Guaranteeing Differential Privacy using Quantitative Information Flow Analysis

Alec Miller

Master of Information Technology
COMP90054 Research Project Presentation

27/05/2024

Presentation Overview

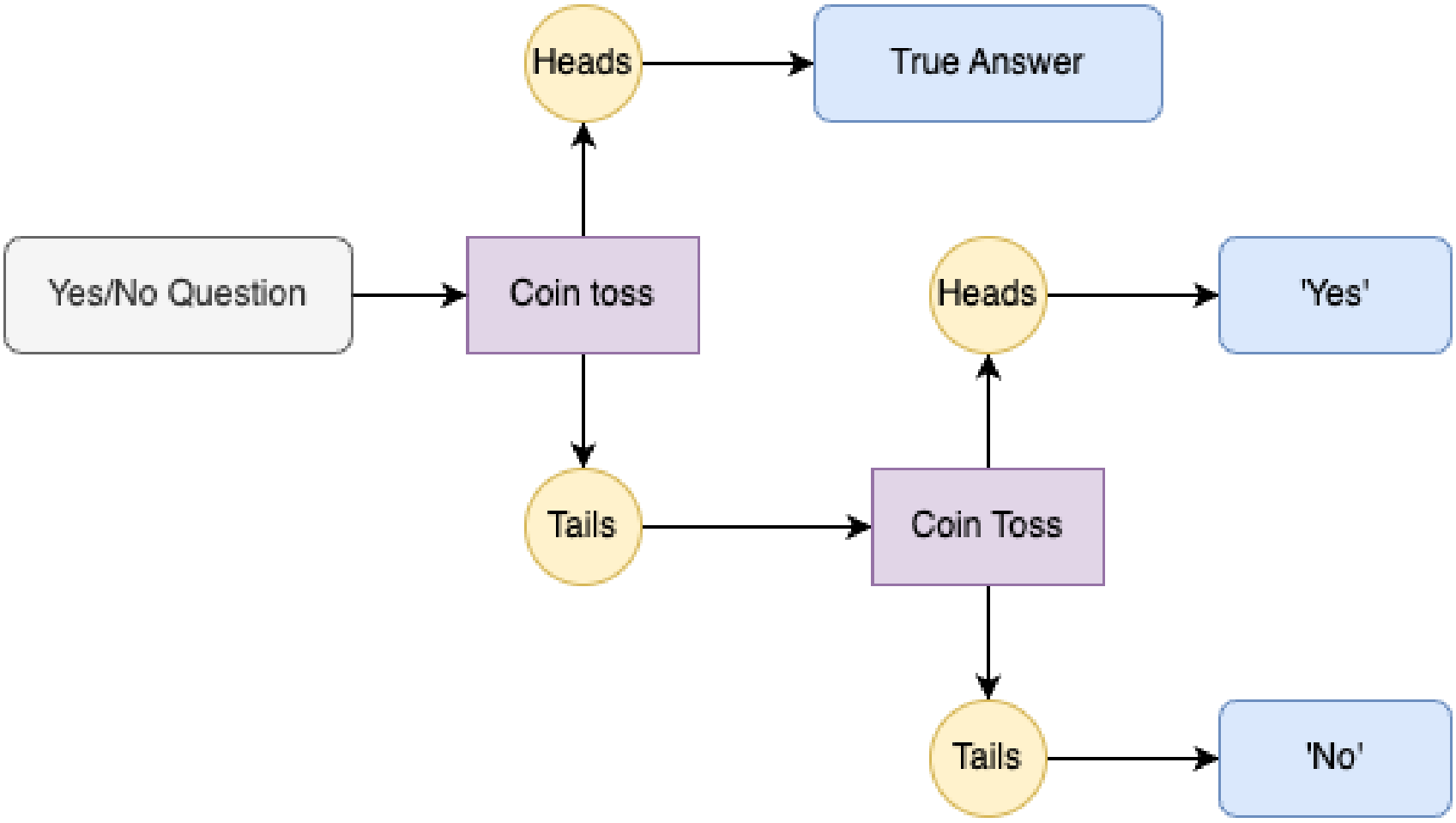
- Review of differential privacy, Quantitative Information Flow (QIF) and the relationship between them
- Explanation of extending differential privacy definitions in QIF
- Kuifje demonstration to experimentally evaluate differential privacy guarantees.

