Introduction to Privacy-preserving Techniques for AI

Tu Hoang (contact@tuhoang.me, anhtu.hoang@uninsubria.it)

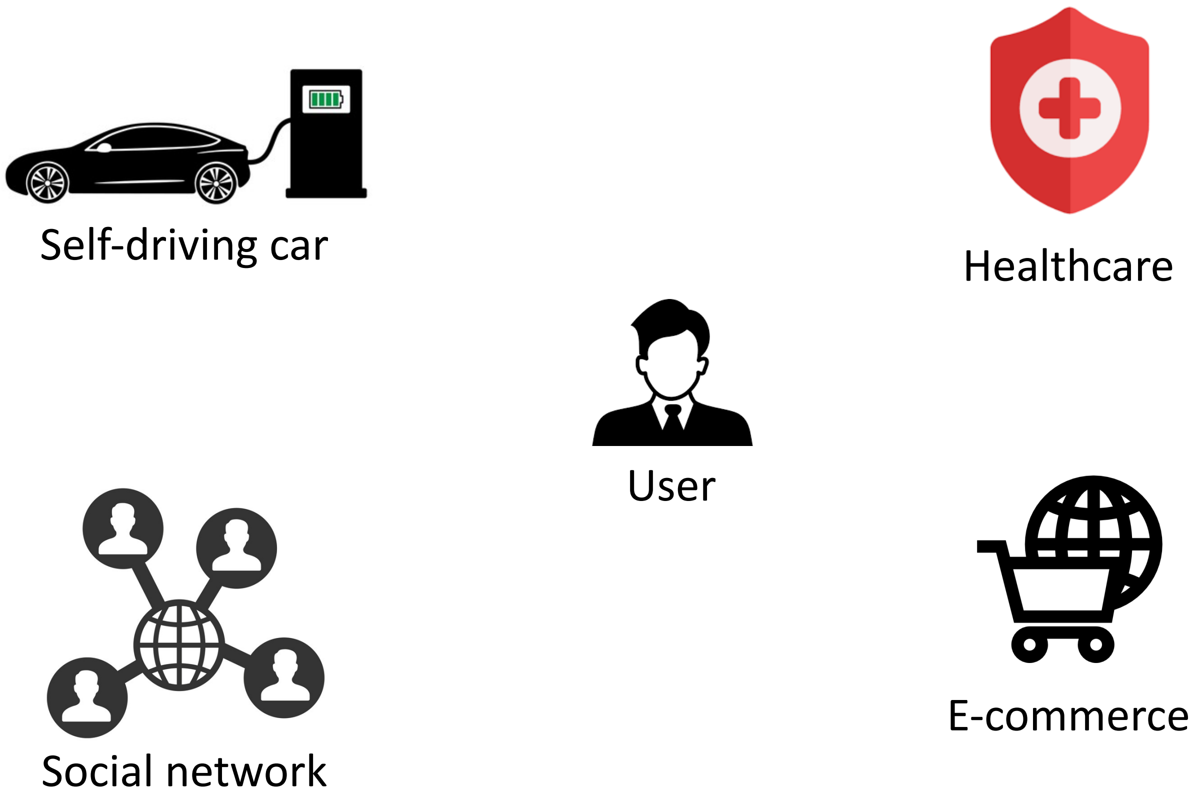
Postdoctoral Researcher

DiSTA, University of Insubria, Italy

Website: tuhoang.me

May 30th, 2023

AI and Data Protection



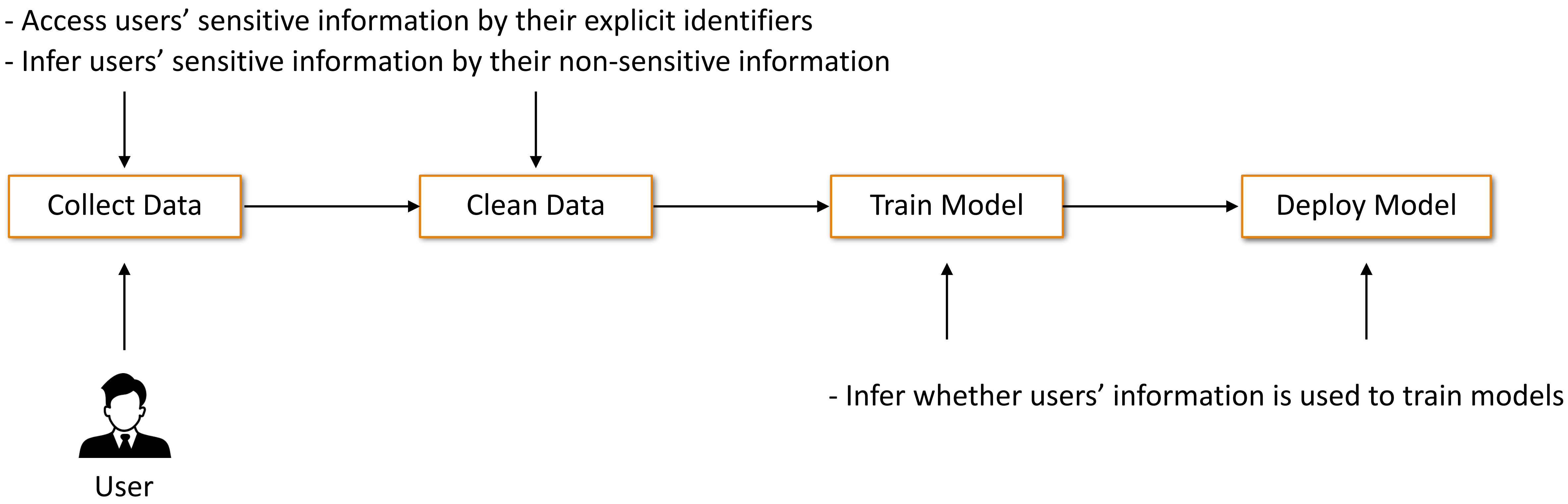
Users' Sensitive Information Leakage:

- Netflix $1-million contest (2010)
- Facebook Cambridge Analytica (2015-2018)
- ...

Data Protection Regulations:

- Europe: GDPR (2016)
- USA:
  + HIPAA (healthcare providers) (2022),
  + GLBA (financial institutes) (2022),
  + FISMA (federal agencies) (2022),
  + CCPA (California residents) (2020).

AI Workflow & Privacy Issues



Privacy-preserving Techniques for AI

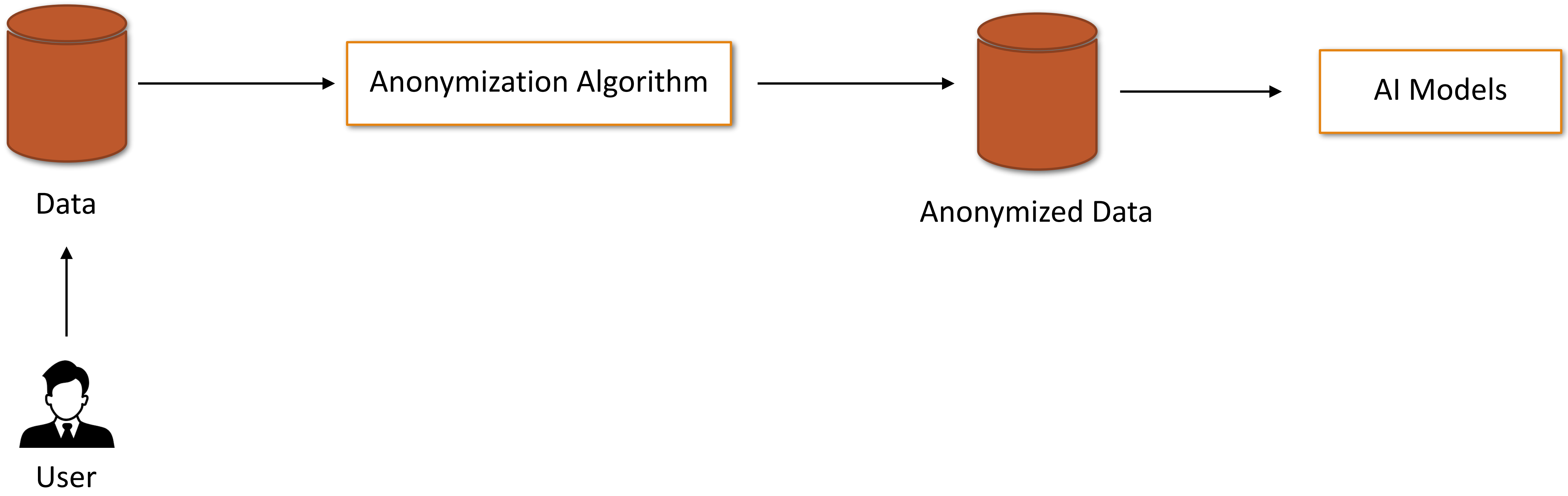
- ❖k-Anonymity
- ❖Differential Privacy
- ❖Homomorphic Cryptography
- ❖Distributed Learning

k-Anonymity

Attributes' Types

Key Attribute	Quasi-identifier			Sensitive attribute
Name	DOB	Gender	Zipcode	Disease
Andre	1/21/76	Male	53715	Heart Disease
Beth	4/13/86	Female	53715	Hepatitis
Carol	2/28/76	Male	53703	Brochitis
Dan	1/21/76	Male	53703	Broken Arm
Ellen	4/13/86	Female	53706	Flu
Eric	2/28/76	Female	53706	Hang Nail

k-Anonymity Workflow



k-Anonymity Protection

- Assume attackers' background knowledge
- Ensure that by using the knowledge, the confidence of inferring users' sensitive information is at least 1/k

Released table

	Race	Birth	Gender	ZIP	Problem
t1	Black	1965	m	0214*	short breath
t2	Black	1965	m	0214*	chest pain
t3	Black	1965	f	0213*	hypertension
t4	Black	1965	f	0213*	hypertension
t5	Black	1964	f	0213*	obesity
t6	Black	1964	f	0213*	chest pain
t7	White	1964	m	0213*	chest pain
t8	White	1964	m	0213*	obesity
t9	White	1964	m	0213*	short breath
t10	White	1967	m	0213*	chest pain
t11	White	1967	m	0213*	chest pain

External data Source

Name	Birth	Gender	ZIP	Race
Andre	1964	m	02135	White
Beth	1964	f	55410	Black
Carol	1964	f	90210	White
Dan	1967	m	02174	White
Ellen	1968	f	02237	White

Attribute Linkgage Protection

Homogeneity attack

Bob	
Zipcode	***Age***
47678	27

Background knowledge attack

Carl	
Zipcode	***Age***
47673	36

A 3-anonymous patient table

Zipcode	Age	Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
4790*	≥40	Flu
4790*	≥40	Heart Disease
4790*	≥40	Cancer
476**	3*	Heart Disease
476**	3*	Cancer
476**	3*	Cancer

Attribute Linkage Protection (l-Diversity)

Bob	
Zip	***Age***
47678	27

A 3-diverse patient table

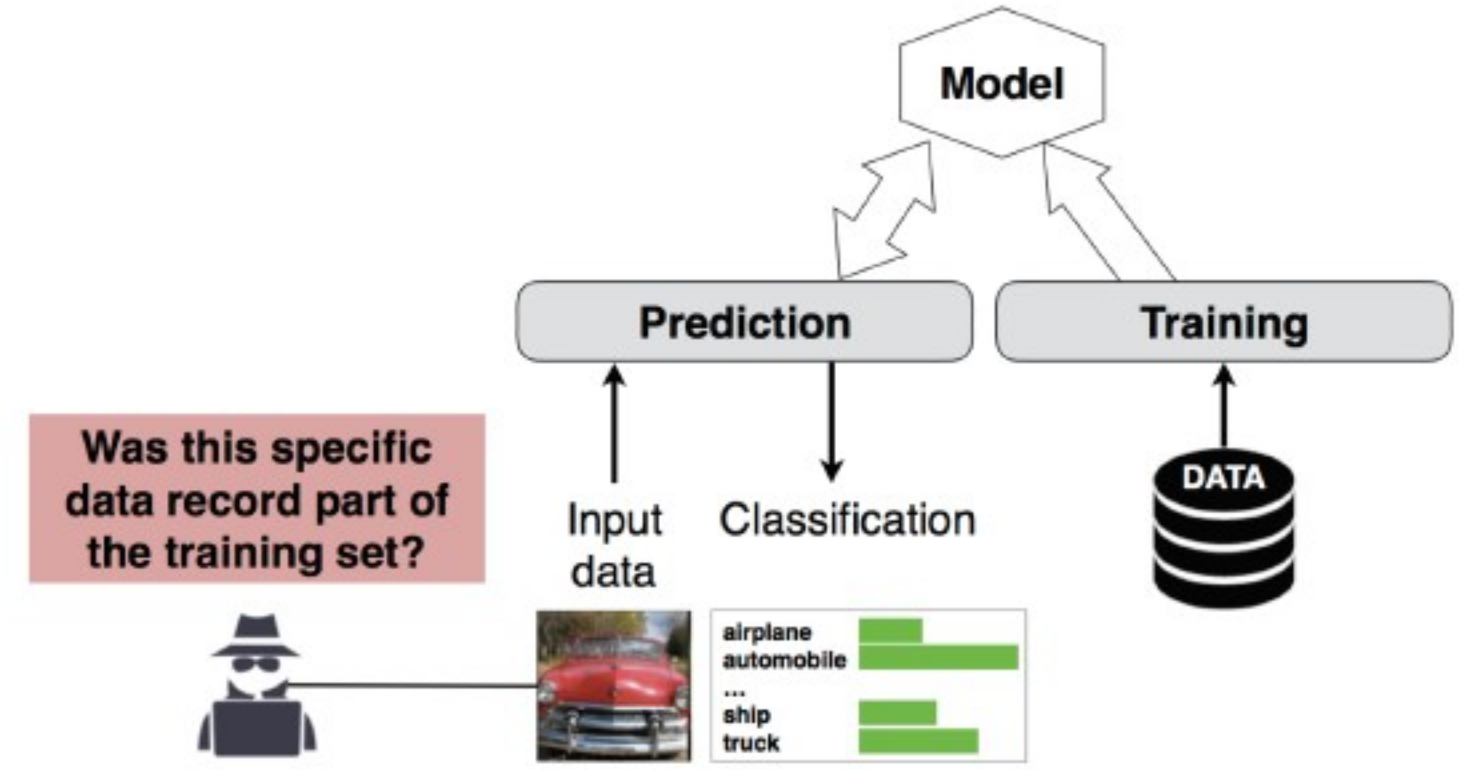
Zipcode	Age	Salary	Disease
476**	2*	20K	Gastric Ulcer
476**	2*	30K	Gastritis
476**	2*	40K	Stomach Cancer
4790*	≥40	50K	Gastritis
4790*	≥40	100K	Flu
4790*	≥40	70K	Bronchitis
476**	3*	60K	Bronchitis
476**	3*	80K	Pneumonia
476**	3*	90K	Stomach Cancer

Challenges

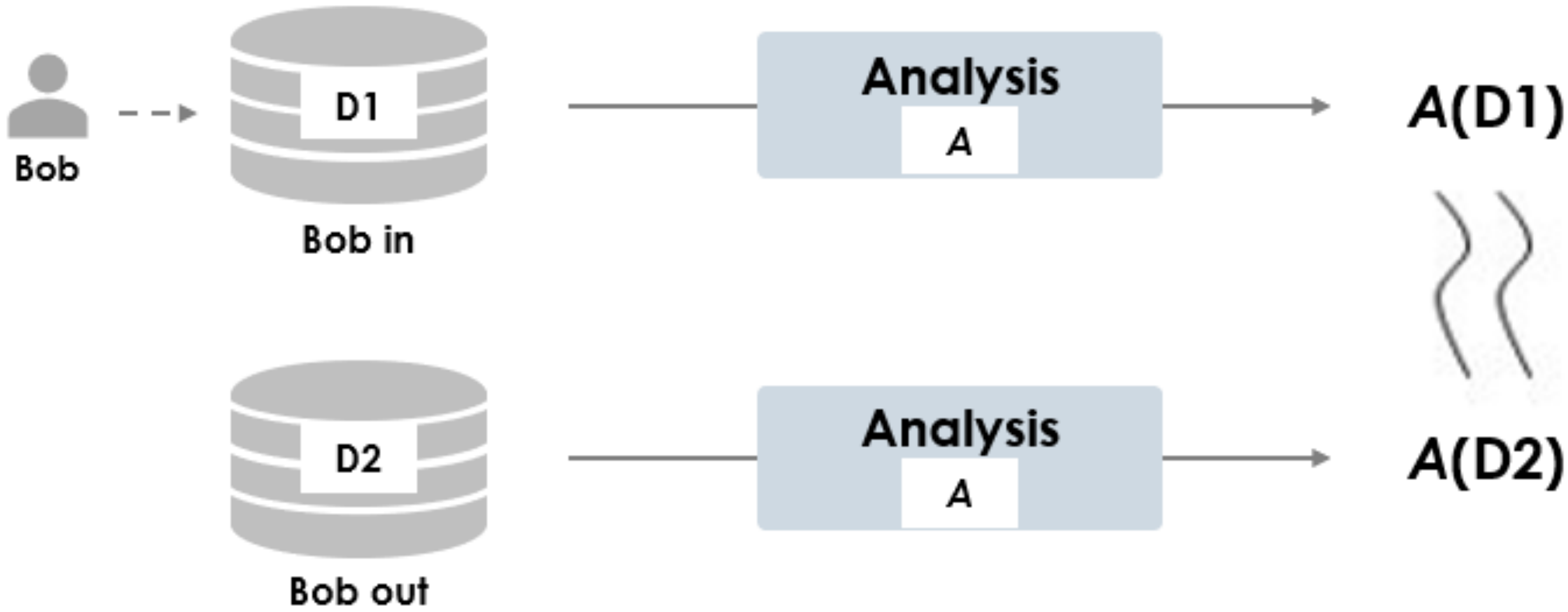
- ❖Assume attackers' background knowledge
- ❖Design optimization algorithms to generate anonymized data:
  - maximize anonymized data's quality
  - maximize performance
- ❖Support other data types:
  - Relational data
  - Text
  - Knowledge graphs

Differential Privacy

Membership Attacks



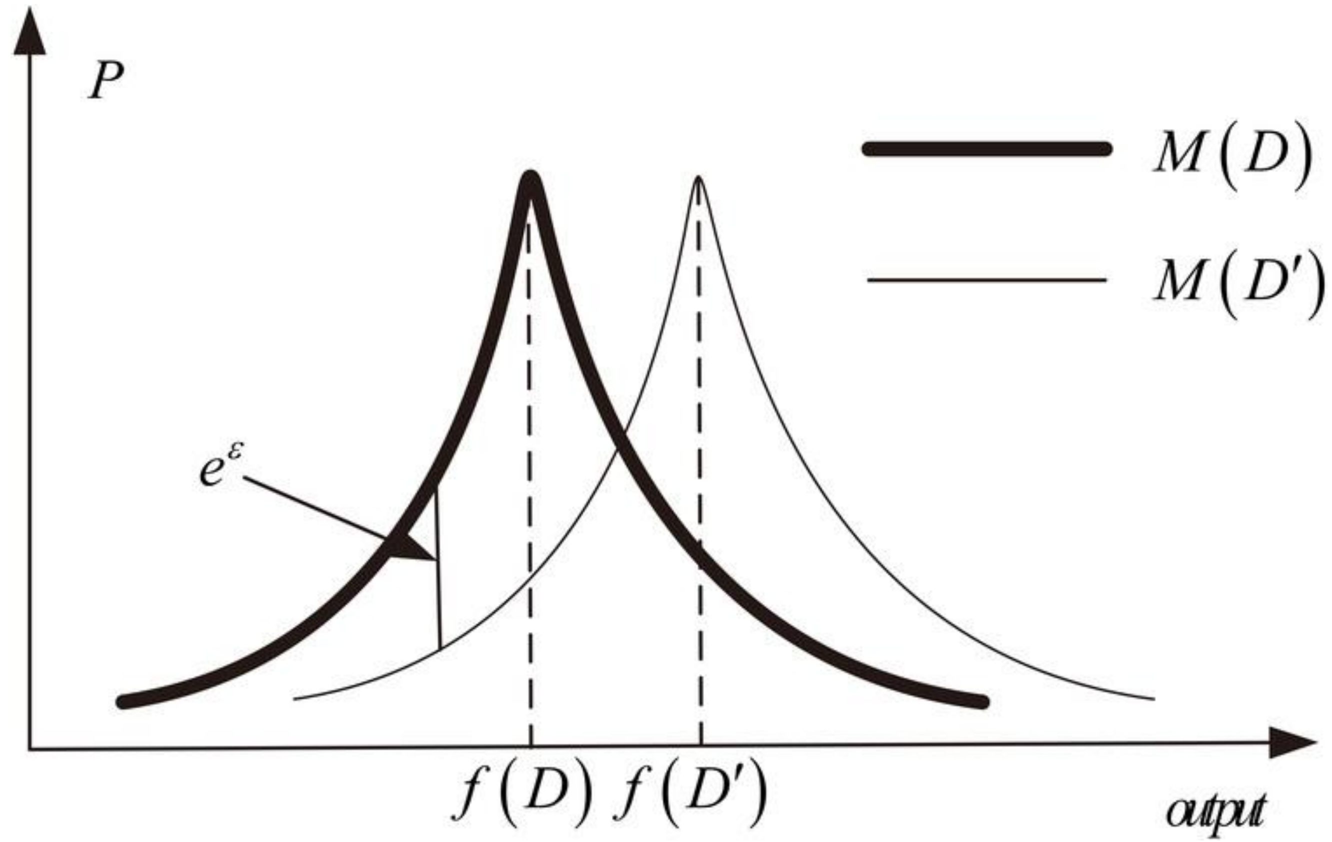
Differential Privacy Protection



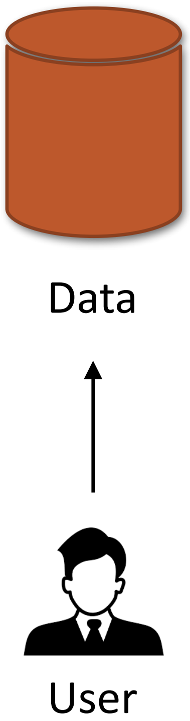
Definition

$$\frac{Pr[\mathcal{M}(x) \in S]}{Pr[\mathcal{M}(x') \in S]} \leq e^{\epsilon}$$

$$ln\left(\frac{Pr[\mathcal{M}(x) \in S]}{Pr[\mathcal{M}(x') \in S]}\right) \leq \epsilon$$

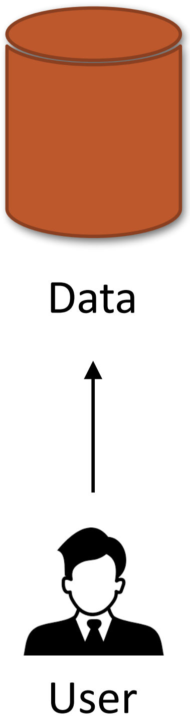


Stochastic Gradient Descent (**SGD**)



1. Initialize model
2. While epoch < num_epoches
    2.1. Get a batch of data
    2.2. Predict the output of the batch on the current model
    2.3. Calculate the gradients
    2.4. Update the current model with the gradients
3. Return model

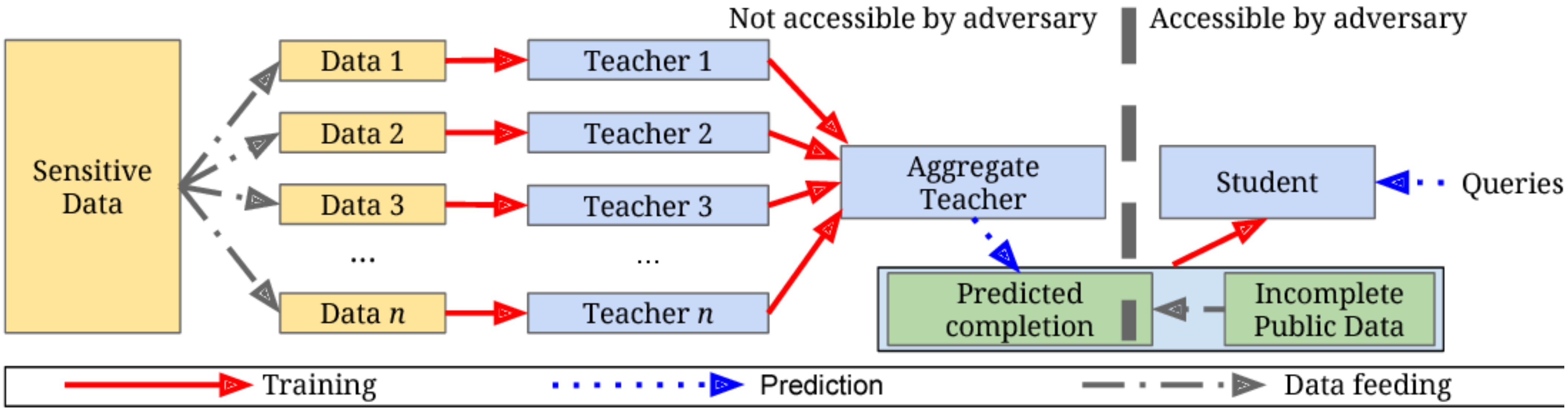
DP-SGD



1. Initialize model
2. While epoch < num_epoches
   2.1. Get a batch of data
   2.2. Predict the output of the batch on the current model
   2.3. Calculate the gradients
   2.4. Calculate noises
   2.5. Update the current model with the gradients and noises
3. Return model

Cannot infer whether a data point is used to train the model

PATE



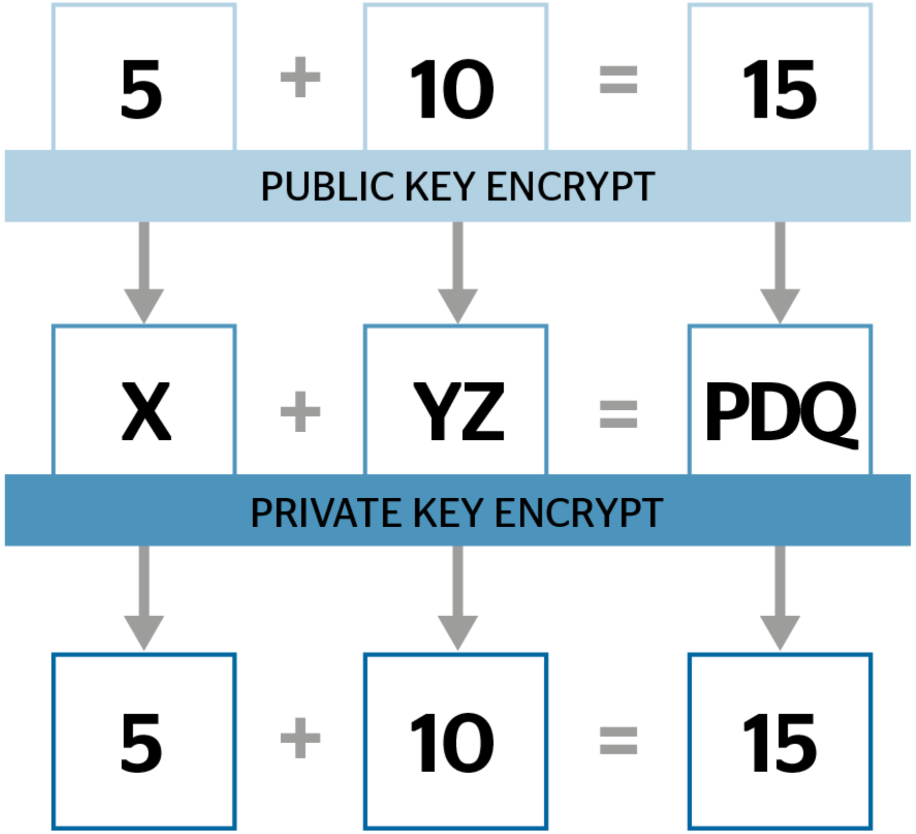
Aggregation satisfies DP

Challenges

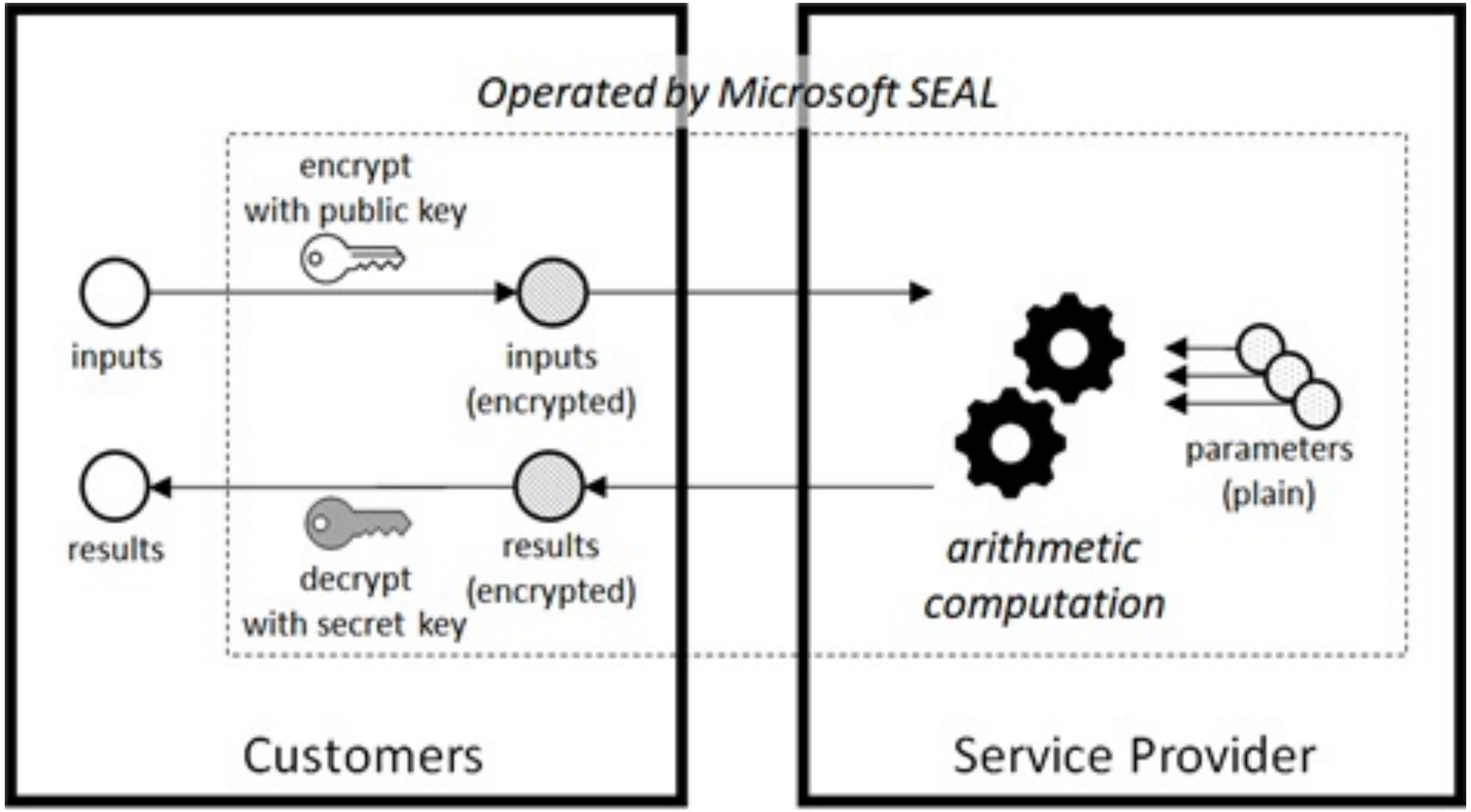
- Quality and privacy trade-off:
  - DP-SGD: the higher number of epochs, the more noises are added.
  - PATE: the more data are used to generate public data, the more noises are added.

Homomorphic Encryption

Homomorphic Encryption



Training Models with Homomorphic Encryption

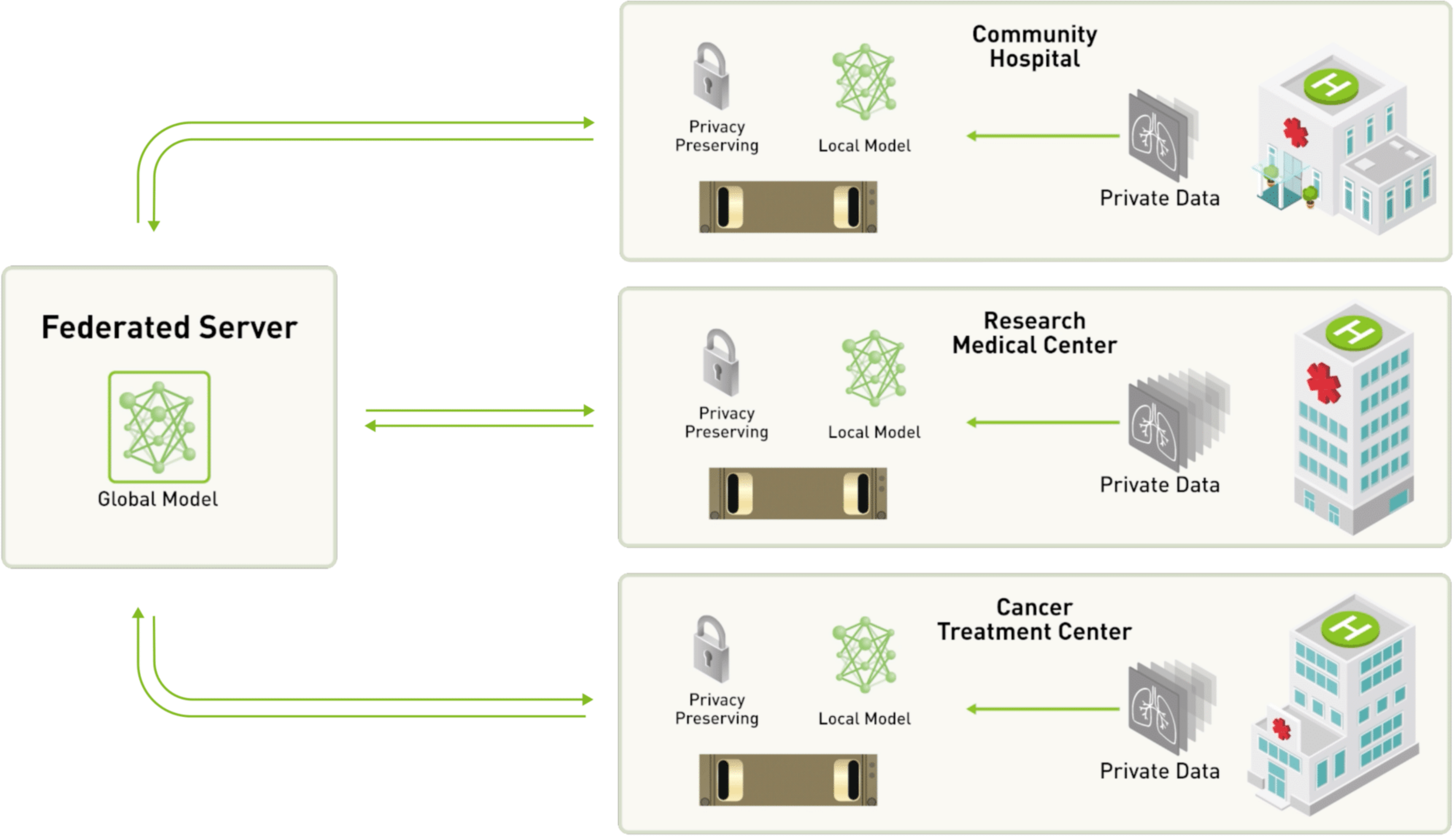


Challenges

- ❖Cannot use standard libraries (Tensorflow, PyTorch) to predict/update training models.
- ❖Increase training time.

Federated Learning

Federated Learning



Even though clients do not share their data, attackers can exploit membership attacks

Privacy-Preserving Federated Learning

❖Differential Privacy

- Add noises to gradients before sending them to the server.

❖Homomorphic Encryption

- Each client encrypts gradients before sending them to the server,
- Server aggregates encrypted gradients and sends the aggregated ones to clients,
- Each client decrypts aggregated gradients and updates their model,

Conclusion

❖Users' privacy must be considered when their data are handled.

❖State-of-the-art Techniques:

- k-Anonymity: is flexible and is used when the usage of the data is unknown,
- Differential Privacy: uses to generate data' statistics and trains models with common optimization algorithms (e.g., SGD),
- Homomorphic Encryption: uses when performance is unimportant,
- Federated Learning: requires transferring a lot of data and needs to combine with differential privacy and homomorphic encryption.

Recommended Resources

❖ k-Anonymity:

- Hoang, A.-T., Carminati, B., and Ferrari, E. 2020. Cluster-Based Anonymization of Knowledge Graphs. Applied Cryptography and Network Security, Springer International Publishing, 104–123.

❖ Differential Privacy:

- Dwork, C. and Roth, A. 2014. The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science 9, 3–4, 211–407.
- Near, J.P. and Abuah, C. Programming Differential Privacy. https://programming-dp.com/book.pdf.
- Zhu, T., Li, G., Zhou, W., and Yu, P.S. 2017. Differential Privacy and Applications. Springer, Cham.

❖ Federated Learning:

- Tensorflow Tutorial on Federated Learning https://www.tensorflow.org/federated/federated_learning

Thank you for your attention