

Research on Privacy Protection Technology in Federated Learning

Zihan Xiang^a

School of Spatial Information and Surveying Engineering, Anhui University of Science and Technology, Huainan, Anhui, 232001, China

Keywords: Federated Learning, Privacy Protection, Differential Privacy, Homomorphic Encryption.

Abstract: The extensive implementation of machine-learning techniques, the exponential expansion of big data, and the reinforcement of global legal provisions regarding data privacy safeguarding have spurred the swift advancement of federated learning. The primary benefit of federated learning is manifested in its capacity to carry out collaborative data training while refraining from the sharing of unprocessed data, which is crucial for protecting user privacy and complying with data protection regulations. This paper first summarizes the basic definition, classification, and algorithm principles of federated learning and then focuses on the applications of differential privacy and homomorphic encryption techniques within the privacy protection domain of federated learning. Differential privacy safeguards data privacy through the addition of noise when updating models. It yields a favorable outcome in terms of privacy protection, particularly within the medical sector. However, it encounters difficulties in achieving a balance between privacy protection and model accuracy, as well as in determining the value of the privacy budget. Homomorphic encryption enables direct calculations to be carried out on ciphertexts, achieving privacy protection throughout the entire process of federated learning. It has strong compatibility and wide applications but has high computational costs and low performance in large-scale distributed systems. In the future, privacy protection technologies for federated learning will develop towards multi-technology integration, adaptation to emerging scenarios, and standardization and normalization to address challenges such as inference attacks, data heterogeneity, and malicious attacks, promote the secure and compliant sharing of data, and facilitate the development of a digital society.


1 INTRODUCTION

In the digital era today, the rapid development of artificial intelligence technologies has significantly transformed the operational mode of society. Among them, federated learning, as an innovative distributed machine learning framework, has received extensive attention and research since it was proposed by Google in 2017 (McMahan, 2017). It allows multiple data holders to collaboratively train machine learning models without directly exchanging raw data, effectively solving many problems in traditional machine learning regarding data privacy protection and centralized training and opening up new ways for data collaborative utilization in the big data era.

As federated learning is increasingly applied, novel privacy problems have been gradually cropping up during the aggregation of distributed intermediate

outcomes. Yin deeply analyzed the privacy leakage risks in federated learning based on a newly proposed 5W scenario classification method and explored privacy protection solutions (Yin, Zhu, & Hu, 2021). This study offers thorough and profound references as well as guidance for research and practice within the realm of privacy- safeguarding federated learning. It propels further advancement and innovation in this domain.

The distributed nature of federated learning makes it face severe security problems, among which model poisoning attacks pose a significant threat to the security and performance of federated learning (Tolpegin et al., 2020). Wang's review concentrated on the countermeasures against model-poisoning attacks within federated learning. It also deliberated on the challenges, including the complexity of discerning attack approaches, the constraints of defense mechanisms, and the susceptibility of model

^a <https://orcid.org/0009-0009-9740-4821>

aggregation. Future research directions lie in studying the impact of different attack strategies on defense mechanisms and finding a balance among resource optimization, privacy protection, and defense effectiveness (Wang et al., 2022).

There are potential privacy leakage risks in various links of federated learning, such as parameter exchange during the training process, unreliable participants, and model release after training completion. For example, attack methods such as data reconstruction from gradients or inferring the source of records based on intermediate parameters have been proven feasible (Hu, Liu, & Han, 2019; Song, Ristenpart, & Shmatikov, 2017). Different from traditional centralized learning, federated learning faces more complex internal attacks, which greatly increases the difficulty of its privacy protection. When studying the privacy protection problem in federated learning, Liu found that the internal attackers in federated learning include the terminals participating in model training and the central server (Liu, 2021). Compared with external attackers, internal attackers have more training information and stronger attack capabilities (Liu, 2021).

At present, the approaches to privacy protection in federated learning are on the rise. Some scholars have put forward three solutions to the privacy-protection issue in federated learning: secure multi-party computation, differential privacy, and homomorphic encryption. This paper predominantly centers on the applications of differential privacy and homomorphic encryption techniques in safeguarding privacy for federated learning. It also summarizes and explores the most recent research advancements of relevant technologies. In addition, the review content of this paper encompasses the methods of applying federated-learning privacy-protection technologies grounded in differential privacy and homomorphic encryption to the medical field and deliberates on the future challenges and development of federated-learning privacy-protection technologies.

