

Enhanced security in federated learning by integrating homomorphic encryption for privacy-protected, collaborative model training

Ganga Rama Koteswara Rao [#]

Department of Computer Science & Information Technology

Koneru Lakshmaiah Education Foundation

Vaddeswaram

Andhra Pradesh

India

Hayder M. A. Ghanimi ⁺

Department of Information Technology

College of Science

University of Warith Al-Anbiyaa

Karbala

Iraq

and

Department of Computer Science

College of Computer Science and Information Technology

University of Kerbala

Karbala

Iraq

V. Ramachandran [§]

Department of Computer Science and Engineering

GITAM School of Computing

GITAM (Deemed to be University)

Bengaluru 561203

Karnataka

India

[#] E-mail: grkraoganga@gmail.com

⁺ E-mail: hayder.alghanami@uowa.edu.iq

[§] E-mail: ramachandran.veer@gmail.com

Dokhyl Al-Qahtani [‡]

*Department of Electrical Engineering
College of Engineering
Prince Sattam bin Abdulaziz University
Al-Kharj 11942
Saudi Arabia*

Pankaj Dadheech [@]

*Department of Computer Science & Engineering
Swami Keshvanand Institute of Technology, Management & Gramothan (SKIT)
Jaipur 302017
Rajasthan
India*

Sudhakar Sengan ^{*}

*Department of Computer Science and Engineering
PSN College of Engineering and Technology
Tirunelveli 627152
Tamil Nadu
India*

Abstract

A significant novel approach in distributed ML, Federated Learning (FL), enables multiple parties to work simultaneously on developing models while securing the confidentiality of their unique datasets. There are issues regarding privacy with FL, particularly for models that are being trained, because private information can be accessed from shared gradients or updates to the model. This investigation proposes SecureHE-Fed, a novel system that improves FL's defense against attacks on privacy through the use of Homomorphic Encryption (HE) and Zero-Knowledge Proofs (ZKP). Before data from clients becomes involved in the learning procedure, SecureHE-Fed encrypts it. The following lets us determine encrypted messages without revealing the data as it is. As an additional security test, ZKP is employed to verify if modifications to models are valid without sharing the true nature of the information. By evaluating SecureHE-Fed with different FL techniques, researchers demonstrate that it enhances confidentiality while maintaining the precision of the model. The results of this work obtained validate SecureHE-Fed as a secure and scalable FL approach, and we recommend its use in applications where user confidentiality is essential.

Subject Classification: 68N30.

Keywords: Homomorphic encryption, Federated learning, ZKP, Privacy, Security, Machine learning.

[‡] E-mail: dm.alqahtani@psau.edu.sa

[@] E-mail: pankajdadheech777@gmail.com

^{*} E-mail: sudhasengan@gmail.com (Corresponding Author)

1. Introduction

With the increasing availability of large amounts of data, Machine Learning (ML) has evolved into an innovative area of research. Improvements in industries that include such as medical care, financial services, and autonomous vehicles have evolved from having the ability to reach enormous amounts of data. Security, such as confidentiality of data and information monopolization, is a significant issue, considering ML models have traditionally employed centrally managed data storage and processing. Learning models that emphasize user security are of paramount importance in the modern day, where the privacy of information is a fundamental right. A technique that is decentralized and allows training ML models across multiple nodes or users while maintaining data local to each node is Federated Learning (FL). Sharing data is especially significant when circumstances where rules about confidentiality or practical restrictions prevent it. One possible approach for privacy-conscious learning is FL, which analyzes data privately and only communicates modifications to models to a centralized server for storage. This does not imply that FL is susceptible to risks, considering the fact that it is decentralized. There is a threat of data extraction and privacy violations due to the improper cybersecurity of the shared model updates, even though they do not contain data in its raw form. A new architecture termed SecureHE-Fed is put forward in this article with the objective of improving FL's privacy and security features. Through the application of Homomorphic Encryption (HE), SecureHE-Fed secures user data before model update participation, allowing computation on data that has been encrypted without compromising the original data. Resolving the crucial problem of data confidentiality, this approach ensures that private inputs to the global system will keep their identities secret.

Also, SecureHE-Fed's Zero-Knowledge Proofs (ZKP) implementation enables additional security by allowing the accuracy of encrypted updates to be confirmed without compromising the contents of the information itself. By ensuring that the server can verify modifications to models without compromising privacy, this approach represents an essential step in securing FL identity. The requirement for adequate information security continues to increase in importance as FL evolves rapidly. In order to protect FL from risks to its privacy, SecureHE-Fed provides an acceptable approach and promotes additional research in the context. The current study intends to provide evidence that SecureHE-Fed works efficiently by

carrying out comprehensive tests and analyses; by doing so, it will define an accepted norm for security in FL environments.

The paper is organized as follows: the literature reviewed is presented in section 2, the problem is defined in section 3, section 4 offers the system model, section 5 presents the experiment analysis, and section 6 concludes the work.

2. Literature Review

The interplay between data privacy and utility in ML is a pivotal concern in the AI community. Recent studies have addressed this through FL, which promises to harness sensitive data without compromising privacy. [1] offer a comprehensive analysis of privacy in collaborative ML, emphasizing the role of Privacy Enhancing Technologies (PETs) and providing a selection guideline for their use in collaborative environments. Expanding on this, [2] introduces a privacy-preserving framework employing partially HE within FL. Their framework, PFMLP, maintains model accuracy while enhancing privacy, showcasing the viability of HE in practical applications. [3] present a novel algorithm, HEFL-LDP, combining semi-homomorphic encryption with local differential privacy to protect user data in FL. Their method addresses the balance between privacy, efficiency, and accuracy, proposing a utility optimization approach to mitigate accuracy decline. [4] tackle data scarcity in fault diagnosis with a vertical FL model that preserves data privacy through HE, demonstrating the model's effectiveness through industrial case studies. The study by [5] developed a secure credit data-sharing architecture that optimizes privacy within a FL and blockchain framework. Their approach is validated through experiments and is crucial for financial stability and economic recovery post-pandemic—SecureHE-Fed builds on this foundation, integrating advanced cryptographic techniques for secure data sharing in collaborative learning environments [6].

3. Problem Definition

In FL, the privacy risk arises from the potential exposure of sensitive data during model training. This is especially pertinent in the context of model updates ΔM_i from clients $\{C_1, C_2, \dots, C_n\}$, where $\Delta M_i = f(D_i, M_i^{prev})$. Here, f is the training algorithm, D_i is the local dataset, and M_i^{prev} is the model's previous state. The challenge is that ΔM_i could be used to infer details about D_i . Further incorporating HE into FL aims to encrypt ΔM_i

for secure aggregation $(E(\Delta M_i) \oplus E(\Delta M_j) = E(\Delta M_i + \Delta M_j))$ without exposing D_i . The integration must address computational efficiency and algorithm compatibility and, mainly, operate without a centralized trusted key authority. This necessitates a decentralized approach for key management within the SecureHE-Fed framework, ensuring intrinsic data and model security throughout the process [7].

4. System Model for SecureHE-Fed with ZKP Integration

The SecureHE-Fed system model is a framework (Figure 1) designed to enhance the security and privacy of FL by integrating HE and ZKP. This model focuses on preserving the confidentiality of data while ensuring the integrity and validity of the model training process in a FL environment. The System Components are listed below:

- 1 *Clients (Data Owners)*: The system comprises a set of clients $\{C_1, C_2, \dots, C_n\}$, each holding a private dataset D_i . These clients are responsible for computing model updates from their data, typically represented as ΔM_i , where M_i is the local model trained on D_i .
- 2 *Central Server (Aggregator)*: The system's central server aggregates client model updates, ensuring data privacy by preventing access to raw data.
- 3 *HE*: In order to protect model updates during transmission and aggregation, the HE scheme is used for computation on encrypted data that the server cannot access. The encryption of a model update ΔM_i is denoted as $E(\Delta M_i)$.
- 4 *ZKP*: ZKP is integrated to improve model update integrity and validity. This system allows the central server to verify that each encrypted model update $E(\Delta M_i)$ is correctly computed and does not contain any malicious or incorrect data without the need to decrypt it. The verification process involves challenges and responses to ensure data validity, adhering to the ZKP principle, EQU (1).

$$R = f(C, S) \quad (1)$$

Here, R is the response to the challenge C , and S is the secret (the correctness of the encrypted model update) that the prover (client) demonstrates without revealing.

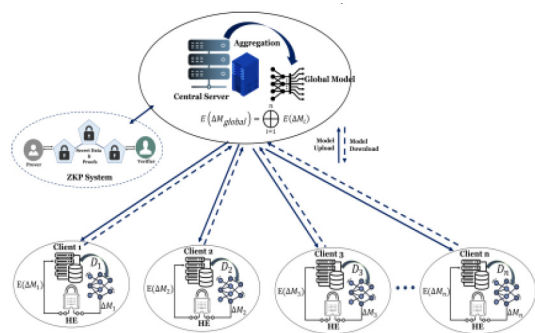


Figure 1
SecureHE-Fed system [8]

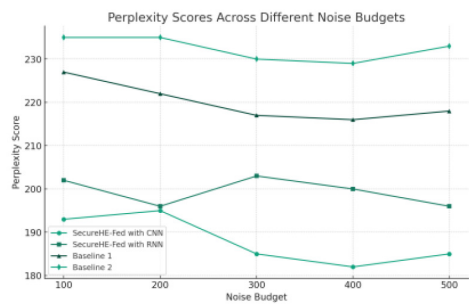


Figure 2
Analysis of perplexity scores

The operational process begins with each client training their local model on their dataset and computing the model update ΔM_i . These updates are then encrypted using the HE scheme and transmitted securely to the central server [9]. The prominent server aggregates encrypted updates using HE to maintain client data confidentiality. The ZKP system verifies the integrity of updates without decrypting them, ensuring data privacy (Figure 2). The system uses a decentralized key management approach to prevent malicious activity or data tampering during model training, reducing reliance on a trusted key distribution center and enhancing overall security [10].

5. Experimental Analysis

In evaluating the SecureHE-Fed system, we implemented it using Python 3.8 with the PySyft library for privacy-preserving aspects,

integrating the Python-Paillier library for homomorphic encryption functionalities. The structure of the simulated scenario was set up on a 4. NVIDIA Tesla V100 GPUs and a 2.50 GHz Intel Xeon Processor on an AWS EC2 server. Applying a convolutional neural network (CNN) and a feedforward neural network (FCN) as a base for the training of the model being suggested, this study selected the CIFAR-10 dataset to assess the accuracy of classification. Baseline 1 [2] and Baseline 2 [4] are initial models that have been contrasted with the model that was suggested. Research has been performed in order to evaluate the performance of the SecureHE-Fed system in a hypothetical FL scenario. The experiments addressed noise tolerance, accuracy of classification, and computational time.

Throughout 50 epochs, the mathematical models were trained, and to ensure reliability, the outcomes were calculated as an average across five training sessions. By testing FL models under multiple levels of cryptography noise tolerance, the SecureHE-Fed system was evaluated. The research effort examined two SecureHE-Fed implementations—one employing a CNN backbone and the other constructed with an RNN backbone—based on complexity scores, which assess the quality of the predictions made by the model. The baseline models were taken into account as well. In the current investigation, authors analyzed confusion scores across multiple signal cost levels in order to assess how encryption noise tolerance impacts the accuracy of models. Figure 2 shows that perplexity scores improved consistently across all models as the noise cost increased, indicating a higher tolerance for encryption noise. The SecureHE-Fed system with the CNN architecture showed notable performance, starting with a perplexity score of 193 at a noise cost of 100 and improving to 185 as the noise budget increased to 500. The RNN variant of SecureHE-Fed showed improvement but was slightly less effective than the CNN counterpart, with scores ranging from 202 to 196 over the same noise cost range. Both implementations outperformed baseline models.

Over the range of noise cost spectrum that was investigated, Baseline 1's scores decreased from 227 to 218 and Baseline 2's values from 235 to 233. In the context of performance measurements, this investigation demonstrates that the SecureHE-Fed system performs higher than the baseline models, especially while using a CNN backbone that can control encryption noise while maintaining or enhancing the precision of the model. The efficient functioning of the SecureHE-Fed design, which combines CNN and RNN backbones over 25 training phases, is shown in the accuracy of the visualization in Figure 3. Demonstrating CNN's

effectiveness in quickly collecting and analyzing features in a privacy-preserving FL context, the SecureHE-Fed platform with a CNN backbone demonstrates outstanding performance, rapidly improving and reaching a turning point. Also, SecureHE-Fed, which RNNs help, demonstrates good learning capabilities, although the enhancement in accuracy is delayed. This might be because RNNs are innately effective at identifying relationships between various periods in the data. While there is a slight advancement in accuracy with the baseline models, it is currently not good enough for success with the SecureHE-Fed system. Because of its resilience, the SecureHE-Fed system can overcome privacy-preserving challenges, implying its FL technique can keep or even enhance the performance of models without reducing data privacy. By employing neural network backbones, the diagram successfully shows how SecureHE-Fed design performs efficient learning in secure FL settings, given their high standards. Accuracy for different FL models as the total number of clients increases from 1 to 10, as can be seen previously in Figure 4. With ten users, the SecureHE-Fed with the CNN algorithm drops from an initial accuracy of 0.9 to a final accuracy of 0.695. The SecureHE-Fed with RNN

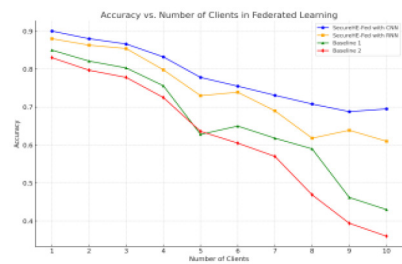


Figure 3

Analysis for Accuracy vs Epoch

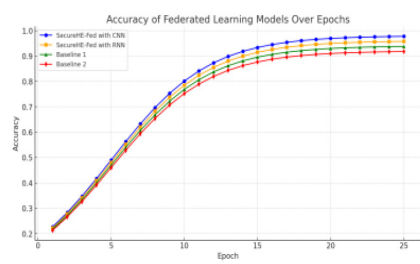


Figure 4

Accuracy analysis against clients

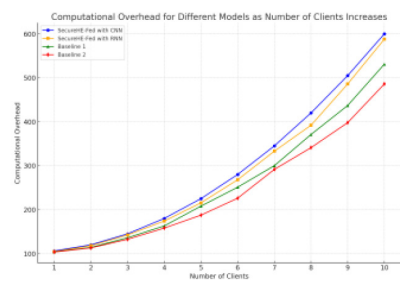


Figure 5

Computational Cost analysis

method exhibits an identical declining trend, beginning at 0.88 and closing at 0.61. Assuming ten users, Baseline 2 decreases from 0.83 to 0.36, while Baseline 1 starts at 0.85 and decreases to 0.43. Even though it can manage more users, the SecureHE-Fed model gradually loses its position in terms of accuracy when compared with baseline models. This could be because the framework is more qualified for dealing with heterogeneous user models or because aggregation techniques are more effective. Still, it would additionally imply that the framework can manage more heterogeneity in data from clients.

Figure 5 illustrates that the SecureHE-Fed simulations, CNN+RNN, have a higher computational cost than the baseline models. This is because SecureHE-Fed systems execute additional computing tasks, like those required to compute HE and ZKP. A significant growth in complexity and computational resources is demonstrated by the fact that the FL method's cost increases non-linearly with a growing number of users. Latency is highest for the SecureHE-Fed with CNN approach, which may be related to its higher computational requests and minimum for the Baseline 2 model, irrespective of the number of users.

6. Conclusion and Future Work

In order to solve the significant privacy concerns that result from distributed ML, the present research presented SecureHE-Fed, a state-of-the-art Federated Learning (FL) system that incorporates HE and ZKP. Comprehensive testing of the suggested approach demonstrates that SecureHE-Fed significantly enhances data privacy without decreasing the algorithm for learning accuracy. This model overcomes the information diffusion problem concerning traditional FL techniques by encrypting information during model development and employing ZKP to validate updates. When data privacy is more significant than the model's effectiveness in practical situations, the outcomes demonstrate that SecureHE-Fed can be an appropriate match. This design is suitable for implementing practical problems in FL contexts because it scales well with increasing client numbers and is robust to computational cost. If secure FL is to be commonly employed, additional investigation into optimizing encrypted computations and evidence procedures for verification is vital, and SecureHE-Fed lays the framework for development.

References

- [1] Soykan, E. U., et al., A survey and guideline on privacy enhancing technologies for collaborative machine learning. *IEEE Access*, 10, 97495-97519 (2022).
- [2] Fang, H., & Qian, Q., Privacy-preserving machine learning with homomorphic encryption and federated learning. *Future Internet*, 13(4), 94 (2021).
- [3] Zhao, J., et al., Local differentially private federated learning with homomorphic encryption. *The Journal of Supercomputing*, 1-31 (2023).
- [4] Xia, L., et al., Privacy-preserving gradient boosting tree: Vertical federated learning for collaborative bearing fault diagnosis. *IET Collaborative Intelligent Manufacturing*, 4(3), 208-219 (2022).
- [5] Yang, F., et al., Privacy-preserved credit data sharing integrating blockchain and federated learning for industrial 4.0. *IEEE Transactions on Industrial Informatics*, 18(12), 8755-8764 (2022).
- [6] S. Zeng, B. Mi, and D. Huang, "Emergency Vehicle Identification for Internet of Vehicles Based on Federated Learning and Homomorphic Encryption," IEEE 12th Data-Driven Control and Learning Systems Conference, Xiangtan, China, pp. 208-213 (2023).
- [7] S. Mohammadi et al., Optimized Paillier Homomorphic Encryption in Federated Learning for Speech Emotion Recognition," IEEE 47th Annual Computers, Software, and Applications Conference, Torino, Italy, pp. 1021-1022 (2023).
- [8] A. Pedrouzo-Ulloa et al., Practical Multi-Key Homomorphic Encryption for More Flexible and Efficient Secure Federated Average Aggregation," IEEE International Conference on Cyber Security and Resilience, Venice, Italy, pp. 612-617 (2023).
- [9] S. P. Sanon et al., Secure Federated Learning: An Evaluation of Homomorphic Encrypted Network Traffic Prediction," IEEE 20th Consumer Communications & Networking Conference, Las Vegas, NV, USA, pp. 1-6 (2023).
- [10] P. Yao, et al., Efficient Federated Learning Aggregation Protocol Using Approximate Homomorphic Encryption," 26th International Conference on Computer Supported Cooperative Work in Design, Rio de Janeiro, Brazil, pp. 1884-1889 (2023).