Autonomous IoT Device Identification Prototype

Nesrine Ammar
Nokia Bell Labs
Nokia Paris-Saclay
Nozay, France
nesrine.ammar@nokia-bell-labs.com

Ludovic Noirie
Nokia Bell Labs
Nokia Paris-Saclay
Nozay, France
ludovic.noirie@nokia-bell-labs.com

Sebastien Tixeuil
Sorbonne Université
CNRS, LIP6
Paris, France
sebastien.tixeuil@lip6.fr

***Abstract*—In this paper, we demonstrate a prototype implementation to help identifying the types of IoT devices being connected to a home network. Our solution is based on a supervised classification algorithm (*decision tree*) trained on 33 IoT devices using relevant information extracted from network traffic. Our demo shows that our proposal is effective to automatically identify the types of IoT devices.**

***Index Terms*—Internet of Things, device identification, network protocols, network traffic, decision tree.**

I. Introduction

With the proliferation of smart objects, people buy Internet of Things (IoT) devices and home appliances to benefit from new services enabling them to be informed about and control their houses anytime and anywhere [1]. Several researches and industrial works are still questioning how to identify devices in IoT environments to better manage them and reduce the user involvement to properly and securely use them.

However, there is no clear solution for autonomous identification of devices connected to home networks that spans sufficiently many types of IoT devices in such heterogeneous environments. Existing hardware-based solutions are not efficient enough because the same hardware components and drivers may be deployed as different device types. Furthermore, other solutions such as web fingerprinting, OS and applications detections, can only handle specific types of devices.

In this demonstration, we present an implementation of a prototype to identify the types or models of wired and wireless IP enabled devices. Our prototype is based on textual information that is obtained from network protocols and relevant traffic flow statistics, using a trained *decision tree* model. The relevant textual information is retrieved from packets sent by the device. Traffic flow characteristics includes packet lengths, packet inter-arrival times and used protocols. The evaluation of our solution on the test set shows high identification accuracy with 98% precision and 98% recall on average.

This paper is structured as follows. First, we discuss in Section II the methodology and the feature selection, and we compare with related works. Section III provides the architecture and the implementation of our device identification prototype. Then, Section IV describes the demonstration itself. Finally, Section V concludes this paper by focusing on the next research work.

This work was partly carried out at the LINCS (Laboratory of Information, Networking and Communication Science, http://www.lincs.fr/)

II. Device classification methodology

Our main objective is to identify the types of devices when they newly connect to the home gateway, to help end users better manage their devices and obtain more services from them. We thus extract features from the first packets emitted or received by the devices during their setup phase, and later use them to classify devices according to their types.

*a) **Features selection**:* Meidan *et al.* [2] used network statistics extracted from a full TCP session to obtain the feature vector. Thus, one has to wait until the end of the TCP session to extract the feature vector. Furthermore, during the setup phase some devices do not use TCP to communicate with the gateway. Miettinen *et al.* [3] used 23 features extracted from the first 12 packets exchanged between the device and the gateway. Their results show good precision for 17 out of 27 devices, but a precision around 0.5 for the remaining 10.

By contrast, our solution is based on features coming from many protocols, it is also scalable and is able to work on encrypted packets. The first set of features is extracted from flow characteristics: packet length inter-arrival time of the flow, flow's size, protocols used by the flow. The second set of features is extracted from device's description shared in network payload, such as the ones used by Ammar *et al.* [4]: manufacturer name from MAC address, device name from DHCP information, manufacturer, model and type from UPnP messages during the discovery process, device local name, services names offered by the device from mDNS records, device OS, model and in some cases type from the user-agent of the HTTP header. Finally, the textual features are presented by a binary bag of words model, set to 1 if the word is present in the device description and to 0 if not.

*b) **Device classification**:* We build a single classifier for each device class, using decision tree[1] models. We tested our solution using 28 IP enabled devices, with traffic captures of 22 devices from Miettinen *et al.* [3] and 6 devices from our laboratory. Based on features selected in this work, our results show a high accuracy (on average 98%). The trained models illustrate the importance of the textual features we use in this work. The models trained to distinguish between devices traffic are used to identify the devices being connected to the home network at a specific moment to feed higher level applications.

[1]We tested other classification algorithms, decision tree showed the best performances, so this is the one we use for this demo.

