Privacy-Preserving Machine Learning with Zero-Knowledge Proofs in Federated Learning

Prajin Chopra

Introduction

Blockchain is a transformative technology that has revolutionized the way data is stored, secured, and transacted. It serves as a decentralized and immutable ledger, ensuring transparency, trust, and tamper resistance. With applications spanning from cryptocurrencies to supply chain management and beyond, blockchain has the potential to reshape various industries. Zero-knowledge (ZK) proofs allow one party with a secret witness to prove some statement about that witness without revealing any additional information. In recent years we have seen massive progress in the efficiency and scalability of ZK proofs based on many different ideas **[1].** Given these advancements, we see significant promise for using Zero-Knowledge (ZK) proofs in the field of machine learning, especially in the context of Federated Learning inference. Federated learning is an engaging framework for large-scale distributed training of deep learning models with thousands to millions of users **[2].** The widespread usage of computing devices, such as mobile phones and tablets, has increased the amount of proprietary user data. The wealth of data raises concerns about user privacy as well as providing opportunities to develop various machine learning (ML) models. To address this issue, a decentralized learning paradigm called federated learning (FL) has been proposed by Yang et al. (2019) **[3].** Yet, the landscape of federated learning is riddled with numerous privacy and security concerns including potential vulnerabilities to model poisoning attacks and other malicious behaviors **[3].**

Literature Survey

However, there are limitations to the zero-knowledge (ZK) system that deserve further exploration in future works. Specifically, the ZK protocol can only prove to one verifier at a time, and the communication cost is fairly high compared to succinct ZK proofs like zk-SNARKs. we also observed a very high overhead for Batch Normalization, which may potentially be further optimized**[1].** It has been acknowledged that existing protocols based on secret sharing, which securely

compute functions among multiple parties, are not practical in federated learning due to the high-dimensional vectors involved. The proposed protocol aims to address this challenge**.** Although there have been studies on zero-knowledge proof (ZKP) protocols based on secure multi-party computation (MPC), they note that they have not been widely used in the machine learning context. The proposed protocol uses MPC protocols to devise a ZKP protocol for detecting poisoned models**.[2]**

While it proposes a defense mechanism that does not require direct access to local model parameters, it does not provide a solution for scenarios where such access is necessary.

This does not solve the problem of not having prior knowledge about specific attack scenarios. The proposed defense framework, DifFense, leverages differential testing and outlier detection without requiring previous knowledge of attack scenarios, but it does not address the issue of not having any knowledge at all.**[3]**

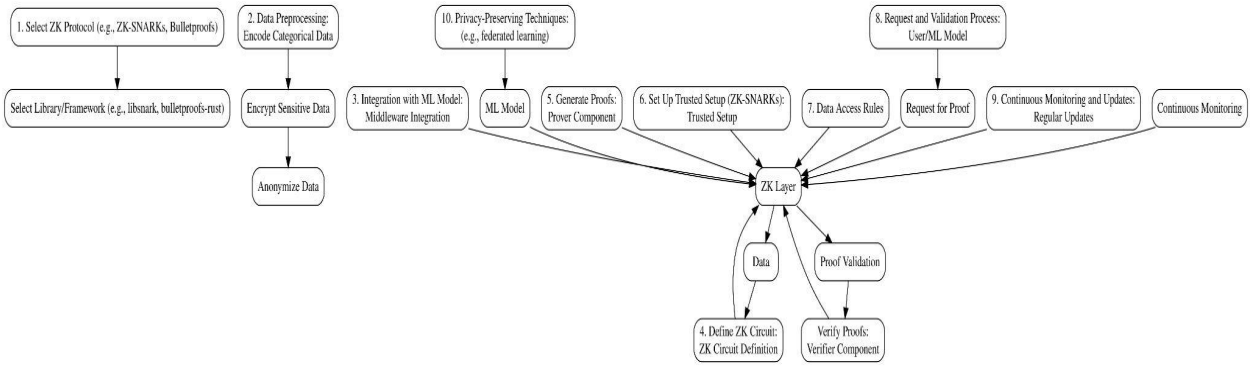
The focus is on the security aspects of federated learning (FL) and does not extensively address privacy vulnerabilities. Privacy vulnerabilities can play a significant role in attacks against FL models. It has been mentioned that proactive defense mechanisms such as anomaly detection and robust aggregation, do not provide in-depth discussions or specific techniques to implement these mechanisms effectively in FL systems**[5].** It has been mentioned that the inherent heterogeneity of IoT devices poses challenges for deploying the federated learning framework in large-scale real-world scenarios. This includes issues related to non-i.i.d. data distribution, unbalanced data, and heterogeneous devices. Further research is needed to address these heterogeneity issues in federated learning**[6]**.

