IIS: Intelligent Identification Scheme of Massive IoT Devices

1st Jie LIU
School of Computer Science
(National Pilot Software Engineering School)
Beijing University of Posts and Telecommunications
Beijing, China
ljbupt@bupt.edu.cn

2nd Yi SUN
School of Computer Science
(National Pilot Software Engineering School)
Beijing University of Posts and Telecommunications
Beijing, China
sybupt@bupt.edu.cn

3rd Fengkai XU
The 6th Research Institute of China Electronics Corporation
Beijing, China
xufk@ncse.com.cn

4th Keping YU
Global Information and Telecommunication Institute
Waseda University
Tokyo, Japan
keping.yu@aoni.waseda.jp

5th Ali Kashif Bashir
Department of Computing and Mathematics
Manchester Metropolitan University
Manchester, United Kingdom
dr.alikashif.b@ieee.org

6th Zhaoli LIU
School of Computer Science
(National Pilot Software Engineering School)
Beijing University of Posts and Telecommunications
Beijing, China
joel_liu@bupt.edu.cn

***Abstract*—Device identification is of great importance in system management and network security. Especially, it is the priority in industrial internet of things (IIoT) scenario. Since there are massive devices producing various kinds of information in manufacturing process, the robustness, reliability, security and real-time control of the whole system is based on the identification of the massive IIoT devices. Previous IIoT device identification solutions are mostly based on a centralized architecture, which brings a lot of problems in scalability and security. In addition, most traditional identification systems can only identify inherent types of devices which is not suitable for the adaptive device management in IIoT. In order to solve these problems, this paper proposes a Intelligent Identification Scheme(IIS) of Massive IoT Devices, a completely distributed intelligent identification scheme of massive IIoT devices. The scheme changes the traditional centralized architecture and realizes more efficient clustering identification of massive IIoT devices. Moreover, IIS can identify more and more types of devices intelligently with the continuous learning ability since the identification model is constantly updated according to the ledger which is maintained by all gateways collaboratively. We also conduct experiments to evaluate the performance of IIS based on the data obtained from real IIoT devices, which proves that IIS is efficient in device identification and intelligent for the adaptive device management in IIoT.**

***Index Terms*—IoT Security, Device Identification, Device Management**

I. Introduction

Generally speaking, the first step of device identification is to extract device fingerprinting [1]–[6], which is to collect device characteristics and characterize them [7]. Device fingerprinting can be divided into two categories: active and passive. Bratus et al. [8] propose an active fingerprinting method for 802.11 wireless device identification, identifying the device types by sending a series of carefully designed non-standard 802.11 frames to the target device and observing the response (or lack of response) of the target device. However, this kind of active fingerprinting method needs to be connected with the device as a premise. With the rapid growth of the number of IIoT devices, its scalability is poor, and it will occupy the bandwidth of the network, resulting in the waste of network resources. Therefore, the most common method is the passive fingerprinting, which does not need to establish real connection and consume any network resource. It has stronger scalability and is more suitable for IIoT scenario. Therefore, we apply the passive fingerprinting method this paper.

In terms of the identification methods, there are many research results [9]–[16]. Meidan et al. [17] firstly used machine learning method for traffic analysis to distinguish devices. They generate a classifier C_i for each device type in the network, and extract sessions (a unique 4-tuples consisting of source and destination IP addresses, port numbers, and connection flags) from the TCP packets of device traffic. Each session is represented by a feature vector S. When identification begins, the feature vector S of the target device is inputted into the classifier C_i of device type D_i, and then a probability value P_i is outputted as the result. When the probability value is greater than a certain threshold value, it means that the target device belongs to the device type D_i; otherwise, it belongs to other devices.

Kotak, Jaidip et al. [18] establish a neural network model [19]–[22] for device identification. They convert TCP session payloads into grayscale images during the data processing phase to represent the ”communication behavior” of IoT devices. After the model is trained, it is only necessary to inject the grayscale image of the target device into the model to identify the device type. Different from previous identification methods, this solution avoids complex feature analysis. Moreover, it focuses on the network traffic payloads of different IoT devices rather than the packet headers, so it is applicable to common IoT devices regardless of the communication protocol used.

Noguchi, Hirofumi et al. [23] designed a device identifier based on feature analysis of traffic packets. In their solution, the similarity is calculated by a feature similarity calculation algorithm and then the identifier compares the similarity values of the traffic features of the target device with those of the devices in the database. When the similarity exceeds the threshold, the match is successful, that is, the device identification is successful; otherwise, the traffic features of the new device will be stored in the database. With the automatic accumulation of traffic features of new devices, the identification performance of the identifier will be improved constantly with time.

