# FL-DP: Differential Private Federated Neural Network

Muhammad Maaz Irfan[1,2], Lin Wang[1,2], Sheraz Ali[1,2], Shan Jing[1,2(✉)], and Chuan Zhao[1,2,3]

[1] School of Information Science and Engineering, University of Jinan, Jinan 250022, China
jingshan@ujn.edu.cn
[2] Shandong Provincial Key Laboratory of Network-based Intelligent Computing, University of Jinan, Jinan 250022, China
[3] Shandong Provincial Key Laboratory of Software Engineering, Jinan, China

**Abstract.** The rapid development of the Internet and machine learning has brought convenience and comfort to users' lives. However, due to various attacks and sensitive data leaks, the large amount of data used in machine learning training has made the issue of personal privacy a growing concern as well. In the era of big data, anyone's information can be stolen, which makes many people feel uneasy. We propose a new approach called FL-DP (Federated Learning Based on Differential Privacy). Based on differential privacy, this approach can effectively restrict the adversary's access to the client model, which has the ability to limit data leakage. In this framework, we use the DP Laplace mechanism. It is ensured that all operations (including the server-side aggregation process) are secure and do not leak any information about the training data. Also, we consider adding multiple noises to the preprocessing process so that the client-side data is secure during the training process. Our approach not only provides client-level privacy, but also balances efficiency and privacy. After evaluation, our approach is highly scalable and can be applied to most machine learning based applications.

**Keywords:** Federated learning · Data integrity · Privacy-preserving · Differential privacy

## 1 Introduction

In recent decades, machine learning-based systems have been successfully applied in various areas of social industry. From engineering solutions to advances in technology to the development of intelligent web-based systems that provide online services, machine learning plays an irreplaceable role [9]. Moreover, in order not to share data directly, the concept of federated learning has been proposed, where federated learning frameworks use distributed storage and processing of data as opposed to centralized approaches. They use sensors to collect data in


W. Shi et al. (Eds.): SPNCE 2021, LNICST 423, pp. 271–281, 2022.
https://doi.org/10.1007/978-3-030-96791-8_20

the physical environment, however, this feature poses new challenges for user privacy.

As attacks become more sophisticated over time, privacy is a huge challenge for machine learning. Powerful computer systems, such as smartphones, laptops, and desktop computers, can use sophisticated methods to observe malicious behavior and malware. However, some lightweight systems have limited computational resources, so they can only detect small attack variations, such as the Zero Day attack [8]. Joint learning (FL) algorithms can be used to enhance the IoT network and improve the efficiency of the whole system.

Research shows that the effect of deep learning model is related to model size and training data set. Deep learning is widely used in image recognition, feature extraction, classification and prediction. With the rapid growth of data and model parameters, it is often shown as a large number of model parameters, which will have high requirements for the amount of calculation. This scenario brings some challenges related to data security and privacy. These challenges come from the lack of effective tools and methods to protect large data sets. In an environment where the Internet is becoming more developed, heterogeneous data is more easily collected and the volume of data is more enormous. Data protection raises more concerns in environments where privacy protection is required.

The most advanced existing solutions consider the training of centralized ML model. But it is stored locally and preprocessed; For example, the machine learns from sensitive data such as personal images, audio, video, etc. Usually, the parameters of the training model must be transformed into secret form, not the plaintext of the training sample. In order to achieve this and provide strong privacy protection when training data sensitivity, the common practice is to use differential privacy technology. In addition, the computing and storage capacity of devices in distributed systems is growing, and powerful local computing resources can be used on each device. This has led to an increasing interest in federated learning [3], which directly explores training statistical models on remote devices. The current research results have made breakthrough progress in the fields of large-scale machine learning, privacy and distribution [1,8].

Jointly learn the general method of "bringing code into data rather than data into code", and solve the problems of data privacy, location and ownership [7]. It is far away from the use of local models, and carries out prediction on mobile devices by bringing model training to mobile devices. In this case, the key challenge is to check when training a single global model, and the data will not be distributed on the device. Overcome these problems. FedAvg [11,12] was introduced. In i.i.d, parallel SGD and related variants similar to fedavg are analyzed, which are updated locally. However, multiple attacks may occur during training and aggregation, which will lead to private data leakage [4]. In this paper, we provide a solution to this problem. Specifically, we propose a new protocol to implement secure computing.

