Federated Learning and its Applications for Security and Communication

Hafiz M. Asif[1]
Department of Electrical & Computer Engineering
Sultan Qaboos University
Muscat, Oman

Mohamed Abdul Karim[2]
Department of Information Technology
University of Technology and Applied Sciences
Suhar Campus, Oman

Firdous Kausar[3]
Department of Electrical & Computer Engineering
Sultan Qaboos University
Muscat, Oman

***Abstract*—The not so long ago, Artificial Intelligence (AI) has revolutionized our life by giving rise to the idea of self-learning in different environments. Amongst its different variants, Federated Learning (FL) is a novel approach that relies on decentralized communication data and its associated training. While reducing the amount of data acquired from users, federated learning derives the benefits of popular machine learning techniques, it brings learning to the edge or directly on-device. FL, frequently referred to as a new dawn in AI, is still in its early stages and is yet to garner widespread acceptance, owing to its (unknown) security and privacy implications. In this paper, we give an illustrative explanation of FL techniques, communication, and applications with privacy as well as security issues. According to our findings, there are fewer privacy-specific dangers linked with FL than security threats. We conclude the paper with the challenges of FL with special emphases on security.**

Keywords*—Federated learning; communication; security; deep learning; Artificial Intelligence

I. Introduction

In the modern era, ubiquitous mobile gadgets are coupled with computation and sensor capabilities that collect large volumes of data. Such massive quantities of data are used to train various learning algorithms. These learning techniques, when combined with Data Mining and AI in other words with Deep Learning (DL) breakthroughs, enable a wide range of beneficial applications, including image analysis, speaker identification, healthcare, vehicular networks, among others. Machine Learning (ML) techniques need to be checked in and generally have to be consolidated on internet-based cloud services. However, due to the enormous volumes of data and privacy-critical nature, login into such cloud services to train supervised learning is cumbersome. As a result, major challenges such as excessive latency and transmission inefficiencies arise. The notion of Federated Training or Learning (FL) has now been proposed in face of emerging privacy rules in various nations. Mobile phone users in Federated Learning (FL) can train a feature map by pooling their native models without disclosing their confidential material. In ML, a model is usually developed by training locally on the user's own server (PC) whereas, in FL, the model is built by training on different machines located at distributed locations and there is no dedicated connection between the servers. They have their dataset or database sample at their ends. In simple terms, a form of machine learning that is decentralized is termed federated learning [1]. While there has been some research on this subject, there is not enough progress in terms of comprehending FL's security and privacy implications. This paper aims to provide a full review of FL in terms of a formal definition, then we compare ML models with salient features and tabulate. We also discuss the pros and cons graphically along with the challenges. Finally, we provide some recommendations, making this unique among previous studies. The following is a summary of this paper's contributions to the field's recent literature:

- Providing a categorization and review of the FL methods and strategies.
- Identifying and examining pros and cons in FL environments.
- Delineating potential applications of FL environments.
- Highlighting challenges faced by FL systems with special emphasis on the security.
- Providing recommendations to enhance the security and privacy of the FL implementation.

II. Federated Model and Critical Analysis

ML techniques traditionally require that all the training data be centralized on a single server in a datacenter or the cloud. With enormous increase in the number of mobile devices and the training data available on various machines, the challenge of assimilating the relevant data arises. Federated learning uses the model training approach that enables a device to train from the collaboration of shared models. Proxy data on the server initially trains the shared model. Subsequently, the model is downloaded on each device and then improved by data locally stored on the device, which is also termed as federated data. ML algorithms assume that all learning data is available and maintained in a centralized dataset. In order to facilitate such training, centralized learning networks are created. These networks have serious privacy concerns, high communication costs, and scalability challenges. Federated Learning (FL) has been introduced to enable remote supervised learning without a centralized training classifier, considering the aforementioned difficulties. It can be observed from Fig. 1(b) that the federated learning network is composed of multiple Edge Devices (EDs)

and servers. According to a survey, in federated learning networks, a machine-learning model is trained with two iterative steps such as local model training and global model aggregation at EDs and server respectively [2]. In the first stage, EDs update the local model with the downloaded model from the server, algorithms of Stochastic Gradient Descent are executed to learn the local model with their dataset and upload the updated model to the server. In the next phase, model updates received by the server are aggregated with weighted average to previous global model and thus the new model is obtained. These two steps involve a training round. In a federated learning model or network, the parameters of the ML model are exchanged instead of data and this prevents and reduces privacy issues with the reduction of communication overhead. Federated learning is deployed with flexibility in multiple environments including the mobile environment that is a complicated one [3]. Comparison of machine learning models with salient features is given in Table I.

