Federated Learning-Based Privacy-Preserving and Security: Survey

Mengna Yang, Yejun He*, Jian Qiao
Guangdong Engineering Research Center of Base Station Antennas and Propagation
Shenzhen Key Laboratory of Antennas and Propagation
College of Electronics and Information Engineering, Shenzhen University, 518060, China
Email:2019115452@qq.com, heyejun@126.com*, 446941582@qq.com

2021 Computing, Communications and IoT Applications (ComComAp) | 978-1-6654-2798-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/COMCOMAP53641.2021.9653016

***Abstract*—Traditional machine learning (ML) algorithms need to collect a large mount of users' data for model training, which result in privacy leak and "data islands" problems emerge in endlessly. In order to solve the above problems, federated learning (FL) has emerged as an outstanding tool. FL is widely used for the sixth generation mobile network (6G) communications, artificial intelligence, and privacy-preserving applications. This article starts from the concept of FL, introduces the research status of FL algorithms and privacy-preserving technology, and further explains some of the current applications and future challenges. Although the FL has brought dawn, it still faces many challenges in terms of enhancing privacy-preserving and training model security. Communication overhead is a problem in the encryption process; the noise threshold of different scenarios needs to be solved in the process of noise handling; how to identify malicious attackers, and reduce malicious attacks is also a worth noticing challenge in modeling process.**

***Index Terms*—Federated learning (FL), Privacy-preserving, Security**

I. Introduction

With the rapid development of artificial intelligence (AI), big data and edge computing, communication technologies have many applications in transportation [1], medical [2] and so on. At the same time, due to the rapid growth of data, a new challenge has been emerged. Different companies have different data type and companies in the same type will generate different data type. Therefore, users pay more attention to the protection of data privacy while enjoying the benefits of artificial intelligence and other technologies.

Traditional machine learning (ML) algorithms require large amount of data for model training, which generates a lot of computational overhead. It is not easy to pursue training on one device, and even complete training requires centralized storage of large-scale data, which causes in some privacy leakage. In some fields, such as medical and bank, more private data is involved in the training of the model. The "Didi Application", which was reported to be temporarily removed for leaking users' privacy data. In fact, most applications also have the risk of privacy leakage. Generally, as long as you register, they are granted to access photo albums and mobile contacts, which can easily cause in information theft opportunities for illegal elements.

In order to solve these problems, federated learning (FL) came into being. In 2016, Google proposed the concept of FL [3]. Later, Yang Qiang *et al.* held that FL refers to the design of a ML architecture that allows users to build a shared learning model [4]. User data is trained locally, then is uploaded to a trusted third party after encrypting or adding some noise, and finally this process is repeated until the model converges. The process is shown in Fig. 1. This not only solves the data islands, but also helps in protecting the data of participants. However, some malicious participants reversely may infer some sensitive data of users according to the difference of the gradient parameters in each round, or some malicious users upload errors to interfere the accuracy of the entire model. At present, a large number of scholars have studied the defense of malicious attacks and privacy-preserving technology. FL is still in early stages of development in some field, and a comprehensive review of the relevant literature may be especially helpful in synthesising the key research insights and unveiling major research trends in this field. Although some scholars have made a comprehensive review of privacy-preserving technology and FL algorithms, few of them have summarized specific applications and future challenges (such as 6G). Hence, this study intends to overview a comprehensive survey on the security and privacy-preserving, sorts out current applications algorithms, and elaborates part of the challenges

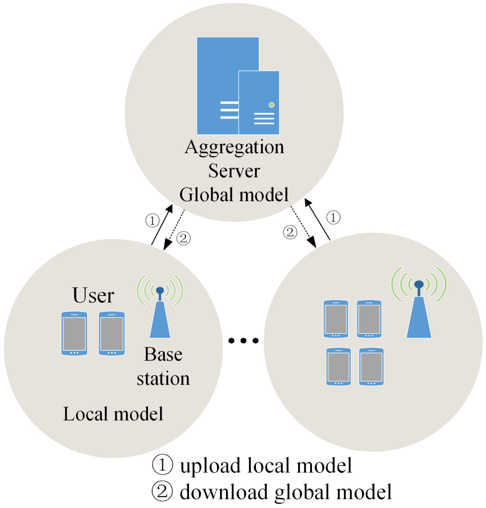


Fig. 1. The process of federated learning.

by reviewing relevant literature.

