Introduction to Differential Privacy

Jeremiah Blocki
CS-555
11/22/2016

Credit: Some slides are from Adam Smith

Google Scholar

differential privacy

About 3,000,000 results (**0.06** sec)

Articles

Case law

My library

Any time

Since 2016

Since 2015

Since 2012

Custom range...

Differential privacy: A survey of results

C Dwork - International Conference on Theory and Applications of …, 2008 - Springer

Abstract Over the past five years a new approach to **privacy**-preserving data analysis has born fruit [13, 18, 7, 19, 5, 37, 35, 8, 32]. This approach differs from much (but not all!) of the related literature in the statistics, databases, theory, and cryptography communities, in that **...**

Cited by 2557 Related articles All 32 versions Web of Science: 365 Cite Save More

Mechanism design via **differential privacy**

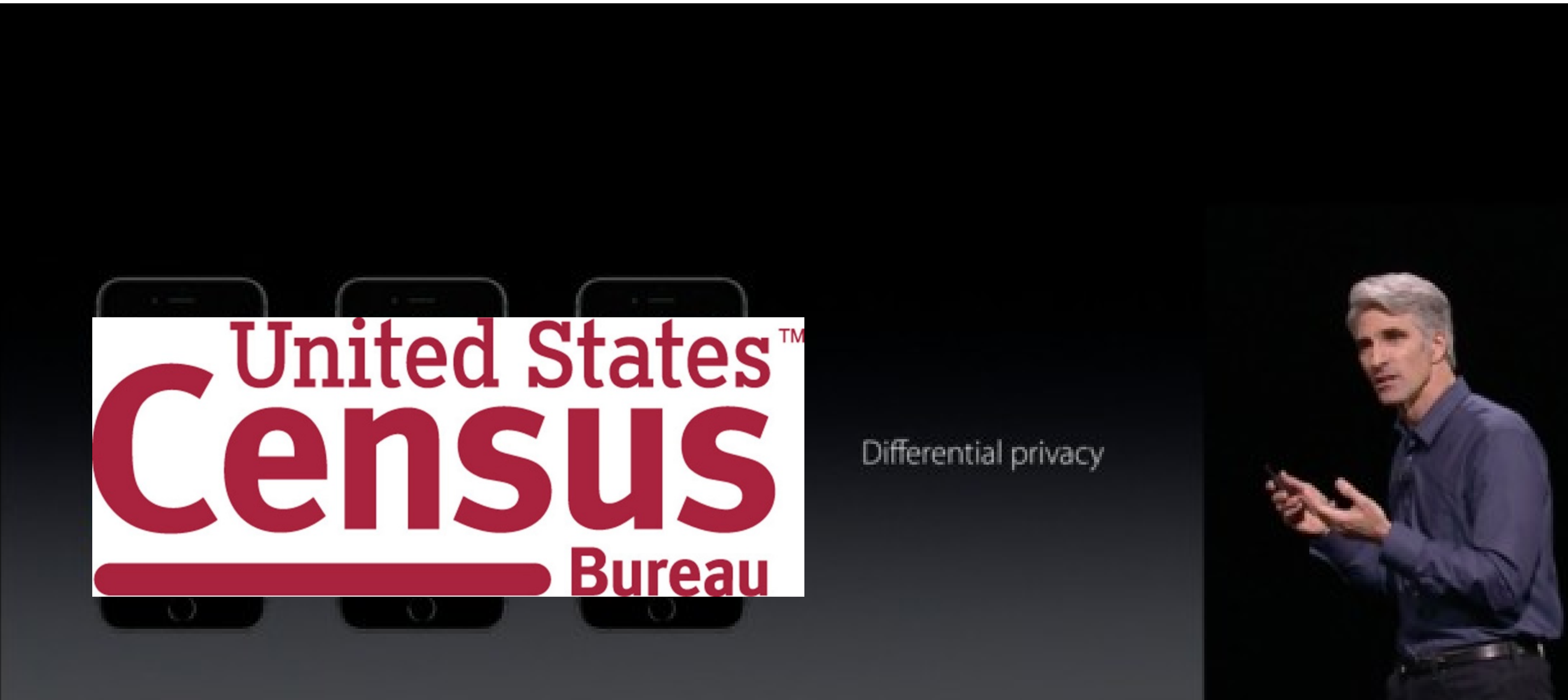
F McSherry, K Talwar - … of Computer Science, 2007. FOCS'07. …, 2007 - ieeexplore.ieee.org

Abstract We study the role that **privacy**-preserving algorithms, which prevent the leakage of specific information about participants, can play in the design of mechanisms for strategic agents, which must encourage players to honestly report information. Specifically, we **...**

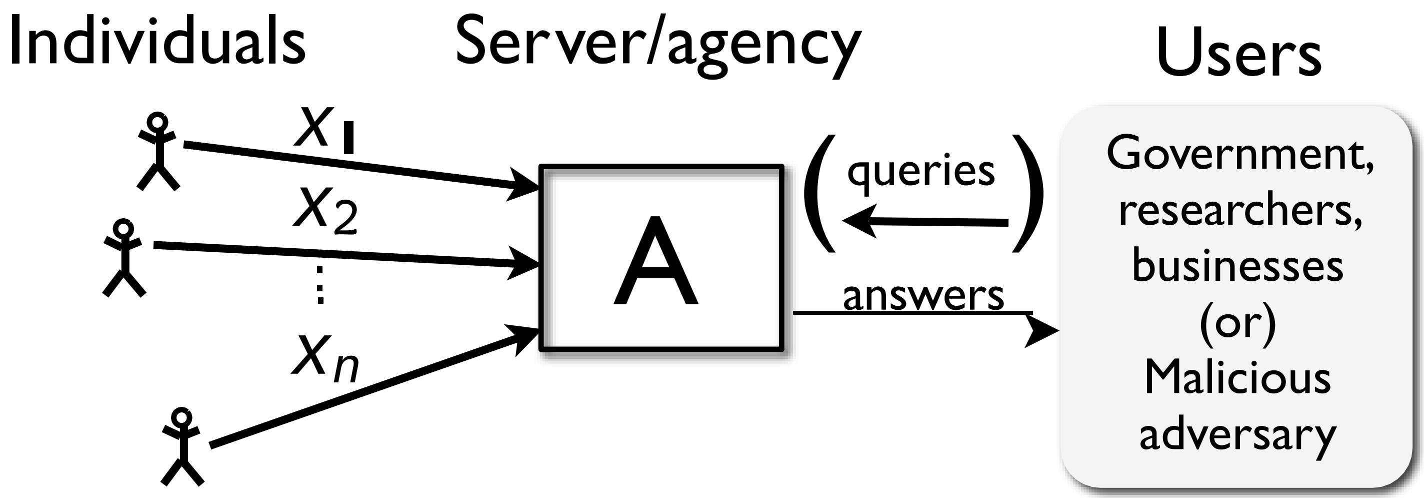
Cited by 708 Related articles All 25 versions Cite Save

Microsoft® Research

Differential Privacy



Privacy in Statistical Databases



- What information can be released?
- Two conflicting goals
  - Utility: Users can extract "global" statistics
  - Privacy: Individual information stays hidden
- How can these be made **precise**?
  - (How context-dependent **must** they be?)

Why not use crypto definitions?

- Cryptography successfully defined concepts such as
  - encryption
  - secure function evaluation
- Recall encryption:
  - "Semantic Security": For any function *f*, distribution on messages and efficient algorithm A, there exists an efficient algorithm A' such that:

$$\Pr[A(PK, Enc_{PK}(m)) = f(m)] \le \Pr[A'(PK) = f(m)] + \varepsilon$$

  - "Indistinguishability": For any message m, no efficient adversary can tell apart encryptions of m and a default message:

$$Enc_{PK}(0)$$

  - Adversary's information quantified precisely
  - Encryption must be **randomized**

$$Enc_{PK}(m)$$

Encryption: Real vs Ideal worlds

- **Real** world: Alice sends Bob encryption of 100-bit message **m**, adversary sees ciphertext
- **Ideal** world: Alice tells adversary "I am sending Bob a message of 100 bits" and nothing else.
- How can you "**simulate**" the ideal world, i.e. make the ideal world look like the real word?
  - Have Alice send encryption of $0^{100} = \underbrace{000000\ldots0}_{\text{100 zeros}}$"
- No adversary can tell the real world from the simulation, and clearly the simulation leaks no information about **m**!

Notes about these definitions

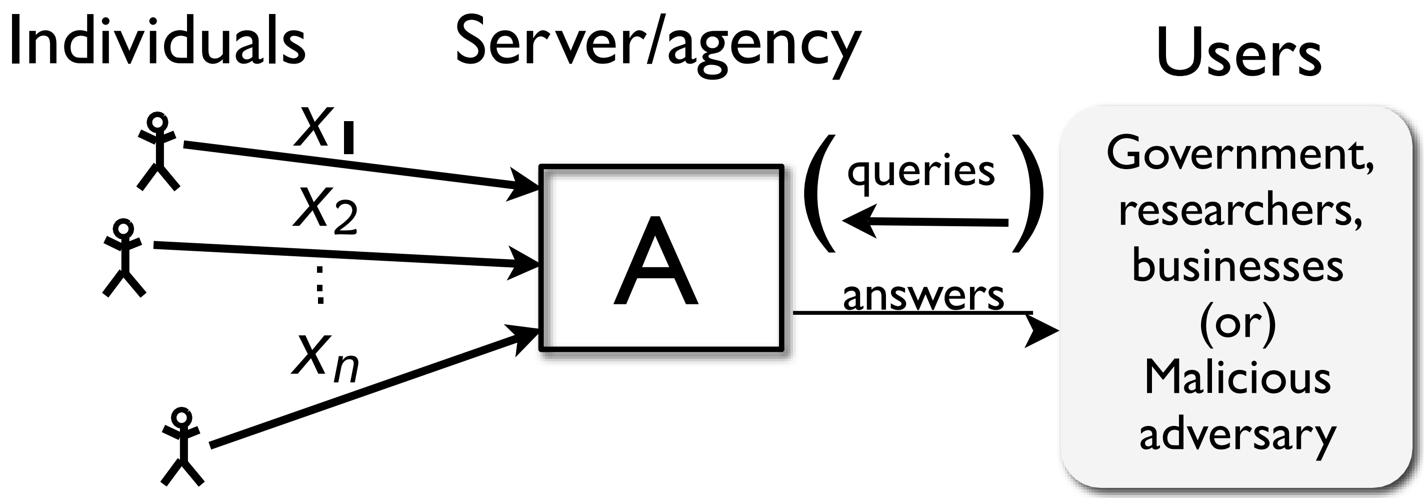
- Security is a property of the **algorithm** used for encryption
  - ➢ You can't point at a particular string and say it is "secure"
- Adversary's information and abilities quantified precisely
- Because we allow adversary side information about the message, all the security resides in the secret key and randomness used for encryption

Secure Function Evaluation

a.k.a. "multi-party computation"

- Several parties, each with input x_i, want to compute a function $f(x_1,x_2,\ldots,x_n)$
- **Ideal world**: all parties hand their inputs to a trusted party who computes $f(x_1,\ldots,x_n)$ and releases the result
- There exist secure protocols for this task
  - ➢ Idea: a simulator can geneerate a dummy transcript given only the value of f
- Privacy: use SFE protocols to jointly data mine
  - ➢ Horizontal vs vertical
  - ➢ Lots of papers (see optional topics)

Privacy in Statistical Databases



- What information can be released?
- Two conflicting goals
  - Utility: Users can extract "global" statistics
  - Privacy: Individual information stays hidden
- How can these be made **precise**?
  - (How context-dependent **must** they be?)

