



OpenMined

WOMEN of OM

Study Group

Intro to Differential Privacy

by Zumurut Muftuoglu

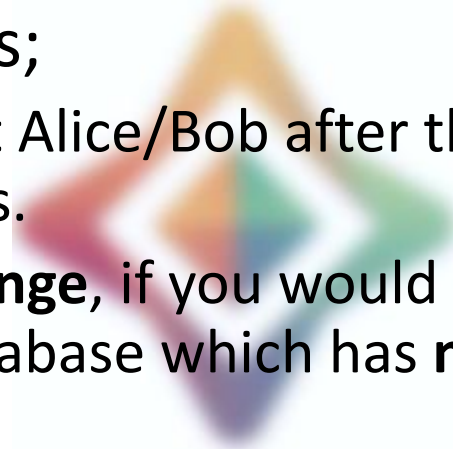
12th April, 2020

Privacy

- Privacy enables us to create boundaries and protect ourselves from unwarranted interference in our lives, allowing us to negotiate who we are and how we want to interact with the world around us.
- Privacy is a fundamental human right. The right to privacy is articulated in all of the major international and regional human rights instruments.
- Privacy ***is not** hiding «personally identifiable information»* or releasing only «aggregate information».

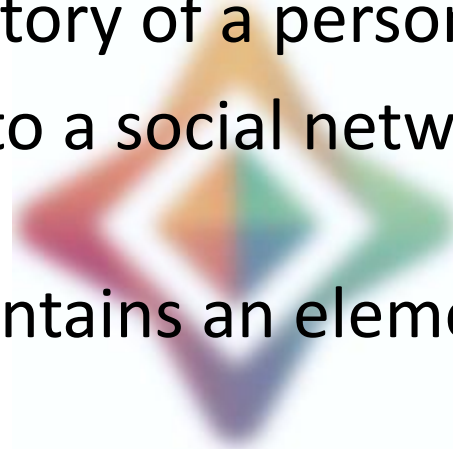
So what is success when it comes to privacy?..

- Suppose that you analyze a dataset D.
- At the end of your analysis;
 - If you know **no more** about Alice/Bob after the analysis than you know about her/him before the analysis.
 - And the result **is never change**, if you would have conducted the same analysis on an identical database which has **no record** about Alice/Bob.



Types and Structures of Data

- ***Location Data***-location history of a person,
- ***Graph Data***-data relating to a social network,communication network or physical network,
- ***Time Series***-Data which contains an element of updating in time, such as census information.



