ABFL: A Blockchain-enabled Robust Framework for Secure and Trustworthy Federated Learning

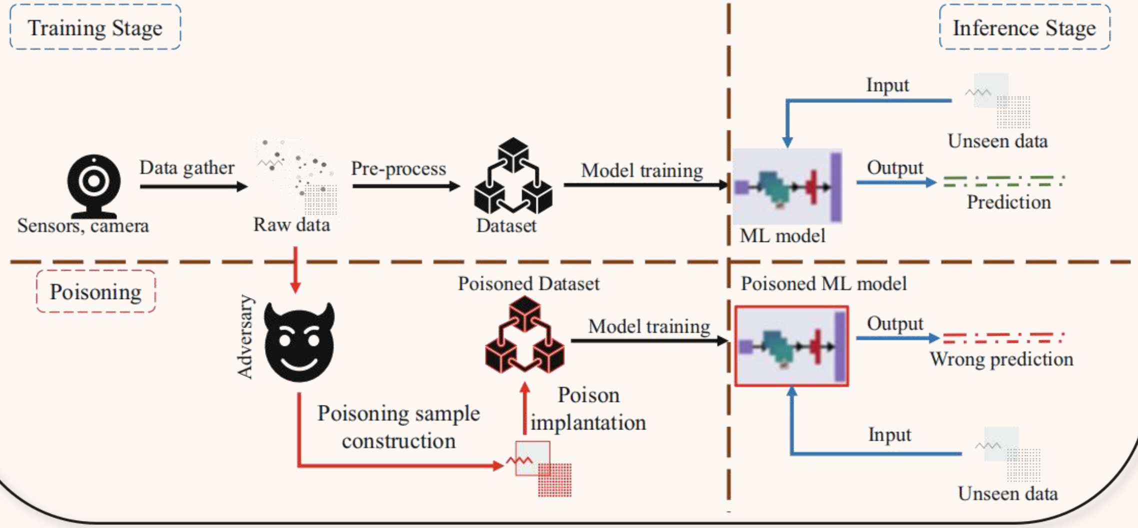
Tianyu Mei, Bo Cui

December 7^{th}, 2023

Poisoning Attacks in Federated Learning (FL)

What Is Poisoning Attack?

> The life cycle of legitimate machine learning (upper part) and poisoning attack (lower part)



