Called Function Identification of IoT Devices by Network Traffic Analysis

Daichi Koike
Graduate School of Information Science and Electrical Engineering, Kyushu University.
Fukuoka, Fukuoka
koike.daichi@arakawa-lab.com

Shigemi Ishida
Graduate School/Faculty of Information Science and Electrical Engineering, Kyushu University.
Fukuoka, Fukuoka
ishida@ait.kyushu-u.ac.jp

Yutaka Arakawa
Graduate School/Faculty of Information Science and Electrical Engineering, Kyushu University.
Fukuoka, Fukuoka
arakawa@ait.kyushu-u.ac.jp

ABSTRACT

IoT devices are currently used in various situations, with the number of applications steadily increasing over the past few years. These IoT devices are connected to outside networks, and as such are often targets of hacking attempts and associated with private user information leaks. The end user is often unaware of how these devices operate, and as a result is unable to notice any unauthorized communication. In response, we aim to create a system to visualize the activity of an IoT device (IoT activity monitor). To create this system, we analyze the network traffic of an IoT device and propose a way to estimate which function it uses. Using WireShark, a packet capture software, we evaluate the data transmitted from the Amazon Echo Spot, Amazon Echo Dot and Amazon Echo Flex, which are smart speakers. We confirmed that we can estimate 8~10 kinds of called functions with 76.1 %, 89.8 % and 85.2 % accuracy for each smart speaker respectively.

CCS CONCEPTS

• **Security and privacy** → *Trust frameworks*; Usability in security and privacy; • **Human-centered computing** → Activity centered design; • **Computing methodologies** → *Supervised learning by classification*; Boosting;

KEYWORDS

IoT, Smart house, Traffic analysis, Trust, Packet capture

ACM Reference Format:
Daichi Koike, Shigemi Ishida, and Yutaka Arakawa. 2021. Called Function Identification of IoT Devices by Network Traffic Analysis. In *The 36th ACM/SIGAPP Symposium on Applied Computing (SAC '21), March 22–26, 2021, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, Article 4, 7 pages. https://doi.org/10.1145/3412841.3441951

SAC '21, March 22–26, 2021, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-8104-8/21/03…$15.00
https://doi.org/10.1145/3412841.3441951

1 INTRODUCTION

These days, IoT devices are used in a lot of different situations. According to the Japanese Ministry of Internal Affairs and Communications, the number of IoT devices in the world in 2017 is about 27 billion and in 2020 they predict it will be 40 billion[1]. Thus the number of IoT device will continue to increase. Famous IoT devices used in homes are, for example, smart speakers like Google Home and Amazon Echo. These devices have a voice user interface (VUI) and users can perform various functions such as Internet Search or Timer using voice commands. In addition, some devices have a camera for video call functionality with other devices. Whilst network cameras are not yet widespread in Japan, they are nevertheless easily available in home improvement stores. One of the features of this kind of IoT devices is integration with the cloud and smartphones. Most IoT devices are connected to their exclusive cloud system through a home WiFi network which collects and analyzes the data created and transmitted by the devices. This data is accessed through the user's smartphone, which is in turn used to control the devices.

IoT devices are designed on the premise that they are to be connected to an outside network meaning that the devices often become the targets of hacking and we have seen various privacy leak incidents happen. For example, we have discovered vulnerabilities with the camera attached to a smart vacuum cleaner[2], and witnessed incidents in which customer information was stolen from a casino through the smart thermometer in a water tank[3], and in which cameras all over the world became accessible because of the hacking of an IP camera. In addition, the number of adversarial attacks which target IoT devices have been increasing recently. For example, in [7][2] and [3], IoT devices are accessed and controlled remotely using sounds inaudible to humans. While the spread of IoT devices makes our lives more convenient, it's important to consider the privacy implications associated with the use of these devices.

