Private mouse and keyboard behavioral data

Mid-term presentation – Bachelor Thesis

Hossam Elfar
Supervisors: Guanhua Zhang and Mayar Elfares

June 16, 2023

Perceptual User Interfaces Group, University of Stuttgart
www.perceptualui.org

Table of Contents

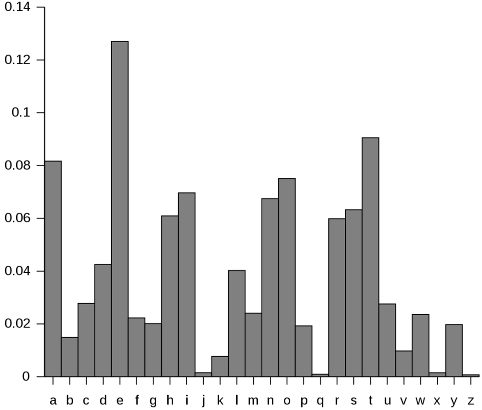
Recap

Motivation

Motivation

- Keyboard and mouse data contain highly sensitive data:
  - Passwords and login credentials
  - Personal messages and communications
  - Banking information

- These keyboard and mouse data are vulnerable to attacks that can potentially expose personal information about individuals in the dataset.
- E.g. Frequency analysis

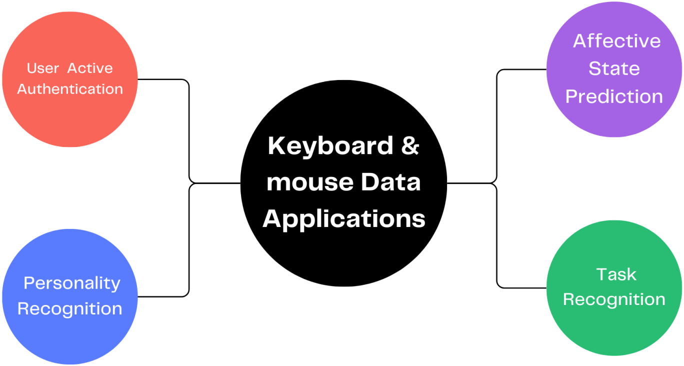


Motivation - Mouse attacks

- By analyzing the patterns, speed, and direction of mouse movements.
- Adversaries can infer:
  - User's activities, interests, or intentions.
  - User's interactions with applications and websites

.

Motivation - Applications



Challenge: Find a privacy-preserving mechanism that protects these sensitive datasets while maintaining their utility.

Recap

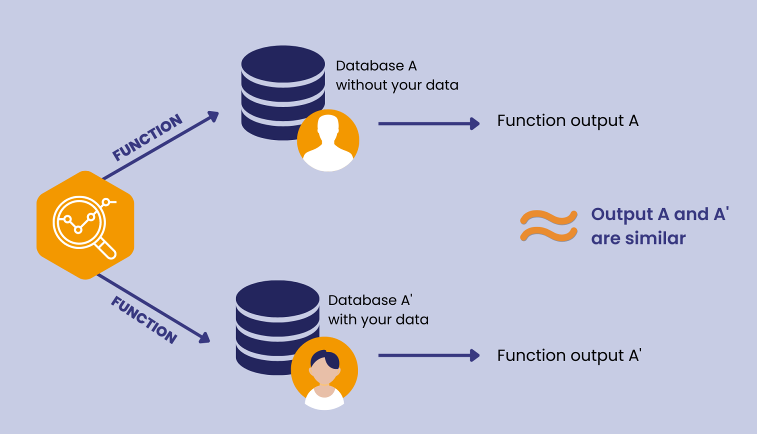
Preliminaries

Definition: Algorithm $\mathcal{M}$ with domain $\mathcal{D}$ satisfies ε-differential privacy if for all pairs of adjacent datasets D and D' that differ in the data of a single individual.

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

- ε: privacy loss (small ε = stronger privacy protection)
- The inequality ensures that the probability of obtaining an output S from dataset D is approximately the same as the probability of obtaining the same output S from a neighboring dataset D', up to a multiplicative factor of e^{ε}.

[1] The algorithmic foundations of differential privacy - Dwork et al. - 2014



Source: Statice

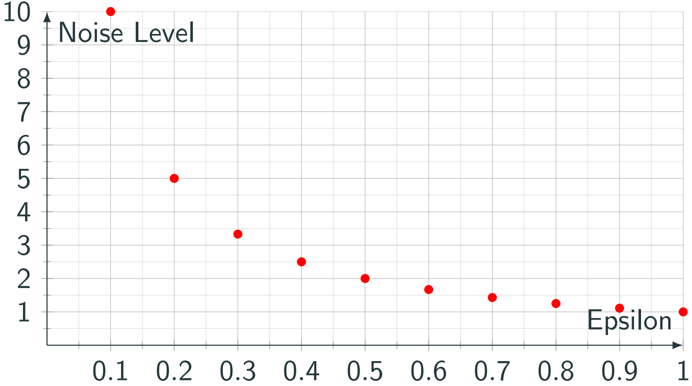
Laplace Mechanism

$$F(x) = f(x) + \mathrm{Lap}\left(\frac{s}{\varepsilon}\right)$$

- s: the sensitivity of the query.
- ε: the privacy loss.
- $\mathrm{Lap}(x)$: a sample from the Laplace distribution with scale parameter x.

