LoRa Devices Identification Based on Differential Constellation Trace Figure

Xiuting Wu[1(✉)], Yu Jiang[1,2(✉)], and Aiqun Hu[1,2(✉)]

[1] Southeast University, Nanjing, China
{220184473,jiangyu,aqhu}@seu.edu.cn
[2] Purple Mountain Laboratories, Nanjing, China

Abstract. With the development of the Internet of Things, the application of LoRa devices is more and more extensive. However, the security access issues of LoRa terminals has not been paid enough attention yet. In this paper, a novel device identification method is proposed according to the modulation principle and frame format of LoRa signals, which is based on the implementation of differential constellation figure on the physical layer. Unique RF fingerprint features can be extracted from each LoRa signal frame or even the preamble part through this method, thus realizing device identification. The method principle is theoretically analyzed and then experiments are conducted with four LoRa devices using the method of pattern classification, which turns out to be a high accuracy rate device identification for every signal frame of these devices, verifying the effectiveness of the identification method.

Keywords: LoRa · RF fingerprint · Differential constellation trace figure · Device identification

1 Introduction

Internet of things (IoT) is a network technique which connects information collection equipments such as radio frequency identification (RFID) cards, sensors, infrared sensors, global positioning system and laser scanners with the Internet for information exchange and communication according to the agreed protocol, so as to realize intelligent identification, positioning, tracking, monitoring and management, etc. [1, 2]. With the continuous development of science and technology, IoT technology is widely used in daily life, urban infrastructure, agriculture and other places, and therefore the variety and the number of IoT devices increases, bringing up many addressing issues, attacks and information leaks [3, 4].

Traditional Internet-based security problems have a lot of mature and effective solutions, such as terminal weak password has a high-strength password design scheme to solve, etc. [5]. However, IoT has its own unique security problems, including weak authentication and authorization mechanism; lack of transmission layer encryption; illegal remote control after terminals online; risks brought by new nodes and local area

X. Sun et al. (Eds.): ICAIS 2020, LNCS 12239, pp. 658–669, 2020.
https://doi.org/10.1007/978-3-030-57884-8_58

networks introduced by the perception layer; security issues of uncontrolled environmental terminals, etc. [6], however there isn't any perfect security scheme yet on the issue of access control.

There is a contradiction between communication distance and power consumption for technologies like Wi-Fi, Bluetooth, ZigBee, 3G, LTE, etc. The emergence of Low Power Wide Area Network (LPWAN) communication technology solves this difficult problem [7]. LoRa (long range) communication technology, as a kind of wireless technology in LPWAN, has the advantages of long transmission distance and low power consumption simultaneously. In addition, LoRa equipment has high autonomy in networking, and its industrial chain is relatively mature and the commercialization application is earlier [8]. LoRaWAN is a set of communication protocol and system architecture designed by LoRa alliance based on LoRa long-distance communication network, which follows the protocol of Low-Rate Wireless Personal Area Networks (IEEE802.115.4-2011) [9].

In recent years, more and more researches show that radio frequency characteristics of the equipment can be extracted from electromagnetic waves emitted by the wireless communication system [10]. Due to the differences of electronic components in the equipments, the electromagnetic wave emitted by these equipments contains their unique RF characteristics, which can be a parameter for device identity authentication, also known as "RF fingerprint" [11]. Feature extraction based on RF fingerprints is on the physical layer of the communication system, which is not easy to be modified, so it can protect the system security from the bottom of the communication system.

At present, the most widely used method of RF fingerprint extraction is based on the transient response and steady-state response of the system [12–14]. In addition to the above two methods, in 2016, Peng et al. effectively completed the identification of ZigBee equipment based on the novel method of differential constellation trace figure (DCTF) [15]. In 2017, robyns et al. proposed and analyzed a new fingerprint recognition method based on supervised machine learning, taking the data preprocessed as the whole identification object, turning out a high recognition accuracy [16].

This paper explores how to realize the device identification based on the fingerprint information of the physical layer in the LoRa network. Theoretically we analyze the modulation mode, frame format and demodulation principle of LoRa signal. Experimentally LoRa signal is collected with a USRP device, and then after preprocessing, signal characteristics of valid data section and preamble section of every signal frame are analyzed to extract RF Fingerprint using the method of DCTF to realize LoRa device identification on the physical layer. Finally through the method of pattern classification, we realize the classification of four LoRa modules, which verifies the effectiveness of the LoRa device identification method.

