International Journal of Applied Science; Vol. 8, No. 2; 2025
ISSN 2576-7240 E-ISSN 2576-7259
https://doi.org/10.30560/ijas.v8n2p110

Analysis and Prospect of Federated Learning and Privacy Protection Technology

Peng Hongye[1,2]

[1] School of Information, Guizhou University of Finance and Economics, Guizhou, China

[2] Key Laboratory of Blockchain and Fintech, Department of Education of Guizhou Province, Guiyang Guizhou 550025, China.

Correspondence: Peng Hongye, Key Laboratory of Blockchain and Fintech, Department of Education of Guizhou Province, Guiyang Guizhou 550025, China.

Received: May 12, 2025; Accepted: May 29, 2025; Published: May 30, 2025

Abstract

As a new type of distributed machine learning technology, federated learning has shown great application potential in the Internet of things, health care, smart home, finance and other fields. Its core advantage is that it can conduct model training without centralized data, effectively reducing the cost of data transmission and storage, and avoiding the risk of privacy disclosure. However, with the wide application of Federated learning, the problems of data security and privacy protection are increasingly apparent, especially in the face of complex network attacks and data leakage risks. This paper deeply analyzes the basic principles and architecture of Federated learning, and discusses the possible privacy threats in data transmission, model updating and participating devices in detail. Combined with the existing security protection technologies, such as differential privacy, encryption algorithm and secure multi-party computing, this paper discusses how to effectively ensure the security of Federated learning. Finally, the article also looks forward to the future development trend of Federated learning in privacy protection, model optimization, computational efficiency and cross domain collaboration, aiming to provide theoretical support and practical guidance for the further development and application of this technology.

Keywords: federal learning, privacy disclosure, safety protection technology, future development

1. Introduction

With the rapid development of smart devices and Internet of Things technology, the volume of data has grown explosively, and at the same time, a large amount of data has also been generated on different user terminals such as smart phones and medical devices. Traditional machine learning methods usually require data to be uploaded to a centralized computing cluster for unified training. This not only increases the cost of data transmission and storage, but also may face the risk of data privacy leakage. To address the poor model training effect on a single device and the privacy leakage problem faced by centralized storage, researchers have proposed federated learning.

Federated Learning (FL, Federated Learning) is a distributed machine learning technology that enables multiple devices or servers to collaboratively train without having to centralize the data in a single server. With the core idea that when multiple data sources jointly participate in model training, only the intermediate parameters of the model are passed for model training without uploading the original data. Unlike traditional centralized machine learning, the data in federated learning does not leave the local area. It collaboratively completes tasks by sharing model parameters. On the basis of ensuring data privacy, security, and compliance with laws and regulations, it achieves data sharing and joint modeling, greatly reducing privacy and security issues, and is widely applied in smart home, medical and health care, and other fields as well.

However, although federated learning can avoid centralized and unified processing of data and has certain advantages in privacy protection, it still faces certain security risks. With the increasingly strict regulations on data privacy, such as the "Personal Information Protection Law of the People's Republic of China" and the "Data Security Law of the People's Republic of China" promulgated by our country, the "General Data Protection Regulation" implemented by the European Union, the "Personal Data Protection Bill" released by India, the data security issue of federated learning have also attracted widespread attention. To further ensure data security, it is necessary to further guarantee the security of federated learning Therefore, this paper conducts research on the privacy and security issues of federated learning. The principles, architectures, privacy threats and corresponding

protection technologies of federated learning were analyzed and summarized, and the future research trends were discussed.

2. Principles of Federated Learning Applications

In traditional federated learning, its global objective is to minimize a global loss function $L(w)$, namely

$$w^* = \arg\min_w \sum_{k=1}^{K} N_k L_k(w)$$

Where w represents the model parameters, $L_k(w)$ is the local loss function of the k device. N_k is the data samples of the k device, and K indicates the total number of devices participating in the training.

When each device trains the model locally and updates the model parameters, the local update formula of the k device is:

$$w_k^{t+1} = w_k^t - \eta \nabla L_k(w_k^t)$$

Among them, η is the learning rate and $\nabla L_k(w_k^t)$ is the gradient of the current model parameter of the k device.