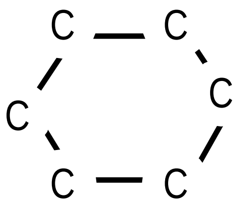
Why not use crypto definitions?

Why not use crypto definitions?

- **Attempt #1**:
  - **Def'n**: For every entry i, no information about x_i is leaked (as if encrypted)
  - **Problem**: no information at all is revealed!
  - Tradeoff privacy vs utility

Why not use crypto definitions?

- **Attempt #1**:
  - **Def'n**: For every entry i, no information about x_i is leaked (as if encrypted)
  - **Problem**: no information at all is revealed!
  - Tradeoff privacy vs utility
- **Attempt #2**:
  - Agree on summary statistics f(DB) that are safe
  - **Def'n**: No information except f(DB)
  - **Problem**: why is f(DB) safe to release?
  - Tautology trap
  - (Also: how do you figure out what **f** is?)



Why not use crypto definitions?

Why not use crypto definitions?

- Problem: Crypto makes sense in settings where the line between "inside" and "outside" is well-defined
  - E.g. psychologist:
    - "inside" = psychologist and patient
    - "outside" = everyone else

Why not use crypto definitions?

- Problem: Crypto makes sense in settings where the line between "inside" and "outside" is well-defined
  - E.g. psychologist:
    - "inside" = psychologist and patient
    - "outside" = everyone else

Why not use crypto definitions?

- Problem: Crypto makes sense in settings where the line between "inside" and "outside" is well-defined
  - E.g. psychologist:
    - "inside" = psychologist and patient
    - "outside" = everyone else
- Statistical databases: fuzzy line between inside and outside

A Problem Case

Question 1: How many people in this room have cancer?

Question 2: How many students in this room have cancer?

The difference (A1-A2) exposes my answer!

Achieving Differential Privacy

- Examples
- Intuitions for privacy
  - Why crypto def's don't apply
- A Partial* Selection of Definitions
  - Two straw men
  - Attribute Disclosure and Differential Privacy
- Conclusions

* "partial" = "incomplete" and "biased"

Achieving Differential Privacy

- Examples
- Intuitions for privacy
  - Why crypto def's don't app
- A Partial* Selection of Definitions
  - Two straw men
  - Attribute Disclosure and Differential Privacy
- Conclusions

Criteria

- Understandable
- Clear adversary's goals & prior knowledge / side information

* "partial" = "incomplete" and "biased"

Straw Man #0

Omit ` ... ıe
data

e.g., N

This h

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov
The University of Texas at Austin

Abstract

We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge.

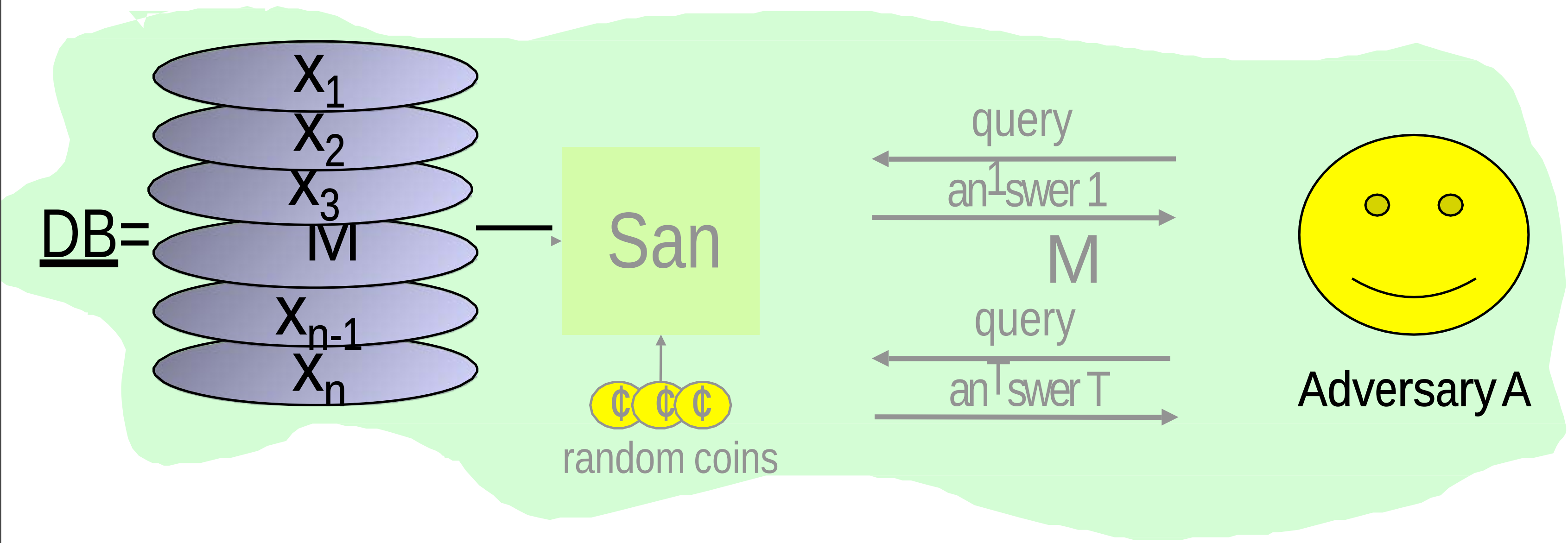
We apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous movie ratings of 500,000 subscribers of Netflix, the world's largest online movie rental service. We demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, we successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.

and sparsity. Each record contains many attributes (*i.e.*, columns in a database schema), which can be viewed as dimensions. Sparsity means that for the average record, there are no "similar" records in the multi-dimensional space defined by the attributes. This sparsity is empirically w ... tail" pl ... record ...

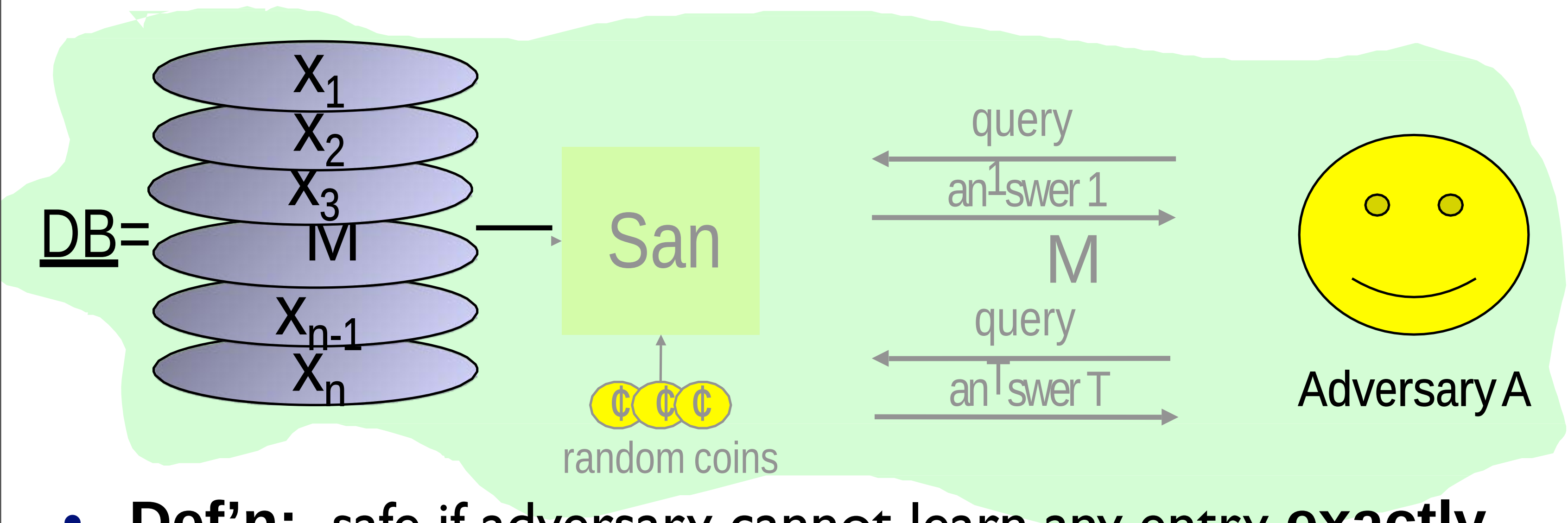
Our c ... model ... (sectio ... probab ... the am ... Unlike ... *ori* tha ... set of " ... compasses a much broader class of de-anonymization attacks than simple cross-database correlation.

Our second contribution is a very general class of de-anonymization algorithms, demonstrating the funda-

