Secure Federated Learning with a Homomorphic Encryption Model

Yasmin Makki Mohialden[1], Nadia Mahmood Hussien[1], Saba Abdulbaqi Salman[2], Mohammad Aljanabi[2]

[1]Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq
[2]Department of Computer Science, College of Education, Al-Iraqia University, Baghdad, Iraq

Abstract

Federated learning (FL) offers collaborative machine learning across decentralized devices while safeguarding data privacy. However, data security and privacy remain key concerns. This paper introduces "Secure Federated Learning with a Homomorphic Encryption Model," addressing these challenges by integrating homomorphic encryption into FL. The model starts by initializing a global machine learning model and generating a homomorphic encryption key pair, with the public key shared among FL participants. Using this public key, participants then collect, preprocess, and encrypt their local data. During FL Training Rounds, participants decrypt the global model, compute local updates on encrypted data, encrypt these updates, and securely send them to the aggregator. The aggregator homomorphic ally combines updates without revealing participant data, forwarding the encrypted aggregated update to the global model owner. The Global Model Update ensures the owner decrypts the aggregated update using the private key, updates the global model, encrypts it with the public key, and shares the encrypted global model with FL participants. With optional model evaluation, training can iterate for several rounds or until convergence. This model offers a robust solution to Florida data privacy and security issues, with versatile applications across domains. This paper presents core model components, advantages, and potential domain-specific implementations while making significant strides in addressing FL's data privacy concerns.

Keywords: Secure Federated Learning; Homomorphic Encryption; Data Privacy; Model Security; Collaborative Machine Learning

Received : September 7, 2023
Received in Revised: October 25, 2023
Accepted: October 28, 2023

Introduction

Federated Learning (FL) has transformed collaborative machine learning by enabling decentralized model training while safeguarding data privacy (Baracaldo & Shaul, 2023; NVIDIA, 2022; Jin, 2023; Wibawa et al., 2023; IEEE, 2021). This paradigm holds promise across diverse domains such as healthcare, finance, and the Internet of Things (IoT). Nonetheless, Florida encounters formidable privacy and security challenges hindering its widespread adoption. This paper introduces a pioneering approach, "Secure Federated Learning with a Homomorphic Encryption Model," which confronts these challenges by integrating homomorphic encryption techniques into the FL framework. Our approach originally merges two robust technologies, federated learning, and homomorphic encryption, yielding notable contributions:

DOI: https://doi.org/10.47667/ijpasr.v4i3.235

Enhanced Data Privacy: Leveraging homomorphic encryption, our model ensures the confidentiality of raw data throughout the FL process. This is a pivotal advancement in protecting sensitive information within collaborative machine-learning scenarios. Model Security: Our approach fortifies model integrity by enabling secure aggregation of model updates without exposing individual contributions from participants. This safeguarding measure enhances the resilience of the global model against potential adversarial attacks and data tampering. Facilitated Collaborative Machine Learning: "Secure Federated Learning with Homomorphic Encryption" fosters collaboration among multiple participants without compromising their data privacy. This has far-reaching implications, particularly in healthcare, finance, and IoT applications, where data sharing and collective model training are imperative.

Problem Statement Despite the merits of federated learning, current implementations fail to ensure data privacy and model security. Traditional FL protocols often entail sharing raw model updates, posing risks to sensitive information. Moreover, the absence of secure aggregation mechanisms makes the model susceptible to attacks. Our problem statement revolves around the quest for a robust solution to address these challenges, bridging the gap between FL's collaborative nature and the critical imperatives of data privacy and model security.

To implement the proposed "Secure Federated Learning with Homomorphic Encryption" model, we suggest several Python libraries and frameworks: PySyft: This Python library facilitates privacy-preserving machine learning, enabling secure multi-party computation and federated learning; TenSEAL: TenSEAL serves as a Python library for homomorphic encryption, empowering secure data processing; PyTorch: A widely-adopted deep learning framework, PyTorch integrates into our model for local training on participants' devices, and NumPy: We employ NumPy for array manipulation and mathematical operations within the model.

These libraries allow us to implement and evaluate our "Secure Federated Learning with Homomorphic Encryption" model, a significant step toward realizing safe and private collaborative machine learning. The outline structure of the paper is as follows: Section 2, related work; Section 3, Homomorphic encryption; Section 4, Federated learning; Section 5, Core Components of the Proposed Model; Section 6, the Proposed Model; Section 7, results and Discussion, and Section 8 Conclusions.

