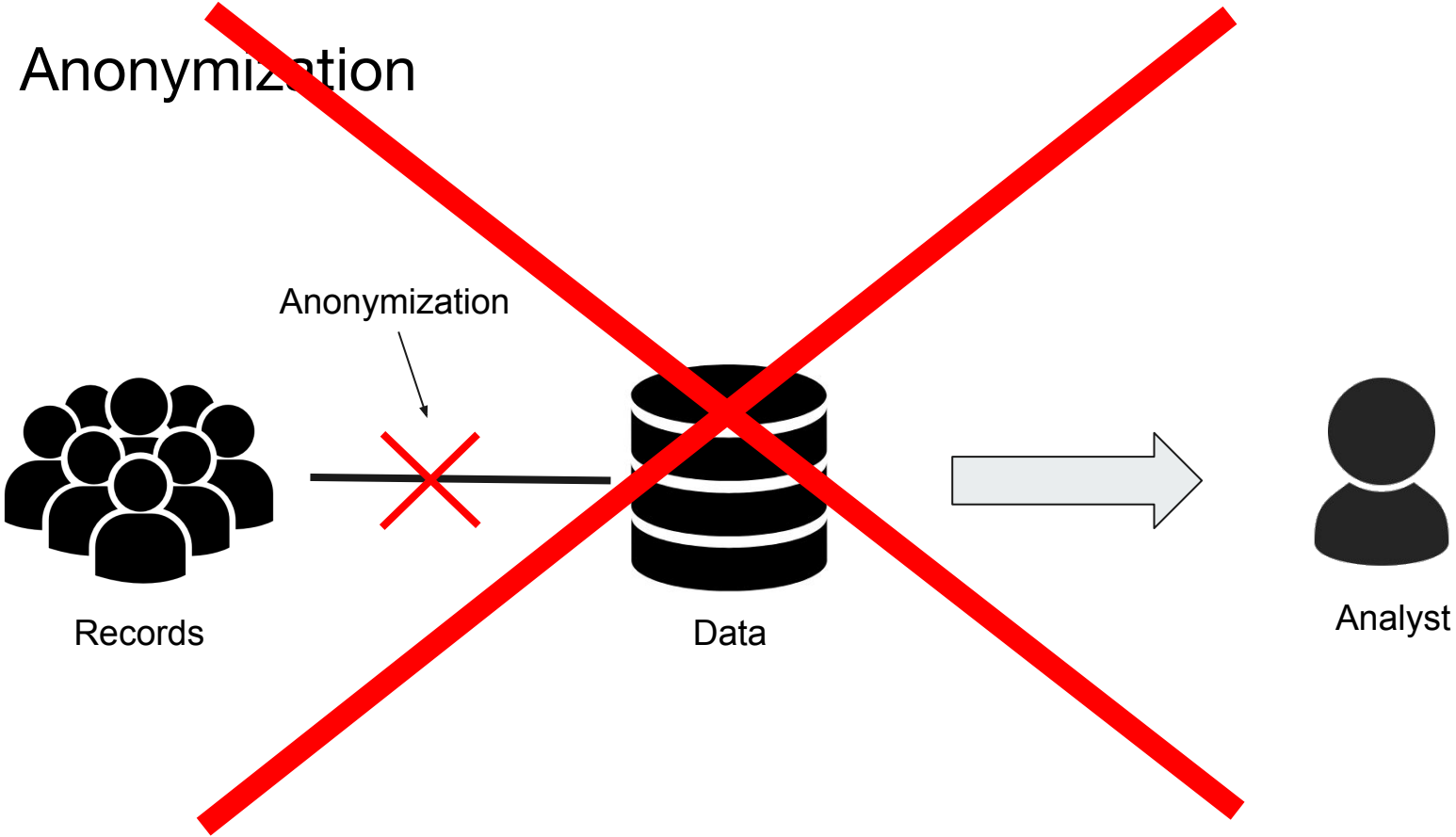
When the Signal is in the Noise: Exploiting Diffix's Sticky Noise

Andrea Gadotti*, Florimond Houssiau*, **Luc Rocher***,
Benjamin Livshits, Yves-Alexandre de Montjoye

