Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

Analysis of RF Device Fingerprinting using Convolutional Neural Network

Parvathi G R
Communication Systems
Electronics and Communication Engineering
TKM College of Engineering
Karicode, Kollam

Preetha Basu
Professor
Electronics and Communication Engineering
TKM College of Engineering
Karicode, Kollam

Abstract—**The authentication of devices and identification of unauthorised signals has been difficult in many scenarios. In this paper, a method for identifying a specific device among a range of similar devices is discussed using the techniques of convolutional neural network. The approach is to use radio fingerprints as samples for the identification process. The variability caused in the manufacture of these wireless transmitters serves as a unique feature, creating repeatable signatures for a particular device. These signatures are used as fingerprints for identification and verification. A framework is then developed for training the convolutional neural network using complex baseband error signals in time domain for varying signal-to-noise ratios (SNRs).**

Keywords—Convolutional Neural Network; device fingerprints; error signals; hardware variability; trained data; Radio Frequency.

I. INTRODUCTION

In the previous couple of decades, there has been an improbable growth within the application of internet and the devices connected to it. However, the security and privacy of these billions of devices could be a preponderating concern within the Internet of Things (IoT) network. Any device that has network property is vulnerable to attacks. Information gathered by such devices are prone to attacks like ID spoofing by intruders. Almost all of those devices have restricted memory capacity and computing power, that makes it troublesome to use complicated cryptographic algorithms that need additional resources than the devices will give. Therefore, there is limited authentication or authorization.

The device identification is the process of relating a received signal to a device employing the radio frequency (RF) features database of certain known transmitters. The most efficient identification systems should be ready to correctly determine emitters, however should even be quick, robust to ever-changing environments, and simply and quickly adjustable. The speed of such systems is particularly essential during a military setting, wherever the system is also accustomed to give early warning. Moreover, in environments where there can be constant changing of channels or when variations occur in the transmitted frequency, modulation schemes and bandwidth or if there occurs presence of other transmitters, it is vital that the developed system works well under these conditions. However, ought to the system have to be compelled to be tuned or changed in order to accommodate a brand new surroundings, such changes ought to be easy and economical to implement.

In several mission crucial eventualities, issues in unauthorised transmissions, authentication of devices and ID spoofing are major considerations. This work on device fingerprinting solves the eventualities by learning device-specific features of the transmitters in an exceedingly pre-deployment training stage, that is then exploited throughout actual network operation. Thus, for general purpose use, any device fingerprinting approach must be computationally straightforward once employed in this field. It is for this sole reason, machine learning (ML) techniques, specifically, the one Convolutional Neural Network (CNN) is proposed and through an experiment demonstrate its radio identification performance for a set of devices.

RF fingerprinting aims to spot transmitters based on the device-specific features found in their emitted analog signals. The main advantage here is the minimal pre-processing performed on the baseband signals that is down-converted which is then sent for identification to neural networks. These features are caused due to the hardware variability within the analog transmitter parts of the device. For this reason, a study on RF fingerprinting supported by applying of deep learning classification strategies is performed and its feasibility is demonstrated. Owing to the huge success of deep learning ways at advanced classification tasks in image and speech recognition, this method tend to apply convolutional neural network (CNN) to the RF fingerprinting drawback.

II. RELATED WORK

A traditional identification system can be represented as in Fig 1. This system consists of an RF module, followed by the data collection module, signal processing module, feature extraction, clustering techniques and finally the identification and verification module [1]. The RF and data collection modules focus on pre-processing the analog signal obtained from the RF receiver, converting it into the digital domain, before the signal processing operation occurs. The stages involved in the signal processing phase mostly deal with the features that need to be extracted in the next stage and primarily include demodulation and filtering. Once all the necessary processing is done, all the data is brought on to the feature extraction stage, where pre-determined features are

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

extracted. The obtained features are further used in clustering and identification purposes.