There are many considerations to take into account when using IoT devices. Concerning security, in particular, there are three main considerations to bear in mind:

- There is a high diversity of IoT devices, making it difficult to update the security of all devices continuously. Currently, new devices are released in quick succession, but the rate of firmware updates doesn't keep up as opposed to what

[1] https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2018/chapter-1.pdf
[2] https://threatpost.com/robot-vacuums-audio-lidarphone-hack/161421/
[3] https://mashable.com/2018/04/15/casino-smart-thermometer-hacked/

we see with more traditional personal computers (PCs). Devices produced by larger, more established companies are more likely to receive consistent and regular firmware updates and support, however this is not always the case with devices created by smaller companies.

- The activity of an IoT device can be likened to a black box and as a result often works outside of the users' intention, who has no knowledge of what data is being transmitted by the device or where the data is being transmitted to. After connecting to a network using the device's initial settings, users usually do not know which server the device connects to, what kind of protocol it uses and how often it connects to the network. We have seen recently that there is a possibility that network communication goes through a certain country as with the recent Zoom incident[4]. The root of network communication is supposed to encrypted but that is not something general users are able to verify.
- Users are not able to install a fraud detection system like anti-virus software for PC.

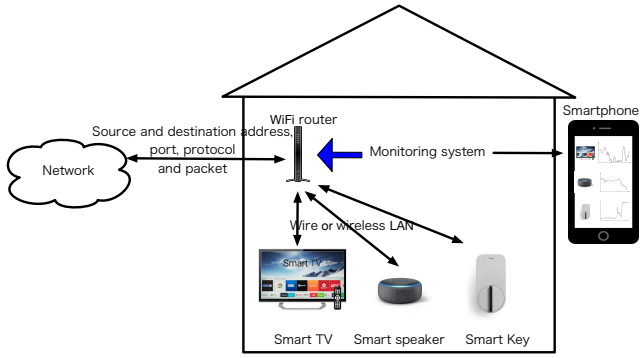
Therefore, in this paper, we propose an IoT activity monitor, which visualizes the activity of IoT devices. The IoT activity monitor aims to help users monitor and understand how an IoT device communicates with the network. We aim to solve the problems outlined above, check the activity of any device connected to a given network, and make a system where users can install a device in their homes with confidence. For example, with smartphones, when users install an application, the device asks whether the application can use a resource such as camera and microphone. In addition, users can check which functionalities are being accessed by any given application at any time: they are able to visualize the operation and behavior of the various installed applications. We aim to design a system that makes these functionalities found in smartphones available for IoT devices. Recently, devices which can perform various functions and work in connection with other devices like smart speakers have become popular. Although there is some research concerning the identification of IoT devices based on their traffic [4][5], no research identifies the called function of an IoT device. It's very difficult to detect unauthorized communication from these IoT devices using only device identification. However, IoT device identification as featured in the above research helps to detect unauthorized communication. Our proposed IoT activity monitor identifies not only the device type but also the function called by each device. To create the IoT activity monitor, we focus on the network traffic of the corresponding IoT device. On the premise that a home router collects all network traffic, we identify each device and its corresponding activity by monitoring the traffic passing through the router. Thus we aim to implement an IoT activity monitor by extending this work and available technologies.

In this paper, we observe network traffic by connecting an IoT device to the PC in which packet capture application "Wireshark" is installed.. In our implementation, we use three smart speakers: the Amazon Echo Spot, the Amazon Echo Dot and Amazon Echo Flex as IoT devices and we call 8~10 kinds of functions at 3~5 times repeatedly. We then collect the data. The size of the data obtained in this manner is 48.9MB. We evaluate identification accuracy by using a supervised learning algorithm, random forest, followed by stratified 10-fold cross-validation. We use 24 features including average, variance and standard deviation of packet length, as well as the number of occurrences of each of the 21 communication protocols. We divide traffic data where we use a function into separate windows, and extract the features from each individual window. We then identify the called functions. We set 30 data as the window size because the time between data isn't equal (WireShark only receives data when network communication occurs). We were able to identify 8~10 kinds of called functions with an average accuracy of 76.1 %, 89.8 % and 85.2 % for each smart speaker respectively.