Imperial College London
COMPUTATIONAL PRIVACY GROUP
UCLouvain
fnrs LA LIBERTÉ DE CHERCHER

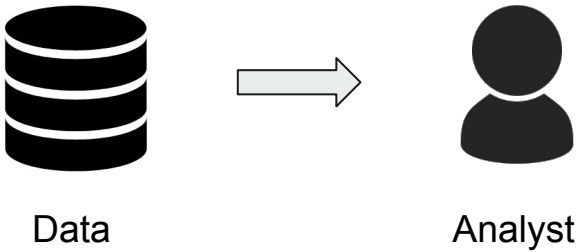
Anonymization



A different model: data query systems

From de-identification...

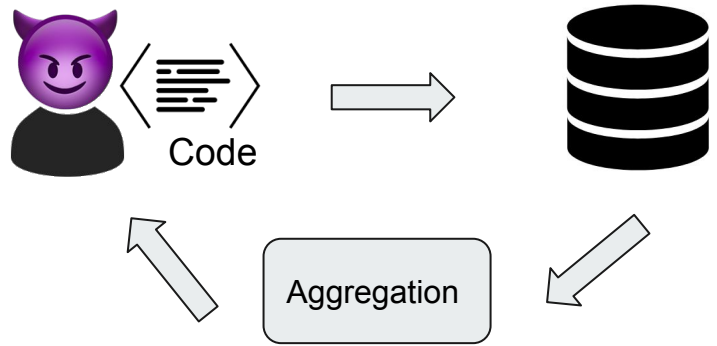
- Individual-level data
- No control over analyses



... to data query systems

- Aggregation
- Additional security and privacy measures

WHAT IF ANALYST IS MALICIOUS?



"How many people named Bob have a salary ≤ £2000"

Q1 = "How many people have a salary ≤ £2000"

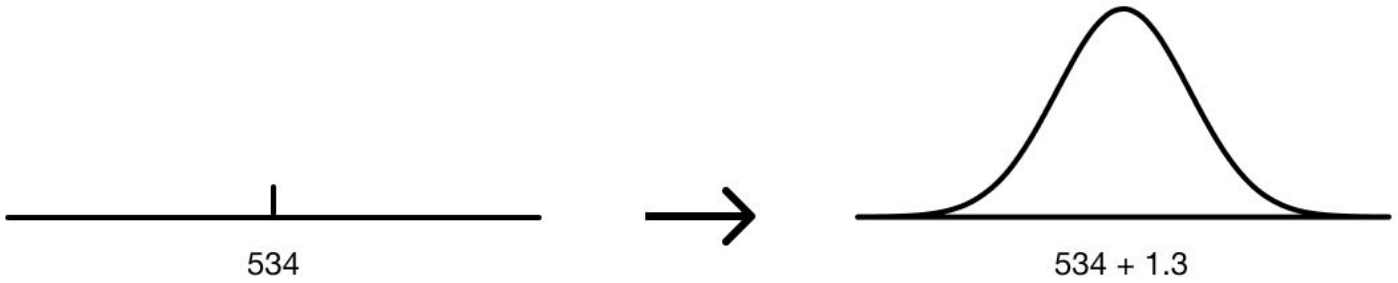
Q2 = "How many people not named Bob have a salary ≤ £2000"

→ Q1 - Q2 = [0 or 1]

This is called *differencing attack*.

Random noise to prevent privacy attacks

“How many people have a salary ≤ £1500?”



Reconstruction attacks and differential privacy

First **reconstruction attack** (Dinur and Nissim, 2003). If noise is not enough → attacker can reconstruct the full dataset in polynomial time.

Since then, the attack has been generalized and improved.

One solution: **differential privacy** (Dwork et al., 2006).

