Systematically Quantifying IoT Privacy Leakage in Mobile Networks

Shuodi Hui, Zhenhua Wang, Xueshi Hou, Xiao Wang, Huandong Wang, Yong Li, *Senior Member, IEEE,*
Depeng Jin, *Member, IEEE,*

Abstract—Privacy leakage of Internet of Things (IoT) has become a great challenge with the popularity of IoT services through mobile networks, such as smart homes, wearables, and healthcare. While previous work summarized general structures to analyze IoT privacy and provide case studies of specific devices or scenarios, it is still challenging to conduct a comprehensive and systematic quantification study of large-scale IoT privacy leakage in real world. To combine systematic analyses with real-world measurements, we provide a method to quantify IoT privacy leakage on a large-scale mobile network traffic dataset containing 47,651 IoT devices. We generate privacy fingerprints, and attribute them to a privacy quantification framework. The framework is constructed based on the semantics of multiple privacy sensitive markers selected from the traffic along with the involved network entity types in IoT (i.e., user, device, and platform), and the fingerprints are generated from sensitive information extracted in the traffic via their markers. Our quantification shows that IoT users, devices, and platforms have considerable risks respectively. Moreover, IoT devices have a larger scale of privacy leakage than users and platforms, and they perform different daily patterns on privacy leakage following their working conditions. In addition, we present three case studies on the leakage of location information, application calling, and voice service, which illustrate that a third party can profile a network entity in both cyberspace and physical space.

Index Terms—IoT, privacy, security, mobile network.

---◆---

1 INTRODUCTION

INTERNET of Things (IoT) brings great convenience to users by connecting various devices to the Internet and enabling them to perform their functions independently and even intelligently. These sensors, actuators, or computing devices connected to IoT and remotely controlled through the Internet are defined as IoT devices. With the rapid proliferation of IoT, increasing devices join the network, and Cisco forecasts that machine-to-machine (M2M) connections will grow to 14.6 billion by 2022, with 1.8 M2M connections for each member of the global population [1]. Through IoT, the devices and operators are collecting, transmitting, processing, and storing all kinds of data, which probably contains sensitive information and thus leads to privacy leakage. For instance, motivated by the growing number of Internet-connected TV devices, Moghaddam *et al.* [2] found that 89% of Amazon Fire TV channels and 69% of Roku channels have trackers to collect users' viewing habits and preferences. In addition, Celik *et al.* [3] identified sensitive data flows in 138 of 230 SmartThings market applications (60%). Wood *et al.* [4] also detected cleartext information that may reveal sensitive medical conditions in network traffic of four popular consumer medical IoT devices.

Meanwhile, even small embedded IoT devices with sensors (e.g., wearable devices) are continuously collecting all kinds of information from its surroundings and transmitting it to remote servers for further analysis [5]. Sikder *et al.* [6] found that the sensors on IoT devices can reveal sensitive data like passwords, secret keys of a cryptographic system, credit card information, etc, which can be used directly to violate user privacy or for future attacks. Apthorpe *et al.* [7] also demonstrated that an Internet service provider (ISP) or other possible attackers can infer privacy sensitive in-home activities by analyzing the changes of Internet traffic rates from several smart home devices even when these devices encrypted their traffic. Moreover, Ling *et al.* [8] found that launching different kinds of attacks can help obtain the user authentication credentials of a smart plug system, which can lead to both security and privacy concerns. Under this circumstance, IoT privacy gains increasing attention.

More extensive than the traditional definition of human privacy, IoT privacy is the ability to dispose the sensitive or valuable information collected, processed, and transmitted by IoT from individuals, groups, environments, workplaces, etc. Concerning privacy issues in IoT and the methods to figure them out, a number of analysis studies have been done from different perspectives. As privacy issues originate from security problems in most cases, some researchers started to mention IoT privacy leakage in their study on IoT security at first: Kozlov *et al.* [9] analyzed the known and new threats for the security, privacy and trust at different IoT architecture levels; Zhao *et al.* [10] and Mahmoud *et al.* [11] discussed the IoT typical architecture of three layers, i.e., perception, network, and application layers. Further, Roman *et al.* [12] expounded the challenges of IoT security and privacy under different distributed approaches of architecture and attack models. For IoT privacy respectively, Porambage *et al.* [13] explored IoT privacy concerns on

S. Hui, Z. Wang, X Wang, H. Wang, Y. Li and D. Jin are with Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing 100084, China.
Xueshi Hou are with Dept. of Electrical and Computer Engineering, University of California, San Diego.
E-mail: liyong07, jindp@tsinghua.edu.cn.

multiple IoT applications (e.g., healthcare, smart homes, public safety, supply management), presented four key IoT privacy aspects of user privacy, data mining, underlying technologies, and legal regulations. Apart from the literature above, privacy analyses on real-world IoT devices and traffic become popular, especially the devices interacting with users directly. Loi *et al.* [14] developed a systematic method to identify the security and privacy disadvantages of various consumer IoT devices by a test suite including lightbulb, switch, camera, printer, smoke alarm, and sleep monitor in smart homes; and Chu *et al.* [15] investigated the security and privacy of Internet-connected smart toys for children via case studies on three commercial products.

The literature above provides general structures and detailed instances on IoT privacy, but there are still some hurdles to conduct a large-scale and systematic quantification study of IoT privacy leakage. The general analyses above presented the security vulnerabilities or attacks leading to privacy leakage in IoT, and found the privacy risks under multiple IoT applications or scenarios. However, they lack support from accurate measurement experiments. The detailed instances illustrated privacy issues for specific kinds of IoT service or device via tests or case studies, but failed to quantify privacy leakage in universal IoT from macro perspective. To fill this gap, we intend to combine a systematic framework with quantitative measurements, which give consideration to both macro structure and accurate details of IoT privacy leakage.

