Robust Federated Meta Learning Framework Against Adversaries

Lancaster University
School of Computing and Communications

Dr. Zhengxin Yu (z.yu8@lancaster.ac.uk)
Prof. Neeraj Suri (neeraj.suri@lancaster.ac.uk)

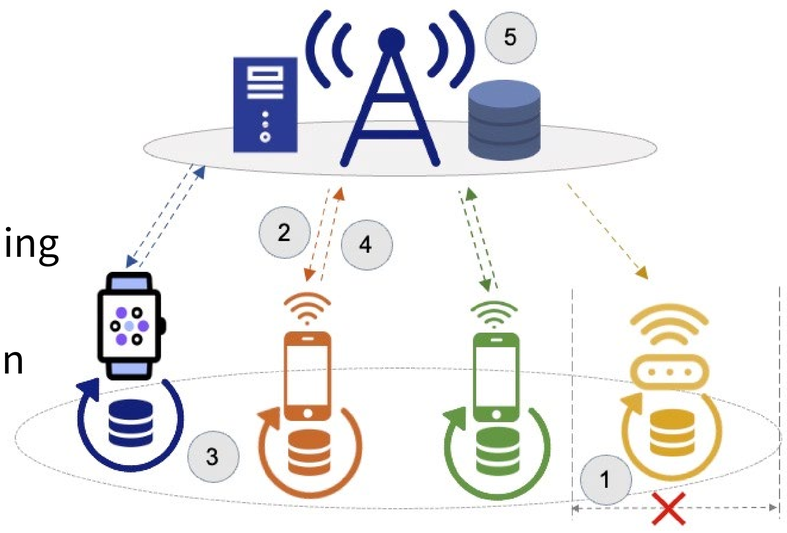
Background : Federated Learning

- Data is born at the edge
- Data processing is moving on devices
- Data silos
- General Data Protection Regulation (GDPR)

Analytics? Learning? → **Federated Learning**

What is Federated Learning (FL)?

- Multiple users collaboratively train a global model
- Keep data decentralised



FL training process:

1) Client selection
2) Download model
3) Local model training
4) Upload model
5) Model aggregation

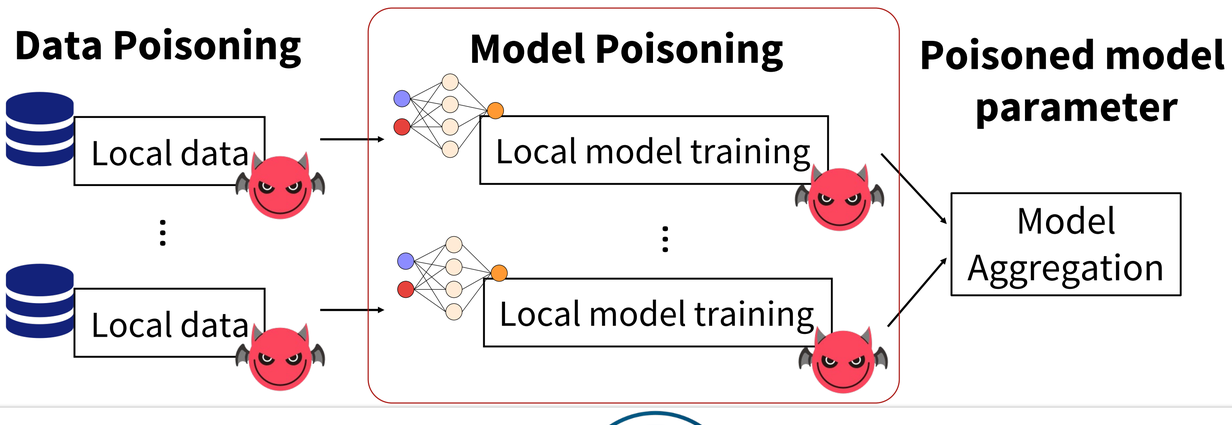
Motivation

- FL systems can be vulnerable to various kinds of failures.