2 LoRa Signal Analysis

2.1 LoRa Modulation Principle

LoRa modulation scheme is improved from chirp spread spectrum (CSS) scheme [17]. Linear frequency modulation (LFM) signal, also known as chirp signal, has a constant amplitude with the frequency changing linearly across the whole bandwidth. LoRa signal modulation mainly depends on chirp pulse to encode information.

LoRa modulation technology mainly has four key parameters: carrier frequency (f_c), bandwidth (BW), spreading factor (SF) and code rate (CR) to realize signal modulation and control of the wireless communication. The signal representation is as follows:

$$s(t) = e^{j(2\pi f_c t + 2\pi \frac{\beta}{2} t^2)} \tag{1}$$

$$\text{Where,} \quad \beta = \frac{BW}{T_{symbol}} \tag{2}$$

$$T_{symbol} = \frac{2^{SF}}{BW} \cdot CR \tag{3}$$

For LFM spread spectrum, the instantaneous frequency of chirp signal is a time-dependent linear function:

$$f(t) = f_c + \mu \cdot \frac{B}{T} \cdot t = f_c + \mu kt \tag{4}$$

Where, f_c is the carrier center frequency, μ represents the instantaneous frequency change slope of chirp signal, $\mu = 1$ represents a up-chirp, $\mu = -1$ represents a down-chirp, B represents bandwidth, k is the frequency modulation slope. A typical chirp signal is shown in Fig. 1.

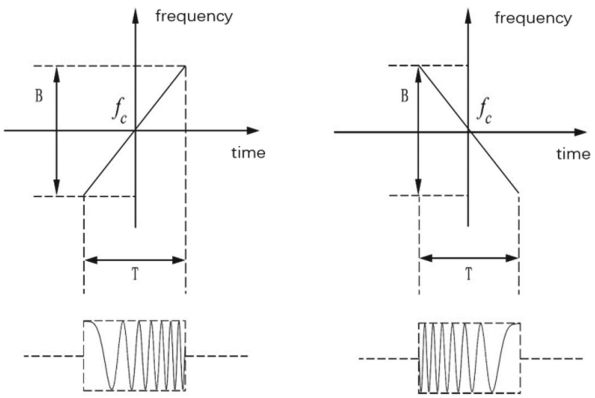


Fig. 1. Typical chirp signal and its instantaneous frequency.

LoRa frame structure is shown in Fig. 2, which starts from the preamble that is used to keep the receiver and transmitter synchronized, including synchronization word (SYN) and start of frame delimiter (SFD). And then following the LoRa physical header (PHDR) plus a header CRC, and the PHDR can explicit or implicit. PHDR includes the length of information data (only 255 bytes at most), error correction coding rate and whether load CRC is carried at the end of the frame. While these messages are known or fixed, implicit mode is recommended to improve efficiency, shorten transmission time and reduce power consumption.

The most significant difference between the protocol of LoRa and LoRaWAN is that part of the data load in LoRaWAN is encrypted by AES128 while the frame format is almost the same.

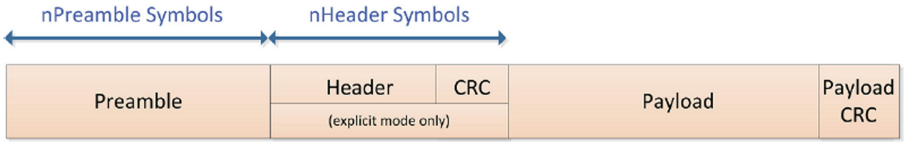


Fig. 2. LoRa frame format in explicit mode.

2.2 Modulation and Demodulation

In LoRa modulation, each chirp symbol is composed of a linear sweep signal. Time from the signal starting position to the frequency mutation is called the symbol duration (SD), up to the value of spreading factor (SF) within the frequency range of band width, determining the symbol value (SV) ranging from 0 to $2^{SF} - 1$, as shown in Fig. 3.

