Efficient Client Selection in Federated Learning

William Marfo, Deepak K. Tosh, Shirley V. Moore

Department of Computer Science, *University of Texas at El Paso*, El Paso, USA
wmarfo@miners.utep.edu, dktosh@utep.edu, svmoore@utep.edu

2025 IEEE 22nd Consumer Communications & Networking Conference (CCNC) | 979-8-3315-0805-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/CCNC54725.2025.10976004

***Abstract*—Federated Learning (FL) enables decentralized machine learning while preserving data privacy. This paper proposes a novel client selection framework that integrates differential privacy and fault tolerance. The adaptive client selection adjusts the number of clients based on performance and system constraints, with noise added to protect privacy. Evaluated on the UNSW-NB15 and ROAD datasets for network anomaly detection, the method improves accuracy by 7% and reduces training time by 25% compared to baselines. Fault tolerance enhances robustness with minimal performance trade-offs.**
[1]

***Index Terms*—Federated learning, Client selection, Distributed machine learning**

I. Introduction

Federated Learning (FL) enables distributed machine learning without centralizing data, addressing critical privacy concerns [1]. However, efficient client selection and robust privacy preservation mechanisms remain key challenges in FL [2]. Poor client selection can degrade model performance, while insufficient privacy safeguards may expose sensitive information. Moreover, heterogeneity in client data and system resources can lead to bottlenecks, slowing down training. In this paper, we propose an efficient client selection method that integrates differential privacy (DP) and fault tolerance mechanisms to enhance model performance and system resilience. The adaptive selection process adjusts the number of clients dynamically based on performance and system constraints, with added noise ensuring data privacy [2]. To evaluate our framework, we apply it to the UNSW-NB15 and ROAD datasets for network anomaly detection [3], [4], demonstrating significant improvements in accuracy and training efficiency compared to baselines.

II. Proposed Client Selection Method

We propose a client selection method for FL that balances accuracy, privacy, and fault tolerance. The method selects a subset of clients based on their potential contribution to the global model while incorporating differential privacy (DP) and fault tolerance mechanisms. Our approach has two main components: (1) an adaptive client selection algorithm and (2) the integration of DP and checkpointing.

In each round, available clients A_t are evaluated based on utility scores, which are computed using factors such as data quality and computational capacity. The top K clients are selected to train local models. Differential privacy is ensured by adding Gaussian noise to the model updates, controlled by the privacy budget ϵ, to prevent the server from inferring individual client data during aggregation. To enhance fault tolerance, a checkpointing mechanism, with intervals determined by t_c^*, allows clients to save and recover from failures during training. Algorithm 1 presents the proposed method, outlining the client selection, privacy protection, and fault tolerance procedures.

Algorithm 1 Client selection with differential privacy and fault tolerance

Require: Set of all clients C, number of clients to select K, privacy budget ϵ, checkpointing interval t_c^*
Ensure: Selected subset of clients S_t for each round t

```
Initialize utility scores for all clients
for each round t do
    Get available clients A_t
    Compute utility scores for each client in A_t
    Select top K clients based on utility scores
    for each selected client do
        Train local model, apply gradient clipping
        Add Gaussian noise to gradients for DP
        Send noisy gradients to the server
        if checkpoint interval reached then Save checkpoint
        end if
        if client failure detected then Recover from checkpoint
        end if
    end for
    Aggregate updates and update global model
end for
```

III. Performance Evaluation

A. Experimental Setup

We used two widely recognized network security datasets: the UNSW-NB15 [3] and ROAD [4]. The experiments were conducted on a system with a 12th Gen Intel Core i9-12900HK, NVIDIA RTX 3080 Ti GPU, and 32GB of RAM, using `Python 3.8`, `TensorFlow 2.6.0`, and `PyTorch 0.5.0`. For baseline comparisons, we used: - **ACFL** [5]: an active learning-based client selection method. - **FedL2P** [6]: a meta-learning approach for personalized fine-tuning.

Evaluation metrics include accuracy and AUC-ROC, which are standard for anomaly detection tasks.

[1]This material is based upon work supported by the United States Department of Energy's (DOE) Office of Fossil Energy (FE) Award DE-FE0031744.

B. Results and Analysis

1) Performance Comparison with Baselines

We evaluated the performance of our method against two baseline approaches, ACFL and FedL2P. As depicted in Fig. 1, our proposed method demonstrates a 7% improvement in accuracy and a 25% reduction in training time compared to the baselines. These gains are particularly pronounced in the ROAD dataset, where our method effectively handles complex anomaly patterns, showcasing its robustness in diverse network conditions.

