Rigorous Foundations for Statistical Data Privacy

Adam Smith
Boston University

CWI, Amsterdam
November 15, 2018
“Privacy” is changing

• Data-driven systems guiding decisions in many areas

• Models increasingly complex

Benefits of data (better diagnoses, lower recidivism…)
Privacy in Statistical Databases

Large collections of personal information
- census data
- medical/public health
- online advertising
- education
Two conflicting goals

• **Utility**: release aggregate statistics
• **Privacy**: individual information stays hidden

How do we define “privacy”?

• Studied since 1960’s in
  - Statistics
  - Databases & data mining
  - Cryptography
• This talk: **Rigorous foundations and analysis**
“Relax – it can only see metadata.”
This talk

• Why is privacy challenging?
  ➢ Anonymization often fails
  ➢ Example: membership attacks, in theory and in practice

• Differential Privacy [DMNS’06]
  ➢ “Privacy” as stability to small changes
  ➢ Widely studied and deployed

• The “frontier” of research on statistical privacy
  ➢ Three topics
First attempt: Remove obvious identifiers

“AI recognizes blurred faces”
[McPherson Shokri Shmatikov ’16]

Name: [blank]
Ethnicity: [blank]

[Gymrek McGuire Golan Halperin Erlich ’13]

Everything is an identifier

[On Taxis and Rainbows
Lessons from NYC’s improperly anonymized taxi logs

[Hospital A]
[Anonymize]
[Attacker]

[Hospital B]
[Anonymize]

[Ganta Kasiviswanathan S ’08]

Images: whitehouse.gov, genesandhealth.org, medium.com
Is the problem granularity?

What if we only release aggregate information?

Statistics together may encode data
- Example: Average salary before/after resignation

- More generally:
  Too many, “too accurate” statistics reveal individual information
  - Reconstruction attacks [Dinur Nissim 2003, …, Cohen Nissim 2017]
  - Membership attacks [next slide]

Cannot release everything everyone would want to know
A Few Membership Attacks

- [Homer et al. 2008] Exact high-dimensional summaries allow an attacker to test membership in a data set
  
  Caused US NIH to change data sharing practices

- [Dwork, S, Steinke, Ullman, Vadhan, FOCS ‘15] Distorted high-dimensional summaries allow an attacker to test membership in a data set

- [Shokri, Stronati, Song, Shmatikov, Oakland 2017] Membership inference using ML as a service (from exact answers)
Membership Attacks

Suppose

• We have a data set in which membership is sensitive
  ➢ Participants in clinical trial
  ➢ Targeted ad audience

• Data has many binary attributes for each person
  ➢ Genome-wide association studies
  $d = 1000000$ (“SNPs”), $n < 2000$
Membership Attacks

- Release **exact** column averages
- Attacker succeeds with high probability when there are **more attributes than people**


**Membership Attacks**

- Release exact distorted column averages ($\pm \alpha$).
- Attacker succeeds with high probability when there are more attributes than people and $\alpha \ll \sqrt{d/n}$.

\[
\begin{array}{cccccccc}
0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
\end{array}
\]

- $d$ attributes
- $n$ people
- Alice's data
- Attacker
- Population
- No matter how distortion performed

\[
\begin{array}{cccccccc}
0.5 & 0.75 & 0.50 & 0.50 & 0.75 & 0.50 & 0.25 & 0.25 \\
0.4 & 0.7 & 0.6 & 0.5 & 0.8 & 0.4 & 0.2 & 0.3 \\
\end{array}
\]
Machine Learning as a Service

Model

Prediction API

Training API

Input from users, apps …

Classification

Sensitive!
Transactions, preferences, online and offline behavior
Exploiting Trained Models

Model

Prediction API

Training API

Input from the training set

Classification

Input not from the training set

Classification

recognize the difference
Exploiting Trained Models

- **Prediction API**
- **Training API**

Train a model to...

recognize the difference

... without knowing the specifics of the actual model!
This talk

• Why is privacy challenging?
  ➢ Anonymization often fails
  ➢ Example: membership attacks, in theory and in practice

• Differential Privacy [DMNS’06]
  ➢ “Privacy” as stability to small changes
  ➢ Widely studied and deployed

• The “frontier” of research on statistical privacy
  ➢ Three topics
Differential Privacy

• Several current deployments

Apple  Google  US Census

• Burgeoning field of research

Algorithms  Crypto, security  Statistics, learning  Game theory, economics  Databases, programming languages  Law, policy
Differential Privacy

- **Data set** \( x = (x_1, \ldots, x_n) \in D^n \)
  - Domain \( D \) can be numbers, categories, tax forms
  - Think of \( x \) as **fixed** (not random)
- **\( A = \) randomized procedure**
  - \( A(x) \) is a random variable
  - Randomness might come from adding noise, resampling, etc.
• A thought experiment
  - Change one person’s data (or add or remove them)
  - Will the probabilities of outcomes change?

