CS 350S: Privacy-Preserving Systems

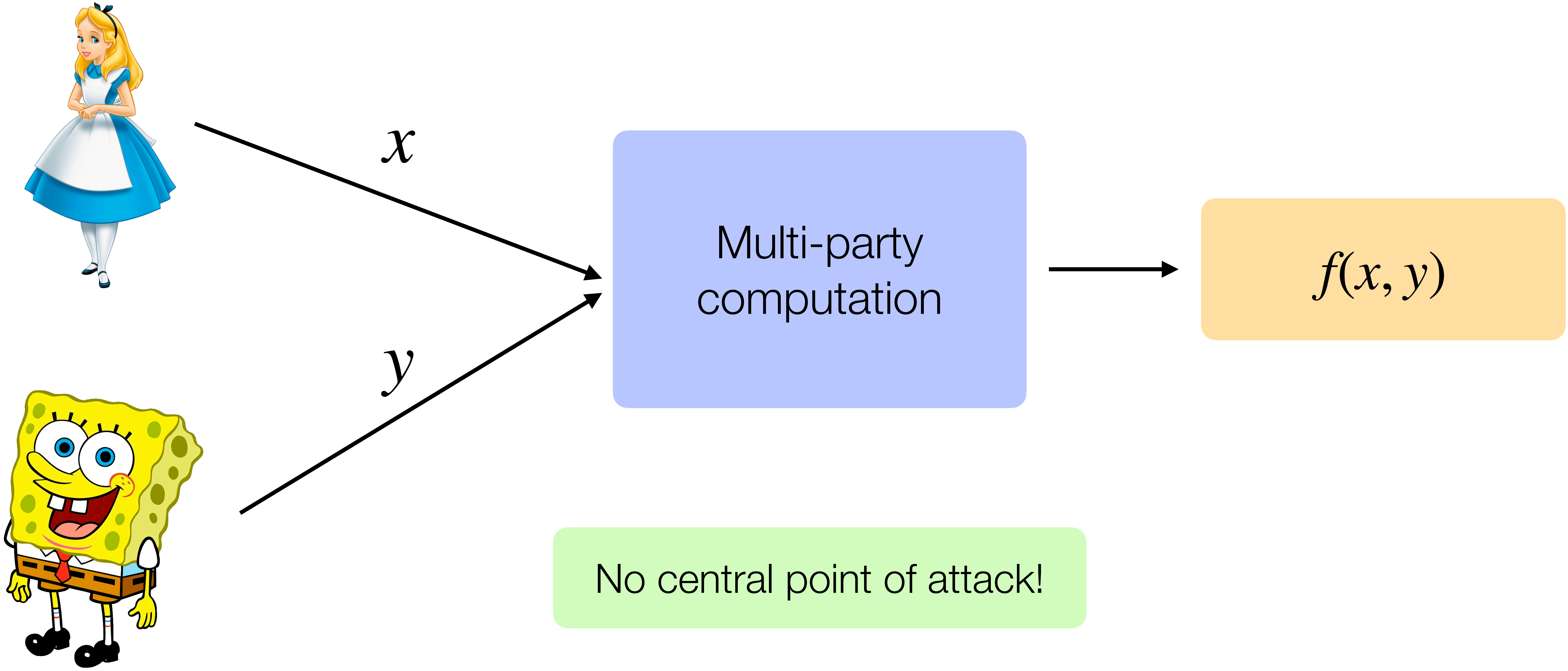
Differential privacy

Outline

1. Differential privacy definition
2. Differential privacy mechanism
3. Differential privacy applications
4. Logistics
5. Student presentation

MPC recap

(Informal) Any computation that can be performed with a trusted third party can be securely computed *without* one!



x

y

Multi-party computation

$f(x, y)$

No central point of attack!

Use case: private aggregate statistics

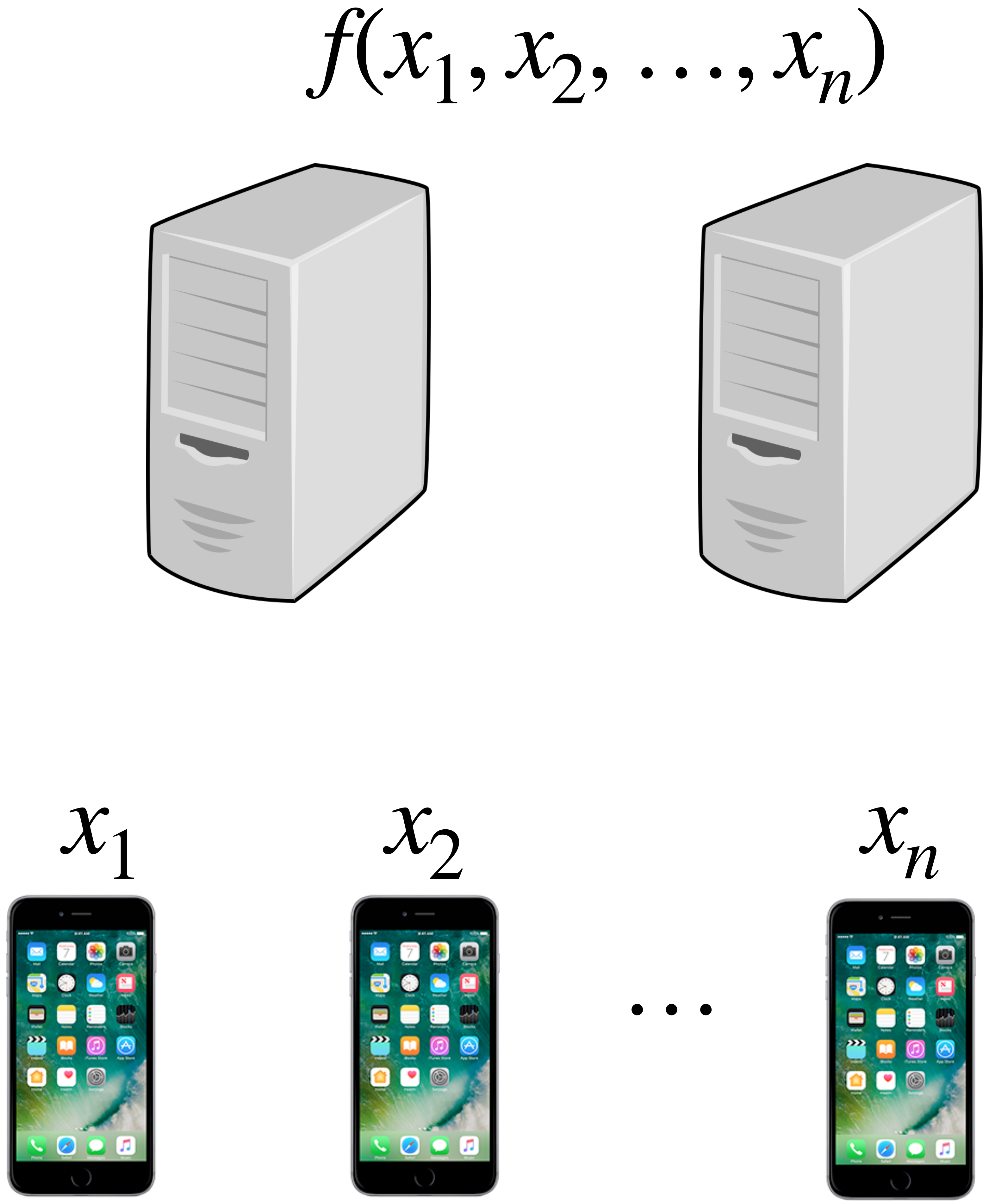
Aggregation function f

Correctness: If all servers are honest, servers learn $f(x_1, x_2, \ldots, x_n)$

Privacy: If one server is honest, servers only learn $f(x_1, x_2, \ldots, x_n)$

- Privacy with 1 malicious server

Robustness: Malicious clients have bounded influence



Examples of when the output of a computation can reveal private information?

Anonymization is not enough (AOL query dataset)

- AOL query dataset had >20M anonymized search queries from 650,000 AOL users over 3 months
- Dataset released where each username was replaced with a random identifier
- Queries for
  - “Landscapers in Lilburn, Ga”
  - Several people with last name Arnold
  - “Homes sold in shadow lake subdivision Gwinnett county Georgia”
  - … other sensitive queries
- Only 14 citizens with last name Arnold in Gwinnett County
- Found that user was Thelma Arnold, 62-year old woman in Georgia

The New York Times

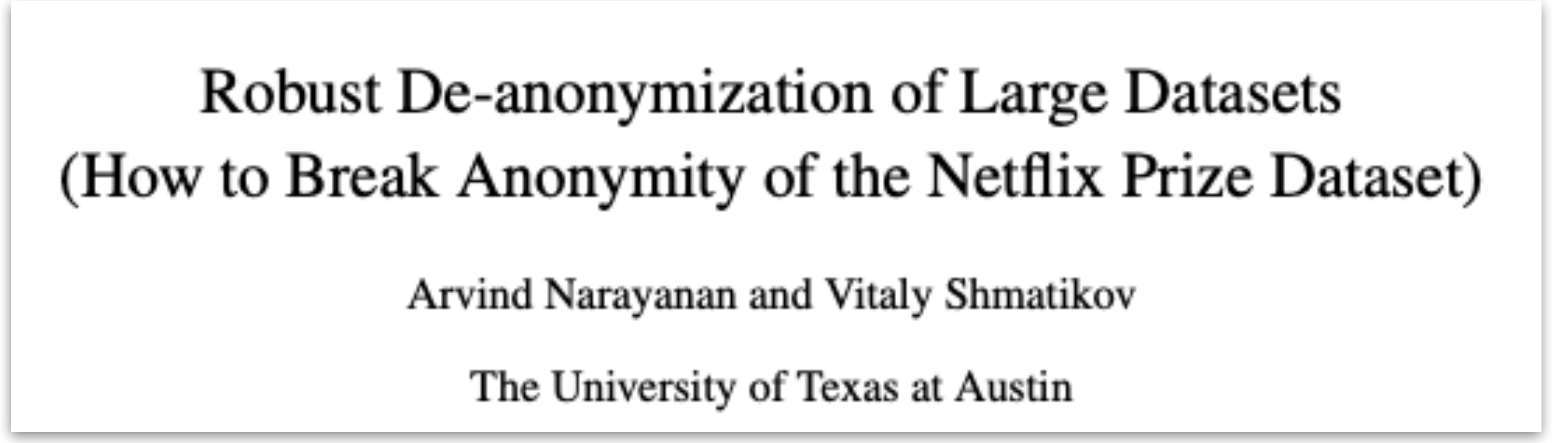
A Face Is Exposed for AOL Searcher No. 4417749

https://www.nytimes.com/2006/08/09/technology/a-face-is-exposed-for-aol-searcher-no-4417749.html

Anonymization is not enough (Netflix prize dataset)

[Narayanan, Shamtikov]

- Netflix prize dataset has anonymized movie ratings for 50K Netflix subscribers
- Adversary who has a small amount of background knowledge (e.g., dates of some ratings within a 14-day window, some approximate / potentially incorrect ratings), can uniquely identify a record in the published dataset
- Used IMDB ratings that users posted publicly under their own name
- A seemingly “harmless” dataset without identifying information can be a privacy violation