The network nodes in IoT, which are electronic devices or communication endpoints attached to IoT, are creating, transmitting, or receiving privacy information through network. Moreover, most nodes in IoT tend to access the mobile network for its ubiquity and convenience, which leads to a potential danger for sensitive data to be captured by a third party along with the space-time coordinates of these nodes from GPS or base stations. Xia *et al.* [16] illustrated that third parties (e.g., hackers, cyber criminals, rogue employees in a cellular service provider (CSP) or ISP) can crawl data and gather digital footprints from mobile network users not simply by tapping into the wire directly, but by extracting information from the Web. Especially for IoT, the data generated by IoT devices usage is gathered and processed by third parties for all kinds of services (e.g., advertising and tracking) without being fully aware [17], [18], [19], and an attacker may eavesdrop on the communication of IoT nodes and extract the available unencrypted contents in the traffic [17], [20]. Then, not only service providers, but also malicious third parties, can collect the footprints in both cyberspace and physical space to profile a node in the IoT, which results in illegal monitoring, financial risks or even personal safety threats. Consequently, we propose to quantify IoT privacy leakage systematically from the traffic, and reveal the privacy risks to IoT consumers. Based on the previous works and their limitations, we find the following challenges of conducting a large-scale and real-world quantification:

- Multifarious IoT services, applications, devices and scenarios lead to various privacy sensitive information, which brings difficulty to construct a complete framework to quantify them coherently. The typical IoT architechture is better suited to security analysis, but unapplicable to privacy quantification. Meanwhile, considering from dozens of IoT applications or devices can hardly contain all of the privacy issues.

- We lack the knowledge of IoT privacy without direct interaction in human activities. In addition to the human-involved scenarios like smart home, privacy risks exist in many other IoT scenarios, such as smart agriculture, automated manufacturing, geological prospecting, safety monitoring, etc. However, most of the existing research focuses on human privacy, and it is hard to describe privacy in unmanned IoT intuitively.

In this paper, we provide a method to quantify real-world IoT privacy leakage in large-scale mobile network traffic systematically. With a three-day IoT traffic dataset containing 47,651 IoT devices from a mobile network operator in China, we generate traffic blocks and select privacy sensitive markers from them. Then, we extract sensitive information via these markers to collect fingerprints involved in the privacy of different IoT network entities. A network entity means a thing connected to the network that has separate and distinct existence and objective or conceptual reality, we find three major types of IoT network entities in our dataset, i.e., user, device, and platform. Based on these three types of entities, we construct a systematic framework to quantificationally analyze IoT privacy leakage through the fingerprints, which contains the basic information, attributes, and behaviors in both cyberspace and physical space. We find that the fingerprints can be used to profile users, devices or platforms in IoT, and present three case studies for instances, i.e., the privacy issues in location service, application calling, and voice service.

Our work reveals some regularities of IoT privacy and studies IoT privacy macroscopically. In conclusion, our contributions are summarized as follows:

- We use a semantic method to extract sensitive information from the real-world IoT traffic dataset, which implies that real information from IoT users, devices and platforms can be extracted and intelligently collected by any third party eavesdropping the IoT traffic.

- Based on the semantics of sensitive information and different IoT entities, we build a systematic framework and generate privacy fingerprints. For IoT users, we consider their basic information like name and gender, and the behaviors of using applications, shopping, travel, and entertainment; for devices, we consider their identifiers, spatio-temporal activities, network access, and data usage information; for platforms, we consider their services, data storage, and log records.

- We quantify the privacy leakage through our framework and fingerprints, analyze the quantification results, and present case studies. Our quantification shows that IoT devices have a larger scale of privacy leakage than other network entities (i.e., platforms, users). However, the leakage of several privacy issues with small scales (e.g., user password) can

lead to badly high risks. The case studies on the leakage of location information, application calling, and voice service data demonstrate that a third party can profile a network entity in both cyberspace and physical space via the aggregation of their privacy fingerprints.

The rest of this paper is organized as follows. First, we introduce some preliminaries, and give an overview of our problem and dataset in Section 2. Next, we present our method of extracting sensitive information about privacy in Section 3, and quantify and analyze IoT privacy leakage in Section 4. Finally, we review related works in Section 5 and conclude our work in Section 6.

2 PROBLEM FORMULATION

Notation	Definition
TF	Traffic flow
P_i	Content of the i-th packet
t_i	Arrival time of the i-th packet
N	Number of packets
$b_i; B$	The i-th traffic block; collection of b_i
T	Threshold of interval between the arrival time of two packets in each traffic block
$s_i; S$	The i-th piece of sensitive information; collection of s_i
$k; K$	Marker that represents the specific type of s_i; collection of k
v	Content gathering from the network traffic of s_i corresponding to k
$f_i; F$	The i-th privacy fingerprint; collection of f_i

TABLE 1: Summary of the main notations.

2.1 Preliminaries

In order to process IoT traffic collected from the mobile networks, we define *Traffic Flow* and *Traffic Block* at first, then we define *Sensitive Information* to describe IoT privacy and define *Privacy Fingerprint* as the basic unit to quantify privacy leakage. Table 1 presents the main notations.

Traffic Flow. The traffic flow, which contains a sequence of Internet packets, is denoted as a set $TF = \{(t_i, P_i)\}_{i=1}^{N}$, where N is the total number of packets, t_i is the arrival time of the i-th packet, and P_i is the content of the i-th packet.