We use differential privacy (DP) to construct an efficient federated learning framework based on Laplace noise. Our state-of-the-art FL iotdp joint learning

Internet of things protects privacy, effectively limits adversaries' access to the client model, and has the ability to limit data leakage. In this framework, we use DP Laplace mechanism. Ensure that all operations (including server-side aggregation process) are secure and will not disclose any private information about training data. At the same time, we consider adding multiple noises in the preprocessing, and each client does not need to share. It is convenient for the client to join the model update or offline at any time, which greatly improves the flexibility and scalability of the system. This is a state-of-the-art framework that provides a trade-off between privacy costs and model efficiency.

The organization of our research paper is organized as follows: we will start with the introduction, and then present related work in Sect. 2, in Sect. 3 we give an overview of technical preliminaries about federated learning and differential privacy, and then in Sect. 4 a novel protocol for FL-NNPP Federated Learning Neural Network for Privacy-Preserving. Furthermore, we give experimental results for the method of our protocol in terms of accuracy, scalability, and security trade-offs in Sect. 5. Finally in Sect. 6 we will conclude our paper with future directions.

## 2 Related Work

The benefits of federated learning emerge as awareness of machine learning increases, as do the number of channels for collecting personal information, and as people become more aware of information security. This is because federated learning allows users to train data locally without exchanging data, thus protecting the privacy of the data. However, as the research progressed, researchers found that the weights uploaded by users also expose some information about the user [10,16].

The most natural way to prevent the leakage of information is to use the method of adding noise, which we know as differential privacy (DP). The existing algorithms include Local Differential Privacy (LDP). [5]adds noise to the local information to protect it from being leaked. The work in [14] proposes a solution for building LDP-compliant SGDs that support a variety of important ML tasks. The work in [15] considers the distributed estimation of client uploaded data by the server while providing protection of these data using LDP.

At present, the research on differential privacy has become mature and has mature applications [11,12]. However, it is still a difficult goal for differentiated private in-depth learning to achieve efficient use under reasonable privacy guarantee [2,4,13]. For example, some works perform well on NIST data, but it is difficult to obtain reasonable privacy parameters on cifar [2].

In 2000, Agrawal and Srikant [2] proposed two methods of privacy protection and data mining. Lindell and Pinkas built an efficient and secure function evaluation protocol for ID3 algorithm. Agrawal and srikkant [2] proposed privacy protection methods to construct a safe and effective function and evaluate the protocol of ID3 algorithm. All parties collect private data during their training and hope to jointly train the decision tree model through these data. In order

to enable the parties to the agreement to calculate this tree without disclosing the personal privacy information they hold. Agrawal and Srikant demonstrated how to understand the probability distribution behind some personal data sets in the presence of interference noise (introduced to maintain privacy).

## 3 Preliminaries

In this section, we will briefly introduce deep learning and its possible attacks. In addition, we have defined various terms and keywords in this section. We summarize the common methods and their implementation to solve the privacy problem of deep learning (especially neural networks) [6].

### 3.1 Deep Learning

Neural network is a branch of artificial intelligence. CNN (convolutional neural network) is a special neural network, which is mainly used in computer vision. In CNN, it receives input in a 2D structure and has a multi-layer structure to create a feature map called sub sampling. Therefore, it has been widely used in image recognition tasks in data preprocessing and deep learning. Today, researchers [11] modify a single pixel during training, or change all pixels to a smaller number, or combine the two methods to attack the model. These attacks have a great impact on the model, and the average confidence is reduced by up to 84% [11]. After reviewing the work of other researchers, we found that they used cifar-10 [2], MNIST [13] or Imagenet datasets, just like Carlini and Wagner [4].