II. Classifications of Federated Learning

The FL can be divided into three categories according to the amount of overlap between the user part or the user-owned data feature, as shown in Fig. 2.

A. User Dimension Segmentation

When the user has less overlap and the user data features have more overlap, the data set can be segmented according to the user dimension. For example, there are two hospitals from different regions, one of which is located in Shenzhen and the other is located in Shanghai. The two hospitals treat diseases of similar types, but have less user intersection since they are far apart. Therefore, only the data features may overlap, so that it can be segmented according to the user dimension to build a FL model.

B. Feature Dimension Segmentation

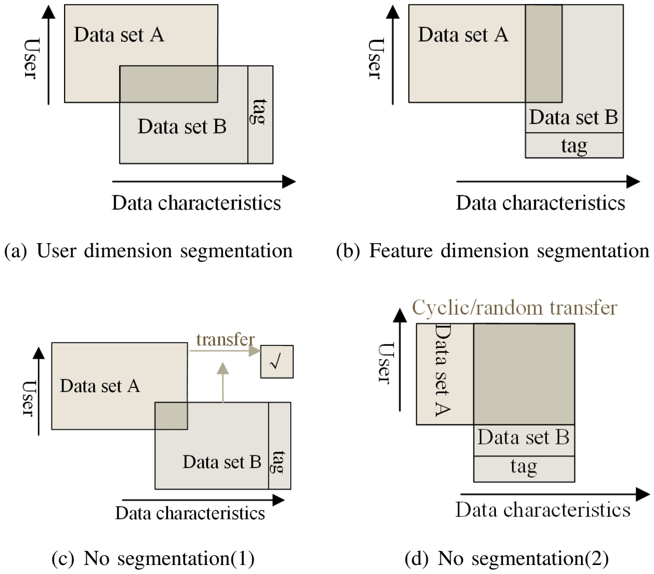
When the user part overlaps more and the data feature overlap is less, the data set can be segmented according to the feature dimension. For example, there are two institutions, one is a bank in Shenzhen, and the other is a large local shopping mall nearby. The bank records the users credit level and income level, and the shopping mall records the users purchasing capability. They share little common in user data features but more in users themselves, because they are geographically close and most of their users are local residents. The situation can be segmented according to the feature dimensions, and these different features data can be encrypted and aggregated to train a better FL model.

C. No Segmentation

When both users and user-owned data share less in common, the transfer learning (TL) can be a good method for model training. However, if they have a big overlaps, the random or cyclic transfer based on peer-to-peer network is more appropriate. For example, there are two institutions from different regions, one is a hospital in Shanghai, and the other is a bank in Shenzhen. These two institutions are of different types and they are located in two distant regions, so the users and data-owned features overlap possibility is very less. In this case, the TL can be used to train the global model. For another example, if we consider China Merchant Bank in Shenzhen and the other is Bank of China in Shenzhen, then these two in similar businesses and users. In this case, you can use a peer-to-peer network or cyclic transmission to train model without third part.

III. Privacy-Preserving and Security Based on FL: Status

This section expands from the two aspects of enhancing user privacy-preserving and data security. Table I lists recent studies related to enhanced privacy-preserving and security technologies.



(a) User dimension segmentation (b) Feature dimension segmentation

(c) No segmentation(1) (d) No segmentation(2)

Fig. 2. Classifications of federated learning.

