PUI Perceptual User Interfaces

University of Stuttgart Germany

Private mouse and keyboard behavioral data

BSc intro

Hossam Elfar

June 16, 2023

Perceptual User Interfaces Group, University of Stuttgart

www.perceptualui.org

Table of Contents

Who am I ?

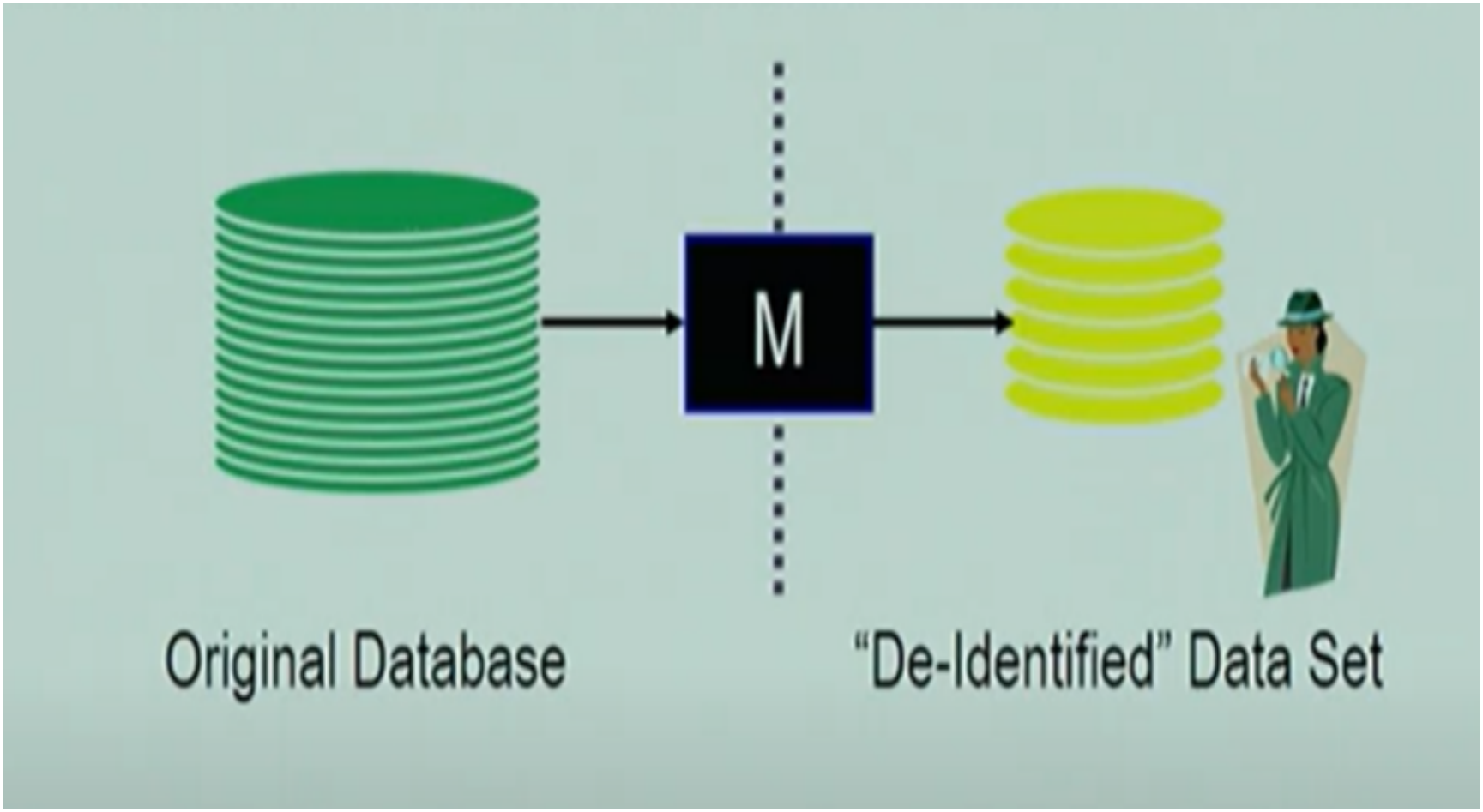
- Hossam Elfar
- Student at the German University in Cairo (GUC).
- Media Engineering and technology (MET).
- hobbies : Playing Chess , Swimming

Table of Contents

Behavioral datasets

Mouse and Keyboard dataset

- Effective behavioural biometrics for data analysis [Hanisch et al., 2021]
- Highly personally sensitive data such as usernames, passwords, banking information, or text messages
- Used as biometrics in active authentication and predicting user's intents [Sun and Upadhyaya, 2015]
- Limited prior work in privacy [Sánchez et al., 2020]


M
Original Database
"De-Identified" Data Set

Netflix Prize

But are these de-identified datasets actually private?

Why anonymisation isn't enough ?

