- Scribes?
- Lecture recording

- Secure computation (how?): given a functionality $F$ to compute, how to do it securely?

- Privacy (what?): what functionalities $F$ are "safe" to compute in the first place?

Given database $D$, want to answer query $q$ on $D$

will give an approximate/noisy answer $q^*(D)$.

Informally: releasing $q^*(D)$ is private if the answer would be "roughly the same" whether or not a particular user's data was in $D$

**Differential Privacy**

Let $D = (x_1, \ldots, x_n) \in \mathbb{X}^n$ ($x_i$ is data of user $i$)

$D, D'$ are neighboring if they differ in data of one user.

Mechanism $M: \mathbb{X}^n \to Y$ is $\epsilon$-diff. private if for all neighboring $D, D'$ and all $T \subseteq Y$,

$$\Pr[M(D) \in T] \leq e^\epsilon \Pr[M(D') \in T]$$
M is $(\varepsilon, \delta)$-differentially private if for $D, D', T$ as above

$$\Pr[M(D) \in T] \leq e^\varepsilon \cdot \Pr[M(D') \in T] + \delta$$

Note: $\varepsilon = \Omega(\frac{1}{n})$, $\delta$ can be cryptographically small

need to look at privacy/utility tradeoff

**Composition**

* If $M$ is $\varepsilon$-differentially private and $D, D'$ differ in data of $K$ users, then for any $T \subseteq Y$,

$$\Pr[M(D) \in T] \leq e^{K \varepsilon} \cdot \Pr[M(D') \in T]$$

* If $M_1, \ldots, M_k$ are $\varepsilon$-differentially private,

then $(M_1 \times \ldots \times M_k)$ is $k\varepsilon$-differentially private.

In fact, if $k \leq \frac{1}{\varepsilon^2}$, then for any $\delta > 0$

$(M_1 \times \ldots \times M_k)$ is $O((k \log \frac{1}{\delta})^2 \cdot \varepsilon, \delta)$-differentially private

**Laplace mechanism**

For a query $q : X^n \to \mathbb{R}$, define global sensitivity:

$$GS_q = \max_{D \sim D'} |q(D) - q(D')|$$

Idea: answer query $q$ by returning $q(D) + \text{noise}$, where

noise depends on $GS_q \pm \varepsilon$.

What distribution to use?
\( \text{Lap}(\sigma) : \Pr[2] \propto e^{-12/\alpha} \)

\( \text{mean } 0, \text{ std. dev } \sqrt{\gamma} \)

\( \Pr[\text{Lap}(\sigma) > \sigma - t] \leq e^{-t} \)

**Laplace mechanism:** return \( g(0) + \text{Lap}(\frac{G \cdot \sigma}{\varepsilon}) \)

**Theorem:** This is \( \varepsilon \)-diff. private

**Proof:** Fix \( D \sim D', t \).

\[
\frac{\Pr[M(0) = t]}{\Pr[M(0') = t]} = \frac{\Pr[\text{Lap}(G \cdot \sigma/\varepsilon) = t - g(0)]}{\Pr[\text{Lap}(G \cdot \sigma/\varepsilon) = t - g(0')]}
\]

\[
= \frac{e^{-\varepsilon \cdot |t - g(0)|/G \cdot \sigma}}{e^{-\varepsilon \cdot |t - g(0')|/G \cdot \sigma}} \leq e^\varepsilon \quad \blacksquare
\]

**Utility:** \( \Pr[M(0) - g(0) \geq \frac{G \cdot \sigma}{\varepsilon} \cdot \log \frac{1}{\beta}] \leq \beta \)

**Exponential mechanism**

Abstract mechanism based on scoring function \( \text{score}(D, y) \)

Let \( G \sigma = \max_y \max_{D \sim D'} |\text{score}(D, y) - \text{score}(D', y)| \).

**Mechanism:** On input \( D \), output \( y \) w/ prob.

\[
\text{proportional to } e^{\varepsilon \cdot \text{score}(D, y)/2 \cdot G \sigma}.
\]

This is \( \varepsilon \)-diff. private for any scoring function

**Utility:** w/ prob. \( O(1) \), the output \( y \) satisfies

\[
\text{score}(D, y) \geq \max_{y'} \text{score}(D, y') - O\left(\frac{G \sigma \log |Y|}{\varepsilon}\right)
\]