Pros:

- provable and meaningful guarantee
- mathematical framework for privacy/utility

Cons (as of today):

- adds too much noise in many cases
- hard to allow many queries
- hard to provide good usability/flexibility

A heuristic-based data query system: Diffix

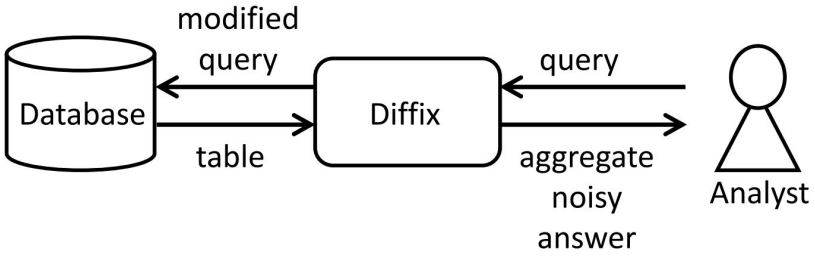
Diffix is a privacy-preserving database system

Diffix is a patented commercial system developed by the company *Aircloak* and researchers at the Max Planck Institute for Software Systems.

aircloak

Diffix operates as an **SQL proxy** between the analyst and the database.

- **Rich SQL syntax**
- **Little noise**
- **Infinite queries**



Diffix's noise mechanism: sticky noise

An analyst submits a (count) SQL query Q to Diffix:

```
SELECT count(*)
FROM table
WHERE condition1 AND condition2 [AND ...]
```

To which Diffix responds with:

output = true count + **static noise** + **dynamic noise**

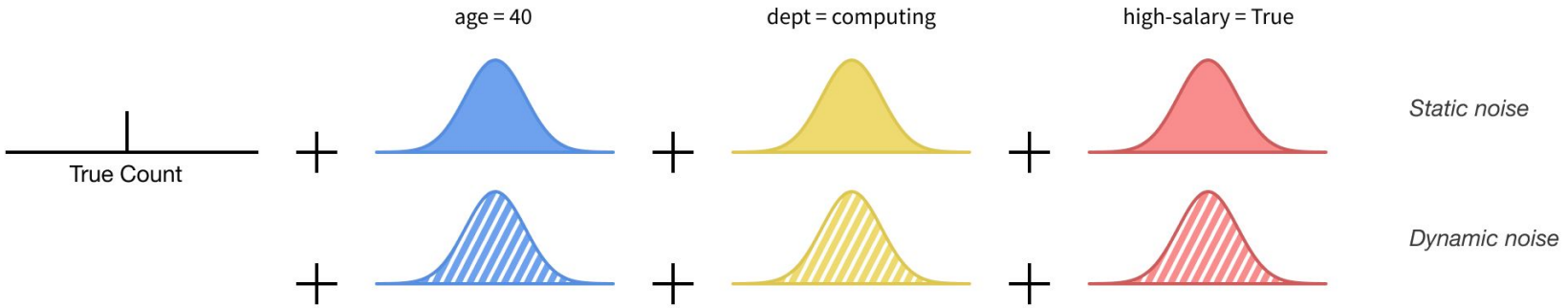
static noise ← **query syntax** of Q

dynamic noise ← query syntax and **user set** of Q

Both noises are **sticky**: issuing the same query gives the same noise

Diffix's noise mechanism: sticky noise

Q = count(age = 40 ∧ dept = Computing ∧ high-salary = True)



Each noise ~ $N(0,1)$

More measures...

Our noise-exploitation attack(s) on Diffix: Exploiting data-dependent noise

Attack model and assumptions

- Dataset has *d* attributes

  $$\{a_1, \ldots, a_{d-1}, \ s\}$$

- One target at a time: Bob
- Attacker wants to infer Bob's attribute *s* (binary).
- Attacker knows:
  - Bob's record is in the dataset
  - The value of *k* attributes about Bob

Example (with *d*=3, *k*=2)

Dataset attributes:

`{age, department, high-salary}`

Secret attribute: `high-salary`

Bob's record: `(40, Computing, true)`

Attacker knows: `(40, Computing)`

Differential attack

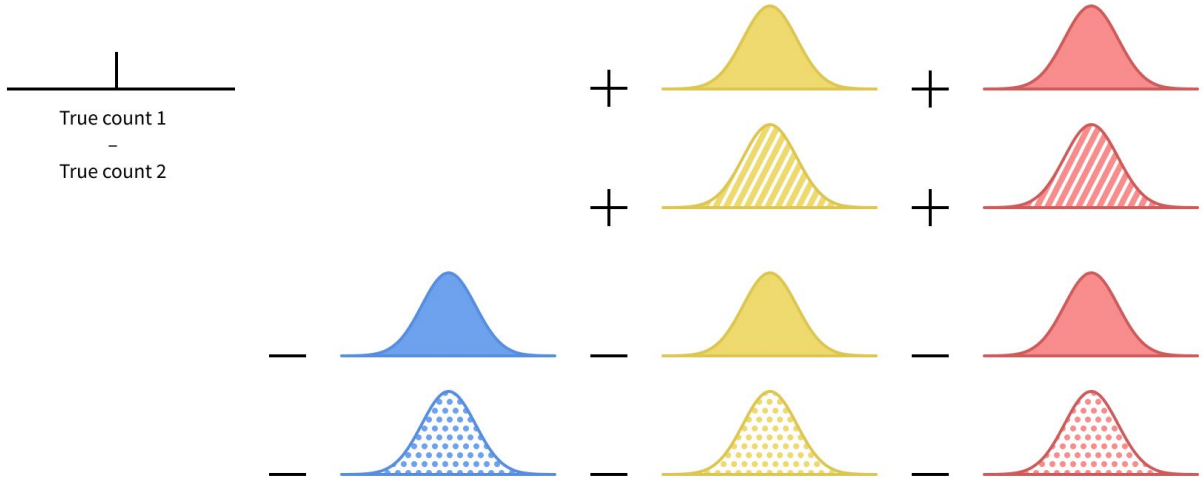
Q1 = count(dept = Computing ∧ high-salary = True)

Q2 = count(age ≠ 40 ∧ dept = Computing ∧ high-salary = True)

Bob :

```
age = 40
dept = computing
high-salary = ?
(unique)
```

Output of Q1 – Q2



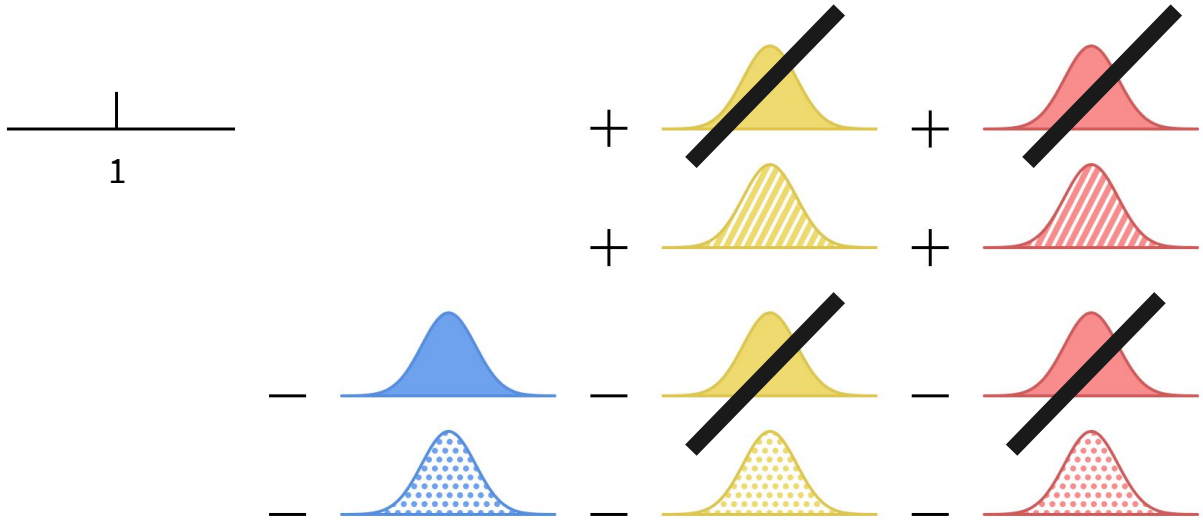
Differential attack

Q1 = count(dept = Computing ∧ high-salary = True)

Q2 = count(age ≠ 40 ∧ dept = Computing ∧ high-salary = True)