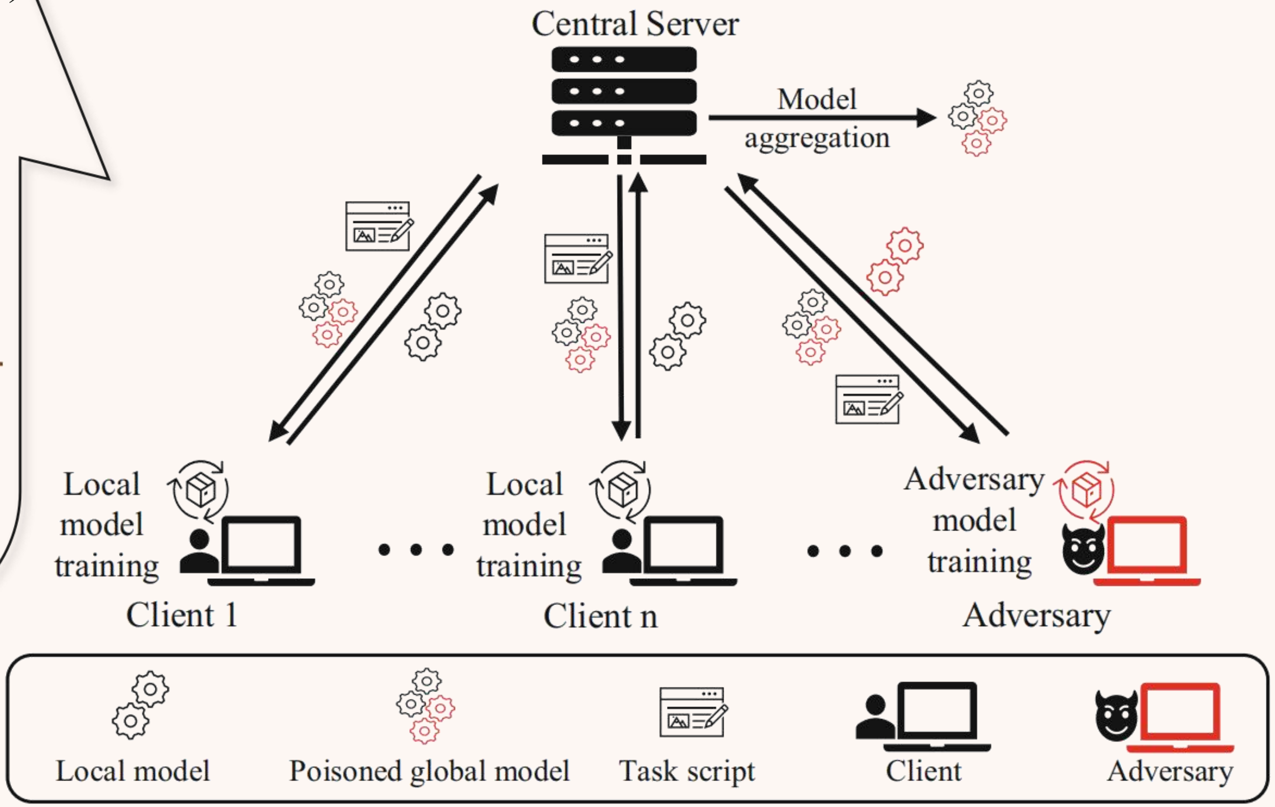
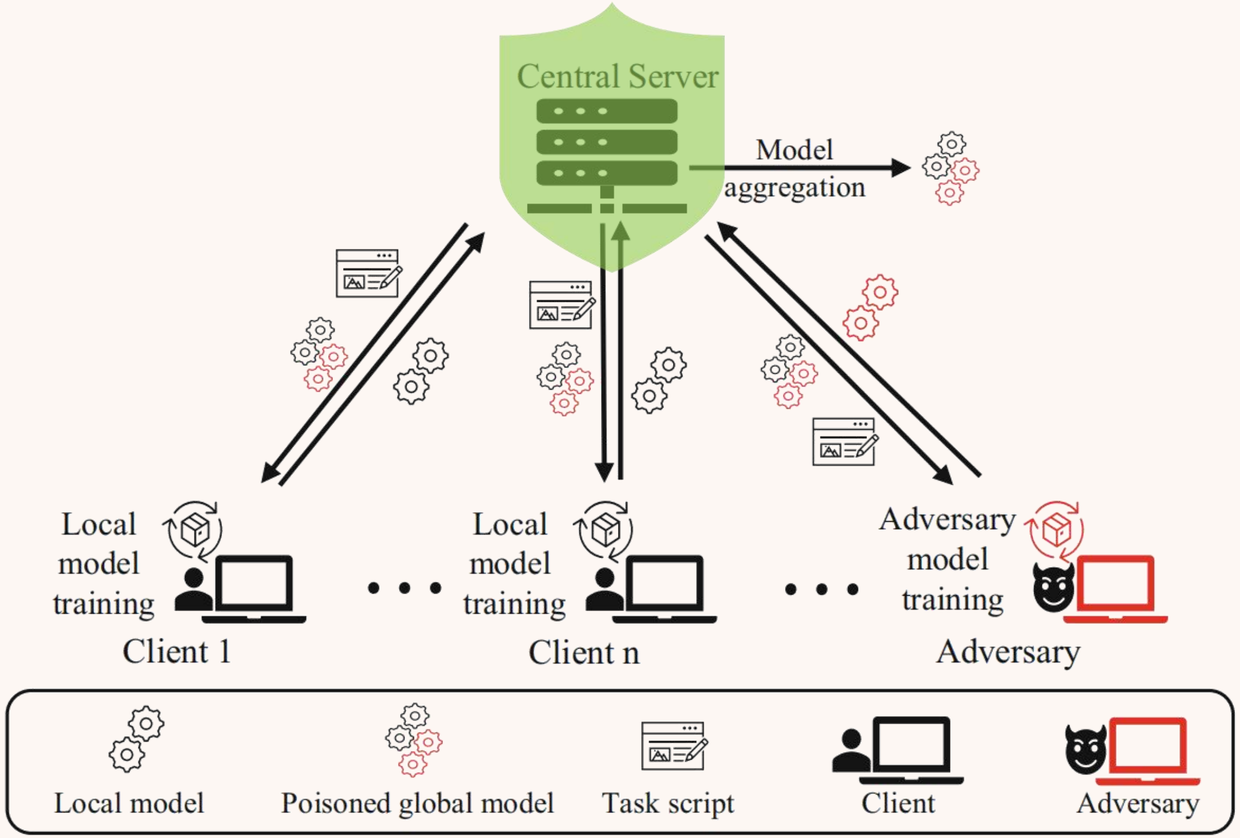