Federated learning systems need to be robust to various challenges, including worker dropouts, Byzantine failures, and adversarial attacks. Ensuring robustness is crucial for real-world deployment but is not detailed in the information provided**[4]**. The proposed model is used to generate and verify proofs without revealing the actual data. This protection can help guard against adversarial attacks that attempt to extract sensitive information from the model or its outputs.The proposed model aims to address the challenge of working with high-dimensional vectors in federated learning. This is crucial because traditional protocols based on secret sharing may not be practical for such high-dimensional data and proposes to leverage secure multi-party computation (MPC) protocols to develop a ZKP protocol for detecting poisoned models. It is to be addressed that the limited use of

ZKPs in the machine learning context and contributes to enhancing the security of federated learning. The proposed model a defense mechanism that does not require direct access to local model parameters. This approach aligns with addressing the issue of limited access to local model parameters in federated learning systems.

Architecture

This privacy-preserving system encompasses data preprocessing, ML model integration, Zero-Knowledge Circuit definition, ZK proof generation and verification, trusted setup, access control, request-validation processes, continuous monitoring, and privacy-enhancing techniques. It enables secure and confidential data access for a privacy-preserving ML model.



1. Data Preprocessing:
   - Encode categorical data into numerical representations.
   - Encrypt sensitive data using homomorphic encryption.
   - Anonymize data to remove personally identifiable information (PII).

2. Integration with ML Model:
   - Integrate the privacy-preserving layer into the ML model architecture.

3. Define Zero-Knowledge (ZK) Circuit:
- Create a ZK circuit that represents the computation you want to prove.
- In ZK-SNARKs, use a domain-specific language (DSL) provided by the library.
- In Bulletproofs, define constraints in code (e.g., Rust) that capture the computation.

4. Generate and Verify ZK Proofs:
- Prover Component: Generates a ZK proof for the given circuit without revealing the actual data.
- Verifier Component: Checks the validity of the proof.

5. Set Up Trusted Setup (ZK-SNARKs):
- Conduct a one-time trusted setup phase.
- Generate a set of public parameters.
- Ensure it's done by a trusted entity and securely stored to prevent compromise.

6. Data Access Rules:
- Define who has permission to request and verify ZK proofs.
- Implement authentication and authorization mechanisms.

7. Request and Validation Process:
- When a user or ML model wants to access data, send a request to the ZK layer.
- The ZK layer retrieves data and circuit, generates a ZK proof, and sends it to the user or model.
- The user or model sends the proof to the verifier within the ZK layer.
- The verifier checks the proof's validity, ensuring that the data used in the computation corresponds to the proof.

8. Continuous Monitoring and Updates:
- Regularly update the ZK layer and its dependencies to address security vulnerabilities.
- Implement continuous monitoring to detect suspicious activity related to data access and proof generation.

9. Privacy-Preserving Techniques:

- Additional techniques like federated learning, differential privacy, etc., can be integrated to enhance privacy.

10. ML Model:
- The privacy-preserving ML model that utilizes the ZK proofs to make predictions while maintaining data privacy.

Methodology

Data Preprocessing: This step involves preparing the input data for the ML model while preserving privacy. It includes encoding categorical data, encrypting sensitive data using homomorphic encryption, and anonymizing data to remove PII.

Integration with ML Model: Integrate the privacy-preserving layer into the ML model's architecture to enable it to interact with the ZK proofs.

