Research Article

An Adaptive Communication-Efficient Federated Learning to Resist Gradient-Based Reconstruction Attacks

Yanbin Li , **Yue Li**, **Huanliang Xu**, and **Shougang Ren**

College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210095, China

Correspondence should be addressed to Shougang Ren; rensng@njau.edu.cn

Received 10 March 2021; Revised 2 April 2021; Accepted 13 April 2021; Published 22 April 2021

Academic Editor: Weizhi Meng

Copyright © 2021 Yanbin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The widely deployed devices in Internet of Things (IoT) have opened up a large amount of IoT data. Recently, federated learning emerges as a promising solution aiming to protect user privacy on IoT devices by training a globally shared model. However, the devices in the complex IoT environments pose great challenge to federate learning, which is vulnerable to gradient-based reconstruction attacks. In this paper, we discuss the relationships between the security of federated learning model and optimization technologies of decreasing communication overhead comprehensively. To promote the efficiency and security, we propose a defence strategy of federated learning which is suitable to resource-constrained IoT devices. The adaptive communication strategy is to adjust the frequency and parameter compression by analysing the training loss to ensure the security of the model. The experiments show the efficiency of our proposed method to decrease communication overhead, while preventing privacy data leakage.

1. Introduction

In recent years, Internet of Things (IoT) has had great popularity in different aspects of modern life and a huge amount of IoT services are emerging. In the IoT area, user devices generate a large amount of data that can be used to improve the user experience of the intelligence system. However, the extensive users' data processing from the IoT device brings some privacy problems [1]. As the IoT devices can be deeply involved in users' private data, the data generated by them will contain privacy-sensitive information [2–4]. To tackle the privacy challenges and encourage clients to proactively participate in IoT services, federated learning enables training a deep learning model across different participants in a collaborative manner. It provides the privacy of clients to keep their original data training on their own devices, while jointly learn a global model by sharing only local parameters with the server.

However, several recent works have shown that the privacy in federated learning is insufficient for protecting the local training data from gradient-based reconstruction attacks [5–7]. The wide malicious devices in IoT make it

vulnerable to these type attacks based on shared parameters. The first type of attack is GAN-based attacks. Hitaj and Perez-Cruz proposed a GAN-based attack against collaborative deep learning on a malicious client, which infers sensitive information from another client successfully [8]. Based on this work, an improved GAN with a multitask discriminator was proposed to enable a malicious server to discriminate category, reality, and client identity of input samples simultaneously [9]. Another type of gradient-based reconstruction attacks is Deep Leakage from Gradients (DLG), which was proposed by Zhu et al. to reveal the training data from gradients [6]. The main idea of DLG is to generate dummy data and labels via matching the dummy gradients to the shared gradients. It has been used in many following works to perform the privacy leakage attacks on federated learning [5, 7]. The GAN-based attack uses GAN to generate pictures that look similar to the training images, while DLG aims at revealing the complete training data from gradients. These two types of attacks both utilize gradient to reconstruction.

To guarantee the privacy of federated learning, there are many privacy techniques for preventing indirect leakage.

Cheng et al. presented a FL-EM-GMM algorithm to make model training without data exchange for protecting privacy [10]. Secure multiparty computation (SMC) involves multiple parties and provides security proof to guarantee complete zero knowledge so that each party knows nothing except its own input and output. It has been used for model training and verification without users revealing sensitive data [11–13]. However, the secure aggregation requires gradients to be integers, which makes it is not compatible with most CNNs. A general method named differential privacy involves adding noise to the training data or obscuring certain sensitive attributes so that the third party could not distinguish the individual information [14–16]. This method usually decreases the accuracy. However, the GAN-based attack is resist against to a certain-level differential privacy [8]. Asad et al. proposed the FedOpt algorithm using homomorphic encryption to protect the privacy of users [17]. But homomorphic encryption increased the model upload time and the system burden, which may increase the system overhead in a bandwidth-limited server system. In addition, they ignore the relationship between efficiency and privacy.

To overcome the performance bottleneck is to apply optimization technologies in federated learning. There are numerous variants of gradient quantization or sparsification and communication delay have been proposed to different distributed deep learning tasks to reduce the communication cost. Han et al. proposed a fairness-aware gradient sparsification method to minimize the overall training time [18]. Zhou et al. proposed a privacy-preserving multidimensional data aggregation scheme, which has great advantages in the communication overhead [19]. In addition, the adaptive communication strategy was adopted to save communication delay and improve convergence speed [20]. There are few works that discuss the impact of these optimization technologies on the security under gradient-based reconstruction attacks. The experiments showed that gradient compression and sparsification could mitigate the leakage of DLG [6, 21].

In this paper, we discuss the security of the federated learning model with different optimization technologies comprehensively. Based on our analysis, the optimization technologies used to reduce communication cost may also improve the resistance against gradient-based reconstruction attacks. To promote the efficiency and security simultaneously, we propose a defence strategy of federated learning without extra high overhead countermeasures, which are not suitable to resource-constrained IoT devices. Our strategy aims at reducing the communication overhead in IoT environment and achieving higher security against gradient-based reconstruction attacks. The experiments on the open source dataset have shown that our method achieves a relative low training loss and prevents from gradient-based reconstruction attacks.

The remainder of the paper is organized as follows. Section 2 describes the optimization technologies to improve the efficiency of FL and the two type gradient-based reconstruction attacks for FL. In Section 3, we discuss the relationship between the optimization and security. Based

on the results, we introduce a new method of FL to improve the efficiency and security simultaneously in Section 4. The experimental results are shown in Section 5. Finally, we provide the conclusion.

