1 Introduction

1. Privacy properties:
   1. Unlinkability
   2. Anonymity and Pseudonymity
   3. Plausible deniability (can't prove the user knows smth or not)
   4. Undetectability (exist item or not) and Unobservability (undetectability + anonymity of all users)
   5. Confidentiality
   6. Content awarness
   7. Policy and consent compliance

2. Fair info. principles (FIPS)
   1.) Collection Limitation
   2.) Data Quality
   3.) Purpose Specification
   4.) Use Limitation
   5.) Security Safeguards
   6.) Openness
   7.) Individual Participation
   8.) Accountability

3. Privacy by design

2 Requirements

1. Naive Anonymization
   - Just replace identifiers (or delete) it is not enough.
   - Example: AOL (cross referencing them with phonebook listings)

Req: #1 Resilience to background knowledge.

2. Privacy by obscurity
   - Data perturbation with secret parameters.
   - Kerckhoff: mechanisms must be secure even if 1. algo is known 2. parameters know 3. keys not.

Req #2. Privacy without obscurity

3. Multiple Releases

- Health records
- Req #3: Composition over multiple releases

4. Post-processing (simple suppression is not enough)

#4 Post-processing the output of a privacy mechanism must not change the privacy gurantee

3 Differential Privacy

1. Dinur-nissim reconstruction attack:

Queries: $S \subseteq [n]$, $S \in \{0,1\}^n$

True answ: $A(S) = d * S$

Options:

1. $r(S) = A(S)$
2. Add noise:

do what you want

return $|r(s) - A(s)| \leq E$

def A - blatatly non-private if an adversary can construct a database $c \in \{0,1\}^n$ such that it matches the true DB in all but $o(n)$ entries.

3. Attacks:

1. Inefficient Attack (Exponential)

- if the analyst is allowed to ask 2^n subset queries and the curators adds noise with some bound E, then based on result, the adv. can reconstract the database in all but 4E position.

- Attack:
  - Ask all 2^n queries (Ex. [1,0,0...] = [0,1,0...]
  - Check every possible DB c
  - $|\sum c_i - r(S)| > E$: yes? rule out c, no? okey we found.
  - All other candidate DB differ from correct one by 4E at most

  □ $I_0 = \{i \mid d_i = 0\}$, $I_1 = \{i \mid d_i = 1\}$

  c: $|\sum c_i - r(I_0)| \leq E$ - adversary side ; $|\sum d_i - r(I_0)| \leq E$ by triangle

  c and d differ by at most 2E.

2. Polynomial

- if the analyst is allowed to ask $O(n)$ queris to a dataset of n users, and the curator adds noise with some bound $E = o(\sqrt{n})$ (ultimately smaller than), then based on the results, a computationally efficient adversary can reconstruct the database in all but $O(n)$ positions.

2. Diff Privacy

$\boxed{1.}$ - Any single element in a dataset should have only a limited impact on the output. (Undetectability example.)

- $\boxed{\text{def}}$ A mechanism $M: X \to Y$ is **ε-diff. private** (ε-DP) if for any two neighboring DB D_1 and D_2:

$$\forall T \subseteq Y : \Pr[M(D_1) \in T] \le e^{\varepsilon} \Pr[M(D_2) \in T].$$

- ε small → strong privacy.
- all known M that diff. private are randomized functions.
- for small ε: $e^{\varepsilon} \approx 1 + \varepsilon$

$\boxed{2.}$ Properties:

1. Safety against post-processing

$M: X \to Y$ is ε-diff. priv.

$A: Y \to Z$, and $A \circ M: X \to Z$ also ε-diff. private. we can not unprivatized.