A. Privacy-Preserving

Regarding the enhancement of privacy-preserving in FL, most of the studies are based on encryption and noise addition, and a small number of hybrid mechanisms have been found so far. It can combine encryption and noise addition. Rui Hu *et al.* added noise to improve the level of privacy-preserving, capture the individual characteristics of heterogeneous users, and train an effective personalization model [5]. Kang Wei *et al.* proposed a framework of adding noise before model aggregation based on dynamic programming to improve privacy-preserving, that is, adding Gaussian noise before model aggregation to disturb local training parameters [6]. In the resource sharing of the Internet of Vehicles, there are also many studies to improve privacy-preserving by adding noise. Y. Lu *et al.* enhanced the privacy by adding noise to the model parameters [7]. In addition to the method of adding noise, the encryption can also be applied for privacy purposes. Bo Yin *et al.* proposed a secure collaboration framework based on federated deep learning (FDL). They considered blockchain and encrypted transmission data to realize multi-party secure computing [8]. This is a decentralized FL process. Recently, Yong Li *et al.* proposed an innovative chain-based privacy-preserving framework based on encryption technology, which achieved a good balance between the communication efficiency and privacy-preserving [9]. Most studies also combined noise and encryption in many industrial scenarios driven by sensitive data. In [10], the author added Gaussian noise to counter data attacks, and used blinding and Paillier homomorphic encryption to defend attacks on the trained model. Compared with separate noise or encryption, the stronger privacy-preserving can be achieved, but it still fails to prevent multiple entities from colluding with each other. Recently, Meng Hao *et al.* proposed a non-interactive privacy-enhanced FL framework, which can prevent data leakage and collusion

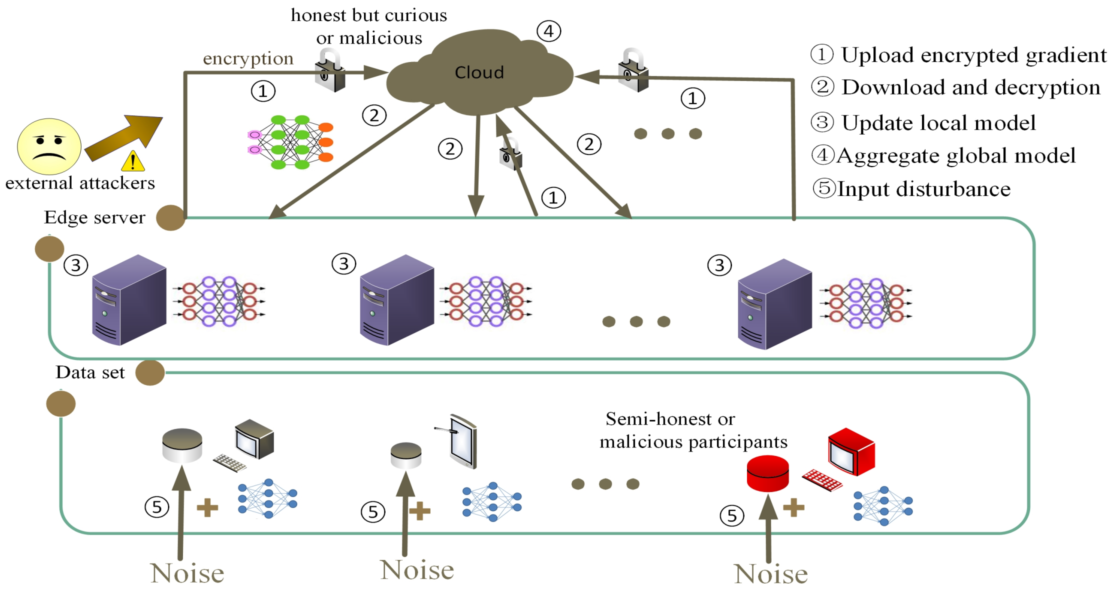


Fig. 3. The process of encryption and noise in federated learning.

between multiple data owners [11]. The process of encryption and noise in FL are shown in Fig. 3.

B. Threat Security Model

Recently, the main security threat models addressed are based on honest but curious or malicious aggregation servers, honest but curious or malicious participants, and malicious external attackers. Some of them maliciously collect the data and steal sensitive information; Some malicious participants collude with each other, and deliberately provide low-quality of data which effects the accuracy of trained model; Some attackers may use the trained model parameters to infer the users' data information. Rui Hu *et al.* proposed a privacy enhancement method based on adding noise to defense honest but curious central servers or data owners [5]. Though it has a significant effect on the user privacy, the attacks from external opponents are ignored. A model proposed by Kang Wei *et al.* added the noise frame before aggregation [6]. They have considered the channel security of the uplink and downlink to prevent attacks from external opponents. Y. Lu *et al.* proposed an asynchronous joint learning scheme based on the local noise addition that can improve the accuracy of the model [7]. It can protect user privacy and prevent malicious clients from deliberately sending wrong or low-quality data. Yong Li *et al.* used encryption technology to propose an innovative chain-based framework made curious opponents not infer any sensitivity information of the output without collusion. They used convolutional neural network (CNN) to realize the model [9]. The model security aggregation method based on encryption and noise can reduce collusion attacks made by malicious users or malicious servers and user entities [10]. However, the above studies didn't consider the communication and computation cost, as well as the practicality and continuous security of the model.

