The Fundamental Limits of Data & Metadata Privacy

Peter Kairouz
ECE Department
University of Illinois at Urbana-Champaign
Communication

- transfer of information from one point in space-time to the other
Wireless communication

- the fundamental limits of wireless communication are well understood
Rise of the planet of the apps!
Rise of the planet of the apps!
Rise of the planet of the apps!
Rise of the planet of the apps!

can we communicate anonymously and privately?
Data and metadata privacy

Bob

Alice

Nikkole Paulun
@nikkoleMTV

Don't broadcast having an abortion on national television & expect to not get crap for it. You're not brave in my eyes.
#sorrynotsorry

7 Jul 12
Part 1: Metadata Privacy
Political activism

some people have important, sensitive things to say
Existing anonymous messaging apps are not truly anonymous!

centralized networks are not truly anonymous!
Anonymous communication

freenet

Tor

OneSwarm
Privacy preserving peer-to-peer data sharing
Anonymous communication

designed for point-to-point communication
Distributed messaging
Distributed messaging
Distributed messaging

what can an adversary do?
Adversary without timing information
Adversary without timing information
Adversary without timing information
Adversary without timing information

adversary can figure out who got the message
Adversary with timing information
Adversary with timing information
Adversary with timing information

Craig

Bob

Mary

David

Alice
Adversary with timing information

adversary can collect **timing information**
Adversary with timing information

- message
- timestamp

adversary can collect timing information
Distributed network forensics

\[ \text{timing} + \text{who has the message} = \text{authorship} \]
Information flow in social networks

- $G$ is the graph representing the social network
Information flow in social networks
Information flow in social networks

- the author passes the message to its neighbors
Information flow in social networks

- its neighbors pass the message to theirs
Information flow in social networks

- the message spreads in all directions at the same rate
Information flow in social networks

- the message spreads in **all directions** at the **same rate**
Information flow in social networks

- the message spreads in **all directions** at the **same rate**
Information flow in social networks

- this spreading model is known as the diffusion model
can we locate the message author?
Concentration around the center

- the message author is in the “center”
Maximum likelihood detection

diffusion spreading = deanonymization

Shah, Zaman 2011
Our goal

engineer the spread to hide authorship
Our goal

\( N_T \): expected number of nodes with the message at time \( T \)
Main Result: Adaptive diffusion
Main Result: Adaptive diffusion
Main Result: Adaptive diffusion
Main Result: Adaptive diffusion
Main Result: Adaptive diffusion
Main Result: Adaptive diffusion
Main Result: Adaptive diffusion
Main Result: **Adaptive diffusion**
Main Result: **Adaptive diffusion**

provides provable **anonymity guarantees**

[Best Paper Award at SIGMETRICS 15]
Part 1.1:
Adversary without timing info
$d$-regular trees: adaptive diffusion
$d$-regular trees: adaptive diffusion

- initially, the author is also the "virtual source"
\(d\)-regular trees: adaptive diffusion

- at \(T = 1\), the author selects one neighbor at random
at $T = 1$, the author selects one neighbor at random
$d$-regular trees: adaptive diffusion

- the chosen neighbor becomes the new virtual source
$d$-regular trees: adaptive diffusion

- at $T = 2$, the virtual source passes the message to all its neighbors
$d$-regular trees: adaptive diffusion

- the new virtual source has two options:
  - keeping the virtual source token
  - passing the virtual source token
Keeping the virtual source token

- all leaf nodes with the message pass it to their neighbors
Passing the virtual source token

- current virtual source selects one of its neighbors at random

new virtual source
Passing the virtual source token

- new virtual source passes the message to its neighbors which in turn pass it to their neighbors
When to keep the virtual source?