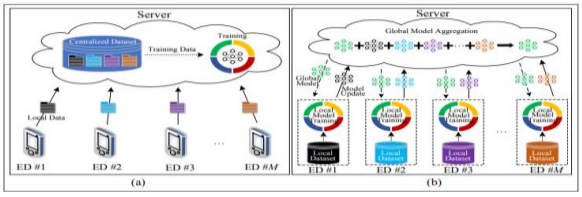

Fig. 1. System Model between Centralized ML Network vs Decentralized FL Network [4].

Federated Learning Network (FLN) has also been adopted to orchestrate various mobile devices across the world for training language models with BERT [7]. All such mobile devices are owned by different users and then connected to multiple types of links such as WiFi, mobile network, etc. Hence, in terms of ownership, capabilities, and computing, Edge Devices (EDs) in federated learning model are heterogeneous [1].

TABLE I. COMPARISON OF MACHINE LEARNING MODELS WITH SALIENT FEATURES [2]

Scheme	Salient features	Used in percentage according to survey
Distributed learning	Provision of holistic estimation of parameters	21%
Parallel learning	Distribution of data in laid fashion	27%
Federated learning	Model training using natural database, massive distribution of data over local learners	45%
Ensemble learning	Production of an optimal model	7%

Federated Learning (FL) is reliable for joint ED's efforts for the training of the ML model. Even with an abnormality of few EDs, Machine Learning model can be tampered. Besides all this, the FL model or network has multiple attack surfaces concerned with the security of federated learning such as malicious EDs, and insecure connections. These attack surfaces are vulnerable to many security issues in FL networks such as data positioning as well as model positioning [4]. Survey analysis of the given graph shows the results of two cases in comparison with using active federated learning framework (Fig. 2).

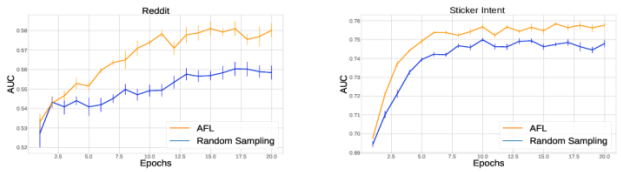

Fig. 2. Comparison of AUC on Reddit and Sticker Intent using Active FL Framework [8].

III. ADVANTAGES OF FEDERATED LEARNING

Diversity of data: Large-scale ML models may be unable to merge datasets from diverse sources. The reasons for impediments are partially due to the information security, reluctance and connection unavailability among the edge devices. On the other hand, Federated learning makes it easier to access diverse data, even when sources of data could only interact at a particular period (Fig. 3).

- Real-time learning continuity: There is no requirement for aggregate data in continuous learning because algorithms are constantly upgraded using client information.
- The efficiency of hardware: Since decentralized learning methods do not require a single, complex cloud database to interpret data, this strategy requires less complex hardware infrastructure [6].

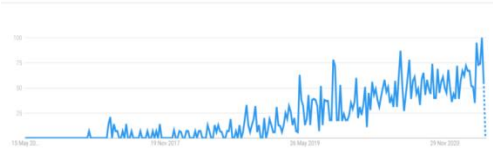

Fig. 3. Interest Overtime Related to use of Federated Learning [5].

IV. APPLICATIONS OF FEDERATED LEARNING

In terms of distributed machine learning, FL is a viable approach with the inherent characteristics of privacy. Without communicating data, many nodes can work together to develop a collaborative learning model. Data access rights, privacy, security, and access to a variety of data types can all be managed in this way. It is believed that FL has applications in a wide range of areas such as Industrial IoT, healthcare, smart transportation, self-driving cars; traffic forecasting; smart buildings, recommender system, Fintech, the insurance sector, and telecommunications [13-14].

FL is a revolutionary technique for machine learning. It has the potential to have a profound impact on the healthcare sector. It can also help healthcare workers in many ways. Sharing health care data raises a number of privacy issues. In addition, strict laws like HIPPA make it more challenging to exchange critical data, which has made it more difficult to conduct studies that could lead to new medical advancements. All parties, including hospitals, AI businesses, and regulatory

authorities, have a responsibility to protect extremely sensitive information. Researchers are currently investigating how FL may be utilized to protect patients' privacy while beneficially utilizing their data. FedHealth [14] is the first federated transfer learning framework for wearable healthcare, capable of providing precise and individualized healthcare without risking patient privacy. A community-based federated learning algorithm (CBFL) [15] proposes a system that clusters distributed data into clinically significant communities based on shared diagnoses and geographical locations and then develops a model for each community. Li et al. [16] develop a brain tumor segmentation FL system using differential privacy to protect patient data. Patients with uncommon tumors will benefit from Owkin's FL-based platform, which will be used in tests to determine drug toxicity, predict disease progression, and assess survival rates [17].