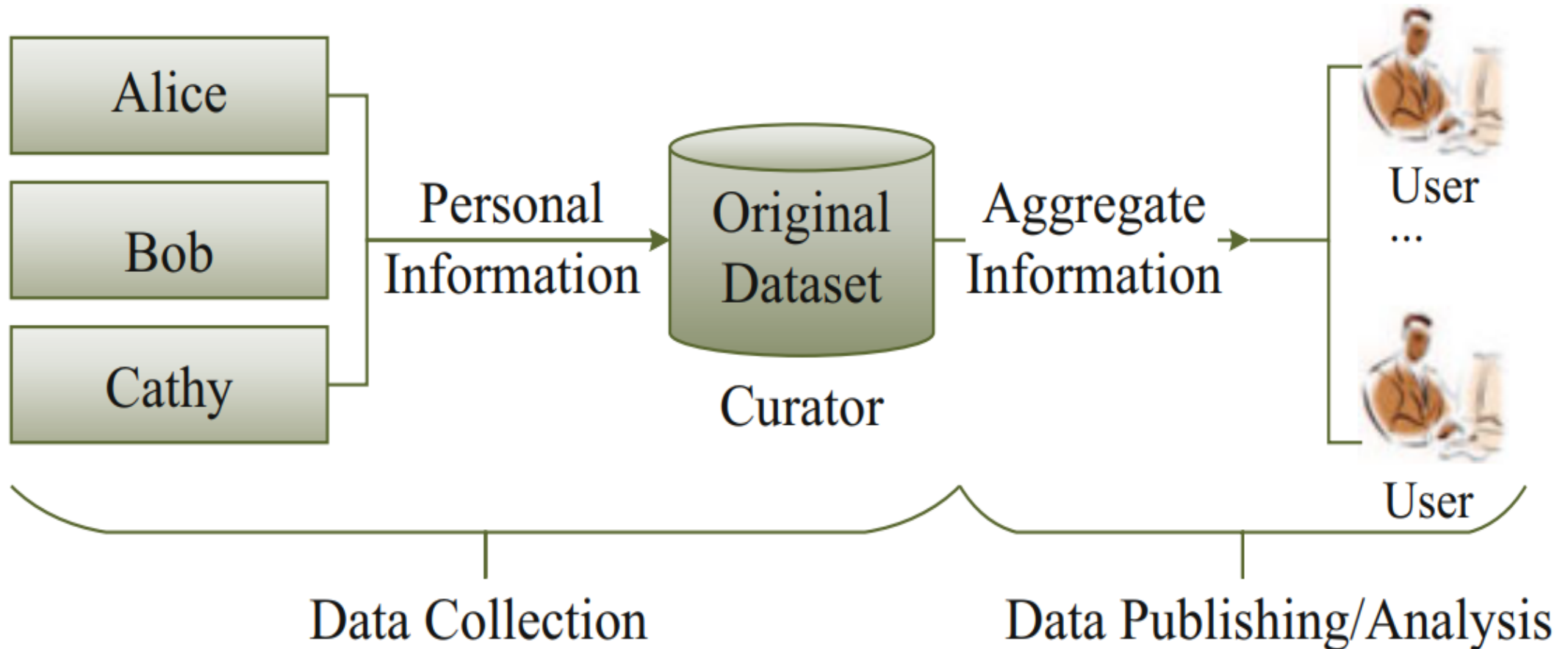
Privacy Failures in History

- Massachusetts Health Records(1990s)
- AOL Search Logs(2006)
- Netflix Prize(2006)
- Facebook Ads(2010)
- New York City Taxi Trips(2014)





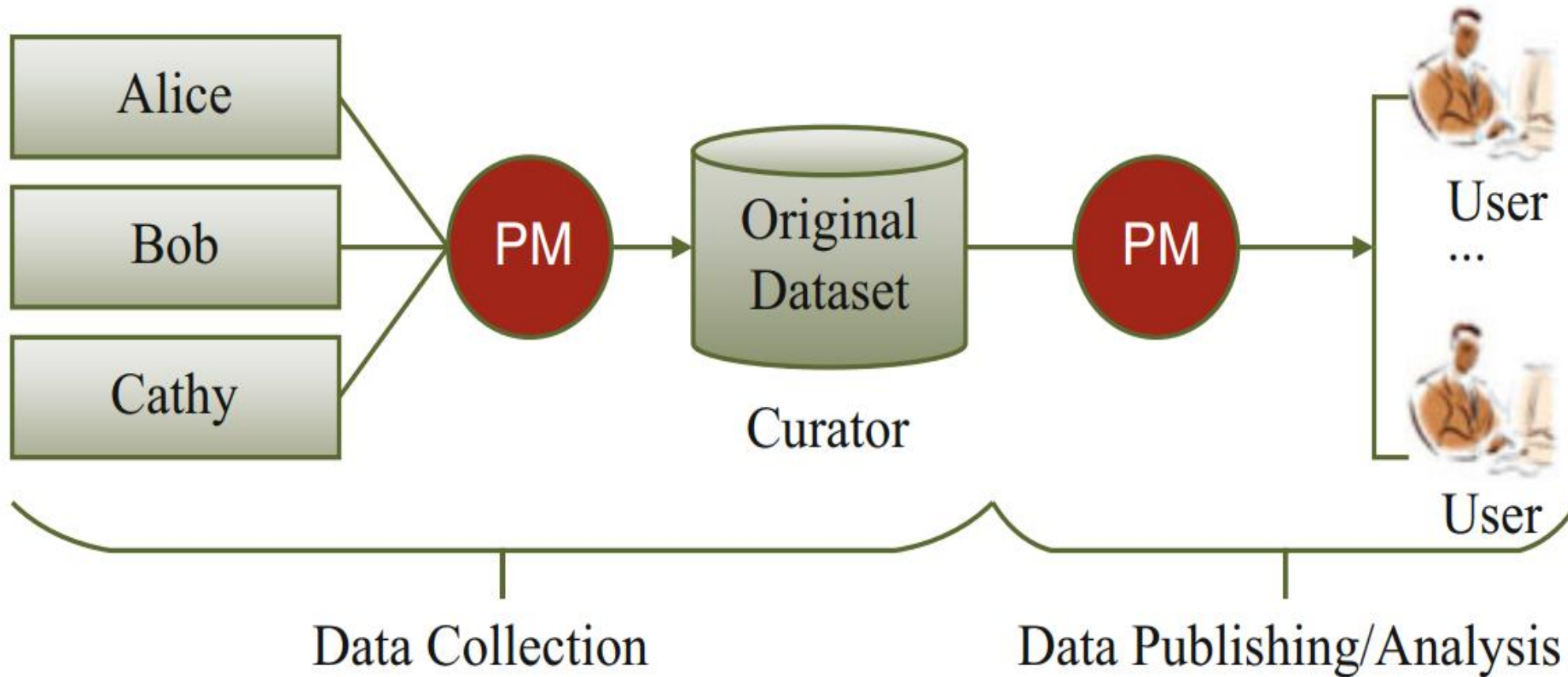
Privacy Preserving Data Publishing and Analysis