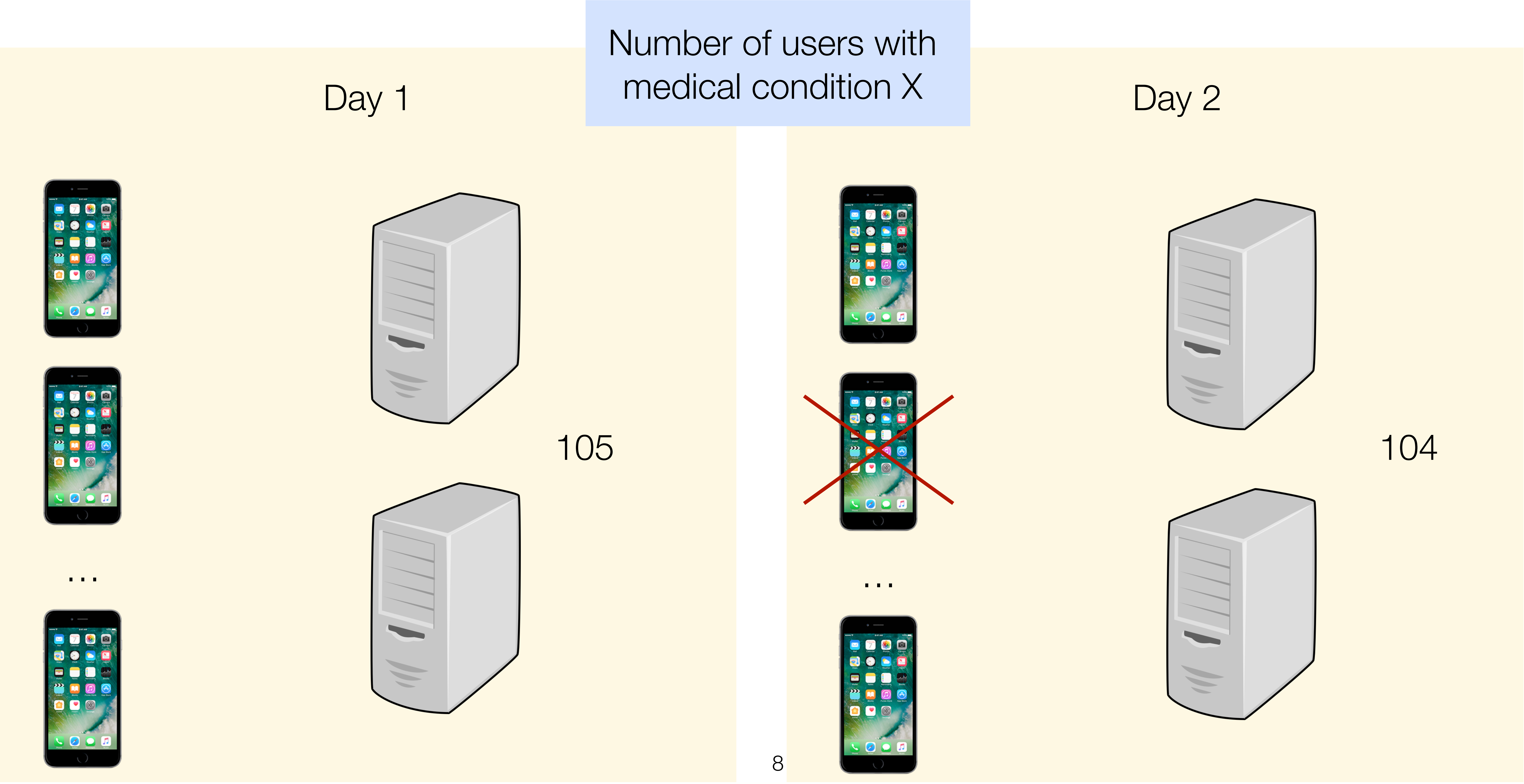


Robust De-anonymization of Large Datasets
(How to Break Anonymity of the Netflix Prize Dataset)

Arvind Narayanan and Vitaly Shmatikov

The University of Texas at Austin

Aggregating alone is not enough



Number of users with medical condition X

Day 1

105

Day 2

104

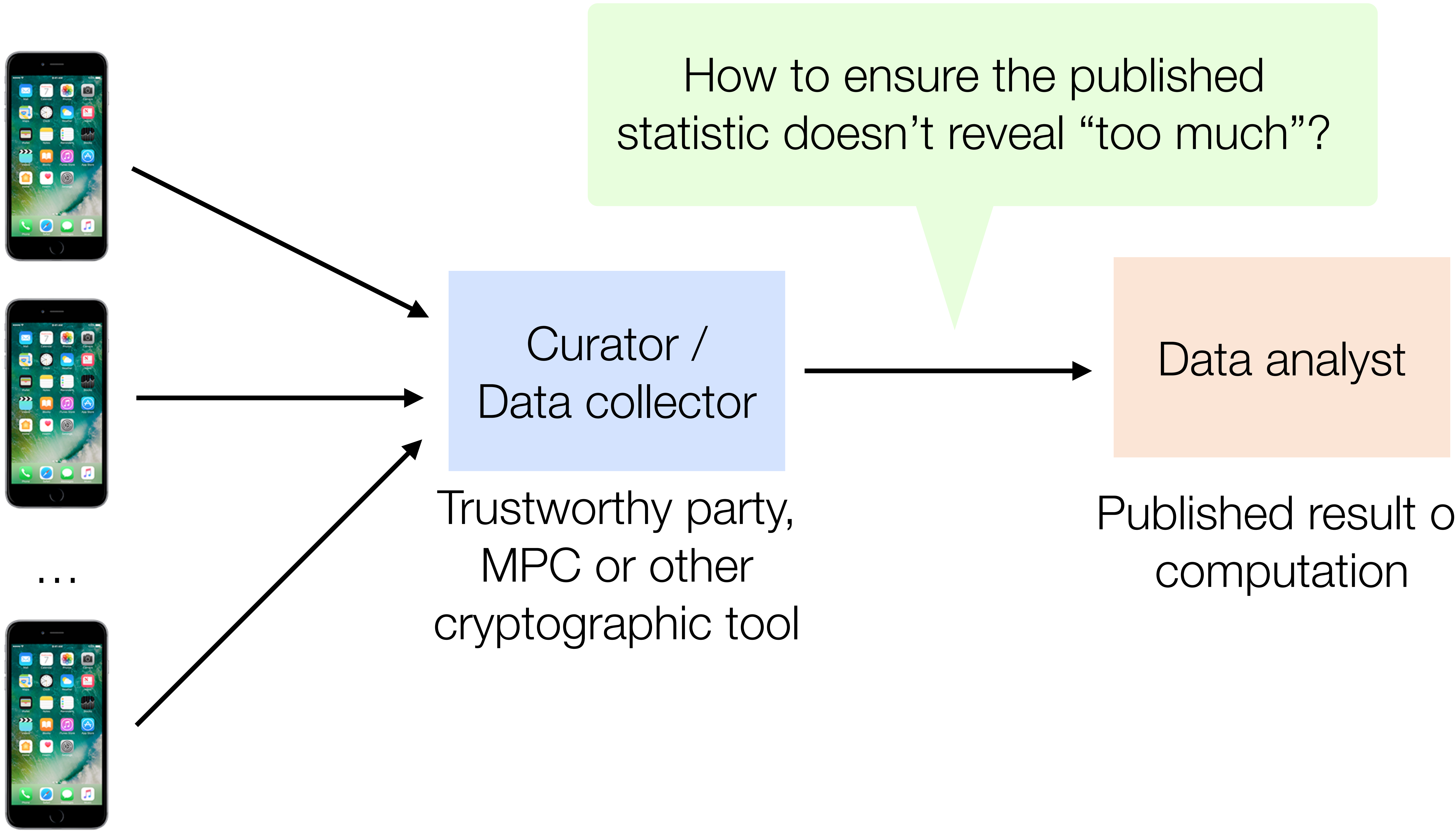
Building a private system

- Cryptographic protocols like MPC often focus on *how* to provide privacy
    - Given a function f, how to compute the output of the function while hiding the inputs and intermediate data

- Today: What function f should we be computing to ensure privacy?
    - How to define privacy?
    - How to provide privacy?

- Systems need both

Model



How to ensure the published statistic doesn't reveal "too much"?

Curator / Data collector

Trustworthy party, MPC or other cryptographic tool

Data analyst

Published result of computation

Attempt #1 at defining privacy

The result of the analysis should not allow an adversary to learn any information about individual users' data that it would not learn without access to the result.

Does this work?

No — the revealed statistic needs to be useful, and so we cannot use some sort of cryptographic definition where the output appears random and reveals nothing about the inputs

Attempt #2 at defining privacy

The result of the analysis should not allow an adversary to learn any information about any individual in the dataset that the adversary could not learn if the individual was not in the dataset.

Does this work?

No —the revealed statistic needs to be useful, and so cannot be independent of all participants

Attempt #2 at defining privacy

The result of the analysis should not allow an adversary to learn **any information** about **any individual** in the dataset that the adversary could not learn if the individual was not in the dataset.

Does this work?

No — the revealed statistic needs to be useful, and so cannot be independent of all participants

Can we weaken part of this statement?

Attempt #3 at defining privacy

The result of the analysis should not allow an adversary to learn **any information** about **most individuals** in the dataset that the adversary could not learn if the individual was not in the dataset.

Does this work?

Kind of — possible to sample a small group of users and then compute statistics from only their data
… but some users have to give up their privacy