ZK Circuit Definition: Create a ZK circuit that represents the computations performed by the ML model. Depending on the chosen ZK protocol (ZK-SNARKs or Bulletproofs), use the appropriate tools and languages to define the circuit.

Generate and Verify ZK Proofs: The prover component generates ZK proofs without revealing the actual data, and the verifier checks the validity of these proofs. This ensures that the ML model can make predictions without knowing the underlying data.

Set Up Trusted Setup (ZK-SNARKs): In the case of ZK-SNARKs, perform a one-time trusted setup to generate public parameters securely. This step is critical for ensuring the security of the system.

Data Access Rules: Define who can request and verify ZK proofs. Implement authentication and authorization mechanisms to control access.

Request and Validation Process: Users or the ML model can request ZK proofs for specific computations. The ZK layer retrieves data and circuits, generates proofs, and verifies them to ensure data privacy and accuracy.

Continuous Monitoring and Updates: Regularly update the ZK layer and its dependencies to address security vulnerabilities and monitor activities related to data access and proof generation.

Privacy-Preserving Techniques: Enhance privacy by incorporating additional techniques like federated learning or differential privacy into the ML model.

ML Model: The final model leverages the ZK proofs and privacy-preserving techniques to make predictions while safeguarding the privacy of the underlying data.

Conclusion

In conclusion, the proposed architecture and methodology offer a robust solution to the multifaceted challenges faced in the domains of privacy-preserving and secure machine learning, particularly in the context of federated learning. By integrating Zero-Knowledge (ZK) proofs, data preprocessing techniques, and stringent data access controls, this model addresses the paramount concerns of data privacy, security, and fairness. It enables the development of machine learning models that can make predictions while safeguarding sensitive information, thereby fostering trust in the deployment of AI systems. Furthermore, the emphasis on continuous monitoring and updates underscores the commitment to staying ahead of evolving security threats. The incorporation of privacy-preserving techniques, such as federated learning and differential privacy, further bolsters the model's ability to navigate complex privacy challenges. In essence, this model represents a substantial step toward the responsible and ethical use of machine learning in real-world applications, where data privacy and security are of paramount importance.

[1]Mystique: Efficient Conversions for Zero-Knowledge Proofs with Applications to Machine Learning Chenkai Weng, Northwestern University; Kang Yang, State Key Laboratory of Cryptology; Xiang Xie, Shanghai Key Laboratory of Privacy-Preserving Computation and MatrixElements Technologies; Jonathan Katz, University of Maryland; Xiao Wang, Northwestern University

[2]Preserving Privacy and Security in Federated Learning; Truc Nguyen, My T. Thai

[3]BACKDOOR DEFENSE IN FEDERATED LEARNING USING DIFFERENTIAL TESTING AND OUTLIER DETECTION Yein Kim University of California, San Diego La Jolla, CA y5kim@ucsd.edu Huili Chen University of California, San Diego La Jolla, CA huc044@ucsd.edu Farinaz Koushanfar University of California, San Diego La Jolla, CA farinaz@ucsd.edu

[4]Marzo, S., Pinto, R., McKenna, L., Brennan, R. (2023). Privacy-Enhanced ZKP-Inspired Framework for Balanced Federated Learning. In: Longo, L., O'Reilly, R. (eds) Artificial Intelligence and Cognitive Science. AICS 2022. Communications in Computer and Information Science, vol 1662. Springer, Cham. https://doi.org/10.1007/978-3-031-26438-2_20

[5]N. Bouacida and P. Mohapatra, "Vulnerabilities in Federated Learning," in IEEE Access, vol. 9, pp. 63229-63249, 2021, doi: 10.1109/ACCESS.2021.3075203.

[6]Z. Chen, P. Tian, W. Liao and W. Yu, "Zero Knowledge Clustering Based Adversarial Mitigation in Heterogeneous Federated Learning," in IEEE Transactions on Network Science and Engineering, vol. 8, no. 2, pp. 1070-1083, 1 April-June 2021, doi: 10.1109/TNSE.2020.3002796