Federated Learning Based Distributed Algorithms for RF Fingerprinting Extraction and Identification of IoT Devices

Weiwei Wu[1], Su Hu[1(✉)], Yuan Gao[2], and Jiang Cao[2]

[1] University of Electronic Science and Technology of China, Chengdu, China
husu@uestc.edu.cn
[2] Academy of Military Science of the PLA, Beijing, China

Abstract. With the development of Internet of things (IoT), exponential data growth and diversified functions and services have dramatically enhanced the importance of user authentication for data access. As a solution to the problem of user authentication, we study the deep-learning based methods for the radio frequency (RF) fingerprinting recognition of mobile devices in this paper. In consideration of the distributed storage of RF signals in practice, instead of using the deep learning algorithms for centralized data training, we employ the federated learning algorithms for distributed RF fingerprinting recognition, where the data of RF signals are distributed in multiple mobile devices for storage and recognition. To reduce the impact of uneven data distribution among mobile devices on the performance of federated learning algorithms, we propose the dynamic sample selection based federated learning algorithms to train the data. In comparison with the traditional federated learning algorithms, our proposed algorithm can improve the system accuracy as well as reduce the computation time.

Keywords: RF fingerprinting · Federated learning · IoT devices

1 Introduction

The amount of data has dramatically increased in the era of Internet of Things (IoT) with the development of sensors, mobile devices, etc. By the forecast of CISCO company, more than 25 billion devices will be connected in the communication systems by 2021 [1]. Because of their inherent nature, mobile devices are vulnerable to malicious attacks when these devices are under untrusted environments. When a mobile device is required to be authenticated, a few traditional authentication methods, e.g., digital signature are not applicable since they are vulnerable to various key-hacking attacks, e.g., invasive attacks, side channel attacks, etc. Also a few mobile devices cannot support high-complexity computation, and thus the traditional authentication methods by using IP or MAC addresses for authentication are not efficient [2].

As a promising authentication method, radio frequency (RF) fingerprinting can identify the unique features of a mobile device by analyzing its transmitted signals, and thus

X. Wang et al. (Eds.): AICON 2021, LNICST 396, pp. 3–18, 2021.
https://doi.org/10.1007/978-3-030-90196-7_1

effectively prevent the impersonation of a mobile device for security credentials. The features of mobile devices which can be extracted from transmitted signals are primarily caused by the difference of RF components in the process of manufacturing, including the imperfections of power amplifiers, the errors of magnitude and phase, carrier frequency differences, phase offset, and clock offset, etc. [3].

Based on the original signals and the extracted features, we can use either machine-learning based algorithms or deep-learning based algorithms to identify mobile devices [4]. With machine-learning based RF fingerprint technologies, we preprocess the collected signals by denoising and normalization, extract various fingerprint features, and identify the mobile devices by analyzing the fingerprint features. In comparison, the deep-learning based radio frequency fingerprint recognition technology does not require the extraction of features from the signals. After the preprocessing, the predicted label is directly compared with the label registered in the fingerprint library in the deep neural network to identify different communication devices. Deep learning method relies on a large amount of data to train a model with existing fingerprints.

Either machine-learning based algorithms or deep-learning based algorithms need to train a large amount of data on the original signals as well as the extracted features. However, the traditional data processing method which needs to load all the data into a centralized node is no longer applicable. Instead, we need to use distributed machine-learning or deep-learning algorithms, e.g., federated learning algorithms. Under the federated learning based distribution architecture, we can train the data in separate nodes instead of performing centralized computations with an aggregation model.

In this paper, we discuss the design of federated learning based RF fingerprinting extraction and identify mobile devices under the distribution architecture. In the following, we overview the RF Fingerprinting and its applications in identifying mobile devices. We then present the design of federated learning based RF Fingerprinting extraction and device identification algorithms. Finally, we discuss a few comparative study results and conclude the paper.

2 Related Work of RF Fingerprinting Recognition

Based on the types of signals, the studies on the RF fingerprinting recognition can be classified as transient signal recognition and steady-state signal recognition, and the primary studies are summarized in Table 1.

A. RF fingerprinting recognition methods for transient signals.

The relevant literature mainly studies on transient signals when the communication process starts or finishes. A transmitter emits transient signals in an unstable working state in which the transmit power fluctuates between approaching zero and rated power when turning on or down transmit power [5]. Transient signal only shows the hardware characteristics of a transmitter, and does not carry any data information. The RF fingerprinting recognition based on the transient signals is independent from data, so it is one of the most commonly studied method in the field of RF fingerprinting recognition Fig. 1.