Traffic Block. We segment the traffic flow into traffic blocks, and the i-th traffic block is denoted as a set $b_i = \{P_j\}_{j=1}^{N_i}$, where N_i is the total number of packets in traffic block b_i and P_j is the content of j-th packet. The packets in each traffic block b_i belong to the same network entity and relevant to the same IoT service. A network entity means a thing connected to the network that has separate and distinct existence and objective or conceptual reality. For IoT, there are three basic types of network entities: user, device, and platform. An IoT service means a specific task or operation in IoT, such as subscribe to a message, give a command, and upload data. These packets satisfy the following principles: 1) packets in the same traffic block have the same source IP address, destination IP address,

and transport protocol; 2) each pair of packets (P_i, P_j) in the same traffic block have an interval less than T, i.e., $|t_i - t_j| \leq T$, where t_i and t_j are the arrival time of packet P_i and P_j, T is a parameter depending on the service duration and dynamic IP assignment scheme used by the network service provider; 3) TCP packets in the same traffic block belong to the same TCP connection. Particularly, we denote the set of traffic block as B.

Sensitive Information. Sensitive information is data that the unauthorized accessing can incur privacy or security risks of an individual or organization. We formulate the i-th piece of sensitive information as a key-value pair denoted as $s_i = (k, v)$, where key k is a marker that represents a specific type of sensitive information, and value v is the corresponding content gathering from the network traffic. In particular, we denote the set of sensitive information pieces as S and the set of markers as K.

Privacy Fingerprint. We denote the IoT privacy fingerprints in network traffic as a set $F = \{f_i\}$, where each privacy fingerprint $f_i = \{s_j\}_{j=1}^{N_i}$ contains several pieces of sensitive information s_j, where N_i is the total number of sensitive information pieces in privacy fingerprint f_i. In a given privacy fingerprint f_i, the sensitive information pieces s_j are generated from the same traffic block b_i, and associated with the same sub-class of markers K.

2.2 Problem Overview

In order to quantify the privacy leakage of IoT in mobile network systematically, we collect IoT traffic from the mobile network and divide the problem into the following two major tasks as shown in Figure 1:
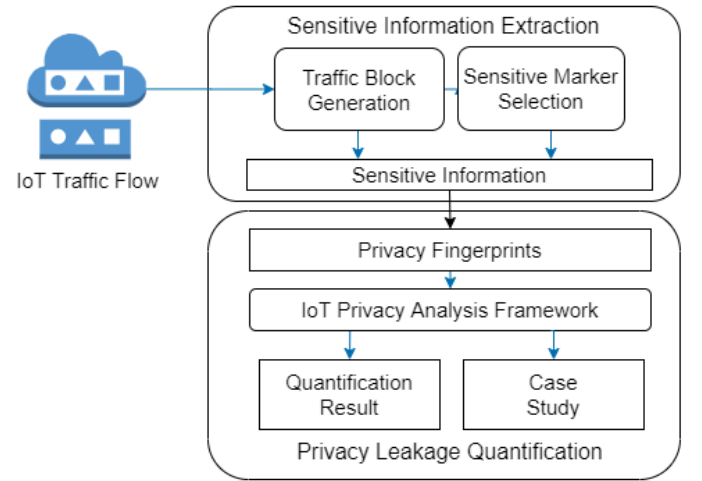

Fig. 1: The quantification framework overview.

1) *Sensitive Information Extraction.*

With a given IoT traffic flow TF, we firstly generate N traffic blocks $B = \{b_i\}_{i=1}^{N}$ by segmenting the packets in TF according to the definition of traffic block. Then, we provide a semantic method to gather sensitive markers into a set K, through which we extract L_i pieces of sensitive information $S_i = \{s_j\}_{j=1}^{L_i}$ from each traffic block b_i. The details of our semantic method are presented in Section 3.2.

2) *Privacy Leakage Quantification.*

Based on the semantics of sensitive information markers K, we divide these markers into different privacy issues, and construct a framework to systematize them according to the network entities of user, device, and platform. Then, we generate M_i privacy fingerprints $F_i = \{f_j\}_{j=1}^{M_i}$ from the sensitive information S_i extracted from each traffic block b_i considering privacy issues and network entities, conduct quantitative analyses by attributing the privacy fingerprints to the framework and discuss typical case studies.

2.3 Dataset

We obtain a dataset which contains three-day IoT traffic flow from a mobile network operator in China. The dataset covers multiple IoT applications and services. To identify the IoT devices in our dataset, we make reference to their Type Allocation Code (TAC) allocated by the Global System for Mobile Communications Alliance (GSMA). In addition, the product descriptions on relevant websites of IoT devices are highly correlated with their device type information [21]. We also refer to the official documents or user guides provided by the manufacturers of these IoT devices, and the product descriptions or instructions given by distributors. We identify 47,651 devices including locating and navigating instruments, monitoring equipment, on-board tablets, industrial sensors, vending machines, POS terminals, wearables, etc. These devices are connected to 37 different IoT platforms, such as logistics and vehicle management platforms, the communication between devices and platforms generates 22,121,216 packets in total. We find all kinds of readable contents in the payload of packets, including the basic information of users, devices, and platforms, the details of instructions, etc. We classify the devices into ten categories, as shown in Table 2, which cover the commonly used functions of IoT devices.

3 SENSITIVE INFORMATION EXTRACTION

We generate *traffic blocks* from the IoT *traffic flow* from our dataset, and design a semantic method to select markers for *sensitive information*. Based on these markers, we extract sensitive information from traffic blocks and give the measurement results.

3.1 Traffic Block Generation

We generate traffic blocks to gather the packets from the same network entity and involved in the same IoT service together, so as to maintain the integrality of sensitive information. In addition, due to the limitation of the Maximum Transmission Unit (MTU), a piece of message may be cut into several different packets in the traffic flow, which need to be solved by traffic block generation.

