Federated Learning in IoT: Privacy Issues and Solutions

FIT Europe Milan, 19 November 2021

FIT Europe - Blue Team

University Politehnica of Bucharest

- Andreia-Irina Ocănoaia
- Teia-Andrada Vava
- Vlad-Stefan Dieaconu

University of Milan

- Marco Pedrinazzi
- Daniel Buruian

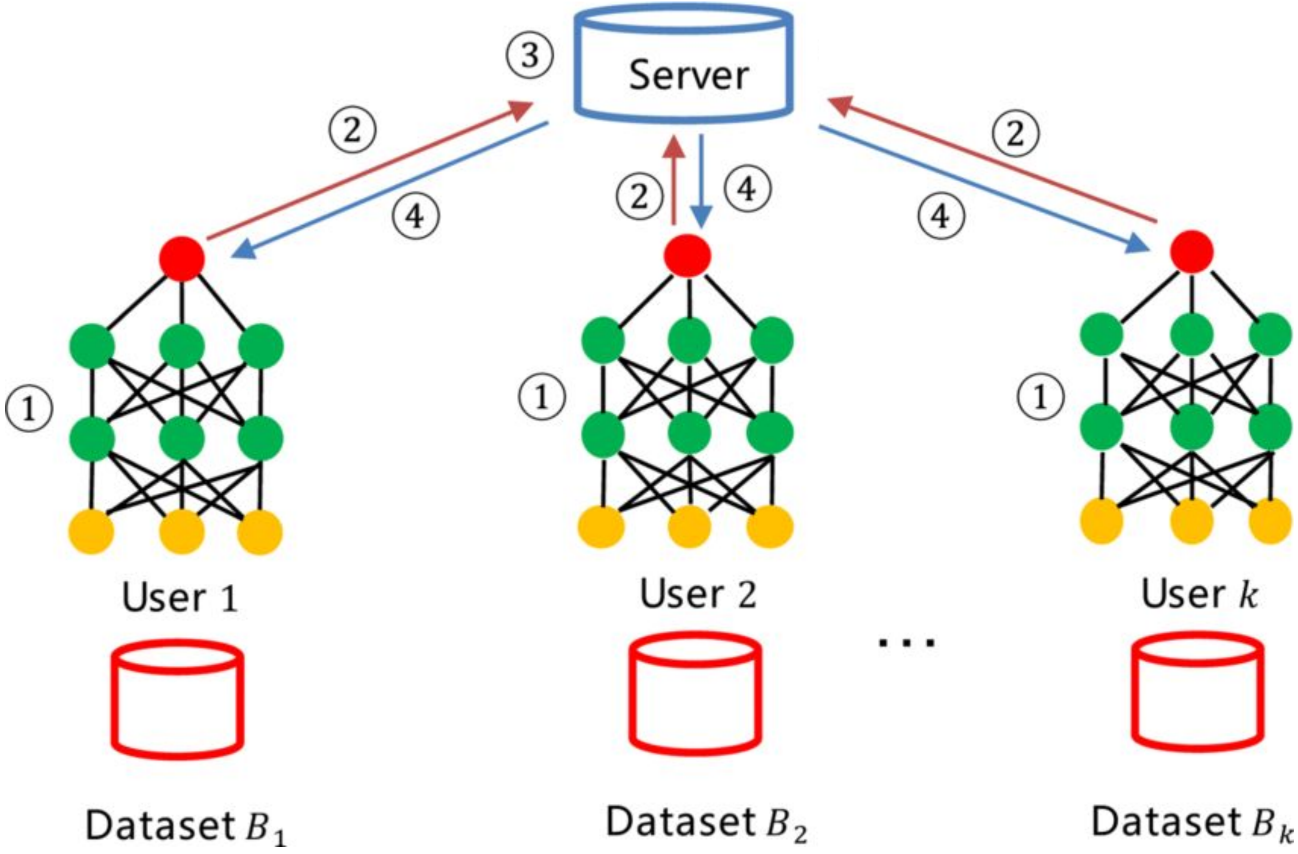
INSA Lyon

- Corentin Forler