When each device is aggregated locally, the server will regularly collect the model parameters of each device and perform aggregation updates. Suppose each device k provides its own model update $\Delta w_k = w_k^{t+1} - w_k^t$. The weighted average of the model updates of each device during the global update on the server side:

$$w^{t+1} = \sum_{k=1}^{K} \frac{N_k}{N} w_k^{t+1}$$

Among them $N = \sum_{k=1}^{K} N_k$ is the total number of all devices.

Ultimately, through multiple rounds of local training and global aggregation, the optimization of the global model can be achieved.

3. Organizational Structure of Federated Learning

In federated learning, each participant conducts model training locally based on their own data and sends the trained parameters such as weights and gradients to the central server or other participants. The central service decrypts these parameters, sets them together and updates the global model. Client-server architecture and end-to-end architecture are two common federated learning architectures.

3.1 Client-Server Architecture

In the client-server architecture as shown in Figure 1, there are mainly two participants: the central Server (Server) and the Client (Client). The client is mainly responsible for storing local data, conducting model training locally, and updating the local model and uploading it to the central server. The central server is responsible for aggregating model parameters from multiple clients in each round of iteration and broadcasting the updated global model to each participating client.

During the training process, the server first initializes a global model and sends it to each participating client. The client uses its data locally for model training, updates the model parameters, and encrypts these parameters before transmitting them back to the central server. After receiving updates from each client, the central server decrypts and aggregates the model parameters to generate a new global model. Subsequently, the updated global model is encrypted and returned to the client for the next round of training. The client decrypts the updated model to update the local model. This process is repeated continuously until the model converges or reaches the expected number of iterations.

This architecture is easy to implement and relatively efficient. It can not only effectively utilize the computing power of each client, but also reduce the reliance on centralized data storage. However, the central server is highly vulnerable to attacks, causing a single point of failure. Furthermore, semi-trusted third parties or even malicious third party central servers are likely to damage or steal the model, steal model parameters and infer source data, which will lead to the risk of data privacy leakage.

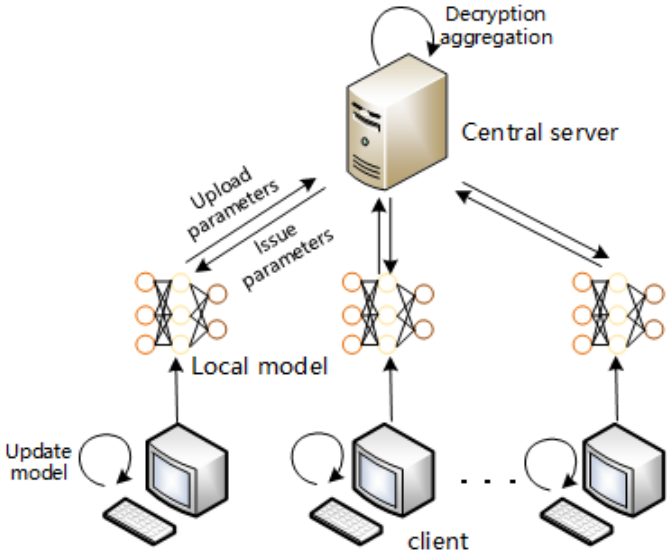

Figure 1. Client-server architecture

3.2 End-to-End Architecture

The end-to-end architecture is a decentralized federated learning architecture, as shown in Figure 2. In the architecture, there is no need to use a central service for model aggregation and update. The parameters of the model are passed among each participating client, and the client jointly aggregates , trains and optimizes the model.

The P2P network is utilized in the end-to-end architecture for data exchange, model aggregation and update. It not only avoids the problem of single point of failure, but also has good scalability and can support highly heterogeneous environments. However, how to guarantee the honesty and trustworthiness of the participants, ensure the effective aggregation of the model and the security of the model are the main problems faced by this architecture.