Firstly, we carry out deep packet inspection (DPI) on the packets to extract key information from each network layer, such as IP quintet (i.e., source and destination IP address, source and destination port, transport layer protocol) from IP layer header, TCP flags and sequence number from transport layer header, and payload from application layer. Then these packets are divided into different blocks according to the definition of traffic block in Section 2.1. We assign $T = 5\ minutes$ as the maximum interval of packets

in each block, and 12,634,271 traffic blocks are generated. We concatenate the payloads of packets in the same block, and use the payload contents of unencrypted protocols (e.g. HTTP and MQTT) for sensitive information extraction.

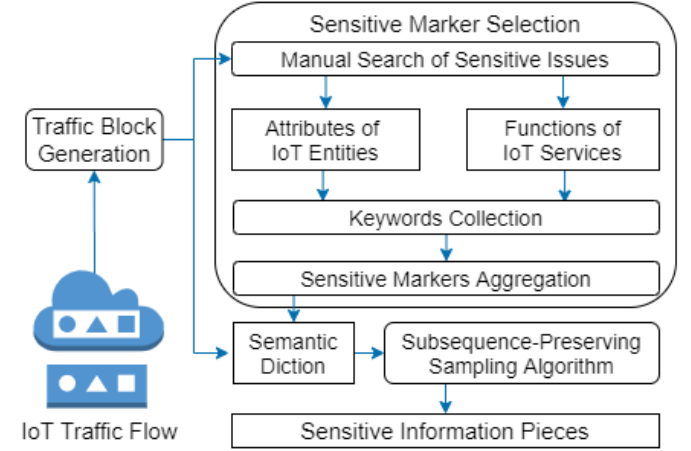## 3.2 Selection of Sensitive Information Markers

Fig. 2: Extraction of sensitive information.

Figure 2 presents an overview of sensitive information extraction. As shown in the figure, we firstly search for sensitive issues in the traffic blocks manually. Due to the lack of privacy standards of IoT devices and platforms, we search the online services and functions provided by common IoT operators, and collect 86 information fields associated with IoT users, devices, and platforms. Based on these fields, we use the method of approximate string matching to narrow the selection of keywords in the unencrypted payload of each traffic block. Then, taking the attributes of IoT network entities and the functions of IoT services into consideration, we collect a number of keywords that are relevant to some specific types of sensitive issues respectively, some keywords also provide indications of certain services in IoT scenarios. Hence, we set these keywords as traffic markers, and aggregate the similar forms of them, such as $'deviceid'$, $'device\_id'$, $'deviceID'$, and $'equipid'$. Finally, we collect all the markers as a semantic diction for sensitive information, and combine them with Subsequence-Preserving Sampling Algorithm [22] to extract sensitive information from each traffic block. Thus we acquire a large set of sensitive information pieces, the subsets of which are expected to profile the related network entities.

We look into our semantic diction, and measure the distribution of markers in the sensitive information pieces. More than 50 categories of traffic markers are selected, including identities or attributes of IoT users and devices, spatial-temporal information, application types, etc. For instance, marker $'imsi'$ falls into the category $'IMSI'$, its corresponding sensitive information is the $'International\ Mobile\ Subscriber\ Identity'$, which is a unique number to distinguish each mobile user; the other markers in the category $'IMSI'$ add some prefixes or suffixes to the core $'imsi'$ marker for deeper filtering, such as $'device\_imsi'$, $'imsi\_code'$, $'imsicode'$, etc. Table 3 shows several top coverage sensitive markers in our dataset. All of them are

Category	Device Count	Description
Locating	9897	An IoT device which locates objects like parcels and cargo containers remotely
Monitoring	1339	An IoT device which monitors remotely and uploads data streams
Portable	256	A handheld IoT device like card scanner
Router	1496	An IoT device which performs the function of routing
Vehicle	545	An on-broad IoT device used for driving assistance, recording data, or entertainment
Dongle	8737	An IoT device which is inserted into another device to provide network connectivity
Module	1809	An IoT device which works as a part of a system or machine
Modem	365	An IoT device which connects one system to another using a telephone line
Terminal	132	An IoT device that joins users or environments to the central system
Others	23075	IoT devices in other categories (e.g., POS terminal, Bluetooth Audio, Metering)

TABLE 2: Details of the dataset.

Marker	Where to find	Coverage
loc	HTTP: payload, json, name	28.47%
platform name	HTTP: payload, URL	7.86%
imsi	HTTP: payload, json, name	4.79%
imei	HTTP: payload, json, name	3.33%
userid	HTTP: payload, json, name	3.05%
appname	HTTP: payload, json, name	2.75%

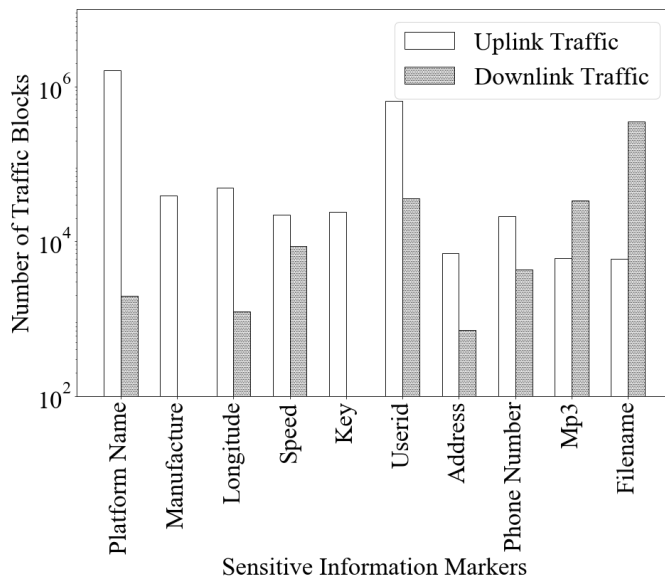TABLE 3: Top 6 coverage sensitive information markers.