Image source:https://link.springer.com/book/10.1007/978-981-19-8692-5

Counter Poisoning Attacks in Federated Learning



Centralized defense method:

> Secure aggregation

> Anomaly detection

Challenge：

> Malicious central server

> Single point of failure

Image source:https://link.springer.com/book/10.1007/978-981-19-8692-5

Blockchain-based Federated learning (BFL)

Challenge:

> Consensus mechanism efficiency

> Security defense mechanism

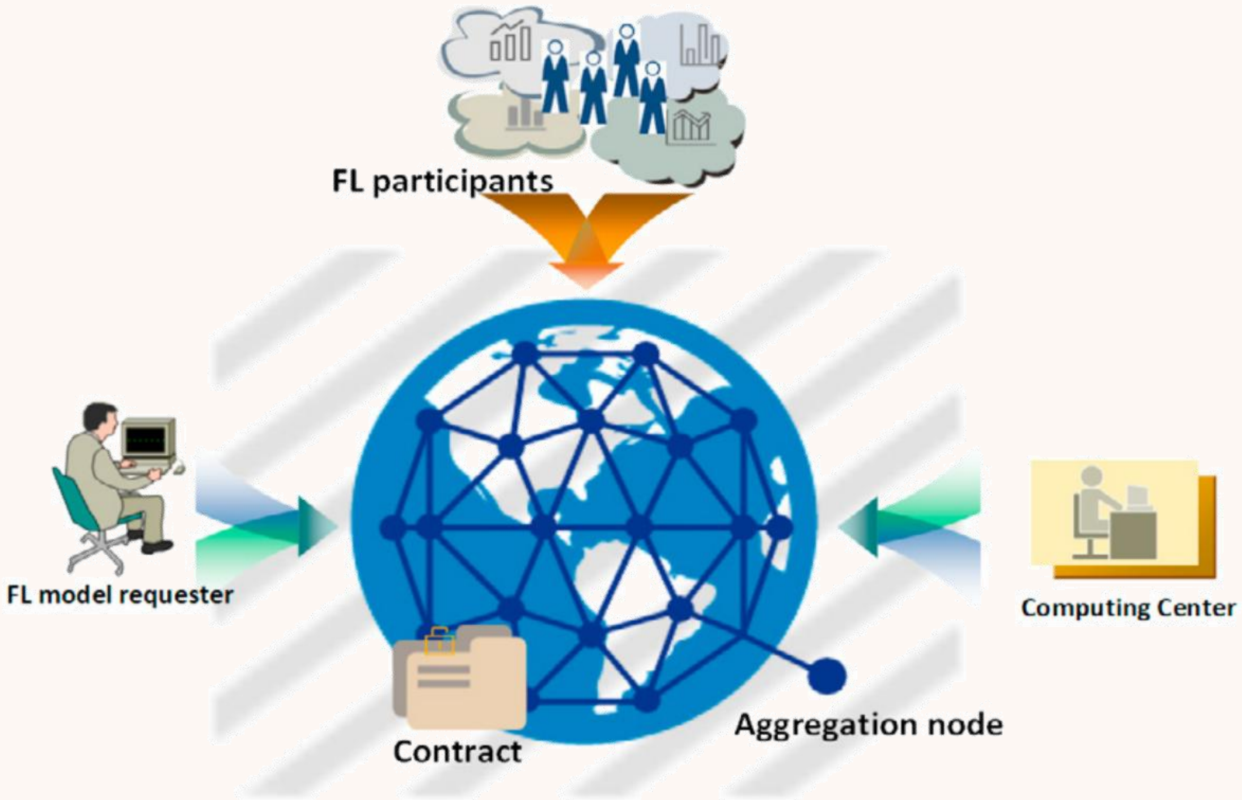
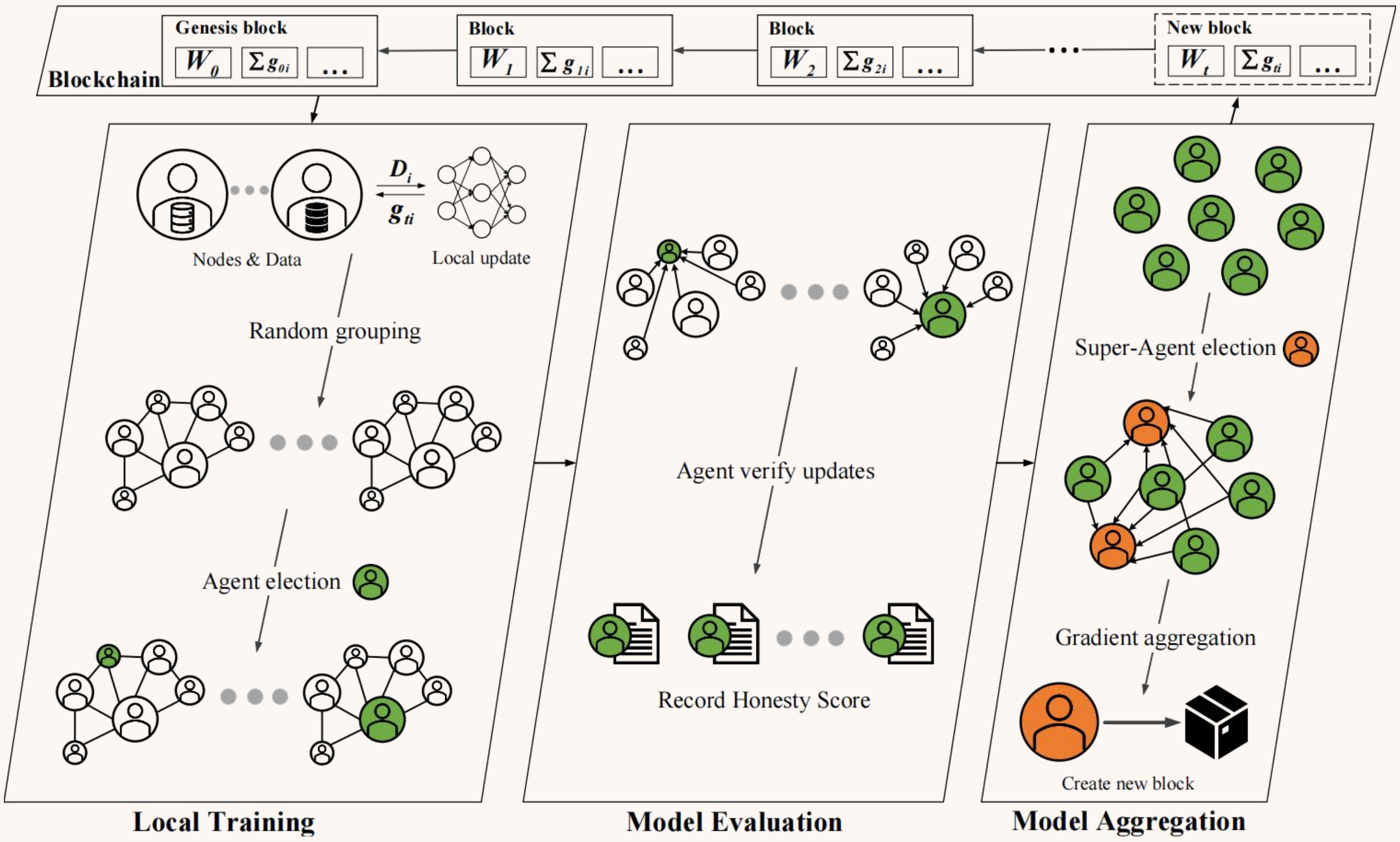