Le, Franck and Ortiz et al. [24] apply natural language processing technology in traffic analysis for the first time combining TF-IDE algorithm to domain name resolution process for identifying device type and manufacturer. But this approach is not suitable for large-scale IoT scenario because some non-commercial entities may deploy devices which will not issue DNS requests. Furthermore, device identification can be combined with some strategies to protect system security. In this respect, Miettinen et al. [25] proposed an IoT sentinel system by combining device identification with vulnerability query and isolation strategy to prevent vulnerable devices from affecting the security of other devices. After identifying the type of the device, the system evaluates the vulnerability of the device, and isolates the device from the network if there is a security threat. In order to prevent illegal devices from stealing legal identity to attack the network and other devices, Yousefnezhad, Narges et al. [26] proposed a device identification framework (MeDI) based on device behavior analysis. It identifies the security of the device by monitoring the data packet sent by the device and protect the server from accepting and spreading false data.

However, although the above methods improve the device identification system in some aspects, they have two common flaws that they are all based on a centralized framework as shown in Fig.1 [27]–[30]and can only identify the fixed device types. In order to solve these two hard problems, Thangavelu et al. [31] proposed DEFT, a distributed device fingerprinting solution, which works in a hierarchical network architecture and contains two entities – control logic and gateway. The identification model of all gateways comes from the control logic, which updates the model and sends it back to the gateway according to the feature vector of the unknown device sent by the gateway. The gateway can identify the device locally, so the efficiency of identification is greatly improved. In addition, the identification model is constantly updated. However, it is still a centralized identification architecture essentially, and there are still single point of failure and lack of scalability problems.

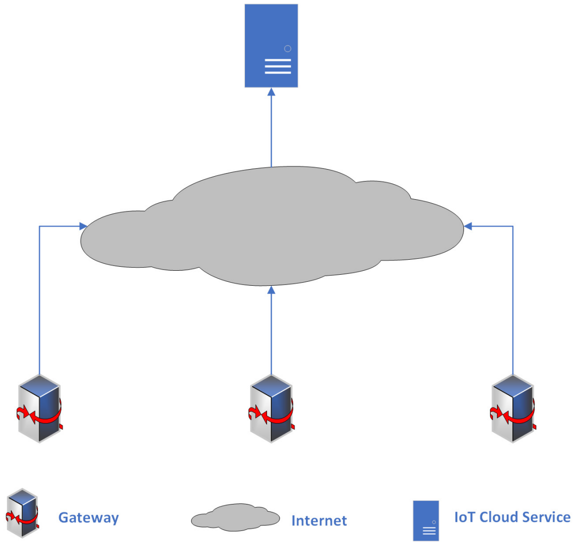

Fig. 1. The Centralized Architecture

II. IIS

In this paper, we propose the decentralized IIS solution. Compared with precious solutions, IIS has two more advantages: Firstly, since the training set of the identification model comes from the distributed nodes, it prevents the identification model of the gateway from external attacks, thus further ensuring the security of the system. Secondly, the completely distributed system ensures the scalability of the system.

Then, we introduce IIS in detail. Fistly, we will show the network architecture and workflow of IIS and then introduce the scheme step by step.

The workflow of IIS can be described in Fig.2. It can be seen that the whole process includes five phases:

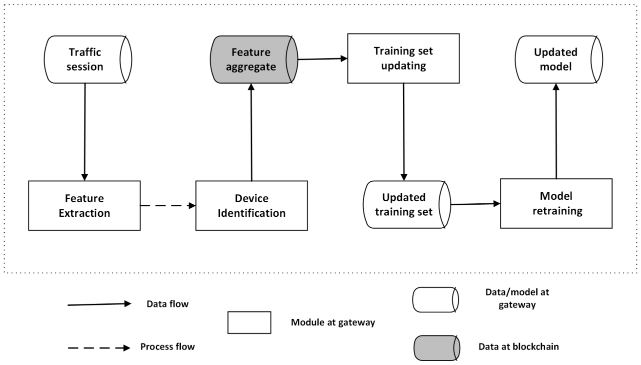

Fig. 2. The Workflow of IIS

- ***Feature Extraction***: Listen to a network traffic session of a device and extract feature vectors to represent the device type from it;
- ***Device Identification***: Identify the extracted feature vectors through deployed identification model and send the feature vector as a transaction to the ledger if it is identified as ”unknown”;
- ***Ledger Sharing***: Through consensus mechanism, the ledger, which is consisted of newly added feature vectors, is updated constantly and shared among all gateways synchronously;
- ***Model Training Set Updating***: All gateways update the model training set according to the ledger synchronously. Each gateway extracts the feature vectors newly added to the ledger, and then clusters these feature vectors with the original training set stored locally in the gateway to obtain a new model training set;
- ***Model Training***: The updated training set is used for model training to generate a new device identification model.