Fig. 3: Occurrences of sensitive markers in the uplink and downlink traffic.

found in the HTTP payload of packets, and the payload is obtained by DPI, as discussed in Section 3.1. Specifically, the marker $'platform\ name'$ is found in the URL in HTTP payload, and the other five markers are found in the names of name/value pairs in json data in HTTP payload. The coverage of each marker means the proportion of traffic blocks containing the marker among all the blocks. We find that marker $'loc'$ has the topmost coverage of 28.47%, which implies that sensitive information leakage on 'location' can be serious; and the other top coverage sensitive markers are relevant to the identity or name of different network

entities, which is conducive to the sensitive information collection and behavior profile of corresponding network entity. Figure 3 shows the occurrence numbers of ten typical sensitive markers in the uplink and downlink traffic blocks. As shown in Figure 3, the markers relevant to basic information and identity authentication, like $'manufacture'$ and $'userid'$, occur more frequently in the uplink traffic; while the markers like $'mp3'$ and $'filename'$ tend to appear in the downlink traffic, which implies that these markers represent the feedback data for users. Moreover, we find that the occurrences of sensitive markers vary between the uplink traffic and downlink traffic, with the uplink traffic sent by devices or users and received by platforms or servers, and downlink traffic on the contrary.

4 PRIVACY LEAKAGE QUANTIFICATION

We design a quantification framework for IoT privacy leakage based on the semantics of *sensitive information* markers selected from mobile network traffic and the three major types of network entities of user, device, and platform. Then *privacy fingerprints* are generated from the sensitive information pieces of each traffic block, and attributed to the framework to quantify privacy leakage. At last, we discuss instances of privacy leakage on location information, application calling, and voice service.

4.1 IoT Privacy Framework Design

As shown in Table 4, we classify the sensitive information markers into different privacy issues based on their semantics, and systematize them by multiple attributes or behaviors of network entities in IoT.

User. IoT users have the most types of privacy issues, we consider their basic information and the behaviors of using applications, shopping, travel, and entertainment. The basic information includes their identifier, name, gender, address, email, phone number, and even password, which can be used to distinguish a user in cyberspace or physical world. Then, the distinguished users can be associated with privacy sensitive data of their behaviors, such as the calling sequences of the applications they used, their payment information in shopping, the time they went out, and their preferred entertainments.

Entity	IoT Privacy Classes	Sub-classes						
User	User Info	User identifier 690630/7.81% 4/53.7814	Phone number 25462/0.29% 5/50.7249	Address 7782/0.09% 5/44.7984	Name 4269/0.048% 5/41.7928	Email 99/0.00047% 5/23.0258	Gender 66/0.00075% 5/21.0324	Password 1/0.00001% 5/3.4657
	Application	App name 576205/6.52% 3/39.7926	Calling sequence 101497/1.14% 3/34.5833	App identifier 76872/0.87% 4/44.9996	Key 24127/0.27% 4/40.3645			
	Shopping	Purchase 28498/0.32% 3/30.7728	Payment 2030/0.023% 4/30.4651	Bank identifier 9/0.0001% 5/11.5129	Customer identifier 4/0.000045% 4/6.4377			
	Travel	Speed 30582/0.34% 3/30.9845	Car phone 15350/0.17% 4/38.5557	Traffic information 5636/0.064% 5/43.1855	Drive 1513/0.017% 3/21.9675	Fuel consumption 17/0.00019% 5/14.4518		
	Entertainment	Music 80085/0.91% 4/45.1634	Game 17027/0.19% 4/38.9704	News 11789/0.13% 4/37.5000	Video 6257/0.07% 4/34.9664	Wechat 5539/0.062% 4/34.4789	Book 46/0.00052% 4/15.4005	
Device	Device Info	IMSI 1003070/11.34% 4/55.2743	IMEI 697499/7.88% 5/67.2762	Client identifier 127995/1.44% 3/35.2792	Device identifier 106249/1.20% 3/34.7206	Manufacturer 39606/0.45% 3/31.7602		
	Spatio-temporal Activity	Location 5958124/67.37% 5/78.0013	Latitude 120467/1.36% 4/46.7965	Longitude 50572/0.57% 4/43.3246	City 57266/0.65% 3/32.8664	Device shutdown 33928/0.38% 3/31.2960	Device startup 30076/0.34% 3/30.9345	Coordinate conversion 4829/0.054% 3/25.4478
	Device data	Channel 1038928/11.75% 3/41.5611	Filename 362347/4.10% 4/51.2014	Voice 114253/1.29% 2/23.2923	UMID 49457/0.56% 3/32.4266	Mp3 39626/0.49% 2/21.1745		
	Access	User-agent 437646/4.94% 3/38.9675	Network-type 10465/0.12% 2/18.5117	Network mode 465/0.0052% 2/12.2884	Network operator 462/0.0052% 2/12.2754			
Platform	Platform Info/Service	Platform name 1644653/18.60% 4/57.2522	Platform service 184890/2.09% 4/48.5101	Firmware upgrade 121293/1.37% 3/35.1179	Cloud api 107021/1.21% 3/34.7423	SMS verification 50/0.00056% 2/7.8637		
	Platform data	Platform log 248533/2.81% 4/49.6933	Command exchange 208474/2.35% 3/36.7427	Session identifier 25423/0.29% 3/30.4303	Data storage 13506/0.15% 4/38.0438			

Legend: High risk (red), Medium risk (yellow), Low risk (green)

TABLE 4: Framework for IoT privacy leakage quantification.