Related work

Park & Lim (2022) This paper proposes a secure federated learning framework that uses homomorphic encryption and verifiable computing to ensure the confidentiality and integrity of data and model parameters. Madi et al. (2021), This article proposes a privacy-preserving federated learning algorithm that uses homomorphic encryption to protect data privacy while training machine learning models.

Fang & Qian (2021) This paper proposes a privacy-protected machine learning algorithm that combines homomorphic encryption and federated learning to protect data and model security during model training. Rahulamathavan (2023), This paper proposes a novel federated learning algorithm based on fully homomorphic encryption that can protect against Byzantine attacks.

Kurniawan & Mambo (2022), This article discusses how federated learning can be combined with homomorphic encryption to preserve privacy while training machine learning models. Park & Lim (2022), This paper proposes a homomorphic encryption-based federated learning scheme to preserve privacy in active learning scenarios.

DOI: https://doi.org/10.47667/ijpasr.v4i3.235

Homomorphic encryption

Homomorphic encryption is a cryptographic technique that allows computations on encrypted data without decrypting it. This technique has several key points, including the ability to perform mathematical operations directly on encrypted data, the existence of three types of homomorphic encryption (partially, somewhat, and entirely), and various use cases such as secure outsourcing, privacy-preserving machine learning, secure multi-party computation, and cloud computing. However, homomorphic encryption faces challenges such as computational overhead, key management, and complexity. Homomorphic encryption has applications in various domains, including finance, healthcare, and secure messaging, and ongoing research aims to improve its efficiency and practicality (Gillis, 2022; SSL2BUY, 2019; Yackel, 2021; IEEE, 2010; Munjal & Bhatia, 2023).

Federated learning

is an approach to machine learning that allows training models without centralizing training data in a single place, for reasons of privacy, security, and efficiency. It involves multiple parties collaborating to train a shared model while keeping their data private. Federated learning frameworks are software tools that enable the implementation of federated learning. Here are some sources for Federated Learning Frameworks: Apheris (2023) highlights the top open-source federated learning frameworks. Federated Learning has seven open-source frameworks and software solutions, including TensorFlow Federated, PySyft, and IBM Federated Learning. IBM Federated Learning (Ludwig, 2020) an enterprise framework white paper that presents Federated Learning (FL) and its advantages. The IBM Federated Learning Framework addresses corporate FL concerns in the article.

The scientific study (Song, 2022) offers a Federated Learning Framework (H2-Fed) that may significantly improve the performance of cooperative intelligent transportation systems (C-ITS). Comparing Open-Source Federated Learning Frameworks for IoT: A Review and Analysis (Kholod et al., 2020) The scientific study compares open-source Federated Learning Frameworks and their usefulness in IoT systems. (Ali et al., 2020; Kholod et al., 2020) TensorFlow Federated, Paddle Federated Learning Framework, and PySyft are open-source federated learning frameworks.

The refined CNN and deep regression forests federated learning framework (Nolte et al., 2023) is a research paper that provides a way to discover and manage population heterogeneity in federated learning contexts. Federated learning frameworks allow federated learning. TensorFlow Federated, PySyft, and IBM Federated Learning are open-source Federated Learning frameworks. These frameworks solve federated learning difficulties in corporate, IoT, and cooperative intelligent transportation systems.

Core Components of the proposed model

Federated Learning Framework: Our model leverages the existing FL framework, which consists of a central server and multiple decentralized participants. Participants communicate with the server during model training rounds, exchanging model updates instead of raw data. Homomorphic Encryption: Homomorphic encryption is a cryptographic technique that enables computations on encrypted data without decrypting it. We integrate HE into the FL framework to enable secure model aggregation.

Key Components: Data Encryption: Each participant encrypts their local data using HE before sharing it with the central server. This ensures that sensitive data remains confidential throughout the FL process. Secure Aggregation: The central server performs model

DOI: https://doi.org/10.47667/ijpasr.v4i3.235

aggregation on the encrypted model updates received from participants. He allows the server to compute the aggregated model while the data remains encrypted.

Decryption on Model: Once the aggregated model is computed, the server sends it back to the participants, who can decrypt and use it for local improvements.

The proposed model

The Algorithm of Secure Federated Learning with a Homomorphic Encryption Model is:

Initialization:

- Initialize a global machine learning model.
- Generate a homomorphic encryption key pair (public and private keys).
- Share the public key with FL participants.

