2022 19th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) | 978-1-6654-9389-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCWAMTIP56608.2022.10016548

PRIVACY-PRESERVING AND MODELS INTRUSION DETECTION FEDERATED DEEP LEARNING CHALLENGES, SCHEMAS AND FUTURE TRAJECTORIES

YANG YU[1,2], LI JIANPING[1,*], DUAN WEIWEI[1]

[1]School of Computer Science and Engineering (School of Cybersecurity), University of Electronic Science and Technology of China, Chengdu 611731, China

[2]School of Information Technology Industry, Yunnan Vocational Institute of Energy Technology, Qujing 655001, China

E-MAIL: yangyu2022@std.uestc.edu.cn, jpli2222@uestc.edu.cn,dww@std.uestc.edu.cn

Abstract:

Deep learning has made remarkable research advancements and wide-ranging applications in the domains of computer vision, multimodal, natural language processing, additionally, other areas. This has caused the academic community to pay increasingly close attention to the attack and defense technology in its training and testing phases, among which the federal deep learning has produced positive results. Federated deep learning models are prone to memorizing private and sensitive terminal participants' data, model parameters, when combined with the model's inherent vulnerability, they will result in privacy leakage, poisoning attack, model inference attack, adversarial attack. We briefly discuss the conception of federated deep learning as well as security challenges and open questions in this paper. In order to facilitate the understanding of these challenges and problems, we further propose a security system model. We also provide an overview and deduce the attack and mitigation approaches to the most sophisticated privacy-preserving and intrusion detection models. in the last two years. To tackle these challenges and enlighten further encryption techniques researches, finally, we discuss and describe current prospects and future trajectories of federated deep learning.

Keywords:

Federated deep learning; Models intrusion detection; Inference attack; Poisoning attack; Adversarial attack; Encryption techniques; Privacy-preserving

1. Introduction

With the era of Artificial Intelligence(AI) and the rapid ascending of deep neural network learning, federated deep learning catches researchers' attention on the privacy-preserving and models intrusion detection issues. Federated deep learning(FDL) has been research hotspots due to its ability to update local participants' model parameters without converging terminal users' raw training data (For model training, conventional machine learning necessitates the acquisition of raw data, which has significant privacy exposure issues). However, since attackers and adversaries can deduct and generate participants' privacy from the shared gradients, FDL is still exposed to a large number of privacy threats and various models intrusion challenges.

A subcategory of Methods termed Deep Learning (DL) is based primarily on multi-layer artificial neural networks. [1]. Federated Learning (FL) is a decentralized distributed data partitioning strategy which ensures user privacy and trains AI models without disclosing the local datasets of terminal participants. Recent assaults, though, show that simply protecting data location privacy during training procedures is insufficient to provide appropriate security. To handle the above problems, FDL uses multifarious Deep Neural Networks (DNN) patterns to dispose privacy leakage and security challenges, is proposed that merging privacy-preserving FL with DL. At present, the core threats to DNN are proneness to leakage of model parameters and the vulnerability of model properties in model training process and model testing process. Taking these into consideration, the model parameters mainly manifestations are vulnerability against gradient leakage attacks, feature extraction, feature fusion, pseudo-labeling and system-based safety detecting, the model properties crucially demonstrations are model sensitivity, generalization and non-linear operations(BatchNorm, MaxPool, ReLU) errors, feature embedding and data contamination.

The following list summarizes our significant and substantial contributions to this paper:

- The paper systematically illuminates security challenges and attack threats are relevant to FDL, i.e.,

978-1-6654-9389-5/22/$31.00 ©2022 IEEE

model poisoning and model inference attacks. It also summarizes the process of different attacks and the corresponding defense techniques.

• Illustrating different variants of defend methods of which comprises Differential Privacy(DP) and masking privacy-preserving techniques, Homomorphic Encryption(HE), Secure Multiparty Computation(SMC) and providing a classification and overview of the prior studied. The paper concisely analyzes the advantages and disadvantages of the most promising defend techniques to resolve or alleviate the above-mentioned attacks.

• Describing a FDL security system model. Our work convinces that FDL security system model is able to solve and alleviate the issues and attacks of the model parameters and model properties.

• Discussing possible prospects and future trends in providing insights into existing privacy-preserving techniques and intrusion detection models in FDL.

The remainder of our work, Section 2 presents the schemas, security challenges and open problems of federated deep learning and elaborate the proposed security system model under considering already exist some preliminary works. Section 3 presents the most promising privacy-preserving techniques and intrusion detection models. Then it describes HE, SMC, DP and masking methods, training data poisoning attacks and defense method, membership inference attacks and defense method, model inference attacks and defense method in detail. Following that, Section 4 expounds the prospects and future trajectories in FDL. Finally, Section 5 reviews and summarizes our paper and draws a conclusion.

2. Federated deep learning schemas and core security challenges

In this section, we first outline the latest schemas of FDL. We proposed a FDL security system model to account for threats and challenges in privacy-preserving and intrusion detection.