- Saad Ouidan

University of Passau

- Shashi Kumar
- Seyed Peyman Hosseini

Federated Learning



Problem Statement

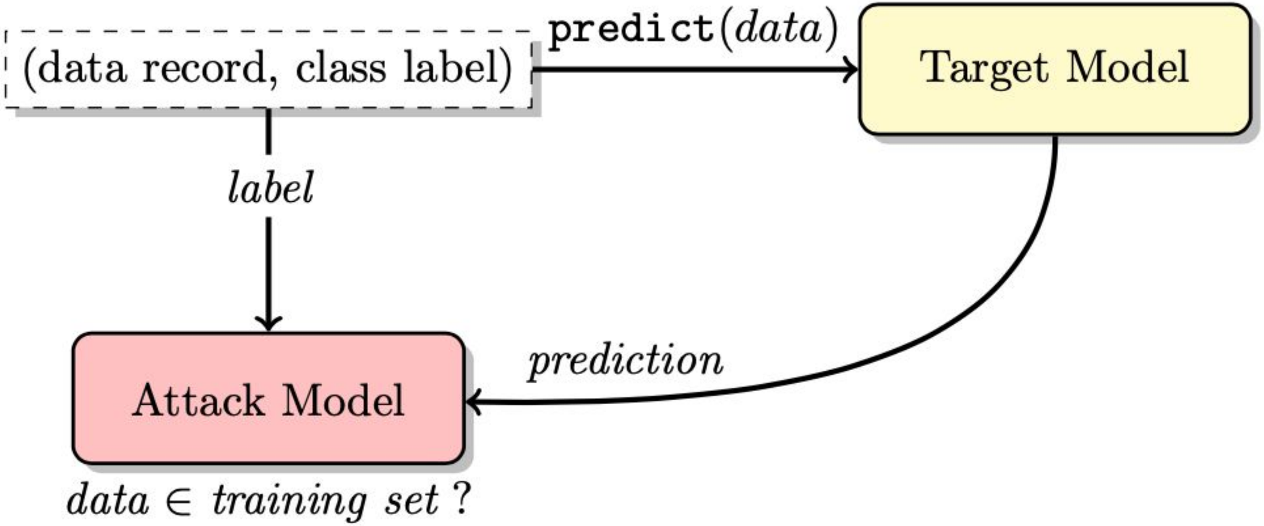
- **FL naturally offers privacy advantages**
  - Real user data is never exchanged

- **FL architectures can be attacked**
  - Sensitive information can be extracted from the local model sent to the server

