Research on the Security Technology of Federated Learning
Privacy Preserving

Juan Mao*, Chunjie Cao1,2, LongJuan Wang1,2, Jun Ye1,2, WenJie Zhong1,2
*1School of Computer Science and Cyberspace Security, Hainan University, No.58, Renmin Road, Haikou 570228, China
2Key Laboratory of Internet Information Retrieval of Hainan Province, Hainan University, No.58, Renmin Road, Haikou 570228, China

*cs_maojuan@163.com

Abstract: With the emergence of data islands and the popular awareness of privacy, federated learning, as an emerging data sharing and exchange model, can realize multi-party collaboration under the premise of protecting data privacy and security because the data distributed in multiple devices cannot be sent locally. To achieve benefits for all parties involved, it has been widely used in many fields such as finance, medical care, and education. However, FL also has various security and privacy issues. Starting from the overview of federated learning, this article describes in detail the threat model and existing security issues, including replay attacks, poisoning attacks, reasoning attacks, etc., and then makes a certain analysis of FL privacy protection security technologies. Compared with SMC and HE, differential privacy is excellent in terms of efficiency. Finally, we discussed the challenges of privacy protection and security issues and future research directions.

Keywords: Deep learning, federated learning, privacy-preserving, security technology, data security, communication.

1. Introduction
With the development and popularization of artificial intelligence technology, the rise of big data, the gradual improvement of policies and regulations, and the strengthening of public privacy protection awareness, considering user privacy and commercial secrets, it is difficult for enterprises to exchange
original data and services between various independent industries, so there is a problem of "data islands" in cross-enterprise collaborative governance cooperation.

Regarding the above problems, in 2017, Google proposed federated learning, an encrypted distributed machine learning technology that can be used to a large extent for mobile devices to complete data model training without infringing on user privacy [1]. The specific implementation strategy of federated learning today is to establish a virtual shared model, which is similar to the optimal model built by aggregating data. Under this federated mechanism, the identity and status of the older brother participants are the same, realizing multi-party cross-data Collaborate and benefit together.

However, there are still huge security risks in federated learning. This article analyzes the possible security problems of federated learning, to ensure the safety of federated learning and promote its further development and popularization.

2. Overview of Federated Learning

2.1. Definition
Federal learning refers to a security solution that meets the requirements of privacy protection and security supervision [1]. That is to design a machine learning framework so that the original data distributed on multiple devices does not need to be shared, and the training results are trained locally and encrypted [2].

As shown in Fig.1 below, the definition of federated learning: [3] Assume that there are N data set owners \( \{F_1, \ldots, F_N\} \), and they all hope to train machine learning by merging their respective data sets \( \{D_1, \ldots, D_N\} \) model. The traditional approach is to put all the data together and use \( D = D_1 \cup \ldots \cup D_N \) to train the model \( M_{SUM} \). Federated learning is the process of each data owner cooperating to train the model \( M_{FED} \). In the model training process, the data owner \( F_i \) does not need to send its own data \( D_i \) to other data holders. At the same time, the accuracy \( V_{FED} \) of the model \( M_{FED} \) should be very close to the performance of the model \( M_{SUM} \). Finally, assume that \( \delta \) is a small positive number, such as \( |V_{FED} - V_{SLM}| < \delta \), that is, the accuracy error of the federated learning algorithm is \( \delta \).
Fig. 1 Federated Learning Architecture

2.2. Classification
According to the distribution characteristics of the data, as Fig. 2:

1) Horizontal federated learning: It is suitable for the situation where the user characteristics of party A and party B overlap more, but the users of both parties overlap less, and the data set is segmented according to the user dimension, and the user characteristics of both parties are the same but the users are not exactly the same. That part of the data for training;

2) Vertical federated learning: It is suitable for the users of A and B to overlap more. According to the fact that the characteristics of the users of both parties are less overlapped, we divide the data set according to the vertical (feature dimension) and extract the same users of both parties. The part of the data with not exactly the same features is trained;

3) Federated transfer learning: Both A and B have relatively few users features and user overlaps, and the data is not segmented, and transfer learning [4] is used to help customers with insufficient data or labels Happening.

(a) Horizontal federated learning  (b) Vertical federated learning   (c) Federated Transfer learning

Fig. 2 Categorization of Federated Learning

3. Security issues
As a new training model with neural networks, the adversary can track and obtain the privacy of participants from the shared gradient, federated learning still faces various security and privacy threats.

3.1. Three main attack types of machine learning
1) Completeness: detect errors and detect the intrusion point as normal.
2) Usability: classification errors (side negative and side positive), which is a broader type of attack.
3) Confidentiality: Leakage of sensitive information.