Zhang et al. [18] propose an Industrial Internet identification using blockchain and federated learning technologies, which provides privacy protection. Liu et al. [19] develop an on-device FL-based deep anomaly detection system for IIoT time series data sensing, which detects edge devices' failure in IIoT industrial product production. Edge device failures adversely affect IIoT industrial product production. Khanal et al. [20] examine the value of proactive content caching in self-driving cars to reduce content retrieval costs and improve QoE with edge cloud infrastructure. It extracts local content popularity patterns in self-driving automobiles utilizing LSTM-based prediction mechanisms in a federated scenario to predict regional content popularity.

Machine learning is constantly growing and reshaping the technological landscape. FL applications, like any other machine learning technique, face challenges. In spite of its flaws, it has the potential to transform numerous industries. There will be tremendous progress in FL and its diverse applications soon. When applied effectively, it can aid in the evolution of numerous sectors and benefit users.

Another area where FL finds its rigorous application is data communication. For instance, the feasibility of FL for its using in 6G communication systems has been investigated in [21]. The FL key challenges for 6G include security, cost-effective systems, and privacy concerns. FL can also be used for data augmentation in wireless communication. For instance, edge users can cooperate by sharing certain parameters, which in turn significantly reduces the communication overhead [22]. FL also finds its application in Wireless Power Transfer (WPT) where Wireless-Power enabled can be enabled. A complete wireless-power enabled FL has been investigated in [23].

V. Challenges of Federated Learning with Special Emphases on Security

There are multiple disadvantages related to security issues of the federated learning model. These include data positioning as well as model positioning. The main aim of the positioning attack is to degrade the accuracy of the machine-learning model. This happens by tampering the aggregation of global models with updates of the poisoned model of federated learning [9]. The attack surfaces for such insecurities are Malicious Edge Devices (MEDs) and insecure connections. MEDs are set by the attackers in smart devices through malware. As new smart devices are more sophisticated and have inescapable flaws. Hence, it is convenient for attackers to join the Federated Learning Networks (FLNs) through malicious EDs. Moreover, security of all connections through which the EDs of federated learning model are connected to the network needs be monitored. Wireless connection has vulnerability through various channels. Through such insecure connections, the uploaded model updates of Federated Learning might be manipulated or hijacked. In data poisoning attacks on the security of federated learning, these intentional attacks intend to achieve low accuracy of machine learning models on certain classes. Attackers to the federated learning securities flip labels of training data in those concerned classes [5].

In model poisoning attack on the security of federated learning, the attack is concerned with the ML model updates that are generated from Gaussian distribution (see Table II). In this, the attacker manipulates updates of the benign model into poisoned updates. To achieve this purpose, attackers use updates of pre-model designs to craft the updates of the poisoned model and replace the ML model with the pre-designed poisoned models. The vulnerability points for all these attackers are insecure connections and malicious Edge Devices (EDs). Furthermore, other disadvantages are performance limitations, indirect leakage of information, and a degree of centralization [10].

TABLE II. Attacks at Securities of Federated Learning [8]

Security attacks	**Description**	**Methodology for attacks**	**Target users**
Data poisoning	Training data is modified by attackers and EDs training is made incorrect as well as poisoned updates of the model are generated.	Labels of training data are flipped intentionally, and labels of training data are also flipped in certain classes.	Unintentionally
Model poisoning	Poisoned updates of the model are created by attackers Benign model updates are manipulated based on pre-designed rules.	• Model updates are generated using pre-designed poisoned model to impact the security of federated learning. • Flipping signs of model updates.	Intentionally

Lots of investment is required for federated learning models with frequent communication and large storage capacity with high bandwidth. Data is not collected on a single entity, which increases attack surfaces [1]. The below Fig. 4 depicts the secure aggregation of private federated learning. In this scenario, aggregator or server builds a global model jointly without revealing the security of training data. Hence, it is powerful in terms of keeping privacy while computing millions of data in parallel [3] (Fig. 5).

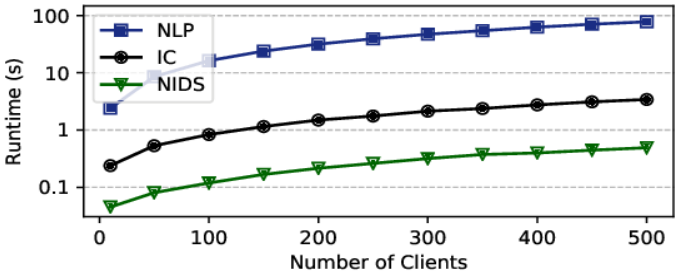

Fig. 4. Secure Aggregation of Private Federated Learning [1].