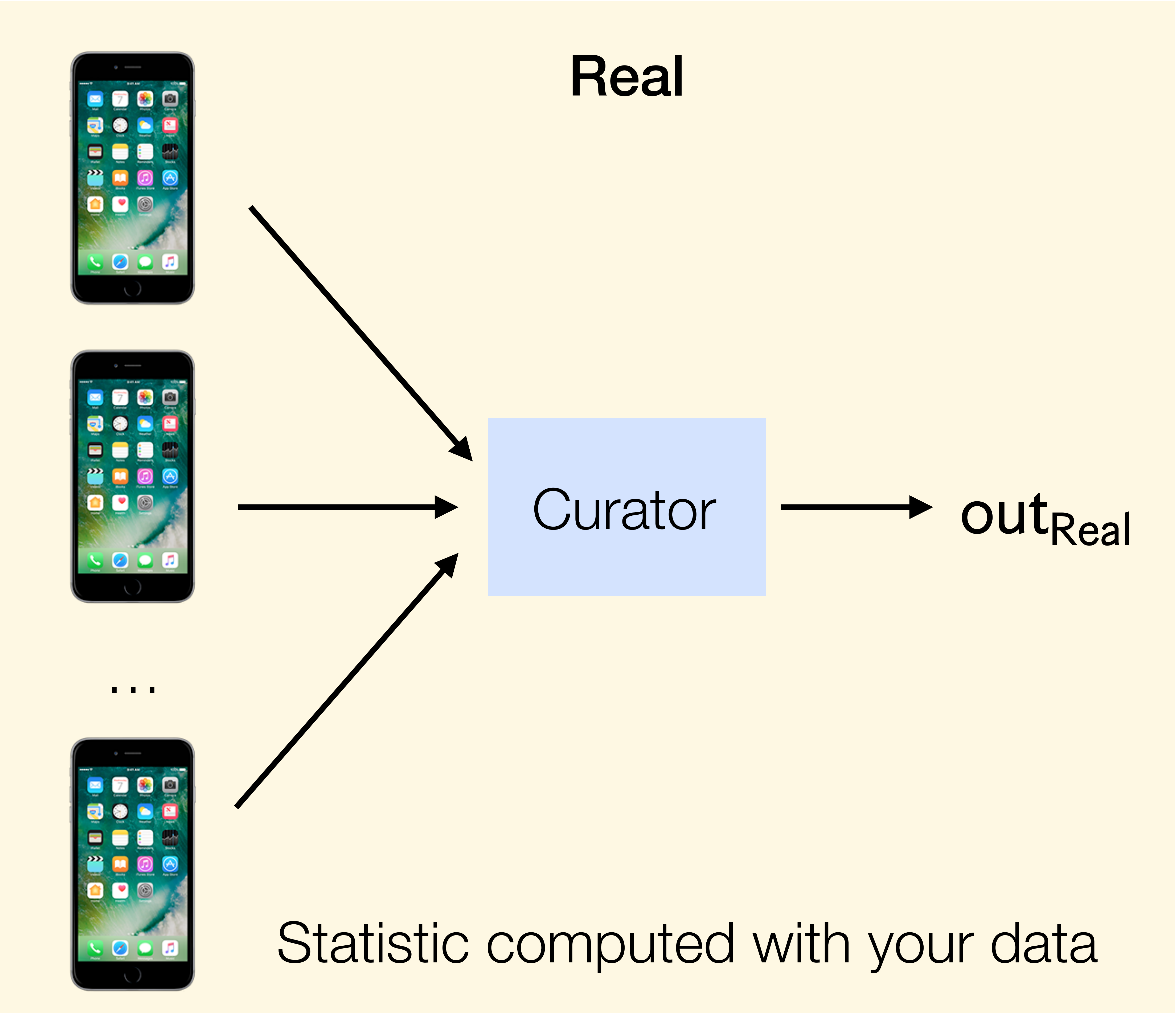
Attempt #4 at defining privacy

The result of the analysis should not allow an adversary to learn **anything new confidently** about **any individual** in the dataset that the adversary could not learn if the individual was not in the dataset.

Does this work?

Yes, if we define what “anything new confidently” means — intuition behind differential privacy

Differential privacy



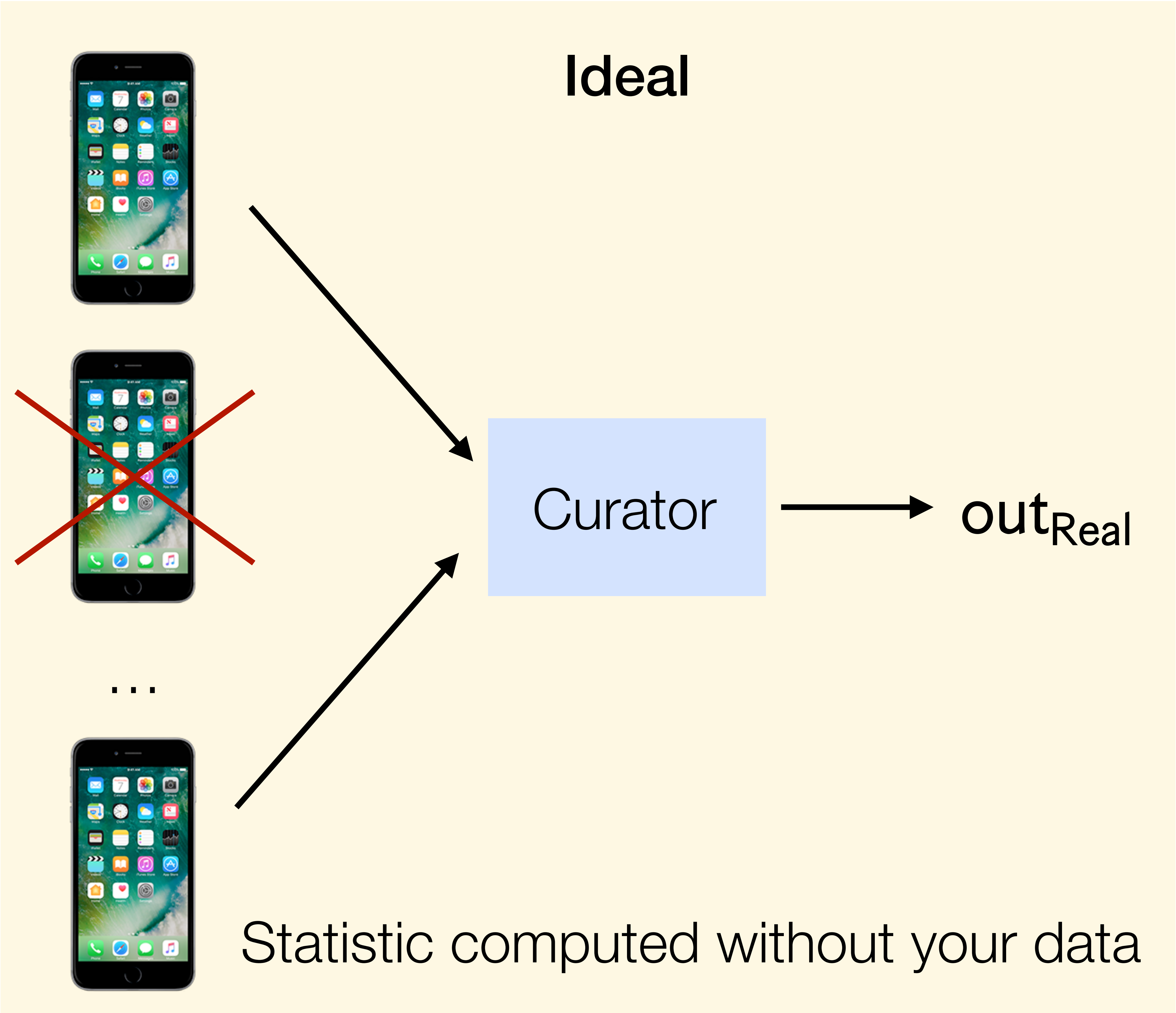
Real

Curator

$\mathsf{out}_{\mathsf{Real}}$

…

Statistic computed with your data



Ideal

Curator

$\mathsf{out}_{\mathsf{Real}}$

…

Statistic computed without your data

$\mathsf{out}_{\mathsf{Real}} \approx \mathsf{out}_{\mathsf{Ideal}}$ (Not cryptographic indistinguishability)

Formalizing differential privacy

[Dwork, Sherry, Nissim, Smith]

Mechanism $\mathscr{M} : \mathscr{X}^n \to \mathscr{Y}$

For database with n rows of type $\mathscr{X}$ and output statistic $\mathscr{Y}$

Two databases are “neighboring” if they differ in at most one row

A mechanism $\mathscr{M}$ is ε differentially private if for all pairs of “neighboring databases” D, D' and every set of values $S \in \mathscr{Y}$:

$$\text{Pr}[\mathscr{M}(D) \in S] \leq e^{\varepsilon} \cdot \text{Pr}[\mathscr{M}(D') \in S]$$

Formalizing differential privacy

[Dwork, Sherry, Nissim, Smith]

Mechanism $\mathscr{M} : \mathscr{X}^n \to \mathscr{Y}$

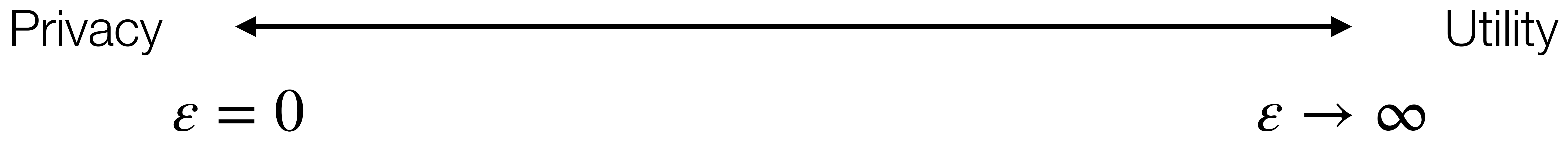
For database with n rows of type $\mathscr{X}$ and output statistic $\mathscr{Y}$

Two databases are “neighboring” if they differ in at most one row

A mechanism $\mathscr{M}$ is ε differentially private if for all pairs of “neighboring databases” D, D' and every set of values $S \in \mathscr{Y}$:

$$\Pr[\mathscr{M}(D) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathscr{M}(D') \in S]$$

How to set ϵ?



Formalizing differential privacy

[Dwork, Sherry, Nissim, Smith]

Mechanism $\mathscr{M} : \mathscr{X}^n \to \mathscr{Y}$

For database with n rows of type $\mathscr{X}$ and output statistic $\mathscr{Y}$

Two databases are "neighboring" if they differ in at most one row

A mechanism $\mathscr{M}$ is ε differentially private if for all pairs of "neighboring databases" D, D' and every set of values $S \in \mathscr{Y}$:

$$\Pr[\mathscr{M}(D) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathscr{M}(D') \in S]$$

Intuition: Any "bad event" when Alice is in the database would have happened with similar probability if she was not in the database

Formalizing differential privacy

[Dwork, Sherry, Nissim, Smith]

Mechanism $\mathscr{M} : \mathscr{X}^n \to \mathscr{Y}$

For database with n rows of type $\mathscr{X}$ and output statistic $\mathscr{Y}$

Two databases are “neighboring” if they differ in at most one row

A mechanism $\mathscr{M}$ is ε differentially private if for all pairs of “neighboring databases” D, D' and every set of values $S \in \mathscr{Y}$:

$$\text{Pr}[\mathscr{M}(D) \in S] \leq e^{\varepsilon} \cdot \text{Pr}[\mathscr{M}(D') \in S]$$

Question: What are some databases where, even if we apply differential privacy, the published statistic still reveals some sensitive information about a user?

Example

Phone number	Timestamp
123-456-7890	11/17 8:00AM
911	11/17 9:00AM
123-456-7890	11/17 9:30AM
…	…

Will a ε differentially private mechanism over this database hide the following with ε-DP:

- Frequent calls to Bob (123-456-7890)? X
- One call to 911? ✓

Properties of differential privacy

Post-processing: Let $\mathscr{M} : \mathscr{X}^n \to \mathscr{Y}$ be ϵ differentially privacy and let $f : \mathscr{Y} \to \mathscr{Z}$ be any (randomized) function. Then $(f \cdot \mathscr{M}) : \mathscr{X}^n \to \mathscr{Z}$ is ε differentially privacy.