TABLE I
CHARACTERISTICS OF ENHANCED PRIVACY-PRESERVING TECHNOLOGY.

Ref.	Encryption	Noise	Insider attacks	Outsider attacks
[5][7]	-	Y	Y	-
[6]	-	Y	-	Y
[8]	Y	-	-	-
[9]	Y	-	-	Y
[10]	Y	Y	Y	Y
[11]	Y	Y	Y	-

IV. Classifications of Federated Learning Algorithms

In the FL system, in order to reduce the network latency and enhance privacy protection, most scholars currently conducted researches based on ML and reinforcement learning (RL), namely federated machine learning (FML) algorithm and federated reinforcement learning (FRL) algorithm. Fig. 4. shows the process of the algorithm. And Table II lists the specific algorithms used in related current researches.

A. Federated Machine Learning Algorithm

Most of the current researches combine FL and ML into FML, which solved some related privacy security issues. Recently, the most commonly used basic ML algorithms are ensemble learning (EL) and deep learning (DL). Rui Hu *et al.* used a differential stochastic gradient descent (SGD) method to collaboratively train multiple personalized ML models on heterogeneous smart devices, found an appropriate noise threshold, and achieved a less privacy loss [5]. Y.Lu *et al.* used the gradient boosting decision tree (GBDT) model for local training [7]. This model requires less training data and can achieve high accuracy in short time. It can achieve balance between efficiency, accuracy and privacy enhancement. Bo Yin *et al.* proposed a secure data collaboration framework

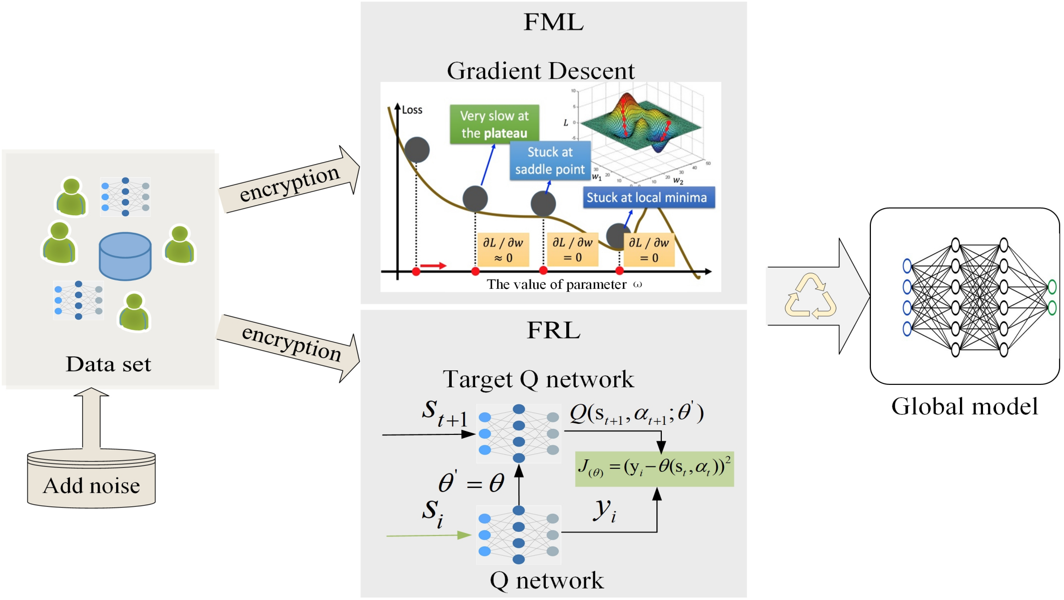


Fig. 4. The process of federated learning algorithms.

based on FDL algorithms, which can realize multi-party data calculation without transferring data from a private data center [8]. The algorithms can improve the privacy security of user data and decrease risk of user sensitivity information leakage. Chunyi Zhou *et al.* used SGD-based back propagation (SGD-BP) algorithms to train models to prevent data attacks, model attacks, and collusion attacks [10]. These models are trained in smart home and smart medical scenarios to prevent data . Although the efficiency of the model is reduced, the final global model can achieve a better privacy-preserving and higher model accuracy. At present, many studies are dedicated to enhance the privacy-preserving and model security, but in the future they may pay more and more attention to the efficiency of model training and the practicality of models.