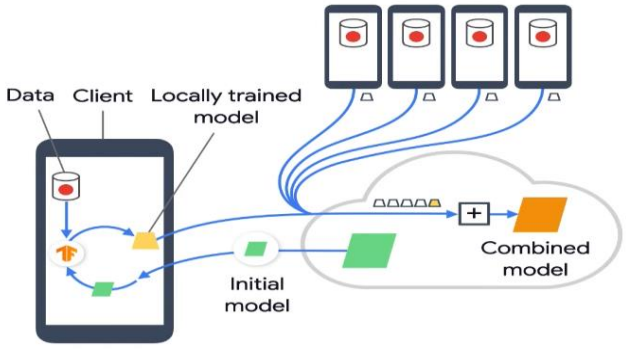

Fig. 5. A Typical Federated Learning Model.

A. Active Federated Learning

The gadgets obtain a training program (which is normally small size in terms of few bytes).

- The gadgets are programmed to learn from local data.
- The sensing notes the computer anonymized updates mostly on variables.
- The data from devices are aggregated by the administrator. The server combines the information it receives from each variety of technologies to conduct an approach with regards to the present system by each grouping.
- The newly added model is delivered to the gadgets with an assessment (again, the idea of decentralization is at work here) as well as a fresh round between training after several rounds of learning [8] (Fig. 6).

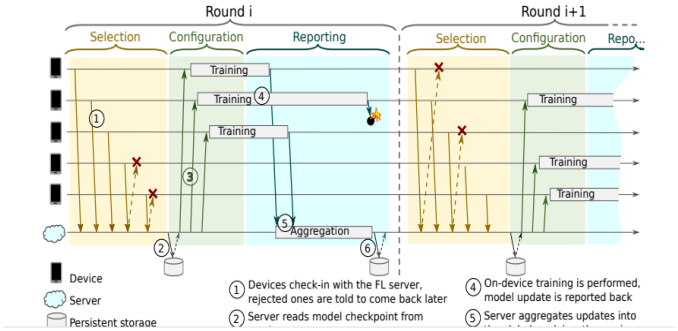

Fig. 6. Federated Learning Protocol [11].

VI. Recommendations and Suggestions

Privacy preservation, safe multiparty processing, and cryptography are examples of confidentiality technologies that can be utilized to improve the data protection possibilities of federated learning. A variety of measures is suggested in this section. First, sharing less information about the generic model updation at the server can maximize the privacy of the Federated-learning model [12]. Moreover, the use of deep neural network also makes complex the use of available gradients. There is also possibility for developers to choose or create an algorithm that has less chance of data breaching and attack on the security of federated learning system. Using more privacy regulations will also inevitably make data acquisition easier and less vulnerable to exploitation [1].

VII. Conclusion

The paper discussed the federated learning techniques and applications with respect to privacy as well as security issues. Federated learning has been successfully implemented in a variety of settings, such as the challenging mobile environment. Despite the advantages of federated learning, there are many privacy and security issues related to the model. When contrasted to exchanging personal data across data centres, federated learning offers certain privacy benefits. The capability to immensely develop machine-learning algorithms depending on user input, while minimizing bandwidth impacts for uploading confidential information over the network is also one of the advantages. Data poisoning and model poisoning are two major security attacks on federating learning networks. Among communication networks, wireless connections are vulnerable. Federated learning system updates can also be altered or hijacked over such unsafe connections. In information poisoning threats on supervised learning of federated model security, these deliberate attacks aim to achieve poor sensitivity of machine learning techniques on specific classes. These attacks are vulnerable through two attack surfaces of federated learning mode such as internet connections and Edge Devices (EDs) of the federated learning model.

Acknowledgment

The authors would like to thank Sultan Qaboos University and University of Technology and Applied Sciences for their support.

REFERENCES

[1] M. Asad, A. Moustafa, T. Ito and M. Aslam, "Evaluating the Communication Efficiency in Federated Learning Algorithms," IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 552-557, 2011.

[2] Uddin, M.P., Xiang, Y., Lu, X., Yearwood J., and Gao L., "Mutual Information Driven Federated Learning," IEEE Transactions on Parallel and Distributed Systems, vol.32, no.7, 1526-1538, 2021.

[3] Asad, M., "Federated Learning Versus Classical Machine Learning: A Convergence Comparison," Journal of scientific research, vol.2, no.2, 23-31, 2019.