2. Background

2.1. Efficiency Optimization of Federated Learning. The surge of massive data has led to significant interest in distributed algorithms for scaling computations in the context of machine learning and optimization. The baseline communication protocol is used in many early federated learning implementations: the client sends a full vector of local training parameter update back to the federated learning server in each round. In this context, the current research is focused on how to reduce the transfer cost of model parameters to make it more efficient in terms of communication, of which the gradient compression and periodic methods are intensively researched.

2.1.1. Gradient Compression. Gradient quantization or sparsification is used to reduce the communication cost through gradient compression. Strom proposed a compression and quantization-based approach to compress single communications and introduced the concept of gradient residuals [22]. Firstly, the participating node k computes the local gradient ΔW_k by adding the local gradient ΔW_k to the previously residual gradient residual ΔW_k^r . If the new gradient is larger than the threshold T , the index and threshold T of that gradient are encapsulated in message M and the gradient residuals are updated: $\Delta W_k^r = \Delta W_k^r - T$. If the new negative gradient is less than threshold $-T$, the index and threshold $-T$ of the gradient are also encapsulated in message M , and the gradient residuals are updated: $\Delta W_k^r = \Delta W_k^r + T$. Finally, the compression message M is sent to other nodes.

Given that the exact selection of the threshold T is difficult in practice, Aji and Heafield proposed a heuristic method for threshold compression [23]. The unique feature of this method is the dynamic selection of thresholds, which reduces the difficulty of threshold selection by setting a discard rate R , sorting the sampled gradient values by absolute values, and taking the number with the $(1 - R\%)$ largest absolute value of the gradient as the current threshold. Tian et al. proposed a novel sketch-based framework (DiffSketch) for distributed learning [24]. The framework can protect privacy using federated learning and compressing the parameters. But some existing attacks can already steal privacy information in federated learning, and only compressing the parameters could not guarantee the data privacy. We would like to seek a balance between compression and communication frequency that protects privacy and ensures accuracy.

2.1.2. Periodic. The communication delay approach is another solution to the above bottleneck, which differs from the gradient sparse and quantitative optimization perspective. The former significantly reduces the number of

communication rounds by increasing the local computational cost appropriately, while the latter is to reduce the cost of communication per round. The two can be complementary, but not contradictory.

Though period-average gradient descent can significantly reduce the number of communication rounds through delayed communication, it also increases the local computational cost, and the appropriate communication frequency is not easy to select. High frequent communication leads to huge communication rounds, but eventually it converges to a smaller loss, while sparse communication reduces the cost of communication, but the results of the federated learning model is worse. Therefore, to solve the above issues, Wang proposed an adaptive communication strategy ADACOMM, which divided the training phase into subphases and tried to find the optimal communication frequency for each phase [20]. Before the start of the new phase, ADACOMM was used to select the frequency by the training loss.

The communication frequency update rule is shown in the following equation:

$$\tau_l = \lceil \sqrt{\frac{F(X_{t=IT_0})}{F(X_{t=0})}} \tau_0 \rceil. \quad (1)$$

As the training proceeds, the loss $F(X_{t=IT_0})$ becomes smaller and the frequency τ_l becomes smaller, i.e., the local computation round becomes smaller and the communication frequency increases gradually. Finally, the model will converge with fewer iterative rounds according to the ADACOMM.

2.2. Gradient-Based Reconstruction Attacks. The original idea of the federated learning was to build global models based on the gradient parameters that are distributed across multiple devices and to prevent data leakage. Potential loopholes are found in some research in the gradients shared by federated learning, which can be divided into two main categories: GAN-based attacks and DLG attacks. The procedure of the two types of attacks is depicted in Figure 1.

2.2.1. GAN-Based Attacks. GAN is proposed to implement a novel class of active inference attacks on deep neural networks in a collaborative setting. Specifically, the generator G attempts to imitate the data from target distribution to make the “fake” data indistinguishable to the adversarial supervisor D . There may be a setup defensible to attacks, which may be achieved by setting stronger privacy guarantees, releasing fewer parameters, or establishing tighter thresholds. However, as proved by the results in this article, such measures lead to models that are unable to learn or worse performance than models trained on centralized data. Therefore, we consider solving the problem from a combination of approaches.

2.2.2. DLG Attacks. Zhu et al. presented an approach which shows the possibility to obtain private training data from the publicly shared gradients [6]. In their Deep Leakage from

Gradient (DLG) method, they synthesized the dummy data and corresponding labels with the supervision of shared gradients. Specifically, they start with random initialization of pseudodata and labels. Virtual gradients are computed on the current shared model in the distributed setup. By minimizing the difference between the virtual gradient and the shared real gradient, they iteratively update the virtual data and labels simultaneously. Although DLG works, we find that it could be affected by a number of factors that affect the quality of the images generated by federated learning efficiency.

2.3. Notation. In order to express with conciseness and standardization, we stipulate the letters' notation of some indicators and show the main hyperparameter settings and notations in Table 1. CE is a communication compression ratio index, which is also one of the most important indexes for evaluating communication efficient algorithms. E_0 is the fixed rounds of updating. $\text{avg}_{\text{parameter}}$ is the mean parameters of all iteration communication. acc_{90} is the accuracy reached 90% for the first time. max_{acc} is the maximum of the accuracy in all iteration.