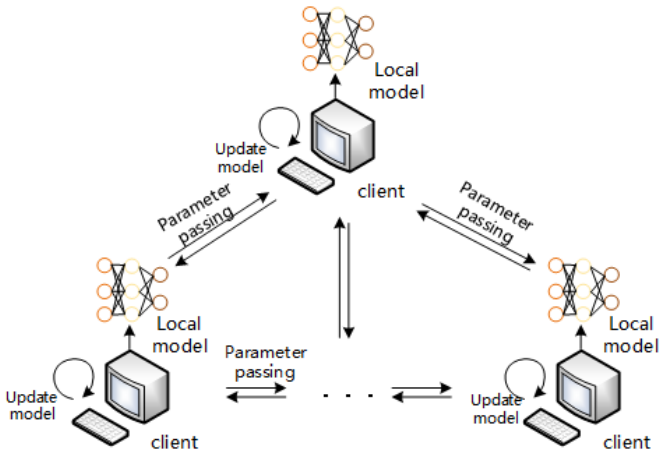

Figure 2. End-to-End Architecture

4. Federated Learning Privacy Attack

4.1 Attribute Inference Attack

Attribute reasoning attack is to infer the existence of a specific attribute in the local data set by analyzing the parameters updated by the model. This attack can be divided into active attack and passive attack. Active attackers usually use multi task learning or other attack strategies to mislead the system by deliberately inducing model training, so as to determine whether the data contains specific target attributes. For example, an attacker may design a task to make the model pay too much attention to the correlation of a certain attribute during the training process, so as to indirectly obtain some privacy information of the data. In contrast, passive attackers are relatively covert. They can only analyze model changes and infer whether there is a certain target attribute by eavesdropping on model parameter updates during data transmission. Although passive attacks usually do not directly interfere with the training process, it is still possible to infer whether the data contains sensitive attributes through precise analysis of the changes in the data transmission process. Both active and passive attribute reasoning attacks can effectively disclose the privacy of local data. Especially in the medical and financial fields, the leakage of sensitive attributes may bring great risks.

4.2 Member Inference Attack

Member inference attack is an attack method in which an attacker tries to infer whether a specific data sample is included in the local training data set by analyzing the parameters or outputs of the model. According to whether the training process is tampered by the attacker, this kind of attack can be divided into passive reasoning attack and active reasoning attack. In the passive reasoning attack, the attacker does not modify any data flow in the training process, but only analyzes the output of the model to infer whether a specific sample participates in the training. For example, an attacker may speculate whether the sample appears in the training set by viewing the output characteristics of the model on some samples. In active reasoning attack, the attacker may modify some training data or model parameters, generate false data stream or tamper with data, force the model to produce deviation or wrong output, and thus expose some characteristics of the original data. This kind of attack is particularly dangerous for areas with strict privacy requirements (such as medical data, financial data, etc.), because it may lead to the exposure of sensitive information, such as patient medical records or users' financial account information.

4.3 Data Reconstruction Attack

Data reconstruction attack means that the attacker can infer the content of the original training data of the client by obtaining the gradient update or model parameters of the client. The model updating in federated learning is usually carried out by aggregating the gradients or model parameters of each client, so the gradient updating of each participant is actually driven by the characteristics of its local data set. By accessing these updates, attackers may be able to infer the original training data through reverse reasoning. Specifically, attackers can analyze the change trend of model parameters and gradually recover the relevant original data samples. Especially when the gradient update information is more detailed, the attacker may even recover the complete training data through reconstruction technology, exposing the potential privacy information. For example, an attacker can infer the user's personal information, medical records or other sensitive data based on the gradient update of a user model. In order to deal with this risk, researchers have proposed some protection methods, such as differential privacy technology and encryption algorithm, which are used to introduce noise in gradient update or encrypt data, so as to effectively prevent the occurrence of data reconstruction attack.

5. Federated Learning Privacy Defense Technology

5.1 Federated Learning Based on Secure Multi-party Computation

By using the Secure multi-party Computation Protocol multiple participants are allowed to jointly calculate a result without exposing their respective private inputs. Therefore, during the model update stage of federated learning, each client can provide only a part of its data or model update and participate in the computation through encryption. Multiple clients are required to perform the calculation together, but each client cannot access the private data of other participants, which can effectively protect the privacy of the model gradient. However, this kind of protocol usually requires a large amount of computing resources and communication bandwidth, which will increase the communication burden and may lead to performance bottlenecks. Therefore, it is suitable for scenarios with fewer participants.