Laplace Mechanism

- privacy vs utility trade-off



Source: Amount of Noise Added for Different Epsilon Values

Table of Contents

Progress

Dataset

Dataset

- Everyday Mouse And Keyboard Interactions dataset [2]

Name	EMAKI dataset
Users	39 users
Data	1.2M Mouse data, 210K Keyboard data
Tasks	Text Entry & Editing, Image Editing, Questionnaire Completion

[2] Exploring Natural Language Processing Methods for Interactive Behaviour Modelling - zhang et al. - 2023

Progress

Remote Data Science

- Domain server: manages the remote study of the data by a Data Scientist and allows the Data Owner to manage the data and control the privacy guarantees of the subjects under study.
- Data owner: provides mouse and keyboard datasets to make available for study by an outside party.
- Data scientist: end users who desire to perform computations or answer a specific question using one or more data owners' datasets.

Source: Remote Data Science

Data Owner

- Deploy a Domain Server
- Upload Private Data
- Manage Privacy Budget

Data Scientist

- Connect to a Domain
- Search for Datasets
- Analyse Data
- Retrieve Secure Results

Remote Data Science - Steps

- Launch domain node.
- Preprocessing of the data.
- Upload datasets to the domain node.
- Create a data scientist account with an initial privacy budget.

- Data scientist view the available datasets of the node.
- Select one of the datasets (Mouse or keyboard).
- Perform a query with noise.
- Review code and approve.
- Data scientist download secure results.

Remote Data Science - Steps

[46]:
```
datasets
```

[46]:

Dataset List

id	name	url
70ad...cbd	mouse data	https://github.com/OpenMined/datasets/tree/main/trade_flow
6038...96a	keyboard data	https://github.com/OpenMined/datasets/tree/main/trade_flow

•[44]:
```
mouse_dataset = datasets[0]
```

•[45]:
```
mouse_dataset
```

[45]:

mouse data

mouse data

Uploaded by: Hossam Elfar

Created on: 2023-06-15 16:36:40

URL: https://github.com/OpenMined/datasets/tree/main/trade_flow

Contributors: to see full details call dataset.contributors

Asset List

id	name	shape
deae...39d	mouse_data	(1000, 13)

Source: Remote Data Science

Remote Data Science - Steps

•[47]:

```
mock = mouse_dataset.assets[0].mock
mock
```

[47]:

	user	session	task	timestamp	X	Y	resolutionX	resolutionY	O	C	E	A	N
2552	1	1	3	1616056500873	0.907339	0.511719	1349	768	0	1	1	1	1
2553	1	1	3	1616056500889	0.932543	0.501302	1349	768	0	1	1	1	1
2554	1	1	3	1616056500906	0.963677	0.492188	1349	768	0	1	1	1	1
2555	1	1	3	1616056500922	0.998517	0.480469	1349	768	0	1	1	1	1
2556	1	1	3	1616056500940	1.011861	0.523438	1349	768	0	1	1	1	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
5592	1	1	3	1616057025996	0.598962	0.467448	1349	768	0	1	1	1	1
5593	1	1	3	1616057026013	0.607858	0.466146	1349	768	0	1	1	1	1
5594	1	1	3	1616057026030	0.616753	0.466146	1349	768	0	1	1	1	1
5595	1	1	3	1616057026047	0.624166	0.464844	1349	768	0	1	1	1	1
5596	1	1	3	1616057026063	0.631579	0.464844	1349	768	0	1	1	1	1

Source: Remote Data Science

Table of Contents

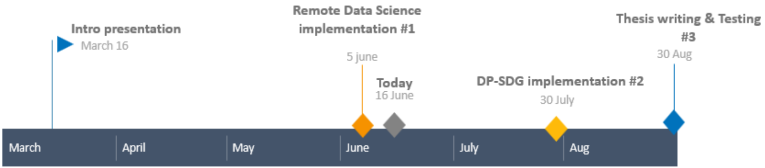
Next steps

- Remote machine learning - DP-SGD[3]
- Thesis writing

[3] Deep learning with differential privacy - Abadi et al. - 2016

Table of Contents

Schedule


Intro presentation
March 16
Remote Data Science implementation #1
5 june
Today
16 June
DP-SDG implementation #2
30 July
Thesis writing & Testing #3
30 Aug
March
April
May
June
July
Aug

Thank you!

Questions?