Netflix Prize & Linkage attacks

Source: image from Simply Explained youtube channel

Table of Contents

Differential privacy

Differential privacy & remote data science

Our key contributions are two-fold:

We propose the first differential privacy [Dwork, 2008] approach for mouse and keyboard datasets and allow remote processing of mouse and keyboard behavioural data by performing remote data analysis.

(a) Mouse dataset of 20 users

Position : X >= 50 , Y >=30

Does John actually clicks on that position on the screen ?

Probability of people clicking on the position

Dataset → Model → 0.55

Dataset +john → Model → 0.57

Maybe john has clicked ?

Plausible deniability

(a) Mouse dataset of 20 users

Position : X >= 50 , Y >=30

Does John actually clicks on that position on the screen ?

Probability of people clicking on the position

Dataset → Model → 0.55

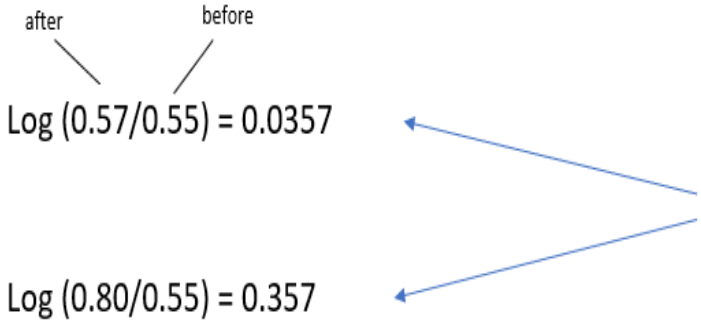
Dataset → Model → ~~0.57~~ 0.88

+john

john has clicked !

Privacy leak !

Privacy Loss



Differential privacy : Aims to limit this privacy loss!

- Log (probability after / probability before) <= **ε**
  Privacy budget
- Smaller budget = more private >>>> Larger budget = Learn more

Remote data science

- Deploy a domain node using HAGrid
- Deploy a network node that collects data from different domain nodes and handles the network requests using PySyft and PyGrid
- Allowing data owners to upload datasets to domain nodes.
- Obfuscating the data once uploaded via differential privacy
- Allow data scientists to log into the network, get a privacy budget, and run machine learning models.