Theorem 6. *Let $M : \mathcal{X}^n \to \mathcal{Y}$ be ε-differentially private, and let $F : \mathcal{Y} \to \mathcal{Z}$ be an arbitrary randomized mapping. Then $F \circ M$ is ε-differentially private.*

Proof. Since F is a randomized function, we can consider it to be a distribution over deterministic functions f. The privacy proof follows for every neighbouring dataset X, X' and $T \subseteq \mathcal{Y}$:

$$\begin{aligned}\Pr[F(M(X)) \in T] &= \mathbf{E}_{f \sim F}[\Pr[M(X) \in f^{-1}(T)]] \\ &\le \mathbf{E}_{f \sim F}[e^{\varepsilon} \Pr[M(X') \in f^{-1}(T)]] \\ &= e^{\varepsilon} \Pr[F(M(X')) \in T].\end{aligned}$$

□

2. Sequential composition

$M_1 : X \to Y_1$ ε_1-DP

$M_2 : X \to Y_2$ ε_2-DP

- $M : X \to Y_1 \times Y_2$ as $M(x) = (M_1(x), M_2(x))$. Then M is $(\varepsilon_1 + \varepsilon_2)$-DP.

3. Parallel composition

- $M: X \to Y$ ε-DP, $X = x_1 \cup x_2 \dots \cup x_k$
- $M: x_1 \to y_1$, ε-DP .... $M: x_k \to y_k$, ε-DP

- applying same M k times by seq. composition it should satisfy $k \cdot \varepsilon$-DP
- But: contributions only once → M sees each individual's data once, so ε

4 Group Privacy

- $M: X \to Y$ - ε- diff private
- D_1 and D_2 differ in k position
- $\forall T \subseteq Y$: $P[M(D_1) \in T] \leq e^{k \cdot \varepsilon} P[M(D_2) \in T]$

Proof. The proof follows by what is known in the business as a "hybrid" argument. Let $X^{(0)} = X$, $X^{(k)} = X'$ – since they differ in k positions, there exists a sequence $X^{(0)}$ through $X^{(k)}$ such that each consecutive pair of datasets is neighbouring. Then, for all $T \subseteq \mathcal{Y}$:

$$\begin{aligned}\Pr[M(X^{(0)}) \in T] &\leq e^{\varepsilon} \Pr[M(X^{(1)}) \in T] \\ &\leq e^{2\varepsilon} \Pr[M(X^{(2)}) \in T] \\ &\dots \\ &\leq e^{k\varepsilon} \Pr[M(X^{(k)}) \in T].\end{aligned}$$

□

$\boxed{3.}$ Mechanisms for DP:

$\boxed{\text{def}}$ Let $f: \mathcal{X}^n \to \mathbb{R}^k$. The ℓ_1- sensetivity of f is ; $q: X \to R$

$$\Delta^{(f)} = \max_{D_1, D_2} \|q(D_1) - q(D_2)\|_1$$, $D_{1,2}$ - neighboring dataset.

1. Laplace Mechanism (the noise level needs to depend on the magnitude of the possible answer).

- Add noise from Laplacian distribution

$$Lap(\mu, b): f(x) = \frac{1}{2b} e^{\left(\frac{-|x-\mu|}{b}\right)}, \quad Var = 2b^2$$

- usually $\mu = 0$
- $b = \frac{S}{\varepsilon}$ then DP.

- Laplace Mechanism: Compute $f(x)$ and output $f(x) + v$ where v follows a Laplace Distribution with $b = \frac{\Delta f}{\varepsilon}$.
- Claim: Laplace mechanism obeys $\varepsilon - DP$.
- Proof: Let x and x' be be neighboring databases, and p_x and $p_{x'}$ are output distributions. Lets also fixed an arbitrary output value $o \in R$. We have:

$$\frac{p_x(o)}{p_{x'}(o)} = \frac{\frac{1}{2b} e^{-|f(x)-o|b}}{\frac{1}{2b} e^{-|f(x')-o|b}}$$

such that $(|f(x') - o| - |f(x) - o|) \geq |f(x') - f(x)|$. Therefore,

$$e^{\frac{(|f(x')-o|-|f(x)-o|)}{b}} \leq e^{\frac{|f(x')-f(x)|}{b}} = e^{\cancel{\Delta f}\frac{\varepsilon}{\Delta f}} = e^{\varepsilon}$$

Example:

a) Counting query, $\ell = 1$

б) Sum query, largest value

в) Average query, $\frac{\text{largest}}{n}$

- Clipping (cut values to enforce the bounds)
  - Need to use DP queries, and include the ε in the total privacy cost.