- Intuition: Whatever computation is performed on the results of a differentially private query, the answer is still differentially private
- Implies resilience to any side information

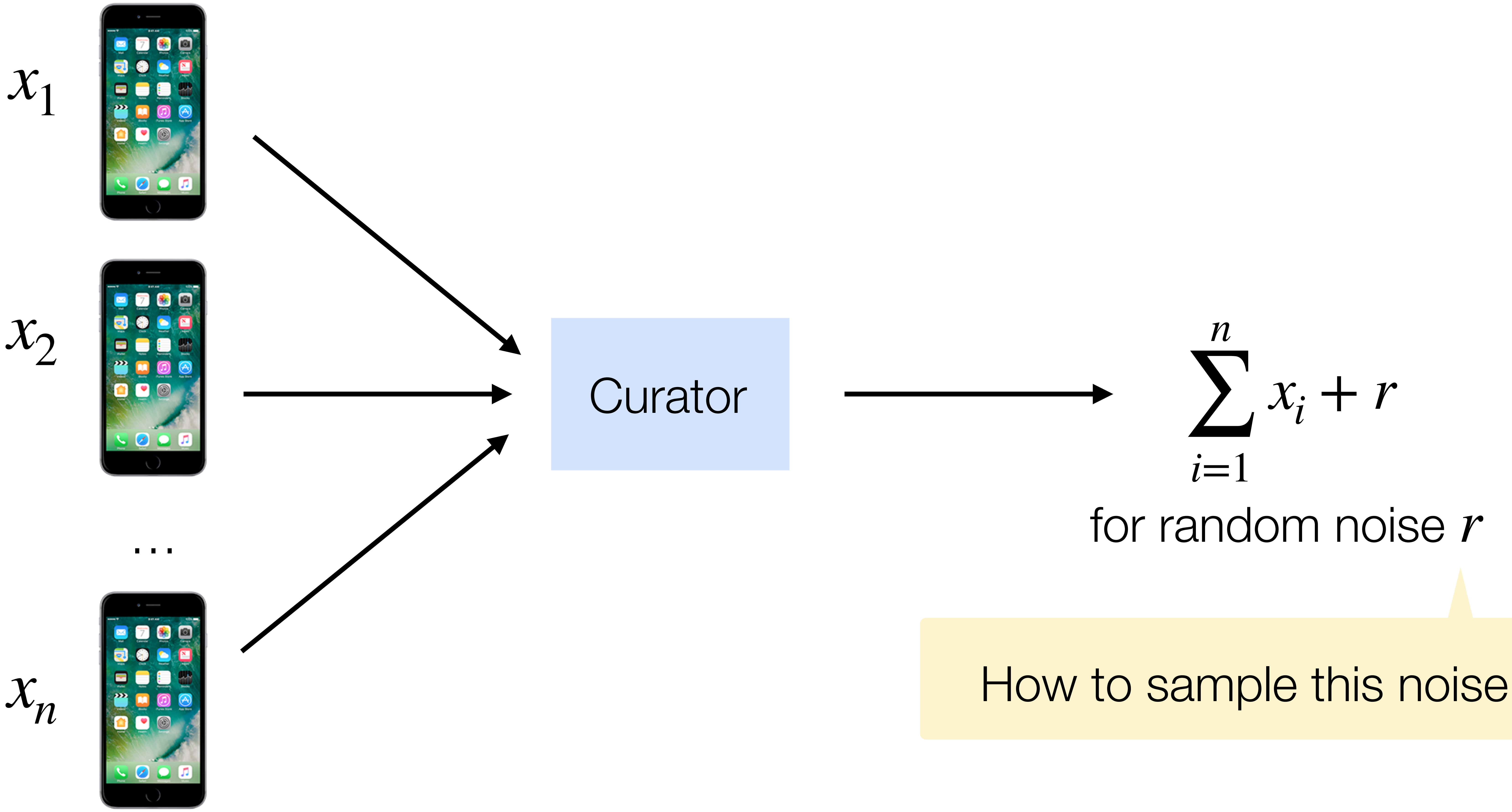
Composition: Let $\mathscr{M}_1, \mathscr{M}_2, \ldots, \mathscr{M}_n$ be mechanisms where $\mathscr{M}_i : \mathscr{X}^n \to \mathscr{Y}_i$ is ε_i differentially private. Then $\mathscr{M}(D) \to (\mathscr{M}_1(D), \mathscr{M}_2(D), \ldots, \mathscr{M}_n(D))$ is ε differentially private for $\varepsilon = \sum_{i=1}^{n} \varepsilon_i$

- Intuition: Answering more queries reveals more information
- Bounds overall privacy risk

Outline

1. Differential privacy definition
2. **Differential privacy mechanism**
3. Differential privacy applications
4. Logistics
5. Student presentation

Differential privacy at a high level



x_1

x_2

…

x_n

Curator

$$\sum_{i=1}^{n} x_i + r$$

for random noise r

How to sample this noise r?

Defining query sensitivity

Sensitivity: For a query $q : \mathscr{X}^n \to \mathbb{R}$, the sensitivity of q is $\Delta q = \max_{D \sim D'} |q(D) - q(D')|$

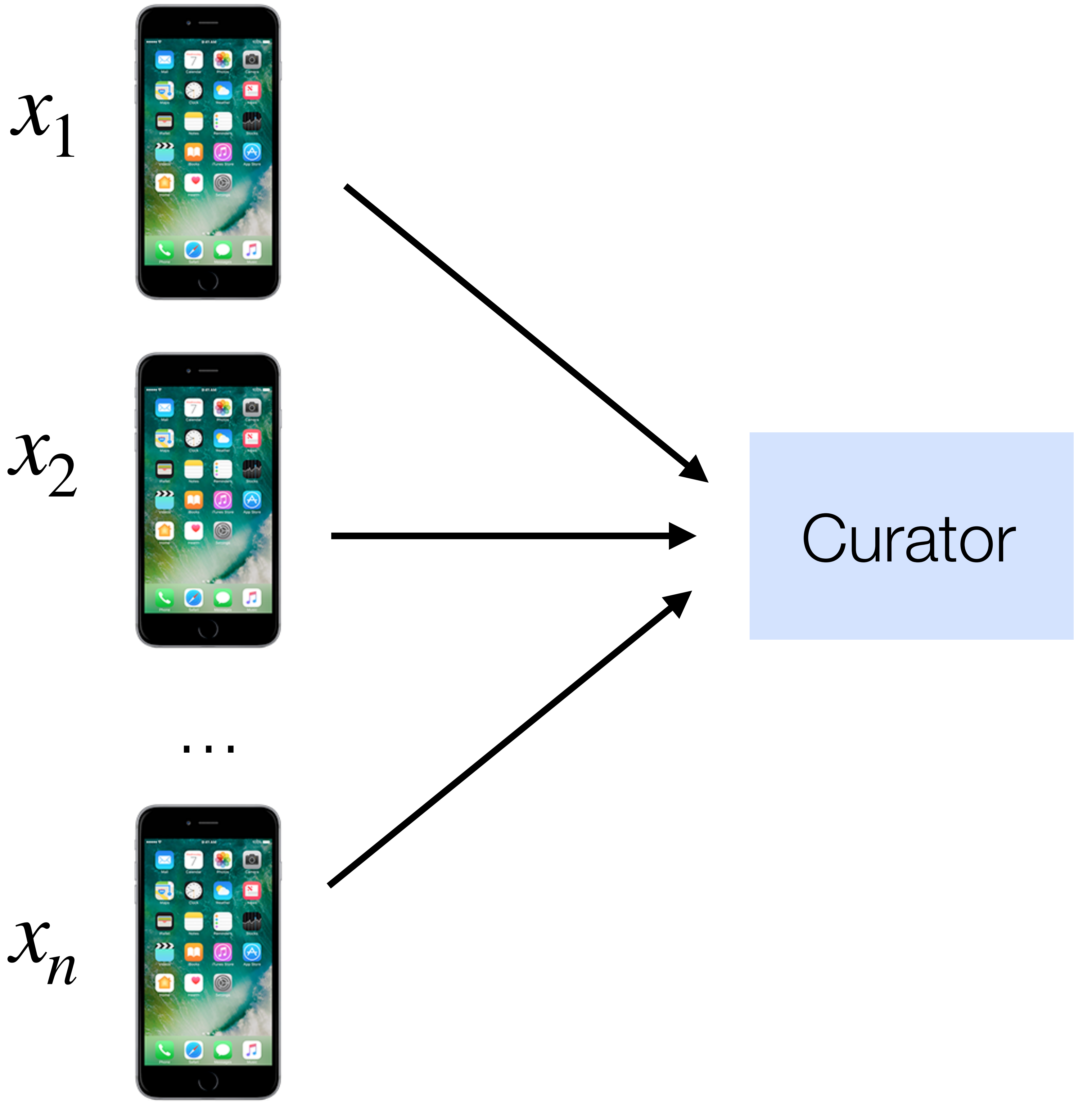
For counting query?

Sensitivity: 1

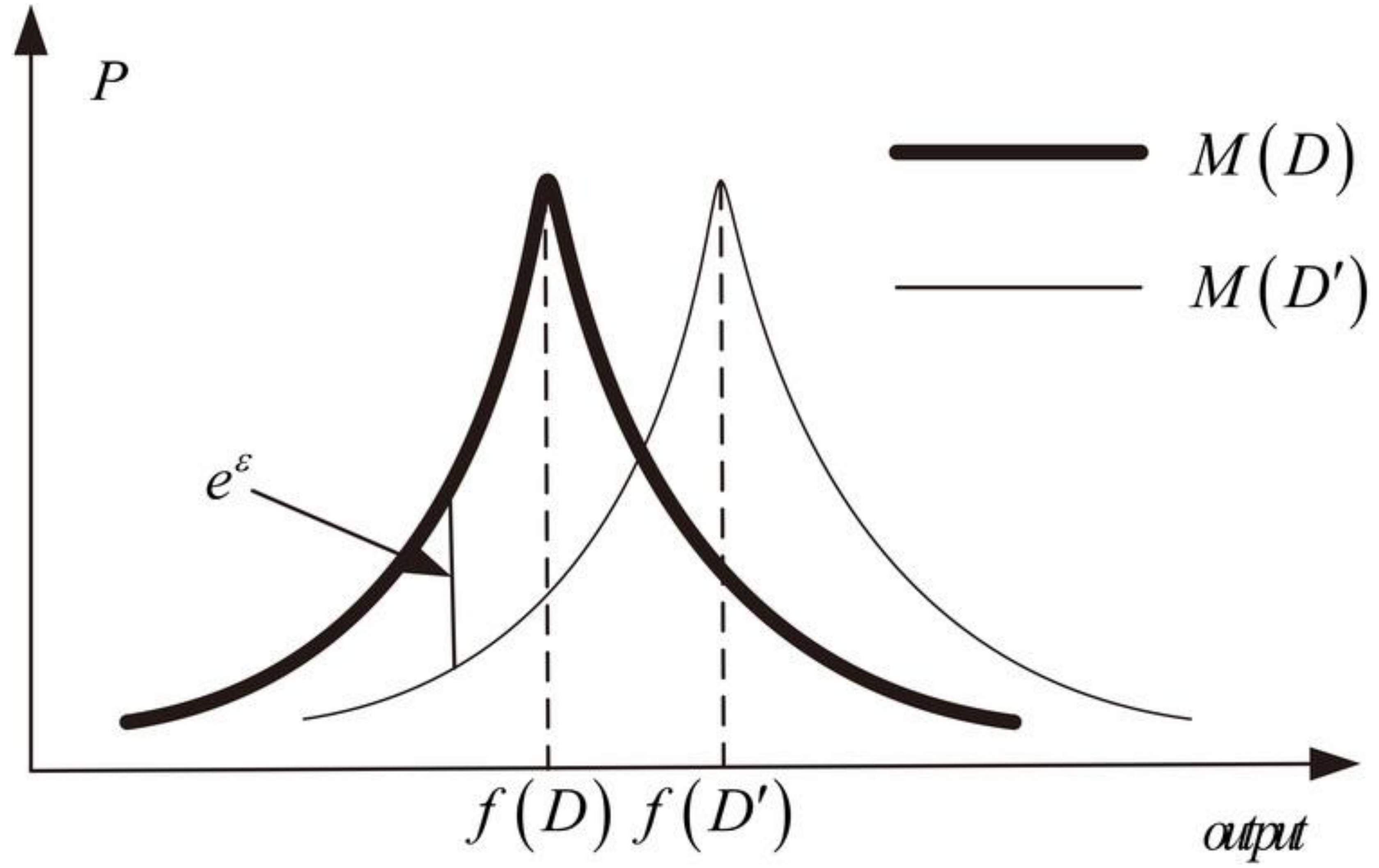
For sums of 8-bit values?