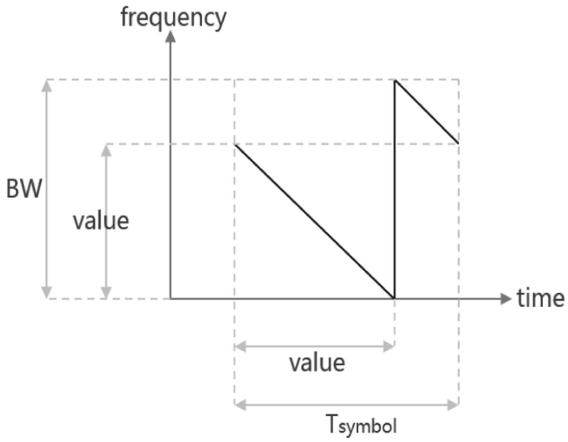


Fig. 3. Definition of chirp signal value.

$$\text{Here,} \quad SD = \frac{sv}{2^{SF}} T_{symbol} \tag{5}$$

The time-frequency diagram of the effective data section of a typical LoRa signal is shown in the Fig. 4, with 10 up-chirp symbols as the preamble part, and 2 down-chirp as SFD, and then the data load part. In the process of demodulation, we find the starting position of the effective data segment through the signal energy threshold method, and then synchronize the signal preamble using the sync word (SYN). Additionally the data synchronization is precisely calibrated based on SFD symbols, and then SD and SV can be calculated.

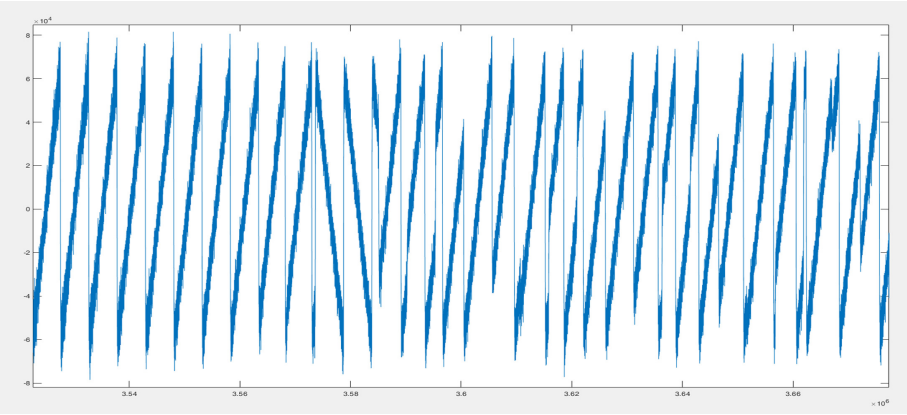


Fig. 4. Time frequency diagram of LoRa effective data segment.

3 LoRa Signal Acquisition and Preprocessing

3.1 Signal Acquisition

In the laboratory environment, USRP Mini B200 is used for data sampling, with LoRa modules placed in a fixed position. USRP is connected to the host through a USB port, and is equipped with a receiving antenna, as shown in Fig. 5.

Fig. 5. USRP Mini B200 receiving device.

Two types of LoRa modules used to generate LoRa signals, among which there are three modules based on LoRa SX1278 chip and one RAKWireless rak811 module following the LoRaWAN protocol within a SX1276 chip. The differences between these LoRa chirps are shown in Table 1.

According to different chip parameters, we set $f_c = 433$ MHz for the SX1278 modules while set $f_c = 868$ MHz for the rak811 module. As for other parameters, SF = 7, BW = 125 kHz, CR = 4/5, PHDR in implicit mode, they are set the same. Besides, all these modules are set to transmit certain content. The modules are shown in Fig. 6.

Table 1. Differences between LoRa chips.

Chip	Frequency	SF	BW	Bit rate
SX1276	137–1020 MHz	6–12	7.8–500 kHz	0.018–37.5 kbps
SX1277	137–1020 MHz	6–9	7.8–500 kHz	0.11–37.5 kbps
SX1278	137–525 MHz	6–12	7.8–500 kHz	0.018–37.5 kbps

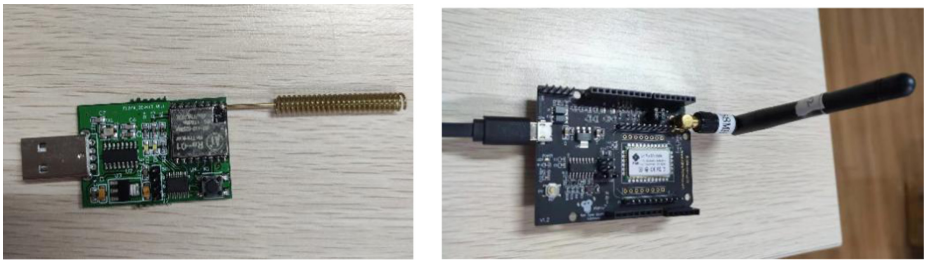


Fig. 6. LoRa SX1278 module and RAKWireless rak811 module.