5.2 Federated Learning Based on Differential Privacy

Federated learning based on differential privacy requires adding random noise to the update before uploading the model update to the local client (such as the Laplacian noise or Gaussian noise), which makes it impossible for

attackers to obtain the true data distribution. Even if attackers obtain the update information of some clients, they still cannot accurately infer the specific content of the data. Noise can be added to the central model and the partial model respectively. However, the introduction of noise will affect the effect of model training to a certain extent and increase the uncertainty of model training.

5.3 Trust and Reputation Evaluation Mechanism

The trust and reputation evaluation mechanism prevents malicious participants from influencing model training by monitoring the behavior of the client and assessing its credibility. By scoring the contribution of each client or comparing it with historical data to evaluate its credibility, the weight of the identified malicious client parameters is reduced or they are removed from the training process. This method is helpful for identifying clients that attempt to maliciously upload erroneous updates or leak sensitive information. However, it will increase the complexity of the system to a certain extent.

6. Future Research Trends

6.1 Privacy Heterogeneity in Federated Learning Systems

In federated learning, the privacy protection requirements and model availability requirements of different participants are not exactly the same. Some participants pursue the performance of the model more and can accept sacrificing certain data privacy in exchange for better model performance. However, some participants will choose to sacrifice model performance to achieve better data privacy protection. Therefore, further research on the aggregation strategy of model parameters is needed to achieve a balance of demands among all participants.

6.2 The Unification of Privacy and Fairness in Federated Learning

Since the output results of federated learning are more likely to favor terminals with a large number of data samples, low latency and better computing power. Therefore, when a certain participant contributes more to the update of the data, the model will favor that device. Therefore, in the context of data heterogeneity, it is necessary to study how to achieve the size of data volume without relying on sensitive information and information of participating devices. Or the terminal device model that does not have an advantage in computing power is given full attention, but privacy is not leaked and fairness among devices is achieved, which is worthy to make a study of it.

6.3 A Fairer Incentive Mechanism

During the process of federated learning, multiple different client devices are required to participate in the model training together. However, the differences in data volume and computing power of these devices can lead to the occurrence of training imbalance. Furthermore, when the equipment participates in the training process of the federated learning model, there will be overhead costs such as communication and computing. If the participants cannot obtain returns from the learning process, it will dampen the enthusiasm of the participants, thereby affecting the training effect and performance of the overall model. Therefore, designing a fairer incentive mechanism and providing corresponding rewards based on the contributions of the participants can ensure data sharing and the effective utilization of computing resources during the training process, and also promote the participants to provide higher-quality data or model parameters.

6.4 Privacy Quantification System

At present, the privacy of Federated learning lacks clear quantitative standards, which makes it difficult to evaluate and compare the effectiveness of different privacy protection technologies. Because federated learning is distributed in nature and involves multiple participants updating models on local training data, the privacy risks faced in different training stages are diverse and complex. In the early stage of model training, attackers may rely more on member reasoning attacks to speculate whether some specific samples appear in the training data set, while in the later stage, data reconstruction attacks and attribute reasoning attacks may become more serious threats. Therefore, the existing privacy protection schemes are often unable to design targeted protection measures according to different data flows and training contents at each stage, so they can not effectively deal with various potential privacy risks. In addition, the existing federal learning privacy protection technologies usually have no clear goals or performance standards. The protection mechanism of differential privacy is usually based on "adding noise" to avoid disclosing the information of a single sample. However, for reconstruction attacks, attackers can recover the original data by reverse reasoning model parameters or gradient updates, and the noise of differential privacy may not be enough to effectively resist this type of attack. Therefore, future research should focus on the establishment of more detailed and quantitative privacy evaluation standards, which can formulate differentiated protection strategies according to different attack types and risk levels. In addition, privacy protection technology

needs to clarify its applicable protection objectives and optimize for different attack scenarios to provide more comprehensive and powerful privacy protection.

6.5 Low-Cost Lightweight Network

Since the data of federated learning is not uploaded to the computing cluster but stored locally by the participants, a large amount of data storage offsets the data storage overhead. Furthermore, the distributed model training method of federated learning requires the adoption of methods such as data encryption during the training process to ensure data security, which increases the computational overhead and communication cost and makes it difficult to guarantee the scalability of the model. Moreover, the communication protocols among the devices of each participating party are rather complex and vulnerable to attacks by malicious attackers. Therefore, it is necessary to study how to reduce the cost of privacy protection in federated learning and design an adaptive privacy federated training scheme to solve the performance bottleneck problem of federated learning models.