### 3.2 Adversarial Attacks in Deep Learning

This section focuses on adversarial attacks and adversarial examples, which were first proposed by szengendy et al. [2]. They discovered the process by which neural networks were attacked. In addition, they also successfully designed a network to misclassify the output images by applying a certain amount of noise. In all these programs, adding noise to the original image to increase the uncertainty of prediction is a typical attack case. In fact, if such attacks occur (for example, modifying traffic signs that may be misunderstood by autonomous vehicles and lead to accidents), it will seriously affect the application of machine learning, and even lead to serious consequences. The target of the attack is to add noise, so that the subjective model misclassifies the given output. There are three main types of adversary attacks: uncertain target attack. In this attack, it deceives the classifier by modifying the target image, makes the model unable to execute, and gives a random class output different from the real image. Target against attack: [3] in this attack, it incorrectly classifies the input image into a specific target category to modify the target image. The result of this model is only a certain class. These attacks may disguise a face as an administrator user, allowing people without permission to obtain permission.

### 3.3 Differential Privacy

For a randomization algorithm a (the so-called randomization algorithm means that for a specific input, the output of the algorithm is not a fixed value, but follows a certain distribution), the two output distributions obtained by acting on two adjacent data sets are difficult to distinguish. The formal definition of differential privacy is:

$$Pr[A(D) = O] \leqq e^{\epsilon} \cdot Pr[A(D') = O] \tag{1}$$

In differential privacy, researchers propose different noise distribution mechanisms to protect the privacy of data and reasoning. Next, we will introduce some of these mechanisms.

**Theorem 1 (Laplace Mechanism).** *For functions, the Laplace mechanism L of dataset S,*

$$L(X) = f(X) + (Lap(\Delta(f)/e))d \tag{2}$$

**Theorem 2 (Exponential Mechanism).**

*Given data set D and an availability function q(D,r) → R, Privacy protection mechanism M satisfied with $\epsilon$-differential privacy, If and only if the following expression holds:*

$$M(D, q) \propto e^{\frac{\epsilon * q(D,r)}{2\Delta q}} \tag{3}$$

This theorem implies that the Exponential mechanism can make high utility outputs exponentially more likely at a rate particularly depends on the utility score like the final outcome would be approximately optimal with respect to $s$, and meantime give rigorous privacy guarantee. More ever, the composition properties of differential privacy yield the privacy guarantee for a series of composition.

**Theorem 3 (Sequential Composition).** *Let $C_1, C_2, C_r$ be a series of mechanisms and each $C_i$ yields $\epsilon$-differential privacy. Let C be some other mechanism that compute $C_1$ (X), $\cdots$, $C_r(X)$ using R independent randomness for every $C_i$. Then C satisfies $\epsilon$-differential privacy.*

**Theorem 4 (Parallel Composition).** *Let Pi that give ei-differential privacy. A series of Pi ($X_i$)'s over the disjoint datasets $X_i$ yeild max $\epsilon$-differential privacy. These mechanisms help us to distribute the privacy budget among r mechanisms to recognize $\epsilon$-differential privacy.*

In the above paragraph, we briefly discussed differential privacy, and it's typing. As mentioned in the introduction machine learning used in various types of regression algorithms to secure users' data, one of the approach researchers used is [11]. They proposed a method using differential privacy-we [1]. The model limits how much information an adversary can gain about particular private value, by observing a function learned from a database containing values even if she knows every other value in the database. The $\epsilon$ privacy parameters in the proposed algorithm are uniform and have low margin data and inseparable data.

The strengthening of the privacy guarantee corresponds to reducing $\epsilon$ it degrades the learning performance in this case. The majority values are tested by authors in given research [11] the majority of $\epsilon$ tested. The method they introduced is superior in managing the tradeoff between privacy and learning performance. By taking care of using very small $\epsilon$ corresponding to extremely strengthen the privacy requirement. This method performs better and gives predication accuracy close to chance which is not good and useful in machine learning purposes.

## 4 System Architect

In this section, we describe the system architecture, objectives and possible privacy leaks in existing privacy protection technologies (such as differential Privacy), and conduct distributed training on the client. In the initial, the global model request for a parameter from clients to perform training. After receiving the K number of parameters from clients, the global model sends a signal and requests for no more parameters required to be uploaded to maintain the communication cost. After receiving a parameter from the clients it updates the gradients by using a gradient decent algorithm. Then it sends a signal to download the updated model parameters. Clients download the updated global model and train the local model using their respective dataset, during the training we insert the Laplace noise to protect user data. After updating the model the clients upload the differentially private models back to the global model. The global models combine all the updated local models and perform aggregation using the federated average theorem.