978-3-903176-17-1 / © 2019 IFIP

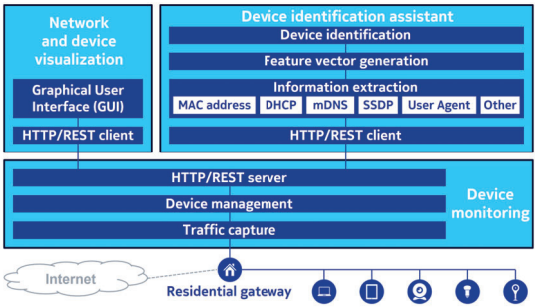

Fig. 1. Architecture of the IoT device identification assistant

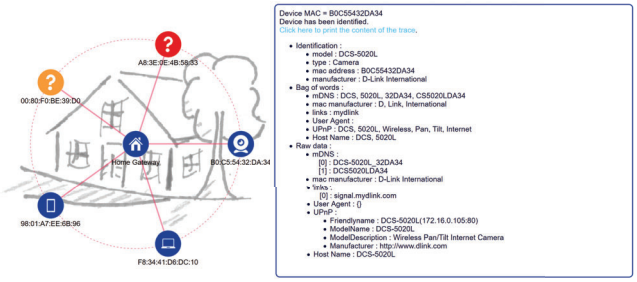

Fig. 2. IoT device identification assistant Graphical User Interface

III. Architecture and Implementation

We implemented the global architecture of our IoT device identification assistant presented in Fig. 1:

1) The *home network* contains several devices that are connected through Ethernet or Wi-Fi to a residential gateway that is connected to the Internet.
2) The *device monitoring* software module monitors the Ethernet and Wi-Fi traffic from the devices connected to the home network. It includes the following components (using *node.js* with *Win10Pcap*):
   - The *traffic capture* captures the first packets from any newly connected device, which is identified by a newly observed MAC address;
   - The *device management* manages the list of the devices connected to the home network;
   - The *HTTP/REST server* allows the communication of information to the other software modules.
3) The *device identification assistant* is the main software module of the demo, with the following components (using Python with *scikit-learn* library):
   - The *HTTP/REST client* links to the device monitoring through HTTP/REST;
   - The *information extraction* extracts the relevant textual information from the first packets sent by newly connected devices and their traffic flow statistics;
   - The *feature vector generation* represents the extracted data as a feature vector, using a Bag of Words (BoW) for textual information;
   - The *device identification* identifies newly connected devices using a supervised classification algorithm (*decision tree*) already trained on 33 IoT devices.
4) Finally, the *network and device visualization* software module has a *HTTP/REST client* component to get the information from the device monitoring server, which is used by the *Graphical User Interface* to visualize the results in a web browser (using HTML and JavaScript).

IV. Demonstration

To illustrate the overall solution and its benefits, the presented demo showcases the automated device-type identification. It involves a smart environment that illustrates a real deployment of a smart home network, with a home gateway and several devices connected through Ethernet or a Wi-Fi.

In a first step, the IoT device identification assistant presents the IoT devices being connected to the network. As shown as an example in Fig. 2, the identification process is illustrated on several devices of our lab: D-Link Camera, Panasonic Camera, iPhone SE Smart phone, Asus Tablet, Lenovo Laptop. The list of connected devices is updated by the device monitoring component each time a newly connected device is detected through its MAC address and the icon of newly connected device is set to red. Once the device monitoring has captured the first packets sent by the newly connected device, the device icon turns orange as illustrated for one device in Fig. 2.

The identification process can then start. Once achieved, the newly connected device is identified, its icon turns blue with the right type presented and its identification information can be displayed with some BoW representation as illustrated for the camera D-Link in Fig. 2.

V. Conclusion

Our device identification assistant aims to automate the identification of devices being connected to a home network in order to help users better manage their devices and benefit from new services. With our demo prototype, we visually and practically demonstrate the efficiency of our solution to identify the connected devices: one simply connects a device to the network and it appears on the GUI with all collected and infered information about its type.

This demo prototype is meant to be further deployed in various home network settings to strengthen its classifying abilities. We plan to facilitate its dissemination for research purposes in other hardware and software IoT environments.

References

[1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer Networks, vol. 54, no. 15, pp. 2787-2805, 2010.

[2] Y. Meidan, M. Bohadana, A. Shabtai, J. D. Guarnizo, M. Ochoa, N. O. Tippenhauerand 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, SAC '17, pages 506-509, New York, NY, USA, 2017. ACM.

[3] M. Miettinen, S. Marchal, I. Hafeez, N. Asokan, A. 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, pages 2177-2184, June 2017.

[4] N. Ammar, L. Noirie, and S. Tixeuil, "Identification du type des objets connectés par les informations des protocoles réseaux," in Rencontres Francophones sur la Conception de Protocoles, l'Évaluation de Performance et l'Expérimentation des Réseaux de Communication, CoRes 2018, Roscoff, France, May 2018.