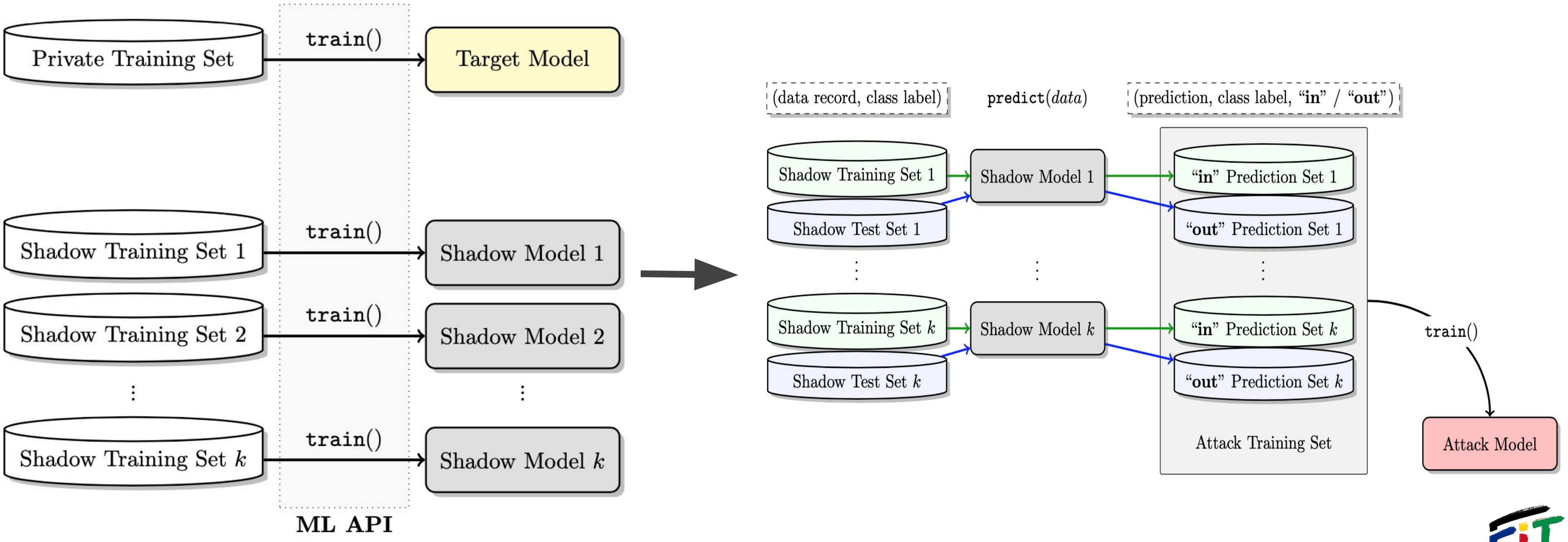
Privacy Attacks

<table>
<tr><th colspan="2">Attack Models</th><th>Privacy-preserving techniques employed at server side</th><th>Privacy-preserving techniques employed at client side</th></tr>
<tr><td rowspan="2">Inference Attacks</td><td>Reconstruction Attacks</td><td rowspan="2">– SMC & Secure Aggregation
– Homomorphic Encryption</td><td rowspan="2">– SMC & Secure Aggregation
– Homomorphic Encryption
– Batch Level DP
– User-level DP</td></tr>
<tr><td>Membership Tracing</td></tr>
</table>

Privacy Attacks - Membership Tracing (1)



Privacy Attacks - Membership Tracing (2)

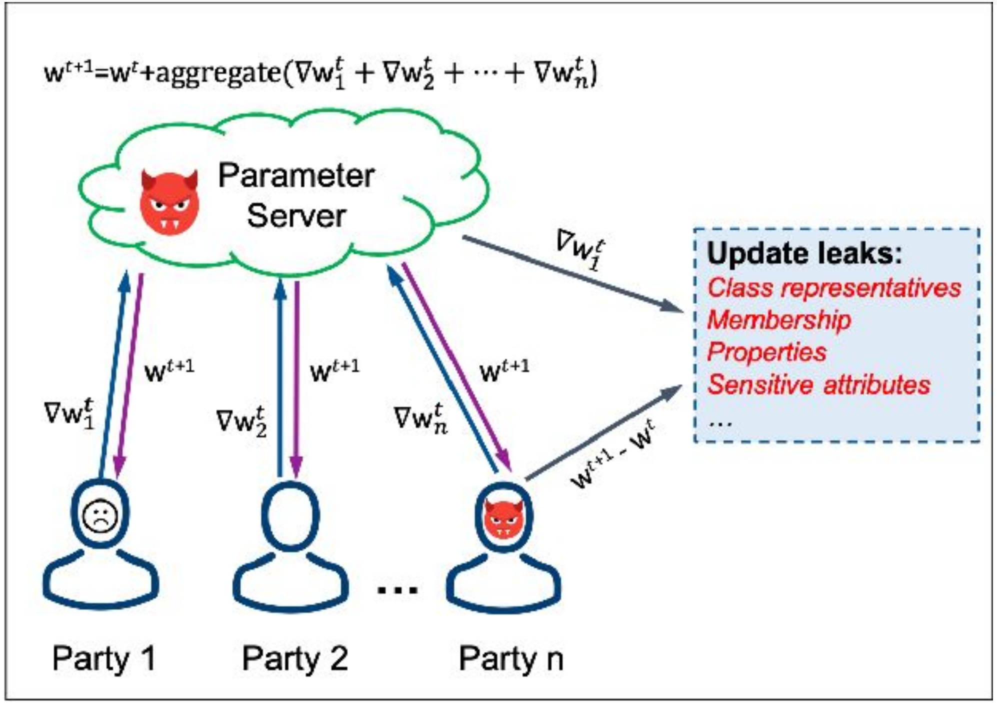


Privacy Attacks - Reconstruction Attacks (1)

Threat Model:

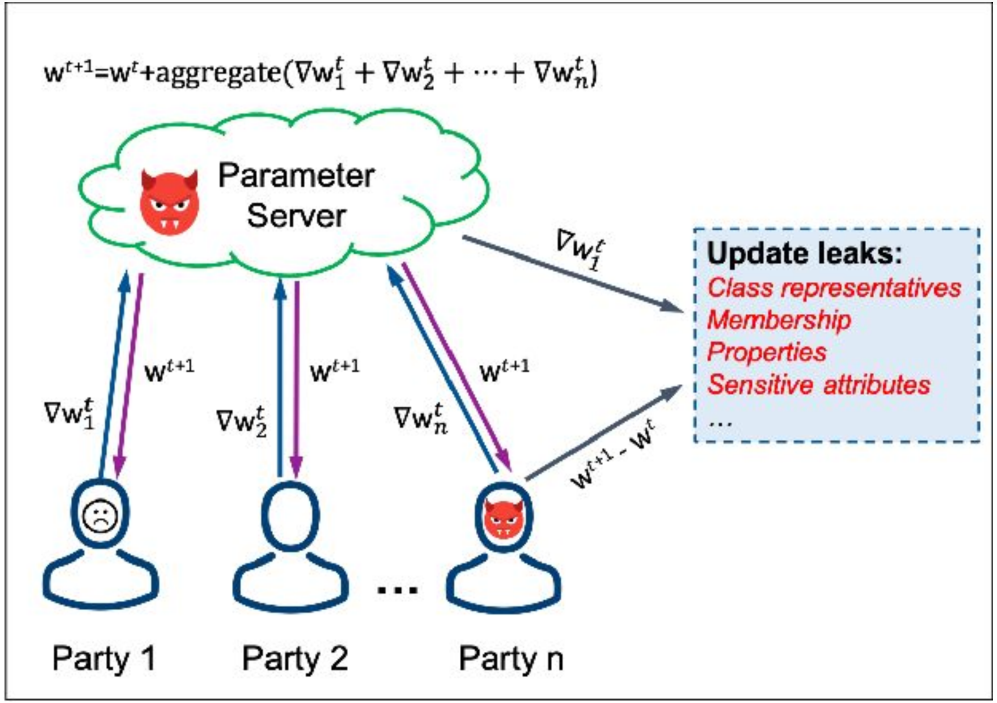
- Access to a trained model
- Access to non-sensitive attributes
- Access to gradients across multiple training iterations
- If the membership of the data is not known, a Membership Inference Attack can be launched.

Privacy Attacks - Reconstruction Attacks (2)



Privacy Attacks - Reconstruction Attacks (2)

$$J = cosinesim(w^{t+1}, w_n^t)$$



SotA - Differential Privacy

- **User-level differential privacy**
  - Hides the participant sensitive data by adding noise to the whole local training dataset.

- **Batch-level differential privacy**
  - Adds noise to the model parameters.
  - Groups of a few parameters are selected randomly or deterministically.
  - Parameters are updated at each communication round.

SotA - Homomorphic Encryption

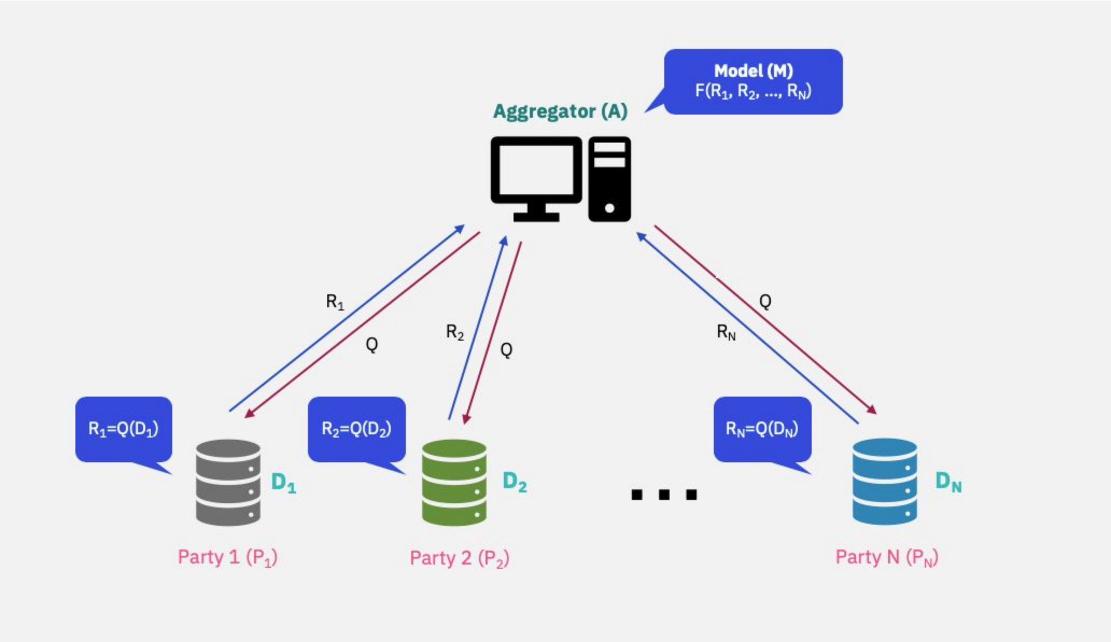
- **A form of encryption that permits users to perform computations on encrypted data without first decrypting it**
  - Resulting computations are left in an encrypted form
  - Sensitive data → Encryption → Computation → Decryption → Result