Straw man #1: Exact Disclosure

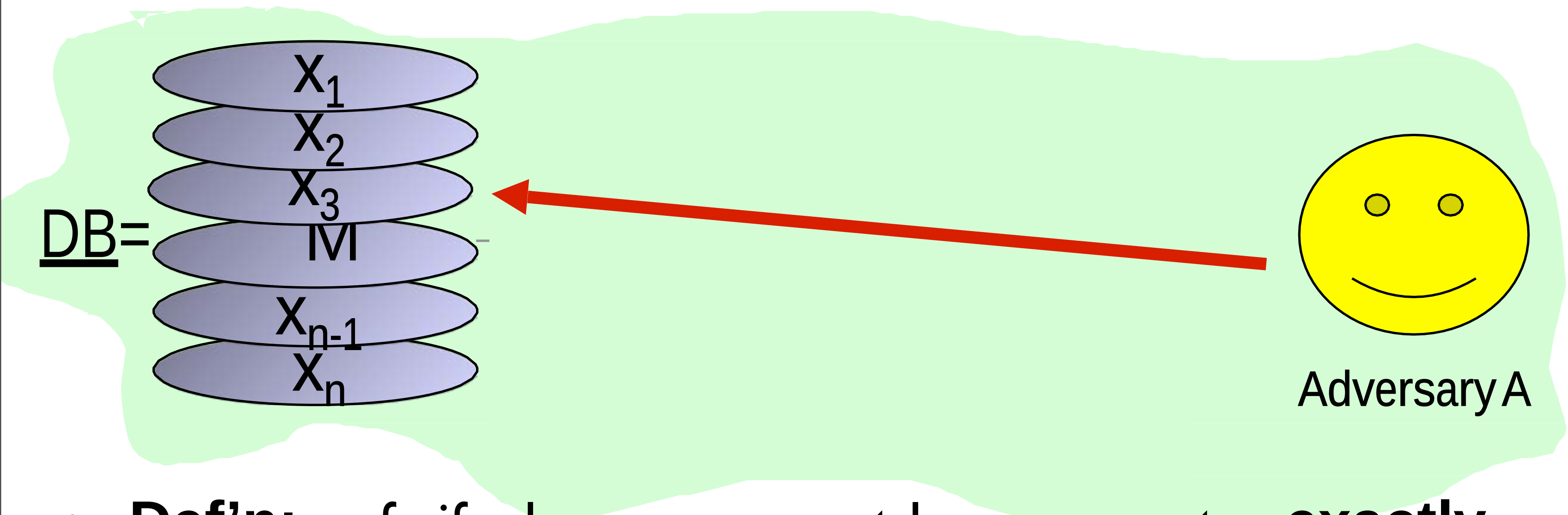


Straw man #1: Exact Disclosure



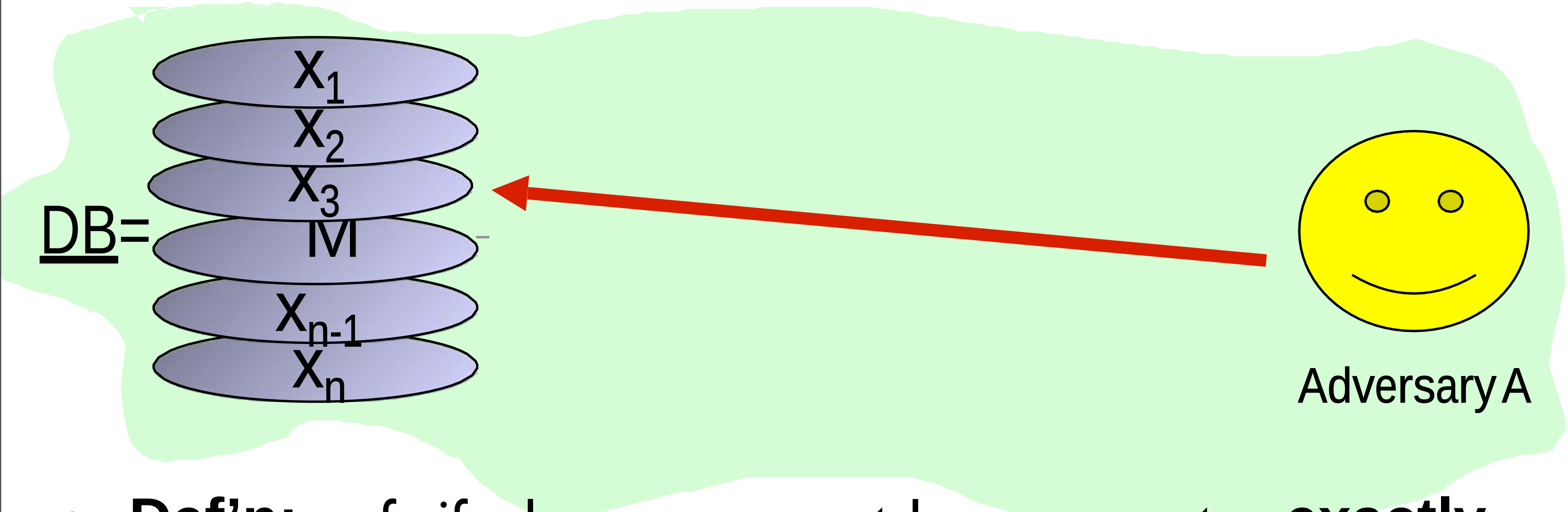
- **Def'n:** safe if adversary cannot learn any entry **exactly**

Straw man #1: Exact Disclosure



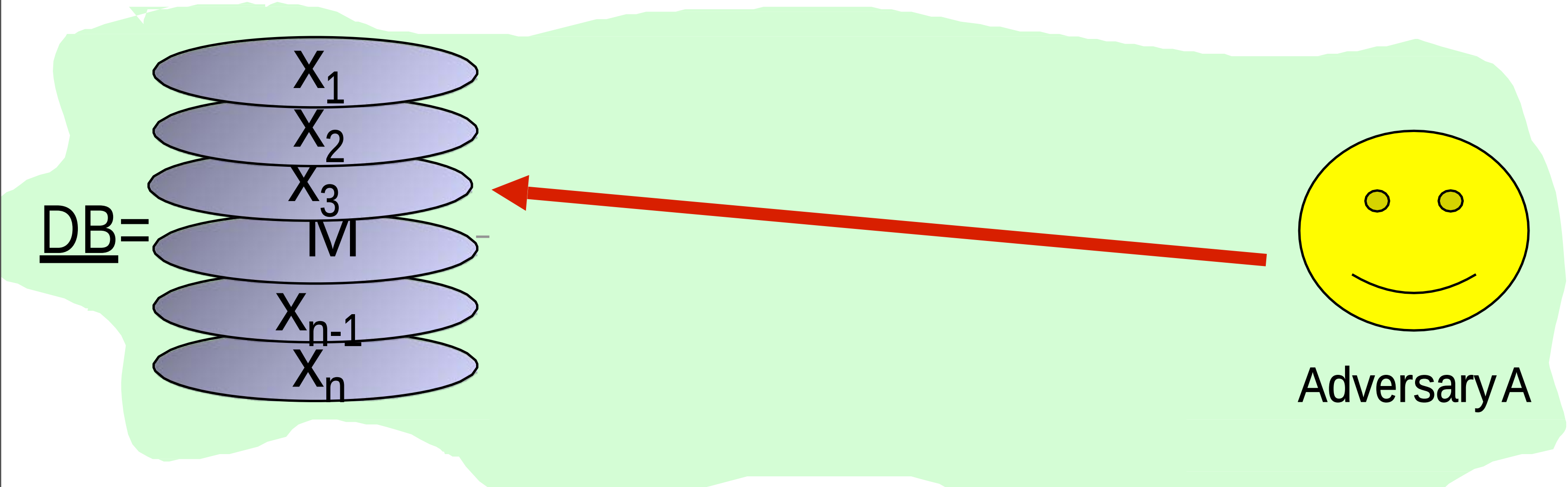
- **Def'n:** safe if adversary cannot learn any entry **exactly**

Straw man #1: Exact Disclosure



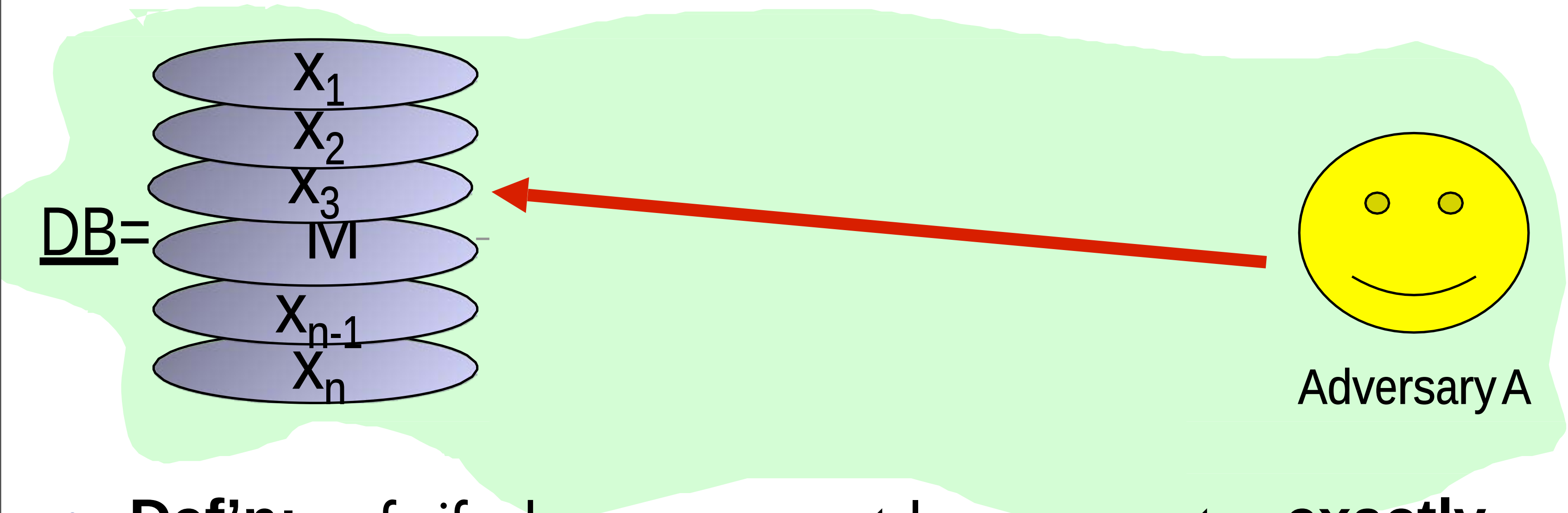
- **Def'n:** safe if adversary cannot learn any entry **exactly**
  - ➢ leads to nice (but hard) combinatorial problems

Straw man #1: Exact Disclosure



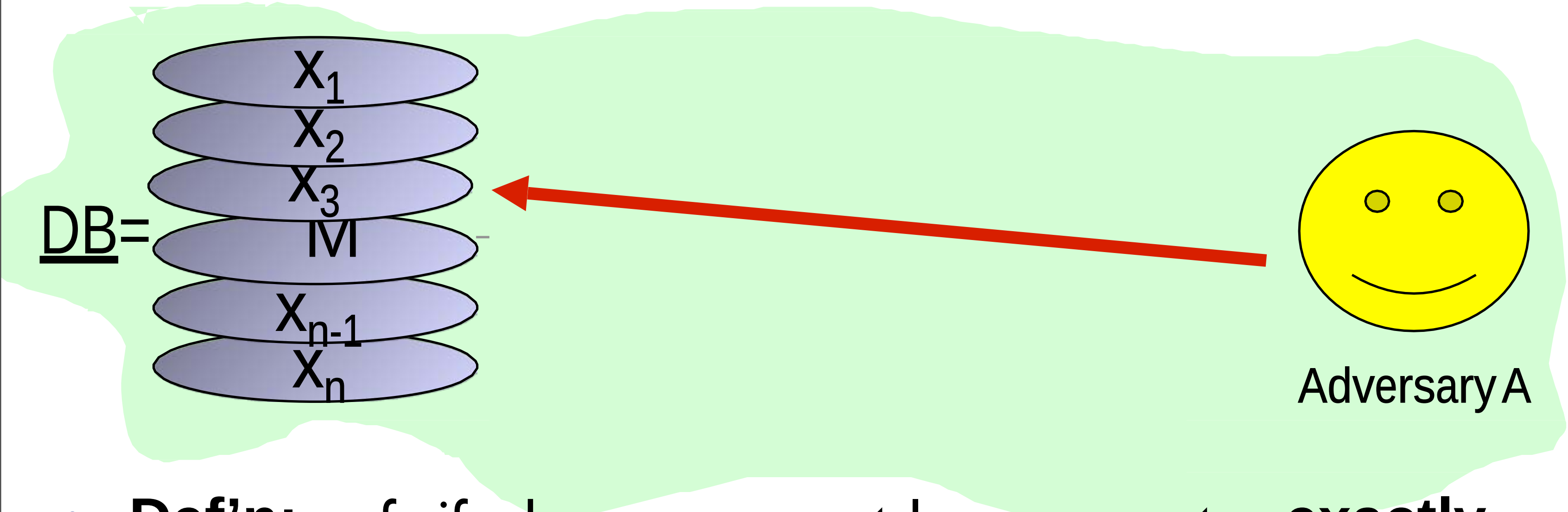
- **Def'n:** safe if adversary cannot learn any entry **exactly**
  - ➢ leads to nice (but hard) combinatorial problems
  - ➢ Does not preclude learning value with 99% certainty or narrowing down to a small interval