Sensitivity: $2^8 = 256$

Laplace mechanism



Curator



$$q(x_1, x_2, \ldots, x_n) + r$$

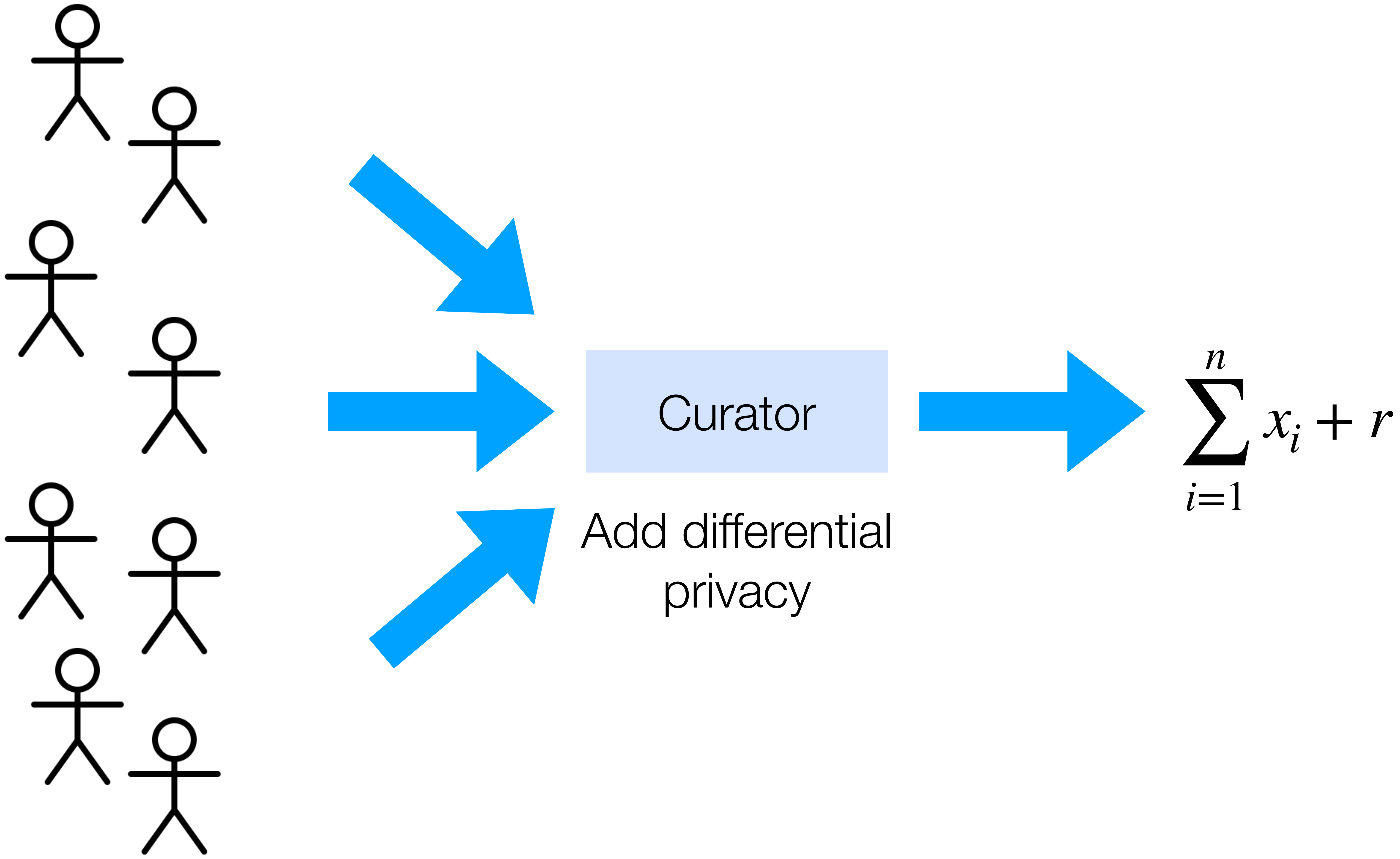
Where $r \sim \text{Lap}\left(\frac{\Delta q}{\varepsilon}\right)$

Image: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/methodologies/applyingdifferentialprivacyprotectiontoonsmortalitydatapilotstudy

Outline

1. Differential privacy definition
2. Differential privacy mechanism
3. **Differential privacy applications**
4. Logistics
5. Student presentation

