ADScAI
Applied Data Science & Artificial Intelligence Symposium

Crowdsourcing in Federated Learning: A Scalable Approach to Collaborative Model Training

D.V. S. Rashmika
School of Computing
Informatics Institute of Technology
Colombo, Sri Lanka
sandaru.20210334@iit.ac.lk

Guhanathan Poravi
School of Computing
Informatics Institute of Technology
Colombo, Sri Lanka
guhanathan.p@iit.ac.lk

Keywords—Federated Learning, Crowdsourcing, Model Aggregation, Differential Privacy, Perturbation Mechanism

I. Introduction

Mobile crowdsourcing, powered by advancing mobile technologies, offers efficient data management but poses privacy challenges. **Federated Learning (FL)** mitigates this by enabling joint model training while keeping data on-device [1]. However, sharing model parameters during training still risks data reconstruction by adversaries [2,3].

Differential Privacy (DP) is often used to protect FL models, yet its uniform deployment across heterogeneous clients can negatively impact performance, especially considering clients' heterogeneous privacy needs. The extended abstract explores two significant issues:

(1). Trade-off between individualized privacy preservation and model utility, and

(2). Constructing efficient perturbation mechanisms with minimal performance degradation.

II. Literature Review

Federated Learning (FL), suggested by McMahan et al. [1], keeps data local and thereby preserves privacy. Attackers can, however, reconstruct data based on model parameters, as demonstrated by Zhu et al. [2] and Geiping et al. [3]. There are various solutions proposed to deal with this issue, i.e., homomorphic encryption [4-6], secure multi-party computation [7], and differential privacy [8-10].

Most federated learning frameworks use an overall privacy budget, thus ignoring client diversity. Jorgensen et al. [11] proposed Personalized Differential Privacy (PDP) to enable client-specific privacy settings; yet, this often ignores model accuracy. Zhao et al. [12] suggested adapting noise scaling based on accuracy variations, and Fu et al. [13] utilized sequential loss reductions. These methods are based on overall statistics, opening the door to server tampering.

III. Materials and Methods

This research presents a Crowdsourced Federated Learning Framework with Personalized Privacy Preservation that includes two primary components:

(1). a two-stage dynamic game to decide personalized privacy budgets, and

(2). a differential privacy **Perturbation Mechanism based on Weight Priority** (PMWP).

A. System Architecture

The system consists of three vital components: task requestor, platform, and clients. The task requestor initiates federated learning activities and also reimburses clients for contributing to model training. The platform acts as the aggregation server, while clients hold data assets and participate in local training activities.

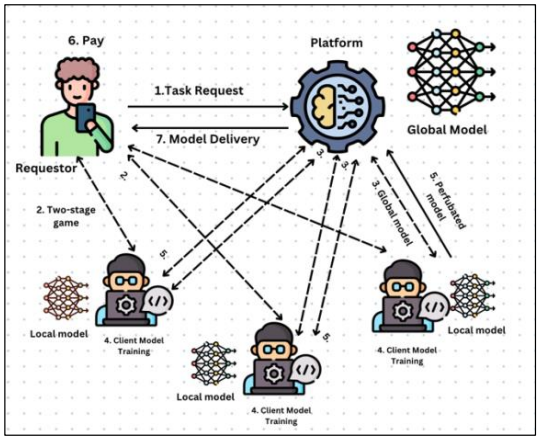

Fig. 1. The architecture of the proposed Crowdsourced Federated Learning Framework with Personalized Privacy Preservation

B. Two-Stage Dynamic Game

Under a Stackelberg game model, leaders (task requestors) set payment rewards, and followers (clients) choose their privacy budgets independently. This guarantee balancing the privacy needs of clients with the FL model's performance and cost-reduction needs. Equilibrium analysis yields a unique and stable Nash equilibrium for privacy budget allocation.

C. Differential Privacy Perturbation Mechanism based on Weight Priority (PMWP)

The authors propose the PMWP mechanism, which adds differential privacy noise selectively to improve privacy while ensuring that the DP noise does not heavily degrade the performance of models. Hence, we classify model weights into strongly correlated and weakly correlated

PMWP considers the relevance between model weights and data. The authors organize and analyze the relevance, distinguishing between weights as strongly related and weakly related We retain for model performance maintenance and add noise to for privacy preservation.

https://doi.org/10.31705/ADScAI.2025.45

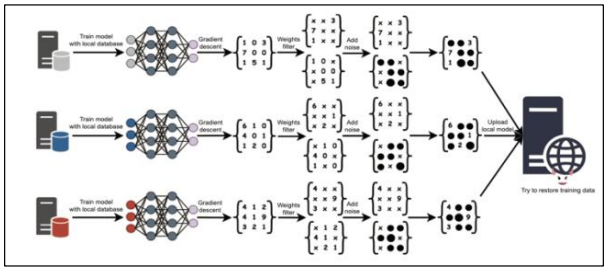

Fig. 2. Federated learning with improved DP with PMWP

IV. Results and Discussion

The authors have evaluated the proposed approach on MNIST and CIFAR10 datasets. For experiments, the authors have used a three-layer CNN for MNIST and VGG11 for CIFAR10. The authors have compared their method with No-DP (without differential privacy) and PFLF.