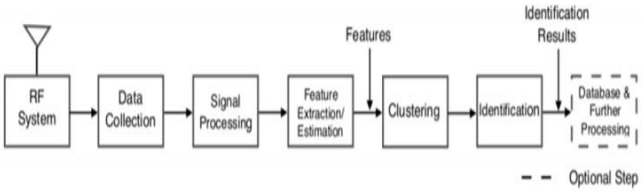

Fig 1: A traditional identification system

Normally, the RF fingerprint of a device, termed as the electronic signature, occurs due to the variations in architecture and RF components employed by manufactured and certain non-idealities seen in the device's hardware [2]. Moreover, RF fingerprints usually does not depend on the signal transmitted or on the data present in that signal. The major success of the traditional identification systems was due to the selection of features which symbolises a certain portion of the RF fingerprints of a set of devices [3]. Often, the features employed in the traditional identification systems either come from the steady-state or the transient portions of the received signal.

The methods used to analyse the steady-state signal are a lot diverse and include usage of modulation-based approaches [4], wavelet-based approaches [5], cyclostationary-based approaches, preamble-based approaches [6]. One of the popular wavelet-based approach, termed as the dynamic wavelet fingerprint approach, utilises wavelet transforms on the steady signal in its time domain [5]. The features generated in the modulation-domain calculate the error between the ideal demodulated signal and the transmitted signal [4], whereas the preamble-based approach analyse the features of the preamble extracted, namely its periodicity [6]. The cyclostationary-based approach, analyse the salient cyclic features present in a signal.

The transient or steady-state techniques are made in use to extract features which are used in combination with the clustering techniques for identification and verification of specific emitters [7]. The most commonly used algorithms include k-Nearest Neighbors (kNN), support vector machines (SVMs) [8], along with principal component analysis (PCA) or linear discriminant analysis (LDA), if required. Furthermore, there can be variations in the frequency-domain and time-domain features according to the channel conditions and noise and which can henceforth be impacted by various channel impairments like the multipath [9].

Even though it is more practical to use the features that are extracted from the received signal that fall under its steady-state portion, there are certain limitations associated with the expert features that are used for the detailing of the steady-state signal. For example, wavelet functions greatly impact the wavelet-based techniques. For cases that lack a pre-defined preamble, the preamble-based techniques fail [10]. Furthermore, methods used to analyse the cyclostationary features of some particular signal are normally seen inconsistent when the phase or frequency uncertainties are present and often consumes a lot of time to compute.

Finally, there are numerous limitations involved in the classification or clustering algorithm of traditional identification systems. Many classification or clustering algorithms such as neural network classifiers, kNNs and SVMs demand to input the cluster number from the user or usage of an iterative method to calculate the required number of clusters for the provided dataset [11]. Other than for a cooperative environment, the number of clusters (devices) cannot be known earlier and is subject to change, causing limitation in the identification of anomalous devices and their behaviors. Some training has been carried out on a probabilistic neural network, to identify eight IEEE 802.11b WiFi cards making use of the transient amplitude signal of WiFi transmissions [12]. Also, various techniques involving wavelet and Hilbert-Huang [13] transforms have been experimented for studying the success of RF fingerprinting.

The major drawbacks of traditional identification approaches involve the extraction and the usage of expert-defined and pre-determined features. The initial step for the design of a traditional identification system involves the need of an "expert" to characterise the quality features under consideration. These defined features remain accurate only over a certain range of parameters and often require consistent and detailed measurement or evaluation in order to facilitate quality identification performance.

III. PROPOSED METHOD

The project consists of two phases, i.e., a training phase and an identification phase. In the first phase, the CNN is trained using raw IQ samples which are collected from each transmitter in order to solve a multi-class classification problem. In the identification stage, those samples belonging to the unknown transmitters are fed as input to the trained neural network and identification of transmitter is done based on values observed at the output layer.

A. Data Generation

The sequence of binary data samples are generated based on particular number of samples for each device, then it modulated with the help of Quadature Amplitude modulation (QAM). After the modulation the signal are converted to Orthogonal Frequency Division Multiplexing (OFDM) data by taking inverse fast Fourier transform, these OFDM signal generated by specific range of symbol to noise ratio and 64 fft size and 52 subcarriers, and pilot is added before conversion of OFDM modulation, similarly we are generating the input for all devices and saving it in a .mat file and loading it for testing.