Our Differentially-Private neural network uses a Laplace mechanism [5], to guarantee a privacy for participants. It can preserve the privacy of data while building an effective model with high learning accuracy. We are proposing a novel approach known as Differentially Private Neural Network for Privacy-Preserving which can effectively restrict the adversary from accessing the clients' model, and having the ability to restrict the data leakage. In this framework, we used a DP Laplace mechanism. To ensures that all operations, including the server-side aggregation process, are secure and do not reveal any private information about the training data. At the same time, we consider adding multiple noises on preprocessing where client data will be secure during training.

Our main goal is joint learning, and the client can join and exit at any time during training. The proposed algorithm can reduce the performance loss of the model on the premise of using differential privacy, because it is completely independent of epochs. At the same time, the training process for the client also uses differential privacy, but because the training model we use on the client uses element level privacy measures, it leads to low efficiency and insufficient privacy budget. In the federated optimization framework [8], the central server aggregates the client model after each round of communication. We use a random mechanism to change and approximate this average. This is to hide the contribution of a single entity in the aggregation model, so as to hide it in the whole distributed learning process. In our proposed federated optimization framework

[17], the server aggregates the client model parameters after each round, and uses a random mechanism to approximate this aggregation. All processes are accomplished by hiding customer contributions and following the entire distributed learning process.

In our proposed framework, more noise is added to the input features with little correlation with the output, and different activation functions can be applied. It ensures that our framework can be applied to various large data sets of machine learning tasks. The problems to be solved by our proposed framework:

- To Protect user's data on distributed computation.
- To improve the efficiency of learning rate during training, methods that can provide better accuracy.
- To balance tradeoff between privacy budget and model efficiency.

First, our goal is to develop a system in which the server randomly selects updates for each customer and aggregates the updates of all customers to obtain a global model. In our federated learning system, the server aggregates the client model parameters after each round, and uses a random mechanism to approximate the aggregation. The whole process is to hide the customer's contribution and follow the whole federal learning process. Second, we send the private model to each client, where noise is immediately inserted and preprocessed.

We have completed the model training on the client, added more noise to the input features that are less related to the model output, and can apply different activation functions to our framework, which ensures that our framework can be extended to a wide range of different data sets in the deep learning model. Our solution improves the limitations of joint learning privacy problems, which usually have higher computational overhead. In addition, our solution improves privacy and protects the model from various attacks, such as reasoning attack, adversarial attack and the recently proposed back door attack in federated learning.

We train safely in a distributed environment, and our framework will automatically adjust the sensitivity analysis and noise insertion on the depth neural network. It is completely independent of the number of training rounds in the privacy budget. This makes our mechanism more practical. In addition, in distributed deep learning, our method can redistribute noise insertion to enhance the utility of the model.

In Fig. 1 we have shown our method to flow chart, the detail mechanism of our method is given below:

- The server sends the model to random IoT clients.
- Update the model parameter at the IoT client device.
- Apply Laplace noise to the updated model.
- Send the noise model back to the server, which will aggregate using fedavg.

Experiments show that differential privacy at the client level is feasible, and can still provide high accuracy when enough clients are involved.

1. By applying LRP to data d using deep neural network, the average correlation of all input features is obtained, expressed as RJ (d).