Central differential privacy

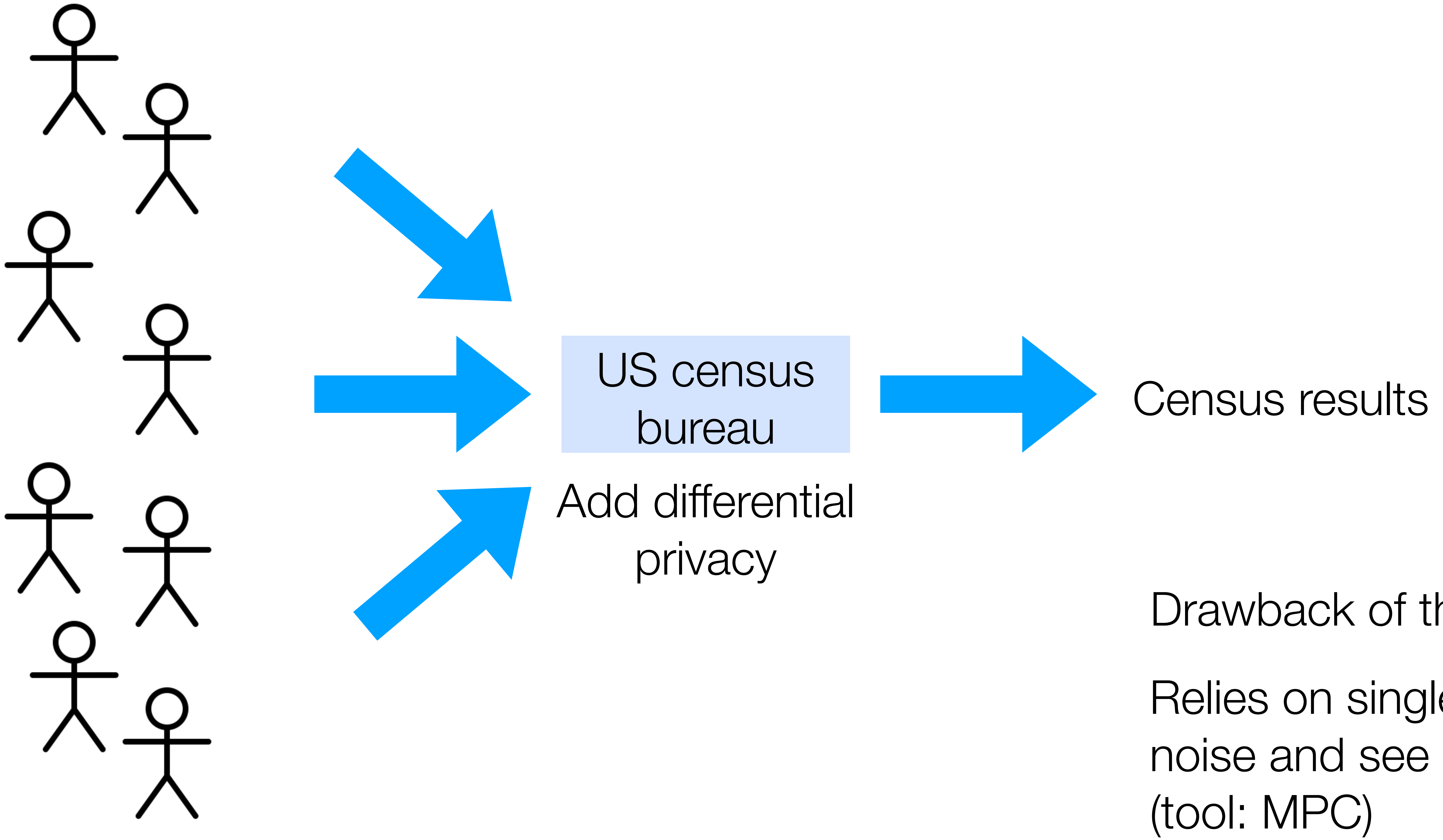


Curator

Add differential privacy

$$\sum_{i=1}^{n} x_i + r$$

Central differential privacy: US census

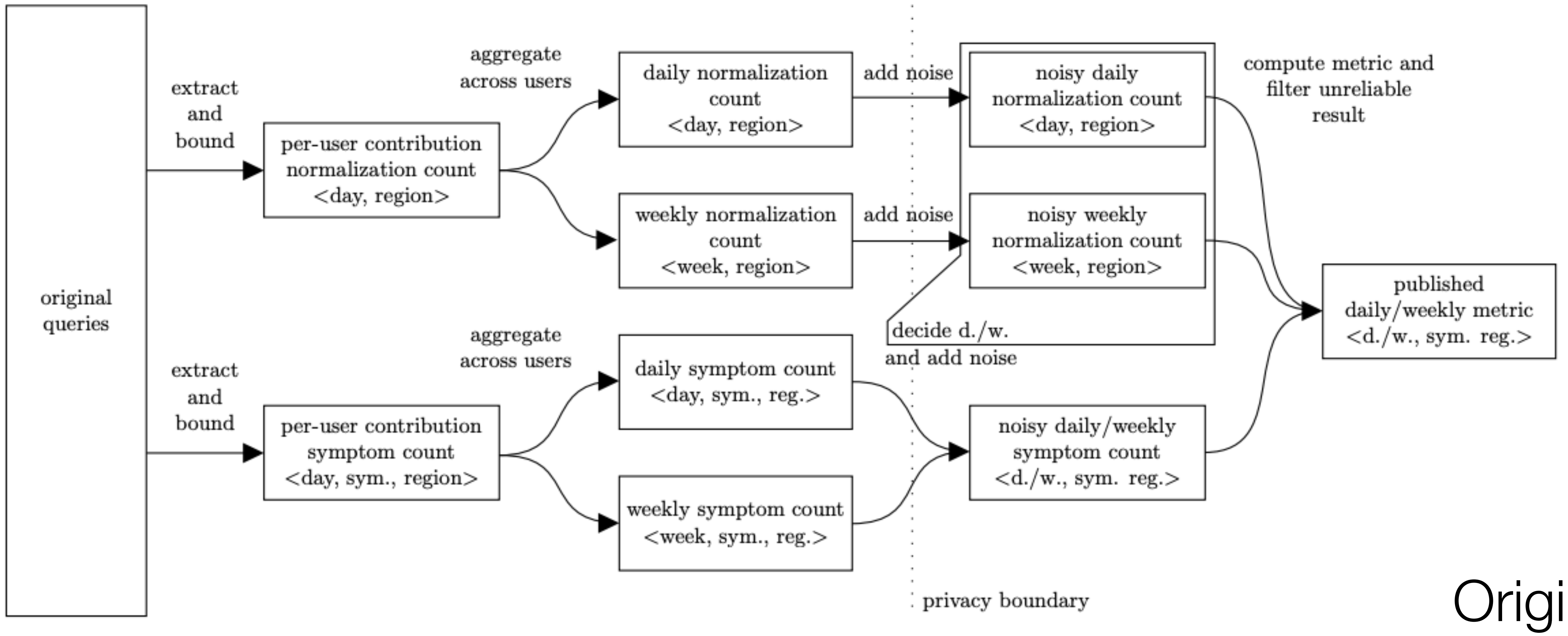


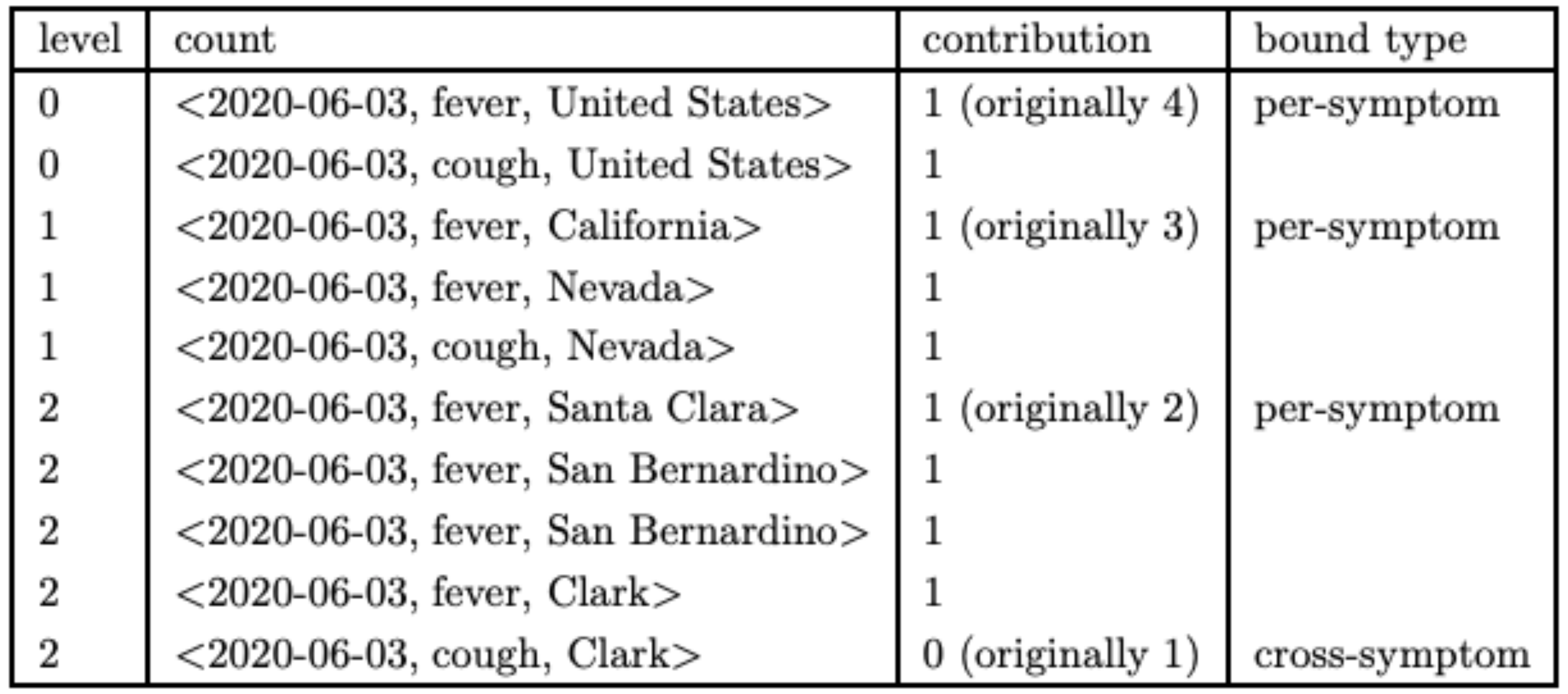
Drawback of this model?

Relies on single party to add DP noise and see individual contributions (tool: MPC)

Google COVID-19 search trends

Also in the central DP model (ε = 1.68)


original queries
extract and bound
per-user contribution normalization count <day, region>
aggregate across users
daily normalization count <day, region>
weekly normalization count <week, region>
add noise
add noise
noisy daily normalization count <day, region>
noisy weekly normalization count <week, region>
compute metric and filter unreliable result
published daily/weekly metric <d./w., sym. reg.>
decide d./w. and add noise
extract and bound
per-user contribution symptom count <day, sym., region>
aggregate across users
daily symptom count <day, sym., reg.>
weekly symptom count <week, sym., reg.>
noisy daily/weekly symptom count <d./w., sym. reg.>
privacy boundary



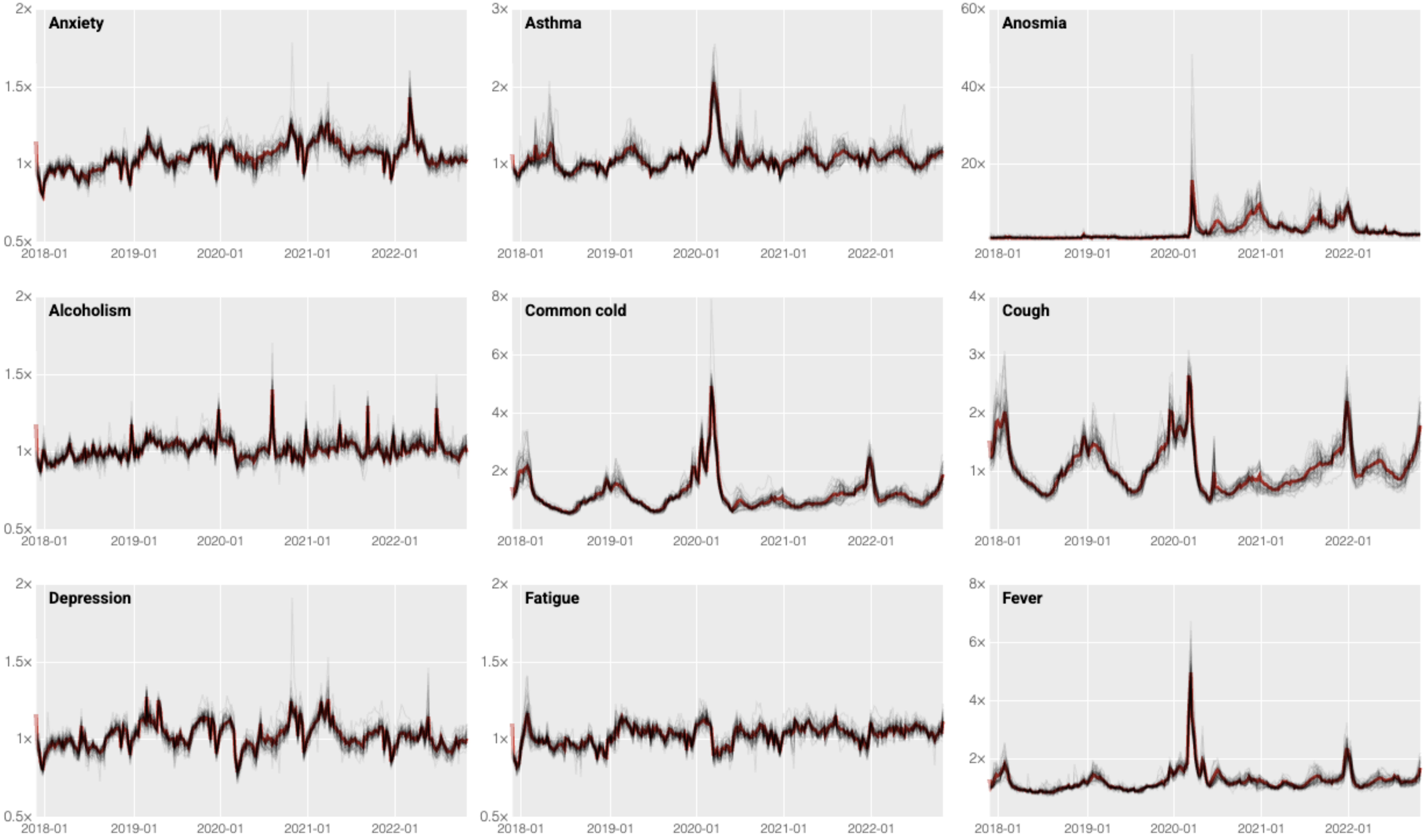
level	count	contribution	bound type
0	<2020-06-03, fever, United States>	1 (originally 4)	per-symptom
0	<2020-06-03, cough, United States>	1	
1	<2020-06-03, fever, California>	1 (originally 3)	per-symptom
1	<2020-06-03, fever, Nevada>	1	
1	<2020-06-03, cough, Nevada>	1	
2	<2020-06-03, fever, Santa Clara>	1 (originally 2)	per-symptom
2	<2020-06-03, fever, San Bernardino>	1	
2	<2020-06-03, fever, San Bernardino>	1	
2	<2020-06-03, fever, Clark>	1	
2	<2020-06-03, cough, Clark>	0 (originally 1)	cross-symptom