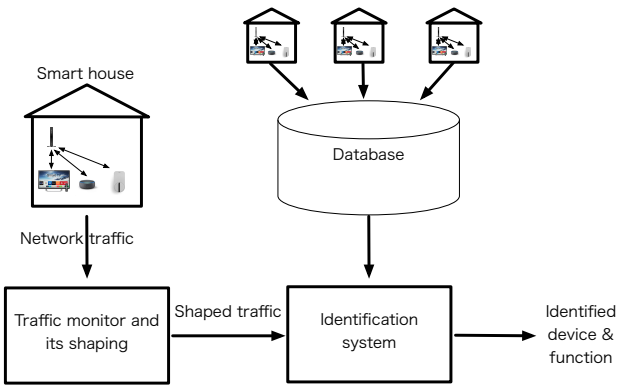
The rest of this paper is organized as follows: Section 2 shows the challenges and merits of creating the system. We present called function identification method of IoT Devices in section 3. Section 4 shows an initial evaluation. Section 5 describes related research in the field of IoT traffic analysis. Finally, we conclude the paper in Section 6.

2 SYSTEM OVERVIEW: CHALLENGES AND MERITS

2.1 Outline of Ideal System



(a) Outline of proposed IoT activity monitor platform



(b) Outline of identification system

Figure 1: Outline of IoT activity monitor

[4]https://techcrunch.com/2020/04/03/zoom-calls-routed-china/

Figure 1 shows the IoT activity monitor platform. In this research, we assume that a certain amount of IoT devices are set in a house and that they are connected via a wired or wireless connection to a router. Focusing on the router which all devices are connected to, we observe traffic information which is sent from the IoT devices and goes through the router. The data is then collected and used to identify the called function of the devices.

We define IoT activity monitor as a platform where we can register our own IoT devices and check their activity. With smartphones, when users install an application a pop-up such as "This application requests access to a function (e.g. microphone)" is displayed to the user, who can choose whether or not to allow it. In addition, they can see a list of applications with access to each function from the device's settings. As mentioned above, a function used by an application is visualized on the smartphone and can be used with trust. Therefore, we aim to visualize the activity of an IoT device in the same way as smartphones and display functions used by the IoT device as well as the time variation of communication packets. This will enable the end user to use the device without any security concerns. In this research, by focusing on the fact that IoT devices are connected to a cloud system through a home router and observing the router which all devices are connected to, we identify the function called by the device. Displaying the identified function to users through the cloud enables them to check the activity of the devices even if they are outside.

Figure 1b shows the system we aim to implement. By introducing the platform described above to each home and sharing an algorithm which identifies the called function of an IoT device, we can collect enough network traffic data and identify it with high accuracy. We use supervised learning as the algorithm. In addition, by carrying out questionnaires about called function times for users, we update the model regularly with collected data and labels. In this paper, we present the identification part of the system outlined previously. One of the challenges of this research is to get device information without direct access to it. Function identification by getting data with direct access to the device is not realistic: each device has different specifications and performing identification for individual devices is extremely expensive. Additionally, we have to make a new identification model at every single release time. Two-step identification is also a challenge of this research: to identify a function, we have to identify the type of device first.

2.2 Merit

One of the merits of this research is making the detection of unauthorized communication straightforward by visualizing IoT device activity. Currently, the activity of a given IoT device is like a black box and unauthorized communication can happen. Visualization of IoT device activity can prevent this. Although the research which identifies the type of IoT device has already been done, an IoT device has many functions these days, and thus function identification enables us to detect unauthorized access in more detail. Even if there is no unauthorized access for a user, invisible and unmonitored activity can make them uncomfortable. In addition, when used with other research techniques, this can help contribute to identify whether communication is intentional or not.