For MNIST, the authors used a three-layer CNN with the SGD optimizer (learning rate = 0.1, batch size = 64, 20 global rounds, 100 local epochs). For CIFAR10, the VGG11 model was trained under the Adam optimizer (learning rate = 0.01, batch size = 64, 30 global rounds, 200 local epochs). Gradient clipping with a value of 1.0 was utilized, and differential privacy noise was added based on a filter proportion (θ) of 0.4 and an activation threshold (ω) of 0.5.

A. Nash Equilibrium in Two-stage Game

Figure 3 shows the utility functions for clients with different privacy costs. The concave nature confirms the existence of optimal privacy budgets for each client and utility is also concave concerning payment, confirming a unique Nash equilibrium.

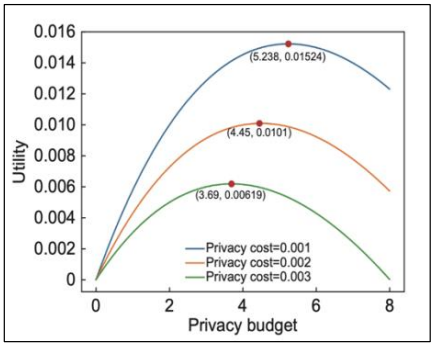

Fig. 3. Utility functions for clients with different privacy costs

B. Performance of PMWP

The below table summarizes the accuracy of different methods. Our PMWP approach consistently outperforms PFLF across all datasets, achieving accuracy closer to No-DP.

TABLE I. Model Accuracy Comparison

Dataset	No-DP	PFLF	PMWP
IID MNIST	0.985	0.961	0.971
Non-IID MNIST	0.967	0.924	0.946
IID CIFAR10	0.707	0.597	0.633
Non-IID CIFAR10	0.614	0.536	0.568

On MNIST, the authors' approach has negligible performance degradation compared with No-DP. The performance difference is larger on CIFAR10 because of its more complex features, where weight perturbation influences feature representation more significantly.

V. Challenges and Limitations

Experimental verification is restricted to two popular benchmarking datasets, MNIST and CIFAR10, which hinders its generalization to real-world data. Furthermore, the two-stage game mechanism and personalized noise injection may lead to increased computational and communication overhead, thereby bringing scalability issues. The use of a semi-trusted server also compromises security in fully adversarial settings.

VI. Conclusion

This extended abstract introduces a new architecture for crowd-sourced federated learning with personalized privacy preservation. Our two-stage dynamic game enables clients to manage the levels of privacy preservation independently while preventing excessive requirements that compromise model usability. The PMWP mechanism cleverly applies varying degrees of perturbation according to the significance of weights to data, thereby ensuring better model performance under identical privacy constraints.

References

[1] H. B. McMahan et al., "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artificial Intelligence and Statistics, 2017, pp. 1273–1282.

[2] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in Proc. 33rd Conf. Neural Information Processing Systems, 2019.

[3] J. Geiping et al., "Inverting Gradients: How easy is it to break privacy in federated learning?" arXiv preprint arXiv: 2003.14053, 2020.

[4] G. Xu et al., "Privacy-preserving federated deep learning with irregular users," IEEE Trans. Dependable Secure Comput., vol. 19, no. 2, pp. 1364–1381, 2022.

[5] L. Zhang et al., "Homomorphic encryption-based privacy-preserving federated learning in IoT-enabled healthcare system," IEEE Trans. Netw. Sci. Eng., vol. 10, no. 5, pp. 2864–2880, 2023.

[6] Z. Jiang, W. Wang, and Y. Liu, "FLASHE: additively symmetric homomorphic encryption for cross-silo federated learning," arXiv preprint arXiv: 2109.00675, 2021.

[7] C. Zhang et al., "Augmented multi-party computation against gradient leakage in federated learning," IEEE Trans. Big Data, DOI: 10.1109/TBDATA.2022.3208736.

[8] S. Weng et al., "Privacy-preserving federated learning based on differential privacy and momentum gradient descent," in Proc. Int. Joint Conf. Neural Networks, 2022, pp. 1–6.

[9] L. Sun, J. Qian, and X. Chen, "LDP-FL: Practical private aggregation in federated learning with local differential privacy," in Proc. 30th Int. Joint Conf. Artificial Intelligence, 2021, pp. 1571–1578.

[10] R. Hu et al., "Personalized federated learning with differential privacy," IEEE Internet Things J., vol. 7, no. 10, pp. 9530–9539, 2020.

[11] Z. Jorgensen, T. Yu, and G. Cormode, "Conservative or liberal? Personalized differential privacy," in Proc. IEEE 31st Int. Conf. Data Engineering, 2015, pp. 1023–1034.

[12] J. Zhao et al., "Utility optimization of federated learning with differential privacy," Discrete Dyn. Nat. Soc., vol. 2021, p. 3344862, 2021.

[13] J. Fu, Z. Chen, and X. Han, "Adap DP-FL: Differentially private federated learning with adaptive noise," in Proc. IEEE Int. Conf. Trust, Security and Privacy in Computing and Communications, 2022, pp. 656–663.

[14] H. Zhou et al., "PFLF: privacy-preserving federated learning framework for edge computing," IEEE Trans. Inf. Forensics Secur., vol. 17, pp. 1905–1918.