What is Differential Privacy?

- A rigorous framework for guaranteeing privacy in data, where privacy is introduced by adding random noise to data.
- Balances accurate analysis and protection of sensitive information.

What is Differential Privacy?

Example: Randomized Response Mechanism



What is Differential Privacy?

Formal Definition

Definition

A randomized algorithm $\mathcal{M}$ is (ϵ, δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all datasets D, D' differing by at most one element:

$$\Pr[\mathcal{M}(D) \in \mathcal{S}] \leq \exp(\epsilon)\Pr[\mathcal{M}(D') \in \mathcal{S}] + \delta$$

where the probability space is over the randomness of $\mathcal{M}$.

If $\delta = 0$, we say that $\mathcal{M}$ is ϵ-differentially private.

What is Quantitative Information Flow?

- A rigorous framework for modelling information leakage in programs.
- Assumes the presence of an adversary that has the goal of exploiting leaked information.
- The QIF paradigm models:
  - Secrets
  - Mechanisms (Channels)
  - Observables
  - Adversaries (Loss functions)

Relating Differential Privacy and QIF

- Natural link between information leakage in programs and the security of data querying a database
- Privacy mechanisms $\rightarrow$ channels, data values $\rightarrow$ secrets. Differential privacy guarantees can be calculated using a differential privacy loss function or by comparing channel rows.

Definition (ϵ-differential privacy of a channel)

A channel M is ϵ-differentially private for secret values x, x' if for all $y \in \mathcal{Y}$,

$$M_{xy} \leq e^{\epsilon} M_{x'y} \text{ and } M_{x'y} \leq e^{\epsilon} M_{xy}$$

Relating Differential Privacy and QIF

Random Response Mechanism Example

$$RRM := \begin{array}{c} \\ \texttt{resp=0} \\ \texttt{resp=1} \end{array} \begin{array}{c} \begin{array}{cc} \texttt{result=0} & \texttt{result=1} \end{array} \\ \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix} \end{array}$$

The smallest ϵ that satisfies $RRM_{xy} \leq e^{\epsilon} RRM_{x'y}$ and $RRM_{x'y} \leq e^{\epsilon} RRM_{xy}$ for all y (`result`) is $\epsilon = \ln(3)$.

Relating Differential Privacy and QIF

Extension to (ϵ, δ)-differential privacy

Definition ((ϵ, δ)-differential privacy of a channel)

A channel M is (ϵ, δ)-differentially private for secret values x, x' if and only if

$$\sum_{y \in \mathcal{S}} M_{xy} \leq e^{\epsilon} \sum_{y \in \mathcal{S}} M_{x'y} + \delta \text{ and } \sum_{y \in \mathcal{S}} M_{x'y} \leq e^{\epsilon} \sum_{y \in \mathcal{S}} M_{xy} + \delta$$

for all $\mathcal{S} \subseteq \mathcal{Y}$.

Using Kuifje to model Differential Privacy

Code:

```
// Database of 0's and 1's
database = [0,1,0,1,1,0,1,0,0];
// New response to add to database
resp <- uniform [0, 1];
// Coin toss
coin <- uniform [0, 1];

// Add data to database depending
// on coin and resp
new_data <- uniform [resp, coin];
database.append(new_data);

// Query the count of 1's in the
// database
count = 0;
for r in database:
        count = count + r;
        leak(count);
```

Output:

```
> Variable resp hyper
0.500000   0.250000   R 0.0
           0.750000   R 1.0
0.500000   0.750000   R 0.0
           0.250000   R 1.0
```

Left column = Outer distribution, probabilities of what the attacker observes (`count`)