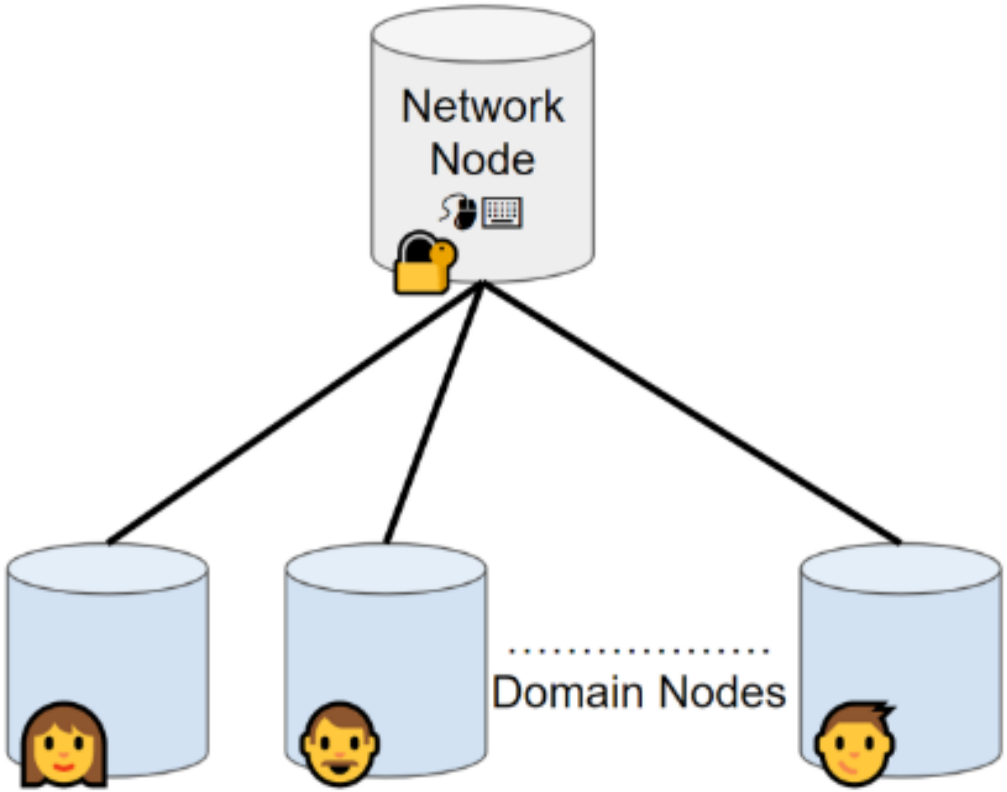
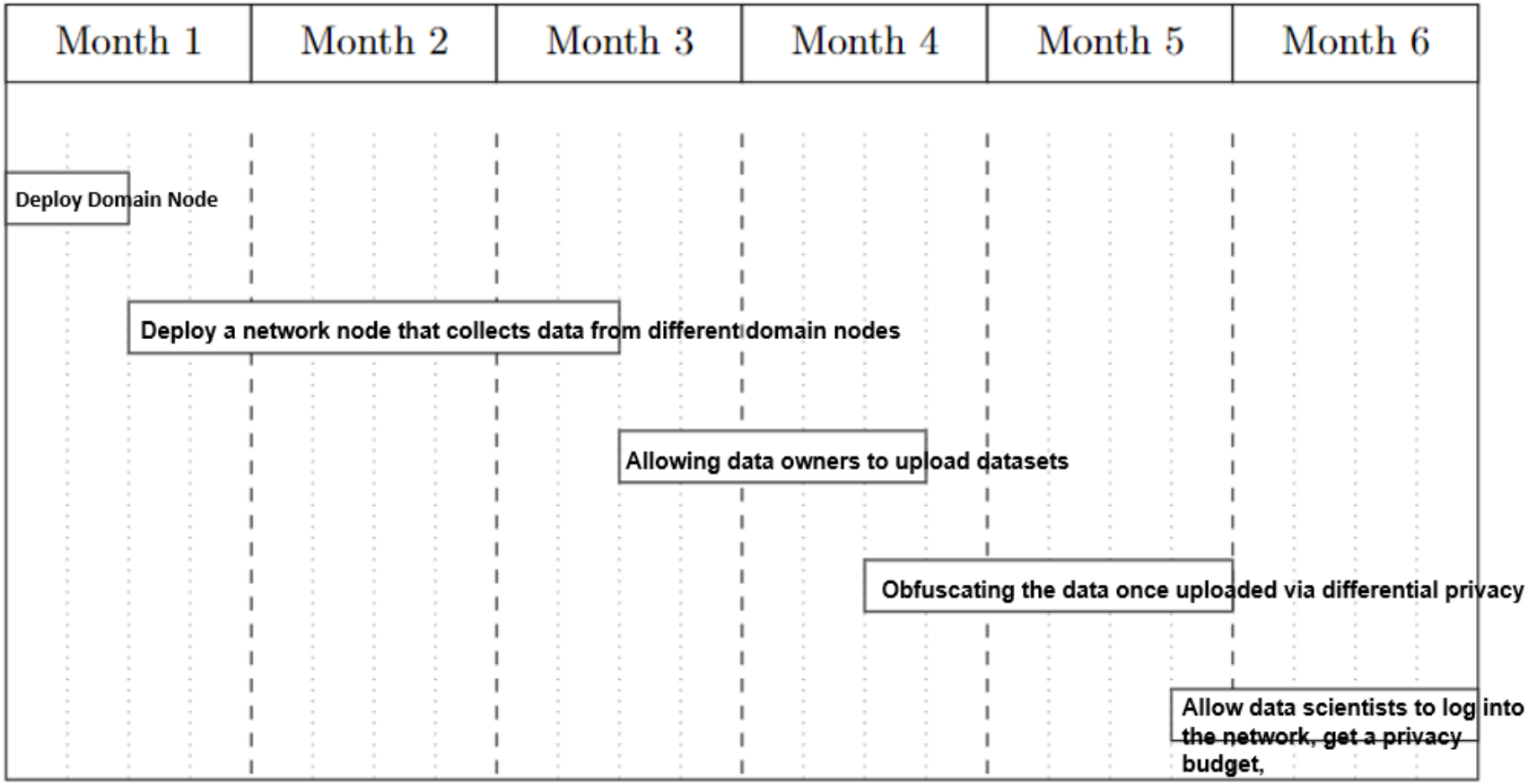

Network
Node
Domain Nodes

Table of Contents

Schedule with milestones



References i

C. Dwork. Differential privacy: A survey of results. In *International conference on theory and applications of models of computation*, pages 1–19. Springer, 2008.

S. Hanisch, P. Arias-Cabarcos, J. Parra-Arnau, and T. Strufe. Privacy-protecting techniques for behavioral data: A survey. *arXiv preprint arXiv:2109.04120*, 2021.

P. M. S. Sánchez, J. M. J. Valero, M. Zago, A. H. Celdrán, L. F. Maimó, E. L. Bernal, S. L. Bernal, J. M. Valverde, P. Nespoli, J. P. Galindo, et al. Behacom-a dataset modelling users' behaviour in computers. *Data in brief*, 31:105767, 2020.

Y. Sun and S. Upadhyaya. Secure and privacy preserving data processing support for active authentication. *Information Systems Frontiers*, 17(5):1007–1015, 2015.