3.2. Threat and Security Model

3.2.1. Reconstruction Attacks
In the local model training phase, if the data structure is consistent, the gradient information is likely to be used, thereby leaking additional information about the training data. You should avoid using machine learning models that store and display feature values [5].

3.2.2. Model inversion attack
An adversary with black box authority may implement an equation solving attack. An adversary with Baihe authority can obtain the plaintext content of the model without storing feature vectors [6].

3.2.3. **Reasoning attack**

Even if the data is independently trained locally, if there is a malicious participant, the hidden content of other partners can be derived through known shared parameters. Literature [5]. An attack is implemented that can infer the type of traffic used in the model building process. The training set information inferred by the model reverse attack can be whether a certain member is included in the training set or some statistical characteristics of the training set.

3.2.4. **Poisoning attack**

According to the source of the poisoning model update, the poisoning attack can be divided into data poisoning and model poisoning:

Model poisoning: Model poisoning does not modify the training data, but directly generates model poisoning according to some predefined rules. For example, [7] believes that the poisoning model update can be sampled from the Gaussian distribution. In addition, an attacker can manipulate a benign model update into a poisoned model update. In addition, in [4], [8], and [9], the attacker uses a pre-designed damaged model deliberated to make a poisoning model update.

3.2.5. **Counter Attack**

Adversarial attacks refer to maliciously constructing input samples, causing the model to output wrong results with high confidence. This input sample generated by adding disturbance to the original sample is called an adversarial sample [10].

4. **Privacy protection technology based on federated learning**

4.1. **Secure multi-party computation**

Secure Multiparty Computing (SMC) is a solution to the problem of collaborative computing, which protects the privacy of a group of untrusted parties without trust between trusted third parties. SMC must ensure the confidentiality, independence, and accuracy of the information of all parties in the calculation.

The SMC security model naturally involves many parties, and all parties do not know anything except their inputs and outputs to ensure a complete zero-knowledge security certification [10].

4.2. **Homomorphic encryption**

In [11], Hardy, S., Henecka, W et al., adopted a homomorphic encryption method to protect user data through parameter exchange using an encryption mechanism. Such as literature [12]. The homomorphic encryption scheme is used to achieve the aggregation of gradients on the honest and curious server, and to ensure that the system reaches the corresponding depth of training on the joint data set with all participants. The same accuracy as the learning system.

4.3. **Differential privacy**
The formula of differential privacy: A randomized algorithm $A(D)$ satisfies $\varepsilon$-differential privacy if for all datasets $D$ and $D'$, that differ by a single record, and for all sets $S \subseteq R$, where $R$ is the range of $A$,

$$\Pr[A(D) \in S] \leq e^{\varepsilon} \Pr[A(D') \in S]$$

where $\varepsilon$, a privacy parameter, is a non-negative number. That is, the output distribution of the algorithm is less affected by any record in the data set [13-14].

In this way, data is usually achieved by adding noise to the updated parameters by a trusted third party to prevent curious participants from inferring training data [15].

4.4. Confrontation training

The real samples and adversarial samples are used as the training set to train the final model. Adversarial training is suitable for a variety of supervision problems [16]. It can make the model learn the characteristics of adversarial samples during the training process and improve the robustness of the model.

Adversarial attacks can help malware evade detection and generate poisoned samples. In terms of attack environment, adversarial attacks can be divided into black-box attacks and white-box attacks. According to the intensity of interference, it can be divided into an infinite norm attack, two norm attacks, and zero norm attack. Nowadays, we are gradually studying attack methods against attacks, such as Least-Likely-Class Iterative Methods [17], DeepFool [18], etc.

5. Future Directions

Federated learning can realize the training and calculation of feature data on the cloud while maintaining the local confidentiality of the data of all parties and can be widely used in environments such as telemedicine in the future [19].

Privacy-protection algorithm: Both theoretically and empirically, privacy protection algorithms should be designed. In addition, the trade-off between privacy level and convergence speed needs further study [20].

Communication security: Such as (DoS) and interference attacks due to the exposed characteristics of wireless media [21].

Heterogeneous model: The model system is different due to different application scenarios and the size of the data volume [22].

6. Conclusion

In this article, we outline federated learning, which is a machine learning paradigm that uses a decentralized method to directly learn from user data while protecting the privacy of local data. The statistical model is based on a distributed network. The fringe was trained. However, there are also data privacy protection issues. We describe the security technical issues of privacy protection on a large scale. Finally, we discussed the challenges of privacy protection and security issues and future research directions.

Federated learning is a new paradigm in big data application and a new way to solve the problem of data privacy protection. With the advent of the 5G era, the future can further unite financial institutions, the Internet, medical care, education, smart cities, etc., to highlighting the advantages of
federated learning data privacy protection, will help break industry data barriers and realize secure socialized big data applications, which have broad application prospects.