3. The Relationship between Communication Efficiency and Security

Recent improvements have been focused on communication cost in federated learning. The main approaches are to reduce the communication overhead and improve the overall efficiency of federated learning. The goal can be achieved by reducing the communication frequency and compressing the parameters. This section introduces the evaluation indexes to measure the security threats to federated learning; based on this, we discuss the relationship between communication optimization methods and security under the gradient-based reconstruction attacks in federated learning.

3.1. Evaluation Metrics. Privacy threats in federated learning are mainly recovery training dataset images and image pixels that imply private information about the user, so the image similarity metric can be referred as a security evaluation metric.

Attack success rate (ASR) refers to the percentage of successful attackers recovering the local training data victim. The metrics for determining the success of the attack are different for various attack strategies. In the GAN-based attacks, the accuracy of the recovered image label category shall prevail; however, in the DLG attacks, the similarity between the reconstructed image and the original image can be used as a criterion for success. The attack success rate (ASR) is the percentage of successfully reconstructed training data to the number of attacked training data.

An iterative attack is a situation in which a malicious attacker recovers the original data attacked through multiple iterative rounds. The criteria for determining the success of an attack is the same as ASR.

Structural Similarity (SSIM) is usually used as an index to measure the similarity of two images. SSIM is based on the

FIGURE 1: The gradient-based reconstruction attacks on federated learning in IoT.

TABLE 1: Hyperparameter settings and letter representation.

Notation	Denote
M	The number of all clients
C	The number of compression
f	The number of frequency
CE	Compression rate of single communication
E_0	The fixed rounds of updating
$\text{avg}_{\text{parameter}}$	The mean parameters of all iterations' communication
acc_{90}	The number of iterations when test accuracy is beyond 90 for the first time in all iterations
max_{acc}	The maximum of the accuracy in all iterations

perception model to measure the structural similarity between two images. Due to the outstanding performance of this indicator, it has been widely used in fields such as measuring video quality and image deblurring. Given two images x and y , then SSIM can be expressed as

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (2)$$

where μ_x and μ_y are estimated as the mean intensity, and the luminance comparison function is then a function of μ_x and μ_y . σ_x and σ_y are the unbiased estimate in the discrete form, and the comparison of the two signal is used as the contrast comparison. The constant c_1 and c_2 are to avoid instability when other signals are close to zero.

MSE is a signal fidelity measure. MSE refers to the root mean square deviation. The mean square error function is used to measure the similarity between the attacker's

reconstructed image y and its true value x . Usually, it is assumed that one of the signals is a pristine original, while the other is distorted or contaminated by errors. The data recovered by the attacker is more similar to the real data with smaller MSE. The following formula is usually used to calculate MSE:

$$\text{MSE}(x, y) = \frac{1}{M} \sum_{i=1}^M (x(i) - y(i))^2, \quad (3)$$

where x and y are two finite-length discrete signals (e.g., visual images), where M is the number of signal samples (pixels, if the signals are images) and $x(i)$ and $y(i)$ denote the values of the i th samples in x and y , respectively.

3.2. Relationships between Efficiency and Security. Hitaj et al. proposed the GAN-based attacks, to which the impact of optimization methods has not been researched. Therefore,

this section is to study the effect of communication performance factors on the security of the two types of attacks [8]. We perform the experiments on the MNIST and AIFAR-100, which were used as the validation datasets in DLG work [6]. To defend the DLG attacks, the author experimented to defend by gradient compression. We reproduce some experiments according to the source code given by the authors, and the results show that parametric compression of the recovered images has obvious artifact pixels at 10% compression. This result is better than that described in that article. Firstly, the effect of a change in communication frequency on the security of federated learning is shown under the two types of attacks. Secondly, we discuss the effect of the two types of attacks under different parameter compression rates. The two evaluative metrics, SSIM and MSE, are used to determine the results of the attacks. Finally, we summarize the defensive effects of the two factors affecting the efficiency of federated learning on its security.

3.2.1. Relationships between Frequency and Security. Since the DLG method is a pixel-level reconstruction, the number of categories in the original datasets has no effect on it. The successful attacks of the DLG attacks are influenced by the pixel size of the original image. The GAN is very different. The method based on GAN-based attacks is label-level image reconstruction. The number of categories in the original data set determines the classification effect of the classifier, which in turn affects the classification effect of the discriminator, and ultimately affects the generator generation image quality. Therefore, the DLG attacks' method can achieve better attack effects on both the MNIST and CIFAR datasets, while the GAN-based attacks' method performs worse in the CIFAR100 dataset.

Figure 2 is the experimental result when the number of DLG attacks' iteration rounds is set to 500, and the communication frequency is 1. The image is reconstructed by printing the attack every 10 rounds. It can be seen from the figure that, after about 60 iterations, the original image can be considered successfully attacked.

Figure 3 shows the 36 three-category reconstructed images generated by the GAN method after 500 rounds of attack. The generated image can be clearly recognized as 3 by the human eyes, so this attack can be considered effective. The GAN method can generate false images in batches, which is very efficient in scenarios where the image quality is not high and only the category requirements are required.

Reducing the communication cost is one of the optimization goals of federated learning. The method to change the local communication frequency can alleviate the bottleneck problem caused by communication effectively. On the contrary, the change of communication frequency also caused a change in the security of federated learning. In this series of experiments, we explore the relationship between communication frequency and DLG attacks. The experiment includes 15 groups, with the communication frequency set to different values from 1 to 50 and the learning rate of 0.001. We count two indicators, SSIM and MSE, and visualize the reconstructed image after the attacks. The specific experimental results are shown in Table 2.