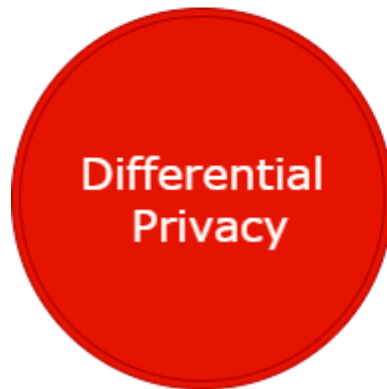
Data Publishing vs Data Analysis

	Differentially private data publishing	Differentially private data analysis
Mechanism	Independent mechanism	Coupled with a particular algorithm
Input	Various data types	Transaction dataset (training samples)
Output	Query answers or datasets	Various models

Privacy Model



Privacy Models

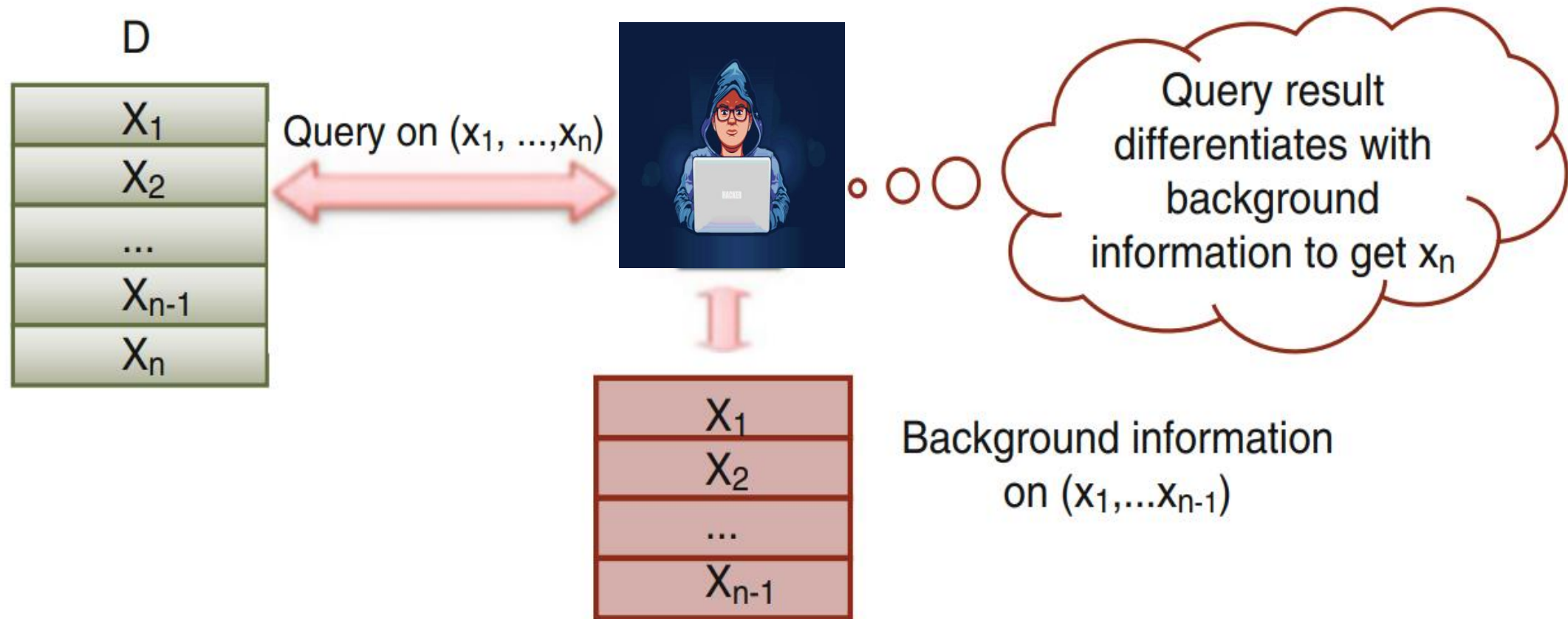


Let's begin...