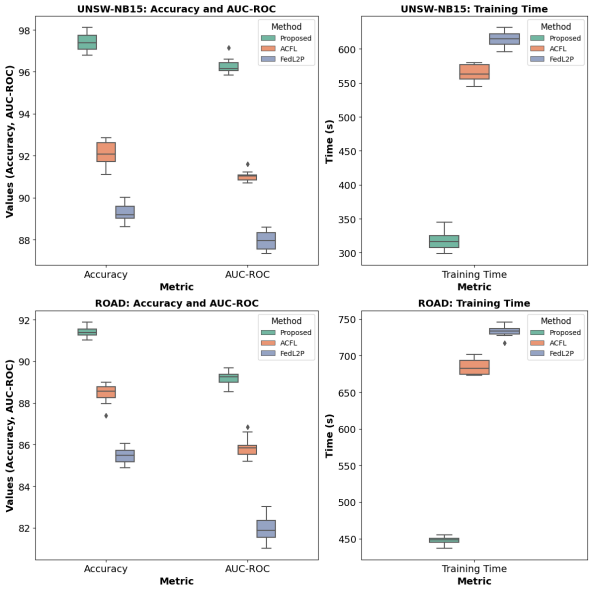

Fig. 1: Performance comparison of the proposed method, ACFL, and FedL2P in terms of accuracy, AUC-ROC, and training time.

2) Impact of Differential Privacy

We analyzed the effect of different privacy budgets (ϵ) on model performance. Fig. 2 shows that increasing the privacy budget leads to improved accuracy and reduced loss across both datasets. For instance, in the UNSW-NB15 dataset, accuracy increased from 86% at $\epsilon = 10$ to 89% at $\epsilon = 100$, while similar trends were observed in the ROAD dataset, where accuracy improved from 73% to 82%. These results demonstrate the balance between maintaining privacy and ensuring model performance.

3) Effect of Fault Tolerance

We evaluated the impact of introducing fault tolerance mechanisms through checkpointing. Table I shows that while accuracy and AUC-ROC experienced a slight decline (approximately 2-3%), the overall system robustness improved by effectively handling client dropouts. Training time increased by around 5-10%, but this is an acceptable trade-off given the enhanced system reliability.

IV. Conclusion

We proposed an efficient client selection method for FL that integrates differential privacy and fault tolerance, validated on network anomaly detection tasks. Our method improved accuracy by 7% and reduced training time by 25% compared to FedL2P, demonstrating a balance between privacy, performance, and robustness. Future work will explore adaptive hyperparameter tuning and other privacy-preserving techniques.

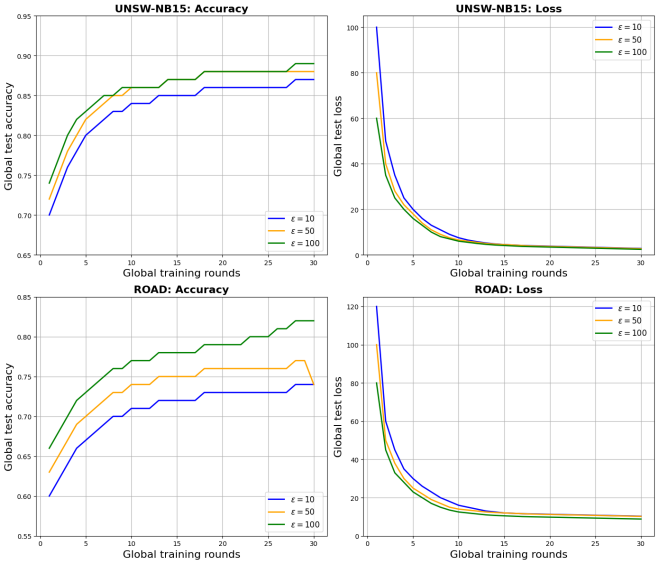

Fig. 2: Impact of privacy budgets on accuracy and loss for UNSW-NB15 and ROAD datasets.

TABLE I: Impact of fault tolerance on model performance (UNSW-NB15 and ROAD).

Configuration	Accuracy (%)	AUC-ROC	Training Time (s)
UNSW-NB15			
Without Fault Tolerance	94.8	0.93	570
With Fault Tolerance	92.1	0.91	600
ROAD			
Without Fault Tolerance	90.3	0.88	680
With Fault Tolerance	88.7	0.86	710

References

[1] W. Marfo, D. K. Tosh, and S. V. Moore, "Network anomaly detection using federated learning," in *MILCOM 2022-2022 IEEE Military Communications Conference (MILCOM)*. IEEE, 2022, pp. 484–489.

[2] Q. Li, X. Li, L. Zhou, and X. Yan, "Adafl: Adaptive client selection and dynamic contribution evaluation for efficient federated learning," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 6645–6649.

[3] N. Moustafa and J. Slay, "Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set)," in *2015 military communications and information systems conference (MilCIS)*. IEEE, 2015, pp. 1–6.

[4] M. E. Verma, R. A. Bridges, M. D. Iannacone, S. C. Hollifield, P. Moriano, S. C. Hespeler, B. Kay, and F. L. Combs, "A comprehensive guide to can ids data and introduction of the road dataset," *PLoS one*, vol. 19, no. 1, p. e0296879, 2024.

[5] G. Yan, H. Wang, X. Yuan, and J. Li, "Criticalfl: A critical learning periods augmented client selection framework for efficient federated learning," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 2898–2907.

[6] R. Lee, M. Kim, D. Li, X. Qiu, T. Hospedales, F. Huszár, and N. Lane, "Fedl2p: Federated learning to personalize," *Advances in Neural Information Processing Systems*, vol. 36, 2024.