Table 1. Studies of RF fingerprinting recognition

	Authors	Type of communications	Methods	Key features
Transient signals	Bihl [5]	VHF FM	Wavelet	Wavelet coefficients
	Zhuo [6]	Wavelet	Wavelet	Statistical and power density characteristics
	Li [7]	VHF FM	Time-frequency analysis	Multi-segment fractal dimension
	Xiao [8]	VHF FM	Time-frequency analysis	Complex envelope, instantaneous amplitude, instantaneous phase and instantaneous frequency
	Shi [9]	VHF FM	Time-frequency analysis	Amplitude and phase characteristics
	Polak [10]	IEEE 802.11b	Wavelet	Transient signal amplitude, phase, in-phase component, quadrature component, power and discrete wavelet transform (DWT) coefficients, etc.
	Wang [11]	433 MHz	Wavelet	Signal duration, normalized amplitude variance, peak number of carrier signal, discrete wavelet transform coefficients of the first signal extracted, and the difference between the average and maximum normalized amplitude
Steady-state signals	Demers [12]	UMTS	FFT	Preamble spectrum

(*continued*)

Table 1. (*continued*)

	Authors	Type of communications	Methods	Key features
	Patel [13]	IEEE 802.11a/g	FFT	Power spectral density of preambles
	Reising [14]	IEEE 802.11b	FFT	Frequency offset, preamble correlation, I/Q offset, amplitude error and phase errors
	Knox [15]	MIMO	FFT	Error vector magnitude, carrier center frequency deviation, OFDM pilot phase deviation, symbol clock deviation, I/Q offset, I/Q phase rotation, I/Q gain imbalance, and preamble correlation
	Yuan [16]	IEEE 802.15.4	FFT	Phase information of the demodulated baseband signal

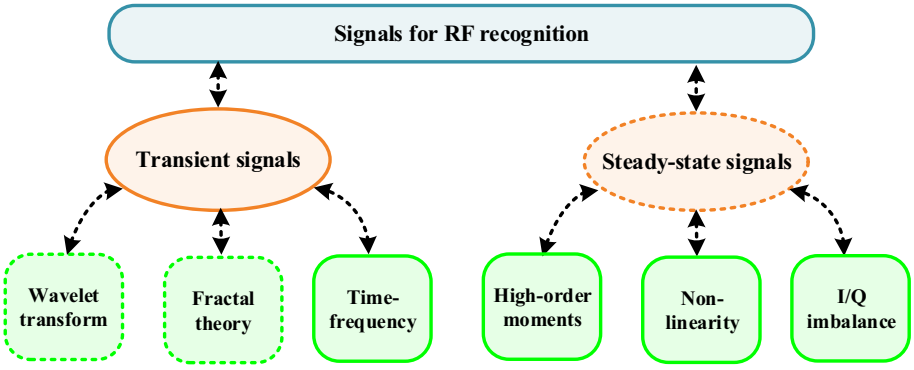


Fig. 1. Signal characteristics for RF fingerprinting analysis

Due to the short duration of a transient signal, it is necessary to determine the starting point of the transient signal in order to obtain the information from the signal. Typical methods of detecting the starting points primarily rely on the comparison of transient signals and noises, and the specific methods include amplitude threshold detection, Bayesian step-start detection, Bayesian rising-point detection, as well as variance estimation detection [6]. Basic principles primarily include Wavelet transform, fractal theory, and time-frequency analysis.

(1) Research on the wavelet transform.

By analyzing transient signals, Li et al. propose a self-recognition method, and it builds neural network models on the non-cooperative radar signals by using wavelet transform features [7]. Aimed at the wireless signals intercepted by a receiver, Li et al. extract the wavelet features of transient signals, and propose the time-frequency information from the signals through wavelet decomposition to identify different signal sources. However, the result shows extremely high signal-to-noise ratio. Through multi-scale and multi-resolution analysis of the signals in the actual environment, Li et al. [7] can achieve a fairly high resolution in the time-frequency domain and high accuracy of classifications. Wavelet analysis reflects the individual characteristics of radiation sources by increasing the number of decomposition layers, at the high cost of time computation.

(2) Research on the fractal theory.

Fractal geometry is a new and emerging geometry theory recently in the field of complex graphs. Xiao et al. [8] study the extraction of multiple dimensions from wireless signals as RF fingerprint features to identify different transmitters, achieving a fairly good recognition performance. Fractal features lack a description of time-varying characteristics of transient signals, and thus the signals with significantly different instantaneous characteristics may have the same fractal complexity. This situation may result in partial misjudgment of RF fingerprinting recognition and reduce the accuracy of recognition. At the same time, due to the different computation methods in different fractal dimensions, we cannot achieve a consistent recognition rate. Xiao et al. [8] analyze the recognition performance when various values of fractal dimensions are used as feature vectors. The testing results show that the recognition performance of the methods with multiple dimensions outperforms the other methods.

(3) Research on the time-frequency analysis.

In the time-frequency analysis of transient signals, Hilbert transform is always used to extract the instantaneous amplitude and phase from the signals. A nonlinear model for transient signal transmission and reception is established, and the characteristics of signal amplitude and phase distortion are used to identify different transmitters [9]. However, in practical applications, the complexity of nonlinear model is extremely high and thus it is difficult to establish with a nonlinear model. Bradford W. Polak et al. construct a kernel function to improve the classification performance, apply time-frequency analysis into the recognition of transmitters, and achieve fairly good recognition performance [10].