Device. Similarly, an IoT device can be distinguished by the hardware and firmware information, including International Mobile Subscriber Identity (IMSI) number, International Mobile Equipment Identity (IMEI) number, customized client or device identifier, and manufacture. Then, we consider their spatio-temporal activities, data information, and the access to mobile networks. In detail, the spatio-temporal activities focus on the whereabouts and working time. The data information includes data channel, filename, Unique Material Identifier (UMID), and specially marked voice data. The access to mobile networks refers to user agent, network operator, and network type.

Platform. Compared with users and devices, IoT platforms have less types of privacy issues, including the platform name, the services and data information. Particularly, for the services of platforms, we find privacy issues on firmware upgrade, Application Program Interface (API), and Short Messaging Service (SMS) verification. Then, the other private data of platforms includes logs, command exchange details, session identifier, and data storage details.

Fig. 4: Number of privacy fingerprints in each IoT privacy class.

4.2 IoT Privacy Leakage Quantification

As discussed in Section 2, we use privacy fingerprint as the basic unit to quantify privacy leakage, and each privacy fingerprint contains several pieces of sensitive information, which are extracted from the same traffic block and associated with the same sub-class of markers. In order to conduct quantification analysis, we generate privacy fingerprints from the sensitive information pieces in each traffic block via attributing their markers to different sub-classes of privacy issues in Table 4 based on the semantics. For example, in a given traffic block, the sensitive information pieces with markers $'deviceid', 'device\_id', 'deviceID',$ or $'equipid'$ generate a privacy fingerprint of the sub-class $'Device\ identifier'$, and the sensitive information pieces with markers $'imsi', 'device\_imsi', 'imsi\_code',$ or $'imsicode'$ generate a privacy fingerprint of the sub-class $'IMSI'$. Then, we count the number of privacy fingerprints
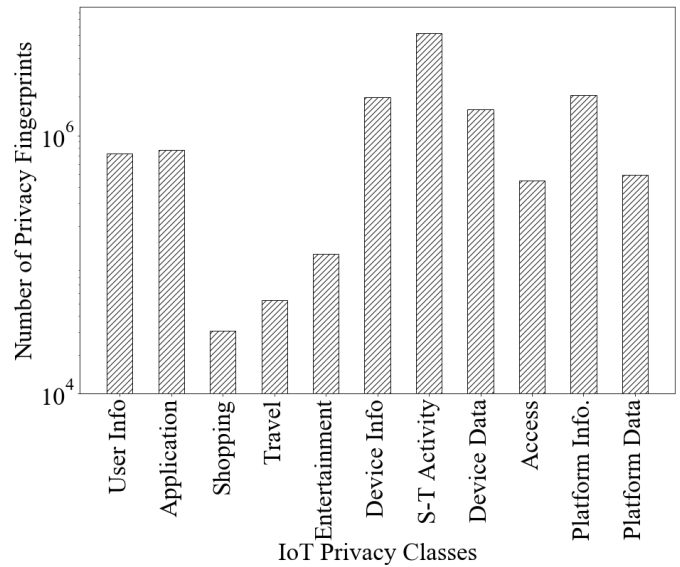
in each sub-class and calculate their proportions in all fingerprints. In particular, we capture 8,843,245 fingerprints in total, and part of them are associated to several different sub-classes at the same time on account of the multiple markers they contain.

Figure 4 shows the number of privacy fingerprints relevant to each IoT privacy class. For IoT users, the numbers of fingerprints on their basic information and the applications they used are both over 700,000, and the numbers of fingerprints on their shopping, travel, and entertainment behaviors are fewer, but still over 30,000 as well. For IoT devices, the numbers of fingerprints on the hardware and firmware information, spatio-temporal activity, and device data reach up to one million, the number of fingerprints on the access to network also reaches 449,038. For IoT platforms, the numbers of fingerprints on the platform basic
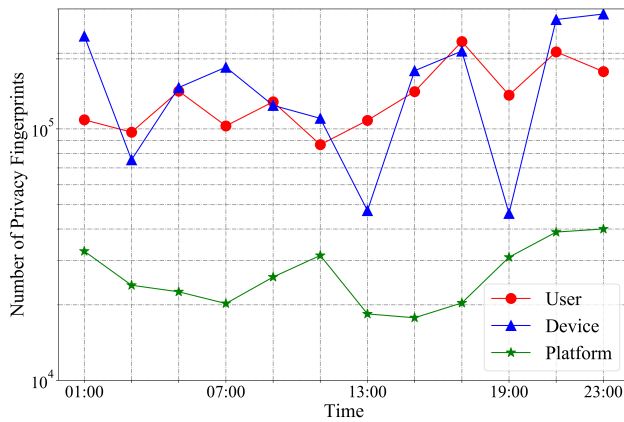
Fig. 5: Daily Patterns of privacy leakage for each type of network entities.

information or service is 2,057,907, and the numbers of fingerprints on the platform data is 495,936. IoT devices have a larger scale of privacy leakage than platforms and users, especially their spatio-temporal activities.