ε=0.01 for seq. and ε=0.99 for the sum.

- $E\,(\text{true answer} - \text{noisy answer})^2 = \mathrm{Var}\left(\mathrm{Lap}\left(\frac{S(q)}{\varepsilon}\right)\right) = 2\,\frac{S(q)^2}{\varepsilon^2}$, $S(q)\uparrow\ \downarrow\varepsilon \to E\uparrow$

2. Exponential Mechanism

- Generalized sensetivity

$$\Delta u = \max_{R,D,D'} |u(D,R) - u(D',R)|$$

- Given input d, set of output R, scoring function, the exp. mechanism samples an output $r\in R$ with p proportional to

$$\exp\left(\frac{\varepsilon \cdot u(d,r)}{2\Delta u}\right)$$

4 Approximate Differential Privacy

1. - $\forall T \in Y,\ P[M(D_1)\in T] \le e^{\varepsilon} P[M(D_2)\in T] + \delta$ ← M is (ε,δ)-DP.
   - δ represents a "failure p". with $p = 1-\delta$ - pure DP, δ - no privacy at all.
   - Properties:
     1. Sequential composition

        $M_1: X \to Y_1$ $(\delta_1, \varepsilon_1)$-DP

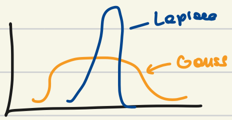
        $M_2: X \to Y_2$ $(\delta_2, \varepsilon_2)$-DP

        - $M: X \to Y_1 \times Y_2$ as $M(x) = (M_1(x), M_2(x))$. Then M is $(\delta_1+\delta_2, \varepsilon_1+\varepsilon_2)$-DP
     2. Post-processing
     3. Parallel composition

2. Gaussian mechanism

   - $G(x) = f(x) + N(\delta^2)$
   - (ε,δ)-DP if $\delta^2 = \frac{2s^2 \ln\left(\frac{1.25}{\delta}\right)}{\varepsilon^2}$

   $$p(x) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{(x-\mu)^2}{2\delta^2}\right)$$



3. Vector-valued functions

   - $f: D \to R^k$
   - Ex: Histograms, ML
   - $\Delta_i^q = \max_{\substack{D_1,D_2:\\ d(D_1,D_2)\le 1}} |q(D_1) - q(D_2)|$ → • L_1 norm • L_2 norm (smaller L_2 sensetivity => better privacy)

To illustrate one difference between the Laplace and Gaussian mechanism, let's consider the problem of estimating the mean of a multivariate dataset. Suppose we have a dataset $X \in \{0,1\}^{n\times d}$, and we wish to privately estimate $f(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$. The largest difference of this statistic between two neighbouring datasets is $\frac{1}{n}\vec{1}$. This is a vector with ℓ_1-norm of $\frac{d}{n}$, and ℓ_2-norm of $\frac{\sqrt{d}}{n}$, which define the ℓ_1 and ℓ_2 sensitivities, respectively. Using the Laplace mechanism to privatize f, we add Laplace noise of magnitude $\frac{d}{n\varepsilon}$ to each coordinate – this gives an ε-DP estimate of f with ℓ_2 error of magnitude $O(\frac{d^{3/2}}{n\varepsilon})$. On the other hand, if we use the Gaussian mechanism, we add Gaussian noise of magnitude $O(\frac{\sqrt{d\log(1/\delta)}}{n\varepsilon})$ to each coordinate – this gives an (ε,δ)-DP estimate of f with ℓ_2 error of magnitude (roughly) $O(\frac{d}{n\varepsilon})$. This example shows that the Gaussian mechanism can add a factor of $O(\sqrt{d})$ less noise (albeit for a marginally weaker privacy guarantee), thus indicating that in some cases it may be better suited for multivariate problems.

④ Homomorphic Encryption

1. Homomorphism
2. UnPadded RSA
3. Semantic Security
4. Partially Homo. Encryption (PHE) $\rightarrow$ ELGamaL
5. Somewhat Homo. Encryption
6. Fully Homo. Encryption: Bootstrapping method