Bob :
`age = 40`
`dept = computing`
`high-salary = ?`
`(unique)`

Output of Q1 – Q2
if high-salary = True



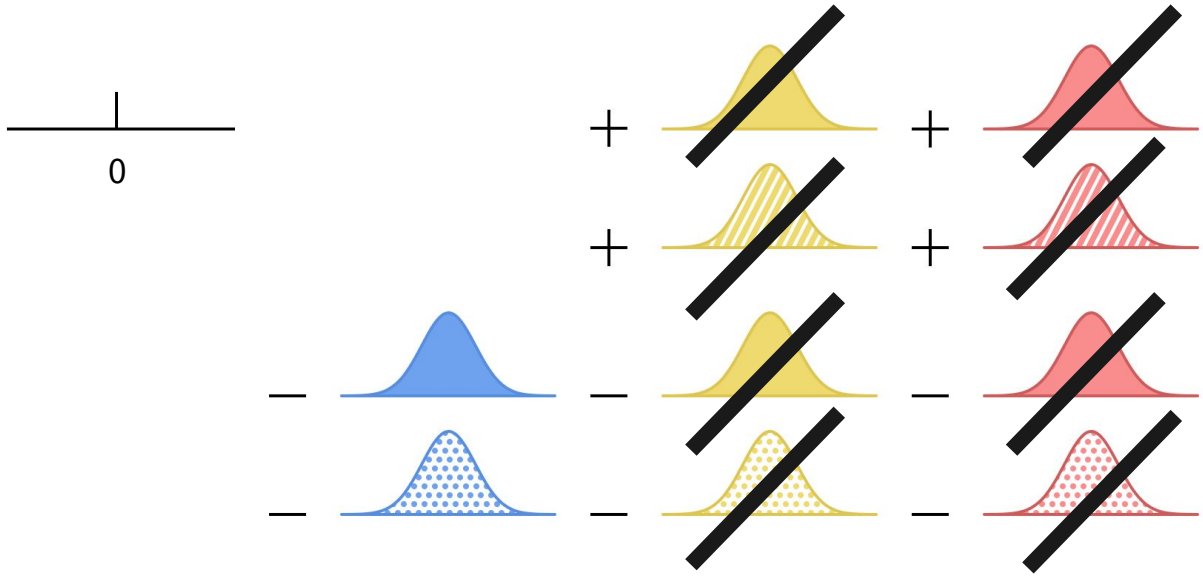
Differential attack

Q1 = count(dept = Computing ∧ high-salary = True)

Q2 = count(age ≠ 40 ∧ dept = Computing ∧ high-salary = True)

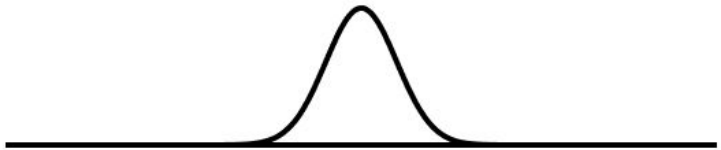
Bob :
`age = 40`
`dept = computing`
`high-salary = ?`
`(unique)`

Output of Q1 – Q2
if high-salary = False



Differential attack

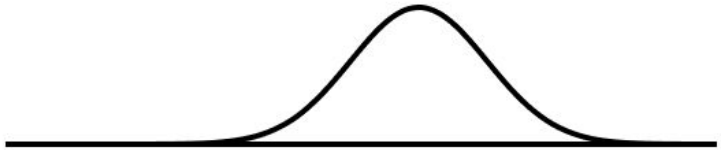
if high-salary = True



0

Q1 - Q2 ~ ***N*(μ=0, σ=2)**

if high-salary = False



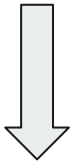
1

Q1 - Q2 ~ ***N*(μ=1, σ=2k+2)**

The cloning attack

Main issues with the differential attack:

1. Assumes that Bob is **unique**
2. Attack queries likely to be **suppressed**



Accuracy not great in some cases

Improved attack: **cloning attack**

- Much better accuracy
- Relies on weaker notion of uniqueness

Value-uniqueness

Definition: *A record is value-unique w.r.t. a set of attributes* $\{a_1, \ldots, a_k\}$ *if all records sharing the same attributes also have the same secret attribute.*