7. Acknowledgments

This work was supported by Natural Science Foundation of Hainan Province under grant no.618MS025 and National Natural Science Foundation of China under grant no. U19B2044.

References

[1] White book of Federated Learning V2.0 https://aisp-1251170195.cos.ap-hongkong.myqcloud.com/wp-content/uploads/pdf/%E8%81%94%E9%82%A6%E5%AD%A6%E4%B9%A0%E7%99%BD%E7%9A%AE%E4%B9%A6_v2.0.pdf

[2] Yang Q 2019 AI and data privacy protection: the solution to Federated learning Research on Information Security 5(11) 961-965

[3] Yang Q, Liu Y, Chen T, and Tong, Y 2019 Federated machine learning: Concept and applications ACM Transactions on Intelligent Systems and Technology (TIST) 10(2) 1-19

[4] Lu, H, Setiono, R, and Liu, H 1996 Effective data mining using neural networks IEEE transactions on knowledge and data engineering 8(6) 957-961

[5] Atieniese G, Felici G, Mancini L V, et al 2013 Hacking Smart Machines with Smarter Ones: How to Extract Meaningful Data from Machine Learning Classifiers J. arXiv preprint arXiv:13064447

[6] Wu, X, Fredrikson, M, Jha, S, and Naughton, J F 2016 A Methodology for Formalizing Model-Inversion Attacks 2016 IEEE 29th Computer Security Foundations Symposium (CSF) IEEE

[7] Xu, G, Li, H, Liu, S, Yang, K, and Lin, X 2019 Verifynet: Secure and verifiable federated learning IEEE Transactions on Information Forensics and Security 15 911-926

[8] Zhang, J, Chen, J, Wu, D, Chen, B, and Yu, S 2019, August Poisoning attack in federated learning using generative adversarial nets In 2019 18th IEEE International Conference on Trust, Security, and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE) IEEE pp 374-380

[9] Du, W, Han, Y S, and Chen, S 2004, April Privacy-preserving multivariate statistical analysis: Linear regression and classification In Proceedings of the 2004 SIAM international conference on data mining Society for Industrial and Applied Mathematics pp 222-233

[10] Bogdanov, D, Laur, S, and Willemson, J 2008, October Sharemind: A framework for fast privacy-preserving computations In European Symposium on Research in Computer Security. Springer, Berlin, Heidelberg pp 192-206

[11] Hardy, S, Henecka, W, Ivey-Law, H, Nock, R, Patrini, G, Smith, G, and Thorne, B 2017 Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption arXiv preprint arXiv:1711.10677

[12] Aono, Y, Hayashi, T, Wang, L, and Moriai, S 2017 Privacy-preserving deep learning via additively homomorphic encryption IEEE Transactions on Information Forensics and
Choudhury, O, Gkoulalas-Divanis, A, Salonidis, T, Sylla, I, Park, Y, Hsu, G, and Das, A 2019
Differential privacy-enabled federated learning for sensitive health data arXiv preprint arXiv:1910.02578

Li Q, Wen Z, Wu Z and et al 2019 A survey on federated learning systems: vision, hype and reality for data privacy and protection J. arXiv preprint arXiv:1907.09693,

Kulkarni V, Kulkarni M, Pant A Survey of Personalization Techniques for Federated Learning J. arXiv preprint arXiv:2003.08673, 2020

Miyato, T, Maeda, S I, Koyama, M, and Ishii, S 2018 Virtual adversarial training: a regularization method for supervised and semi-supervised learning IEEE transactions on pattern analysis and machine intelligence, 41(8), 1979-1993

Kurakin, A, Goodfellow, I, and Bengio, S 2016 Adversarial examples in the physical world

Moosavi-Dezfooli, S M, Fawzi, A, and Frossard, P 2015 Deepfool: a simple and accurate method to fool deep neural networks

Li, T, Sahu, A K, Talwalkar, A, and Smith, V 2020 Federated learning: Challenges, methods, and future directions IEEE Signal Processing Magazine, 37(3), 50-60

Lo, S K, Lu, Q, Wang, C, Paik, H, and Zhu, L 2020 A systematic literature review on federated machine learning: From a software engineering perspective arXiv preprint arXiv:200711354

Lim, W Y B, Luong, N C, Hoang, D T, Jiao, Y, Liang, Y C, Yang, Q, and Miao, C 2020 Federated learning in mobile edge networks: A comprehensive survey IEEE Communications Surveys and Tutorials

Aledhari, M, Razzak, R, Parizi, R M, and Saeed, F 2020 Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications. IEEE Access 8 140699-140725