- virtual source token is kept with probability

\[
\frac{(d-1)^\frac{T}{2}-h-1-1}{(d-1)^{\frac{T}{2}+1}-1}
\]
Symmetry properties

- the graph is **always symmetric** around the **virtual source**
can we locate the message author?
Maximum likelihood detection

- All nodes except for the final virtual source are equally likely
Main result: adaptive diffusion

1. We spread fast: $N_T \approx (d - 1)^2$

2. All nodes except for the final virtual source are equally likely to be the source, hence

$$P(\hat{v}_{ML} = v^*) = \frac{1}{N_T - 1}$$

3. The expected distance between the estimated and true source is at least $\frac{T}{2}$. 
General graphs

can we extend adaptive diffusion for **general graphs**?
Simulation results: Facebook graph

Probability of detection

- likelihoods can be approximated numerically

\[ \frac{1}{N} \] bound

- adaptive diffusion

number of nodes with the message

- likelihoods can be approximated numerically
Part 1.2: Adversary with timing info
Spy adversarial model

adversary can collect **timing information**
Adversary with timing

![Diagram showing an adversary with timing]

- T = 3
- T = 1
- T = 5
- T = 1
Results on \(d\)-regular trees

**MAIN RESULT:** Probability of detection \(P_D = p + o(p)\)
What about general graphs?
Part 2: Data Privacy
Recent data privacy leaks

Anonymized Netflix data + Public, incomplete IMDB data

= Identified Netflix data

On average, four movies uniquely identify user

Second round of Netflix competition postponed

-deanonymizing Netflix data, identifying personal genomes, etc.

= SSN

Image credit: Arvind Narayanan

Image Credit: Alessandro Acquisti
Global vs. local models

Users ➔ Data ➔ Trusted Curator ➔ Query ➔ Released info ➔ Malicious Analyst

Users ➔ Data ➔ Malicious Analyst ➔ Data ➔ Users
Global vs. local privacy

- $Q$ is a privacy mechanism
Part 2.1: Global Privacy
Global privacy

Database

<table>
<thead>
<tr>
<th>Age</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: 20</td>
<td>A: 20</td>
</tr>
<tr>
<td>B: 35</td>
<td>B: 35</td>
</tr>
<tr>
<td>C: 43</td>
<td>C: 43</td>
</tr>
<tr>
<td>D: 30</td>
<td>D: 30</td>
</tr>
</tbody>
</table>

Query function: $q$

How many people are older than 32?

$q(D) = 2$

Privacy preserving query

- say 0 w.p. 1/10
- say 1 w.p. 2/10
- say 2 w.p. 4/10
- say 3 w.p. 2/10
- say 4 w.p. 1/10
Global privacy

Database

<table>
<thead>
<tr>
<th>Age</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A: 20</td>
<td></td>
</tr>
<tr>
<td>B: 35</td>
<td></td>
</tr>
<tr>
<td>C: 43</td>
<td></td>
</tr>
<tr>
<td>D: 30</td>
<td></td>
</tr>
</tbody>
</table>

Privacy preserving mechanism

Given that the query output is \( q(D) \), say \( Y \) with probability \( Q(Y|q(D)) \).
Global differential privacy

- presence or absence of any individual should not have a significant impact on $Y$

Dataset: $D_1$

<table>
<thead>
<tr>
<th>Age</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A: 20</td>
<td></td>
</tr>
<tr>
<td>B: 35</td>
<td></td>
</tr>
<tr>
<td>C: 43</td>
<td></td>
</tr>
<tr>
<td>D: 30</td>
<td></td>
</tr>
</tbody>
</table>

Dataset: $D_2$

<table>
<thead>
<tr>
<th>Age</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A: 20</td>
<td></td>
</tr>
<tr>
<td>B: 35</td>
<td></td>
</tr>
<tr>
<td>D: 30</td>
<td></td>
</tr>
</tbody>
</table>