2 THE CONCEPT OF FEDERATED LEARNING

Traditional machine learning methods gather the data of all clients for learning. However, with data privacy and data security becoming issues, it is considered unsafe to centralize the original data of clients. To solve these problems, a new type of machine learning method called federated learning, which protects client data, has been proposed.

Federated learning is a distributed machine learning technology. Its core feature is that during the process of training a model, the original data of participants always remains local, and collaborative training is achieved only by exchanging model-related intermediate data (such as model update information, gradients, etc.) with the central server. This is in sharp contrast to the "model remains stationary, data moves" mode of traditional centralized learning, and it is a new learning paradigm of "data remains stationary, the model moves" (Liang, 2022). Its purpose is to break down data silos, enabling all parties to fully utilize the knowledge contained in multi-party data without exposing their own data privacy, enhancing the model's performance and maximizing the utilization of data value. For example, in the medical and financial fields, different institutions can jointly train models to improve diagnostic accuracy or risk assessment capabilities while protecting the privacy data of patients or clients.

Based on the distribution disparities in the feature space and sample space of the participants' datasets, federated learning can be categorized into horizontal federated learning, vertical federated learning, and federated transfer learning. Horizontal federated learning is suitable in scenarios where the feature spaces of the datasets of all parties exhibit substantial overlap while the sample spaces have only a minor degree of overlap. It usually involves joint training of data with similar features from different users. For example, numerous Android phone users, under the coordination of a cloud server, train a shared global input method prediction model based on their local data, making use of the data diversity of different users in the same feature dimension to improve the adaptability of the model to different users' input habits and prediction accuracy.

Vertical federated learning is fitting for circumstances where the sample spaces display a high degree of overlap, whereas the feature spaces have a relatively small amount of overlap. It generally involves the joint use of data generated by the same batch of users in different institutions or business scenarios. For example, a bank holds users' income and expenditure records, while an e-commerce platform possesses users' consumption and browsing records. The two parties conduct joint training based on the data of common users but different features to build a more accurate model for tasks such as customer credit rating, achieving cross-industry data integration, and collaborative modeling.

Federated transfer learning mainly focuses on datasets with little overlap in both the sample space and the feature space. It uses transfer learning

algorithms to transfer the trained model parameters or knowledge of one party to another party to assist in training its model, especially applicable in cases of scarce data or insufficient labeled samples. For example, among institutions in different countries or industries, transfer learning is used to overcome data distribution differences and sample shortages, expanding the application scope and generalization ability of the model.

Within the structure of federated learning, the key entities involved are participants and a central server. Participants use local data to build and train local models and send model-related information (such as gradients, loss values, etc.) to the central server after encrypting it according to specific encryption protocols. The central server receives information from all parties. It then employs secure aggregation methods like the Federated Averaging algorithm (FedAvg) and the Federated Prox algorithm to aggregate this information. By doing so, it generates global model update information, which is subsequently broadcast back to the participants. After receiving the global model update information, participants decrypt it and update their local models accordingly. This process iterates until preset stop conditions are met, such as model convergence or reaching a certain number of training rounds. Throughout the process, data privacy is ensured through encryption technologies and local data storage. At the same time, through the interaction and aggregation of model information, collaborative learning is achieved while avoiding the privacy risks caused by data centralization, the performance of the global model is gradually improved to get close to the effect of centralized learning.

3 PRIVACY PROTECTION TECHNOLOGIES IN FEDERATED LEARNING

3.1 Application of Differential Privacy Technology in Privacy Protection for Federated Learning

Differential privacy is designed to guarantee that during data analysis or model training, no sensitive information related to individual data elements is disclosed. It represents a technique for safeguarding data privacy (Dwork, 2008). By introducing noise, it renders the influence of an alteration in the original data on the output outcome insignificant. Federated learning safeguards data privacy by enabling

numerous devices to carry out local computations and share model updates, all without relaying raw data to the central server. However, in this distributed training, the gradients or model updates calculated by each participant may reveal sensitive information about local data. Even if the data itself is not directly exchanged, in some cases, the gradient updates of the model can still reflect the characteristics of the data. Therefore, how to effectively conduct joint model training without leaking data privacy has become a major challenge in federated learning (Xiao et al., 2023). To solve this problem, differential privacy technology provides a feasible solution. It prevents gradients and parameters from revealing detailed information about local data by adding noise during the model update process.