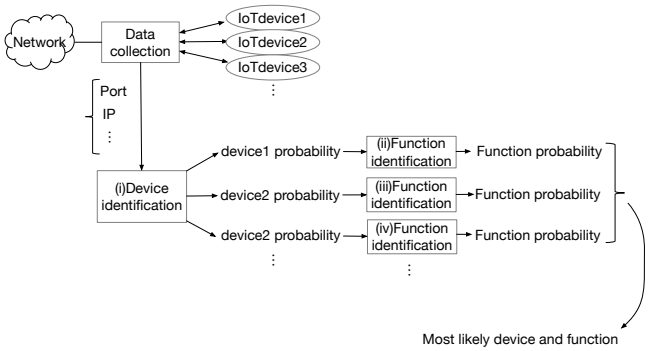


Figure 2: Detail of identification system

Here, we propose a called function identification model, but the vendor of a given IoT device may make its traffic model available to users in the future, in which case people may think our identification model is unnecessary. However, it is quite unlikely that all vendors will make available the models for all their devices. Currently, there are many devices for which vendors don't offer firmware updates, and are unlikely to in the future. The system proposed in this paper acts as a defense mechanism for devices that are not covered buy "in-built" defensive functionalities.

3 CALLED FUNCTION IDENTIFICATION BY NETWORK TRAFFIC ANALYSIS

3.1 Key Idea and System Overview

A key idea of called function identification by traffic analysis is getting the traffic data of an IoT device from the home router which all network communication goes through in conjunction with a machine learning algorithm. Here, the traffic pattern and connecting IP address differ depending on the function in use, even though the device is the same. Focusing on this difference, we extract features from the traffic data obtained from the home router and identify the called function using the algorithm.

The proposed system consists of three blocks, which are a data collection block, a device identification block and a function identification block. Firstly, in the data collection block, we get the data which is needed to identify the called function and reshape it. Then we identify the called function using the rest of the blocks. Here, before identifying the called function, we first need to identify the device type. Thus we make model (i) which identifies the type of IoT devices and model,and model (ii)∼(iv) which identifies called functions. Secondly, in the device identification block, we first input collected traffic data into model (i) and calculate the probabilities for each device. Then, in the function identification block, we input the traffic data to model (ii)∼(iv) and calculate the probabilities for each function. Finally, by comparing the values obtained by multiplying these two different probabilities, we are able to identify the most likely called function and used device. Figure 2 shows this three-step identification process.

3.2 Data Collection

In our proposed method, we identify the called function by analyzing the network traffic pattern using machine learning. To create the identification model, we collected data from an Amazon Echo Spot, Amazon Echo Dot and Amazon Echo Flex, i.e. smart speakers. In the data collection block, we capture packets in a WiFi router to which smart speakers connect while we call a variety of functions from the smart speakers.

The functions we identify this time are Kindle, Amazon Music, video call, question, news, Amazon Prime Video, restaurant search, Spotify, TuneIn and voice call. However, for Amazon Echo Dot and Amazon Echo Flex, we did not get data for video call and Amazon prime video as these devices don't have a screen and cannot call these functions. We call each function 3~5 times for about 155 seconds. We define the data of Amazon Echo Spot, Amazon Echo Dot and Amazon Echo Flex as the letters "a", "b" and "c" respectively. We also define the data of Kindle, Amazon Music, question, news, restaurant search, Spotify, TuneIn, voice call, video call and Amazon Prime Video as a number from 1 to 10. Then we express the letter and number together, for example, the data corresponding to the kindle function for the Amazon Echo Spot is expressed as a-1.

3.3 Function Identification

During function identification, we use the data of each device. For example, when we train function identification on the Amazon Echo Spot, we use a-1~a-10. Firstly, we exclude the error packets and put a label such as "video call" and "voice call" on it. Secondly, we extract idle time from the data and create an "idle time" label as a seperate function. Here, we define idle time as the time when a function is not called before the first call or between each call, which means 5 seconds from the initially received data communication where the size is less than 75 bytes.

We then calculate the features for machine learning which can be divided into two categories. The first category is features contained in each individual packet and the other is features contained in each individual window. In the first category, we use protocols, port, source address and destination address. We use one-hot encoding for this traffic data and create data whether the communication exists or not on each data point. In the second category, we calculate average, variance, standard deviation, maximum, minimum and the difference between the maximum and the minimum packet length for 1 window. The reason we apply a window is because the time variation of network traffic is unique for each function. We set 30 packets as 1 window because we want our packets to be as long as possible to reflect the time variation, however we are limited by our amount of data. Thus we think 30 packets is optimal. Our windows are created by sequentially grouping packets into overlapping groups of 30: our first window consists of packets 1 to 30, our second window of packets 2 to 31, and so on.