[4] McMahan, B., & Ramage, D., "Federated learning: Collaborative machine learning without centralized training data," Google Research Blog, 3, 2017, https://ai.googleblog.com/2017/04/federated-learning-collaborative.html accessed on 11 March 2022.

[5] Dilmegani, C., "What is Federated Learning(FL)?," 2020, Techniques & Benefits in 2021. [Online] https://research.aimultiple.com/federated-learning/ accessed on 11 March 2022.

[6] Tan, J., Liang, Y. –C., Luong, N.C., and Niyato, D., "Toward Smart Security Enhancement of Federated Learning Networks. IEEE Network," vol.35, no.1, pp.340-347, 2021.

[7] lmkil, A., Callh, S., Barbieri, M., S¨utfeld, L.R., Zec, E.L., Mogren, O., "Scaling federated learning for fine-tuning of large language models," Métais, E., Meziane, F., Horacek, H., Kapetanios, E. (eds) Natural Language Processing and Information Systems, 2021.

[8] Jack Goetz, Kshitiz Malik, Duc Bui, Seungwhan Moon, Honglei Liu, Anuj Kumar, "Active Federated Learning," arXiv preprint arXiv:1909.12641, 2019.

[9] Li, T., Sahu, A.K., Talwalkar, A. and Smith, V., "Federated learning: Challenges, methods, and future directions," IEEE Signal Processing Magazine, vol.37, no.3, pp.50-60, 2020.

[10] Fung, C., Yoon, C.J. and Beschastnikh, I., "Mitigating sybils in federated learning poisoning," arXiv preprint arXiv:1808.04866, 2018.

[11] Aledhari, M., Razzak, R., Parizi, R.M. and Saeed, F., "Federated learning: A survey on enabling technologies, protocols, and applications," IEEE Access, 8, 140699-140725, 2020.

[12] Xu, R., Baracaldo, N., Zhou, Y., Anwar, A. and Ludwig, H., "Hybridalpha: An efficient approach for privacy-preserving federated learning," In Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, 13-23, 2019.

[13] IEEE Guide for Architectural Framework and Application of Federated Machine Learning. IEEE Std 3652.1-2020,1-69, 2021.

[14] Shaheen, M.; Farooq, M.S.; Umer, T.; Kim, B.-S., "Applications of Federated Learning; Taxonomy, Challenges, and Research Trends," Electronics. Vol.11, 670. https://doi.org/10.3390/electronics11040670, 2022.

[15] Chen Y., Qin X., Wang J., Yu C. and Gao W., "FedHealth: a federated transfer learning framework for wearable healthcare," IEEE Intelligent Systems, vol.35, no.4, pp.83-93, 2020.

[16] W. Li, et al., "Privacy-preserving federated brain tumour segmentation," Machine Learning in Medical Imaging, doi:10.1007/978-3-030-32692-0_16, 2019.

[17] Online, "Federated Learning—OWKIN," Available online: https://owkin.com/federated-learning/ (accessed on June, 2022).

[18] Zhang X., Hou H., Fang Z., and Wang Z., "Industrial Internet Federated Learning Driven by IoT Equipment ID and Blockchain," Wireless Communications and Mobile Computing, Article ID 7705843, 9 pages, 2021.

[19] Liu Y., Garg S., Nie J., Zhang Y., Xiong Z., Kang J., Hossain M., "Deep Anomaly Detection for Time-Series Data in Industrial IoT: A Communication-Efficient On-Device Federated Learning Approach," IEEE Internet of Things Journal, vol.8, no.8, pp.6348-6358, 2021.

[20] Khanal, S., Thar, K., Hossain, M. D., and Huh, E.N., "Proactive Content Caching at Self-Driving Car Using Federated Learning with Edge Cloud," Twelfth International Conference on Ubiquitous and Future Networks (ICUFN), pp.129-134,2021.

[21] Liu, Y., Yuan, X., Xiong, Z., Kang, J., Wang, X., and Niyato, D., "Federated learning for 6G communications: Challenges, methods, and future directions," China Communications, vol.17, no.9, 105-118, 2020.

[22] Yan, M., Chen, B., Feng, G., and Qin, S., "Federated Cooperation and Augmentation for Power Allocation in Decentralized Wireless Networks," IEEE Access, vol.8, pp.48088-48100, 2020.

[23] B. Clerckx, B., Huang, K., Varshney, L. R., Ulukus, S., and Alouini M.S., "Wireless Power Transfer for Future Networks: Signal Processing, Machine Learning, Computing, and Sensing," IEEE Journal of Selected Topics in Signal Processing, vol.15, no.5, 1060-1094. 2021.