But it is worth noting that the RF fingerprinting recognition technology based on transient signals requires extremely high accuracy of the recognition instrument,

and the signal energy is quite weak at the receive end [11]. Intercepting the transient signal from the received signal is also the difficulty of transient-signal recognition technology. RF fingerprinting features of transient signals include signal duration, transient spectrum, etc. The transient signal duration is extremely short, and the channel environment (e.g., noise, temperature, etc.) has a great impact on the transient signals. All the above-mentioned factors may influence the performance in practice.

B. RF fingerprinting recognition methods for steady-state signals.

In addition to a few transient signals, most wireless signals are steady-state signals. In the steady-state part of a signal, the transmitter is typically stable throughout the communication process, and the information part is easy to be separated from the received signal. Therefore, researchers start to study the RF fingerprint recognition on the steady-state signals.

(1) Research on high-order moments and high-order spectra.

A high-order moment feature recognition method is designed on envelope features, but its recognition rate cannot meet the actual demand [12]. Specifically, the recognition method uses the power spectral density of a preamble sequence as feature vectors to identify different transmitters. However, the default signals of conventional second-order cumulants follow Gaussian distribution, which is always assumed in practical applications. Patel et al. in [13] propose a method of computing the rectangular integral bispectrum of a signal, and it can achieve fairly good recognition performance. With higher-order spectral features, Patel et al. can achieve high performance by extracting the RF fingerprinting features from non-Gaussian signals.

(2) Research on the non-linearity of devices.

Due to the appearance of a large number of non-linear devices such as power amplifiers, researchers employ non-linear characteristics for RF fingerprinting recognition [14]. Specifically, a method of nonlinear dynamics is proposed to spatially reconstruct the received signal, and the results show that spatial reconstruction has a good identification performance for weakly nonlinear devices. Also fractal geometry is used to characterize the nonlinearity of a steady-state signal, and the results show that different transmitters have different fractal characteristics [15].

(3) Research on the imbalance of I/Q signals.

Researchers study the use of frequency estimation to identify different transmitters with instantaneous frequency as the characteristic parameter [16]. By extracting the signal's frequency offset, phase offset, IQ offset, and preamble-related modulation errors, the proposed method can identify different transmitters, and the experimental results show that the proposed method can achieve high anti-noise capability and strong robustness. The study uses IQ imbalance to identify the relay system, and the results show that this method has greatly improved the recognition performance [16].

3 Deep Learning Based RF Fingerprinting Recognition Algorithms

In the process of recognition, we can classify the methods of extracting RF fingerprinting features into two types (see Fig. 2). The former is a traditional machine-learning recognition technology, while the latter is a deep-learning recognition technology.

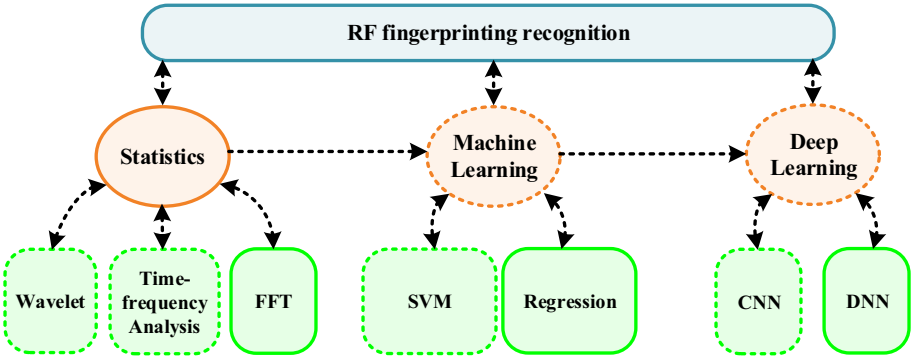


Fig. 2. Types of algorithms for RF fingerprinting recognition

Before 2018, the research on RF fingerprinting focuses on using machine learning algorithms, e.g., support vector machines (SVM) to recognize the identity document (ID) of each mobile device. A few studies employ multiple SVM algorithms to recognize the ID of mobile devices, including the Platt's Minimization Optimization (SMO) algorithm, the PolyKernel algorithms, the Pearson VII Universal Kernel (PuK) algorithms [17]. Typically, the PuK algorithms are more effective in RF Fingerprinting recognition, attaining high performance in recognition and dramatically reducing the computation time. The traditional machine-learning RF fingerprinting recognition technology first preprocesses the collected signals, including power normalization, noise reduction, and label setting. Then, with different algorithms to extract RF fingerprinting features from the pre-processed signal data, we can store the marked features into the fingerprint library. Finally, the extracted RF fingerprint features are used to identify the mobile devices. In the traditional machine-learning recognition technology, the key is to select appropriate signal characteristics from RF fingerprints.