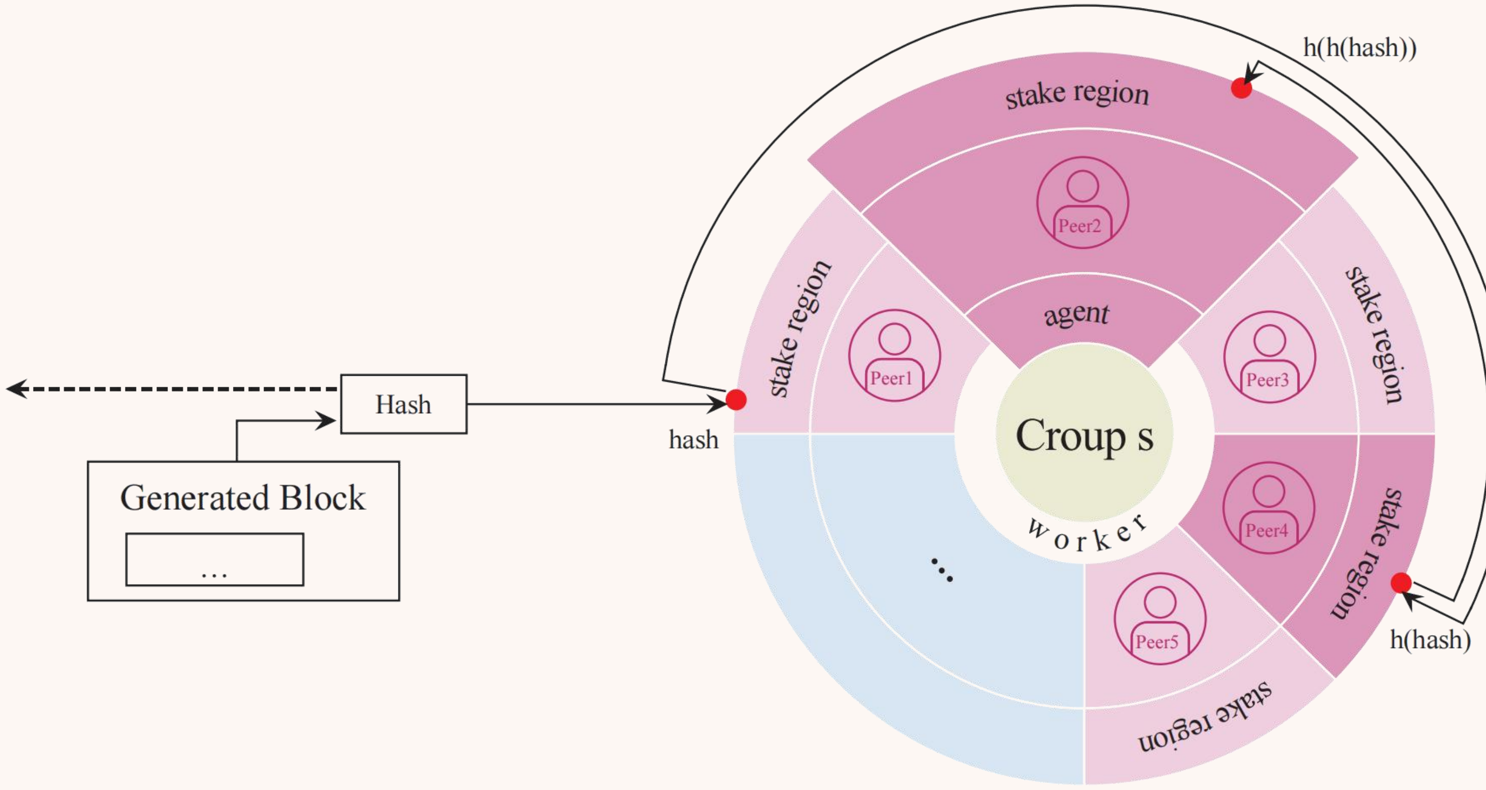
Image source： Wang Z, Yan B, Dong A. Blockchain Empowered Federated Learning for Data Sharing Incentive Mechanism. 2022