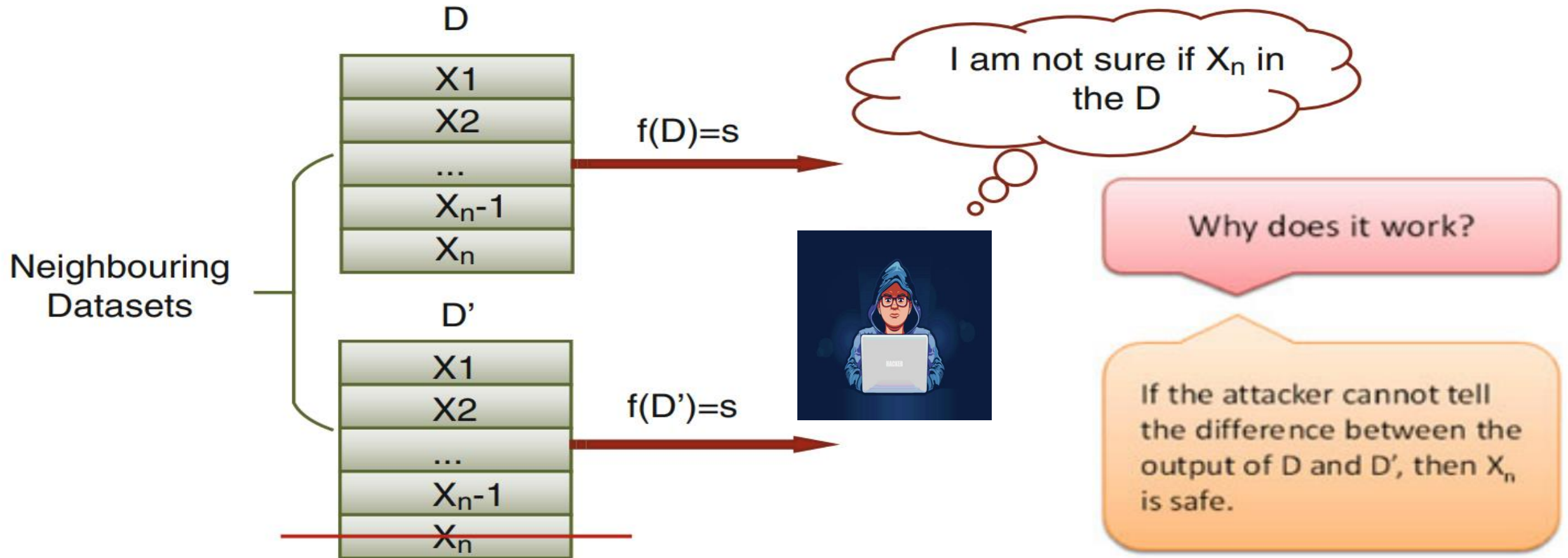
- Notation
- Definition Differential Privacy
- Theoretical View



Attacker Model



Differential Privacy



Notations

Notations	Explanation	Notations	Explanation
\mathcal{X}	Universe	D	Dataset; training sample set
D'	Neighboring dataset	\mathcal{D}	Dataset distribution
r, x	Record in dataset; training sample	d	Dataset dimension
n	The size of dataset	N	The size of a histogram
f	Query	F	Query set
m	The number of queries in F	M	Mechanism
\hat{f}	Noisy output	k	Represent some small value of constant
ϵ	Privacy budget	Δf	Sensitivity
G	Graph data	t, T	Time, time sequence, or iterative round
\mathbf{w}	Output model, or weight	$VC(\cdot)$	VC dimension
$\ell(\cdot)$	Loss function	α, β, δ	Accuracy parameter