After 2018, the study of using deep learning in the field of RF fingerprinting recognition has gradually appeared. Sankhe et al. in [18] propose to use a 2-layered convolution neural networks (CNN) to train the RF fingerprinting data from 16 of X310 USRP SDRs. Wu et al. in [19] employ a deep neural network (DNN) with rectified linear units (ReLU) to run the training model for the RF fingerprinting recognition of 12 Ettus USRP N210. Also Wu et al. in [19] propose an incremental learning based neural networks to train the data in multiple stages and modify the learning model with new-arrival data to accelerate the process of training. Compared with traditional machine learning algorithms, deep-learning based radio frequency fingerprinting recognition technology does not require the procedure of feature extraction. After preprocessing, we directly compare the predicted labels with the labels registered in the fingerprint library in a deep learning

network to identify different mobile devices. Deep learning relies on a large number of data to train the models with existing RF fingerprints.

As the most successful artificial intelligence method in the field of computer vision, CNN has been widely used in classification, recognition, etc. In our research of RF fingerprinting recognition, we load the original signals from various mobile devices into a CNN model for RF fingerprinting recognition. Specifically, we design the CNN model by referring to LeNet-5, which is composed of the operations of convolution and pooling. By using 2 groups of convolution and pooling, we can build up a 5 layered CNN model as Fig. 3. As shown in Fig. 3, the 5 layers of networks include 2 convolution layers, 2 pooling layers, as well as 1 fully connected layer before output. The input data flows through the first convolution layer, the first pooling layer, the second convolution layer, the second pooling layer, as well as a fully connected layer before output.

As the core of a CNN, the convolution layer primary plays the role of extracting RF features from the original signals, and it is composed of a few kernels to operate the convolution computation on the input data. Specifically, a kernel employs a filter in the size of 2×2 to slide and convolve with the input data, creating a feature map with the dimension determined by the sliding interval of the filter. In our CNN architecture, we employ an activation function on the elements of a feature map through a pre-determined transformation, and a typical activation function includes sigmoid, tanh, etc. In our model, we use a rectified linear unit (ReLU) to compose CNN networks, and a ReLU represents the maximum between the input value and zero, i.e. setting each of the negative values to be 0.

In the CNN model for RF fingerprinting recognition, the input data are 2×128 I/Q samples of RF signals. The first convolution layer is in the size of $50 \times 1 \times 3$ with the kernel of ReLU, the first pooling layer is a Max pooling layer, the second convolution layer is in the size of $50 \times 2 \times 3$ with the kernel of ReLU, the second pooling layer is also a Max pooling layer, the last layer is a fully-connected layer with the kernel of Softmax. The output is the ID of a mobile device for recognition.

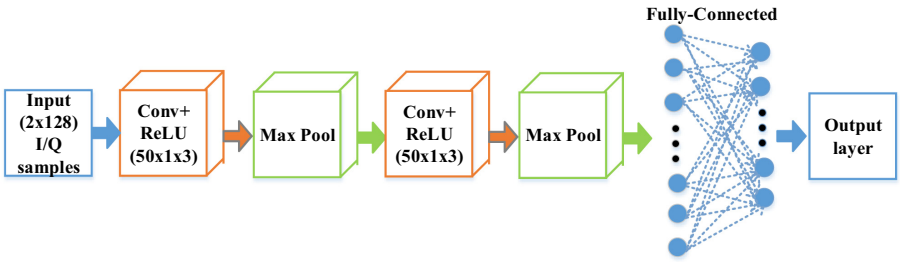


Fig. 3. CNN model for RF fingerprinting recognition

4 Federated Learning Based RF Fingerprinting Recognition

A. Preprocessing data.

The input data of a training model is a sequence of I/Q samples, and usually it represents a time-series of collected data through Rayleigh fading channels. Before we can train raw data, we need to preprocess these data. First, we need to use the channel estimation methods with the objective of minimum mean square error (MMSE), and then normalize the data valued in a range of [0,1]. Before we load the whole sequence of data into the training model, we need to partition data into a few subsequences. Assume that the length of input I/Q sequence is T and each of the subsequences is t, then the number of partitioned subsequences is $M = T/t$. Also we set a window with the length of t, and slide the window over the I/Q sequence with the length of T.

Instead of loading the sequence of I/Q samples with the length of T into our training model, we put multiple subsequences into the model one by one. It is critical but difficult to select the length of subsequence, since a short subsequence can result in a low variance within a subsequence but high bias between subsequences, while a long subsequence can lead to a high variance within a subsequence but low bias between subsequences. In practice, we need to balance the variance within a subsequence and the bias between subsequences when selecting the length of subsequences.

In view of the variation of wireless channels, we can assume that the channels are invariant when the duration of a subsequence is short, i.e. the length of subsequence is small. Thus, we select fairly short subsequences of input samples to train the model and it can simplify the estimation of coefficients in the wireless channels. Specifically, we train the real part and the imaginary part of the I/Q subsequences in a 2×1 vector, respectively.

B. RF Fingerprinting recognition.

In this section, we establish a federated learning based distributed computing model, in which each mobile device trains its own sample data and one server is used to finalize the model by collecting the parameters from each device and modifying the model at each of the device end. In the following, we first present the federated learning based distributed computing model. Then, we present the potential impact of unbalanced amount of sample data on the recognition performance of federated learning. Finally, we address the method of dynamic sample gradient to mitigate the impact of unbalanced amount on the decrease of recognition performance and save the computation time.