Based on our previous investigations and related research work, we summarize the challenges faced by FDL as follows:

• Security challenge 1: Managing the trade-offs between HE and communicating and storage overhead.

• Security challenge 2: Regulating the trade-offs between SMC and the accuracy and privacy risks model.

• Security challenge 3: Balancing the trade-offs the DP and masking methods and the utility of modes.

• Security challenge 4: Controlling the trade-offs between the model poisoning and inference attacks and the effectiveness and detectability of modes.

2.1. Federated deep learning schemas

FDL enable large-scale DNN model secure testing and training and deployment to have been possible. Instead of sending data directly, FDL uses participating terminal users data to do local model training, distribute partial gradients, and update and aggregate model task parameters without disclosing private and sensitive information, hence, it preserves privacy and protects against attacks throughout the FDL testing and training phase. According to the categorization Federated Learning and different parties' data from feature space and independent distribution space, FDL can be categorized into horizontal, vertical and transfer federated deep learning.

There already exist some preliminary FDL schema studies. Sun et al. [2] develops a distributed quantized gradient approach for communication reduction from the terminal users to the server to update the local model parameter. Feng et al. [3] proposed a semi-supervised federated heterogeneous transfer learning framework reducing model overfitting by overlapping data insufficiency. To solve the issue of communication overhead against quantum attacks, a lattice-based multi-use secret sharing method was created [4].

On all accounts, the distributed quantized gradient approach has an advantage over merely strongly convex and nonconvex learning problems in FDL. However, more robust schemas are needed for communicating and storage overhead. The further studies should pay more attention to manage the trade-offs between HE and collaboratively communicating and storage overhead.

2.2. Threats and attacks in privacy-preserving and model intrusion detection

Apart from outside the vulnerability of Federated Learning model properties, to solve threats and attacks issues, most scholars are more currently concerned about privacy-preserving and model intrusion defense. Considering the foundation of foremost research work, we propose a FDL security system model. As according Fig. 1, the security system model shown from the overall of terminal participants privately train the local models are uploaded to cloud server securely aggregates the local model parameters and distributes the aggregated global model parameters to the selected terminal participants.

The security system model is divided into the following six processing steps.

①The terminal participants are chosen on account of their contribution rate to the cloud server separately train local model.

②After the training, the terminal participants upload their local gradients by cryptographic algorithms(for instance, HE) or DP techniques. Fig. 1 exploits HE algorithm technology as a schematic.

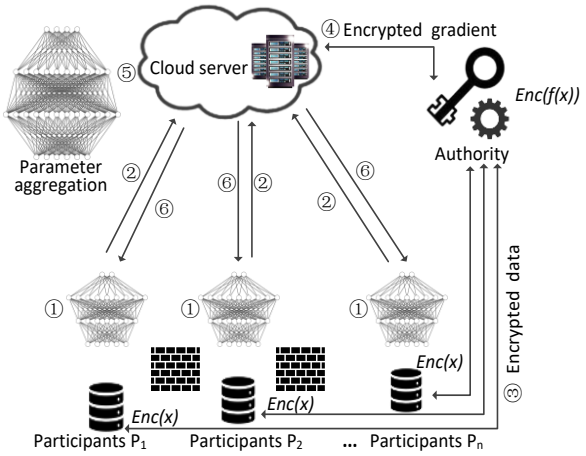

Fig.1 Federal deep learning security system model

③ The terminal participants independently encrypt respectively private gradient rather than data and send to the trusted third parties.

④The trusted third parties unequivocal compute the encrypted gradients. Additionally, the trusted third parties preliminarily and simply verify whether exists gradient leakage and model intrusion detection.

⑤The cloud server aggregate local gradients. Based on the composite parity data, the FDL cloud server assists in computing a portion of the partial gradients and combines those computations with partial gradients obtained from FDL participants.

⑥Ultimately, the cloud server distribute the new cloud server parameters. Launch the subsequent local model train cycle until the global model convergence.

The proposed security system model mainly balancing the trade-offs between model accuracy and privacy risks. Make sure that all the terminal participants in a secure training and testing environment.

3. The state-of-the-art privacy-preserving techniques and intrusion detection models

In section 3, the paper elaborates the most upbeat works related to privacy-preserving techniques and intrusion detection models under the FDL framework in recent years.

3.1. Privacy-preserving techniques of federated deep learning

There already exist some excessively matured and preliminary FDL privacy-preserving defense method studies. Ng et al. [5] built a two-level incentive system to effectively distribute the resources of data terminal users and federated learning workers in order to complete the coded federated learning training assignments，which uses an auction system based on deep learning to determine allocation. A comprehensive framework for adversarial training and evaluation that incorporates two new optimization techniques for optimizing the trade-off between target task performance and the associated privacy budgets is proposed [6]. Idan et al. [7] introduced an attribute-based encryption method with unaware attribute translation for intraorganizational data exchange. Kang et al. [8] designed a federated adversarial domain adaption framework with privacy protection that allows terminal users to collaboratively conduct domain adaptation without exposing private data.