B. Federated Reinforcement Learning Algorithm

With a large number of studies using reinforcement learning to solve numerous practical problems, the combination of RL and FL algorithms have emerged as a prominent solution for privacy check. It can realize the quick convergence and more accurate training of the learning model. Meng Hao *et al.* added augmented learning with error terms (A-LWE) to their encryption model [11]. The trained model not only can prevent privacy leakage and collusion between multiple entities, but has a good advantage on efficiency and accuracy. In addition, the author of [13] considered the adaptability of the dynamic network environment, using deep Q-learning (DQL) and Markov decision to solve the integer linear programming problem. They designed a weighted FDL model to solve the problem of model aggregation between heterogeneous user devices. Xiaofei Wang *et al.* considered the excessive consumption of network resources and complex dynamic control [14]. They proposed a collaborative edge cache framework for federated deep reinforcement learning (FDRL) and used double deep Q-learning (DDQL) to solve the problem of discontinuous data sampling in large spaces. In addition, the proposed algorithms may fail to face attacks from external or internal attackers. Therefore, exploring a new FDL mechanism is an important issue to be solved in the future.

TABLE II
CLASSIFICATIONS OF FEDERATED LEARNING ALGORITHMS.

	Type	Specific algorithms	Features
FML	EL	GBDT [7]	Less training data, high accuracy, balanced between efficiency, accuracy and privacy enhancement.
	DL	CNN [9]	Improve data privacy.
		SGD-BP [10]	Model efficiency may low, but privacy and model accuracy is high.
		A-LWE [11]	Decreased privacy leakage and collusion of multiple entities, balanced efficiency and accuracy.
FRL	RL	DQL [13]	Enhanced privacy-preserving and solved the problem of model aggregation between heterogeneous devices.
		DDQL [14]	Reduced information leakage and network resource consumption.

V. CURRENT APPLICATIONS

FL has become a popular research topic in AI, and attracted great attention in smart medical care and smart city construction. This section summarizes the current applications of FL in various fields.

A. Google Gboard System

At first, Google designed an input prediction model that could collaboratively train a global model without uploading user data, and protect the security and privacy of each users' data during the training process. The FL model is based on the federated average algorithm (FedAvg), which can realize the next word prediction, emoji prediction, keyboard search suggestions, and learning outside the vocabulary.

B. Smart Medical

In recent years, existing AI programs are used in heart disease and radiology. They can help in diagnosing heart disease and identify early cancer cells. To deal with more user data and to ensure model accuracy and protect data privacy, FL brings development to this field. For example, the NVIDIA CLARA framework used by NVIDIA in the field of medical imaging includes 13 state-of-the-art classification and segmentation of AI. Recently, Tencent Tianyan Laboratory and WeBank deeply integrated FL and medical care, and built a big data concentration and mining platform for the diagnosis of stroke diseases. The accuracy of the trained model can be achieved as eighty percent.

C. Smart Financial

At present, FL is in a pilot state in financial field, involving applications in bank credit risk control and anti-money laundering. Through FL, multi-dimensional feature data such as customer transaction data and tax information can be combined to establish a new data cooperation model to help financial institutions in filtering out credit blacklist customers. Thereby, it can reduce credit review cost. For example, Bank of Jiangsu uses encryption algorithms to ensure data security, and integrates the characteristic variables of Tencent's security black and gray production library with the characteristic variables of the Bank of Jiangsu credit card. It protects user privacy and adapts to market changes more quickly.

D. Smart Transportation

With the increase in transportation, traffic congestion and and control of traffic lights have become an urgent problem which need to be addressed in a timely manner. At present, some people use a gated recurrent unit neural network traffic prediction algorithm based on FL [15] to predict traffic flow, which not only protects user privacy but also makes accurate traffic flow predictions. In addition to neural networks, the graph convolutional networks are also applied to FL to train more accurate prediction model.