Straw man #1: Exact Disclosure



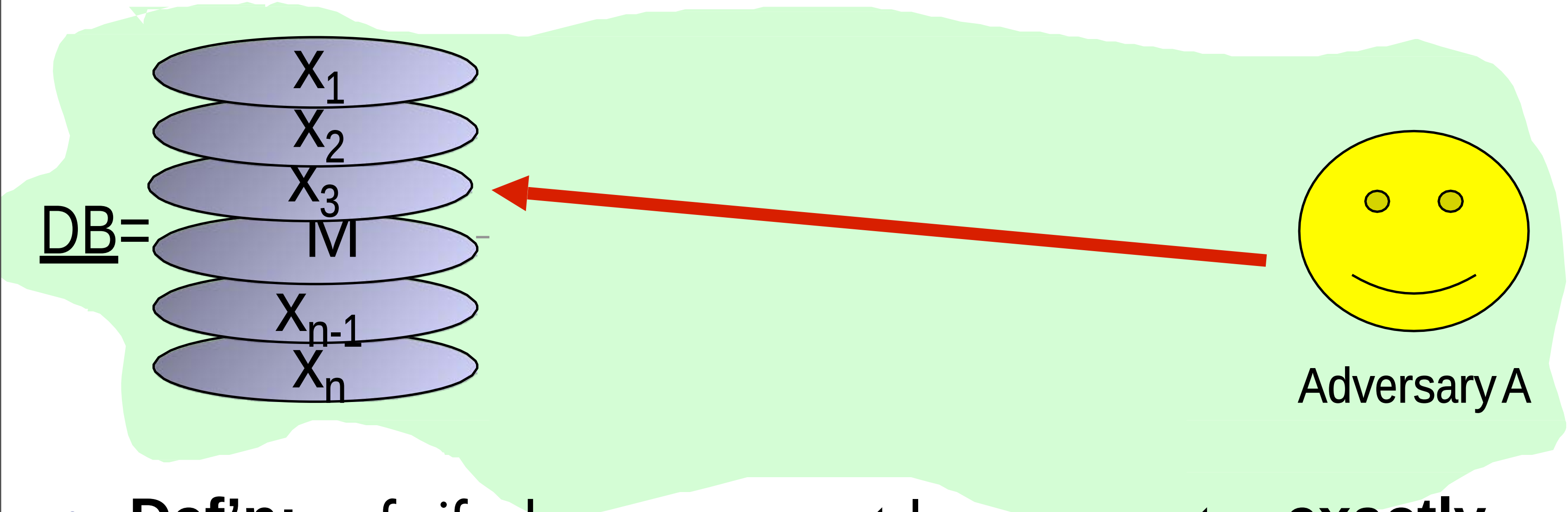
- **Def'n:** safe if adversary cannot learn any entry **exactly**
  - ➢ leads to nice (but hard) combinatorial problems
  - ➢ Does not preclude learning value with 99% certainty or narrowing down to a small interval
- Historically:

Straw man #1: Exact Disclosure



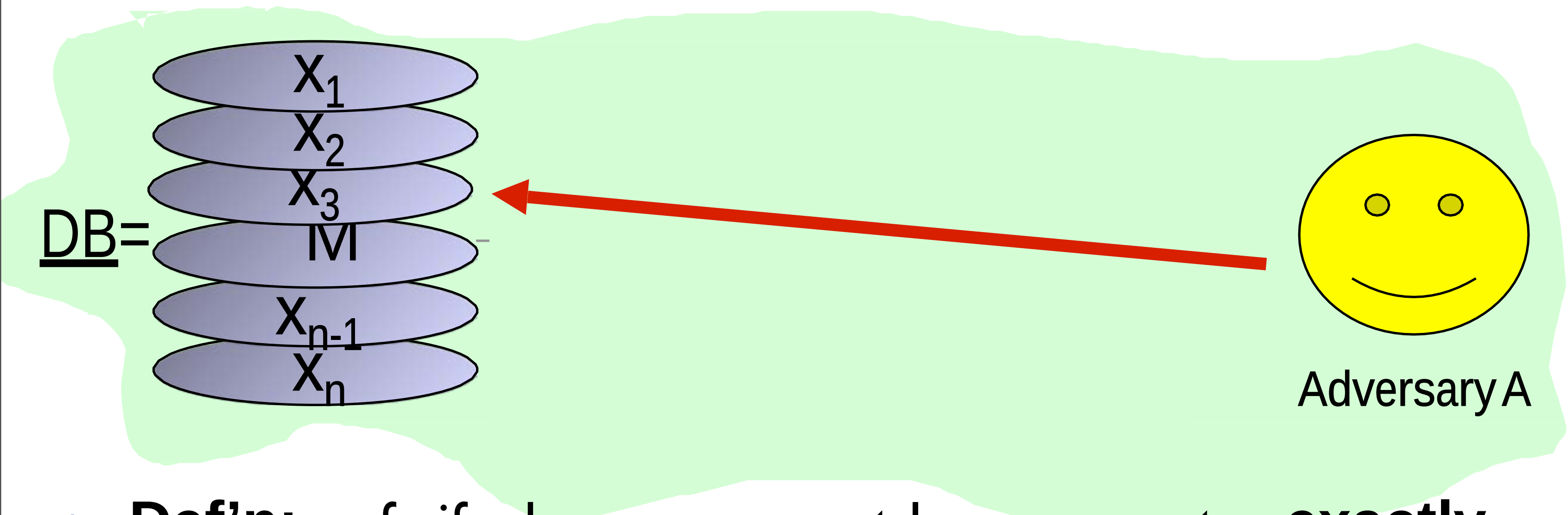
- **Def'n:** safe if adversary cannot learn any entry **exactly**
  - leads to nice (but hard) combinatorial problems
  - Does not preclude learning value with 99% certainty or narrowing down to a small interval
- Historically:
  - Focus: auditing interactive queries

Straw man #1: Exact Disclosure



- **Def'n:** safe if adversary cannot learn any entry **exactly**
  - ➢ leads to nice (but hard) combinatorial problems
  - ➢ Does not preclude learning value with 99% certainty or narrowing down to a small interval
- Historically:
  - ➢ Focus: auditing interactive queries
  - ➢ Difficulty: understanding relationships between queries

Straw man #1: Exact Disclosure



- **Def'n:** safe if adversary cannot learn any entry **exactly**
  - ➢ leads to nice (but hard) combinatorial problems
  - ➢ Does not preclude learning value with 99% certainty or narrowing down to a small interval
- Historically:
  - ➢ Focus: auditing interactive queries
  - ➢ Difficulty: understanding relationships between queries
  - ➢ E.g. two queries with small difference

Two Intuitions for Data Privacy

- "If the release of statistics *S* makes it possible to determine the value [of private information] more accurately than is possible without access to *S*, a disclosure has taken place." [Dalenius]
  - ➢ Learning more about me should be hard

- Privacy is "protection from being brought to the attention of others." [Gavison]
  - ➢ Safety is blending into a crowd

A Problem Example?

Suppose adversary knows that I smoke.

Question 0: How many patients smoke?

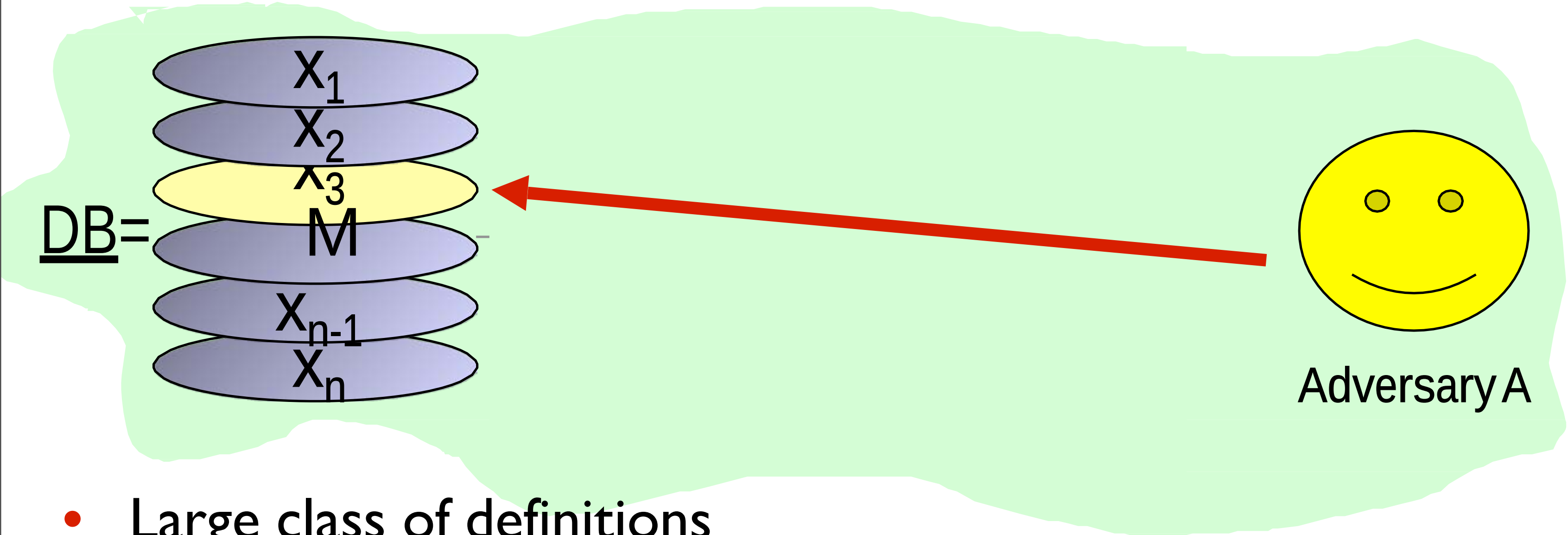
Question1: How many smokers have cancer?

Question 2: How many patients have cancer?

If adversary learns that smoking → cancer then he learns my health status.

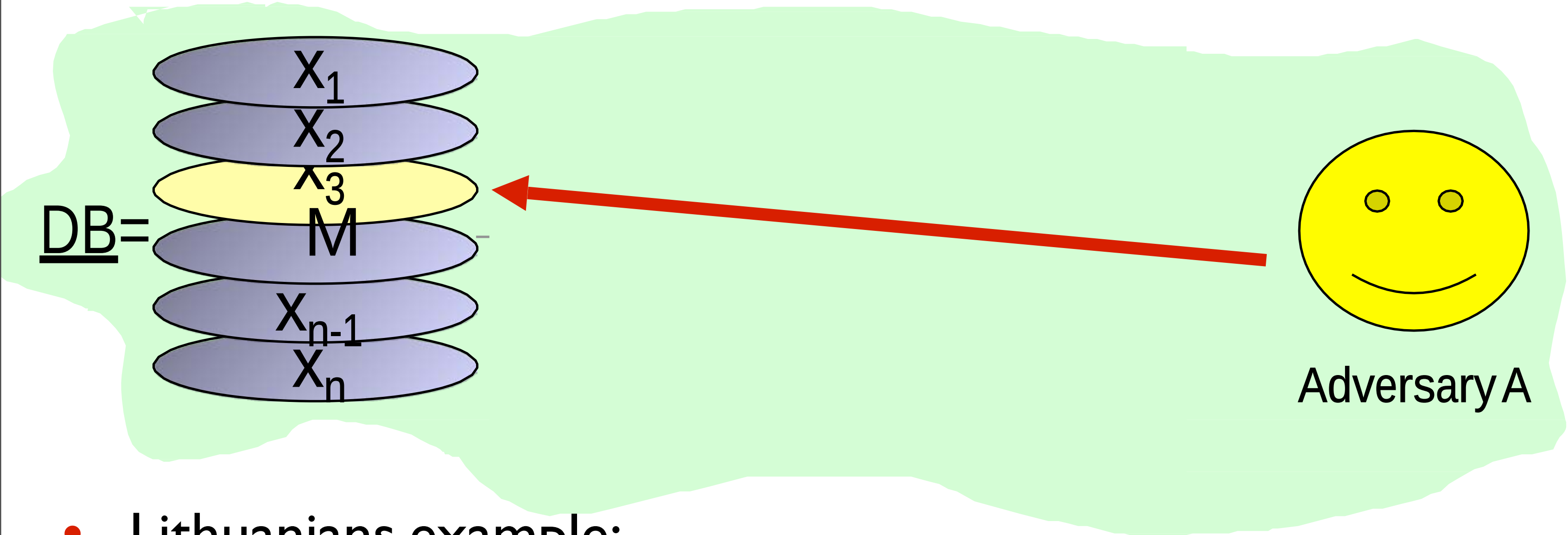
Privacy Violation?