(1) Federated learning based heterogeneous data computing model.

Based on the distributed computing model of federated learning, different amount of data is distributed to different devices for computation, and a server is used to coordinate with multiple devices. Each of the devices can update its own federated learning model based on local data, and communicate with the server regularly to achieve the global minimum of learning loss. The flow chart of model training for federated learning is shown in Fig. 4. The federated learning process is composed of three steps: encrypted

sample alignment, encrypted models training, and incentive effect. Encrypted sample alignment refers to the use of encryption-based sample alignment to find the common samples of two parties on the premise that each device does not disclose data in order to combine the characteristics of these samples for modeling; encryption model training refers to using the distribution of public keys, intermediate encryption effects, and model updating to establish learning models; effect incentive refers to recording the effects of the established model through a permanent data recording mechanism such as blockchain to further optimize the parameters in the federated-learning model.

The distributed computing model based on federated learning is designed to enable the server to aggregate the updates of all the mobile devices. The goal of optimization in the model is defined as the overall learning loss, which equals to the weighted average of learning loss of each mobile device and the weight is the proportion of samples trained at each device.

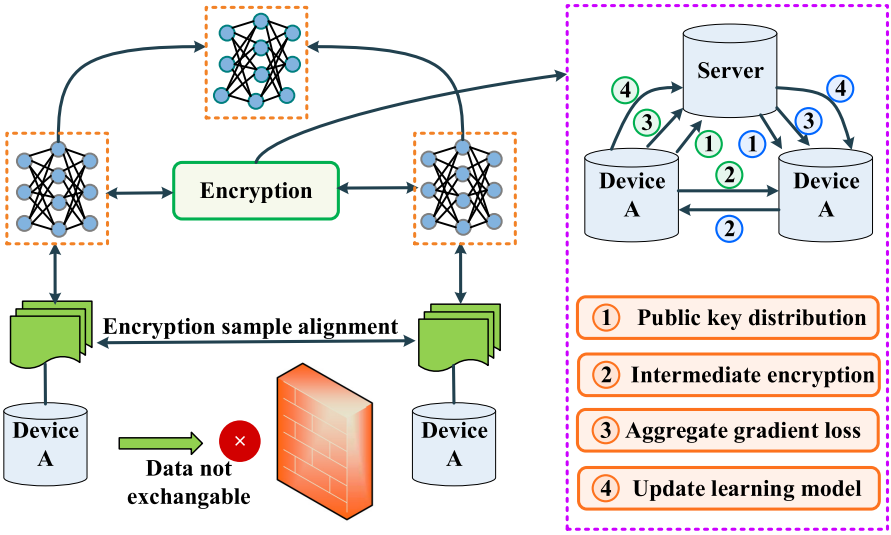


Fig. 4. Flow chart of training in federated learning model

The process of solving the above-mentioned optimization problem can be summarized as follows: we first select n_k mobile devices and use federated learning to train the data distributed at each device. Given the learning rate of η, we can locally compute the average gradient $g_k = \nabla F_k(w_k)$ at each device and send the computation results to the server. For mobile device k, it can update the weight as $w_{t+1}^k \leftarrow w_t^k - \eta g_k$.

In federated learning, when data is heterogeneously distributed at each mobile device, it may cause a few devices to take long time to compute the data, incapable of completing the local update and uploading to the server within the allowed time. In view of the unbalanced distribution of data among heterogeneous mobile devices, we need to quantify the amount of data into several levels for the analysis of federated learning performance. By setting a few thresholds to scale the data amount of mobile devices, we can differentiate the devices which need to handle huge amount of data with those which

only need to process small amount of data. Assume that the capacity of all the mobile devices is almost the same and the sample size of each mobile device can be arranged in an ascending order as $\{n_1, n_2, \cdots, n_k\}$, then the local update time at each mobile device can be computed as $t_1 \leq t_2 \leq \cdots \leq t_k$, since the local update time is linearly increasing with the sample size.

Data is independently generated at each of the distributed mobile devices. The data from different devices have different distribution characteristics, and the amount of training data used when performing local learning is very unbalanced. We intend to use a dynamic sample selection algorithm to mitigate the impact of unbalanced data distribution on the performance of learning model. This method allocates processing tasks according to the data processing capabilities of heterogeneous devices. By guaranteeing the accuracy of learning model, the running efficiency of model can be dramatically improved.

(2) Dynamic sample selection algorithm model.

Mobile devices have different data processing capabilities. When a few devices need to process a large amount of data, they may lead to a dramatic increase in time consumption when performing local update. Previous studies have shown that the use of dynamic sample selection algorithms can meet the challenges of heterogeneous mobile devices to handle large amount of data, and accelerate the convergence rate of data processing algorithms at mobile devices.

Specifically, the dynamic sample selection algorithm computes the estimation of variance obtained by batch gradient to increase the training sample size. The algorithm can dynamically increase the training sample size when setting the initial size to be a small value, and achieve a relatively low computation cost while guaranteeing the expected accuracy of our algorithm. For the local sample set S_k established at device k, the objective function of our dynamic sample selection algorithm can be characterized as the average loss function $l(f(w, x_k), y_k)$ at each device k, given w to be the algorithm parameters, x_k to be the input data and y_k to be the output data.

