Securing Federated Learning Systems: A Case Study of Tokenized Incentive Mechanisms

Olamide T. Tawose, Ph.D.
Assistant Professor,
Department of Computer Science

Amount Requested: $7,000

2/23/2024

1. Description

The purpose of this grant is to provide summer faculty funding to enable me to carry out a research project in my area of expertise i.e., distributed computing. Federated learning (FL) has emerged as a new distributed computing paradigm to both enrich the available training data and protect the data privacy of participating clients. Due to the critical importance of client participation leading to the success of FL systems, multiple incentive mechanisms have been proposed to attract and retain clients in FL; in particular, a tokenized incentive was recently proposed (Han et al., 2022), which was believed more practical than the existing monetary-based, offline incentive mechanisms. However, this work aims to demonstrate that, under mild assumptions, the tokenized incentive mechanism for FL systems can be effectively compromised by a fraction of colluded clients who share their local training models with deliberate Gaussian noises. Hence, we would design an effective detection protocol to enhance the fairness and reliability of tokenized incentive mechanism for federated learning systems.

2. Measurable Goals and Objectives for the Project

Goal 1: Investigate Model-Poisoning Attacks in Federated Learning

Objective 1: Develop and implement Gaussian model-poisoning attacks targeting federated learning systems.

- *Key Result*: Successfully integrate and execute the proposed attacks on federated learning models across multiple datasets and aggregation algorithms.

Objective 2: Evaluate the effectiveness of the model-poisoning attacks in degrading model performance.

- *Key Result*: Measure the impact of the attacks on key performance metrics such as accuracy, convergence rate, and loss function compared to baseline performance.

Goal 2: Propose a Blockchain-Based Auditing Protocol

Objective 1: Design a blockchain-based auditing protocol for tracking local models submitted by clients in federated learning systems.

- *Key Result:* Develop a detailed protocol specification outlining the data structure, consensus mechanism, and auditing procedures.

Objective 2: Implement and test the auditing protocol in simulated federated learning environments.

- *Key Result:* Deploy the auditing protocol to monitor and detect suspicious activities, such as model poisoning attacks, during federated learning experiments.

Goal 3: Evaluate the Proposed Solutions

Objective 1: Assess the detection accuracy and false positive rate of the auditing protocol.

- *Key Result:* Quantify the protocol's performance in accurately identifying and flagging anomalous behavior without generating excessive false alarms.

Objective 2: Measure the training overhead incurred by integrating the auditing protocol into federated learning systems.

- *Key Result:* Evaluate the computational resources, communication bandwidth, and training time required with and without the auditing protocol.

3. Timeline

May: Planning and Preparation

- Review relevant literature on federated machine learning, model poisoning attacks, and auditing protocols.
- Develop a detailed project plan including research methodology and experimentation design.
- Set up the necessary infrastructure for conducting experiments, including software environments, datasets, and blockchain frameworks.

June: Experimentation and Analysis

- Begin conducting experiments with the proposed Gaussian model-poisoning attacks on federated learning systems using selected datasets and aggregation algorithms.
- Collect experimental data and evaluate the effectiveness of the attacks in different scenarios.
- Implement and test the auditing protocol for tracking local models submitted by clients.
- Analyze experimental results, including the impact of the attacks on model performance and the effectiveness of the auditing protocol in detecting suspicious activities.

July: Reporting and Documentation

- Compile and organize experimental results, including any insights or observations.
- Draft the project report, including an introduction, methodology, results, discussion, and conclusion.
- Review and revise the project report based on feedback from peers.
- Finalize the project report, prepare any supplementary materials (e.g., presentations, documentation), and submit the project deliverables.

4. How the Project Will Enhance Teaching and Research at Lincoln University

The proposed project has significant potential to contribute to both teaching and research at Lincoln University in several ways:

Cutting-edge Research Contribution: This project delves into the intersection of federated machine learning and security. By investigating novel model-poisoning attacks and proposing an auditing protocol, the project offers new insights and solutions to pressing challenges in the field. Such cutting-edge research can enhance the university's reputation as a hub for innovation and expertise in emerging technologies.

Curriculum Enrichment: The findings and methodologies developed throughout the project can enrich teaching curricula across various disciplines. Faculty members can incorporate the project's outcomes into courses related to machine learning, cybersecurity, and data science. Students will benefit from learning about real-world applications and challenges, preparing them for careers in industries where federated learning and cybersecurity are increasingly crucial.

External Engagement and Funding Opportunities: Successful completion of the project can attract attention from industry partners, government agencies, and funding bodies interested in advancing research and innovation in federated learning security. Collaborative projects with external stakeholders not only provide additional resources and expertise but also offer opportunities for technology transfer and knowledge exchange, benefiting both academia and industry.

5. Measured Success and Shared Outcomes

The outcomes of the project can be measured through several key metrics and indicators, including:

Effectiveness of Model Poisoning Attacks: The success of the proposed Gaussian model-poisoning attacks can be measured by evaluating their impact on the performance and integrity of federated learning models. Metrics such as accuracy, convergence rate, and loss function can be used to assess the extent of model degradation caused by the attacks compared to baseline performance without attacks.

Detection Accuracy of Auditing Protocol: The performance of the blockchain-based auditing protocol in detecting suspicious activities and mitigating model poisoning attacks can be measured by evaluating its detection accuracy, false positive rate, and false negative rate. The protocol should effectively identify anomalous behavior and trigger appropriate responses to maintain the security and integrity of federated learning systems.

Training Overhead: The overhead incurred by integrating the blockchain-based auditing protocol into federated learning systems can be measured in terms of computational resources, communication bandwidth, and training time. Comparing the training overhead with and without the auditing protocol provides insights into its efficiency and scalability.

Generalization Across Datasets and Algorithms: The generalizability of the proposed attacks and auditing protocol across different datasets and federated learning aggregation algorithms can be assessed by conducting experiments on multiple datasets and varying aggregation strategies. Consistent performance across diverse scenarios demonstrates the robustness and applicability of the proposed solutions.

6. Form of Final Presentation

We propose to deliver our findings as required by the proposal call in the form of an oral presentation at an officially scheduled department or university meeting. We also propose to submit a project and summary report to the FRCC chair and CETL, respectively. In addition, we plan to submit our findings to reputable conferences within the field of machine learning.

Budget

I am only requesting the maximum summer salary of $7,000.

References

1. Jingoo Han, Ahmad Faraz Khan, Syed Zawad, Ali Anwa, Nathalie Baracaldo Angel, Yi Zhou, Feng Yan, and Ali R. Butt. Tokenized incentive for federated learning. In Proceedings of the AAAI Conference on Artificial Intelligence, 2022.