Preventing Attribute Disclosure



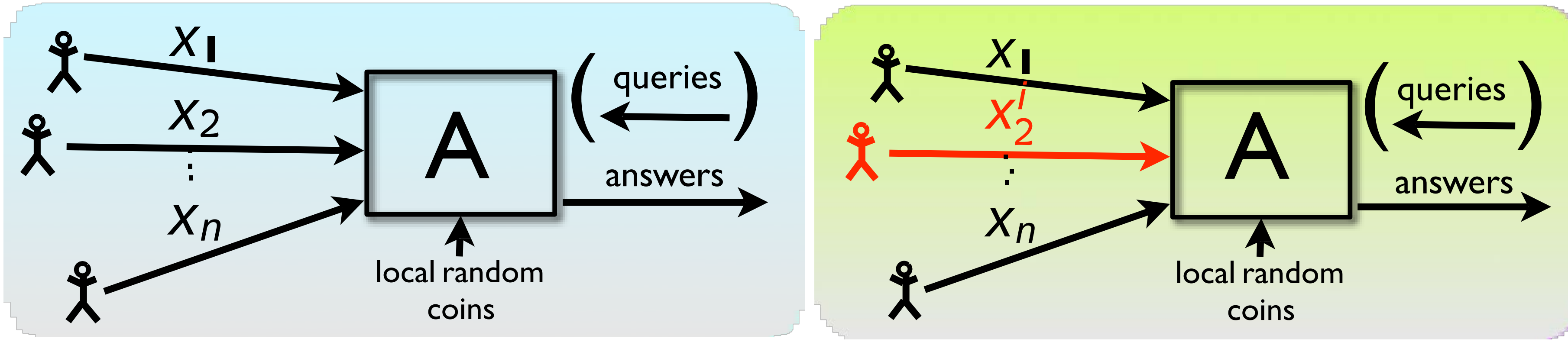
- Large class of definitions
  - safe if adversary can't learn "too much" about any entry
  - E.g.:
    - Cannot narrow $\mathbf{X_i}$ down to small interval
    - For uniform $\mathbf{X_i}$, mutual information $I(X_i;San(DB))$ · ε
- How can we decide among these definitions?

Differential Privacy



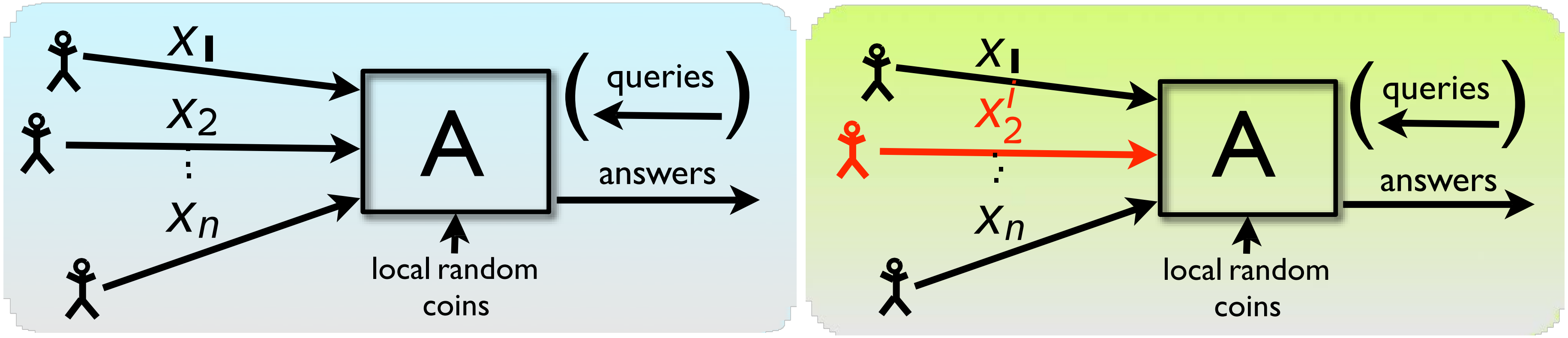
- Lithuanians example:
  - ➢ Adv. learns height even if Alice not in DB
- Intuition [DM]:
  - ➢ "Whatever is learned would be learned regardless of whether or not Alice participates"
  - ➢ Dual: Whatever is already known, situation won't get worse

Approach: Indistinguishability



x' is a neighbor of x
if they differ in one row

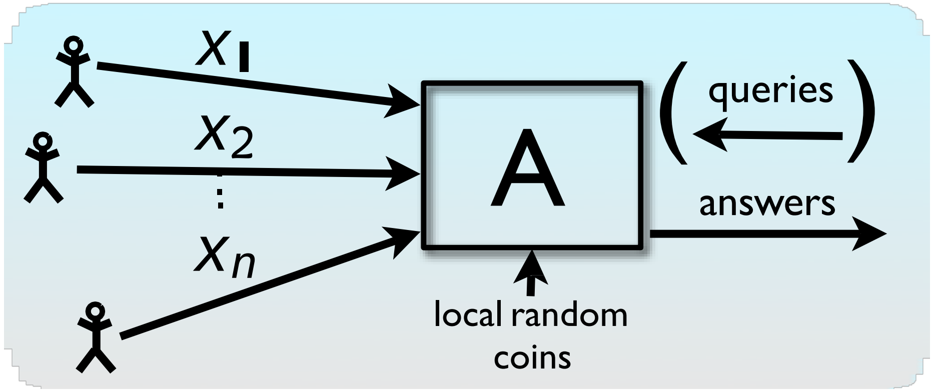
Approach: Indistinguishability

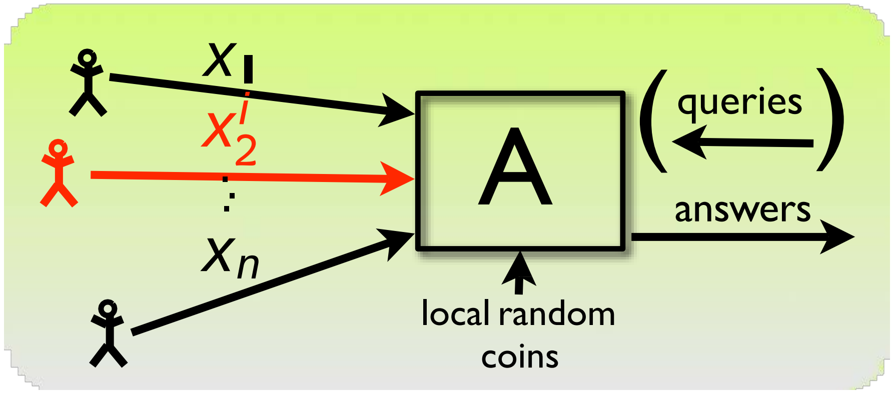


x' is a neighbor of x
if they differ in one row

Neighboring databases induce **close** distributions on transcripts

Approach: Indistinguishability





x' is a neighbor of x
if they differ in one row

Neighboring databases induce **close** distributions on transcripts

Definition: A is indistinguishable if,
for all neighbors x,x',
for all subsets S of transcripts

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

Approach: Indistinguishability

- Note that ε has to be non-negligible here
  - Triangle inequality: **any** pair of databases at distance $< \varepsilon n$
  - If $\varepsilon < 1/n$ then users get no info!
- Why this measure?
  - Statistical difference doesn't make sense with $\varepsilon > 1/n$
  - E.g. choose random i and release i, x_i
  - This compromises someone's privacy w.p. 1

Definition: A is indistinguishable if,
for all neighbors x,x',
for all subsets S of transcripts

$$\Pr[A(x) \in S] \leq (1 + e)\Pr[A(x') \in S]$$

Neighboring databases induce **close** distributions on transcripts

Differential Privacy

- Another interpretation [DM]:

  You learn the same things about me
  regardless of whether I am in the database

- Suppose you know I am the height of median Canadian
  - You could learn my height from database!
    But it didn't matter whether or not my data was part of it.
  - Has my privacy been compromised? No!

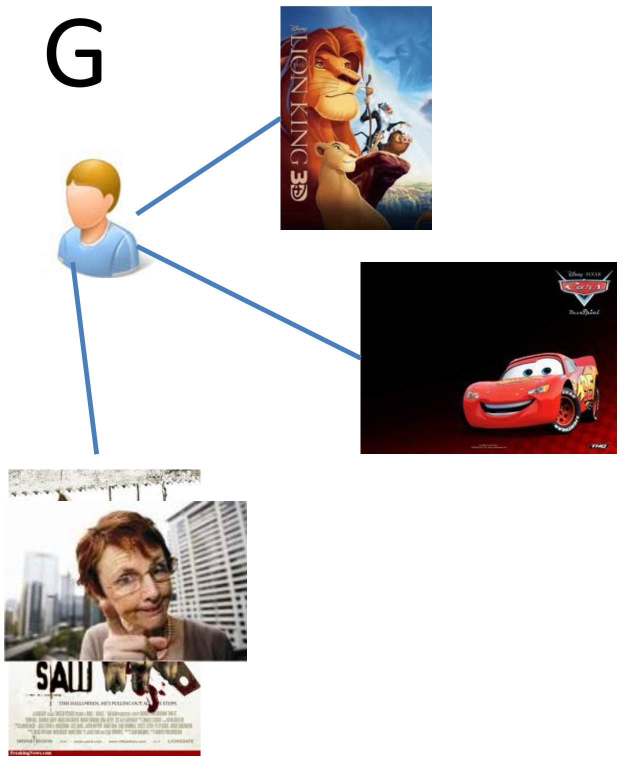
Neighboring databases induce **close** distributions on transcripts

Definition: A is indistinguishable if,
for all neighbors x,x',
for all subsets S of transcripts