For any set of outcomes, (e.g. I get denied health insurance) about the same probability in both worlds.
**Definition:** A is $\epsilon$-differentially private if, for all neighbors $x, x'$, for all subsets $S$ of outputs

$$\Pr(A(x) \in S) \leq (1 + \epsilon) \Pr(A(x') \in S)$$

$\epsilon$ is a leakage measure

**Neighboring databases induce close distributions on outputs**
Randomized Response [Warner 1965]

- Say we want to release the proportion of diabetics in a data set
  - Each person’s data is 1 bit: \( x_i = 0 \) or \( x_i = 1 \)
- Randomized response: each individual rolls a die
  - 1, 2, 3 or 4: Report true value \( x_i \)
  - 5 or 6: Report opposite value \( \overline{x_i} \)
- Output is list of reported values \( Y_1, \ldots, Y_n \)
  - Satisfies our definition when \( \epsilon \approx 0.7 \)
  - Can estimate fraction of \( x_i \)’s that are 1 when \( n \) is large
Laplace Mechanism

- Say we want to release a summary $f(x) \in \mathbb{R}^d$
  - e.g., proportion of diabetics: $x_i \in \{0,1\}$ and $f(x) = \frac{1}{n} \sum_i x_i$

- Simple approach: add noise to $f(x)$
  - How much noise is needed?
  - Idea: Calibrate noise to some measure of $f$’s volatility
### Laplace Mechanism

- **Global Sensitivity:**
  \[
  GS_f = \max_{\text{neighbors } x, x'} \| f(x) - f(x') \|_1
  \]

- Example: \( GS_{\text{proportion}} = \frac{1}{n} \)}
Laplace Mechanism

- Global Sensitivity: \( \text{GS}_f = \max_{\text{neighbors } x, x'} \| f(x) - f(x') \|_1 \)

  - Example: \( \text{GS}_{\text{proportion}} = \frac{1}{n} \)

**Theorem:** If \( A(x) = f(x) + \text{Lap} \left( \frac{\text{GS}_f}{\epsilon} \right) \), then \( A \) is \( \epsilon \)-differentially private.

- Laplace distribution \( \text{Lap}(\lambda) \) has density
  \[
  h(y) \propto e^{-|y|/\lambda}
  \]
- Changing one point translates curve
Attacks “match” differential privacy

- Can release $d$ proportions with noise $\approx \frac{\sqrt{d}}{\epsilon n}$ per entry
- Requires “approximate” variant of DP
A rich algorithmic field

Noise addition

Exponential sampling

\[ Y \sim p(y|x) \propto \exp(\epsilon \cdot \text{quality}(y, x)) \]

Local perturbation
Interpreting Differential Privacy

• A naïve hope:
  Your beliefs about me are the same after you see the output as they were before

• Impossible
  ➢ Suppose you know that I smoke
  ➢ Clinical study: “smoking and cancer correlated”
  ➢ You learn something about me
    • Whether or not my data were used

• Differential privacy implies:
  No matter what you know ahead of time,
  You learn (almost) the same things about me whether or not my data are used
  ➢ Provably resists attacks mentioned earlier
Research on (differential) privacy

- **Definitions**
  - Pinning down “privacy”
- **Algorithms**: what can we compute privately?
  - Fundamental techniques
  - Specific applications
- **Usable systems**
- **Attacks**: “Cryptanalysis” for data privacy
- **Protocols**: Cryptographic tools for large-scale analysis
- **Implications for other areas**
  - Adaptive data analysis
  - Law and policy
This talk

• Why is privacy challenging?
  - Anonymization often fails
  - Example: membership attacks, in theory and in practice

• Differential Privacy [DMNS’06]
  - “Privacy” as stability to small changes
  - Widely studied and deployed

• The “frontier” of research on statistical privacy
  - Three topics
Frontier 1: Deep Learning with DP

[Abadi et al 2016, …]

Sensitive Data

Deep Learning

Parameters

d -0.7 -0.9 0.04
d -1.5 -0.8 -0.85
d -1.0 -0.4 -1.2
d 0.1 0.04

d 6.3 0.1

d -0.3 0.3

d 1.3 1.3

d 6.7 5.11

d -2.5 -5.4

d -1.0

d -0.6

d -0.7

d -1.0

c

e -1.2

Model

Thought of as private now, but better to reason as if public

Revealed now, but should be hidden
Frontier 2: From Law to Technical Definitions

Two central challenges

1. Given a body of law and regulation, what technical definitions comply with that law?
   - E.g., what suffices to satisfy GDPR?

2. How should we write laws and regulations so they make sense given evolving technology?
   - E.g., Surveillance ≠ physical wiretaps

• Technical research must inform these questions
  - E.g. ”personally identifiable information” is meaningless

• [Nissim et al. 2016] When tradeoffs are inherent, mathematical formulations play an important role
  - E.g. formal interpretation of FERPA (a US law) mirrors DP
  - “Singling out” in GDPR is challenging to make sense of
Frontier 3: Privacy and overfitting

- Problem: In modern data analysis, data are re-used across studies
  - Choice of what analysis to perform can depend on outcomes of previous analyses

- Differentially private algorithms help prevent overfitting due to adaptivity
This talk

- **Why is privacy challenging?**
  - Anonymization often fails
  - Example: membership attacks, in theory and in practice

- **Differential Privacy [DMNS’06]**
  - “Privacy” as stability to small changes
  - Widely studied and deployed

- **The “frontier” of research on statistical privacy**
  - Three topics
Beyond privacy

• Data increasingly used to automate decisions
  ➢ E.g.: Lending, health, education, policing, sentencing

• Traditional security: controlling intrusion

• Modern security must include trustworthiness of data-driven algorithmic systems

• Differential privacy formalizes one piece of modern security
  ➢ What other areas need such scrutiny?