Mao's research pointed out that the core concept of differential privacy is the "privacy budget ϵ value," which determines the noise intensity and the effect of privacy protection (Mao, 2024). In real-world applications, differential privacy commonly ensures that the particulars of a particular data point remain undisclosed within the data analysis findings. This is accomplished by incorporating noise into the data, either via the Laplace mechanism or the Gaussian mechanism. It should be emphasized that a lower ϵ value is associated with more robust privacy protection, while a greater quantity of noise implies less effective privacy protection (Tang, 2023).

Privacy protection for model training in the medical field is essential, and differential privacy technology is very effective in protecting patients' privacy data. Medical institutions can share training models of medical images or patients' health data through federated learning without actually exchanging any sensitive data of patients. By adding differential privacy noise to each device, the privacy of patients can be ensured not to be leaked. Liu proposed a medical data sharing and privacy protection scheme based on federated learning. By combining blockchain technology, decentralized model aggregation is achieved, and a hybrid on-chain and off-chain storage method is used to reduce communication costs. To protect model parameters, differential privacy noise is introduced at the local model training stage simultaneously (Zhang, 2024). Experiments indicate that the model attains the highest and most consistent accuracy when noise is introduced prior to the activation function of the second layer within the fully - connected layer. This approach is capable of achieving high accuracy while safeguarding privacy. It effectively addresses the issue of medical data silos and fortifies the security of data sharing (Liu, 2023). Zhang combined the

differential privacy method and proposed a differential - privacy-based, decentralized federated learning protocol (PADFL). This protocol realizes anonymous authentication and privacy protection of nodes by combining blockchain and smart contract technologies (Zhang, 2024). The immutability and traceability of the blockchain are utilized to ensure the security and transparency of the model training process in a situation where a central server is lacking. Simultaneously, this protocol incorporates differential privacy technology. Through the addition of Gaussian noise to local model updates, it effectively thwarts malicious or inquisitive nodes from deducing the privacy information of other nodes via model parameters. Through a decentralized approach, it improves the security and flexibility of the chest disease classification system. Additionally, it is integrated into this system to preserve the privacy of patients during the training procedure (Zhang, 2024).

Although differential privacy technology has made significant progress in federated learning, there are still some challenges. First, although adding noise can effectively protect privacy, in tasks that require high accuracy, the model accuracy may also be affected. Therefore, the current research focus is on how to improve the performance of the model while ensuring privacy. Second, the selection of the privacy budget (ϵ value) is crucial. Although a small ϵ value can provide strong privacy protection, it may lead to a decline in model performance due to its smallness. On the other hand, a large ϵ value may reduce the effect of privacy protection. Consequently, the issue of how to adaptively modify the privacy budget in line with diverse scenarios and demands in real-world applications will emerge as a crucial subject in the future.

3.2 Application of Homomorphic Encryption Technology in Federated Learning

Homomorphic Encryption (HE for short) is an encryption technology that allows mathematical operations to be performed on encrypted data in the ciphertext state without decrypting it first (Li et al., 2020). This means that data is processed in an encrypted state, and the result obtained after decryption is the same as that obtained by directly operating on the original data. Different from traditional encryption methods, homomorphic encryption not only protects data privacy but also ensures that data is not exposed during the entire processing process. Based on the kinds of operations

that homomorphic encryption can support, it can be categorized into two types: partial homomorphic encryption and fully homomorphic encryption. Fully homomorphic encryption is a cryptographic technology that allows arithmetic operations to be directly performed on encrypted data without decrypting it (Li, 2024). The fundamental concept of fully homomorphic encryption is to perform operations on encrypted data. The outcome is identical to that achieved by encrypting the result of the same operation carried out on plain-text data. In this way, computational tasks can be accomplished while safeguarding data privacy.