From the statistical data in the table, we can clearly see that, as the number of local training rounds increases (the communication frequency decreases), the similarity between the image generated by the statistical reconstruction and the original image becomes smaller and smaller, showing an opposite linear relationship. The experimental results show that, within a certain limit, reducing the communication frequency cannot only reduce the communication cost of federated training but also increase the difficulty of the DLG attacks against other client data attacks, which improves the security of the federated learning system.

Figure 4 shows the initial messy image, the original image, and 15 groups of attacks' reconstruction images under different communication frequencies. The visualization results are consistent with Table 2, and the image quality recovered by the attacker is getting worse. Similarly, we count the experimental results of this method on the MNIST dataset, and the above experimental phenomenon can also be found.

We also count the experimental results of the GAN-based attacks on the MNIST dataset. Since this method is more sensitive to the communication frequency, the experiment only sets 5 different frequency values. From the experimental results in Table 3, we can find that the SSIM value is smaller and the MSE value is larger. That is, the image quality reconstructed by the GAN-based attacks' method is average, but the category information is still there, so the applicability of the two indicators of image similarity becomes weaker here.

Figure 5 is the reconstructed images corresponding to the above settings. The image is tending to get blurred, and its category information is gradually lost. The experimental results are consistent with the experimental results of DLG.

From the above experiments, we can find that changing the communication frequency is one of the key factors affecting the attacker's success in the federated learning environment. We can find that the greater the number of local training cycles, the more difficult the attacks, that is, the more secure the client data during the training of the federated learning system.

3.2.2. Relationships between Compression and Security. Parameter compression (gradient sparseness) is often used in federated learning algorithms to reduce the amount of communication between the client and the parameter server, thereby to improve training efficiency. Previous studies have pointed out that this strategy will also affect the difficulty for potential attackers to recover other client data. In order to further explore the potential relationship between federated learning performance and security, we also set up multiple sets of comparative experiments and make statistics on relevant indicators and visualized reconstructed images.

Table 4 is the statistics of comparative experiments conducted under the same communication frequency and different communication compression ratios. We control different communication compression ratios by setting different thresholds. During the communication process, the client only passes the parameters (gradients) that exceed the

FIGURE 2: The results of DLG attacks on images from CIFAR-100.

FIGURE 3: The reconstructed images generated by the GAN-based attacks after 500 rounds.

TABLE 2: The evaluation metrics of the DLG attacks with the CIFAR100 dataset in different frequencies.

Method	Iterations	Communication frequency	SSIM	MSE
DLG	500	1	0.9985	0.3300
DLG	500	2	0.9961	0.5437
DLG	500	3	0.9892	1.4399
DLG	500	4	0.9815	2.7646
DLG	500	5	0.9731	4.3394
DLG	500	6	0.9631	6.3177
DLG	500	7	0.9521	8.3981
DLG	500	8	0.9427	10.8416
DLG	500	9	0.9325	12.7527
DLG	500	10	0.9215	15.2681
DLG	500	15	0.8824	23.8498
DLG	500	20	0.8427	30.3106
DLG	500	30	0.7856	36.5652
DLG	500	40	0.7440	40.5796
DLG	500	50	0.7170	42.4485

FIGURE 4: The results of the DLG attacks with the CIFAR100 dataset in different frequencies.

TABLE 3: The evaluation metrics of the GAN-based attacks with the MNIST dataset in different frequencies.

Method	Iterations	Communication frequency	SSIM	MSE
GAN	100	1	0.6289	41.9540
GAN	100	2	0.6347	40.3287
GAN	100	3	0.6110	43.9704
GAN	100	4	0.6009	42.4333
GAN	100	5	0.5914	46.7876

threshold to the parameter server for aggregation. The experiment set 6 thresholds of different levels, and the learning rate was uniformly set to 0.001. After 500 rounds of attack iterations, the image similarity index was counted, and the attacks' results were visualized.

Figure 6 shows that parameter compression plays a role in suppressing the GAN-based attacks' mode as well. Since

GAN-based attacks perform more frequent interactions for federated learning, the effect of the image is no longer evident when the parameters are compressed to 90. Although the applicability of the two security metrics is weak, the overall trend in Table 5, and the reconstructed image in Figure 7 can reflect the progressively worsening effect of the attack.

It can be seen from the above experimental results that proper parameter compression can effectively avoid the leakage of local data and also reduce the single communication cost. However, excessive compression will adversely affect the training of the global model. When changing the degree of sparsity, we can see that the attacks still cannot be successful when the compression rate reaches 90%. So we can achieve a balance between compression and security by setting appropriate parameter compression thresholds.

FIGURE 5: The results of the GAN-based attacks with the MNIST dataset in different frequencies.

TABLE 4: The evaluation metrics of the DLG attacks with the CIFAR100 dataset in different compression.

Iterations	CE	SSIM	MSE
500	1	0.9987	0.2948
500	0.98	0.9825	2.7182
500	0.96	0.9158	17.5191
500	0.94	0.8589	28.4093
500	0.92	0.7963	37.3693
500	0.90	0.6201	46.2852

FIGURE 6: The results of the DLG attacks with the CIFAR100 dataset in different compression.