Figure 5 plots the average number of privacy fingerprints at different times of a day for each type of network entities. IoT users, devices, and platforms have different daily patterns on privacy leakage. We count the number of fingerprints in each two-hour consecutive interval, and calculate the average of each time period in three days. IoT users leak 96,210 privacy fingerprints per hour in average, and leak most fingerprints around 17:00. IoT devices leak 108,980 fingerprints and platforms leak 18,801 fingerprints per hour in average, and they both leak most fingerprints around midnight. We observe that IoT users leave more fingerprints off from work than at work, and devices tend to work round the clock in three shifts, but the two peaks of the number of platforms' privacy leakage appear in midday and midnight each. The above patterns of users and devices imply that the number of fingerprints represents the inter-activity between different network entities to some extent. In addition, to understand the two peaks of platforms' privacy leakage, we look into the semantics of their privacy fingerprints in the day and night respectively. We find that the services for users are more active during the daytime, while the upgrade and maintenance are more active at night.

Table 4 presents the number of privacy fingerprints and risk level relevant to all the sub-classes. In each cell of Table 4, the first line shows the name of sub-class, and the second line shows the number of privacy fingerprints and the corresponding proportion in all 8,843,245 privacy fingerprints. We put the sub-classes in each IoT privacy class by their proportions from left to right, and the sum of these proportions does not add to 100% because part of fingerprints are associated to several different sub-classes at the same time by the multiple markers they contain. In addition to the leakage quantity, the closeness of the correlation between each privacy sub-class and IoT entity makes a difference to the risk level. As shown in the third line of each cell in Table 4, we measure the closeness of correlation

in values from 1 to 5, and calculate the risk score of each sub-class by multiplying the logarithm of the number of privacy fingerprints by this closeness value. According to the risk scores, we divided all the IoT privacy sub-classes into three risk levels in different colors. The sub-classes with risk scores higher than 50 are marked as high risk in red color, the sub-classes with risk scores less than 50 and higher than 30 are marked as medium risk in yellow color, and the sub-classes with risk scores less than 30 are marked as low risk in green color. We observe that IoT users, devices, and platforms have considerable risks respectively: (i) For IoT users, the most serious privacy leakage appear in the sub-classes of user identifier, application name, and application calling sequence, the numbers of privacy fingerprints in these sub-classes are over 0.1 million, and their proportions are all over 1%. User identifier and phone number have high privacy risk on account of their large leakage scales and close correlations with users. Meanwhile, although the leakage scales of several privacy sub-classes for users are small, the information can lead high risks to a particular user, for example, user password leakage surrenders the whole account to others. (ii) For IoT devices, IMSI, location, and channel of device data have largest scales of leakage, with the numbers of fingerprints over one million and the proportions over 10%. Most of the other sub-classes also leak a lot on the spatio-temporal activity, data, and network access of devices, with seven sub-classes have proportions over 1%. IMSI, IMEI, location, and the file name of device data have high privacy risk, which warns us that the IoT devices have privacy risks no less than users. (iii) For IoT platforms, the number of privacy fingerprints of their names is 1.64 million, with the largest proportion of 18.60%, which leads to high risk. In addition, the leakage proportions of platform service, firmware upgrade, API, log, and command exchange are over 1%.

In general, IoT devices have the largest number of leaked privacy fingerprints, which remind us to pay more attention on preventing device privacy leakage. Then, the privacy leakage of users has a smaller amount, because human privacy gains more concern from both developers and users, and receives more protection as a result. Meanwhile, the privacy sensitive data of platforms found in our dataset is so homogeneous as to fall into limited sub-classes. In consideration of the three-day duration of our dataset, collecting IoT traffic in mobile networks for a longer duration can reveal more privacy risks.

4.3 Case Study

As the privacy fingerprints on location information of IoT devices and application calling of IoT users are both abundant and perceptible, we choose them as the first two cases. For the last case, we find a voice service platform which leaks user information along with their speech content, then provides convenient access to users' daily life for eavesdroppers.

Location Information. Location information leakage has the largest number of privacy fingerprints. We find all kinds of location information, including the longitude and latitude, the city name, and specific address. The locations are relevant to the vehicle trajectories, real-time locations

and delivery addresses of logistics, starting and destination points in navigation services, correspondence address maintained at the bank, etc. The snippet below is an example of specific address leakage, the address content is encoded in UTF-8.

<PAYLOAD:****{...
"locationInfo":{"longitude":##.######,
                "latitude":##.######,
                "address":*************}...}>

For instance, on an IoT cloud platform for vehicles, more than 100 users' tracks are exposed in the form of longitude and latitude. We glean the privacy fingerprints of a selected user, and the snippets below are part of the footprints in the tracks:

<PAYLOAD:****{
"cmd":"overspeed",
"param":{"limit":40,
         "speed":41,
         "lat":##.######,
         "lng":##.##,
         "time":############}}>

<PAYLOAD:****{
"cmd":"overspeed",
"param":{"limit":50,
         "speed":64,
         "lat":##.######,
         "lng":##.##,
         "time":############}}>

We plot one of his tracks during 16/10/2018 09:08:11 and 16/10/2018 11:03:24 in Figure 6, where warmer color represent higher speed as shown in the colorbar. This track is 57.37 kilometers long and passed through five towns in Hubei Province, the average speed is 48 km/h.

Fig. 6: One of the leaked tracks of a selected user.