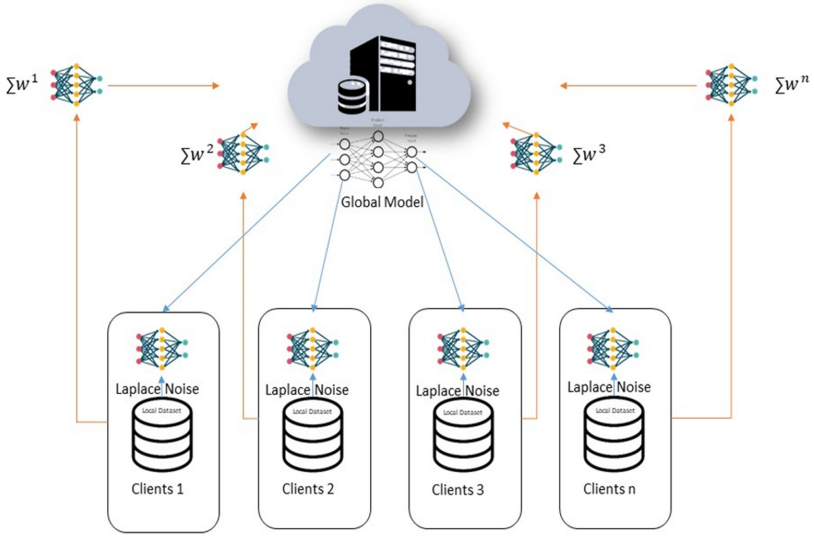


**Fig. 1.** Framework of differentially private federated neural network

2. Laplacian noise is injected into RJ in proportion through different correlation degrees of features
3. After local training, different IOT customers have different weights.
4. The weight of the added noise is calculated using the fedavg algorithm to generate a new weight.
5. The new weight will be transferred to the customer, and the customer will conduct a new round of training and testing.
6. The rounding value can be determined according to the values of efficiency, accuracy, etc.

## 5 Experiments

We have written procedures to evaluate the proposed framework. We used MNIST data sets for testing and divided the sorted MNIST sets into fragments. Each client obtains two fragments; Most customers will only have a two digit sample. Therefore, a single customer cannot train the model according to its data to obtain an available model. The following parameters are cross verified for all $K \in \{100, 200, 500, 1000\}$ scenarios: the number of batches per client. Epochs to run on each client. Number of customers participating in each round. Use Laplacian to insert noise on each client.

Digital classification accuracy of non-IIDMNIST data kept by clients during decentralized training in Fig. 2 and Fig. 3 (Figs. 4, 5, Table 1).

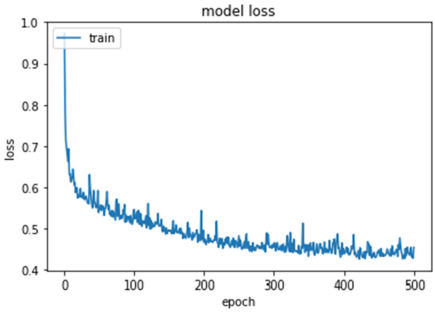


**Fig. 2.** Model loss during training

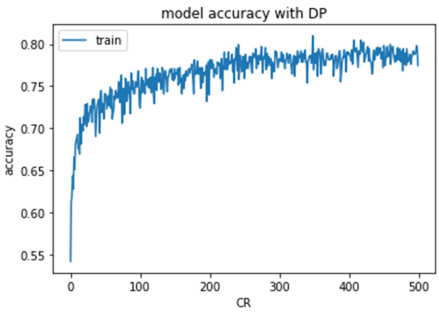


**Fig. 3.** Acc of clients during training

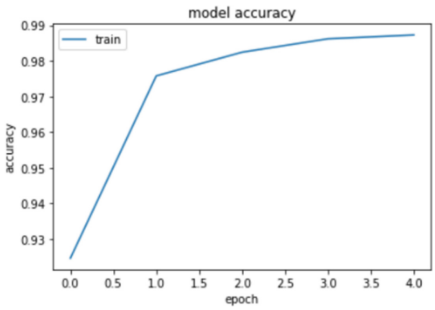


**Fig. 4.** Accuracy of model output with MNIST dataset without decentralization

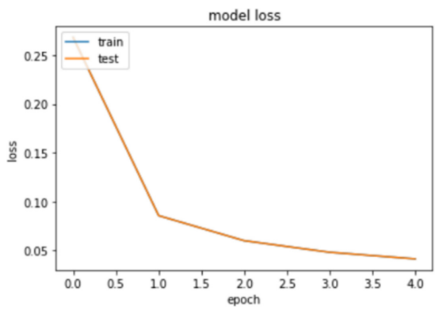


**Fig. 5.** Loss of model output with MNIST dataset without decentralization

**Table 1.** Differentially private federated learning (DP) experimental findings and reference to non-differentially private federated learning (Non-dp).