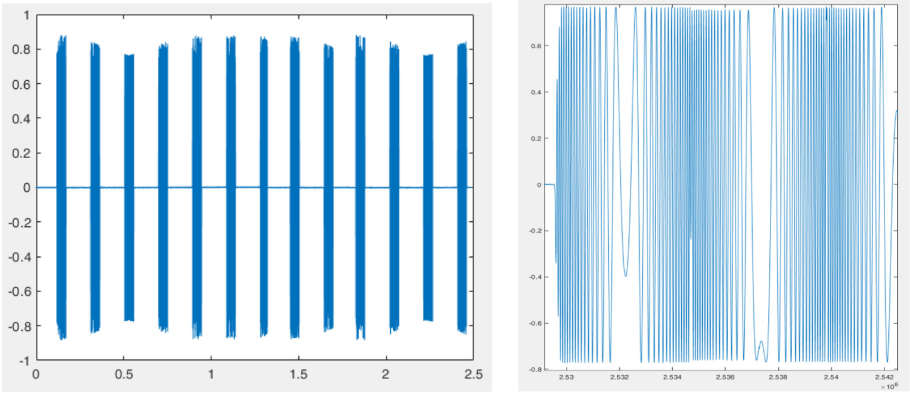
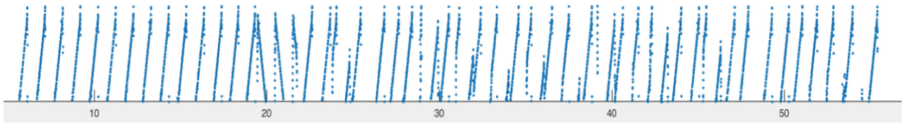


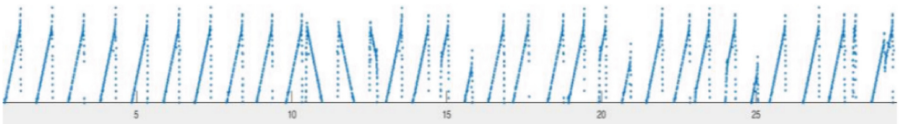
Fig. 7. The original signal and its effective data fragment.

Matlab is used to process the received signal. The original signal and its effective data fragment are shown in the Fig. 7.

After preprocessing, draw the time-frequency graph of the effective data segment of each type of module, and they are shown in the Fig. 8(a) and Fig. 8(b) respectively. It can be seen that different modules have different frame structures in the preamble. The signal of sx1278 module starts with a preamble composed of 14 up chirp signals while rakwireless rak811 module consists of 10 up-chirp preamble. This is because different LoRa modules can be designed according to different standards as long as they follow the protocol. The longer the preamble is sent, the longer the power consumption will be, and the channel resource occupation will also be affected.



(a) LoRa SX1278



(b) rak811

Fig. 8. Time-frequency graph of the effective data segment for the two types of modules.

4 Feature Extraction and Module Recognition

4.1 Principle of Differential Constellation Trace Figure

The constellation diagram can be obtained if we draw the I/Q channel data on the coordinate plane, which reflects the characteristics and relationships between the signals and can be used for studying the digital communication system through the image with an image recognition based method. However, when the constellation diagram is used to analyze the RF characteristics directly, the receiving symbol deviates from its position due to the influence of frequency offset, and as time accumulates, a concentric ring is finally drawn, covering up valid information. By performing differential processing on the received baseband signal, the rotation of the received symbol due to the frequency offset can be eliminated.