Our Proposed Architecture - ABFL

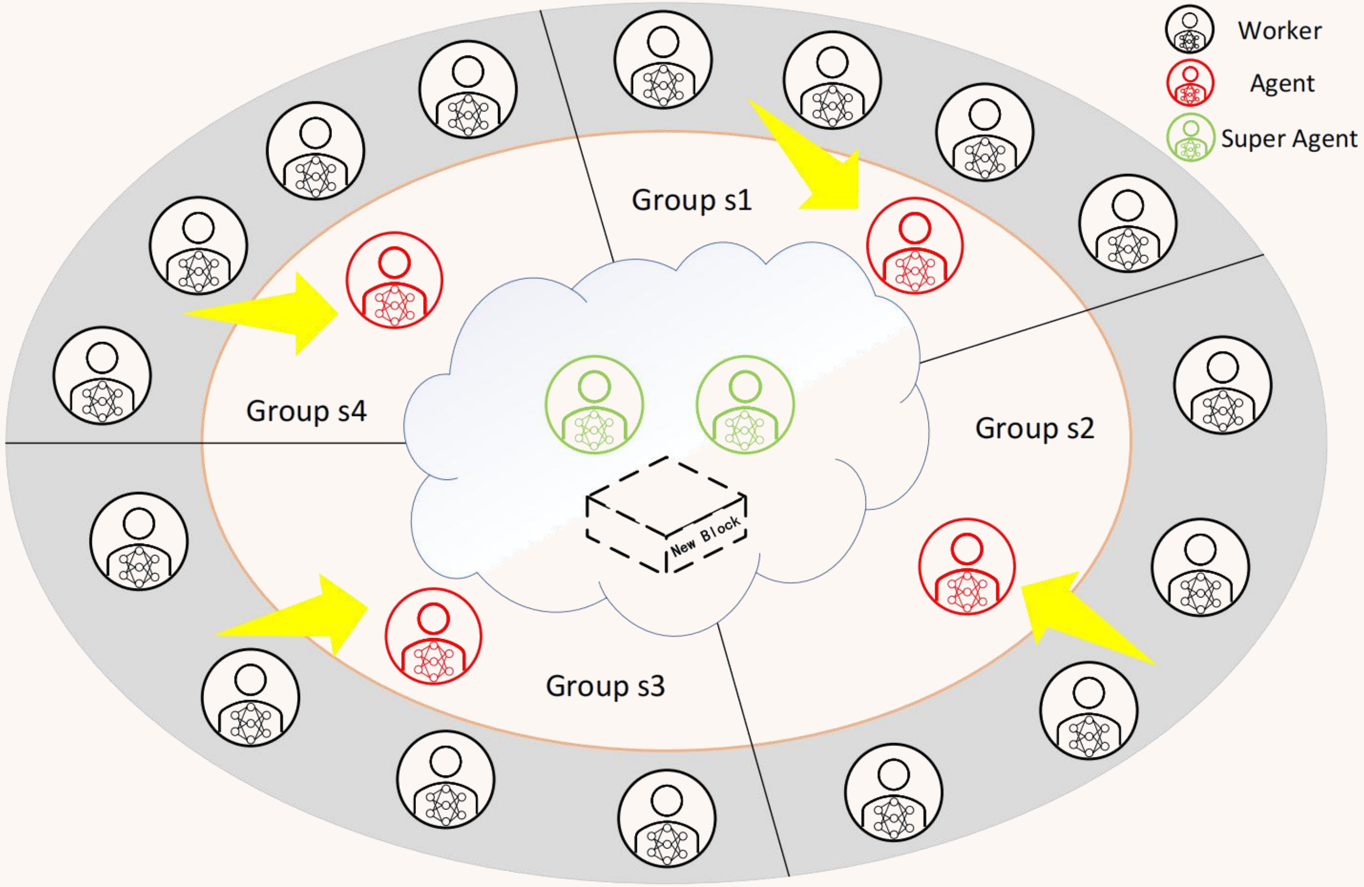


Role selection

> Consistent hashing algorithm with honesty score weighting



Proof of Honesty Score-based Agent Consensus Mechanism



Model Evaluation

> Dual-benchmark robustness algorithm based on cosine similarity

Benchmark 1: Local model updates for agent nodes

Benchmark 2: Model updates based on historical data predictions

At the beginning of each iteration t， the server first sends the current global model w_t to the client.Client i computes the gradient g_t of its loss $f(D_i, w)$ with respect to w_t and sends g_i^t back to the server, where g_i^t is the model update from client i at iteration t .

$$g(w_{t+1}) \approx g(w_t) + \nabla g(w_t) \cdot (w_{t+1} - w_t)$$

Experiments

> Experimental Setup

BFL settings：

> 20 nodes (including 16 workers, 4 agents, and 2 super agents)

> MNIST, FEMNIST and CIFAR-10

Evaluation metrics

> Detection Accuracy (DACC)

> Test Accuracy (TACC)