B. Convolutional Neural Network

The network consists of three convolutional layers and three dense layers. The goal was to first understand stacking of layers and the functions and operations of the various layer. The most challenging task in building the CNN network is to identify how many layers to be defined, number of filters needed in each layer, the filter sizes, values to be given for padding and number of stride. These are not standard values and the network complexity depends on the type of data and the processing required. A lot of effort needs

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

to be put on experimenting with various parameters and finally identifying the exact combination of these hyper parameters that defines the data well. The input given to the CNN is a windowed sequence of raw I/Q samples with length 4320. Further, each of the complex values are represented as three-dimensional real values, which results in the dimension of the input data to be to $1\times 4320 \times 1$.

The first parameter under consideration is the number of filters to be used in the convolutional layers. It was observed that the number of filters within a range of 30 – 256 could provide somewhat similar performance. However, as the number of filters increases, the number of computations also increases. Hence, the first convolutional layer was made of 128 filters whereas the second and third had 32 and 16 respectively for balancing the performance and maintaining computational cost. But, the filter size in all three convolution layer was set as 1×1, since larger filter size does not provide any significant performance improvement in the network. Moreover, on increasing the number of convolution layers from three to four shows no improvement in the performance of the network, which demands continuation with three convolution layers. This is followed by a max pooling layer of dimension 1×1 and having stride 2×2. This layer helps in the extraction of sharp features by reducing the dimensionality through downsampling through reduction of parameters which in turn helps in low computation in the network.

The number of neurons in the dense layer is 10 since ten samples from ten devices are taken for classification. It is found that increasing the number of neurons does not improve the performance. In the parameters selection, it is observed that increasing the number of neurons, increases the complexity and in turn makes the training slower. One of the major problems during network training is over fitting, during which the network weights gets tuned so well to the training examples while the network fails to perform well when given the unseen data. Thus measures need to be taken to alleviate the problem of over fitting. A dropout layer is used, whose main function is to drop out a set of activations in that specific layer by setting them to zero. By doing this, it is made sure that the network is robust and ensure that it does not get too fitted to the training data. A dropout rate of 50% is used at the dense layer.

IV. RESULTS

The work deals with training of ten devices. The transmitted signal from each device is modelled with the help of the MATLAB software. On obtaining the error signal from each device, it is fed on to the CNN as input for training. A total of 200 samples are taken for the training dataset, 20 for each devices and are classified into 10 categories which helps in the identification procedure. Random samples from the training set are given as input for identification process. The important function during training is finding right set of weights that fit the data well and classifies devices correctly by reducing the error at the output layer. This is done using optimizers and here, Adam optimizer is employed. The output predictions for two mat files belonging to device 1and 5 respectively are shown. As the first input mat file (belonging to that of device 1) is fed on to CNN and training is done, a plot of mini-batch accuracy versus number of iterations (=50) is done as shown in Fig 5.2(a). It is clear on plotting that as the number of iterations and learning rate is increased, the identification becomes more accurate. As number of iterations are increased, it is observed that the accuracy increases and the loss decreases (Fig 2(b)). The identification is made as device 1.

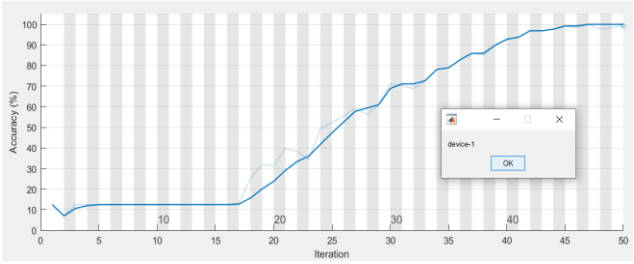

Fig 2(a): Accuracy versus number of iterations (input 1)