Generally, the transmitter and the receiver have frequency offset. If the transmitter carrier frequency is f_{ct1} and the baseband signal is $X(t)$, the transmitted is:

$$S(t) = X(t)e^{-j2\pi f_{ct1}t} \tag{6}$$

For an ideal channel, there is R(t) = S(t), However the signal received is:

$$Y(t) = R(t)e^{-j(2\pi f_{ct2}t+\phi)} = S(t)e^{-j(2\pi f_{ct2}t+\phi)} \tag{7}$$

Where f_{ct2} is the receiver carrier frequency and ϕ is the received signal phase error. Since the transmitter and receiver have a deviation $\Delta f = f_{ct2} - f_{ct1}$, so

$$Y(t) = X(t)e^{j2\pi \Delta ft+\phi}. \tag{8}$$

While received signal contains a rotation factor $e^{j2\pi \Delta ft}$, in order to solve this problem and reflect the frequency offset on the picture, the data need to be differentially operated:

$$D(t) = Y(t) * Y^*(t+n) = X(t) \cdot X(t+n)e^{-j2\pi\theta n} \tag{9}$$

The result of the difference still has a rotation factor $e^{-j2\pi\theta n}$, but it is a stable value, which can be used to directly reflect the frequency offset characteristic of the signal in the constellation trace figure, making it feasible to extract RF fingerprint based on the method of subsequent differential constellation trace figure.

4.2 Feature Extraction and Module Recognition

According to the parameters set during signal acquisition in Sect. 3.1, set the rakwireless rak811 module FC = 868.1 MHz, SF = 7, BW = 125 kHz, Cr = 4/5, set the PHDR to the hidden header mode, and send a fixed section of data irregularly The difference between the parameters set by sx1278 and rak811 is that the carrier frequency fc is 433 MHz, and a fixed period of data is sent at a certain interval. The center frequency point set by USRP equipment at the receiving end is consistent with the carrier frequency of LoRa transmitting module, and the sampling frequency FS is set to 5 MHz. During sampling, USRP and LoRa modules are fixed at a distance of about 1 m, and there is no shelter between them.

According to the parameters set by LoRa module, the signal transmission rate Rs can be obtained by:

$$R_s = \frac{BW}{2^{SF}} = 976.5625\, symbol/s \tag{10}$$

Combined with the sampling frequency, the number of sampling points of each LoRa symbol can be calculated:

$$N = \frac{f_s}{R_s} = 5120 \tag{11}$$

Taking the differential interval to 5120, the sampling data is differential processed, and the differential constellation trace figure is drawn. After visual processing, the density of symbols in different positions is represented by colors on the constellation figure. The larger the symbol density is, the closer the color is to red. The final DCTF for each device is as shown in the Fig. 9.

It can be seen that for LoRa modulated signals, the DCTF clustering center is obvious, and the symbol distribution is centralized with small noise interference. Based on the differences of these images, the methods of image processing and pattern classification can be applied to feature extraction. In this experiment, the method of pattern classification is used to realize LoRa device recognition and the specific operation is as follows:

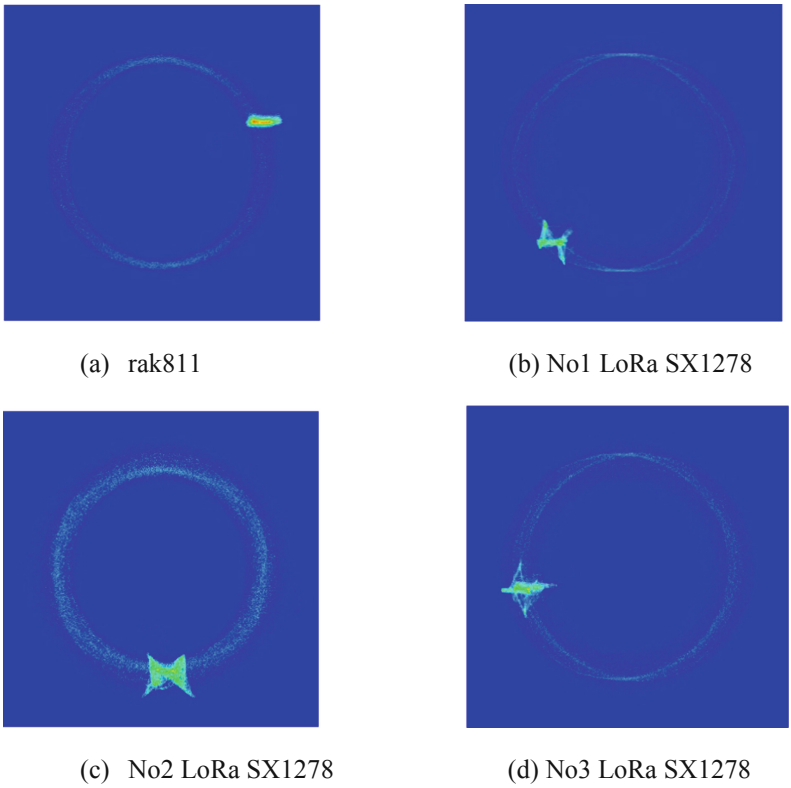
First, in the training data set of a module I, find the maximum value of the point density in the DCTF for each effective data segment;

Second, find the location of all the qualified points reaching 0.95 of the maximum, as shown in the figure.