TABLE 5: The evaluation metrics of the GAN-based attacks with the MNIST dataset in different compression.

Method	Iterations	CE	Communication frequency	SSIM	MSE
GAN	500	1	1	0.6248	37.3750
GAN	500	0.97	1	0.6138	40.1095
GAN	500	0.95	1	0.6107	45.3234
GAN	500	0.93	1	0.5861	45.6194
GAN	500	0.90	1	0.6144	39.9441

We can draw the following conclusions. (1) Changing communication frequencies is one of the key factors affecting the success in an attack. The more local training iteration, the more difficult it is to be attacked, i.e., the more secure the client data is in the training process of the

federated learning system; (2) compressing the weights (parameters) cannot only avoid data leakage but also affect the security of federated learning, and the more the parameters and the smaller the compressing rate, the higher the security. Therefore, the communication frequency and

FIGURE 7: The results of the GAN-based attacks with the MNIST dataset in different compression.

parameter compression are two important factors that affect the security of federated learning. If a single value is changed, it will increase the security, but the quality of the federated learning model will be sacrificed.

4. Adaptive Frequency-Compression Federated Learning

In order to improve the security of the federated learning model and reduce the effect on the quality of the global model, we propose an adaptive frequency-compression federated learning (AFC-FL) by adjusting the communication frequency and parameter compression. The weights of the two factors are adjusted to ensure the accuracy of federated learning adaptively, while providing higher security. This calls for AFC-FL to start from a larger frequency and minimal compression and adjust them gradually as the model reaches closer to convergence. Such an adaptive strategy will offer a win-win in system operation by ensuring communication efficiency and security.

4.1. Adaptive Strategy. This approach of AFC-FL is to change the communication frequency and parameter compression rates in each iteration round, according to the loss value in the model during training. However, the fixed iteration rounds are difficult to be determined without prior knowledge. Therefore, we divide the entire training process into multiple identical iteration rounds. At the beginning of each iteration round, we determine the communication frequency based on the difference between this round and the previous. The parametric compression rate is then affected by the communication frequency. The strategy of AFC-FL is to estimate the choice of two factors accurately and to make the federated learning model more efficient and secure. It will be described in details in the following sections.

During the training phase of federated learning, it is difficult to select the accurate communication period. An alternative is proposed to obtain the basic communication period update rule. Based on this rule, we adjust it to our strategy with fixed iteration rounds. The improved rule is as follows:

$$f_l = \lceil \sqrt{\frac{F(X_{e=lE_0})}{F(X_{e=0})}} f_0 \rceil, \quad (4)$$

where E_0 is the fixed rounds of updating, $F(X_{e=lE_0})$ is the objective function values of the l th update, and $F(X_{e=0})$ is the initial loss value. The frequency f_l of the next round is guided by the training loss value. When the loss $F(X_{e=lE_0})$ becomes smaller, the frequency f_l decreases, i.e., local computation rounds are fewer, and the communication frequency gradually increases.

It can be concluded that both communication frequency and parameter compression have inhibitory effects on the accuracy and security of federated learning. Although the lower frequency and compression can achieve higher resistance against gradient-based reconstruction attacks, the accuracy will be decreased seriously. There is a need for an adaptive strategy to trade off the accuracy and security. Through the results in Section 2.1, we found that there is a linear relationship between frequency and loss and between compression and loss under a certain constraint: $f \approx k * \text{Loss}$ and $C \approx q * \text{Loss}$, where $0 < f < 5$, $90 < C < 100$.

Therefore, we consider whether we can find a balance between frequency, compression, and loss so that the algorithm can guarantee both the accuracy and the security of the model under the joint influence of compression and frequency. We try to find the appropriate Loss to make our algorithm achieve the most effect by giving different values of the Loss interval. We analyze by the following assumptions:

$$\text{Loss} = \frac{1}{k} * f + \frac{1}{q} * C, \quad (5)$$

where Loss is the set constant and k and q are the two influencing factors. The purpose of the formula is to make the obtained communication frequency to influence the parameter compression rate so that the two inhibiting factors do not overlap each other to achieve the effect of adaptive parameter change. Therefore, the formula is organized as follows:

$$C_l = C_0 - \lfloor f_l * D \rfloor, \quad (6)$$

where C_0 is the initial parametric compression rate, C_l is the parametric compression rate after the l th update, and D is the constant used to control the rate of decline. It is found that a low parameter compression rate makes federated learning worse, so we set a minimum threshold for the parameter compression rate ($C_{\min}$). The improved formula is as follows:

$$C_l = \max(C_{\min}, C_0 - \lfloor f_l * D \rfloor). \quad (7)$$

It has been shown that the communication frequency f_l and the parameter compression C can be adjusted mutually when the appropriate parameter D is set, which affects the completion of the federated learning training and makes the attacks fail.

4.2. Adaptive Communication-Efficient Federated Learning (AFC-FL). To improve the security of the system, we combine multiple influences into a federated learning model, where communication frequency and parameter compression jointly affect the security of the model. Through experimental analysis and research, we propose a method for improving the security of the system, AFC-FL. AFC-FL is comprised of one adaptive frequency model and adaptive compression model. The adaptive frequency model is used to change the frequency by model loss. And, the adaptive compression model is designed to change the parameter compression value by changing the frequency. In the following, we present the network architecture and then analyze the procedure of the distributed optimization.

An overview of the proposed architecture is shown in Figure 8. There are N clients and a central server. The central server aggregates the parameters uploaded by each client. Each client updates parameters according to our proposed AFC-FL.