E. Smart Educational System

FL can provide solutions for privacy-preserving issues in the process of educational data mining. At present, this field uses the support vector machine (SVM) algorithm of FL to evaluate teaching quality, uses deep neural network (DNN) to realize the recommendation of learning resources, and analyzes student performance with the help of K-means clustering algorithm. In addition, FL can also help students to develop a learning plan without the sharing of data, which can be suitable for individuals based on their learning abilities, specialties, hobbies and so on. In the process of advancing smart education, the use of FL is a better choice, which can give full play to the value of educational data on the basis of solving the problem of personal privacy-preserving.

VI. Future Challenges

From the aforementioned discussion, although FL has the great advantages of protecting privacy and multi-party local data security, it still faces numerous challenges of interpretability and quick convergence, etc. Based on the overview of this article, this section addresses the following challenges, as shown in Table III.

A. Model Accuracy

So far, basic FL may fail to avoid attacks from internal or external attackers on the original data or trained model parameters. Existing studies used encryption, noise addition, and and the combination of both these approaches to enhance the accuracy of the model. Too much noise or too little noise will reduce the accuracy of the model. Therefore, the data should be properly noised for applications in different environments, and the accuracy of the model should be enhanced without affecting the protection of data privacy in the future.

B. Personalization and Practicality

The model based on FL was only a shared global model that contains all the content of all the participants, so it ignored the personal characteristics of participants. So far, a novel differential privacy FL scheme has been developed [5]. A personalized model that meets the needs of users personalized services is trained to ensure the users' privacy, but the practicality of the model, such as communication costs and resource allocation are ignored. And in actual network, there might be a lot of incorrectly marked or unmarked data. Therefore, the personalization and practicality of the models are big issues which need to be considered.

C. Prevent Malicious Attacks

In the process of model training, although some methods can enhance the protection of user data, it is still easy to deduct the users' private data from the gradient or model parameters. So far, few studies are based on preventing these malicious attacks. Simple defense may fail to reconstruct after the model is damaged. Therefore, we should focus on more malicious detection mechanisms to prevent internal or external malicious attacks from the sources and reduce the loss after the model is damaged.

D. Sixth Generation Mobile Network

The sixth generation (6G) networks have more data forms and penetrate into various industries in the Industrial Internet of Things (IIoT), which means that a large amount of data in the network contain extremely sensitive personal information. For example, smart homes, smart cars, virtual butlers and so on. These smart services need to collect more people private data. In addition, the future of autonomous driving will have higher requirements for network security. Therefore, in the future, 6G data security and privacy-preserving will rise to new heights. In view of the new features and application scenarios of 6G, different methods such as blockchain can be adopted to further improve the protection of user privacy.

TABLE III
FUTURE CHALLENGES.

Items	Specific description	Possible solution
Model Accuracy	too much noise or too little will decrease the accuracy of the model.	explore different noise thresholds in different application scenarios.
Personalization and practicality	personal features of participants,communication costs and resource allocation are ignored.	a novel differential privacy FL scheme [5].
Prevent malicious attacks	just defense,not prevent attacks from the source.	focus on more malicious detection mechanism and how to recover after model is damaged.
The 6G	contain extremely sensitive information.	different methods such as blockchain can be adopted.
Federated learning + Edge computing	less bandwidth resource, and expensive communication costs.	reasonable choice of weights and participants.

E. Federated Learning + Edge Computing

In the future, the combination of FL and edge computing will be a good development direction. Although it has been studied by some scholars, it is still in immature. Moreover, the weighted update method, participant selection, and bandwidth allocation are the hot topics which need to considered in future.

1) the weights update methods: For each user, assume $v = E[Xw - w]$, where w is the current local weights and Xw is the local weights averaged over a certain historical iteration rounds. The w that minimize the expected value v will be uploaded to the server. After receiving w, the server performs weighted average find the median to get $w*$, and each user downloads the updates $w*$ for the next round of model training. Repeat the above operations, and until the required model accuracy is reached. Although this can reduce the communication cost, how to find the ideal weight to be achieved locally as well as balance the communication cost and the local training time is still a challenge.