Concretely, when a new device accesses a gateway, the gateway listens to the traffic session and extracts the feature vector and then uses the identification model to identify it. If the type of the feature vector can be identified, the task is accomplished. If the type of the feature vector cannot be identified, the gateway will sends it to the blockchain network for recording on the ledger. Later, after the ledger of the gateway is updated, the gateway extracts the newly added feature vectors from the ledger and clusters these feature vectors with the original training set stored locally (the feature vector of the same new type are clustered in a feature set so that each feature set in the final training set uniquely represents a device type) to form a new model training set. Finally, the gateway stores this new training set locally and conducts model training on the training set to obtain a new identification model.

III. PERFORMANCE EVALUATION

In this section, we design experiments to evaluate the performance of proposed IIS. Firstly, we introduce the experimental parameters including the types we used and then analyze the cost of each phases. Finally, we show the identification performance of IIS.

A. Experimental Setup

The data set used in our experiment is from the one used by Miettinen, Markus et al. [1]. This data set contains 21 device types with 20 samples for each type, and each sample is a PCAP packet captured by the author at the initial stage of the device. Considering that the feature extraction phase proposed in this paper is to extract the traffic packets within one minute of the initial stage of the device but the sample packet capture time of several device types in [1] is less than one minute, we remove these device types and finally select ten devices as the data set of our experiment.

In order to further evaluate the identification performance of the identification model on each device. We take the whole data set as a training set, and then take all the samples of each device as the test set and inject them into the identification model after training. Finally, the ratio of the number of successful samples of the device to the total number of samples is the identification accuracy of the device. The results are shown in Fig.3. From the figure, we can see that only D-Linksensor’s identification accuracy does not reach 1, but its accuracy is also 0.95. Therefore, it can be concluded that the identification model in this paper has good identification performance on these 10 devices.

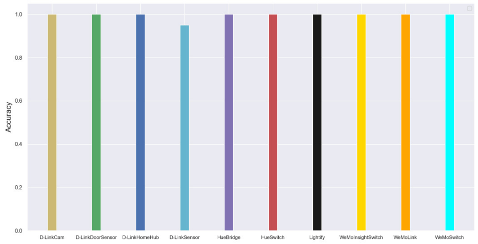

Fig. 3. Accuracy for 10 Device Types

IV. CONCLUSIONS

In this work, we propose IIS, an IIoT device identification scheme. This scheme solves the scalability and security problems in traditional centralized framework. The experiments show that the device identification system based on this scheme is effective for IIoT device type identification, and the identification system can continuously identify new device types.

REFERENCES

[1] M. Miettinen, S. Marchal, I. Hafeez, N. Asokan, A.-R. Sadeghi, and S. Tarkoma, “Iot sentinel: Automated device-type identification for security enforcement in iot,” in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2017, pp. 2177–2184.

[2] N. Aluthge *et al.*, “Iot device fingerprinting with sequence-based features,” 2018.

[3] H. Bojinov, Y. Michalevsky, G. Nakibly, and D. Boneh, “Mobile device identification via sensor fingerprinting,” *Computer Science*, 2014.

[4] S. Mandelli, D. Cozzolino, P. Bestagini, L. Verdoliva, and S. Tubaro, “Cnn-based fast source device identification,” *IEEE Signal Processing Letters*, vol. PP, no. 99, pp. 1–1, 2020.

[5] Jahoon, Koo, Se-Ra, Oh, Young-Gab, and Kim, “Device identification interoperability in heterogeneous iot platforms.” *Sensors*, 2019.

[6] L. Fan, S. Zhang, Y. Wu, Z. Wang, and J. Yang, “An iot device identification method based on semi-supervised learning,” in *2020 16th International Conference on Network and Service Management (CNSM)*, 2020.

[7] Z. A. Yin, A. Hw, A. Fq, A. Zw, and B. Ha, “ibike: Intelligent public bicycle services assisted by data analytics - sciencedirect,” *Future Generation Computer Systems*, vol. 95, pp. 187–197, 2019.

[8] S. Bratus, C. Cornelius, D. Kotz, and D. Peebles, “Active behavioral fingerprinting of wireless devices,” in *Proceedings of the first ACM conference on Wireless network security*, 2008, pp. 56–61.

[9] L. Hyun-Seong, L. Jae-gwang, L. Jae-pil, and L. Jae-kwang, “Design of automatic identification gateway system for different iot devices and services,” in *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 2018, pp. 2022–2025.

[10] Ericsson, “Ericsson mobility report, june 2020,” https://www.ericsson.com/en/mobility-report/reports/ Accessed October 5, 2020.

[11] S. Marchal, M. Miettinen, T. D. Nguyen, A.-R. Sadeghi, and N. Asokan, “Audi: Toward autonomous iot device-type identification using periodic communication,” *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1402–1412, 2019.