Algorithm 1 describes the execution process of the AFC-FL algorithm. The initial parameters include the number of clients M , training epochs E_t , updating epochs E_u , optimization function learning rate η , batch size B , and the optimization function is Adam. AdaptFreq is the function of change of the frequency. It can be expressed by equation (4). The function AdaptComp, which can be expressed by equation (7), is to change the compression. Comp can decrease the number of parameters by compression. The algorithm is divided into two parts, client side and server side. The server side is responsible for controlling the global model generation, while the client side performs the adaptive algorithm updates and the local model uploads.

The execution process on the server side is (1) initialize the model $W_{g,0}$; (2) at round i , collect the sparse parameters $W_{m,i}$ uploaded by m clients and find the next round of global model $W_{g,i+1}$ by means of mean aggregation; (3) send the new round of the global model down to each client.

The execution process of the clients is (1) first download the global model $W_{g,i}$ sent by the parameter server; (2) determine whether it is an update interval before each iteration, and if so, perform the update function to update the communication frequency f and the parameter compression ratio C ; (3) then, train each client node locally according to the new communication frequency; (4) obtain the locally compressed model $W_{m,i}$ according to the parameter compression ratio C and by compressing the parameters of the locally trained model ω ; (5) upload the model $W_{m,i}$ to the server side.

In Algorithm 1, lines 8–12 execute the AFC-FL algorithm after a certain number of rounds through the code, adjusting

the communication frequency of the local model as well as the parameter compression rate after the training is completed. When the communication frequency is higher, the more the parameters of the model trained by each client change, the less effective the attacker's attack will be. At this time, 15 lines of parameter compression will not need too much compression to ensure the accuracy of the model training. When the training reaches the late convergence, the communication frequency increases to correct the accuracy and reduce the model upload parameters. Our parameter compression and communication frequency change are calculated on the client side to ensure that the local model parameters are trimmed before uploading. Meanwhile, it avoids joint attacks by the server and the attacker on the client to ensure the security of the system.

5. Experiment Results

To verify the efficiency and security of AFC-FL, we perform experiments using MNIST datasets. In principle, however, AFC-FL can be extended to other types of data, such as medical records. We first show the advantages of our approach by comparing the experiments in Section 3.2; secondly, we perform the GAN-based attacks' experiment to compare the effect of recovered images after the attacks, and we judge the success of our approach by observing the imaging characteristics of the images artificially, combined with the accuracy of the final model.

5.1. Experiment Setup. We mimic the ideas provided by the authors of the GAN article and use Tensorflow to implement the attacks in the privacy scenario of a federated learning client. And, we set up the adaptive frequency parameter compression scheme to further extend in terms of efficiency and security.

5.1.1. Platform. All experiments are completed in the same experimental environment, including Intel (R) Xeon (R) CPU E5-2620 v4 @2.1 GHz, Nvidia 1080Ti GPU (11 GB) *2, and 32 GB RAM. Due to the limitation of experimental conditions, the uploading and downloading of shared parameters in the iteration process of the federated model are implemented by the same machine simulation. Obviously, the statistical indicators of the experimental results have nothing to do with the communication method, so the evaluation is still accurate and effective.

5.1.2. Dataset. The MNIST dataset is stored in bytes. The training set contains 60,000 0–9 digital pixel samples and labels, and the test set contains 10,000 0–9 digital pixel samples and labels. Each image is composed of 32×32 pixels. This dataset is one of the deep learning benchmark datasets. For each client in the experiment, we use a non-IID approach, i.e., each client has only one class of images.

5.1.3. Model. We choose to use the architecture of the classical convolutional neural network LeNet5, which is the

FIGURE 8: A conceptual scheme for the AFC-FL.

FIGURE 9: The loss of the experimental results.

FIGURE 10: The accuracy of the experimental results.

FIGURE 11: The amount of parameters during training.

basis of many networks such as AlexNet, VGGNet, and ResNet, and it is general with great effect. LeNet5 has seven layers, namely, C1 convolutional layer, S2 pooling layer, C3 convolutional layer, S4 pooling layer, C5 convolutional layer, F6 fully connected layer, and output fully connected layer. Each layer contains trainable parameters; each layer has multiple Feature Maps, and each feature map extracts one feature of the input through a kind of convolutional filter.

5.1.4. Hyperparameter Choice. We choose to use ADAM as optimization algorithms and the batch size of 50; for setting a stable learning rate, we conduct some experiments. In the end, we set the learning rate to 0.01.

5.1.5. Metrics. We compare the performance of proposed AFC-FL with the following methods at a fixed frequency or compression period. (1) Baseline: fully compression and one communication iteration; (2) manually adjust the frequency, i.e., using the same frequency for each iteration and for multiple experimental comparisons; (3) manually tuned the compression case where compression is changed by frequency before new training epochs. We train all methods for a long time to convergence and compare the results of 500 iterations.