Example

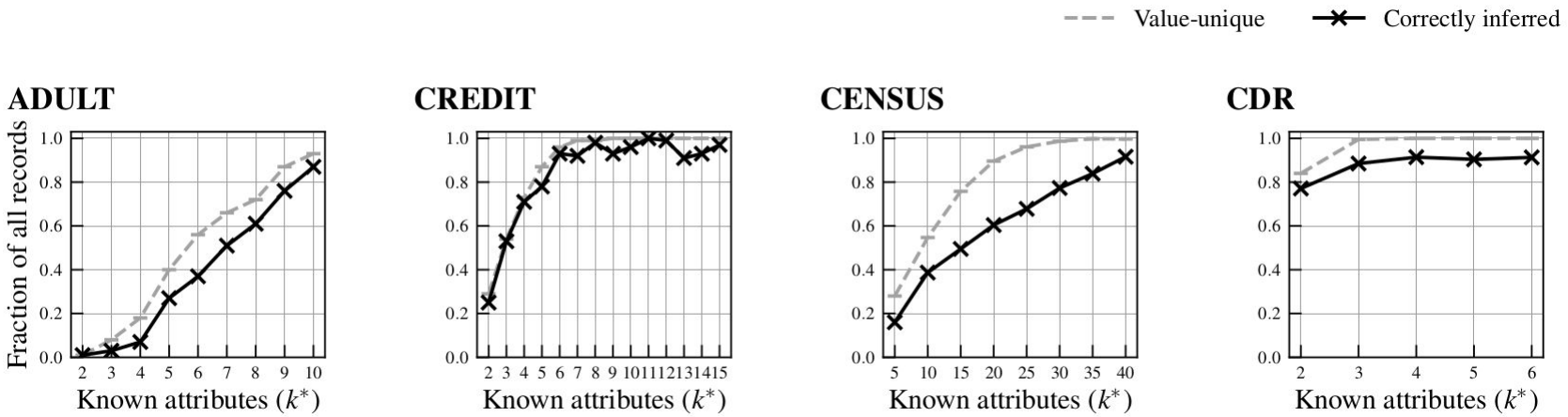
age	dept	high-salary
40	computing	true
40	computing	true
34	math	false
34	math	true

Bob's record
(value-unique)

Alice's record
(not value-unique)

Note:
Value-uniqueness is detected automatically by the cloning attack

Results for the cloning attack



- Attacked and correctly inferred: **~90% of all users**
- Modified attack: **32 queries/user**

Aircloak’s proposed patch

Aircloak’s patch

Remove “dangerous” (low effect) conditions from queries (depending on data).

Comment. Does not address core vulnerability and potentially introduces new one.

Expected patch date: *Q4 2019*

Other attacks on Diffix

Membership inference attack

by A. Pyrgelis, C. Troncoso, E. De Cristofaro

idea: infer whether an individual is in the dataset by training a classifier to tell this from aggregate data.

based on: Apostolos Pyrgelis, Carmela Troncoso, Emiliano De Cristofaro, “Knock Knock, Who’s There? Membership Inference on Aggregate Location Data”, *25th Network and Distributed System Security Symposium (NDSS)*, 2018.

Linear reconstruction attack

by A. Cohen, K. Nissim

idea: send queries targeting “random enough” sets of users and use the results to build a linear system, then reconstruct the database from it.

based on: Dinur, Irit, and Kobbi Nissim. "Revealing information while preserving privacy." *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, 2003.

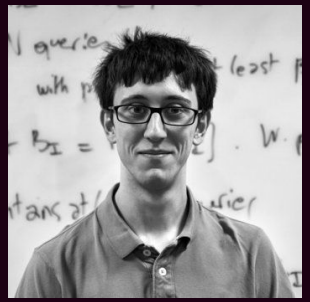
Conclusions

- Anonymization 👎 Data query systems 👍
- Relying on single mechanism is risky
- Defense-in-depth (e.g. query auditing, query rate limiting, etc.)

but also...

- alternatives to differential privacy are useful
- transparency is fundamental

Thank you for your attention!


Find out more at: https://cpg.doc.ic.ac.uk/blog/aircloak-diffix-signal-is-in-the-noise/