Attack settings

> Label Flip Attack

> Gaussian Noise Attack

> Local Model Poisoning Attack

Comparison methods：

> ABFL

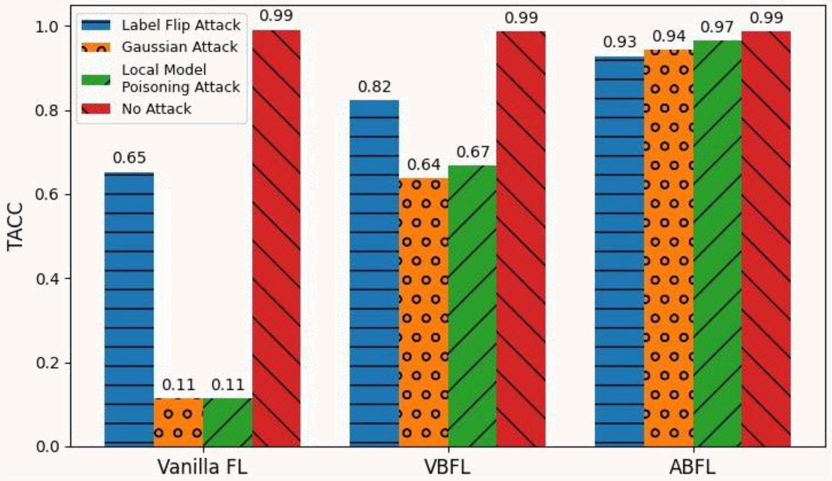
> VBFL

> Vanilla FL

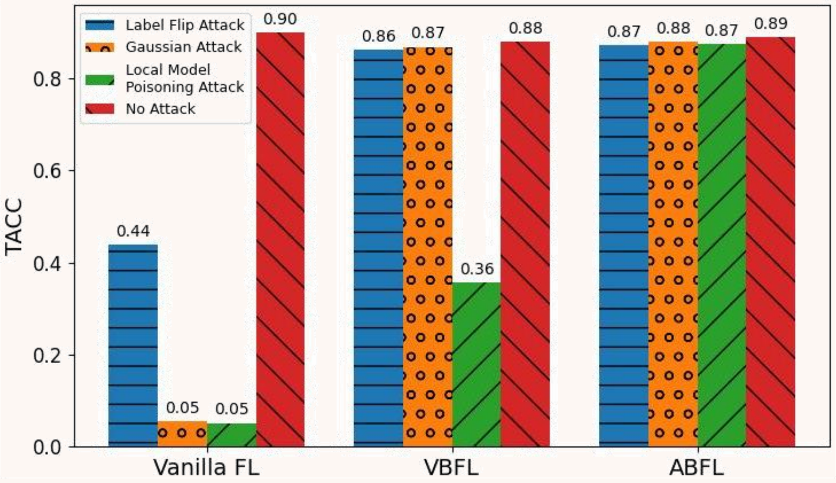
[VBFL] Hang Chen, Syed Ali Asif, Jihong Park, Chien-Chung Shen, and Mehdi Bennis. Robust blockchained federated learning with model validation and proof of-stake inspired consensus. 2021
[Vanilla FL] Paritosh Ramanan and Kiyoshi Nakayama. Baffle: Blockchain based aggregator free federated learning. 2020