5.2. Efficiency Experiment. In our experiments, we evaluate the results using the different learning rates or batch sizes. We find that higher learning rates make the federated


```

Input: The number of clients  $M$ ;
Input: Training epochs  $E_t$ ;
Input: Updating epochs  $E_u$ ;
Input: Learning rate  $\eta$ ;
Input: Local mini-batch size  $B$ ;
Input: Local optimization function Adam;
Output: A global model  $W_g$ ;
(1) Procedure ServerExecute:
(2)   initialize  $W_{g,0}$ ;
(3)   for each Iteration  $i \in [1, E_t]$  do
(4)      $W_{g,i+1} \leftarrow 1/M \sum_{m=1}^M W_{m,i}$ ;
(5)   end for
(6) end procedure
(7) Procedure ClientUpdate ( $m, W_{g,i}$ ):
(8)   for each Iteration  $i \in [1, E_t]$  do
(9)     if  $i \% E_u == 0$  then
(10)       $f \leftarrow \text{AdaptFreq}(L): f = \lceil \sqrt{F(X_{e=iE_u})/F(X_{e=0})} f_0 \rceil$ 
(11)       $C \leftarrow \text{AdaptComp}(f): C = \max(C_{\min}, C_0 - \lfloor f * D \rfloor)$ 
(12)    end if
(13)    for each Client  $m \in [1, M]$  do
(14)       $\omega \leftarrow \text{Adam}(W_{g,i}, f, \eta)$ 
(15)       $W_{m,i} \leftarrow \text{Comp}(\omega, C)$ 
(16)    end for
(17)  end for
(18) end procedure

```

ALGORITHM 1: An adaptive communication-efficient federated learning (AFC-FL).

TABLE 6: The comparative experiment results.

Function	avg. _{parameter}	acc ₉₀ (epochs)	The parameters of acc ₉₀	max _{acc}
AFC-FL	570,434	86	5.265×10^7	0.9597
$F = 1$, CE = 100	620,060	98	6.076×10^7	0.9738
$F = 1$, CE = 90	558,042	329	1.836×10^8	0.9166
$F = 5$, CE = 100	620,060	221	1.370×10^8	0.9622
$F = 5$, CE = 90	558,039	459	2.561×10^8	0.9092

learning model unstable and suffer from model oscillations, and smaller batch sizes result in worse convergence of the model. After analysis of several results, we set a learning rate of 0.01 and a batch size of 50 as the hyperparameter. Meanwhile, we also conduct several experiments on the variation range of the adaptive frequency and parameter compression rate. It shows that the frequency is greater than 5, and the parameter information uploaded by each client node is more vague, which makes the federated learning less effective; when the parameter compression rate is lower than 90, the loss of critical information of the local model will also lead to the reduction of the quality of the federated learning model. Therefore, we set thresholds to control the frequency and parameter compression range when using AFC-FL.

We compare the results between the AFC-FL function and the threshold set at a critical value after the communications of 500 epochs. From Figure 9, we can find that our method converges nearly as fast as the comparative experiments' scheme. In Figure 10, the accuracy of our method is better than the experiment that the frequency is 5 and no matter whether it has parameter

compression or not. Although the experimental accuracy which with a frequency of 1 and no parameter compression is slightly higher than ours, the cost of communication is much higher than ours, and our method provides more security.

We use the cost of communication as an evaluation criterion, i.e., the number of uploader parameters in the same communication round determines the upload time of the local model to the server, which affects the efficiency of the global model. The amount of client traffic handled by the server in the same round can be used to represent the throughput of the federated training system. The user participation will be low in a bandwidth-constrained communication environment. Our algorithms make the system to allow more users to participate in training at the same time by reducing the number of uploaded parameters, which improves the throughput. In order to succinctly compare the cost of communication, we use the average number of parameters uploaded in each communication round as the evaluation criterion. The number of parameters uploaded at each epoch is

FIGURE 12: The reconstruction results of GAN-based attack on AFC-FL.

shown in Figure 11. We record the epoch number when the accuracy of model achieving 90%. In addition, we also focus the accuracy after 500 rounds.

From Table 6, we can find that AFC-FL uses the fewest epochs to achieve 90% accuracy for the first time, where $\text{avg}_{\text{parameter}}$ is the mean parameters of 500 epochs' communication, acc_{90} is the accuracy reached 90% for the first time, the parameters of acc_{90} are the total parameters of the accuracy reached 90% for the first time, and max_{acc} is the maximum of the accuracy in 500 epochs. We verify that the compression and frequency can impact the global model to achieve high accuracy in Section 3. Thus, the results of comparative experiments prove the superiority of our method. Since the model also has the compression to ensure the security of the federated learning model, the accuracy is slightly reduced, but the small reduction in

accuracy is acceptable in exchange for the improvement in communication efficiency and the security of the whole system.

5.3. Security Experiment. We now evaluate the security of our AFC-FL against GAN-based attacks. We partition the MNIST dataset into 10 clients by numbers 0–9, with each client having only one of the numerical datasets. We preprocess each client before the formal iteration to avoid any failure to converge due to obscure model features. We use LeNet5 as the generator (G) and the model training network, and we perform the training using Algorithm 1. We observe the generative effect in Figure 12, showing the reconstruction results of every five rounds of attacks during 500 iteration rounds. We can find that the picture cannot be

FIGURE 13: The reconstructed images generated by the GAN-based attacks on AFC-FL.

reconstructed in most cases. It is also unclear which numbers are actually identified. Figure 13 shows the reconstructed image generated by our method after performing 500 rounds of attacks. The generated image is not recognizable to the human eye, and it can be assumed that our method is effective.