$$\Pr[A(x) \in S] \leq (1+e)\Pr[A(x') \in S]$$

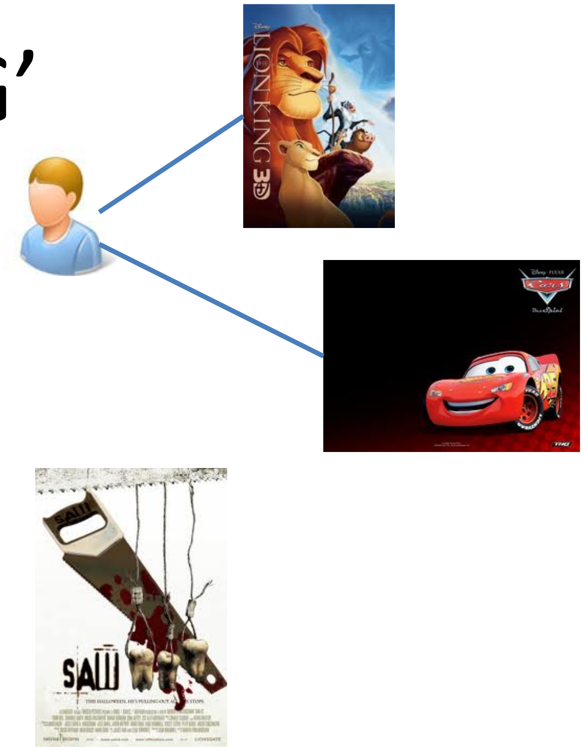
Graphs: Edge Adjacency

G



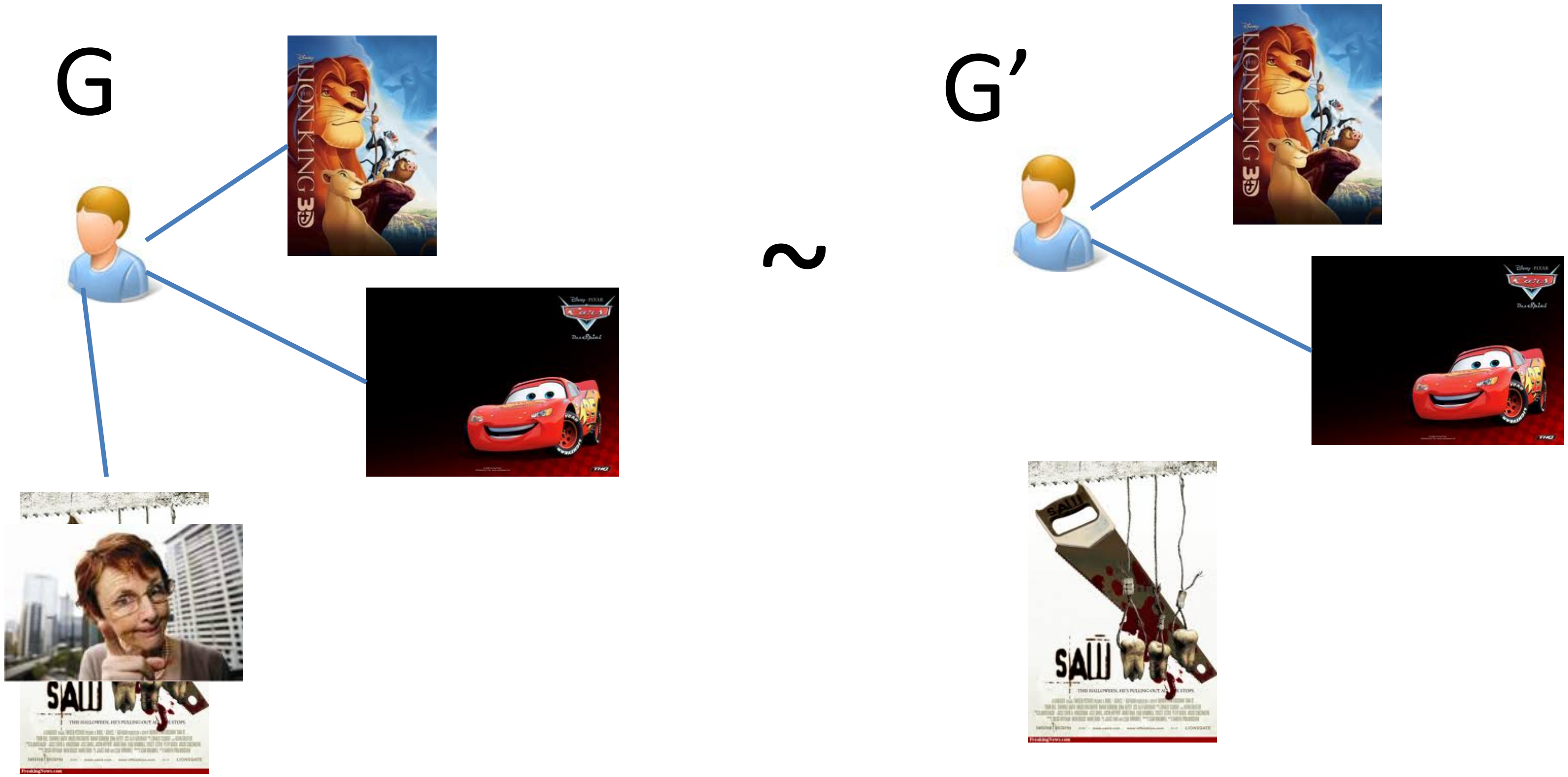
~

G’



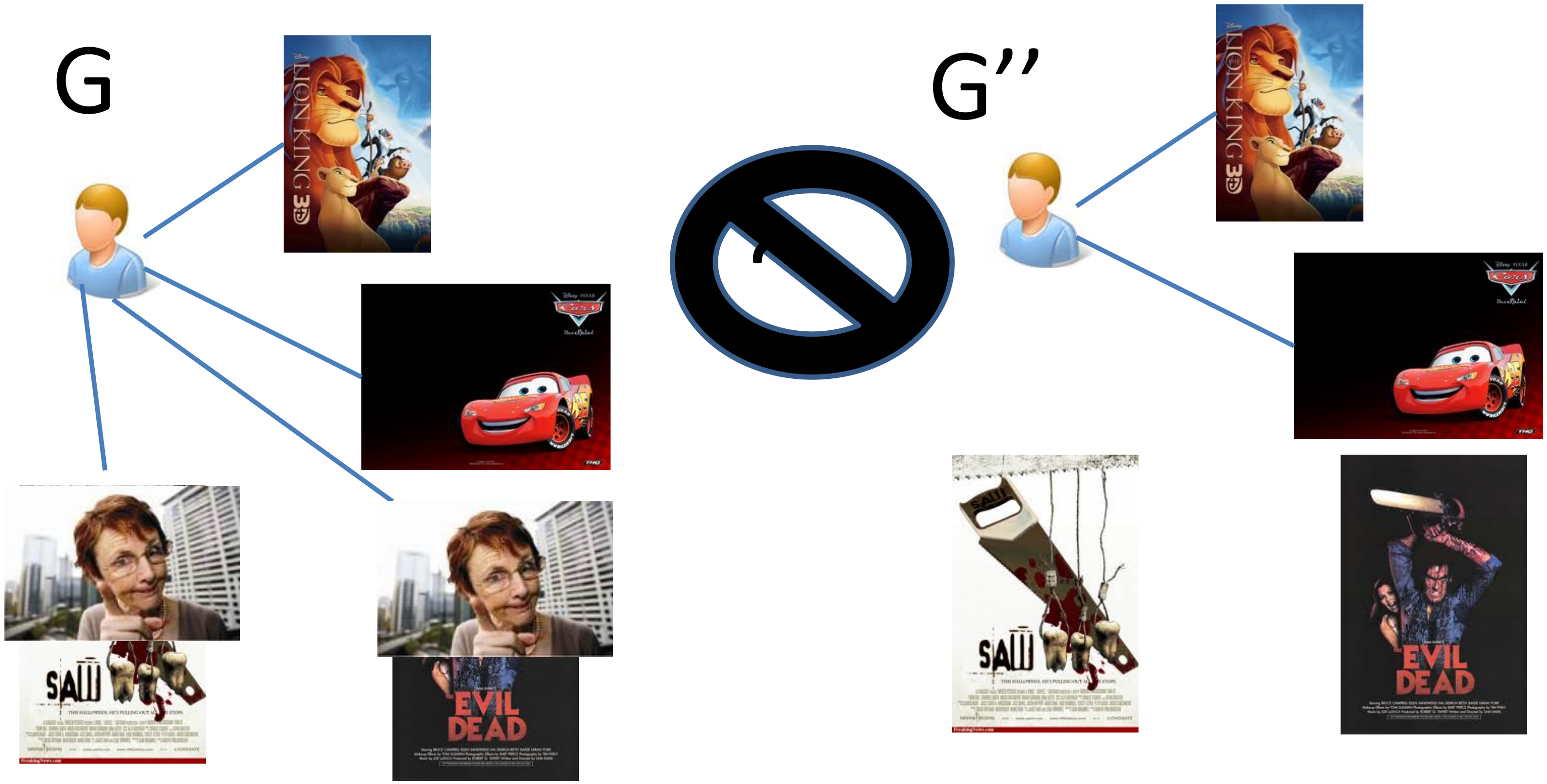
Pr[A(G)∈ ] $\leq e^{\varepsilon}$Pr[A(G’)∈ ] + δ

Graphs: Edge Adjacency



Johnny’s mom does not learn if he watched Saw from the output A(G).

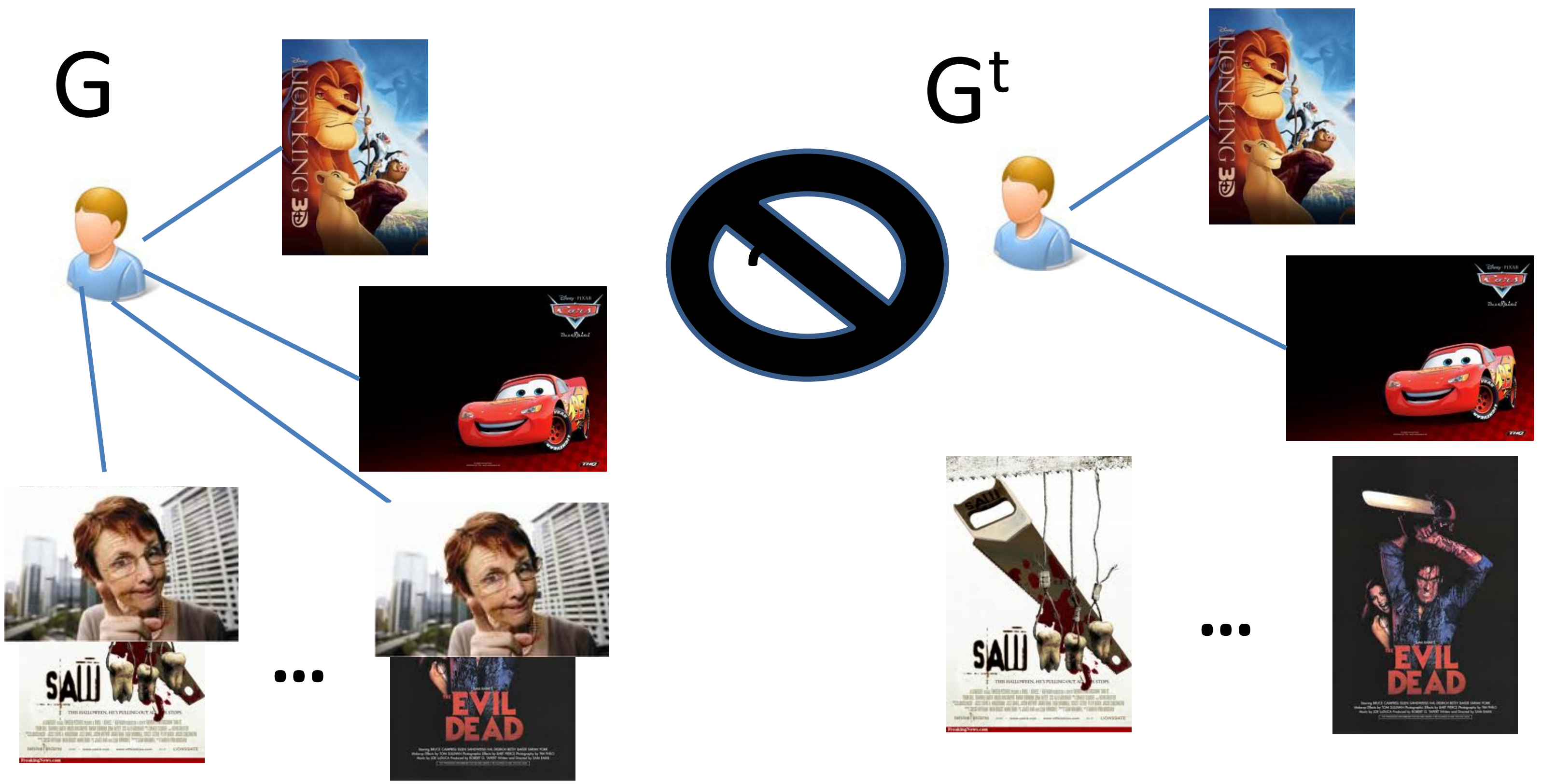
Privacy for Two Edges?