We use the above data as features for machine learning. We created estimation models based on a selection of supervised learning algorithms and evaluated the model with test data. Finally, we compared these models and chose the most accurate model as our estimation model. We use the random forest, XGBoost, LightGBM and CatBoost machine learning algorithms. Then we choose the most accurate algorithm for each identification and model.

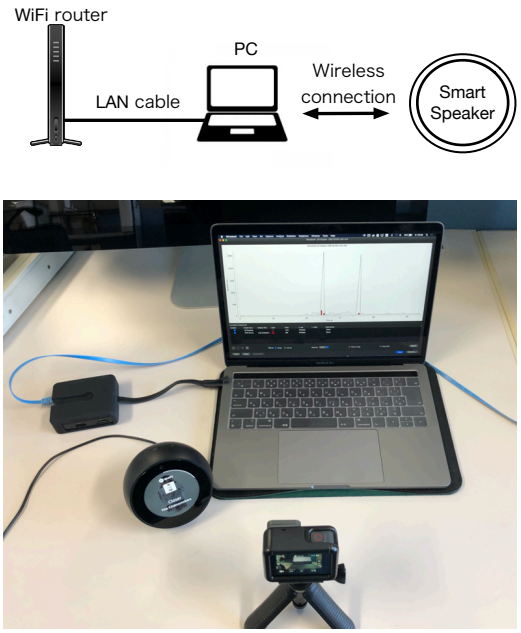


Figure 3: Experiment environment

3.4 Device Identification

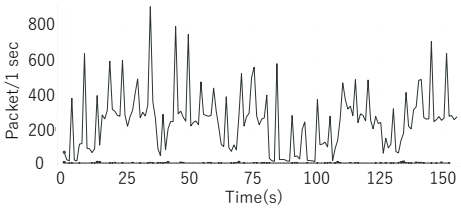
During device identification, we use all the data. Here, the only difference between making models for function identification and device identification is the labeling. To make the function identification model, we use function labels such as video call and question. On the other hand, to make the device identification model, we use device labels such as Amazon Echo Spot and Amazon Echo Dot.

4 EVALUATION

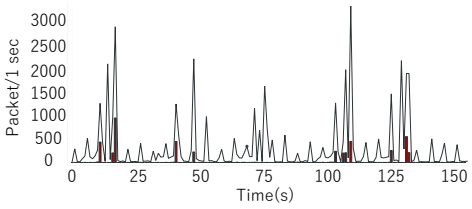
4.1 Data Acquisition

Figure 3 shows the experimental environment. We use the Internet sharing function of macOS to make a WiFi access point which can capture packets. Setting a PC equipped with WireShark, i.e. a packet capture application, as an access point, we can capture all network traffic which goes through the PC. We connect a smart speaker Amazon Echo Spot to the PC to capture all network traffic.

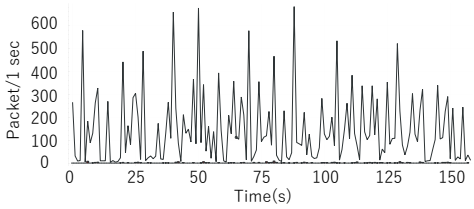
During data collection, we called various functions of Amazon Echo Spot 3 to 5 times and captured network traffic for 155 seconds on each trial. Figures 4a~4h show the number of packets per second as a function of time for the functions Kindle, Amazon Music, question, news, restaurant search, Spotify, TuneIn, voice call, video call, and Amazon Prime Video, respectively. We can confirm that when we call Amazon Music, packet length highly increases immediately after calling the function and then decreases to a lower value. On the other hand, when we call video call, packet length is much lower compared with that of Amazon Music and keeps to an almost constant value. We can assume that the traffic pattern, i.e., the traffic volume as a function of time, is dependent on functions we call.