Query: $q$

$Q(Y|q(D_1)) \approx Q(Y|q(D_2))$
Global differential privacy

$Q$ is $\varepsilon$ differentially private if $\forall$ neighboring datasets $D_1, D_2$ and all $Y$

$$e^{-\varepsilon} \leq \frac{Q(Y|q(D_1))}{Q(Y|q(D_2))} \leq e^{+\varepsilon}$$

$\varepsilon$ controls the level of privacy

- large $\varepsilon$, low privacy
- small $\varepsilon$, high privacy
Two types of privacy mechanisms

Data independent privacy mechanisms:

\[ Y = q(D) + N \]

Data dependent privacy mechanisms:

\[ Y = q(D) + N(q(D)) \]
The Laplace mechanism

\[ f_N(x) = \text{Lap}\left( \frac{\Delta}{\epsilon} \right) \]

\[ \Delta = \max_{D_1, D_1 \text{neighbors}} |q(D_1) - q(D_2)| \]
Privacy-utility tradeoff

- measuring the performance of a privacy mechanism

\[ L(x - y) = |y - x| \]

\[ L(x - y) = (y - x)^2 \]

true query output
\[ q(D) = x \]

noised query output
\[ q(D) + N(q(D)) = y \]

average loss when \( q(D) = x \):
\[ C(Q, x) = \int L(x - y)Q(y|x)dy \]

- loss functions considered are symmetric and increase with \(|x - y|\)
Privacy-utility tradeoff

- the more private you want to be, the less utility you get
- there is a fundamental tradeoff between privacy and utility

\[
\begin{align*}
\text{minimize} & \quad \sup_{Q} C(Q, x) \\
\text{subject to} & \quad Q \in \mathcal{D}_{\varepsilon}
\end{align*}
\]

worst case loss

set of all differentially private mechanisms
Optimality of staircase mechanisms

- **data independent mechanisms** are optimal

- **staircase mechanisms** are optimal

*Figure showing a flowchart with Trusted Curator and Malicious Analyst.*

*Graph showing a function $f_N$ with various segments.*
∀ neighboring datasets $D_1, D_2$

differentially private mechanisms

$$e^{-\varepsilon} \leq \frac{Q(Y|q(D_1))}{Q(Y|q(D_2))} \leq e^{\varepsilon}$$

staircase mechanisms

$$\frac{Q(Y|q(D_1))}{Q(Y|q(D_2))} \in \{e^{-\varepsilon}, 1, e^{+\varepsilon}\}$$
Multidimensional queries

$q(D) = (q_1(D), ..., q_d(D))$

$Y = (Y_1, ..., Y_d)$

Loss functions:

$L(Y - q(D)) = ||Y - q(D)||_1$

$L(Y - q(D)) = ||Y - q(D)||_2$

average loss when $q(D) = x$:

$C(Q, x) = \int ... \int L(x - y)Q(y|x)dy$

minimize $\sup_x C(Q, x)$

subject to $Q \in D_\varepsilon$
Optimality of staircase mechanisms

- the two dimensional **staircase mechanism** is **optimal** for 2 dimensional queries
- we conjecture that the same result holds for any number of dimensions
Part 2.2: Local Privacy
Local privacy

have you ever used illegal drugs?

say yes

answer truthfully

[Warner 1965]
Privacy via plausible deniability

- instead of \( X = x \), share \( Y = y \) w.p. \( Q(y|x) \)
- we focus on discrete alphabets

\[ Q : |\mathcal{X}| \times |\mathcal{Y}| \] row stochastic matrix
Local differential model

$Q$ is $\varepsilon$ differentially private if $\forall X, X'$, and $Y$

$$e^{-\varepsilon} \leq \frac{Q(Y|X)}{Q(Y|X')} \leq e^{+\varepsilon}$$

$\varepsilon$ controls the level of privacy

large $\varepsilon$, low privacy
small $\varepsilon$, high privacy

[Duchi, et. al., 2013]
Privacy-utility tradeoff

- the more private you want to be, the less utility you get
- there is a fundamental tradeoff between privacy and utility

\[
\text{maximize } \quad U(Q) \\
\text{subject to } \quad Q \in D_\varepsilon
\]

application dependent utility function

set of all differentially private mechanisms
Binary alphabets
Utility functions

Utility functions obeying the data processing inequality:

\[ T = Q \circ W \implies U(T) \leq U(Q) \]

