Differential privacy, entropy and security in distributed control of cyber-physical systems
General Question

For distributed control systems, how expensive is it to preserve privacy? How to optimize?

Navigation
- Routing delays vs location privacy

Smart Grid
- Peak demand vs schedule privacy
Section I: On Differential Privacy of Distributed Control System
Distributed control

Consider a network of vehicles evolving in a shared environment (road congestion)

State of each agent (vehicle) $x_i$
- Evolve with coupled dynamics (delays)

Agents want to share state to estimate delays

Private preferences $p_i$,
- Initial states + sequence of waypoint

Report value $z_i = x_i + noise$

Dynamics of agent:

\[
\begin{align*}
    z_i &= x_i + w_i \\
    u_i &= g(x_i, p_i, z) \\
    x_i^+ &= f(x_i, x, u_i)
\end{align*}
\]
Some notations

\[ z_i = x_i + w_i \]
\[ u_i = g(x_i, p_i, z) \]
\[ x_i^+ = f(x_i, x, u_i) \]

- Sensitive data set: \( D = \{p_i\}_{i \in [N]} \) collects agent preference
  - Two data set \( D, D' \) are adjacent if they differ in one agent’s data
- Observation sequence: \( O = \{z(t)\}_{t \in [T]} \in \mathbb{R}^{nNT} \)
- Trajectory: \( \xi = \{x(t)\}_{t \in [T]} \),
  - Fully defined by a data set \( D \) and observation \( O, \xi_{D,O} \)
\( \varepsilon \)-differential privacy

**Definition:** The randomized communication is \( \varepsilon \)-differentially private with \( \varepsilon > 0 \), if for all adjacent datasets \( D \) and \( D' \) for all subset of observations \( S \),

\[
\Pr[O_D \in S] \leq e^{\varepsilon} \Pr[O_{D',} \in S]
\]

- Difference in one agent’s data doesn’t change the output distribution much
- Small \( \varepsilon \), high privacy; \( \varepsilon \rightarrow 0 \), no communication; \( \varepsilon \rightarrow \infty \), no privacy

- How to design the noise to achieve \( \varepsilon \)-differential privacy?

[Dwork2006], [Ny2014], [Huang2012]
Laplace mechanism for one-shot queries [Dwork06]

No dynamics involve, just exchanging initial states

- $p_i \in \mathbb{R}$ is the initial state of agent $i$

Laplace mechanism: $z_i = p_i + \text{Lap}\left(\frac{1}{\epsilon}\right)$ gives $\epsilon$-differential privacy for any $\epsilon$

- $\text{Lap}\left(\frac{1}{\epsilon}\right)$ has p.d.f. $f(x) = \frac{\epsilon}{z} e^{\epsilon|x|}$
- $\forall x, x': \frac{f(x)}{f(x')} \leq e^{\epsilon |x-x'|}$
- The average reported value is $\sum z_i$ which gives DP with accuracy bounds
When dynamics come into the picture

Definition: the sensitivity of the system is supremum 1-norm between agent trajectories

\[ S(t) = \sup_{\text{adj}(D,D')} \| \xi_{D,O,i}(t) - \xi_{D',O,i}(t) \|_1 \]

- Sensitivity is a property of dynamics of the network
- It can be computed [HiCoNS2014], [CAV2014]
Laplace Mechanism for dynamical systems

**Theorem:** The following distributed control system is $\epsilon$-differentially private:

- at each time $t$, each agent adds an vector of independent Laplace noise $\text{Lap}\left(\frac{S(t)T}{\epsilon}\right)$ to its actual state:

  $$z(t) = x_i(t) + \text{Lap}\left(\frac{S(t)T}{\epsilon}\right)$$

- Larger time horizon, higher privacy level, larger sensitivity $\Rightarrow$ more noise $\Rightarrow$ worse accuracy
Cost of Privacy

Average Cost: $Cost_p = \frac{1}{N} \sum_{t=0}^{T} \sum x_i(t) - p_i(t) |^2$

Baseline cost $\overline{Cost}_p$: the cost when $z_i(t) = x_i(t)$
• No noise

The Cost of Privacy of a DP mechanism $M$ is:
$$CoP = \sup_p \mathbb{E}[Cost_p - \overline{Cost}_p]$$
CoP for linear dynamical system