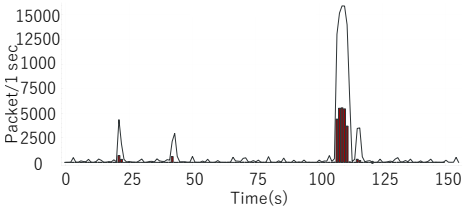
(a) Kindle



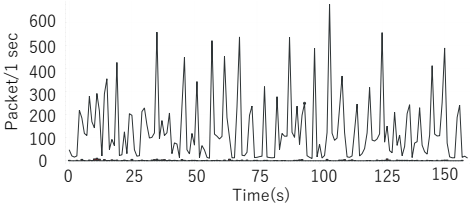
(b) AmazonMusic



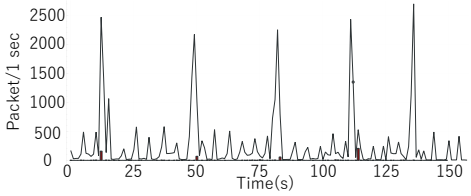
(c) Question



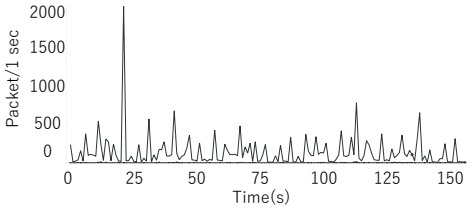
(d) News



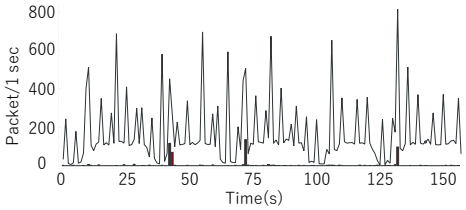
(e) Restaurant search



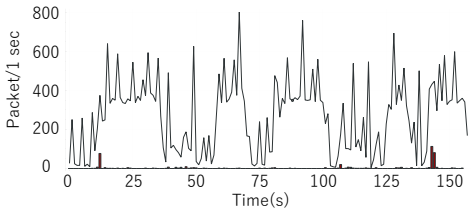
(f) Spotify



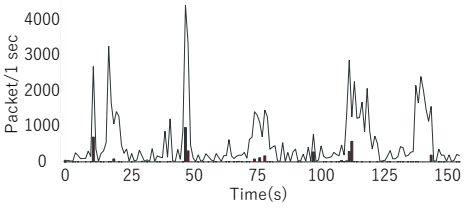
(g) TuneIn



(h) voice call



(i) video call



(j) Amazon Prime Video

Figure 4: packet graphs

4.2 Evaluation for Function Identification

To confirm the effectiveness of our called function identification method, we evaluated function identification accuracy through stratified 10-fold cross-validation using the data a-1~a-10. We use random forest, XGBoost, LightGBM and CatBoost as machine learning algorithms in the function identification block. We then choose the most accurate random forest algorithm as the function identification model.

Figures 5a~5c show heat maps of function identification result confusion matrices for the Amazon Echo Spot, Amazon Echo Dot and Amazon Echo Flex respectively. The total accuracy values of function identification for the Amazon Echo Spot, Amazon Echo Dot and Amazon Echo Flex are 76.6 %, 89.8 % and 85.2 % respectively. We can identify video call and voice call with relatively high accuracy. This is because of close-to-constant packet lengths during calling of the functions. Function identification accuracy is slightly lower for the question function. This is because we used various types of questions. We ask four kinds of questions, which are about weather, translation, calculation and general questions such as "Who is the Google CEO?" These questions require different types of information. For example, when we ask about weather, the device needs to access location information, while other functions require no location information.

We believe that the relatively low accuracy of function identification is mainly caused of insufficient data compared to the number of target functions. In our future work, we will improve the accuracy by collecting more data and will further consider proper preprocessing.

4.3 Evaluation For Device Identification

To confirm the effectiveness of our device identification method, we evaluated device identification accuracy using stratified 10-fold cross-validation with the random forest algorithm. Firstly we combine all the data. Then we split the data to equalize the amount of data for each label.