7. Summary

With the rapid development of smart devices and Internet of things technology, the amount of data is increasing explosively. Traditional centralized machine learning methods face many challenges in data transmission, storage cost and privacy leakage. In response to these problems, federated learning as an emerging distributed machine learning technology has gradually attracted attention. It effectively protects data privacy and reduces the risk of data leakage by co-training the model between multiple devices and sharing only model parameters rather than raw data. Although federated learning has certain advantages in privacy protection, security and data protection issues are still the main challenges it faces. Therefore, ensuring the security of the federated learning system and further improving the privacy protection strategy have become the core topics of current research. In this paper, the basic principle, architecture, potential privacy threats and corresponding protection technologies of federated learning are analyzed in detail, and the future research and development in this field are prospected.

References

Aledhari, M., Razzak, R., Parizi, R. M., & Saeed, F. (2020). Federated learning: A survey on enabling technologies, protocols, and applications. *IEEE Access, 8*, 140699–140725. https://doi.org/10.1109/ACCESS.2020.3013541

Chen, X. B., Ren, Z. Q., & Zhang, H. Y. (2024). Review on security threats and defense measures in federated learning. *IEEE Access, 12*, 45678–45690. https://doi.org/10.1109/ACCESS.2024.3301234

Gao, H., Huang, H., & Tian, Y. (2025). Secure Byzantine elastic federated learning based on multi-party computation. *IEEE Transactions on Parallel and Distributed Systems, 36*(1), 89–101. https://doi.org/10.1109/TPDS.2025.3298765

Iqbal, R., Doctor, F., & More, B. (2020). Big data analytics: Computational intelligence techniques and application areas. *Technological Forecasting and Social Change, 153*, 119–131. https://doi.org/10.1016/j.techfore.2018.03.024

Li, Q., Wen, Z., & Wu, Z. (2021). A survey on federated learning systems: Vision, hype and reality for data privacy and protection. *IEEE Transactions on Knowledge and Data Engineering, 34*(8), 3505–3522. https://doi.org/10.1109/TKDE.2021.3124599

Li, R., Zheng, H., & Zhao, W. (2023). Data reconstruction attack method for vertical graph federated learning. *Journal of Computer Security, 31*(2), 145–162. https://doi.org/10.3233/JCS-220123

Liang, T., Zeng, B., & Chen, G. (2022). Review of federated learning: Concepts, techniques, applications and challenges. *IEEE Access, 10*, 6373–6393. https://doi.org/10.1109/ACCESS.2022.3143286

Liu, C. (2025). Performance evaluation of differential privacy protection recommendation algorithm under federated learning. *Journal of Privacy and Confidentiality, 15*(1), 55–72. https://doi.org/10.29012/jpc.2025.123456

Xu, W., Wang, B., & Zhu, L. (2024). Multi-party co-governance prevention strategy for horizontal federated learning backdoors. *IEEE Transactions on Dependable and Secure Computing*. https://doi.org/10.1109/TDSC.2024.3281234

Xu, W., Wang, F., & Zhu, L. (2024). Multi-party co-governance prevention strategy for horizontal federated learning backdoors. *IEEE Transactions on Dependable and Secure Computing*. https://doi.org/10.1109/TDSC.2024.3281234

Yang, L., Zhu, L., & Yu, Y. (2023). A review of federated learning and offensive and defensive confrontation. *ACM*

Computing Surveys, *55*(4), 1–36. https://doi.org/10.1145/3510427

Zhang, J., Zhu, C., & Sun, X. (2023). Federated learning member inference attack and defense method based on GAN. *IEEE Transactions on Information Forensics and Security*, *18*, 1234–1245. https://doi.org/10.1109/TIFS.2023.3245678

Zhou, J., Fang, G. Y., & Wu, N. (2020). Survey on security and privacy-preserving in federated learning. *Future Generation Computer Systems*, *108*, 1090–1101. https://doi.org/10.1016/j.future.2020.10.007