→ **Degrade the learning performance of FL**

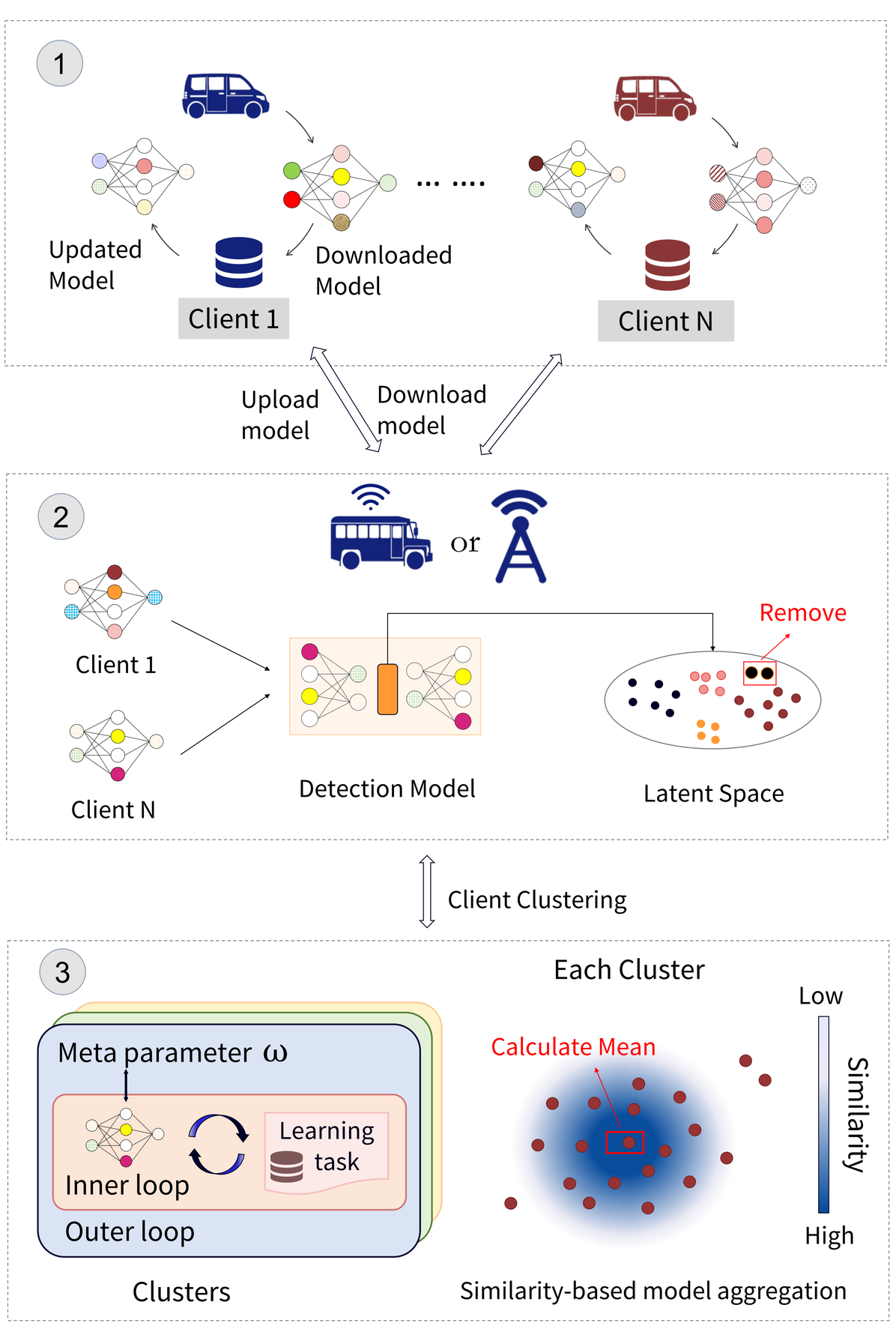
Major Challenges

- Clients upload unreliable model updates intentionally or unintentionally.
- Local resource heterogeneity (Non-IID data distribution).
- Limited influence of attackers in a single round, but attackers can couple their attacks over time.



Insight behind RFML

- RFML: A Robust-by-design FL Meta framework against adversaries that is capable of reducing negative impact from malicious clients on non-IID data.
- A variational autoencoder based anomaly detection model to cluster clients and remove malicious clients.
- Similarity-based adaptive model aggregation method for each clusters.



Conclusion and Future Work

- Introduce a new robust-by-design framework that is able to defend against model poisoning attacks in FL.
- Conduct extensive experiments to evaluate the RFML and demonstrate that the RFML outperforms the existing defence-based methods in terms of model accuracy.
- Explore the applicability of the RFML to multi-attacks and consider more advanced ML models.

This work is supported by the Engineering and Physical Sciences Research Council [grant number: EP/V026763/1]