6. Conclusions

In this paper, we propose a federated learning optimization algorithm (AFC-FL) with adaptive frequency and compression selection in IoT. The sparsification or communication delay technique significantly reduces the communication cost for clients, improving the security during gradient transmission. Meanwhile, the adaptive strategy also decreases the communication costs of clients. Verified analysis of the algorithms and experimental results using MNIST datasets conclude that AFC-FL is effective in resisting gradient-based reconstruction attacks. Extensive experiments are conducted to verify the

effects of resisting attacks and communication time of our algorithm compared to fixed frequency or fixed compression. Experimental results show that AFC-FL not only significantly reduces the communication traffic but also keeps the client data safe to increase the security of the federated learning model, while preserving the convergence. Future works can also consider asynchronous collection of the client parameter, as well as the selection of different update strategies for each client depending on the size of the parameter. It is the goal of our future research studies to ensure security, while speeding up the convergence rate of the model. In addition, we may also consider improvements in homomorphic encryption and differential privacy. How to improve the efficiency using a low additional overhead is also important to research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant no. 62072247.

References

- [1] F. Mármol, C. Sorge, O. Ugus, and G. Pérez, "Do not snoop my habits: preserving privacy in the smart grid," *IEEE Communications Magazine*, vol. 50, no. 5, pp. 166–172, 2012.
- [2] Z. Tian, C. Luo, J. Qiu et al., "A distributed deep learning system for web attack detection on edge devices," *IEEE Transactions on Industrial Informatics*, vol. 16, pp. 1963–1971, 2019.
- [3] Z. Tian, W. Shi, Y. Wang et al., "Real-time lateral movement detection based on evidence reasoning network for edge computing environment," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4285–4294, 2019.
- [4] Z. Guan, Y. Zhang, L. Zhu, L. Wu, and S. Yu, "Effect: an efficient flexible privacy-preserving data aggregation scheme with authentication in smart grid," *Science China Information Sciences*, vol. 62, Article ID 32103, 2019.
- [5] J. Geiping, H. Bauermeister, H. Drge et al., "Inverting gradients—how easy is it to break privacy in federated learning?," 2020, <http://arxiv.org/abs/2003.14053>.
- [6] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," *Advances in Neural Information Processing Systems*, pp. 14774–14784, 2019.
- [7] B. Zhao, K. R. Mopuri, and H. Bilen, "IDLG: Improved deep leakage from gradients," 2020, <http://arxiv.org/abs/2001.02610>.
- [8] B. Hitaj and G. A. F. Perez-Cruz, "Deep models under the gan: information leakage from collaborative deep learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 603–618, Dallas, TX, USA, October 2017.
- [9] Z. Wang, M. Song, Z. Zhang et al., "Beyond inferring class representatives: user-level privacy leakage from federated learning," in *Proceedings of the 2019-IEEE Conference on Computer Communications IEEE INFOCOM*, pp. 2512–2520, Paris, France, 2019.
- [10] W. Cheng, W. Ou, X. Yin et al., "A privacy-protection model for patients," *Security and Communication Networks*, vol. 2020, Article ID 6647562, 12 pages, 2020.
- [11] O. Goldreich, "Secure multi-party computation," *Manuscript Preliminary Version*, vol. 78, 1998.
- [12] B. Ghazi, R. Pagh, and A. Velingker, "Scalable and differentially private distributed aggregation in the shuffled model," 2019, <http://arxiv.org/abs/1906.08320>.
- [13] K. Bonawitz, V. Ivanov, B. Kreuter et al., "Practical secure aggregation for federated learning on user-held data," 2016, <http://arxiv.org/abs/1611.04482>.
- [14] M. A.A. Chu et al., "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318, New York, NY, USA, October 2016.
- [15] Q. Li, Z. Wu, Z. Wen, and B. He, "Privacy-preserving gradient boosting decision trees," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 1, pp. 784–791, 2020.
- [16] S. Song, K. Chaudhuri, and A. D. Sarwate, "Stochastic gradient descent with differentially private updates," in *Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing*, pp. 245–248, Austin, TX, USA, 2013.
- [17] M. Asad, A. Moustafa, and T. Ito, "FedOpt: towards communication efficiency and privacy preservation in federated learning," *Applied Sciences*, vol. 10, no. 8, pp. 1–17, 2020.
- [18] P. Han, S. Wang, and K. K. Leung, "Adaptive gradient sparsification for efficient federated learning: an online learning approach," 2020, <http://arxiv.org/abs/2001.04756>.
- [19] Y. Zhou, X. Chen, and M. Chen, "Privacy-preserving multidimensional data aggregation scheme for smart grid," *Security and Communication Networks*, vol. 2020, Article ID 8845959, 14 pages, 2020.
- [20] J. Wang and G. Joshi, "Adaptive communication strategies to achieve the best error-runtime trade-off in local-update SGD," *Proceedings of Machine Learning and Systems*, vol. 1, pp. 212–229, 2019.
- [21] W. Wei, L. Liu, M. Loper et al., "A framework for evaluating gradient leakage attacks in federated learning," 2020, <http://arxiv.org/abs/2004.10397>.
- [22] N. Strom, "Scalable distributed DNN training using commodity GPU cloud computing," in *Proceedings of the Sixteenth Annual Conference of the International Speech Communication Association*, Dresden, Germany, September 2015.
- [23] A. F. Aji and K. Heafield, "Sparse communication for distributed gradient descent," 2017, <http://arxiv.org/abs/1704.05021>.
- [24] L. Tian, Z. Liu, V. Sekar et al., "Privacy for free: communication-efficient learning with differential privacy using sketches," 2019, <http://arxiv.org/abs/1911.00972>.