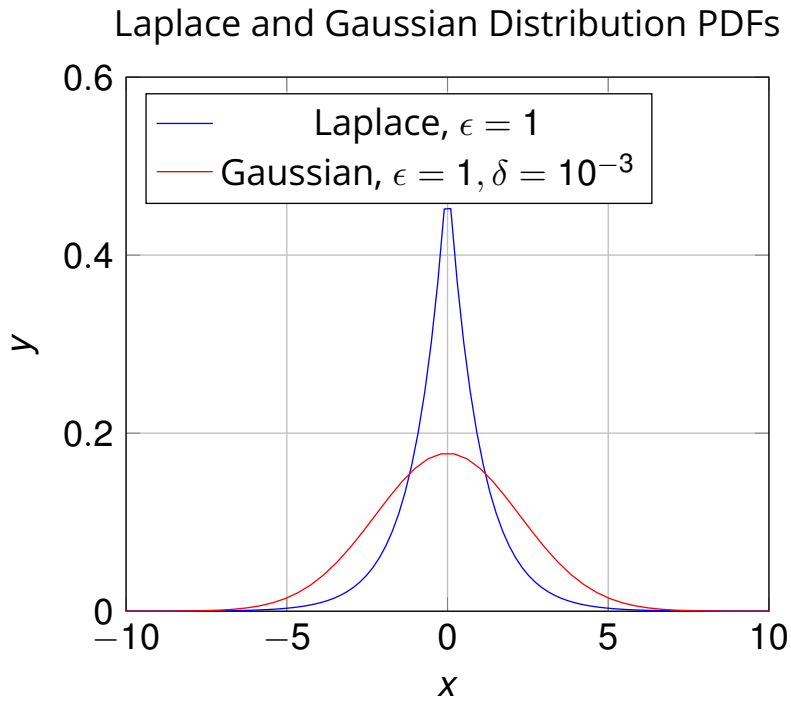
Middle column = Inner distribution, probabilities of the secret (`resp`) given the observation

Right column = Secret value

$$M_{xy} = \Pr[\text{observation}=y \mid \text{secret}=x]$$

$$= \frac{\Pr[\text{secret}=x \mid \text{observation}=y]\Pr[\text{observation}=y]}{\Pr[\text{secret}=x]}$$

The Laplace and Gaussian Mechanisms



$$M(X) = f(X) + Y\ ,\ Y \sim \text{Lap}(1/\epsilon) \text{ or } \text{N}(0, \sigma^2)$$

f is commonly a counting query, i.e. $f(X) =$ the number of entries in X satisfying a certain property.

In practice it is best to use the Discrete Laplace and Discrete Gaussian distributions, as well as it being convenient for our discussion on channel differential privacy.

Differential Privacy Results

Discrete Laplace Mechanism with parameter $\epsilon = \frac{1}{3}$ on a database with 30 0's, 33 1's and 1 unknown `resp`.

Sampling method adapted from (Canonne et al., 2020).

Range is limited to $[-33, 33]$ which introduces a δ.

We look at the set of observations which violate $\frac{1}{3}$-differential privacy.

```
0.000180    0.417233    R 0.0
            0.582767    R 1.0
0.001333    0.417413    R 0.0
            0.582587    R 1.0
0.009848    0.417427    R 0.0
            0.582573    R 1.0
0.000003    1.000000    R 1.0
```

$$\begin{array}{l} \texttt{r=0} \\ \texttt{r=1} \end{array} \begin{pmatrix} \cdots & 0.000150 & 0.001113 & 0.008222 & 0.000000 \\ \cdots & 0.000210 & 0.001553 & 0.011474 & 0.000006 \end{pmatrix}$$

With $\mathcal{S}$ being this set of observations, we have

$$\sum_{y \in \mathcal{S}} M_{xy} - e^{\frac{1}{3}} \sum_{y \in \mathcal{S}} M_{x'y} \approx 0.000007$$

The smallest δ for which we have $(\frac{1}{3}, \delta)$-differential privacy is $\delta \approx 0.000007$.

Conclusions

- Generalization of ϵ-differential privacy of a mechanism in QIF to (ϵ, δ)-differential privacy.
- Models of differential privacy mechanisms in Kuifje, including the well known discrete Laplace and Gaussian mechanisms.
- Demonstrated general method to verify or calculate the differential privacy guarantees using Kuifje, which should be able to be automated.

Questions