Privacy-Preserving AI: Privacy-preserving machine learning techniques. Federated learning and differential privacy

Artificial Intelligence System Engineering

Data driven AI

- AI Systems and applications that heavily rely on data
  - Make predictions
  - Learn
  - Perform tasks
- Data is the foundation
- Models use datasets to learn patterns, relationships, and correlations
- Models enhance their performance as additional relevant data becomes accessible

Data driven AI

- Adaptability
- Data is the core input
- The process is itterative

Data driven AI

- Data driven organizations (Klas Haller)

Privacy-Preserving Machine Learning (1)

- Machine Learning as a Service (MLaaS)
- Facebook–Cambridge Analytica scandal (2018)
- Facebook ML-based facial recognition technology lawsuit (2020)

- Google developed Randomized Aggregatable Privacy-Preserving Ordinal Response (RAPPOR)

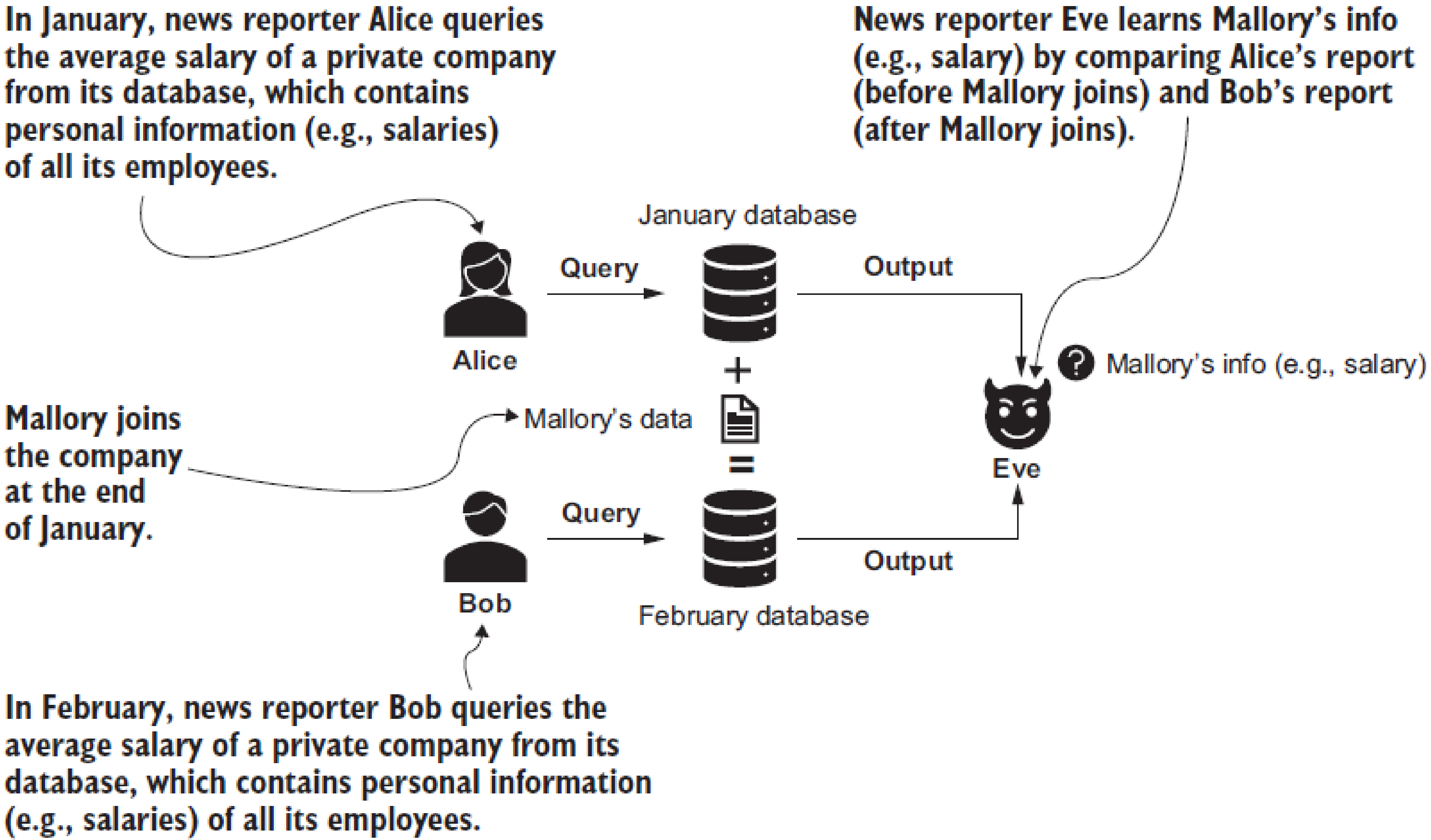
Privacy-Preserving Machine Learning (2)