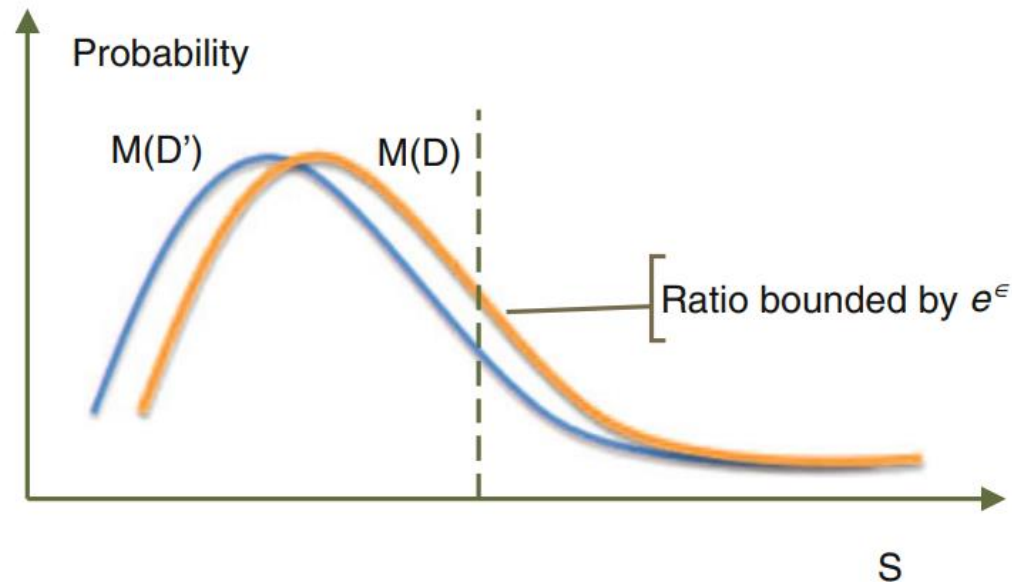
Formal Definition of Differential Privacy

- A randomized mechanism M gives (ϵ, δ) -differential privacy for every set of outputs S , and for any neighbouring datasets of D and D' , if M satisfies:

$$Pr[M(D) \in S] \leq \exp(\epsilon) \cdot Pr[M(D') \in S] + \delta.$$

The Privacy Budget(ϵ)

- Parameter ϵ is defined as the privacy budget ,which controls the privacy guarantee level of mechanism M .
- A smaller ϵ represents a stronger privacy



The Sensitivity

- Sensitivity determines how much perturbation is required in the mechanism.
- Two types of sensitivity are employed in differential privacy:
 - the global sensitivity
 - the local sensitivity.



The Global Sensitivity

- The global sensitivity is only related to the type of query f .

$$\Delta f_{GS} = \max_{D, D'} ||f(D) - f(D')||_1$$

The Local Sensitivity

- Local sensitivity calibrates the record-based difference between query results on neighboring datasets.

$$\Delta f_{LS} = \max_{D'} ||f(D) - f(D')||_1.$$


Differential Privacy Mechanisms

- Currently, three basic mechanisms are widely used to guarantee differential privacy:
 - The Laplace Mechanism,
 - The Gaussian Mechanism,
 - The Exponential Mechanism




Differential Privacy Mechanisms

- The Laplace Mechanism


$$\mathcal{M}(D) = f(D) + \text{Lap}\left(\frac{\Delta f}{\epsilon}\right)$$

- The Gaussssian Mechanism



$$M(D) = f(D) + \mathcal{N}(0, \sigma^2).$$

- The Exponential Mechanism

$$\mathcal{M}(D) = \left\{ \text{return } r \text{ with the probability } \propto \exp\left(\frac{\epsilon q(D, r)}{2\Delta q}\right) \right\}$$

Mechanism Example

- Suppose we have medial records



Name	Job	Gender	Age	Disease
Alen	Engineer	Male	25	Flu
Bob	Engineer	Male	29	HIV
Cathy	Lawyer	Female	35	Hepatitis
David	Writer	Male	41	HIV
Emily	Writer	Female	56	Diabetes
...
Emma	Dancer	Female	21	Flu

Medical record privacy mechanism output

- Suppose we have medical records

Options	Number of people	$\epsilon = 0$	$\epsilon = 0.1$	$\epsilon = 1$
Diabetes	24	0.25	0.32	0.12
Hepatitis	8	0.25	0.15	4×10^{-5}
Flu	28	0.25	0.40	0.88
HIV	5	0.25	0.13	8.9×10^{-6}

Thank you! 😊

