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

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

\[ Q1 - Q2 = [0 \text{ or } 1] \]

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.

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:

\[
\begin{align*}
\text{SELECT} & \quad \text{count(}*\text{)} \\
\text{FROM} & \quad \text{table} \\
\text{WHERE} & \quad \text{condition}_1 \ \text{AND} \ \text{condition}_2 \ [\text{AND} \ldots]
\end{align*}
\]

To which Diffix responds with:

\[
\text{output} = \text{true count} + \text{static noise} + \text{dynamic noise}
\]

\[
\begin{align*}
\text{static noise} & \leftarrow \text{query syntax of } Q \\
\text{dynamic noise} & \leftarrow \text{query syntax and user set of } Q
\end{align*}
\]

Both noises are sticky: issuing the same query gives the same noise
Diffix’s noise mechanism: sticky noise

\[ Q = \text{count}( \text{age} = 40 \land \text{dept} = \text{Computing} \land \text{high-salary} = \text{True} ) \]

Each noise \( \sim N(0,1) \)

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

\[
\begin{align*}
Q1 &= \text{count}(\text{dept} = \text{Computing} \land \text{high-salary} = \text{True}) \\
Q2 &= \text{count}(\text{age} \neq 40 \land \text{dept} = \text{Computing} \land \text{high-salary} = \text{True})
\end{align*}
\]

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

\[ \text{Q1} = \text{count}(\text{dept} = \text{Computing} \land \text{high-salary} = \text{True}) \]
\[ \text{Q2} = \text{count}(\text{age} \neq 40 \land \text{dept} = \text{Computing} \land \text{high-salary} = \text{True}) \]

Output of Q1 – Q2 if high-salary = True

Bob: age = 40
depth = computing
high-salary = ? (unique)
Differential attack

Output of $Q_1 - Q_2$ if $\text{high-salary} = \text{False}$

$$Q_1 = \text{count}(\text{dept} = \text{Computing} \land \text{high-salary} = \text{True})$$

$$Q_2 = \text{count}(\text{age} \neq 40 \land \text{dept} = \text{Computing} \land \text{high-salary} = \text{True})$$

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

if high-salary = False

Q1 - Q2 ~ N(μ=0, σ=2)

if high-salary = True

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**

<table>
<thead>
<tr>
<th>age</th>
<th>dept</th>
<th>high-salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>computing</td>
<td>true</td>
</tr>
<tr>
<td>40</td>
<td>computing</td>
<td>true</td>
</tr>
<tr>
<td>34</td>
<td>math</td>
<td>false</td>
</tr>
<tr>
<td>34</td>
<td>math</td>
<td>true</td>
</tr>
</tbody>
</table>

**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.


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.

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/