In the first iteration, we select a data set in a relatively small size and determine whether the sample size can optimize the objective function. If the sample set can enhance the value of an objective function, the sample set is maintained in the next iteration, and new samples in the same size are selected to complete the iteration. Conversely, if the sample set does not increase the value of an objective function, the algorithm will increase the sample size and reselect a new sample based on a higher value to perform the next iteration. In the gradient descent process of a dynamic sample selection algorithm, the vector $\nabla J_S(w)$ represents the descent direction of our objective function J with the parameter w. In order to achieve the convergence of our objective function, we can represent the deterministic conditions as $\|\nabla J_S(w) - \nabla J(w)\|_2 \leq \theta \|\nabla J_S(w)\|_2$.

When the amount of data to be processed by a mobile device is large, the iteration cost in our federated learning algorithm is high. If all the local samples are trained on a single mobile device, we have to experience a high cost of updating our model in each round of iteration. To solve the above-mentioned problem, we intend to use the dynamic sample selection strategy. First, we use a small number of samples to train the model, and then increase the sample size gradually to achieve a higher model accuracy. The

process is summarized in Algorithm 1, where K represents the set of mobile devices, D_k represents the threshold between the size of a large data set and the size of a small data set, N represents the total number of iterations, t represents the iteration round, η represents the iterative learning rate.

Algorithm 1 Dynamic sample selection algorithm (DASA)

Input K, D_k, N, η
Output ω_{t+1}
Initialize ω_0, D_k
for $t = 1, \cdots N$ do
  select S_t where $S_t \subset K$
    for $k \in S_t$ in parallel do
      $\omega_{t+1}^k = DeviceUpdate(k, \omega_t, D_k)$
    end for
    $\omega_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} \omega_{t+1}^k$
end for
function $DeviceUpdate(k, \omega_t, D_k)$
    Initialize n_k
    If $n_k > D_k$
      Initialize $S_0 \subset \{1, \cdots, n_k\}$
      for $i = 1, \cdots N$ do
        $g_k = -\nabla J_S\left(\omega_t^k\right)$
        $\omega_{t+1}^k = \omega_t^k - \eta g_k$
        $i = i + 1$
            Compute the sample variance
      end for
    else
      for $i = 1, \cdots N$ do
        $\omega_{t+1}^k = \omega_t^k - \eta g_k$
      end for
    end if
    return ω_{t+1}^k
end function

In Algorithm 1, the server sends D_k to each of the mobile devices, and each device compares its own sample size n_k with D_k. If $n_k \leq D_k$, we can use all the sample resources without any adjustment and perform the local gradient descent process based on the

stochastic gradient algorithm. Otherwise, a few mobile devices can be selected to adjust their gradient descent parameters. The sample size at each iteration is determined by the estimation of variance obtained by the computation of batch gradient. The complexity of this algorithm can be denoted as $F = O(\omega/\varepsilon)$, where ε refers to the allowed computation error, ω represents the data processing task allocation among mobile devices.

5 Simulation Results

In the federated learning based RF fingerprinting algorithm, we first consider the convergence of our proposed algorithm (shown in Algorithm 1), and compare it with the convergence of a distributed training model without dynamic sample selection, i.e., using the whole set of data at a local mobile device as training data. Also we consider the accuracy of our proposed algorithm with different parameters in the model.

We complete the distributed RF fingerprinting recognition task based on the experiment shown in Fig. 5, which contains the records of I/Q samples from 4 of X310 USRP SDRs as the transmitters and a X310 USRP SDR as the receiver through a Rayleigh fading channel with the signal noise ratio (SNR) of 5 dB. The records are received at the rate of 5 M/s around the frequency of 2.45 GHz, and the total amount of data is 20 million recored in the PXIE 8840 for one mobile device [20].