Figure 6 shows the heat map of IoT device identification result confusion matrix. The accuracy of device identification is 99.1 %. Compared with function identification, we can identify device type with high accuracy because we use the IP address as a feature which is critical in device identification. In addition, the number of types to identify in device identification is much smaller than in function identification, which results in identification at a higher accuracy.

5 RELATED WORK

5.1 Research Describing End User Security and Privacy Concerns with Smart Homes

The work in [6] shows end user security and privacy concerns with smart homes. They conduct interviews with fifteen people living in smart homes to learn about how they use their smart homes and to understand their security and privacy related attitudes, expectations, and actions. In their opinion, based on these interviews, users are not so interested in the security of a smart home device. However, they claim that creating a device information visualization system would be a potential way to incite interest in device-related security concerns for the end user. Thus our research helps to not only detect unauthorized communication but also decreases the number of users who are not so interested in the security.

5.2 Research Describing Vulnerabilities of IoT Communication Privacy

The work in [1] shows privacy vulnerabilities of encrypted IoT traffic. By analyzing four smart home devices available for commercial purposes (Sense:sleep monitor, Nest cam:indoor security camera, Wemo switch:remote switch, Amazon Echo:smart speaker), they show that the rate of network traffic can reveal users' activity. This is because users' behavior can be estimated using only the transmission and reception rate of encrypted traffic, as IoT devices transform real-world information into network traffic. Thus they warn the users about any potential privacy threats. Of course while it's important to protect traffic information which could enable potential attackers to estimate users' behavior, it's also important to visualize activity information and report it to users for security monitoring purposes. We think this research, i.e. presenting activity information to users by analyzing network traffic of IoT devices, helps to detect suspicious network communication.

5.3 IoT Device Identification by Network Traffic Analysis

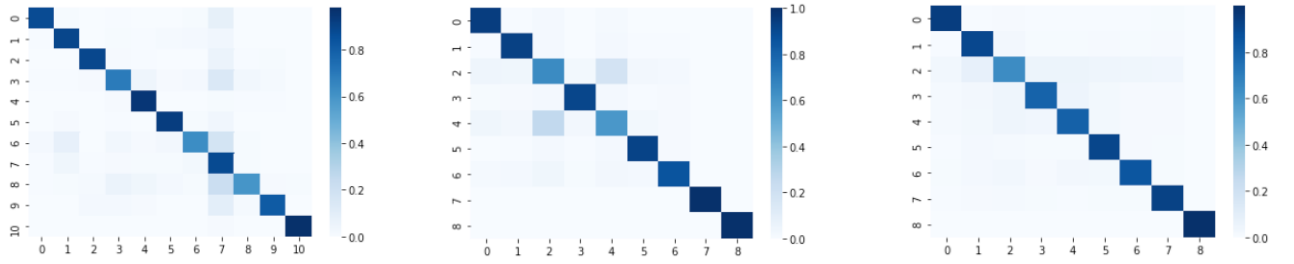
While we identify a function by analyzing the network traffic of an IoT device in our research, there exists previous research which identifies IoT devices.

The work in [4] shows a method for IoT device and non-IoT device identification using network traffic analysis with machine learning. Analyzing a saved file which contains traffic information of devices connected to WiFi, the authors identify the devices in two stages using supervised machine learning while abstracting features such as source address, destination address and port number. In the first stage, they identify whether it's a IoT device or not. In the second stage they identify a device class from a list of registered identified IoT devices. In this research, they identify types of IoT devices with 99.281 % accuracy but don't identify the functions used by the devices.

The work in [5] shows a method of IoT device identification in a smart city and and on a campus. They set 21 IoT devices on a campus and collect traffic data for 3 weeks. Then, analyzing wide network traffic (e.g. traffic load, types of signal and distribution of active time), they identify the devices through supervised learning. In this research they identify types of IoT devices with 95 % accuracy but again don't identify the functions used by the devices.