To more reliable tackle these privacy-preserving open problems， the two-level incentive mechanism merely address linear regression tasks, which should be considered the non-linears operations, such as BatchNorm, MaxPool, ReLU. The adversarial training framework needs to define privacy leakage risk, which involves some non-technical uncertainties. The interorganizational data-sharing reckons the semi-trusted proxies instand of the malicious.

Besides, we also provide a summary of the federated deep learning model testing phases in Table 1.

Table 1 The summary of federated deep learning model testing phases

	Model testing process		
Attack methods	**Model extraction attacks**	**Model intrusion attacks**	**Adversarial attacks**
Defense methods	Homomorphic Encryption、Secure Multiparty Computation、Gaussians Mixture Model	Distributed K-means Clustering、Distributed Random Forest、Traffic Anomaly Detection	Intrusion detection、Digital Tokens and Local Credibility
Deep learning models	Neural Network	Neural Network、LSTM	Logistic Regression Model、CNN、LSTM、GAN
The references	[10]、[25]	[36]、[45]	[14]、[15]、[17]、[20]

3.1.1 Homomorphic encryption techniques

Deep learning with privacy protection HE is a cutting-edge and dynamic research field that strives to develop deep learning systems that work while protecting the privacy of terminal participants. By facilitating computing over encrypted data, HE was developed as a revolutionary cryptographic solution which guarantees participants' privacy at the cloud side. Several related research are presented: Hariss et al. [9] designed a homomorphic hybrid symmetric encryption scheme for resisting the known plain-text/cipher-text attacks. Zhou et al. [10] invented a homomorphic encryption-based system and enhanced it with an oblivious transmission mechanism. While maintaining the privacy of private data, HE algorithms develop innovative ways to support manipulations on encrypted data.

The addition and multiplication encrypted operations require to increase the excessive computation communication overhead and computation power without leaking the information of random matrix for HE. Besides, although HE approaches don't leak information during training, they are unable to fend off membership inference attacks and lack message authentication to provide source authentication and data integrity protection.

3.1.2 Secure multiparty computation

To overcome these issues HE faced and investigate ways to make the augmented ciphertexts for the fully homomorphic encryption minimal. There already exist some preliminary SMC studies.

For secure multiparty computation scenarios, a modified completely homomorphic encryption approach was put out in the plain model [11]. Tan et al. [12] leveraged GPUs to speed up deep learning that protects privacy, and they demonstrated that safe multiparty computation can be done fully on the GPU. The secure multiparty computation is more effective relieve HE symmetric encryption scheme which lacks message authentication deficiency, whereas it is vulnerable to inference and adversarial sample attack.

3.1.3 Differential privacy and masking

The differential privacy is comparatively mature and widely used noises perturbation method that prevents privacy leakages from gradients in FDL.

To overcome these issues, Zhang et al. [13] protected client-level differential privacy for federated learning algorithms in the neural network. Liang et al. [14] introduced the laplacian smoothing scheme to the differentially private federated learning that exploited differential privacy statistical protection precision without losing privacy budget for reducing variance. To address fairness problem and mitigate privacy leakage, To ensure privacy in generative adversarial networks, a differentially private decentralized deep learning system with differential privacy was designed [15]. Wu et al. [16] stochastic gradient descent was utilized to reduce the fitness cost of the machine learning model utilizing noisy, differentially-private gradients. Chen et al. [17] introduced the federated learning and analytics Poisson binomial mechanism for federated learning and analytics to decrease with the privacy budget. Liu et al.[18] developed a two-stage system using local differential privacy and federated stochastic gradient descent. Currently, the primary theoretical underpinnings of DP mathematics are as follows:

• (ϵ, δ) -differentially private on a randomized mechanism $\mathcal{M}$ algorithm[14].

$$\mathrm{P}(\mathcal{M}(\mathcal{D}) \in \mathcal{S}) \leq \mathrm{e}\text{^}\epsilon\, \mathrm{P}(\mathcal{M}(\mathcal{D}') \in \mathcal{S}) + \delta \tag{1}$$

• (α, ρ) -Rényi differential privacy on a randomized mechanism $\mathcal{M}$ algorithm[15].

$$D_{\alpha}\left(\mathcal{M}(D) \| \mathcal{M}(D')\right) := \frac{1}{\alpha - 1} \log \mathbb{E}\left(\mathcal{M}(D) / \mathcal{M}(D')\right)^{\alpha} \leq \rho \tag{2}$$

• ϵ_{ℓ}-differentially private[18].

$$\begin{aligned} &\mathbb{P}\left\{\left(\bar{\mathfrak{Q}}_{\ell}(\mathcal{D}_{\ell}; k)\right)_{k=1}^{T} \in \mathcal{Y}\right\} \\ &\leq \exp\left(\epsilon_{\ell}\right) \mathbb{P}\left\{\left(\bar{\mathfrak{Q}}_{\ell}(\mathcal{D}'_{\ell}; k)\right)_{k=1}^{T} \in \mathcal{Y}\right\} \end{aligned} \tag{3}$$

• $(\epsilon_1 + \epsilon_2)$-local differential privacy[20].