Pr[A(G)∈ ] $\leq e^{2\varepsilon}$ Pr[A(G’’)∈ ] + 2δ

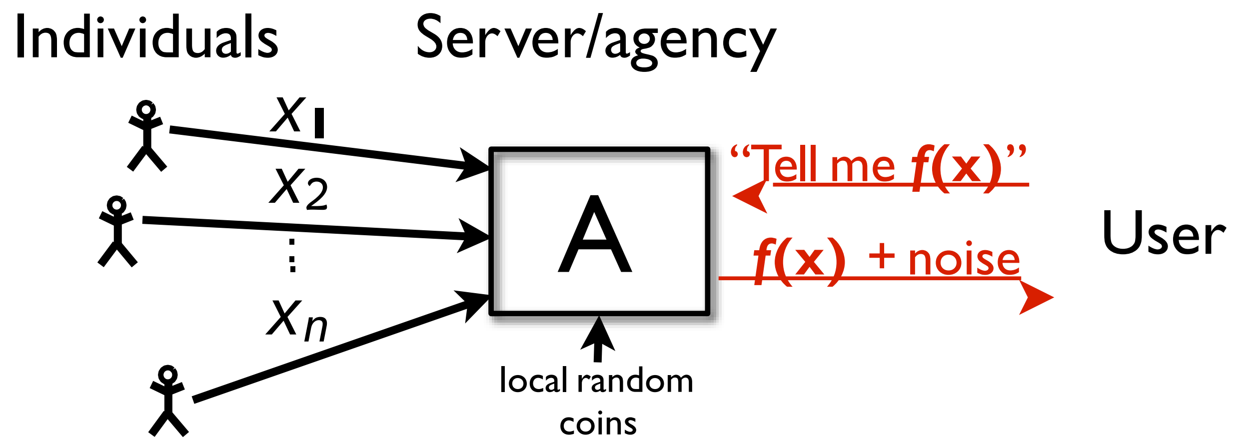
71

Limitations



Johnny's mom may now be able tell if he watches R-rated movies from A(G).

Output Perturbation



- Intuition: $\boldsymbol{f}(\mathbf{x})$ can be released accurately when $\boldsymbol{f}$ is insensitive to individual entries $x_1, x_2, \ldots, x_n$

Global Sensitivity

$$\Delta Q := \max_{G \sim G'} |Q(G) - Q(G')|$$

Global Sensitivity

$$\Delta Q := \max_{G \sim G'} |Q(G) - Q(G')|$$

- What does G~G' mean?
- Example: Change one attribute

- $Q_1(G)$ = #users who watched Lion King

- ΔQ_1 = ?

Global Sensitivity

$$\Delta Q \coloneqq \max_{G \sim G'} |Q(G) - Q(G')|$$

- What does G~G' mean?
- Example: Change one attribute

- $Q_2(G)$ = #users who watched Toy Story

- $\Delta Q_2 = 1$

Global Sensitivity

$$\Delta Q := \max_{G \sim G'} |Q(G) - Q(G')|$$

- What does G~G' mean?
- Example: Change one attribute

- $Q(G) = Q_1(G) + Q_2(G)$

- ΔQ_2 = ?

Global Sensitivity

$$\Delta Q := \max_{G \sim G'} |Q(G) - Q(G')|$$

- What does G~G' mean?
- Example: Change one attribute

- $Q_1(G)$ = #users who watched Lion King

- ΔQ_1 = ?

Global Sensitivity

$$\Delta Q := \max_{G \sim G'} |Q(G) - Q(G')|$$

- What does G~G' mean?
- Example: Add/delete one row?

Global Sensitivity

$$\Delta Q := \max_{G \sim G'} |Q(G) - Q(G')|$$

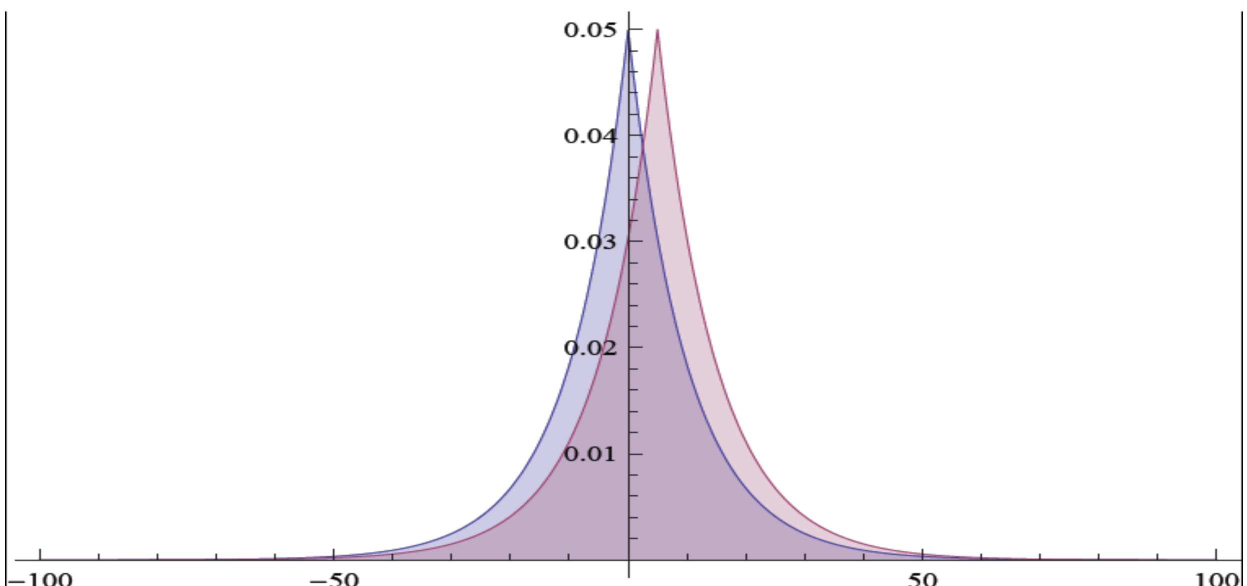
- Example: Add/delete one row?
- Q(G) = Q1(G)+Q2(G)
- $\Delta Q = ?$

Traditional Differential Privacy Mechanism

Fact: The Laplacian Mechanism:

$$\mathrm{A}(G) = Q(G) + \mathrm{Lap}\left(\frac{\Delta Q}{\varepsilon}\right),$$

satisfies $(\varepsilon, 0)$-differential privacy.

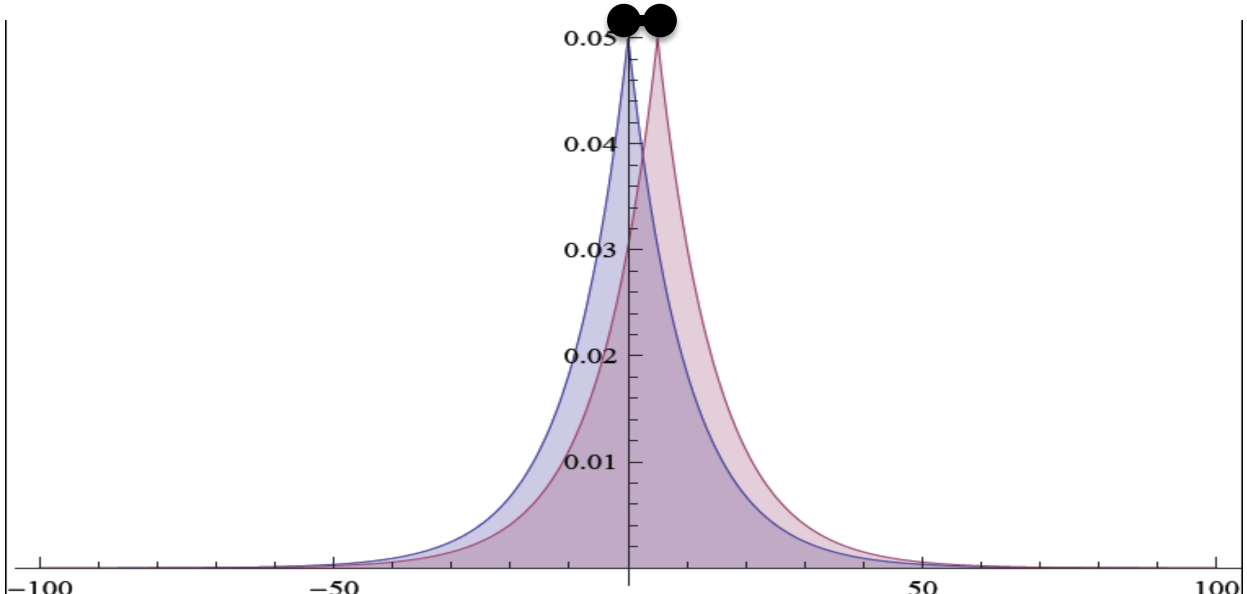


Traditional Differential Privacy Mechanism

$$PDF_{G}(x\ ) \propto e^{-|x\varepsilon|}$$

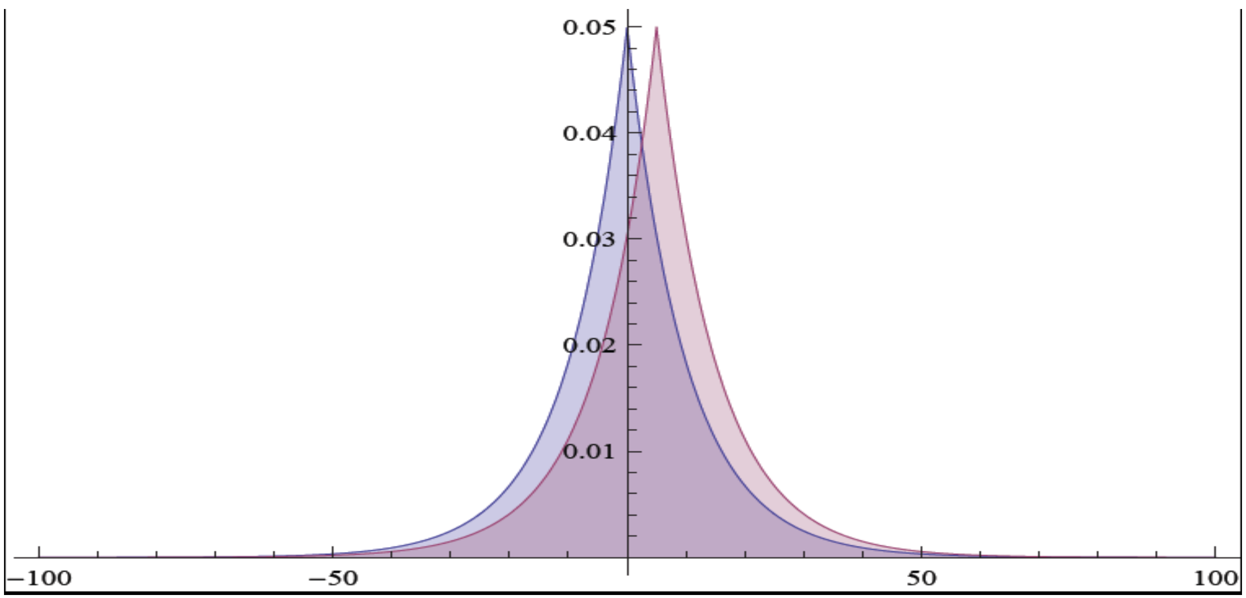
$$PDF_{G'}(x\ ) \propto e^{-|(x-1)\varepsilon|}$$