2) participants selection: At present, most scholars use heuristic algorithms and stochastic matching, which may have the problem of local optima. Therefore, a small number of scholars have adopted deep RL to achieve better results between model convergence and accuracy. Besides, the allocation of bandwidth resources have a significant effect on communication time. All in all, it will face more challenges in reducing communication costs.

VII. Conclusion

This paper reviewes the techniques for enhancing the privacy-preserving and security in FL, and listed the applications of FL in various scenarios. We also summarized some challenges. According to this survey, a fully decentralized FL model model can be considered as an ideal state of FL. However, the current development is concerned, there are still many obstacles to achieve this ideal state. Chain-FL model would be a good direction in 6G. We believe, with the development of fifth generation mobile network (5G) and beyond 5G, the privacy-preserving enhancement technologies are able to take a step forward and FL can play a greater role in the future.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant No. 62071306, and in part by Shenzhen Science and Technology Program under Grants JCYJ20200109113601723 and JSG-G20210420091805014.

References

[1] B. Ai, A. F. Molisch, M. Rupp and Z. D. Zhong, “5G Key Technologies for Smart Railways,” *Proceedings of the IEEE*, vol. 108, no. 6, pp. 856-893, June 2020.

[2] B. D. Deebak and F. Al-Turjman, “Smart Mutual Authentication Protocol for Cloud Based Medical Healthcare Systems Using Internet of Medical Things,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 2, pp. 346-360, Feb. 2021.

[3] KONENY. J, MCMAHAN. H. B, RAMAGED, *et al.*, Federated optimization: distributed machine learning for on-device intelligence[J]. *arXiv preprint*, 2016, *arXiv:161002527.*

[4] YANG QIANG, *et al.*, “Federated Learning” [M]. Beijing: Publishing House of Electronics Idustry, 2020.

[5] R. Hu, Y. Guo, H. Li, Q. Pei and Y. Gong, “Personalized Federated Learning With Differential Privacy,” *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9530-9539, Oct. 2020.

[6] K. Wei *et al.*, “Federated Learning With Differential Privacy: Algorithms and Performance Analysis,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454-3469, 2020.

[7] Y. Lu, X. Huang, Y. Dai, S. Maharjan and Y. Zhang, “Differentially Private Asynchronous Federated Learning for Mobile Edge Computing in Urban Informatics,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2134-2143, March 2020.

[8] B. Yin, H. Yin, Y. Wu and Z. Jiang, “FDC: A Secure Federated Deep Learning Mechanism for Data Collaborations in the Internet of Things,” *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6348-6359, July 2020.

[9] Y. Li, Y. Zhou, A. Jolfaei, D. Yu, G. Xu and X. Zheng, “Privacy-Preserving Federated Learning Framework Based on Chained Secure Multiparty Computing,” *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6178-6186, April 2021.

[10] C. Zhou, A. Fu, S. Yu, W. Yang, H. Wang and Y. Zhang, “Privacy-Preserving Federated Learning in Fog Computing,” *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10782-10793, Nov. 2020.

[11] M. Hao, H. Li, X. Luo, G. Xu, H. Yang and S. Liu, “Efficient and Privacy-Enhanced Federated Learning for Industrial Artificial Intelligence,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6532-6542, Oct. 2020.

[12] W. Y. B. Lim *et al.*, “Hierarchical Incentive Mechanism Design for Federated Machine Learning in Mobile Networks,” *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9575-9588, Oct. 2020.

[13] X. Wang, R. Li, C. Wang, X. Li, T. Taleb and V. C. M. Leung, “Attention-Weighted Federated Deep Reinforcement Learning for Device-to-Device Assisted Heterogeneous Collaborative Edge Caching,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 154-169, Jan. 2021.

[14] X. Wang, C. Wang, X. Li, V. C. M. Leung and T. Taleb, “Federated Deep Reinforcement Learning for Internet of Things With Decentralized Cooperative Edge Caching,” *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9441-9455, Oct. 2020.

[15] Y. Liu, J. J. Q. Yu, J. Kang, D. Niyato and S. Zhang, “Privacy-Preserving Traffic Flow Prediction: A Federated Learning Approach,” *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7751-7763, Aug. 2020.