Offline
- Fix $p_r = 0.9$, $p_{onb} = 0.1$, $p_{off} = 0.1$
- Goal: Reach a joint state where every server is nominal and have at least two servers running at any given time



Initial config:	One server in need of repair.	Two servers in repair.	Two in repair, two offline.