Then, calculate the clustering center Ci of the module I taking all the eligible points in all valid data segments of this device into consideration, and the clustering centers of other three devices are calculated by this method.

Finally, analyze the valid data segments of four modules in the test data set by calculating the average position X of all the eligible points in each valid data segment. And then calculate the Euclidean distance Di between X and each cluster center Ci, and classify this processing data segment as the category where Di is the minimum value.

In the experiment, we trained 304 pieces of valid data segments of four modules to get the clustering center of each module, and then classify 149 pieces of valid data segments of these four modules, which are all successful, as shown in the Fig. 10.



(a) rak811 (b) No1 LoRa SX1278

(c) No2 LoRa SX1278 (d) No3 LoRa SX1278

Fig. 9. Typical DCTF of a valid data segment for each device. (Color figure online)

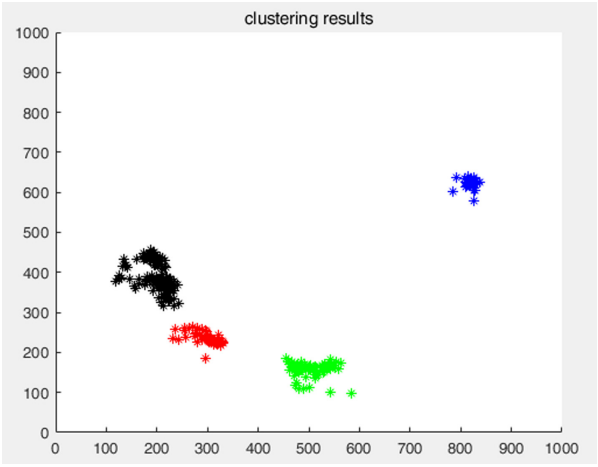


Fig. 10. Clustering results using valid data segments of four modules based on DCTF.

The above experimental results are obtained when all modules send specified data, while the data of the rak811 module is partially encrypted and can be regarded as random data. From section, it is known that preamble of each module is unchanging, therefore we try to draw the DCTF just using the preamble part, and the result is as shown in the Fig. 11, corresponding to the same module in Fig. 9.

Fig. 11. DCTF of the preamble part for each device.

It can be seen that, based on the DCTF of the preamble part, the data noise interference is basically absent, and the data is more centralized. In addition, the contour similarity between the DCTF drawn by a whole segment of data and the DCTF drawn by the preamble is very high.

For the same test data set, pattern classification is executed based on the DCTF generated from the preamble parts of all valid data segments, and the classification accuracy is 98.66%, which is shown in Fig. 12.

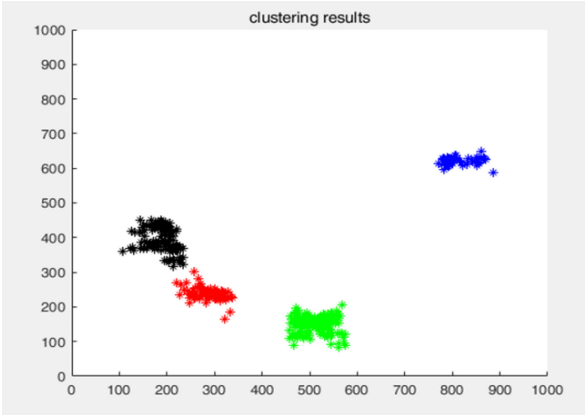


Fig. 12. Clustering results using data preamble parts of four modules based on DCTF.

From the clustering results in Fig. 10 and Fig. 12, we can conclude that the feature extracted from DCTF can achieve the purpose of module identification, and has no relation with the content transferred. Considering the high contour similarity between the DCTF drawn by a whole valid data segment of data and its preamble, so it can be considered that when there are many devices to be identified, the characteristic of DCTF contour similarity can be applied for further precise classification.