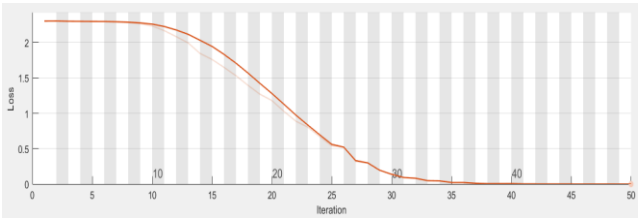

Fig 2(b): Loss versus number of iterations (input 1)

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:01	11.72%	2.3026	0.0100
50	50	00:00:26	100.00%	0.0002	0.0100

Fig 2(c): Improvement in accuracy and loss (input 1)

When another input data (belonging to device 5) is given, it is identified to be of device 5 and is shown in Fig 3(a). The improvement in accuracy and loss as the iteration progresses is further analysed from the obtained results as shown in Fig 3(b) and Fig 3(c). The light and darkened lines indicate the unsmoothed and smoothed curves respectively. If it is smoothed, it is less noisy and the trends can be easily spotted.

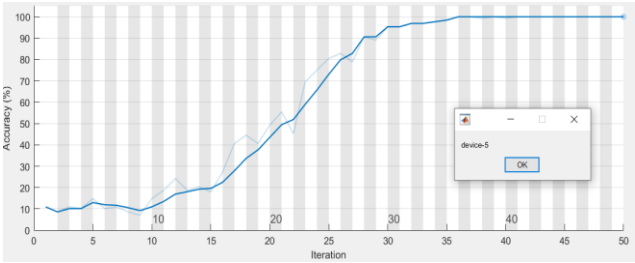

Fig 3(a): Accuracy versus number of iterations (input 2)

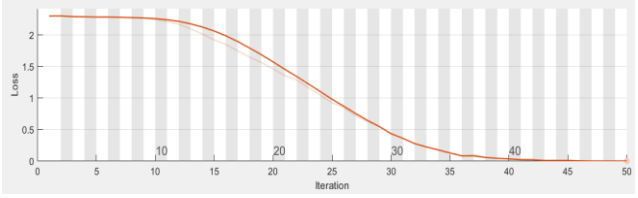

Fig 3(b): Loss versus number of iterations (input 2)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 07, July-2020

```
|=========================================================================|
|  Epoch  |  Iteration  |  Time Elapsed  |  Mini-batch  |  Mini-batch  |  Base Learning  |
|         |             |   (hh:mm:ss)   |   Accuracy   |     Loss     |      Rate       |
|=========================================================================|
|       1 |           1 |       00:00:01 |       10.94% |       2.3026 |          0.0100 |
|      50 |          50 |       00:00:27 |      100.00% |       0.0039 |          0.0100 |
|=========================================================================|
Enter data : '5.mat'
```

Fig 3(c): Improvement in accuracy and loss (input 2)

When a random unknown signal is given as input to the CNN, it does not find any match with the already trained dataset and the MATLAB would plot the input data, but the CNN fails to identify the device to which it has correlation and hence displays it as invalid instead of classifying it into the 10 class of devices. This helps in authentication of unauthorised signals.

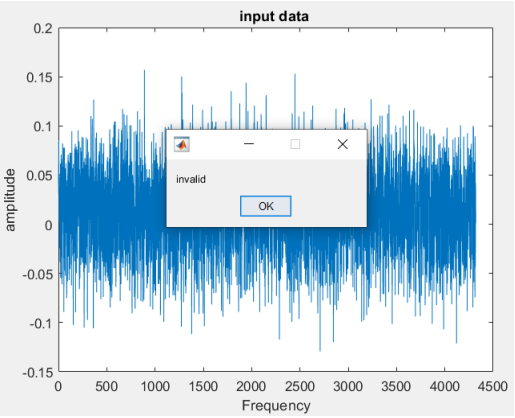

Fig 4: Unknown signal unidentified

The method demonstrates the identification of devices whose transmitted signal samples are present in the trained dataset of CNN. Any other random signals provided as input would be unidentified. This technique is used in military for applications like early warning systems, device tracking, and device location.