[12] V. Thangavelu, D. M. Divakaran, R. Sairam, S. S. Bhunia, and M. Gurusamy, “Deft: A distributed iot fingerprinting technique,” *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 940–952, 2018.

[13] M. Köse, S. Taşcioğlu, and Z. Telatar, “Rf fingerprinting of iot devices based on transient energy spectrum,” *IEEE Access*, vol. 7, pp. 18 715–18 726, 2019.

[14] T. Hiraoka, Y. Miyaji, and H. Uehara, “Device identification based on distortion of power amplifiers excited by swept sine,” *IEICE Communications Express*, 2020.

[15] S. Reinders, L. Lin, W. Chen, Y. Guan, and J. Newman, “Score-based likelihood ratios for camera device identification,” *Electronic Imaging*, vol. 2020, no. 4, pp. 215–1–215–8, 2020.

[16] A. Higaki, T. Kurokawa, T. Kazatani, S. Kido, and H. Okayama, “Image similarity-based cardiac rhythm device identification from x-rays using feature point matching,” *Pacing and Clinical Electrophysiology*, 2021.

[17] Y. Meidan, M. Bohadana, A. Shabtai, J. D. Guarnizo, M. Ochoa, N. O. Tippenhauer, and Y. Elovici, “Profiliot: a machine learning approach for iot device identification based on network traffic analysis,” in *Proceedings of the symposium on applied computing*, 2017, pp. 506–509.

[18] J. Kotak and Y. Elovici, “Iot device identification using deep learning,” *arXiv preprint arXiv:2002.11686*, 2020.

[19] S. M. Raphael, “Neural network model,” in *Image and Signal Processing for Remote Sensing XXVI*, 2020.

[20] Y. Liu, M. Peng, M. R. Swash, T. Chen, and H. Meng, “Holoscopic 3d microgesture recognition by deep neural network model based on viewpoint images and decision fusion,” *IEEE Transactions on Human-Machine Systems*, vol. PP, no. 99, pp. 1–10, 2021.

[21] J. Wu, J. She, Y. Wang, and C. Y. Su, “Position and posture control of planar four-link underactuated manipulator based on neural network model,” *IEEE Transactions on Industrial Electronics*, vol. 67, no. 6, pp. 4721–4728, 2020.

[22] M. Tahir, M. Hayat, and K. T. Chong, “Prediction of n6-methyladenosine sites using convolution neural network model based on distributed feature representations,” *Neural Networks*, vol. 129, 2020.

[23] H. Noguchi, M. Kataoka, and Y. Yamato, “Device identification based on communication analysis for the internet of things,” *IEEE Access*, vol. 7, pp. 52 903–52 912, 2019.

[24] F. Le, J. Ortiz, D. Verma, and D. Kandlur, “Policy-based identification of iot devices’ vendor and type by dns traffic analysis,” in *Policy-Based Autonomic Data Governance*. Springer, 2019, pp. 180–201.

[25] M. Miettinen, S. Marchal, I. Hafeez, N. Asokan, A.-R. Sadeghi, and S. Tarkoma, “Iot sentinel: Automated device-type identification for security enforcement in iot,” in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2017, pp. 2177–2184.

[26] N. Yousefnezhad, M. Madhikermi, and K. Främling, “Medi: Measurement-based device identification framework for internet of things,” in *2018 IEEE 16th International Conference on Industrial Informatics (INDIN)*. IEEE, 2018, pp. 95–100.

[27] S. A. Hamad, W. E. Zhang, Q. Z. Sheng, and S. Nepal, “Iot device identification via network-flow based fingerprinting and learning,” in *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. IEEE, 2019, pp. 103–111.

[28] A. Sivanathan, H. H. Gharakheili, F. Loi, A. Radford, C. Wijenayake, A. Vishwanath, and V. Sivaraman, “Classifying iot devices in smart environments using network traffic characteristics,” *IEEE Transactions on Mobile Computing*, vol. 18, no. 8, pp. 1745–1759, 2018.

[29] Y. Meidan, M. Bohadana, A. Shabtai, J. D. Guarnizo, M. Ochoa, N. O. Tippenhauer, and Y. Elovici, “Profiliot: a machine learning approach for iot device identification based on network traffic analysis,” in *Proceedings of the symposium on applied computing*, 2017, pp. 506–509.

[30] Y. Zhang, Y. Li, R. Wang, J. Lu, and M. Qiu, “Psac: Proactive sequence-aware content caching via deep learning at the network edge,” *IEEE Transactions on Network Science and Engineering*, vol. PP, no. 99, pp. 1–1, 2020.

[31] V. Thangavelu, D. M. Divakaran, R. Sairam, S. S. Bhunia, and M. Gurusamy, “Deft: A distributed iot fingerprinting technique,” *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 940–952, 2018.