CNS Final Project Proposal

r11921a16 何秉學 r11922193 張歐華 r11944034 許智翔 b09902060 馮楷

1 Problem Description

1.1 Terminology

Federated Learning is a decentralized machine learning method that enables training models without exposing data. Traditional machine learning methods require all data to be centralized in one location for training, but Federated Learning enables models to be trained on many distributed devices, such as smartphones, tablets, or embedded devices, with each device training its own local data. This greatly reduces data transmission and storage requirements and better protects user privacy.

Privacy Preserving is a method of designing and implementing computing systems aimed at protecting the privacy of data and personal information. It is typically a technique used in the exchange or processing of data to ensure the confidentiality, integrity, and availability of data.

Privacy preserving techniques can help ensure that data is protected and that sensitive information is not disclosed even during data sharing, analysis, or storage. For example, techniques such as data encryption, differential privacy, and multi-party computation can be used to protect data privacy, and these techniques have wide applications in data analysis and machine learning.

1.2 Research Problem

Although federated learning seems perfect at protecting local data, an attack called gradient inversion attack can restore the local data from local model weights. Thus recent research aims to encrypt model weights without affecting federated learning process.

In this paper [2], they proposed a method to prevent a malicious server as an attacker as their threat model. However, each client in the same group shared the same public/private key so that the attacker can pretend a benign user and exploit part of the packages. Thus, we modify the threat model to assume that the attacker is one of the user in a group and he can decrypt a part of the packages transferred in this group and the confidentiality property is gone.

2 Related Work

2.1 FedML-HE

The paper proposes an efficient privacy-preserving federated learning system based on homomorphic encryption and a universal two-stage optimization scheme: (1)Parameter Efficiency and (2)Parameter Selection.

Limitations

The system still incurs a high computation overhead compared to plaintext federated learning. Since a universal key-pair is shared by every clients in this scheme, one compromised client will lead to the leakage of every local models. Moreover, it requires on a server to publish the key-pair, once the server is malicious, local models can be easily reversed by gradient inversion attack.

2.2 Decentralized Threshold Homomorphic Encryption

The authors provided a (t, n)-theshold HE scheme based on CKKS HE scheme, where a center server is no longer needed for decryption. Instead, the decryption process can be done, when t out of n parties agree to decrypt.

Limitations

To ensure the joint key security, smudging errors are required which inevitably enlarge the entire parameter size, resulting in a huge computational overhead.

2.3 Proxy Re-Encryption

The scheme can re-encrypt ciphertext for multiple receivers at a time by generating the re-encryption key from private and other users' public identities, which meets the requirement of FL.

Limitations

This scheme relies on a trusted authority KGC(key generation center) for initialization. Moreover, it assumes that all participants in the federated learning system are honest and follow the protocol correctly.

3 Plan

Our method would be mainly based on the general scheme proposed for the practical deployment of homomorphic-encryption-based federated learning [2]. Several security issues remain unsolved since the scheme aims to keep generality. We will apply two or more possible schemes to improve data security.

3.1 Key Management

In the FedML-HE scheme, there exists only one pair of public/private keys. Once a client is compromised or is malicious, they can easily decrypt the gradient from another client and possibly reverse the gradient to the original data using the gradient inversion attack. To defend against this type of attack, we may not allow clients to share the same pair of public/private keys.

On the other hand, the key pair is published by a server. If the server is compromised or malicious, then gradients from clients can be decrypted. By the gradient inversion attack, local data can be reversed. Thus decentralization is important to any federated learning scheme.

The following are two latest solutions to the problems.

- Proxy re-encryption [4]
  This method establishes a key management held by a trusted third party and allows all clients to use distinct key pairs. The first problem is solved, while decentralization is not fulfilled.
- Threshold homomorphic encryption [3]
  This method establishes a decentralized scheme by using threshold cryptography. At the same time, a single client can not decrypt any gradient from another, which solves the first problem.

3.2 Poisoning attack

Poisoning attack is still one severe threat to federated learning. A malicious or compromised client can upload a bad gradient to the server and reduce the accuracy of the training model [5]. We will discuss this attack and try to find a scheme to mitigate poisoning attacks.

3.3 Plan

We will compare the security and efficiency of each scheme. For the security part, we will introduce several threats and evaluate the security level of a scheme by checking if the scheme can defend against these threats.

4 Timeline

5/8: Finish studying paper
5/15: Apply Decentralized Threshold HE on the FedML system
5/22: Apply Proxy Re-Encryption on the FedML system
5/29: Finish the report
6/5, 6/12: Oral Presentation

5 Deliverables

The final deliverable of this project will be a comprehensive analysis of the security vulnerabilities in the FedML-HE system, as well as enhancements through the addition of threshold homomorphic encryption and proxy re-encryption. These enhancements will be presented in a detailed report, as well as their potential benefits and limitations in a practical scenario.

References

[1] L. T. Phong, Y. Aono, T. Hayashi, L. Wang and S. Moriai, *"Privacy-Preserving Deep Learning via Additively Homomorphic Encryption,"* in IEEE Transactions on Information Forensics and Security, vol. 13, no. 5, pp. 1333-1345, 2018.

[2] Weizhao Jin and Yuhang Yao and Shanshan Han and Carlee Joe-Wong and Srivatsan Ravi and Salman Avestimehr and Chaoyang He, *FedML-HE: An Efficient Homomorphic-Encryption-Based Privacy-Preserving Federated Learning System*, Arxiv, 2023.

[3] E. Kim, J. Jeong, H. Yoon, Y. Kim, J. Cho and J. H. Cheon, *"How to Securely Collaborate on Data: Decentralized Threshold HE and Secure Key Update,"* in IEEE Access, vol. 8, pp. 191319-191329, 2020.

[4] Chun-I Fan, Ya-Wen Hsu, Cheng-Han Shie, and Yi-Fan Tseng. 2022. ID-Based Multireceiver Homomorphic Proxy Re-Encryption in Federated Learning. ACM Trans. Sen. Netw. 18, 4, Article 55 (November 2022).

[5] Sagar, S., Li, C. S., Loke, S. W., & Choi, J. (2023). *Poisoning Attacks and Defenses in Federated Learning: A Survey.* arXiv preprint arXiv:2301.05795.