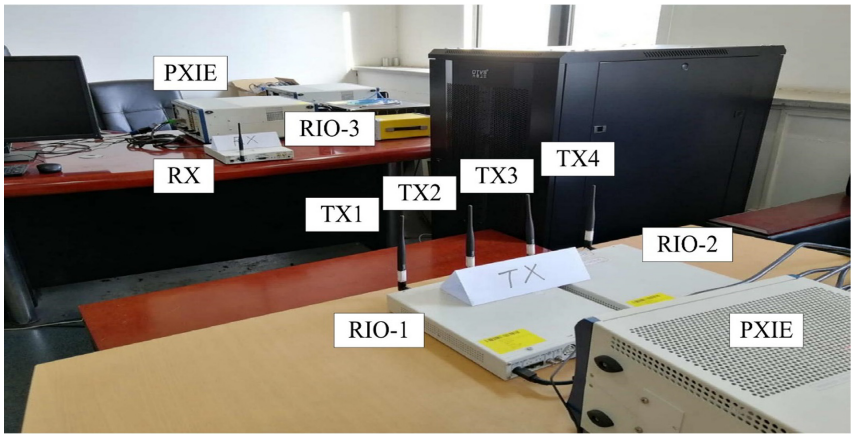


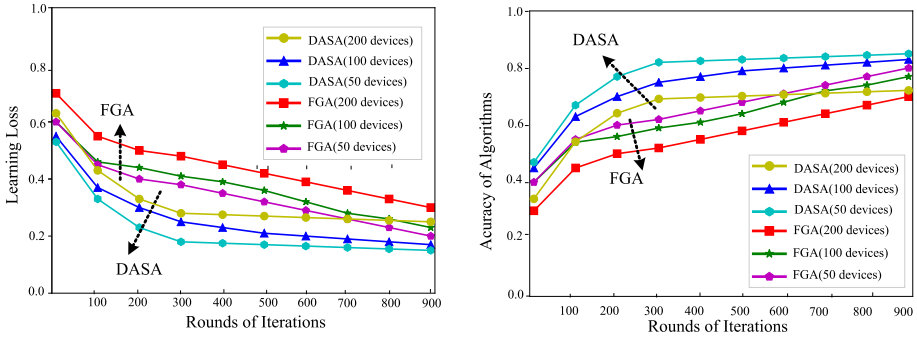
Fig. 5. Experiment environment for RF fingerprinting recognition

A. Convergence Analysis.

Based on the collected data in our experiment, we set the learning rate as 0.2, and set the number of mobile devices to be 50, 100, and 200, respectively. Figure 6 (a) shows the learning loss of the dynamic sample selection algorithm (DASA) in Algorithm 1 and the fixed gradient algorithm (FGA) [21] in which no sample selection is used. In comparison with the FGA, our proposed DASA can effectively reduce the learning loss with different

numbers of mobile devices. In addition, the learning loss of both algorithms increases with the rise of the number of mobile devices, which indicate that the performance gap between the distributed learning algorithms and the centralized algorithm dramatically increases with the number of mobiled devices rising. This is the cost of distributed learning algorithms when loading and training data in the memory of a single device is not applicable.

Figure 6 (b) shows the accuracy of RF fingerprinting recognition in the DASA algorithm (Algorithm 1) and the FGA algorithm. Compared with the FGA algorithm, the DASA algorithm can achieve higher recognition accuracy with different numbers of mobile devices. In addition, the DASA algorithm can converge to the final accuracy results in 400–600 rounds of iterations, while the FGA algorithm needs to complete the convergence after 1000 rounds of iterations. This shows that the DASA algorithm can accelerate the convergence process, reduce the computation cost, as well as guarantee the accuracy of recognition.



(a) Loss of learning (b) Accuracy of algorithms

Fig. 6. Accuracy of federated learning based recognition algorithms

B. Performance Evaluation.

In this section, we compare the system performance of our proposed DASA algorithms with different parameters of η and N_{K}. The former represents the iterative learning rate, while the later represents the number of mobile devices in the set of K.

As shown in Fig. 7, the recognition accuracy decreases with the rise of iterative learning rate η. The internal rationale of this result is that when we enhance the learning rate η, the gradient parameters $\omega_{t+1}^{k} = \omega_{t}^{k} - \eta g_{k}$ in each iteration changes at a high speed, and thus has a high risk of missing the optimum of $\nabla J_S(w)$. Once the optimal $\nabla J_S(w)$ is not achieved, the accuracy of the recognition will be reduced.

Also shown in Fig. 7, the recognition accuracy decreases with the rise of N_{K}. In other words, it is more difficult to recognize the mobile devices when more devices are required to identify. With the subtle differences of RF signals emitted by mobile devices, the process of RF fingerprinting recognition is difficult to complete.

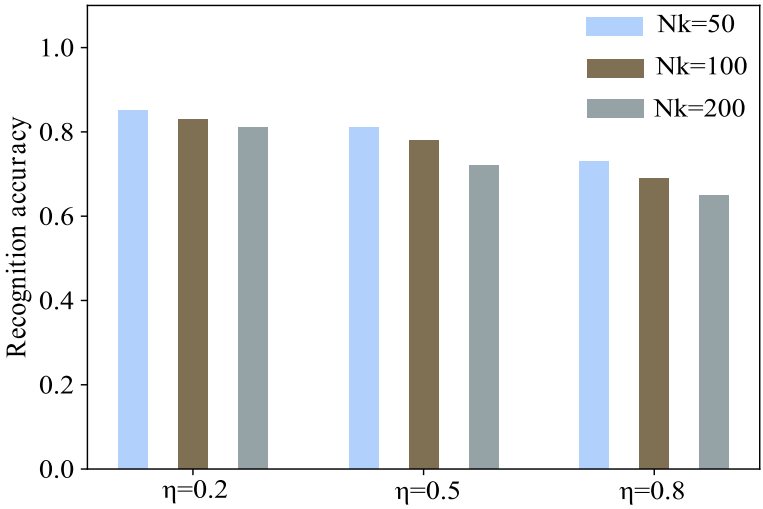


Fig. 7. Accuracy of federated learning algorithms with different parameters

6 Conclusion

To solve the problem of user authentication, we study the federated learning algorithms for the RF fingerprinting recognition of mobile devices in this paper. In consideration of the impact of uneven data distribution on the performance of federated learning algorithms, we propose a dynamic sample selection algorithm to train the RF signals. In comparison with fixed-gradient federated learning algorithms, our proposed algorithm can improve the system accuracy by 10%–20%, while converging to the solution at a higher speed.