V. CONCLUSION

There arises the need for novel techniques that can identify devices and also help detect malicious activity. The existing approaches of device fingerprinting require expert feature extraction and are not capable enough to train large datasets. Hence a radio fingerprinting approach is studied based on deep learning CNN architecture to train the input samples. This approach enables learning features embedded in the signal transformations of wireless transmitters, and helps to identify specific devices. Furthermore, it is clear that this approach of device identification using CNN outperforms traditional machine learning techniques such as logistic regression and SVM for the identification of fewer devices. Finally, it is experimentally validated that the performance of this design improves on increasing the learning rate and the number of iterations.

The future work revolves around increasing the robustness of the CNN architecture so to allow scaling up to correct identification of 1000s of similar devices. Another challenge is realising the right balance between training time and the classification accuracy. The classifier used, performs well on limited number of devices, however, to identify large number of devices (in the range of 1000s), it may require some major changes in the architecture and the need for new optimum parameters.

REFERENCES

[1] Talbot, Kenneth I., Paul R. Duley, and Martin H. Hyatt, "Specific emitter identification and verification", Technology Review 113, 2003.

[2] K. Kim, C. M. Spooner, I. Akbar, and J. H. Reed, “Specific emitter identification for cognitive radio with application to IEEE 802.11,” in 2008 IEEE Global Telecommunications Conference, Nov 2008, pp. 1–5.

[3] K. A. Remley, C. A. Grosvenor, et al. “Electromagnetic signatures of WLAN cards and network security,” in Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology, 2005., Dec 2005, pp. 484–488.

[4] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, “Wireless device identification with radiometric signatures,” in Proceedings of the 14th ACM International Conference on Mobile Computing and Networking. New York, NY, USA: ACM, 2008, pp. 116–127.

[5] C. Bertoncini, K. Rudd, B. Nousain, and M. Hinders, “Wavelet fingerprinting of radiofrequency identification (RFID) tags,” IEEE Tranactions on Industrial Electronics, vol. 59, no. 12, pp. 4843–4850, Dec 2012.

[6] H. L. Yuan and A. Q. Hu, “Preamble-based detection of Wi-Fi transmitter RF fingerprints,” Electronics Letters, vol. 46, no. 16, pp. 1165–1167, August 2010.

[7] C. Song, Y. Zhan, and L. Guo, “Specific emitter identification based on intrinsic timescale decomposition,” 6th Int. Conf. on Wireless Comm. Netw. and Mobile Comp., Sept 2010, pp. 1–4.

[8] G. Huang, Y. Yuan, X. Wang, and Z. Huang, “Specific emitter identification based on nonlinear dynamical characteristics,” Canadian Journal of Electronics and Computer Engineering, vol. 39, no. 1, pp. 34–41, winter 2016.

[9] H. Weng, X. Dong, X. Hu, D. G. Beetner, T. Hubing, and D. Wunsch, “Neural network detection and identification of electronic devices based on their unintended emissions,” in 2005 International Symposium on Electromagnetic Compatibility, Chicago, IL, Aug. 2005.

[10] T. J. O’Shea, N. West, M. Vondal, and T. C. Clancy, “Semi-supervised radio signal identification,” in 2017 19th International Conference on Advanced Communication Technology, Feb 2017, pp. 33–38.

[11] J. Matuszewski and K. Sikorska-ukasiewicz, “Neural network application for emitter identification,” in 18th International Radar Symposium, June 2017, pp. 1–8.

[12] O. Ureten and N. Serinken, “Wireless security through RF fingerprinting,”Canadian Journal of Electrical and Computer Engineering, vol.32, no. 1, pp. 27–33, 2007.

[13] Y. Yuan, Z. Huang, H. Wu, and X. Wang, “Specific emitter identification based on hilbert-huang transform-based time-frequency-energy distribution features,” IET Communications, vol. 8, no. 13, pp. 2404–2412, Sep. 2014.