- Privacy-preserving ML (PPML) algorithms
- Threats and attacks:
  - De-anonymization (re-identification)
  - Reconstruction attacks
  - Parameter inference attacks
  - Model inversion attacks
  - Membership inference attacks

Privacy-Preserving Machine Learning (3)

- Securing privacy:
    - Differential privacy (DP)
    - Local differential privacy
    - Privacy-preserving synthetic data generation
    - Privacy-preserving data mining techniques
        - Privacy-preserving data mining (PPDM)
        - Data collection stage: randomization techniques
        - Data publishing and processing: remove certain attributes, data sanitization (generalization, suppresion, anotamization, perturbation)
        - Output: Association rule hiding, Downgrading classifier effectiveness,Query auditing and restriction
        - Compressive privacy

Differential privacy



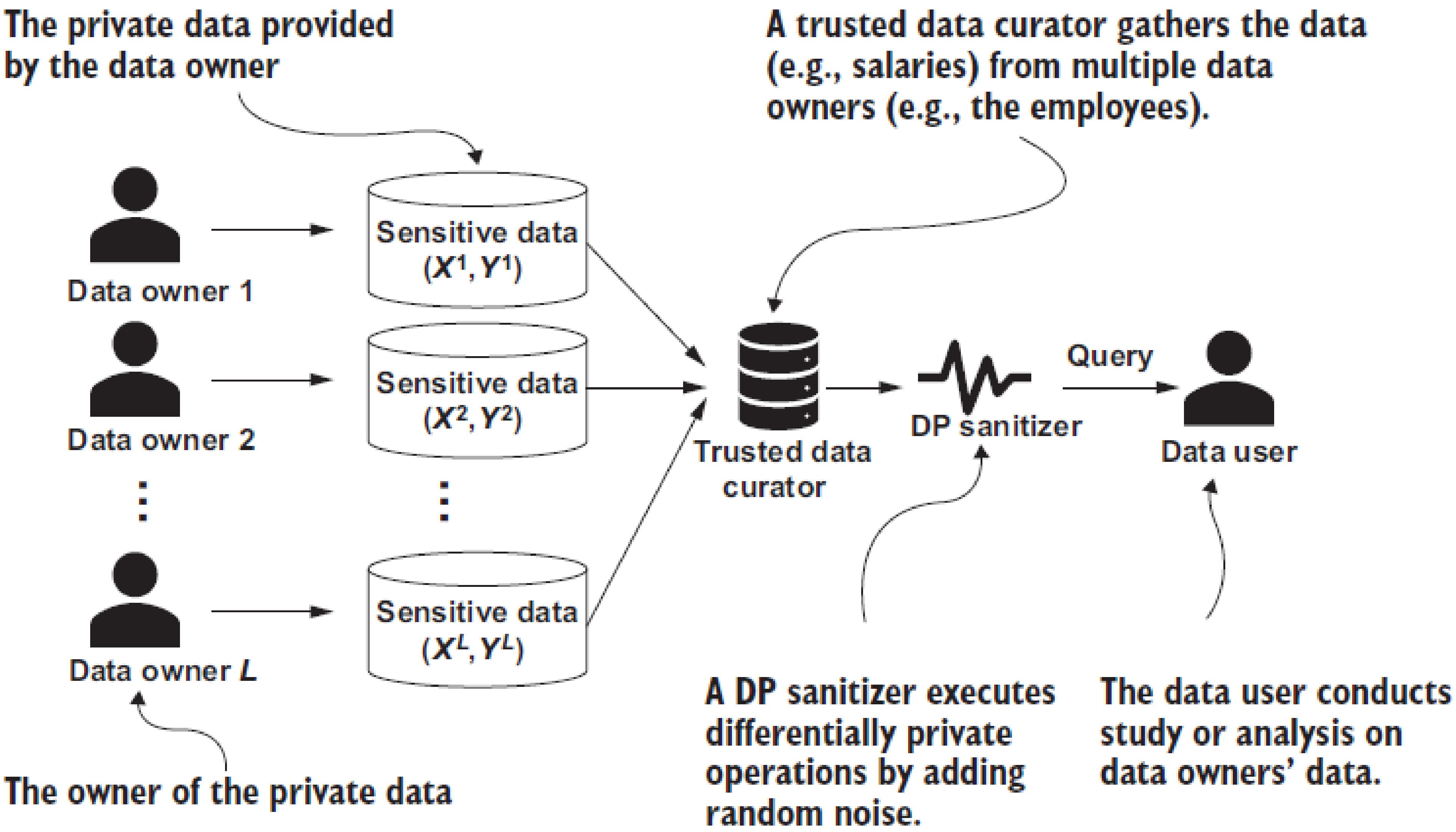
In January, news reporter Alice queries the average salary of a private company from its database, which contains personal information (e.g., salaries) of all its employees.

News reporter Eve learns Mallory's info (e.g., salary) by comparing Alice's report (before Mallory joins) and Bob's report (after Mallory joins).

Mallory joins the company at the end of January.

In February, news reporter Bob queries the average salary of a private company from its database, which contains personal information (e.g., salaries) of all its employees.

J. Morris Chang, Di Zhuang, G. Dumindu Samaraweera - Privacy-Preserving Machine Learning, 2023



The private data provided by the data owner

A trusted data curator gathers the data (e.g., salaries) from multiple data owners (e.g., the employees).

The owner of the private data

A DP sanitizer executes differentially private operations by adding random noise.

The data user conducts study or analysis on data owners' data.

J. Morris Chang, Di Zhuang, G. Dumindu Samaraweera - Privacy-Preserving Machine Learning, 2023

Mechanisms of differential privacy

- Binary mechanism (randomized response)
- Laplace mechanism
- Exponential mechanism
- Etc...

- https://diffprivlib.readthedocs.io/en/latest/modules/mechanisms.html
- https://rbcborealis.com/research-blogs/tutorial-12-differential-privacy-i-introduction/

Applying differential privacy in machine learning

- Input perturbation
- Algorithm perturbation
- Output perturbation
- Objective perturbation
  - (https://kronosapiens.github.io/blog/2017/03/28/objective-functions-in-machine-learning.html)

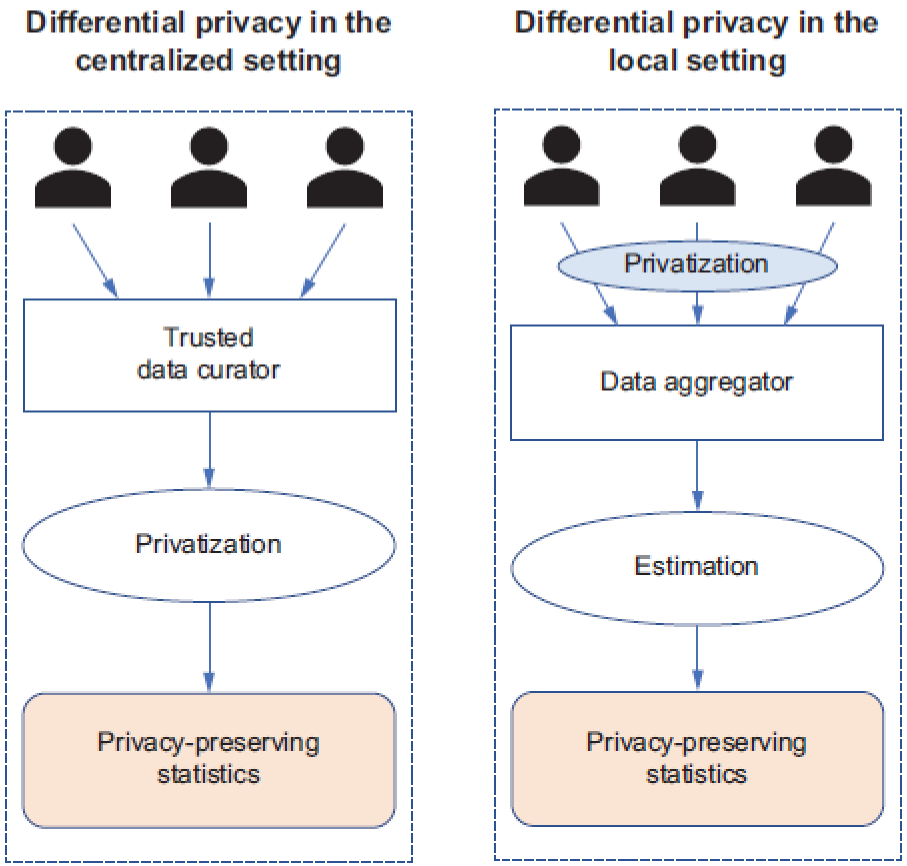
Differentially private supervised learning algorithms

- Differentially private naive Bayes classification (https://arxiv.org/abs/1905.01039)
- Differentially private logistic regression (https://systems.cs.columbia.edu/private-systems-class/papers/Chaudhuri2009Privacy.pdf)
- Differentially private linear regression (https://arxiv.org/abs/2007.05157)

Differentially private unsupervised learning algorithms

- Differentially private k-means clustering
  - https://dl.acm.org/doi/10.1145/2857705.2857708
  - https://arxiv.org/abs/2406.11649
  - https://ieeexplore.ieee.org/abstract/document/9064731

Local differential privacy


Differential privacy in the centralized setting
Trusted data curator
Privatization
Privacy-preserving statistics
Differential privacy in the local setting
Privatization
Data aggregator
Estimation
Privacy-preserving statistics

J. Morris Chang, Di Zhuang, G. Dumindu Samaraweera - Privacy-Preserving Machine Learning, 2023

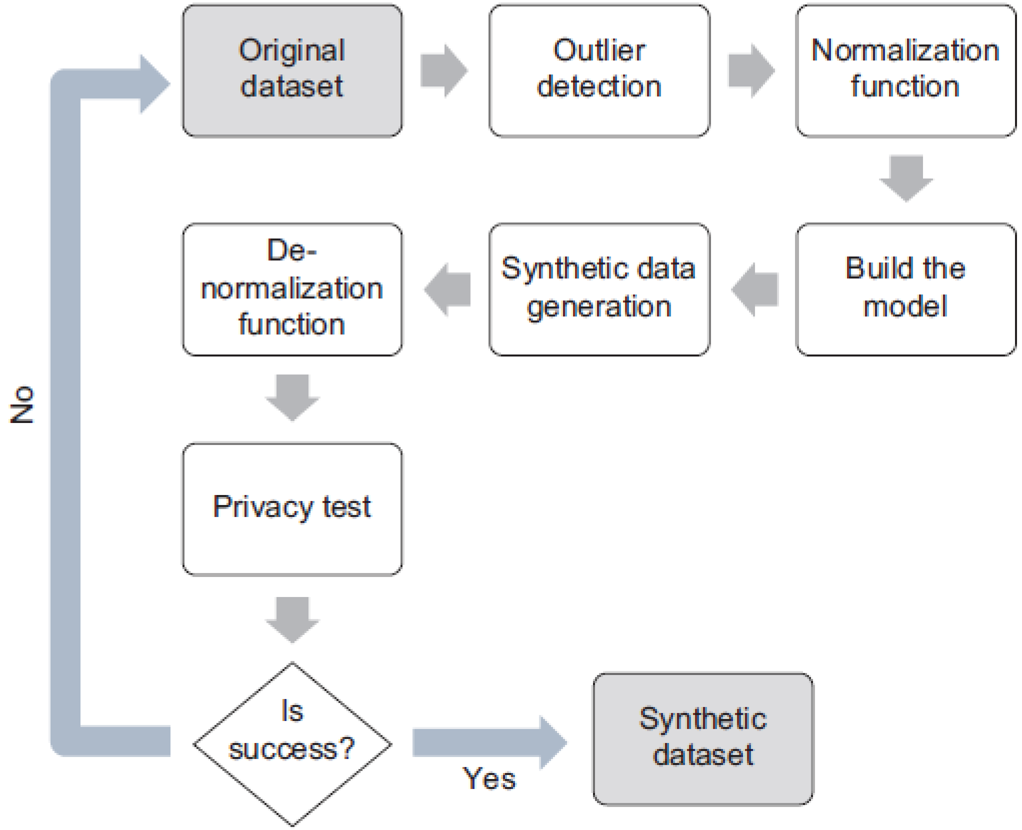
The mechanisms of local differential privacy

- Direct encoding
- Histogram encoding
- Unary encoding
- Examples with code:
  - https://programming-dp.com/ch13.html
- Survey (paper): https://onlinelibrary.wiley.com/doi/10.1155/2020/8829523

Advanced LDP mechanisms

- The Laplace mechanism for LDP
- Duchi’s mechanism for LDP
- The Piecewise mechanism for LDP

Privacy-preserving synthetic data generation

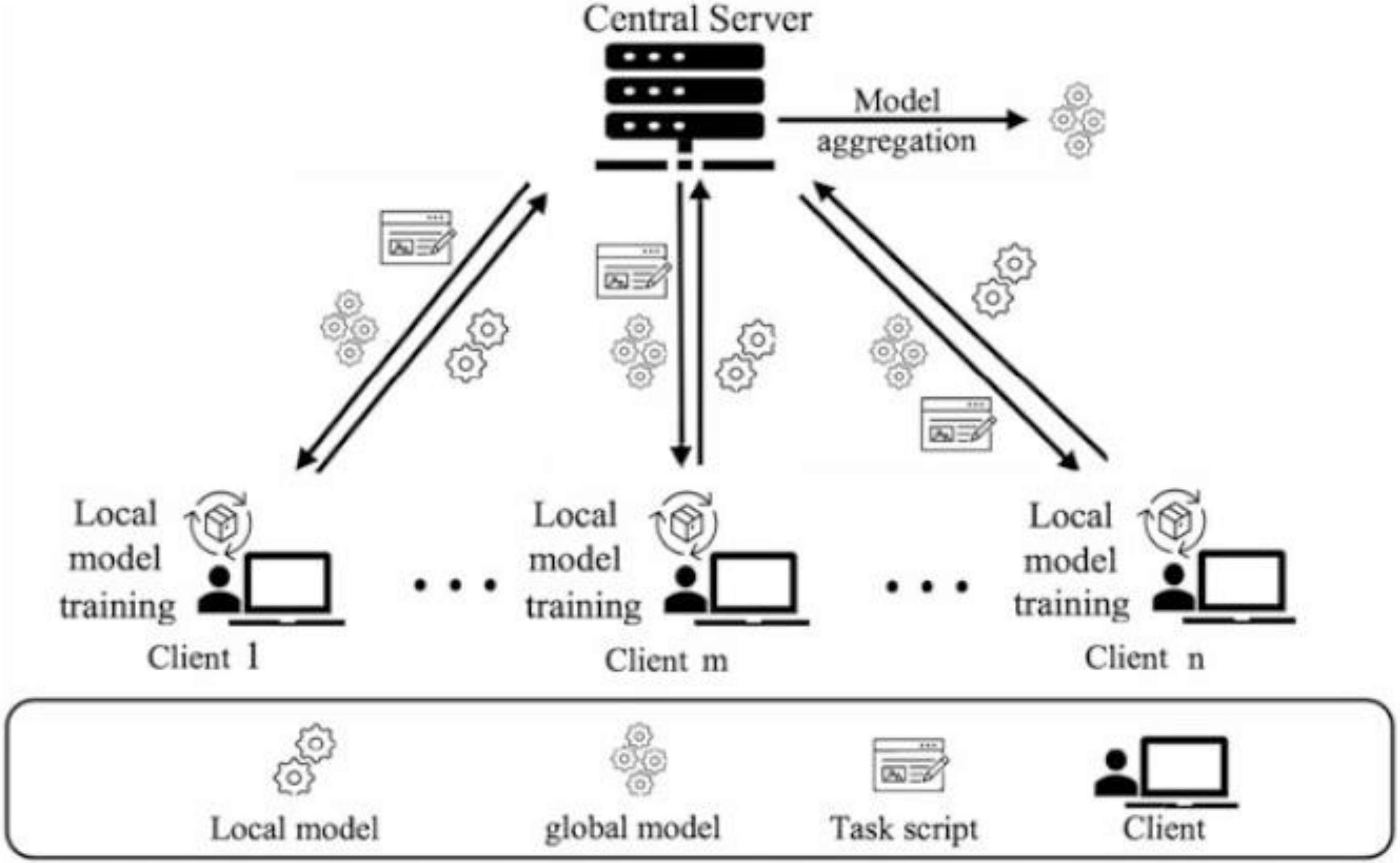


J. Morris Chang, Di Zhuang, G. Dumindu Samaraweera - Privacy-Preserving Machine Learning, 2023

Federated learning

- A decentralized approach to machine learning where models are trained across multiple devices or servers (referred to as "clients") while keeping the data localized on those devices

Federated learning



Shui Yu, Lei Cui - Security and Privacy in Federated Learning, 2024

Federated learning: Vulnerabilities

- Clients
- Server
- Aggregator
- Communication

Attacks in Federated Learning

- Inference attacks
- Poisoning attacks
- GAN-Based attacks

Defense techniques

- Differential Privacy (DP)
- Secure Multi-party Computation (SMPC)
- Secure Data Aggregation
- Anonymous Communication and Shuffle Model

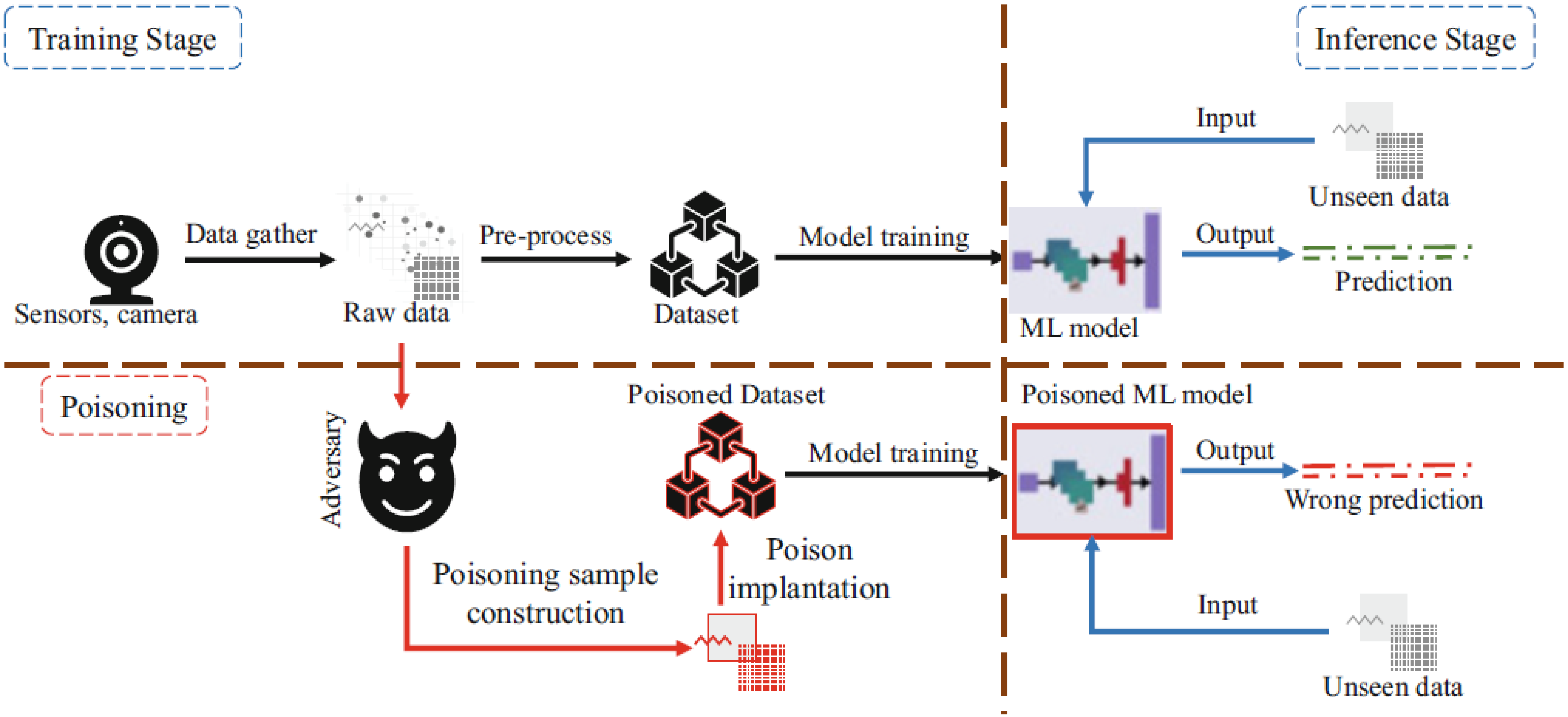
Inference Attacks in FL

- Model Inversion Attacks:
  - Attackson aggregated gradient
  - Attackson global model
- Property Inference Attacks
- Membership Inference Attacks:
  - Data knowledge
  - Training knowledge
  - Model knowledge
  - Output knowledge
- Model Inference Attacks

Counter-Inference Attacks

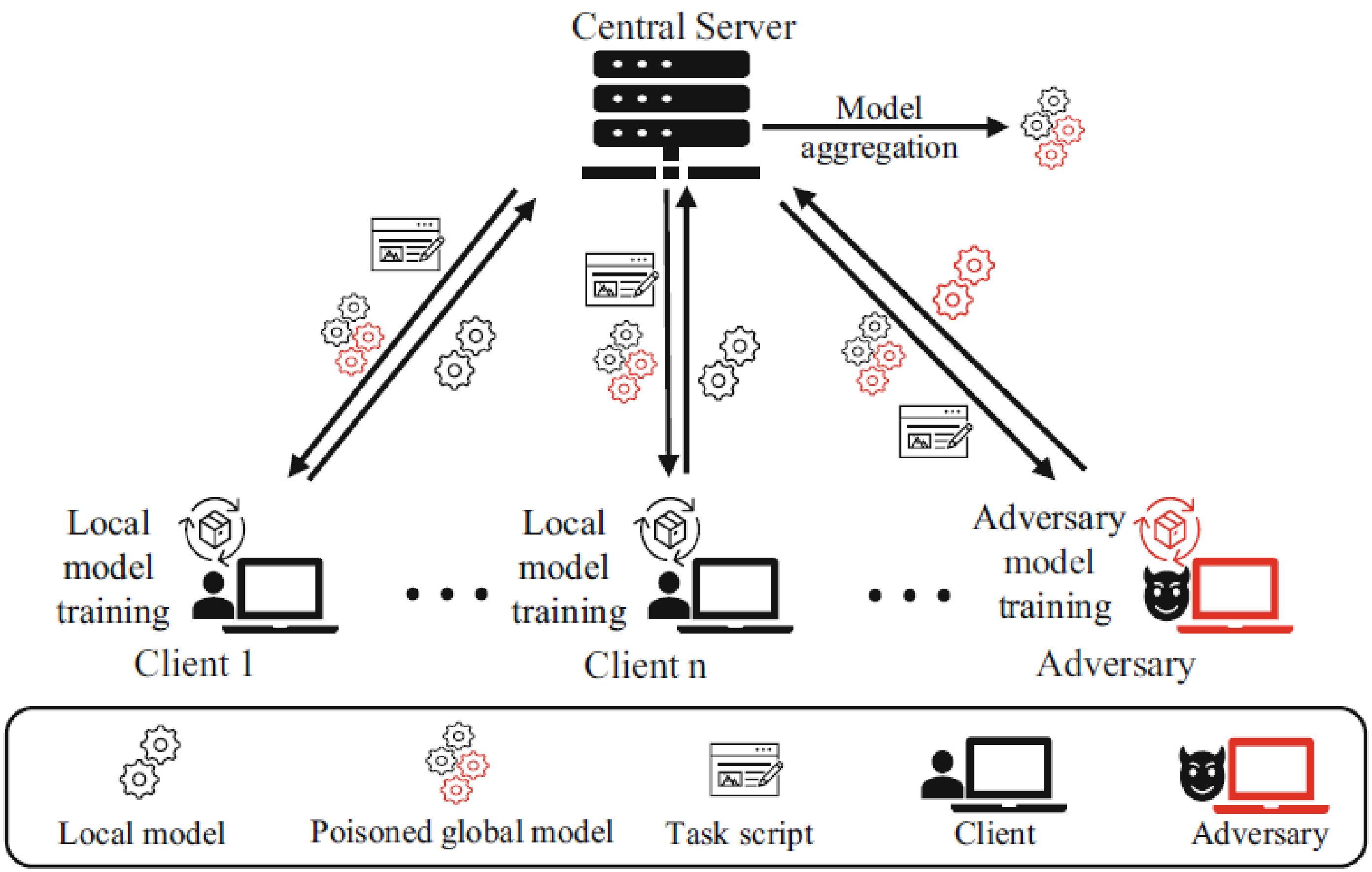
- Machine Learning Optimization-Based Defense
- Perturbation-Based Defense
- Knowledge Distillation
- AdversarialMachine Learning
- Encryption-Based Methods

Poisoning attacks



Shui Yu, Lei Cui - Security and Privacy in Federated Learning, 2024

Poisoning attacks in FL



Shui Yu, Lei Cui - Security and Privacy in Federated Learning, 2024

Poisoning attacks in FL

- Targeted Poisoning Attacks
- Untargeted Poisoning Attacks
- Backdoor Poisoning Attacks
- https://www.researchgate.net/publication/347178320_Threats_to_Federated_Learning

Counter Poisoning Attacks

- Counterattacks from Data Perspective
  - Byzantine-resilient algorithm for distributed SGD
  - Trimmed mean
  - Bulyan
- Counterattacks from Behavior Perspective
- Other (research papers)

Differential Privacy in FL

- Centralized Differential Privacy
- Local Differential Privacy
- Distributed Differential Privacy
- Variant Differential Privacy
- The Combinationof Differential Privacy and OtherMethods