- **Impractical in an IoT setting**
  - Contains practical limitation in performing computations
  - Some techniques can still potentially leak data and more secure variants are still being developed

SotA - Secure Multiparty Computation and Secure Aggregation

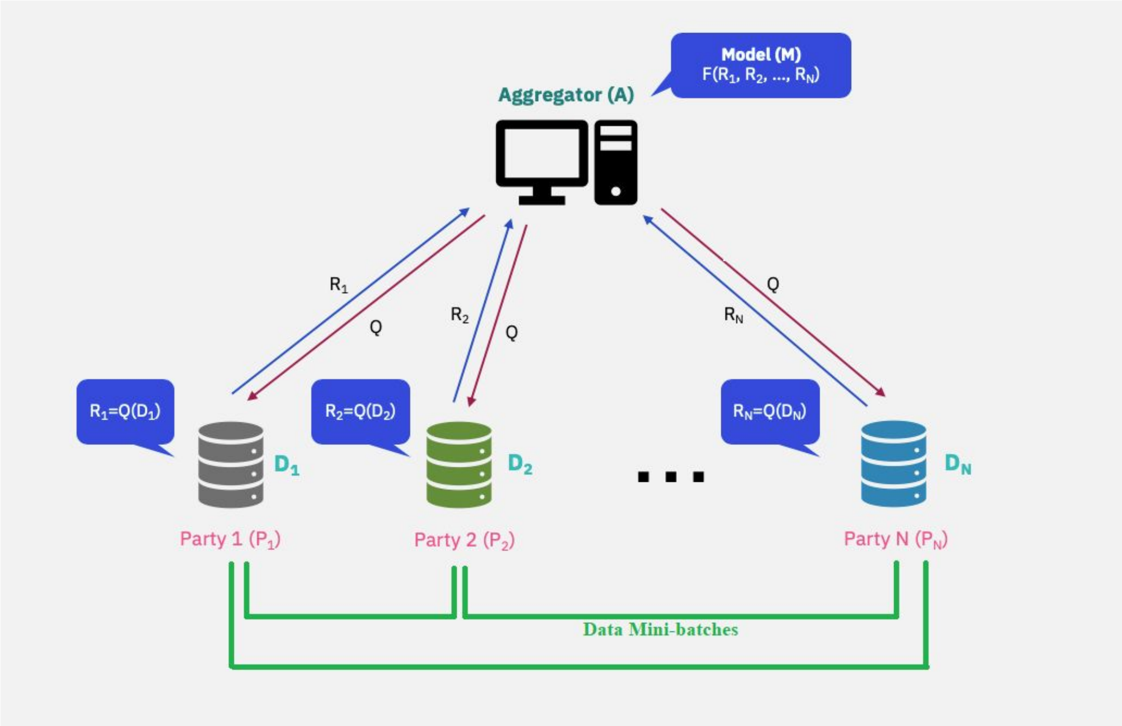
- **SMC is a protocol that distributes computations across multiple parties**
  - A central server can get an aggregate of data from local nodes
  - And no individual party can see the other parties' private data
- **Secure Aggregation is a form of SMC**
  - Local model parameters do not have to be revealed
  - The server cannot infer private data from them
  - Downsides: communication overhead, computational complexity