Experiments

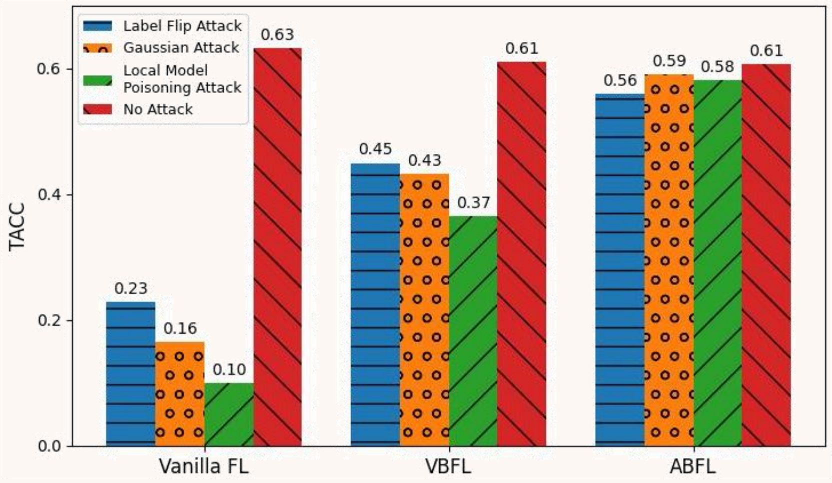
> Performance of the global models



(a) MNIST



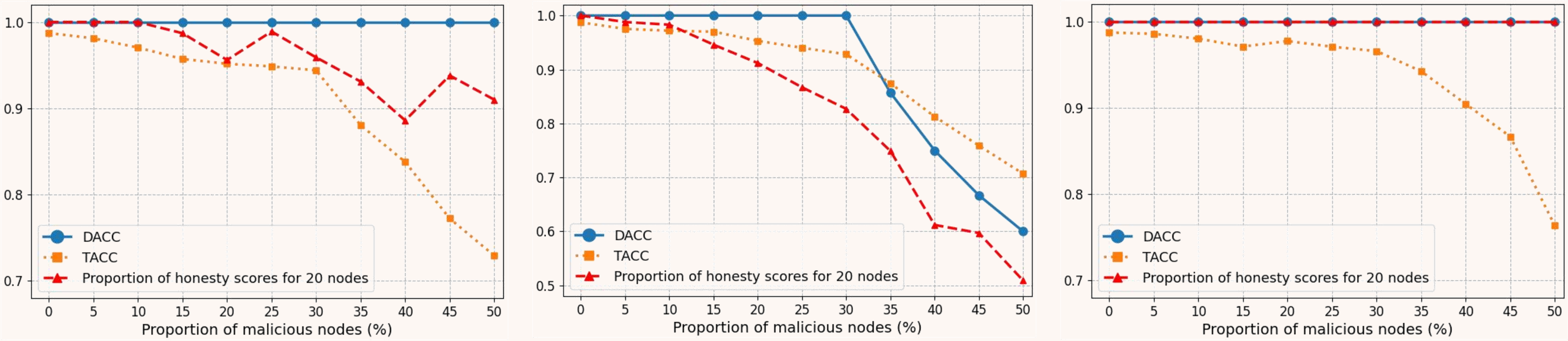
(b) FEMNIST



(c) CIFAT-10

Experiments

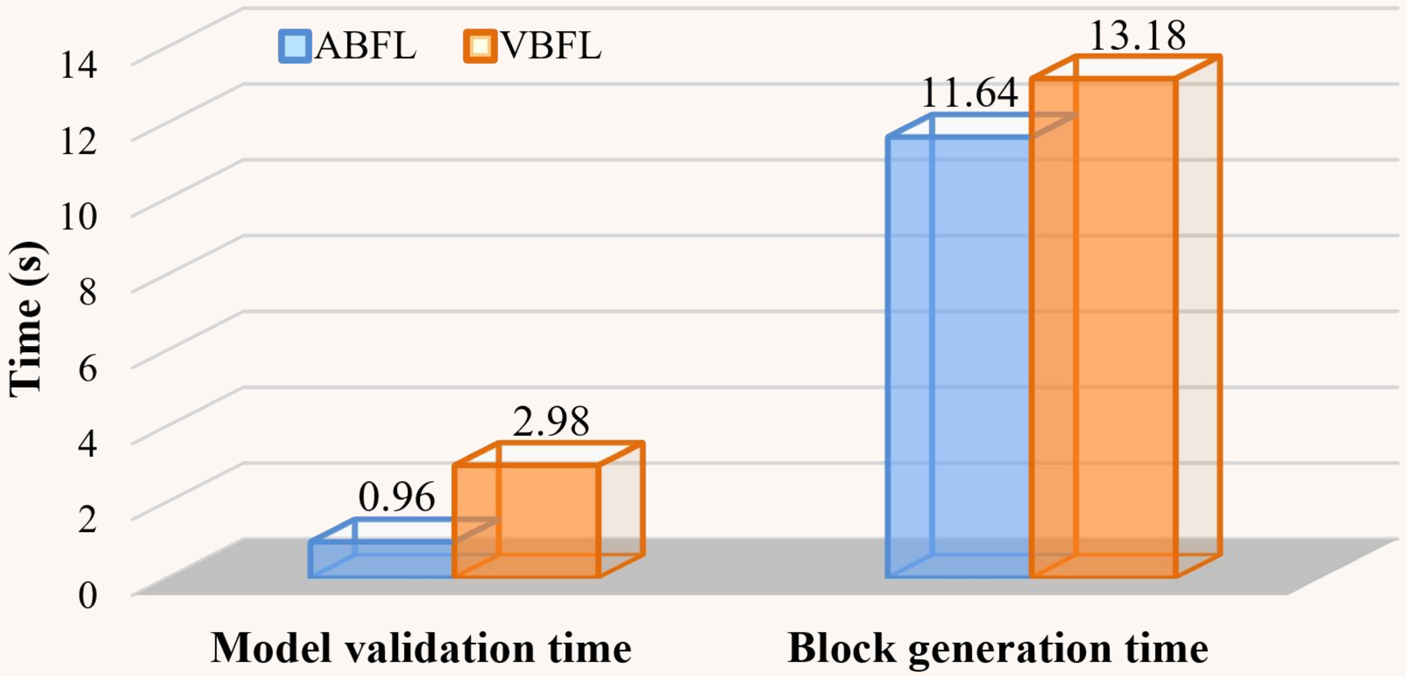
> Impact of the proportion of malicious nodes



(a) Gaussian Attack (b) Label Flip Attack (c) Local Model Poisoning Attack

Experiments

> Consensus efficiency



The average of model validation time and individual block generation time in 500 training rounds

Conclusion

Our Work：

> We propose a blockchain-based FL framework, ABFL, which defines in detail the training process and an efficient agent consensus mechanism.

> We propose a dual-benchmark robustness algorithm based on cosine similarity to identify malicious nodes by checking the consistency of model updates.

> We perform a comprehensive evaluation of the proposed ABFL framework on three benchmark datasets using various advanced poisoning attack methods to demonstrate the resilience of ABFL to various poisoning attacks, as well as the ability to maintain high model performance and improved consensus efficiency.

Future...

> Vertical Federated Learning

> Asynchronous Federated Learning

> ...

Thanks！

Presenter: Tianyu Mei <tiany.mei@gmail.com>