Δ*Q*=1



Traditional Differential Privacy Mechanism

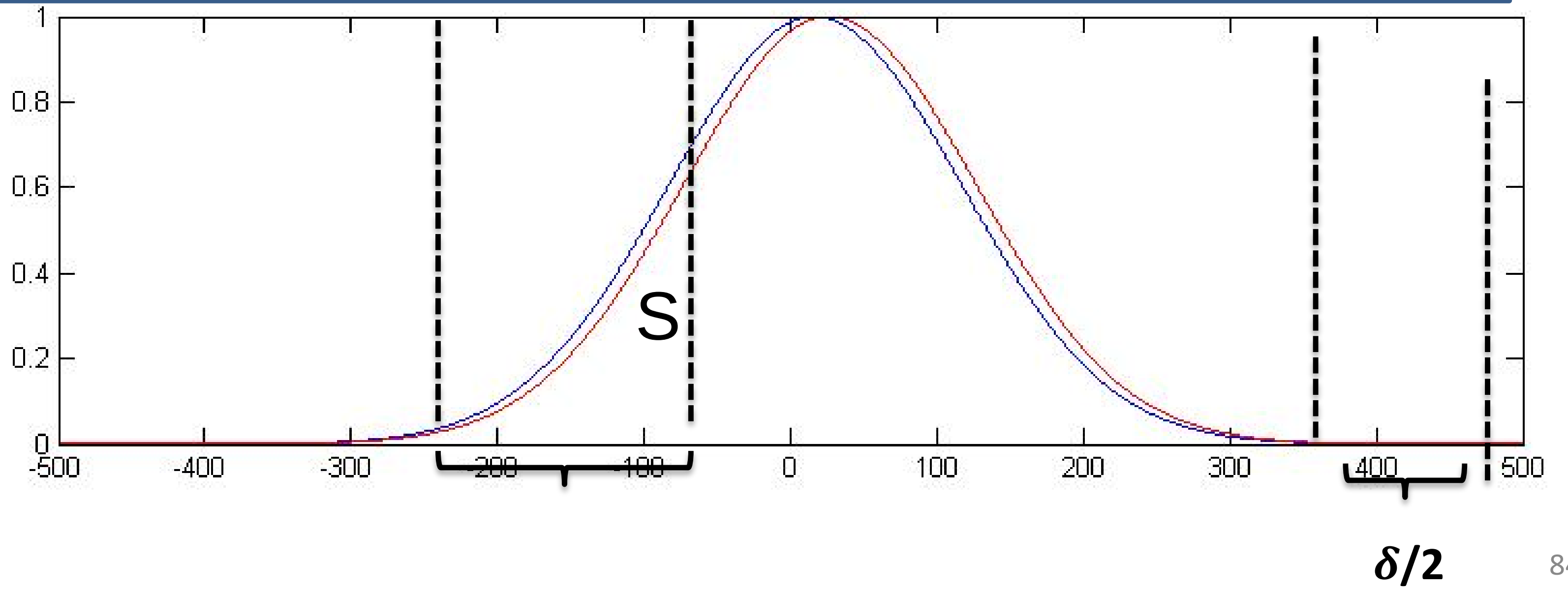
$$\forall x, \frac{PDF_G(x\ )}{PDF_{G'}(x\ )} = \frac{e^{-|x\varepsilon|}}{e^{-|(x-1)\varepsilon|}} \leq e^{-\varepsilon}$$



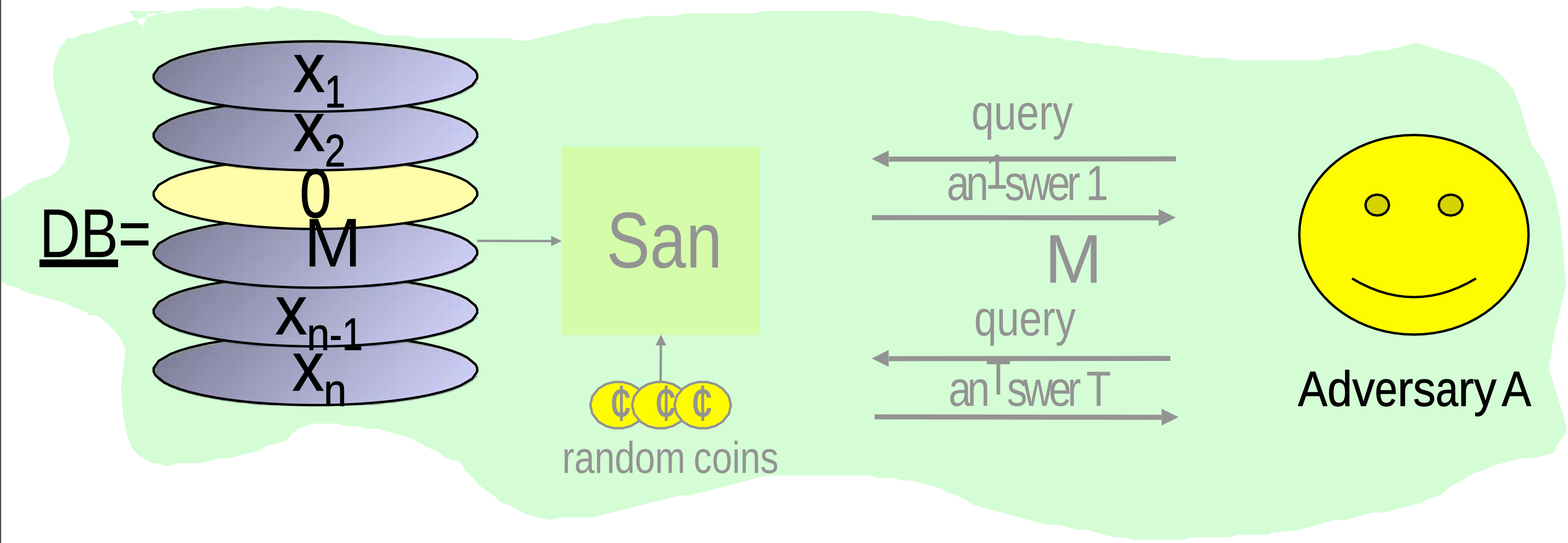
Traditional Mechanism #2

Fact: The Gaussian mechanism preserves (ε, δ)-differential privacy

$$\mathrm{A}(G) = Q(G) + \mathrm{N}\left(0, \frac{2\,(\Delta Q)^2 \log\left({}^{1.25}/_{\delta}\right)}{\varepsilon^2}\right).$$



Differential Privacy



Examples of low global sensitivity

- Example: $GS_{average} = \frac{1}{n}$ if $x \in [0,1]^n$
  - Add noise Lap($\frac{1}{n}$)
  - Comparison: to estimate a frequency (e.g.proportion of diabetics) in underlying population, get sampling noise $\frac{1}{\sqrt{n}}$
- Many natural functions have low GS, e.g.:
  - Histograms and contingency tables
  - Covariance matrix
  - Distance to a property
  - Functions that can be approximated from a random sample
- [BDMN] Many data-mining and learning algorithms access the data via a sequence of low-sensitivity questions
  - e.g. perceptron, some "EM" algorithms, SQ learning algorithms

Why does this help?

With relatively little noise:

- Averages
- Contingency tables
- Matrix decompositions
- Certain types of clustering
- …

Differential Privacy

Differential Privacy

Protocols

Differential Privacy

Protocols

- Output perturbation

  (Release f(x) + noise)

  - Sum queries
    - [DiN'03,DwN'04,BDMN'05]
  - "Sensitivity" frameworks
    - [DMNS'06,NRS'07]

Differential Privacy

Protocols

- Output perturbation
  (Release f(x) + noise)
  - Sum queries
    - [DiN'03,DwN'04,BDMN'05]
  - "Sensitivity" frameworks
    - [DMNS'06,NRS'07]
- Input perturbation
  ("randomized response")
  - Frequent item sets [EGS'03]
  - (Various learning results)

Differential Privacy

Protocols

- Output perturbation

  (Release f(x) + noise)

  - Sum queries
    - [DiN'03,DwN'04,BDMN'05]
  - "Sensitivity" frameworks
    - [DMNS'06, NRS'07]

- Input perturbation

  ("randomized response")

  - Frequent item sets [EGS'03]
  - (Various learning results)

Lower bounds

Differential Privacy

Protocols

- Output perturbation
  (Release f(x) + noise)
  - Sum queries
    - [DiN'03,DwN'04,BDMN'05]
  - "Sensitivity" frameworks
    - [DMNS'06, NRS'07]
- Input perturbation
  ("randomized response")
  - Frequent item sets [EGS'03]
  - (Various learning results)

Lower bounds

- Limits on communication models
  - Noninteractive [DMNS'06]
  - "Local" [NSW'07]

Differential Privacy

Protocols

- Output perturbation (Release f(x) + noise)
  - ➢ Sum queries
    - [DiN'03,DwN'04,BDMN'05]
  - ➢ "Sensitivity" frameworks
    - [DMNS'06, NRS'07]
- Input perturbation ("randomized response")
  - ➢ Frequent item sets [EGS'03]
  - ➢ (Various learning results)

Lower bounds

- Limits on communication models
  - ➢ Noninteractive [DMNS'06]
  - ➢ "Local" [NSW'07]
- Limits on accuracy
  - ➢ "Many" good answers allow reconstructing database
    - [DiNi'03,DMT'07]

Differential Privacy

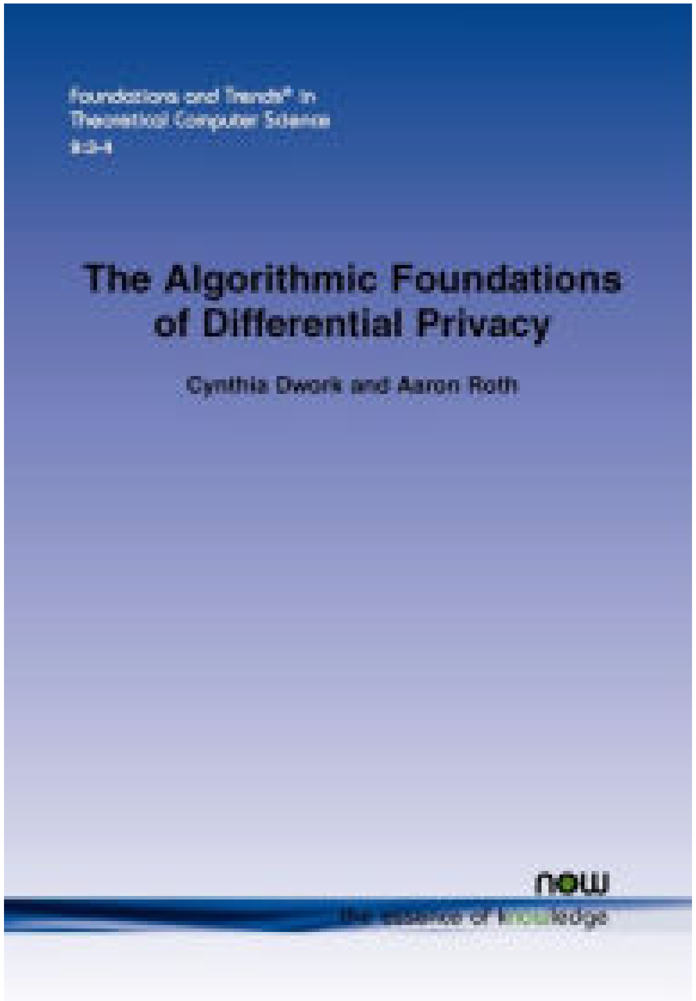
Protocols

- Output perturbation (Release f(x) + noise)
  - Sum queries
    - [DiN'03,DwN'04,BDMN'05]
  - "Sensitivity" frameworks
    - [DMNS'06, NRS'07]
- Input perturbation ("randomized response")
  - Frequent item sets [EGS'03]
  - (Various learning results)

Lower bounds

- Limits on communication models
  - Noninteractive [DMNS'06]
  - "Local" [NSW'07]
- Limits on accuracy
  - "Many" good answers allow reconstructing database
    - [DiNi'03,DMT'07]
- Necessity of "differential" guarantees [DN]

Resources



BARNES & NOBLE

$99

Free PDF:
https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf