Article Review

PPFL: Privacy-preserving Federated Learning with Trusted Execution Environments

Summary

The paper presents a privacy-preserving framework for federated learning that utilizes *trusted execution environments* (TEE) and *greedy layer-wise training*. The proposed framework is designed to address the challenge of protecting users' privacy in federated learning settings against various malicious attacks, including data reconstruction attacks, property inference attacks, and membership inference attacks. To achieve this, the authors leverage TEE on both the server and client sides and perform layer-wise training within TEE to prevent any potential information leakage in the gradients. The experimental results demonstrate that the proposed framework can preserve user privacy, albeit with additional communication and computation overhead.

Strength and Weakness

Strength:

1. The author demonstrates the practicality of covering the gradients using TEE, making it a promising avenue for future research.
2. The paper examines multiple types of privacy threats in federated learning, which presents a more realistic scenario than other studies.

Weakness:

1. Aside from using TEE to conceal gradient information from attackers, the PPFL design lacks novelty. It appears to be a combination of existing techniques, such as layer-wise training and freezing pre-trained layers.
2. PPFL design has a significant flaw. Training the model in TEE on the client side seems unnecessary when private data is already stored in the REE and could be exposed if the device is compromised, rendering gradient protection pointless.
3. The proposed framework inherits the drawbacks of greedy layer-wise training, limiting its applicability to shallow and layer-based models due to increased overheads for deeper models and difficulty in partitioning non-layer-based models into blocks. Additionally, partition settings become additional hyper-parameters that require careful tuning, resulting in a less efficient training process. This is impractical for the trend of AIoT services that require larger-scale clients and more complex models.
4. The experiment has several issues, including unclear implementation of attacks, a lack of complexity in the task (image-based classification only), and a questionable comparison between layer-wise and end-to-end FL in terms of computation and communication overheads.

Questions

1. If attackers cannot obtain the gradients of some layers, what is the method to execute the three types of attacks mentioned in the paper?
2. The correctness of Equation 2 in the complexity analysis section is limited as it does not account for the variation in epoch E that occurs when utilizing layer-wise training.

Conclusion

In my opinion, the use of TEE in federated learning to improve privacy protection is a revolutionary idea. Nevertheless, the possibility of layer-wise training in the real world is uncertain due to various uncontrollable factors that may affect the results. On the other hand, a slight modification of PPFL that includes end-to-end training and TEE-based aggregation on the server side is more feasible. This approach is similar to previous works that have used secure aggregation.