| | Clients | ACC | CR | CC |
|---|---|---|---|---|
| Without Noise(Non-DP) | 100 | 98 | 500 | 35550 |
| With Noise(DP) | 200 | 80 | 500 | 500 |
| // | 300 | 81 | 350 | 1051 |
| // | 500 | 82 | 350 | 150 |
| // | 1000 | 80 | 350 | |

ACC accuracy
CR number of communication
CC communication costs
DP differential privacy

## 6 Conclusion

By using our proposed framework, the accuracy error of the expected results is reduced to a low level. Experiments show that federated learning of differential privacy at the client level is feasible, and high model accuracy can be achieved when a sufficient number of users are involved. Our framework will provide differential privacy at the user local level, which is feasible and easily scalable for scenarios where many users train models together. Our framework operates deep neural networks for sensitivity analysis and noise insertion. It is completely independent of the number of training in the privacy budget consumption. This makes our mechanism more practical. In the future, we will use blockchain for secure multi-party computation and use secure multi-party computation to make encrypted data more secure. Our model will be more focused on adversarial attacks.

**Acknowledgement.** This work is supported by National Natural Science Foundation of China (No. 61702218, 61672262), China Scholarship Council (No.201808370046), Shandong Provincial Key Research and Development Project (No.2019GGX101028, 2018CXGC0706), Shandong Provincial Natural Science Foundation(No. ZR2019LZH015), Shandong Province Higher Educational Science and Technology Program (No. J18KA349), Project of Independent Cultivated Innovation Team of Jinan City (No. 2018GXRC002).

## References

1. Abadi, M., et al.: Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 308–318 (2016)
2. Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (2000)
3. Bonawitz, K., et al.: Towards federated learning at scale: System design. arXiv preprint arXiv:1902.01046 (2019)
4. Li, P., et al.: Multi-key privacy-preserving deep learning in cloud computing. Futur. Gener. Comput. Syst. **74**, 76–85 (2017)
5. Dwork, C., Roth, A., et al.: The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci. **9**(3–4), 211–407 (2014)
6. Friedman, A., Seminar, T.: Privacy Preserving Data Mining. Springer International Publishing, Heidelberg (2014)
7. Hoekstra, M., Lal, R., Pappachan, P., Phegade, V., Del Cuvillo, J.: Using innovative instructions to create trustworthy software solutions. In: Proceedings of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy (2013)
8. Kamm, L.: Privacy-preserving statistical analysis using secure multi-party computation (2015)
9. Kuflik, T., Kay, J., Kummerfeld, B.: Challenges and solutions of ubiquitous user modeling. In: Krüger, A., Kuflik, T. (eds.) Ubiquitous Display Environments. CT, pp. 7–30. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-27663-7_2

10. Ma, C., et al.: On safeguarding privacy and security in the framework of federated learning. IEEE Network **34**(4), 242–248 (2020)
11. Mcmahan, H.B., Moore, E., Ramage, D., Hampson, S., Arcas, B.: Communication-efficient learning of deep networks from decentralized data (2016)
12. Nikolaenko, V., Weinsberg, U., Ioannidis, S., Joye, M., Boneh, D., Taft, N.: Privacy-preserving ridge regression on hundreds of millions of records. In: Security and Privacy (SP), 2013 IEEE Symposium on (2013)
13. Phan, N., Wu, X., Hu, H., Dou, D.: Adaptive laplace mechanism: differential privacy preservation in deep learning. arXiv (2017)
14. Wang, N., et al.: Collecting and analyzing multidimensional data with local differential privacy. In: 2019 IEEE 35th International Conference on Data Engineering (ICDE), pp. 638–649. IEEE (2019)
15. Wang, S., et al.: Local differential private data aggregation for discrete distribution estimation. IEEE Trans. Parallel Distrib. Syst. **30**(9), 2046–2059 (2019)
16. Wang, Z., Song, M., Zhang, Z., Song, Y., Wang, Q., Qi, H.: Beyond inferring class representatives: user-level privacy leakage from federated learning. In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications, pp. 2512–2520. IEEE (2019)
17. Xu, J.: Edgence: a blockchain-enabled edgecomputing platform for intelligent iot-based dapps. China Commun. **17**(4), 78–87 (2020)