Original database (with bounded contributions)

https://arxiv.org/pdf/2009.01265

https://pair-code.github.io/covid19_symptom_dataset/?country=US

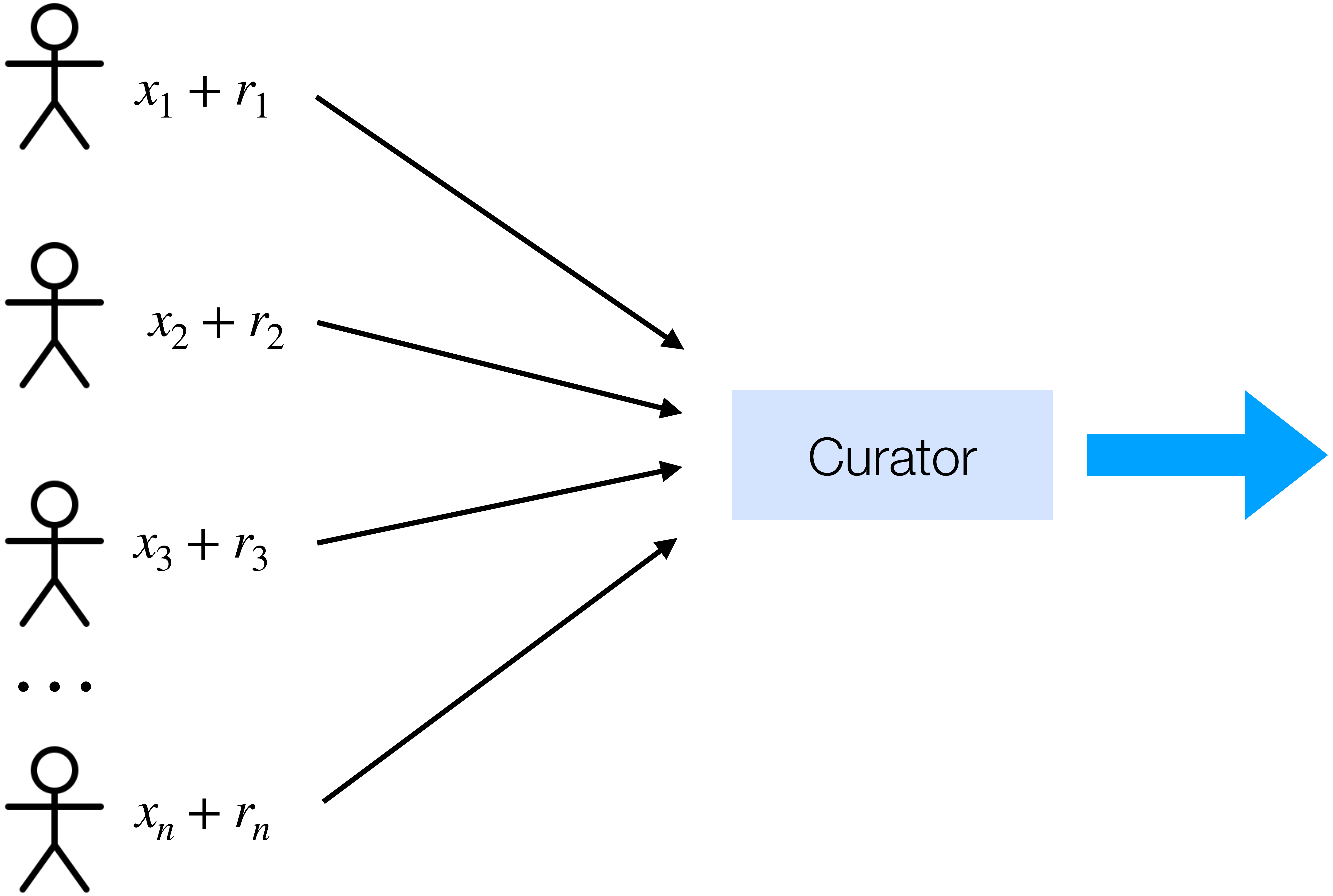
Google COVID-19 search trends



https://arxiv.org/pdf/2009.01265

https://pair-code.github.io/covid19_symptom_dataset/?country=US

Local differential privacy



Advantages?

No central point of privacy failure

$$\sum_{i=1}^{n} (x_i + r_i)$$

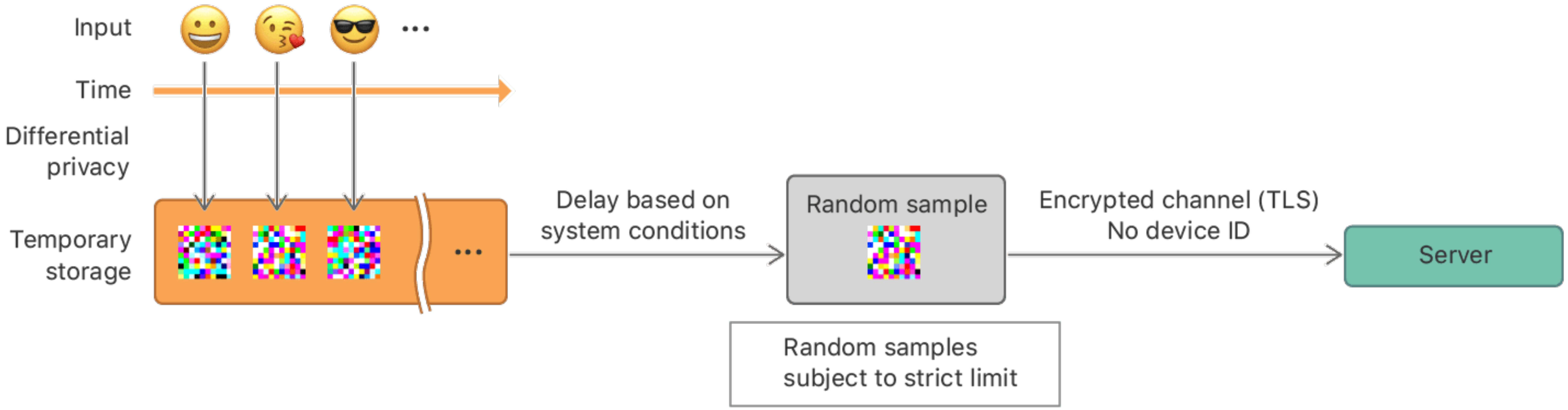
Disadvantages?

Need more noise for privacy.

If we make ε too small, then the final statistic is not very useful

MPC + central DP is better utility vs privacy tradeoff

Local differential privacy: Apple


Input
Time
Differential privacy
Temporary storage
...
Delay based on system conditions
Random sample
Random samples subject to strict limit
Encrypted channel (TLS) No device ID
Server

https://machinelearning.apple.com/research/learning-with-privacy-at-scale

Count mean sketch with differential privacy (Apple)

Setup:

- Server samples a list of hash function $H_1, H_2, \ldots, H_k$
- Server initializes a matrix M of size $m \times k$ to be all 0s

https://machinelearning.apple.com/research/learning-with-privacy-at-scale

Count mean sketch with differential privacy (Apple)

Setup:

- Server samples a list of hash function $H_1, H_2, \ldots, H_k$
- Server initializes a matrix M of size $m \times k$ to be all 0s

To upload val, the client:

- Client samples a hash function H_i from list $H_1, H_2, \ldots, H_k$
- Client computes $H_i(\mathsf{val}) = x$
- Client generates a one-hot vector of length m that is 0 everywhere and 1 at x
- Client adds differential privacy to vector entries and sends the vector and i to the server
- The server adds the client's vector to row i of its matrix M

https://machinelearning.apple.com/research/learning-with-privacy-at-scale

Count mean sketch with differential privacy (Apple)

Setup:

- Server samples a list of hash function $H_1, H_2, \ldots, H_k$
- Server initializes a matrix M of size $m \times k$ to be all 0s

To upload **val**, the client:

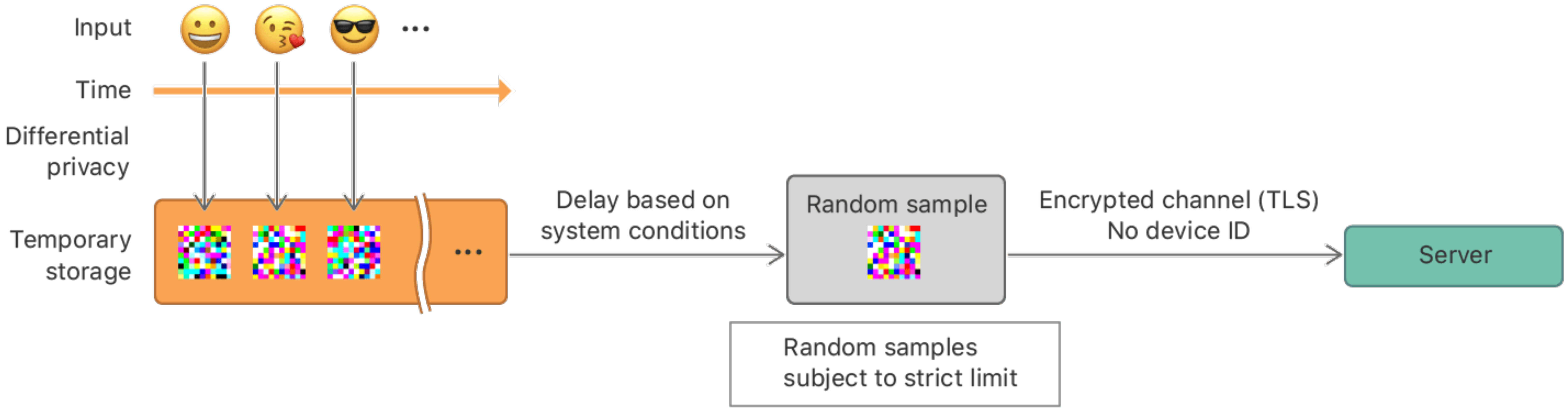
- Client samples a hash function H_i from list $H_1, H_2, \ldots, H_k$
- Client computes $H_i(\text{val}) = x$
- Client generates a one-hot vector of length m that is 0 everywhere and 1 at x
- Client adds differential privacy to vector entries and sends the vector and i to the server
- The server adds the client's vector to row i of its matrix M

To compute the mean of **val**:

- Server computes the average over $M[i, H_i(\text{val})]$ for $i \in [k]$
- Note: requires the server to know the dictionary of domain elements