5 Conclusion

After analyzing the security situation of IoT, this paper proposes a physical layer based device identification method—DCTF for LoRa devices. According to the modulation principle and frame format of LoRa signal, DCTF can be used to extract unique RF fingerprint features of each device and experimentally realize a high accuracy rate in device identification for per valid data frame of 4 LoRa devices with the whole valid data segment or just the preamble part. The follow-up work will focus on how to uniquely identify LoRa devices according to the preamble of LoRa signals taking SNR or sample rate into account, so as to improve the stability and robustness of the device identification system.

References

1. Huang, Y., Li, G.: Descriptive models for Internet of Things. In: 2010 International Conference on Intelligent Control and Information Processing, pp. 483–486. IEEE, August 2010
2. Wang, J., Gao, Y., Liu, W., Wu, W., Lim, S.J.: An asynchronous clustering and mobile data gathering schema based on timer mechanism in wireless sensor networks. Comput. Mater. Contin. **58**, 711–725 (2019)
3. Conti, M., Dehghantanha, A., Franke, K., Watson, S.: Internet of Things security and forensics: challenges and opportunities. Future Gener. Comput. Syst. **78**, 544–546 (2018)
4. Liu, W., Luo, X., Liu, Y., Liu, J., Liu, M., Shi, Y.Q.: Localization algorithm of indoor Wi-Fi access points based on signal strength relative relationship and region division. Comput. Mater. Contin. **55**(1), 71–93 (2018)
5. Cheng, J., Xu, R., Tang, X., Sheng, V.S., Cai, C.: An abnormal network flow feature sequence prediction approach for DDoS attacks detection in big data environment. Comput. Mater. Contin. **55**(1), 95–119 (2018)
6. Mahmoud, R., Yousuf, T., Aloul, F., Zualkernan, I.: Internet of Things (IoT) security: current status, challenges and prospective measures. In: 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST), pp. 336–341. IEEE, December 2015
7. Li, L., Ren, J., Zhu, Q.: On the application of LoRa LPWAN technology in Sailing Monitoring System. In: 2017 13th Annual Conference on Wireless On-demand Network Systems and Services (WONS), pp. 77–80. IEEE, February 2017
8. Neumann, P., Montavont, J., Noël, T.: Indoor deployment of low-power wide area networks (LPWAN): a LoRaWAN case study. In: 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 1–8. IEEE, October 2016
9. de Carvalho Silva, J., Rodrigues, J.J., Alberti, A.M., Solic, P., Aquino, A.L.: LoRaWAN—a low power WAN protocol for Internet of Things: a review and opportunities. In: 2017 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech), pp. 1–6. IEEE, July 2017
10. Danev, B., Zanetti, D., Capkun, S.: On physical-layer identification of wireless devices. ACM Comput. Surv. (CSUR) **45**(1), 6 (2012)
11. Honglin, Y., Aiqun, H.: Fountainhead and uniqueness of RF fingerprint. J. Southeast Univ. (Nat. Sci. Ed.) **39**(2), 230–233 (2009)
12. Demers, F., St-Hilaire, M.: Radiometric identification of LTE transmitters. In: 2013 IEEE Global Communications Conference (GLOBECOM), pp. 4116– 4121. IEEE, December 2013

13. Remley, K.A., 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, pp. 484–488. IEEE, December 2005
14. Romero, H.P., Remley, K.A., Williams, D.F., Wang, C.M.: Electromagnetic measurements for counterfeit detection of radio frequency identification cards. IEEE Trans. Microw. Theory Tech. **57**(5), 1383–1387 (2009)
15. Peng, L., Hu, A., Jiang, Y., Yan, Y., Zhu, C.: A differential constellation trace figure based device identification method for ZigBee nodes. In: 2016 8th International Conference on Wireless Communications & Signal Processing (WCSP), pp. 1–6. IEEE, October 2016
16. Robyns, P., Marin, E., Lamotte, W., Quax, P., Singelée, D., Preneel, B.: Physical-layer fingerprinting of LoRa devices using supervised and zero-shot learning. In: Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks, pp. 58–63. ACM, July 2017
17. Vangelista, L.: Frequency shift chirp modulation: the LoRa modulation. IEEE Signal Process. Lett. **24**(12), 1818–1821 (2017)