6 CONCLUSION

In this paper, we proposed a method to identify the called functions of IoT devices using traffic analysis with the aim of visualizing device activity. Our system is able to identify both the devices, and the types of functions being called. We obtained traffic data from the Amazon Echo Spot, Amazon Echo Dot and Amazon Echo Flex, which are all smart speakers. We identified 8~10 types of called functions with an accuracy of 76.1 % for Amazon



(a) Amazon Echo Spot (b) Amazon Echo Dot (c) Amazon Echo Flex

Figure 5: Heat maps of function identification result confusion matrices (0:Kindle, 1:AmazonMusic, 2:question, 3:News, 4:restaurant search, 5:Spotify, 6:TuneIn, 7:voice call, 8:idle, 9:video call, 10:AmazonPrimeVideo)

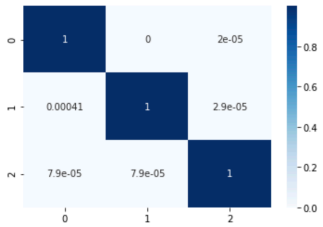


Figure 6: Heat map of device identification result confusion matrix(0:EchoSpot, 1:EchoDot, 2:EchoFlex)

Echo Spot, 89.8 % for Amazon Echo Dot, 85.2 % for Amazon Echo Flex. We identified device and function by choosing the combination with the highest accuracy probability from the values obtained by multiplying the two kinds of probabilities. This increased system accuracy. In addition, the three devices we use this time are all smart speakers developed by Amazon, which implies that the devices might have similar traffic patterns. In actual home environments, installing similar devices is rare and people install different kinds of IoT devices. Thus in the actual case, we expect that our identification system can achieve higher accuracy. Furthermore, as we mentioned in Section 2, we can further improve the accuracy by updating the identification models with the data acquired in many homes.

In future research, we plan to identify the called function for other IoT devices, and to monitor network traffic and notify users about which function is currently being used in real time by connecting the system to a messaging application such as Slack. Additionally, we plan to identify whether the network communication is intentional with other research techniques such as behavior recognition using WiFi. Through the use of activity visualization, we hope that users will be able to use IoT devices without any security concerns.

ACKNOWLEDGMENTS

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP19KT0020.

REFERENCES

[1] Noah Apthorpe, Dillon Reisman, and Nick Feamster. 2017. A smart home is no castle: Privacy vulnerabilities of encrypted iot traffic. *arXiv preprint arXiv:1705.06805* (2017). https://arxiv.org/abs/1705.06805

[2] Anirban Chakraborty, Manaar Alam, Vishal Dey, Anupam Chattopadhyay, and Debdeep Mukhopadhyay. 2018. Adversarial attacks and defences: A survey. *arXiv preprint arXiv:1810.00069* (2018). https://arxiv.org/abs/1810.00069

[3] Jian Mao, Shishi Zhu, Dai Xuan, Qixiao Lin, and Jianwei Liu. 2020. Watchdog: Detecting Ultrasonic-based Inaudible Voice Attacks to Smart Home Systems. *IEEE Internet of Things Journal* (2020). https://ieeexplore.ieee.org/abstract/document/9099849

[4] Yair Meidan, Michael Bohadana, Asaf Shabtai, Juan David Guarnizo, Martín Ochoa, Nils Ole Tippenhauer, and Yuval Elovici. 2017. ProfilIoT: a machine learning approach for IoT device identification based on network traffic analysis. *Proceedings of the symposium on applied computing* (2017), 506–509. https://dl.acm.org/doi/10.1145/3019612.3019878

[5] Arunan Sivanathan, Daniel Sherratt, Hassan Habibi Gharakheili, Adam Radford, Chamith Wijenayake, Arun Vishwanath, and Vijay Sivaraman. 2017. Characterizing and classifying IoT traffic in smart cities and campuses. *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (2017), 559–564. https://ieeexplore.ieee.org/abstract/document/8116438

[6] Eric Zeng, Shrirang Mare, and Franziska Roesner. 2017. End user security and privacy concerns with smart homes. *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)* (2017), 65–80. https://www.usenix.org/conference/soups2017/technical-sessions/presentation/zeng

[7] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2017. Dolphinattack: Inaudible voice commands. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (2017). https://dl.acm.org/doi/abs/10.1145/3133956.3134052