Federated learning aims to collaboratively train a machine-learning model across multiple distributed devices. By refraining from uploading local data to the central server, it safeguards data privacy. Even though federated learning has the ability to prevent direct data exchange, there is still a risk that the local data of participants could be divulged via the uploaded model parameters or gradients. Homomorphic encryption technology provides a feasible solution for solving this problem. It can perform encrypted processing and calculations without decrypting the data.

Homomorphic encryption technology plays an important role in federated learning. By allowing direct computation on ciphertexts without decryption, it achieves comprehensive privacy protection in the entire process of client-side data encryption and upload, server-side aggregation calculation, and global model distribution, effectively preventing the leakage of local data and model parameters during transmission and processing. In addition, homomorphic encryption technology has a high degree of compatibility and can be seamlessly integrated into existing federated learning frameworks, such as the Federated Averaging algorithm, without the need for large-scale modifications to the basic process. It also has a wide range of applicability and is not only suitable for federated learning scenarios but also for other fields that require privacy protection, such as cloud computing and distributed machine learning (Jiang, 2024).

In practical applications, many industries have adopted homomorphic encryption technology. Protecting patient privacy is very important in the medical industry, especially in scenarios such as medical data sharing and joint diagnosis. To ensure that the sensitive medical data of hospitals is not leaked, medical institutions can use homomorphic encryption technology to encrypt data. For example, hospitals can use homomorphic encryption and

masking protocols to protect the privacy of medical data model parameters. Through aggregating these encrypted updates, the central server is able to train a more precise diagnostic model, all the while ensuring the complete protection of patients' privacy (Niu, 2024).

Although homomorphic encryption has great application potential in federated learning, its computational cost in large-scale distributed systems is high, and its performance is also inefficient. An important direction for future development is the optimization of homomorphic encryption algorithms and the reduction of computational overhead. Currently, many studies are exploring how to reduce the costs of encryption and decryption operations while improving the overall efficiency of federated learning. With the continuous improvement of hardware performance and the optimization of homomorphic encryption algorithms, homomorphic encryption is expected to be applied in more fields, such as medical and finance, especially in industries with high requirements for data privacy. More flexible and efficient solutions for the development of privacy protection in federated learning can be provided by, at the same time, combining with other privacy protection technologies such as differential privacy and secure multi-party computation.

4 CONCLUSIONS

Federated learning, regarded as an innovative framework for distributed machine learning, breaks down data silos while protecting data privacy and has broad application prospects in many fields. However, with the deepening of applications, privacy protection issues have become prominent. The differential privacy and homomorphic encryption technologies focused on in this paper have become the key paths to solving this problem. In the privacy protection of federated learning, differential privacy effectively prevents the leakage of sensitive information by adding noise to model updates, especially showing remarkable effects in medical data-sharing scenarios. However, it encounters challenges when it comes to striking a balance between privacy protection and model accuracy. The determination of the privacy budget value is of utmost importance. A too-small value will reduce model performance, while a too-large value will weaken the effect of privacy protection. Subsequent research needs to focus on accurately and dynamically adjusting the privacy budget according to different tasks and data characteristics to maximize model performance while

protecting privacy. Homomorphic encryption technology allows direct operations on ciphertexts, achieving comprehensive privacy protection in the entire process of federated learning. It has good compatibility with existing federated learning frameworks and a wide range of application scenarios. However, it has high computational costs and low performance in large-scale distributed systems. In the future, the optimization of encryption algorithms and the reduction of computational overhead will be the key development directions. With the improvement of hardware performance and algorithm improvements, homomorphic encryption is expected to be widely applied in high-privacy-demand industries such as medicine and finance. Despite the fact that federated learning boasts substantial advantages in safeguarding data privacy, it still encounters numerous challenges.