Acknowledgements. This work was supported by Sichuan Science and Technology Program, and supported by the Fundamental Research Funds for the Central Universities (no. ZYGX2019J076).

References

1. Li, S., Xu, L.D., Zhao, S.: 5G Internet of Things: a survey. J. Ind. Inf. Integr. **10**, 1–9 (2018)
2. Li, Y.S., Xie, F.Y., Chen, S.L., et al.: Feature Extraction and Recognition of Radio Frequency Fingerprint Signal Suitable for Terminal. Commun. Technol. **251**(001), 63–66 (2018)
3. Zhao, F., Jin, Y.: An optimized radio frequency fingerprint extraction method applied to low-end receivers. In: International Conference on Communication Software and Networks, pp. 753–757 Chongqing (2019)
4. Ding, G., Huang, Z., Wang, X.: Radio Frequency Fingerprint Extraction Based on Singular Values and Singular Vectors of Time-frequency Spectrum. In: International Conference on Signal Processing, pp. 1–6 (2018)
5. Bihl, T.J., Bauer, K.W., Temple, M.A.: Feature selection for RF fingerprinting with multiple discriminant analysis and using ZigBee device emissions. IEEE Trans. Inf. Forensics Secur. **11**(8), 1862–1874 (2016)

6. Zhuo, F., Huang, Y., Chen, J.: Radio frequency fingerprint extraction of radio emitter based on I/Q imbalance. Proc. Comput. Sci. **107**, 472–477 (2017)
7. Li, Y.S., Xie, F.Y., Chen, S.L., et al.: Feature extraction and recognition of radio frequency fingerprint signal suitable for terminal. Commun. Technol. **251**(1), 63–66 (2018)
8. Shi, Y., Jensen, M.A.: Improved radiometric identification of wireless devices using MIMO transmission. IEEE Trans. Inf. Forensics Secur. **6**(4), 1346–1354 (2011)
9. Xiao, Z., Yan, Z.: Radar Emitter Identification Based on Feedforward Neural Networks. In: Electronic and Automation Control Conference, pp. 555–558 Chongqing (2020)
10. Polak, A.C., Dolatshahi, S., Goeckel, D.L.: Identifying wireless users via transmitter imperfections. Sel. Areas Commun. **29**(7), 1469–1479 (2011)
11. Wang, W., Sun, Z., Ren, K., et al.: User capacity of wireless physical-layer identification. IEEE Access **5**, 3353–3368 (2017)
12. Demers, F., ST-Hilaire, M.: Radiometric identification of LTE transmitters. In: IEEE Global Communications Conference (GLOBECOM), pp. 4116–4121. IEEE (2013)
13. Patel, H.J., Temple, M., Baldwin, R.O.: Improving ZigBee device network authentication using ensemble decision tree classifiers with radio frequency distinct native attribute fingerprinting. IEEE Trans. Reliab. **64**(1), 221–233 (2015)
14. Reising, D.R., Temple, M., Jackson, J.: Authorized and rogue device discrimination using dimensionally reduced RF-DNA fingerprints. IEEE Trans. Inf. Forensics Secur. **10**(6), 1180–1192 (2015)
15. Knox, D.A., Kunz, T.: Wireless fingerprints inside a wireless sensor network. ACM Trans. Sens. Netw. **11**(2), 1–30 (2015)
16. Yuan, Y., Huang, Z., Wang, F., et al.: Radio Specific Emitter Identification based on nonlinear characteristics of signal. In: IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), pp. 77–81. IEEE (2015)
17. Hu, S., et al.: Machine learning for RF fingerprinting extraction and identification of soft-defined radio devices. In: Liang Q., Wang W., Mu J., Liu X., Na Z., Chen B. (eds) Artificial Intelligence in China. LNEE, vol. 572, pp. 189–204. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-0187-6_22
18. Sankhe, K., Belgiovine, M., Zhou, F., et al.: Oracle: Optimized radio classification through convolutional neural networks. In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications, pp. 370–378. IEEE (2019)
19. Wu, Q., Feres, C., Kuzmenko, D., et al.: Deep learning based RF Fingerprinting for device identification and wireless security. Electron. Lett. **54**(24), 1405–1407 (2018)
20. Sankhe, K., Belgiovine, M., Zhou, F., Riyaz, S., Ioannidis, S., Chowdhury, K.R.: ORACLE: optimized radio classification through convolutional neural networks. In: IEEE INFOCOM, Paris (2019)
21. McMahan, H., et al.: Communication-efficient learning of deep networks from decentralized data. In: International Conference on Machine Learning (ICML) (2017)