https://machinelearning.apple.com/research/learning-with-privacy-at-scale

Local differential privacy: Apple



Input

Time

Differential privacy

Temporary storage

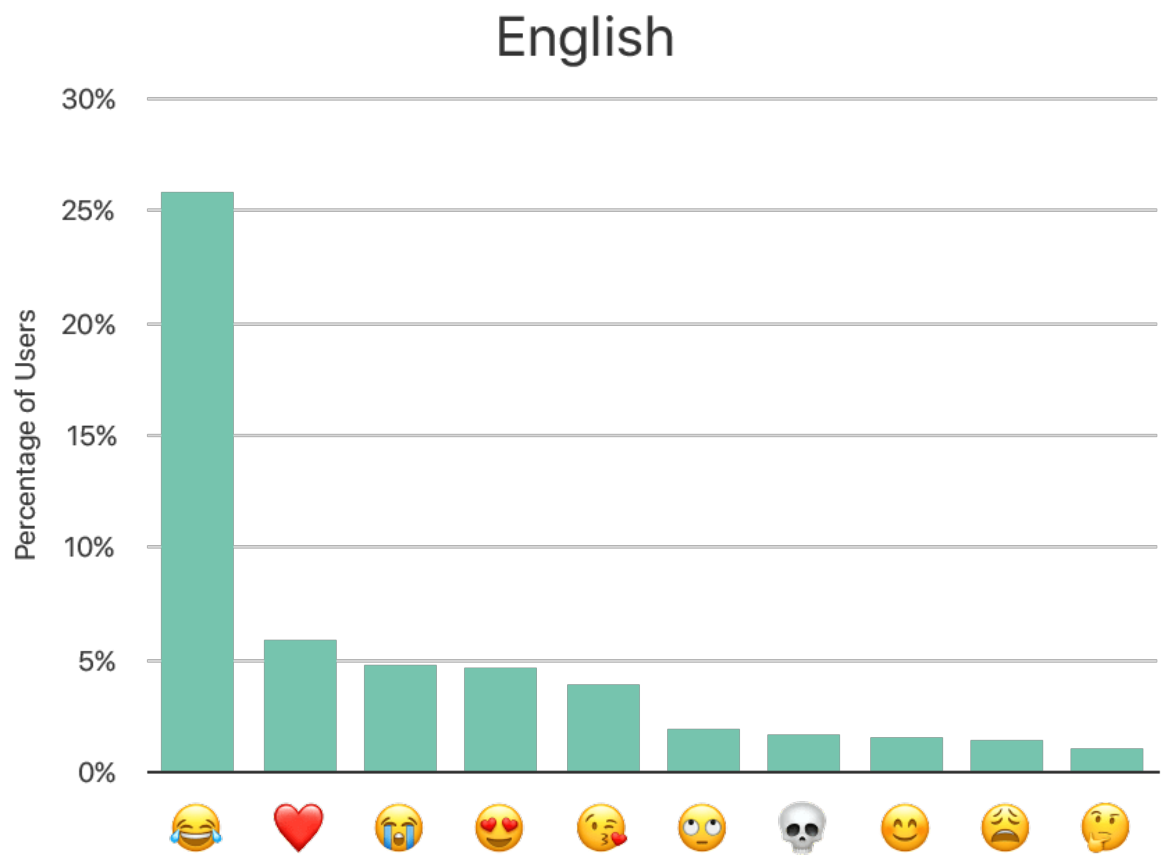
Delay based on system conditions

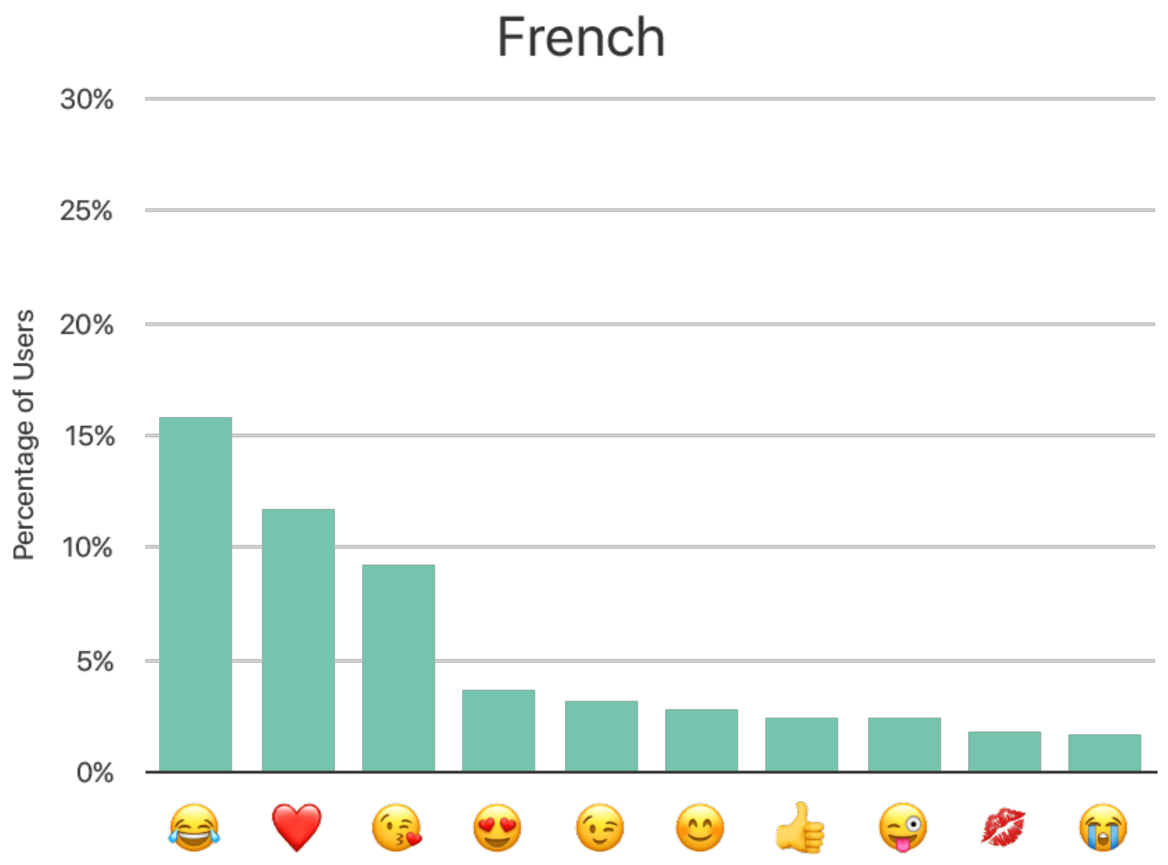
Random sample

Random samples subject to strict limit

Encrypted channel (TLS) No device ID

Server



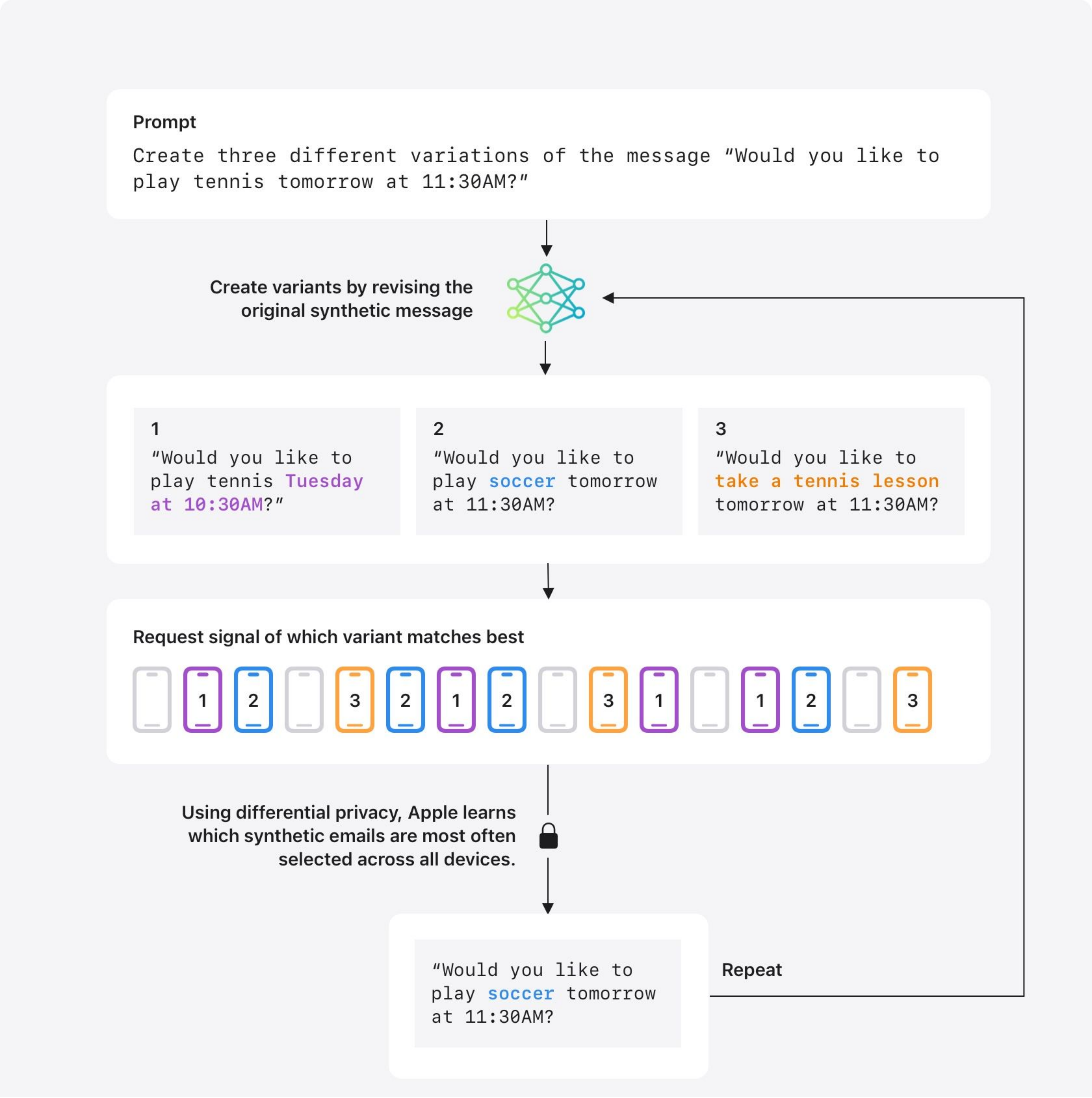


(From 2017)

$\varepsilon = 4$

https://machinelearning.apple.com/research/learning-with-privacy-at-scale

Local differential privacy: Apple



Prompt

Create three different variations of the message "Would you like to play tennis tomorrow at 11:30AM?"

Create variants by revising the original synthetic message

1 "Would you like to play tennis Tuesday at 10:30AM?"

2 "Would you like to play soccer tomorrow at 11:30AM?

3 "Would you like to take a tennis lesson tomorrow at 11:30AM?

Request signal of which variant matches best

Using differential privacy, Apple learns which synthetic emails are most often selected across all devices.

"Would you like to play soccer tomorrow at 11:30AM?

Repeat

https://machinelearning.apple.com/research/differential-privacy-aggregate-trends

Outline

1. Differential privacy definition
2. Differential privacy mechanism
3. Differential privacy applications
4. **Logistics**
5. Student presentation

Logistics

Comments for project reports are in Gradescope

Final project presentations (12/2 and 12/4)

- 9 minute slots (hard cutoff time)
- Time for a few questions as the next group is setting up

Final project report (due 12/2)

Outline

1. Differential privacy definition
2. Differential privacy mechanism
3. Differential privacy applications
4. Logistics
5. **Student presentation**

References

Bavadekar, Shailesh, Andrew Dai, John Davis, Damien Desfontaines, Ilya Eckstein, Katie Everett, Alex Fabrikant et al. "Google COVID-19 search trends symptoms dataset: Anonymization process description (version 1.0)." *arXiv preprint arXiv:2009.01265* (2020).

Dwork, Cynthia, and Aaron Roth. "The algorithmic foundations of differential privacy." *Foundations and trends® in theoretical computer science* 9, no. 3–4 (2014): 211-407.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. "Calibrating noise to sensitivity in private data analysis." In *Theory of cryptography conference*, pp. 265-284. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

Narayanan, Arvind, and Vitaly Shmatikov. "Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset). The University of Texas at Austin." In *Proceedings of the 29th IEEE Symposium on Security and Privacy, Oakland, CA, USA*, pp. 18-21. 2008.

https://6893.csail.mit.edu/lec15.pdf

https://systems.cs.columbia.edu/private-systems-class/lectures/02-differential-privacy.pdf

https://crypto.stanford.edu/cs355/19sp/lec9.pdf

https://www2.census.gov/about/policies/2020-03-05-differential-privacy.pdf

https://machinelearning.apple.com/research/differential-privacy-aggregate-trends

https://machinelearning.apple.com/research/learning-with-privacy-at-scale