Although the data is retained locally, the analysis of the uploaded gradient and model update data may infer the user's data, thus leaking local sensitive information. Data heterogeneity increases the complexity of privacy protection. The inconsistent data distribution of different participants may lead to data leakage or a decline in model performance. In addition, federated learning also faces malicious attack risks such as model poisoning and data poisoning attacks, and the current research focus is on how to prevent these attacks. Overall, privacy protection technologies for federated learning are in a stage of rapid development. In the future, multi-technology integration will become the main development trend. By organically combining differential privacy, homomorphic encryption, and other privacy protection technologies, complementary advantages can be achieved, and a complete privacy protection system can be constructed for federated learning. At the same time, with the continuous evolution of artificial intelligence technologies, privacy protection technologies for federated learning need to continue to innovate to actively adapt to the requirements of emerging application scenarios such as edge computing and the Internet of Things and explore more suitable privacy protection strategies. In addition, strengthening relevant standardization and normalization research and establishing unified evaluation standards and security specifications will strongly promote the wide application of privacy protection technologies for federated learning in various industries, realizing the secure and compliant sharing of data and laying a solid foundation for the stable development of a digital society.

REFERENCES

- Dwork, C., 2008. Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation* (pp. 1–19). Springer Berlin Heidelberg.
- Jiang, H., 2023. Research on Privacy-Preserving Federated Learning Based on Homomorphic Encryption (*Master's thesis, Nanjing University of Science and Technology*). Master's Degree.
- Liang, T., Zeng, B. and Chen, G., 2022. A Review of Federated Learning: Concepts, Technologies, Applications, and Challenges. *Journal of Computer Applications*, 12, pp. 3651–3662.
- Li, T., Sahu, A.K., Talwalkar, A. and Smith, V., 2020. Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), pp. 50–60.
- Li, Q. and Zhou, Q., 2024. Research on Privacy Protection Technology for Federated Learning Based on Fully Homomorphic Encryption. *Modern Information Technology*, 23, pp. 170–174.
- Liu, Y., Chen, H., Liu, Y. and Li, C., 2021. Privacy protection technologies in federated learning. *Journal of Software*, 33(3), pp. 1057–1092.
- Liu, Z., Li, H., Wu, L. and Qin, Y., 2024. Medical data sharing and privacy protection based on federated learning. *Computer Engineering and Design*, 45(9), pp. 2577–2583.
- McMahan, B., Moore, E., Ramage, D., Hampson, S. and Arcas, B.A., 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics* (pp. 1273–1282). PMLR.
- Mao, Y., 2024. Research on Topology Optimization Federated Learning Algorithm Based on Differential Privacy (*Master's thesis, Nanjing University of Information Science and Technology*). Master's Degree.
- Niu, S., Wang, N., Zhou, X., Kong, W. and Chen, L., 2024. A secure federated learning scheme based on secret sharing and homomorphic encryption in smart healthcare. *Computer Engineering*, 1–13.
- Song, C., Ristenpart, T. and Shmatikov, V., 2017. Machine learning models that remember too much. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 587–601).
- Tang, L., Chen, Z., Zhang, L. and Wu, D., 2023. Research progress on privacy issues in federated learning. *Journal of Software*, 34(1), pp. 197–229.
- Tolpegin, V., Truex, S., Gursoy, M.E. and Liu, L., 2020. Data poisoning attacks against federated learning systems. In *Computer Security–ESORICS 2020: 25th European Symposium on Research in Computer Security*, ESORICS 2020, Guildford, UK, September 14–18, 2020, Proceedings, Part 1 (pp. 480–501). Springer International Publishing.
- Wang, Z., Kang, Q., Zhang, X. and Hu, Q., 2022. Defense strategies toward model poisoning attacks in federated learning: A survey. In *2022 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 548–553). IEEE.
- Xiao, X., Tang, Z., Xiao, B. and Li, K., 2023. A review of privacy protection and security defense in federated learning. *Chinese Journal of Computers*, 46(5), pp. 1019–1044.
- Yin, X., Zhu, Y. and Hu, J., 2021. A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions. *ACM Computing Surveys (CSUR)*, 54(6), pp. 1–36.
- Zhang, M., 2024. Research on Decentralized Federated Learning Protocols Based on Differential Privacy (*Doctoral dissertation, University of Electronic Science and Technology of China*).
- Zhu, L., Liu, Z. and Han, S., 2019. Deep leakage from gradients. *Advances in Neural Information Processing Systems*, 32.