$$\Pr\left[\mathcal{M}_1(z) = j\right] \leq e^{\epsilon_1} \cdot \Pr\left[\mathcal{M}_1(z') = j\right] \tag{4}$$

$$\Pr\left[\mathcal{M}_2(v_j) = v_j^*\right] \leq e^{\epsilon_2} \cdot \Pr\left[\mathcal{M}_2(v_j') = v_j^*\right. \tag{5}$$

• (ϵ', δ) differentially private [22].

$$\epsilon' = \frac{c_1 q \sqrt{T \log\left(1/\delta\right)}}{\sigma} = c_2 q \sqrt{T} \epsilon \tag{6}$$

The differential privacy guarantees the terminal participants' privacy in the repeated iteratively collection of model training parameters within the paradigm of DNN.

3.2. Models intrusion detection in federated deep learning

In this section, we focus on two security taxonomies that are poisoning attacks and inference attacks in models intrusion detection. In this paper, poisoning attacks and inference attacks can be categorized into three categories: training data poisoning attacks, membership inference attacks and model inference attacks. To quite large extent, model intrusion will cause terminal participants' data information to be inferred, damage the accuracy of the model, lead to the final classification error of the deep learning model, and even the failure of model training. Therefore, model intrusion resistance is a research hotspot technology in FDL. As a result, model intrusion detection is one of the methods to resist.

Besides, we also provide a summary of the federated deep learning model training phases in Table 2.

Table 2 The summary of federated deep learning model training phases

Model training process			
Attack methods	**Training data poisoning attacks**	**Membership inference attacks**	**Model inference attacks**
Defense methods	Differential Privacy 、Noise-adding Mechanism、Double-masking Protocol	Gradient Descent Method、	Hybrid Preserving、
Deep learning models	GRU、Deep Neural Networks	CNN 、Neural Network 、Deep Learning	Decision Trees、CNN、Linear Support Vector Machines
The references	[4]、[8]、[29]	[16]、[23]、[30]	[24]、[35]

3.2.1 Training data poisoning attacks and defense method

Because of cross-device and cross-silo statistical heterogeneity in federated learning, timely and accurately detecting models intrusion is quite important in terms of preventing privacy leakage and ensuring information.

The training data poisoning attacks arise training data collection duration in FDL. There already exist some basic investigations of training data poisoning attacks studies. Hahn et al. [19] brought out a personalized federated learning method that is well-calibrated to potential label noise under diverse non-independent identically distributed scenarios. Zhou et al. [20] proposed vertically federated graph neural network to solve data isolation problem. The private horizontal federated learning, vertical federated learning and transfer federated learning for the non-independent isodistribution data has been proposed. When used with adaptive gradient descent method for model parameter updating, it can set different privacy budgets depending on the unbalanced data of various terminal users and experimentally achieve data privacy-preserving with higher accuracy [21]. Xiong et al. [22] proposed a novel algorithm (2 differential privacy federated learning) for independent identically distributed scenario data privacy leakage which defends privacy inference attack by including noise when distributing the global model and training the local models. The method strikes a balance between the trade-offs of convergence, privacy-preserving, and server global parameters correctness and integrity. Hahn et al. [23] brought out a approximate bayesian computation-based gaussians mixture model for solving class imbalance problems and avoiding privacy leakage. Jiang et al. [24] proposed an anonymous authentication approach and signature scheme with detection function for data poisoning attacks. Xue et al. [25], two backdoor attack techniques have been proposed, in which some well-designed backdoor instances are injected into the deep neural network model's training set without knowledge of the parameters or architectures. These attacks present new dangers to deep learning models and new difficulties for existing defenses.

Assuming the attacker has controlled of some terminal participators' devices and has manipulated the local training data model parameters, training data poisoning attacks is more susceptible. To solve the above problems, how to effective against those back-door attacks and detect the internal attacks. Considering the terminal participants' non-IID data, due to the potential for local models to overfit local data, the global cloud server model suffers. Designing a hierarchical learning system using the edge computing paradigm to ensure that gradient descent on the user-edge layer occurs without influencing the hyperparameter aggregation layer [26].

3.2.2 Membership inference attacks and defense method

Member inference attack refers to the attack that affects the accuracy of the model by influencing the data distribution or manipulating the data label of the regular terminal participants.

Qi et al. [27] suggested a consortia blockchain-based federated learning framework that employs a differential privacy technique with a masking data(noise-adding) mechanism and the numerical findings that can prevent

membership inference attacks and inaccurate model updates from dispersed vehicles. Yang et al. [28] proposed a model perturbation method in which random numbers have been applied against membership inference attacks and launching reconstruction. An unattacked and unpoisoned terminal participants selection approach that can assess the data quality of terminal users and protect them from attacks by nefarious federal members has been presented [29]. With the use of obtained high-fidelity data extraction in both cross-device and cross-silo federated deep learning threat models, an attacker can use current methods to fish for different data points in arbitrary big batches without changing the model [30].