A New Approach (1) - Classic Federated Learning



A New Approach (2) - Diversity-Infused Federated Learning

- P2P
- Homomorphic encryption
- SMC




YOU GET PRIVACY!!
AND YOU GET PRIVACY!!
EVERYBODY GETS PRIVACY!!
2004

Summary

Privacy Attacks

Federated Learning

Diversity Infused Federated Learning

GDPR

Secure Aggregation

Differential privacy

Homomorphic encryption

IoT

Referenced Papers

[0] **[OUR WORK]** **Federated Learning in IoT: Privacy Issues and Solutions**
[1] Federated Learning: Collaborative Machine Learning without Centralized Training Data
[2] Privacy Preservation in Federated Learning: An insightful survey from the GDPR Perspective
[3] Federated Learning for Internet of Things: Recent Advances, Taxonomy, and Open Challenges
[4] Membership Inference Attacks Against Machine Learning Models
[5] Source Inference Attacks in Federated Learning
[6] A Novel Attribute Reconstruction Attack in Federated Learning
[7] Privacy and Robustness in Federated Learning: Attacks and Defenses
[8] Privacy-Preserving Deep Learning via Additively Homomorphic Encryption
[9] Homomorphic Encryption and Network Coding in IoT Architectures: Advantages and Future Challenges
[10] Differential Privacy Has Disparate Impact on Model Accuracy
[11] Practical Secure Aggregation for Privacy-Preserving Machine Learning
[12] Decentralized Collaborative Learning of Personalized Models over Networks
[13] Federated Learning with Cooperating Devices: A Consensus Approach for Massive IoT Networks
[14] Assisted Learning: A Framework for Multi-Organization Learning
[15] What is Secure Multiparty Computation?
[16] Differential Privacy in TFF
[17] Toward Scalable Fully Homomorphic Encryption Through Light Trusted Computing Assistance


Federated Learning in IoT: Privacy Issues and Solutions

Authors

Andreia-Irina Ocănoaia, University Politehnica of Bucharest, andreia.ocanoaia@gmail.com
Teia-Andrada Vava, University Politehnica of Bucharest, vavateiaandrada@gmail.com
Vlad-Stefan Dieaconu, University Politehnica of Bucharest, vladstefandieaconu@gmail.com
Marco Pedrinazzi, University of Milan, marco.pedrinazzi1@studenti.unimi.it
Daniel Buruian, University of Milan, daniel.buruian@studenti.unimi.it
Corentin Forler, INSA Lyon, corentin.forler@insa-lyon.fr
Saad Ouidan, INSA Lyon, saad.ouidan@insa-lyon.fr
Shashi Kumar, University of Passau, ravula01@ads.uni-passau.de
Seyed Peyman Hosseini, University of Passau, hossei03@ads.uni-passau.de

Abstract

Federated Learning (FL) has emerged as a promising privacy-aware paradigm that allows multiple clients to jointly train a model without sharing their private data. Nevertheless, active research, both specifically related to FL and in general to ML/DL, highlights issues regarding the privacy touted by the FL approach. In this paper we discuss FL in the IoT environment and focus on the privacy issues that it creates. We begin with a brief introduction to FL and immediately proceed to analyze possible attacks that represent serious threats to privacy by allowing server-side inferences. In the successive section we explore possible solutions and propose a novel approach to the problem. Finally, we embed the FL paradigm in the GDPR legal framework and, after a brief introduction to its contents, proceed to analyze how compliance with the latter could be guaranteed by a centralised FL system.

Backup Slides

GDPR compliance in a centralised FL system

The GDPR defines 6 core principles as guidelines for service providers to manage personal data:

- **Lawfulness, Fairness and Transparency**
- **Purpose Limitation**
- **Data Minimisation**
- **Accuracy**
- **Storage Limitation**
- **Integrity and Confidentiality**

Rights of data subjects

The GDPR requires Data Controllers to provide the following rights for Data Subjects:

- **Right to be informed**
- **Right of access**
- **Right to erasure**
- **Right to restrict processing**
- **Right to data portability**
- **Right to object**
- **Right in relation to automated decision making and profiling**

GDPR compliance investigation and demonstration

1. Systematic **description** of data processing operations and associated purposes
2. **Assessment** of the necessity and proportionality of each operation, given its associated purposes.
3. **Assessment** of the data security and privacy risks that might be introduced by each operation