- Further randomization can only reduce utility
Main result: binary data

- for binary alphabets: $|\mathcal{X}| = 2$

- lie w.p. $\frac{1}{e^\varepsilon + 1}$

- say the truth w.p. $\frac{e^\varepsilon}{e^\varepsilon + 1}$

- optimal for all values of $\varepsilon$

- optimal for all $U(Q)$ obeying the data processing inequality

[JMLR 2015]
General alphabets
Private hypothesis testing

\[ X_1 \sim \begin{cases} P_0 \\ P_1 \end{cases}, \quad Y_1 \sim \begin{cases} M_0 \\ M_1 \end{cases}, \quad Y_2 \sim \begin{cases} M_0 \\ M_1 \end{cases}, \quad X_2 \sim \begin{cases} P_0 \\ P_1 \end{cases} \]

Users \rightarrow Q \rightarrow \text{Malicious Analyst} \rightarrow \hat{\nu} \in \{0, 1\} \rightarrow \text{Users}

\[
\text{maximize} \quad D_f\left( \begin{array}{c} QP_0 \\ M_0 \end{array} \right) \parallel \begin{array}{c} QP_1 \\ M_1 \end{array} \right)
\]

subject to \quad Q \in \mathcal{D}_\varepsilon

\( f \)-divergences include KL and total variation distances
Staircase mechanisms

$Q$ is $\varepsilon$ differentially private if $\forall x, x' \in \mathcal{X}$ and all $y \in \mathcal{Y}$

$$e^{-\varepsilon} \leq \frac{Q(y|x)}{Q(y|x')} \leq e^{\varepsilon}$$

$Q$ is a staircase mechanism if $\forall x, x' \in \mathcal{X}$ and all $y \in \mathcal{Y}$

$$\frac{Q(y|x)}{Q(y|x')} \in \{ e^{-\varepsilon}, 1, e^{+\varepsilon} \}$$
Example of staircase mechanisms

\[ Q^T = \frac{1}{1+e^{\varepsilon}} \begin{bmatrix} e^{\varepsilon} & e^{\varepsilon} & 1 & e^{\varepsilon} & 1 \\ 1 & 1 & e^{\varepsilon} & 1 & e^{\varepsilon} \end{bmatrix} \]

\[ Q^T = \frac{1}{3+e^{\varepsilon}} \begin{bmatrix} e^{\varepsilon} & 1 & 1 & 1 \\ 1 & e^{\varepsilon} & 1 & 1 \\ 1 & 1 & e^{\varepsilon} & 1 \\ 1 & 1 & 1 & e^{\varepsilon} \end{bmatrix} \]

Binary Mechanism

Randomized Response
Main result: general case

- for $|\mathcal{X}| = k > 2$

- staircase mechanisms are optimal for all $\varepsilon$
- BM optimal for small $\varepsilon$
- RR optimal for large $\varepsilon$

staircase mechanisms

all differentially private mechanisms

[BIPS 14, JMLR 15]
More general utility functions

- same results hold for a rich class of **convex utility functions**:

\[
\begin{align*}
\text{maximize} & \quad U(Q) = \sum_{y \in \mathcal{Y}} \mu(Q_y) \\
\text{subject to} & \quad Q \in \mathcal{D}_{\varepsilon}
\end{align*}
\]

- $Q_y$: the column of $Q$ corresponding to $Q(y \mid .)$
- $\mu$: any sub-linear function

includes all $f$-divergences and **mutual information**
Rise of the planet of the apps!
Acknowledgments

Quan Geng  Giulia Fanti  Pramod Viswanath  Sewoong Oh