To sum up, the paper convices that the communication overhead for federated learning the performance of the GRU model need to improve. In this attack, the follow-up research work need focus more on controlling the trade-offs between the model poisoning and inference attacks and the effectiveness and detectability of modes, such as, Shamir's scheme to share symmetric key, hybrid secure member selection solution.

3.2.3 Model inference attacks and defense method

The local model inference attack occurs in model training process in federated deep learning.

To deal with these threats, Yehezkel et al. [31] presented auto-encoder losses transfer learning to provide a nonparametric invariant model for network anomaly detection by reconstruction loss. Truex et al. [32] presented a comprehensive scheme in allusion to privacy protection for protecting against inference threats which uses secure multiparty computation in combination with differential privacy to produce models, which can be used to train a variety of machine learning models, such as neural network model. Jia et al. [33] designed an application model and a data protection aggregation scheme of blockchain-enabled federated learning which employs for model extraction attack and model reverse attack in federated learning. Ryffel et al. [34] brought a new addition for conventional neural networks that supports n-party private federated learning for private inference between distant servers. Ma et al. [35] proposed a two-trapdoor homomorphic encryption strategy and the Byzantine-tolerance mechanism aggregation against encrypted model poisoning attacks. Zhang et al. [36] exploited one-line modification to existing to progressively continue the durability from test process to deployment phase for achieving adversarial backdoor attacks. Generative Adversarial Networks (GAN), where the sample adversary can convey the leaked information under the model extraction attack, were the target of investigated model extraction assaults [37]. Aydın et al. [38] designed a distributed denial of service system based on long short-term memory (LSTM)， which employs signature-based attack detection approach for the intrusion attacks.

Some academics study on local model poisoning attacks to Byzantine-robust federated learning[39]. Xu et al. [40] presented a federated learning system that is verifiable and protects privacy which focuses on the problems of local gradients leakage and server colludes with multiple terminal users during training process. Wang et al. [41] propose a graph-based behavioral modeling paradigm for behavioral anomaly detection problem. For the purpose of preventing nonanthropogenic anomalous data interference, a traffic data anomaly detection system on account of the self-coding of long- and short-term memory networks has been developed [42]. Qayyum et al. [43] presented a hybrid learning-based method that detects malicious neural network parameter updates against for demolishing training model aggregation process.

Some researchers argue that it is important to focus research on defenses in FDL onto impactful attacks, because the simplicity of method means it is feasible for attackers to have discovered and deployed this attack already. What's more, the learning-based methods are powerless to withstand targeted and untargeted adversarial machine learning attacks. The secure exchange and sharing of the application model exists some positive impacts. However, numerous works are limited to untargeted poisoning attacks, random poisoning attacks are defenseless. To solve model extraction attacks, high-quality samples are obtained by subsampling generated samples and excluding some subpar ones based on some prior knowledge.

4. Prospects and future trajectories

4.1. Prospects

The latest findings from FDL have been transferred and put into use to all kinds of niche application sectors, where they have produced positive security and privacy protection outcomes as well as notable model intrusion detection defense effects.

The collaborative adversarial sample attack and security and privacy evaluation was used for privacy protection video rectangular frame retrieval. Liu et al. [44] reported a deep learning engineering framework called FedVision helps to create vision-based safety monitoring systems for smart cities. Without having to employ centrally maintained, huge training datasets from terminal users, object identification models. Hamza et al. [45]

provided some theoretical and practical implications for utilizing machine learning privacy protection approaches in Big Data analytics making use of variant homomorphic encryption algorithms and incorporating multi-party security encryption technologies into Big Data applications. Liu et al. [46] offered a synthetic data generation model with user privacy controls to offer privacy guarantees that correspond to the user's wishes in recommendation systems. Chen et al. [47] introduced scPrivacy, which allows single-cell types to be identified automatically, as a prototype to help with single cell annotations in a way that protects data privacy for the single-cell transcriptomics area.. Subramaniyaswamy et al. [48] proposed somewhat homomorphic encryption schemes for deploying healthcare application in the edge environment to guarantee privacy in transmitting sensitive patient data. Li et al. [49] brought out a federated deep generative learning framework for decision-making in power systems in generative adversarial networks. Some decentralized federated architecture that enables the aggregation of models for clinical decision support systems under Internet-of-Things-enabled E-health systems. Zhou et al. [50] proposed a method of encoding for deception attacks in multi-sensor networks that incorporates linear transformation and synthetic noise. Deep Adversarial Neural networks tasks for machine defect diagnostics using data privacy-preserving federated transfer learning.

To utilize these defense techniques, entrepreneurs and academia ought to balance the trade-offs the privacy-preserving and models intrusion detection methods and the utility of modes. With the intelligent innovation of information, privacy-preserving and security defense require synchronous planning, synchronous construction and synchronous deployment with information system. At the same time, the application prospect of FDL security field is necessarily extended to and deep learning fields and traditional applications of which exists vulnerable security, such as distributed communication security of in-memory database, adversarial sample training process, intrusion detection and prevention of critical infrastructure application system model, etc.