For stable dynamics: \( \text{CoP} \sim O\left(\frac{T^3}{N^2\epsilon^2}\right) \), otherwise exponential in \( T \)
Summary

Extend the notion of differential privacy to dynamical systems

Generalize Laplace mechanism to dynamical observation using sensitivity of trajectories

For stable dynamics $\text{CoP} \sim O\left(\frac{T^3}{N^2\varepsilon^2}\right)$, otherwise, exponential in $T$
Section II: Entropy-minimization of Differential Privacy
Feedback control system

\[ z = x + w \]
\[ x^+ = f(x, z) \]

- Feedback control of agent:
  - Sensitive data: \( x_0 \) initial state of agent
    - Protecting the initial state is equivalent to protecting the whole trajectory
  - Observation sequence: \( O = \{z(t)\}_{t \in [T]} \)

- Question: how much information is lost by adding noise? How to minimize the information loss?
Estimation & Entropy

**Definition.** An *estimate* of the agent’s initial state is the expectation of the initial state given the history of the agents’ report

\[ \tilde{x}_t = \mathbb{E}[x_0 | z_0, z_1, ..., z_t] \]

**Definition.** The *entropy* of a random variable \( x \) with probability distribution function \( f(x) \) is defined as

\[ H(x) = -\int f(x) \ln x \, dx \]
Entrophy-minimization problem

For minimizing the amount of information loss for achieving differential privacy, we design the noise $w$ to be added:

\[\text{Minimize } H(\tilde{x}_t)\]
\[\text{Subject to: } \forall a, b: \ P[\tilde{x}_t = a] \leq e^{\epsilon|a-b|}P[\tilde{x}_t = b]\]
Result for one-shot case

\[ z = x + w \]

The estimate \( \tilde{x} \in \mathbb{R}^n \) is computed by the first observation \( z \in \mathbb{R}^n \), no dynamics is involved.

**Theorem**: The lower-bound of estimate entropy is

\[ n - n \ln \frac{\epsilon}{2} \]

which is achieved by adding Laplace noise \( w \sim \text{Lap}(1/\epsilon) \).
Sketch of proof [CDC14]

• Let $p(x, z)$ be the joint distribution of initial state $x$ and report $z$, we find a symmetric property

• Claim 1: for any $x$, $p(x, z - x)$ is even
  • Since the noise to add is $n = z - x$, the noise is mean-zero

• Claim 2: for any $c$, $p(x, z) = p(2c - x, 2c - z)$
  • The noise added is independent of the state

• We can define $f(w) = f(z - x) = p(x, z)$

• Claim 3: $f(w)$ is non-decreasing
Extension with dynamics

\[
\begin{align*}
z &= x + w \\
x^+ &= f(x, z)
\end{align*}
\]

The estimate \( \hat{x}_t = \mathbb{E}[x_0|z_0, z_1, ..., z_t] \) is computed by the first \( t \) observation \( \{Z_s\}_{s \in [T]} \)

• **Theorem**: The lower-bound of estimate entropy is still \( n - n \ln \frac{\epsilon}{2} \), which is achieved by a Laplace mechanism.
Optimal Laplace mechanism

\[ z = x + n \]
\[ x^+ = f(x, z) \]

- The first noise to add is the same as the one-shot case:
  \[ w_0 \sim \text{Lap}(1/\epsilon) \]
- In the following round \( t > 0 \), the noise to be added is by evolving the initial noise with the dynamics:
  \[ w_t = \xi(w_0, t) \]
Summary

• Formulate a general estimation problem for which we want to minimize the entropy of estimate

• Prove a lower bound of estimation entropy $n - n \ln \frac{\epsilon}{2}$

• The lower bound is achieved by Laplace mechanism
Section III: Differential Privacy of Distributed Optimization
Architecture

- Local objective functions
- Global constraints
- Communication via the cloud

How to keep objective functions differentially private in communication?
Algorithm

\[ x_i \leftarrow \Pi_{X_i} \left[ x_i - \gamma_t \left( \frac{\partial f_i}{\partial x_i} + \mu^T \frac{\partial g}{\partial x_i} + \alpha_t x_i \right) \right] \]

\[ \mu \leftarrow \Pi_M [\mu + \gamma_t (g(x) - \alpha_t \mu)] \]

\[ \mu \leftarrow \mu + v(t) \]

\[ \gamma_t = \gamma_1 t^{-c_1} \]

\[ \alpha_t = \alpha_1 t^{-c_2} \]

\[ c_1 > c_2, c_1 + c_2 < 1 \]