FL Participant Setup:

- FL participants collect and preprocess their local data.
- Encrypt their local data using the public key.

FL Training Rounds:

- For each training round:
  - FL participants:
    - Decrypt the current global model using the private key.
    - Compute local updates based on their encrypted data.
    - Encrypt the local updates.
    - Send encrypted updates to the aggregator.
  - Aggregator:
    - Collect encrypted updates from FL participants.
    - Homomorphically aggregate the updates without decryption.
    - Send the encrypted aggregated update to the global model owner.

Global Model Update:

- Global model owner:
  - Decrypt the aggregated update using the private key.
  - Update the global model with the aggregated update.
  - Encrypt the updated global model with the public key.
  - Share the encrypted global model with FL participants.

Repeat training rounds:

- Repeat the training rounds for a predefined number of iterations or until convergence.

DOI: https://doi.org/10.47667/ijpasr.v4i3.235

Model Evaluation:

- Optionally, evaluate the performance of the trained global model.

Termination:

- End the FL process.

Results and Discussion

The Secure Federated Learning with a Homomorphic Encryption Model algorithm is a privacy-preserving approach to federated learning that leverages homomorphic encryption to encrypt both data and model updates. Here, we present a consolidated section that combines both the benefits and advantages of this algorithm:

Benefits and Advantages

Confidentiality and Integrity Protection: The algorithm provides confidentiality and integrity protection for the data and model updates. Ensures that sensitive information remains secure. Data Privacy: Our method guarantees data privacy throughout the Federated Learning (FL) process by ensuring that raw data is never exposed. All computations are performed on encrypted data, preserving data confidentiality.

Model Security: The model remains secure as it is never fully exposed to any party. Only encrypted model updates are exchanged and aggregated, reducing the risk of model leakage or theft. Enhanced Collaboration: Secure FL encourages more participants to join collaborative efforts without fearing data breaches. This fosters a more diverse and representative dataset for model training. Regulatory Compliance: Our approach aligns with data protection regulations, making it easier for organizations to comply with privacy laws such as GDPR(General Data Protection Regulation).

Results

The experimental results demonstrate that the proposed homomorphic encryption-based Federated Learning scheme effectively preserves privacy in active learning while maintaining accuracy.

Applications

The algorithm finds applications in various settings, including enterprise, IoT-enabled healthcare systems, and deep active learning (Kurniawan & Mambo, 2022; Song, 2022; Sattar & Gaata, 2017). Additionally, it can train machine learning models to extract knowledge from training data that cannot be directly accessed (Baracaldo & Shaul, 2023).

Conclusion

The Secure Federated Learning with a Homomorphic Encryption Model algorithm is a privacy-preserving approach to federated learning that uses homomorphic encryption to protect data and model updates. It protects data confidentiality and integrity and allows secure collaboration among multiple parties in training a shared model while maintaining the confidentiality of their data. Applyable in multiple scenarios, "Secure Federated Learning with a Homomorphic Encryption Model" addresses data privacy and security issues in federated learning. Homomorphic encryption secures decentralized cooperation and protects sensitive data. This privacy-preserving technique has significant applicability across many disciplines and might transform collaborative machine learning.

DOI: https://doi.org/10.47667/ijpasr.v4i3.235

Acknowledgment

Mustansiriyah University (https://uomustansiriyah.edu.iq/) and Al-Iraqia University in Baghdad, Iraq, supported this effort.

References

Ali, J. J., Shati, N. M., & Gaata, M. T. (2020, March). Abnormal activity detection in surveillance video scenes. TELKOMNIKA (Telecommunication Computing Electronics and Control), 18(5), 2447-2453.

Apheris. (2023, June 21). Top 7 Open-Source Frameworks for Federated Learning. https://www.apheris.com/resources/blog/top-7-open-source-frameworks-for-federated-learning.

Baracaldo, N., & Shaul, H. (2023, January 4). Federated Learning Meets Homomorphic Encryption. IBM Research Blog.

Fang, H., & Qian, Q. (2021, April 8). Privacy Preserving Machine Learning with Homomorphic Encryption and Federated Learning. Future Internet. https://doi.org/10.3390/fi13040094.

Gillis, A. S. (2022, August 24). Homomorphic Encryption. Security. https://www.techtarget.com/searchsecurity/definition/homomorphic-encryption.