4.2. Future trajectories

On account of the challenges we summarized in the section 2, we outline privacy-preserving and models intrusion detection trajectories that would inspire further research in FDL. To overcome the trade-offs between HE and communicating and storage overhead, further research trajectories on HE can focus on reducing the interaction times between the terminal participants and the cloud server, optimizing encryption and decryption algorithms, improving encryption efficiency, and reducing storage overhead. To dispose the trade-offs between SMC and the accuracy and privacy risks model, optimizing the full homomorphic encryption algorithm to improve the efficiency of encryption, moreover, combining with security hardware to accelerate the efficiency of computing. To balance the trade-offs the DP and masking methods and the utility of modes, more effective and robust model fusion methods need to further study. To control the trade-offs between the model poisoning and inference attacks and the effectiveness and detectability of modes, training data poisoning defense method, membership inference defense method, model inference defense method, adversarial samples attacks defense method, to alleviate these problems, it is necessary to further study the interpretability of the model training process and model testing process of the DNN learning model, and integrate various technologies from the source of data to the application of the model for systematic, comprehensive and multi-directional intrusion detection and privacy protection, settling down to further enhance the model robustness, effectiveness, integrity and availability of the FDL model.

5 Conclusions

FDL has been increasingly used in a multitude of DNN application areas, for example, privacy-preserving video action recognition, Big Data analytics and applications, privacy-preserving recommendation systems, secure single-cell type identification prototype, private healthcare application, sensitive decision-making and clinical decision support systems, extensive multi-sensor networks and deep adversarial neural networks, and etc., etc.

According for the resisting approaches and avoiding strategies of privacy-preserving, we summarized the defense encryption methods and perturbation techniques from four aspects: homomorphic encryption, secure multiparty computation, differential privacy and masking. We then investigated three types of attacks: training data poisoning attacks, membership inference attacks, model inference attacks, along with adversarial samples attacks, and their excellent performance, effective prevention methods and detection technology. Whereafter, the academic research direction and application prospects field of FDL are pointed out, eventually, the open problems and the future trajectories are discussed.

Acknowledgements

This work was supported by the National Natural Science Foundation of China, the National High Technology Research and Development Program of China, the project of Science and Technology Department of Sichuan Province(Grant No. 2021YFG0322), the project of Science and Technology Department of Chongqing Municipality, the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-K202114401)，Chongqing Qinchengxing Technology Co., Ltd., Chengdu Haitian Digital Technology Co., Ltd., Chengdu Chengdian Network Technology Co., Ltd., Chengdu Civil-military Integration Project Management Co., Ltd., and Sichuan Yin Ten Gu Technology Co., Ltd.

References

[1] S.-W. Lee, H. Mohammed sidqi, M. Mohammadi, S. Rashidi, A. M. Rahmani, M. Masdari, and M. Hosseinzadeh, “Towards secure intrusion detection systems using deep learning techniques: Comprehensive analysis and review,” *Journal of Network and Computer Applications,* vol. 187, 2021.

[2] J. Sun, T. Chen, G. B. Giannakis, Q. Yang, and Z. Yang, “Lazily Aggregated Quantized Gradient Innovation for Communication-Efficient Federated Learning,” *IEEE Trans Pattern Anal Mach Intell,* vol. 44, no. 4, pp. 2031-2044, Apr, 2022.

[3] S. Feng, B. Li, H. Yu, Y. Liu, and Q. Yang, “Semi-Supervised Federated Heterogeneous Transfer Learning,” *Knowledge-Based Systems,* vol. 252, 2022.

[4] P. Xu, M. Hu, T. Chen, W. Wang, and H. Jin, “LaF: Lattice-Based and Communication-Efficient Federated Learning,” *IEEE Transactions on Information Forensics and Security,* vol. 17, pp. 2483-2496, 2022.

[5] J. S. Ng, W. Y. B. Lim, Z. Xiong, X. Cao, D. Niyato, C. Leung, and D. I. Kim, “A Hierarchical Incentive Design Toward Motivating Participation in Coded Federated Learning,” *IEEE Journal on Selected Areas in Communications,* vol. 40, no. 1, pp. 359-375, 2022.

[6] Z. Wu, H. Wang, Z. Wang, H. Jin, and Z. Wang, “Privacy-Preserving Deep Action Recognition: An Adversarial Learning Framework and A New Dataset,” *IEEE Trans Pattern Anal Mach Intell,* vol. 44, no. 4, pp. 2126-2139, Apr, 2022.

[7] L. Idan, and J. Feigenbaum, “PRShare: A Framework for Privacy-preserving, Interorganizational Data Sharing,” *ACM Transactions on Privacy and Security,* vol. 25, no. 4, pp. 1-38, 2022.

[8] Y. Kang, Y. He, J. Luo, T. Fan, Y. Liu, and Q. Yang, “Privacy-preserving Federated Adversarial Domain Adaptation over Feature Groups for Interpretability,” *IEEE Transactions on Big Data*, 2022.

[9] K. Hariss, and H. Noura, “Towards a fully homomorphic symmetric cipher scheme resistant to plain-text/cipher-text attacks,” *Multimedia Tools and Applications,* vol. 81, no. 10, pp. 14403-14449, 2022.