For \( v(t) = 0 \), the algorithm converges to optima.
Assumptions

- Linear objective functions $f_i(x_i) = a_i x_i$
- Lipschitz Constraints $\left\| \frac{\partial g_j}{\partial x_k} \right\| \leq l_{j,k}$
- Completely correlated noise $\nu(t)$

\[ \text{Constraints } g_1(x), \ldots, g_m(x) \]

\[ \mu + \nu(t) \quad x_1 \quad \mu + \nu(t) \]

Agent 1 $f_1(x_1)$

\[ \ldots \]

Agent n $f_n(x_n)$
Privacy

Two sensitive data $D = \{a_1, \ldots, a_n\}$ and $D' = \{a_1', \ldots, a_n'\}$ are adjacent if they differ only in the $i$th element. The distance between them is $||D - D'|| = ||a_i - a_i'||$.

The algorithm is $\varepsilon$-differentially private if given initial state $x(0), \mu(0)$, the sequence of public multiplier generated by two adjacent sensitive data satisfies

$$Pr \left[ \mu_D^{x(0),\mu(0)} \in O \right] \leq e^\varepsilon ||D - D'|| Pr \left[ \mu_{D'}^{x(0),\mu(0)} \in O \right]$$
Accuracy

\[ x_i \leftarrow \Pi_{X_i} \left[ x_i - \gamma_t \left( \frac{\partial f_i}{\partial x_i} + \mu^T \frac{\partial g}{\partial x_i} + \alpha_t x_i \right) \right] \]

\[ \mu \leftarrow \Pi_{\mathcal{M}} \left[ \mu + \gamma_t (g(x) - \alpha_t \mu) \right] \]

\[ \mu \leftarrow \mu + \nu(t) \]

The loss of accuracy is defined by

\[ \Lambda_D(T) = \max_{x(0) \in X, \mu(0) \in \mathcal{M}} \text{Var} \left[ \mu_{D, \nu(T)}^{x(0), \mu(0)}(T) - \mu_{D, 0}^{x(0), \mu(0)}(T) \right] \]
Sensitivity

Sensitivity: influence of perturbing the sensitive data on observation

For temporary perturbation on $a(s)$, the noise should be

$$\Delta_s(t) = \begin{cases} 
0, & 1 \leq t \leq s \\
\gamma_s \gamma_{s+1} l, & t = s + 1 \\
\gamma_s \gamma_t \Pi_{k=s}^{t-1}(1 - \alpha_k \gamma_k) l, & t \geq s + 2
\end{cases}$$
Noise-adding Mechanism

Mechanism: Add noise

\[ v(t) = \begin{cases} 
0, & t = 1 \\
\gamma_1 \gamma_2 l w, & t = 2 \\
\gamma_t \left( \gamma_{t-1} + \sum_{s=1}^{t-1} \gamma_s \prod_{k=s+1}^{t-1} (1 - \alpha_k \gamma_k) \right) l w, & t \geq 3
\end{cases} \]

\[ w \sim \text{Lap} \left( \frac{1}{\xi} \right) \]

Asymptotics

\[ v(t) \preceq \frac{\gamma_1 l w t^{-(c_1 - c_2)}}{\alpha_1}, \]
Trade-off

The loss of accuracy is bounded asymptotically by

\[ \Lambda_D(T) \leq \frac{2T^{2c_2}l}{\alpha_1^2 \varepsilon^2} \]

higher privacy level ↔ smaller \( \varepsilon \) ↔ larger \( \Lambda_D \) ↔ larger error
Simulations

The dual trajectory $\mu_{D,\nu}^{x(0),\mu(0)}(T)$ and $\mu_{D,0}^{x(0),\mu(0)}(T)$

$\frac{|\mu_{D,\nu}^{x(0),\mu(0)}(T) - \mu_{D,0}^{x(0),\mu(0)}(T)|}{\Lambda_D(T)}$
Summary

- Privacy in distributed optimization
- Trade-off between privacy and accuracy