IEEE Digital Privacy. (2019). Types of Homomorphic Encryption. https://digitalprivacy.ieee.org/publications/topics/types-of-homomorphic-encryption.

IEEE Xplore. (2021, December 1). Secure Aggregation in Federated Learning via Multi-party Homomorphic Encryption. IEEE Conference Publication. https://ieeexplore.ieee.org/document/9682053.

Jin, W. (2023, March 20). FedML-HE: An Efficient Homomorphic-Encryption-Based Privacy-Preserving Federated Learning System. arXiv.org. https://arxiv.org/abs/2303.10837.

Kholod, I., Yanaki, E., Fomichev, D., Shalugin, E. D., Novikova, E., Filippov, E., & Nordlund, M. (2020, December 29). Open-Source Federated Learning Frameworks for IoT: A Comparative Review and Analysis. Sensors. https://doi.org/10.3390/s21010167.

Kholod, I., Yanaki, E., Fomichev, D., Shalugin, E. D., Novikova, E., Filippov, E., & Nordlund, M. (2020, December 29). Open-Source Federated Learning Frameworks for IoT: A Comparative Review and Analysis. Sensors. https://doi.org/10.3390/s21010167.

Kurniawan, H., & Mambo, M. (2022, October 27). Homomorphic Encryption-Based Federated Privacy Preservation for Deep Active Learning. Entropy. https://doi.org/10.1007/s40747-022-00756-z.

Ludwig, H. (2020, July 22). IBM Federated Learning: An Enterprise Framework White Paper V0.1. arXiv.org. https://arxiv.org/abs/2007.10987.

Madi, A., Stan, O., Mayoue, A., Grivet-Sebert, A., Gouy-Pailler, C., & Sirdey, R. (2021, May 18). A Secure Federated Learning Framework Using Homomorphic Encryption and Verifiable Computing. https://doi.org/10.1109/rdaaps48126.2021.9452005.

Munjal, K., & Bhatia, R. (2023). A systematic review of homomorphic encryption and its contributions to the healthcare industry. Complex Intell. Syst., 9, 3759–3786.

DOI: https://doi.org/10.47667/ijpasr.v4i3.235

https://doi.org/10.1007/s40747-022-00756-z.

Nolte, D., Bazgir, O., Ghosh, S., & Pal. (2023, March 22). Federated Learning Framework Integrating Refined CNN and Deep Regression Forests. Bioinformatics advances. https://doi.org/10.1093/bioadv/vbad036.

NVIDIA Technical Blog. (2022, September 2). Federated Learning with Homomorphic Encryption. NVIDIA Technical Blog. https://developer.nvidia.com/blog/federated-learning-with-homomorphic-encryption/.

Park, J., & Lim, H. (2022). Privacy-Preserving Federated Learning Using Homomorphic Encryption. Applied Sciences, 12(2), 734. https://doi.org/10.3390/app12020734.

Park, J., & Lim, H. (2022, January 12). Privacy-Preserving Federated Learning Using Homomorphic Encryption. Applied Sciences. https://doi.org/10.3390/app12020734.

Rahulamathavan, Y. (2023, June 8). FheFL: Fully Homomorphic Encryption-Friendly Privacy-Preserving Federated Learning with Byzantine Users. arXiv.org. https://arxiv.org/abs/2306.05112.

Sattar, I. A., & Gaata, M. T. (2017, March). Image steganography technique based on adaptive random key generator with suitable cover selection. In 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT) (pp. 208-212). IEEE.

Song, R. (2022, April 1). Federated Learning Framework: Coping with Hierarchical Heterogeneity in Cooperative ITS. arXiv.org. https://arxiv.org/abs/2204.00215.

SSL2BUY. (2019). Homomorphic Encryption: Everything You Should Know About It. https://www.ssl2buy.com/wiki/homomorphic-encryption.

Wibawa, F., Catak, F. O., Sarp, S., Kuzlu, M., & Cali, U. (2022). Homomorphic Encryption and Federated Learning-based Privacy-Preserving CNN Training: COVID-19 Detection Use-Case. ArXiv. /abs/2204.07752.

Yackel, R. (2021, July 6). What Is Homomorphic Encryption, and Why Isn't It Mainstream? Keyfactor. https://www.keyfactor.com/blog/what-is-homomorphic-encryption/.

DOI: https://doi.org/10.47667/ijpasr.v4i3.235