[10] Z. Zhou, Q. Fu, Q. Wei, and Q. Li, “LEGO: A hybrid toolkit for efficient 2PC-based privacy-preserving machine learning,” *Computers & Security,* vol. 120, 2022.

[11] W. Xu, B. Wang, Q. Qu, T. Zhou, and P. Duan, “Modified Multi-Key Fully Homomorphic Encryption Scheme in the Plain Model,” *The Computer Journal*, 2022.

[12] S. Tan, B. Knott, Y. Tian, and D. J. Wu, “CryptGPU: Fast Privacy-Preserving Machine Learning on the GPU,” in 2021 IEEE Symposium on Security and Privacy (SP), 2021, pp. 1021-1038.

[13] X. Zhang, X. Chen, M. Hong, S. Wu, and J. Yi, "Understanding clipping for federated learning: Convergence and client-level differential privacy." pp. 26048-26067.

[14] Z. Liang, B. Wang, Q. Gu, S. Osher, and Y. Yao, “Differentially private federated learning with Laplacian smoothing,” *arXiv preprint arXiv:2005.00218*, 2020.

[15] L. Lyu, Y. Li, K. Nandakumar, J. Yu, and X. Ma, “How to Democratise and Protect AI: Fair and Differentially Private Decentralised Deep Learning,” *IEEE Transactions on Dependable and Secure Computing*, pp. 1-1, 2020.

[16] N. Wu, F. Farokhi, D. Smith, and M. A. Kaafar, “The Value of Collaboration in Convex Machine Learning with Differential Privacy,” in 2020 IEEE Symposium on Security and Privacy (SP), 2020, pp. 304-317.

[17] W.-N. Chen, A. Ozgur, and P. Kairouz, "The Poisson Binomial Mechanism for Unbiased Federated Learning with Secure Aggregation." pp. 3490-3506.

[18] R. Liu, Y. Cao, M. Yoshikawa, and H. Chen, "Fedsel: Federated sgd under local differential privacy with top-k dimension selection." pp. 485-501.

[19] S.-J. Hahn, M. Jeong, and J. Lee, “Connecting Low-Loss Subspace for Personalized Federated Learning,” in Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 505-515.

[20] J. Zhou, C. Chen, L. Zheng, H. Wu, J. Wu, X. Zheng, B. Wu, Z. Liu, and L. Wang, "Vertically federated graph neural network for privacy-preserving node classification," *arXiv preprint arXiv:2005.11903*, 2020.

[21] X. Huang, Y. Ding, Z. L. Jiang, S. Qi, X. Wang, and Q. Liao, "DP-FL: a novel differentially private federated learning framework for the unbalanced data," *World Wide Web,* vol. 23, no. 4, pp. 2529-2545, 2020.

[22] Z. Xiong, Z. Cai, D. Takabi, and W. Li, "Privacy Threat and Defense for Federated Learning With Non-i.i.d. Data in AIoT," *IEEE Transactions on Industrial Informatics,* vol. 18, no. 2, pp. 1310-1321, 2022.

[23] S.-J. Hahn, and J. Lee, "Privacy-preserving federated bayesian learning of a generative model for imbalanced classification of clinical data," *arXiv preprint arXiv:1910.08489*, 2019.

[24] Y. Jiang, K. Zhang, Y. Qian, and L. Zhou, "Anonymous and Efficient Authentication Scheme for Privacy-Preserving Distributed Learning," *IEEE Transactions on Information Forensics and Security,* vol. 17, pp. 2227-2240, 2022.

[25] M. Xue, C. He, J. Wang, and W. Liu, "One-to-N & N-to-One: Two Advanced Backdoor Attacks Against Deep Learning Models," *IEEE Transactions on Dependable and Secure Computing,* vol. 19, no. 3, pp. 1562-1578, 2022.

[26] N. Mhaisen, A. A. Abdellatif, A. Mohamed, A. Erbad, and M. Guizani, "Optimal User-Edge Assignment in Hierarchical Federated Learning Based on Statistical Properties and Network Topology Constraints," *IEEE Transactions on Network Science and Engineering,* vol. 9, no. 1, pp. 55-66, 2022.

[27] Y. Qi, M. S. Hossain, J. Nie, and X. Li, "Privacy-preserving blockchain-based federated learning for traffic flow prediction," *Future Generation Computer Systems,* vol. 117, pp. 328-337, 2021.

[28] X. Yang, Y. Feng, W. Fang, J. Shao, X. Tang, S.-T. Xia, and R. Lu, "An Accuracy-Lossless Perturbation Method for Defending Privacy Attacks in Federated Learning." pp. 732-742.

[29] K. Zhao, W. Xi, Z. Wang, J. Zhao, R. Wang, and Z. Jiang, "SMSS: Secure Member Selection Strategy in Federated Learning," *IEEE Intelligent Systems,* vol. 35, no. 4, pp. 37-49, 2020.

[30] Y. Wen, J. Geiping, L. Fowl, M. Goldblum, and T. Goldstein, "Fishing for user data in large-batch federated learning via gradient magnification," *arXiv preprint arXiv:2202.00580*, 2022.

[31] A. Yehezkel, E. Elyashiv, and O. Soffer, "Network anomaly detection using transfer learning based on auto-encoders loss normalization." pp. 61-71.

[32] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou, "A hybrid approach to privacy-preserving federated learning." pp. 1-11.

[33] B. Jia, X. Zhang, J. Liu, Y. Zhang, K. Huang, and Y. Liang, "Blockchain-Enabled Federated Learning Data Protection Aggregation Scheme With Differential Privacy and Homomorphic Encryption in IIoT," *IEEE Transactions on Industrial Informatics,* vol. 18, no. 6, pp. 4049-4058, 2022.

[34] T. Ryffel, P. Tholoniat, D. Pointcheval, and F. Bach, "Ariann: Low-interaction privacy-preserving deep learning via function secret sharing," *arXiv preprint arXiv:2006.04593*, 2020.

[35] Z. Ma, J. Ma, Y. Miao, Y. Li, and R. H. Deng, "ShieldFL: Mitigating Model Poisoning Attacks in Privacy-Preserving Federated Learning," *IEEE Transactions on Information Forensics and Security,* vol. 17, pp. 1639-1654, 2022.

[36] Z. Zhang, A. Panda, L. Song, Y. Yang, M. Mahoney, P. Mittal, R. Kannan, and J. Gonzalez, "Neurotoxin: Durable backdoors in federated learning." pp. 26429-26446.

[37] H. Hu, and J. Pang, "Stealing Machine Learning Models: Attacks and Countermeasures for Generative Adversarial Networks," in Annual Computer Security Applications Conference, 2021, pp. 1-16.

[38] H. Aydın, Z. Orman, and M. A. Aydın, "A long short-term memory (LSTM)-based distributed denial of service (DDoS) detection and defense system design in public cloud network environment," *Computers & Security,* vol. 118, 2022.

[39] M. Fang, X. Cao, J. Jia, and N. Gong, "Local model poisoning attacks to {Byzantine-Robust} federated learning." pp. 1605-1622.

[40] G. Xu, H. Li, S. Liu, K. Yang, and X. Lin, "VerifyNet: Secure and Verifiable Federated Learning," *IEEE Transactions on Information Forensics and Security,* vol. 15, pp. 911-926, 2020.

[41] C. Wang, and H. Zhu, "Wrongdoing Monitor: A Graph-Based Behavioral Anomaly Detection in Cyber Security," *IEEE Transactions on Information Forensics and Security,* vol. 17, pp. 2703-2718, 2022.

[42] J. Pei, K. Zhong, M. A. Jan, and J. Li, "Personalized federated learning framework for network traffic

anomaly detection," *Computer Networks,* vol. 209, 2022.
[43] A. Qayyum, M. U. Janjua, and J. Qadir, "Making federated learning robust to adversarial attacks by learning data and model association," *Computers & Security,* vol. 121, pp. 102827, 2022.
[44] Y. Liu, A. Huang, Y. Luo, H. Huang, Y. Liu, Y. Chen, L. Feng, T. Chen, H. Yu, and Q. Yang, "Fedvision: An online visual object detection platform powered by federated learning." pp. 13172-13179.
[45] R. Hamza, A. Hassan, A. Ali, M. B. Bashir, S. M. Alqhtani, T. M. Tawfeeg, and A. Yousif, "Towards Secure Big Data Analysis via Fully Homomorphic Encryption Algorithms," *Entropy (Basel),* vol. 24, no. 4, Apr 6, 2022.
[46] F. Liu, Z. Cheng, H. Chen, Y. Wei, L. Nie, and M. Kankanhalli, "Privacy-Preserving Synthetic Data Generation for Recommendation Systems." pp. 1379-1389.
[47] S. Chen, B. Duan, C. Zhu, C. Tang, S. Wang, Y. Gao, S. Fu, L. Fan, Q. Yang, and Q. Liu, "Privacy-preserving integration of multiple institutional data for single-cell type identification with scPrivacy," *bioRxiv*, 2022.
[48] V. Subramaniyaswamy, V. Jagadeeswari, V. Indragandhi, R. H. Jhaveri, V. Vijayakumar, K. Kotecha, and L. Ravi, "Somewhat Homomorphic Encryption: Ring Learning with Error Algorithm for Faster Encryption of IoT Sensor Signal-Based Edge Devices," *Security and Communication Networks,* vol. 2022, 2022.
[49] Y. Li, J. Li, and Y. Wang, "Privacy-preserving spatiotemporal scenario generation of renewable energies: A federated deep generative learning approach," *IEEE Transactions on Industrial Informatics,* vol. 18, no. 4, pp. 2310-2320, 2021.
[50] J. Zhou, W. Ding, and W. Yang, "A Secure Encoding Mechanism Against Deception Attacks on Multisensor Remote State Estimation," *IEEE Transactions on Information Forensics and Security,* vol. 17, pp. 1959-1969, 2022.