Application Calling. We observe a large number of privacy fingerprints on application calling in our dataset. The following application calling list is an example, and we cover the details of corresponding user and platform for privacy protection:

"param":[

1   {"dev":"livebroadcast","type":"***","cap":"#" }
2   {"dev":"com.autonavi.amapautolite","cap":####},
3   {"dev":"com.aliyun.bluetoothphone","cap":#},
4   {"dev":"com.i**.***","cap":##},
5   {"dev":"com.android.settings","cap":########},
6   {"dev":"com.android.settings","cap":########},
7   {"dev":"com.yunos4car.update","cap":########},
8   {"dev":"com.yunos.weatherservice","cap":####},
9   {"dev":"com.mediatek.filemanager","cap":#},
10 {"dev":"com.ximalaya.ting.android.car","cap":#},
11 {"dev":"com.i**.***","cap":#},
12 {"dev":"com.aliyun.filemanager","cap":#},
13 {"dev":"com.i**.***","cap":###},
14 {"dev":"com.yunos4car.music","cap":##},
15 {"dev":"com.i**.updater","cap":##},
16 {"dev":"com.aispeech.aios","cap":###},
17 {"dev":"com.aispeech.aios.wechat","cap":##},
18 {"dev":"com.i**.***","cap":#},
19 {"dev":"com.i**.***","cap":##},
20 {"dev":"cloudapi","cap":"##"},
21 {"dev":"_adas","cap":"none"},
22 {"dev":"_bluetooth","cap":"none"},
23 {"dev":"_voice","cap":"none"},
24 {"type":"general","dev":"***"}]

The application callings containing keywords "*android*" and "*car*" tell us this is an android on-board device; then keywords "*aliyun*" and "*yunos*" represent services from Aliyun (https://cn.aliyun.com), i.e., a Chinese company which provides cloud computing services to online businesses; "*autonavi*" (https://www.autonavi.com) is a Chinese web mapping, navigation and location-based services software; and "*ximalaya*" (https://www.ximalaya.com/) is a Chinese online audio sharing platform; while "*_adas*", "*_bluetooth*", and "*_voice*" are system applications. Referring to the hidden details, we speculate that the above list is a sequence of test application calling behavior from a new user or developers. In practice, a single privacy fingerprint of application calling list tends to contain less applications, but we can concatenate the privacy fingerprints from the same user to trace the temporal behaviors of real application calling.

Voice Service. We find a cloud platform that provides voice services through intelligent devices, e.g., smartphones, on-board tablets. The traffic contains both user information and texts converted from human speeches, which reveal all kinds of users' interactions. For instance, a user requests the navigation service and set her destination in the first text, then requests the weather service to obtain the weather information of her destination place in the next text. At last, the weather forecast for the following three days is fed back. The snippets below are part of her requests:

<PAYLOAD:****{...
"speakerResult":{"message":"success",
                 "minAge":##,
                 "gender":"female",
                 "ageGroups":"senior",
                 "code":"####",
                 "operator":"recognize"},
"html5Url":"**************",
"semantic":{"intent":{"name":"***",

...}...}...}>

<PAYLOAD:****{...
"service":"cn.********.****",
"code":"ANSWER",
"general":{"quitDialog":"true",
          "type":"T",
          "title":"*****",
          "text":"**************************",
          "url":"**************************"},
"responseId":"*********"...}>

5 RELATED WORKS

Privacy is defined on account of users customarily. With regard to IoT, following the traditional definition, Ziegeldorf *et al.* [23] defined privacy as "the threefold guarantee to the subject for: 1) awareness of privacy risks imposed by smart things and services surrounding the data subject; 2) individual control over the collection and processing of personal information by the surrounding smart things; 3) awareness and control of subsequent use and dissemination of personal information by those entities to any entity outside the subject's personal control sphere". Alpár *et al.* [24] also present that "IoT privacy is the right of individuals to determine for themselves when, how and to what extent information about them is collected, processed and communicated". In their definition, individuals not only have the right to determine these aspects within their control area, but also trust their right to be respected when control is not available. Similar to their definitions, IoT privacy is regarded as the extension of human privacy in most works. Hence, literature on universal privacy leakage quantification provides us experience to analyze IoT privacy leakage. In addition to analysis studies, increasing researchers devote to preventing privacy leakage in IoT, which inspires us to figure out the weak points and conduct corresponding analysis.

5.1 Privacy Leakage Quantification in Mobile Networks

Most of the previous works quantified privacy leakage in mobile networks based on mobile services and mobile devices, the typical examples are online social network (OSN) and smartphone.

Mobile Service based Quantification. With the ubiquitous use of mobile devices, European regulators warned of higher privacy losses as a result of searching in the mobile Internet, Krishnamurthy *et al.* [25] started to examine the privacy leakage in mobile network from mobile OSNs. They collected data by creating accounts on some mobile OSNs and observing the private information requested by each mobile OSN. Then, they presented a taxonomy of ways to study privacy leakage, and reported on the current status of existing leakages. Their report shows that mobile OSNs exhibit private information to third parties. Later, Xia *et al.* [16] extended the dataset by collecting traffic from a CSP and called more attention to the privacy leakage in mobile network data than OSN. They sorted and organized the traffic into groups according to the users' names and personal information (e.g., political views, browsing habits, favorite apps with the users). Then they quantified privacy leakage in mobile networks based on these sorted groups. Their methodologies of generating privacy fingerprints and classifying OSN privacy leakage provide us the instances of sensitive information extraction and the framework of user privacy depicting.

Mobile Device based Quantification. Apart from OSN in mobile network, Yang *et al.* [26] presented an analysis framework called AppIntent to analyze sensitive data transmission and detect privacy leakage in android, they considered sensitive data from device ID, phone information, location, contacts, and short message service (SMS); Das *et al.* [27] conducted a measurement study on privacy leakage of BLE communication between smartphone and fitness tracker, they found that the fitness trackers traffic is correlated with user's activity and can be used to speculate user's gait. Their works show us privacy leakage samples from android smartphone and wearable devices.

In addition, some researches also concentrated on specific mobile network access methods to privacy leakage like public hotspots [28] and certain categories of privacy such as geographical locations [29], [30]. Unlike the studies on mobile OSN, smartphone, or wearable device, our work focuses on privacy leakage in IoT mobile network, and give comprehensive quantification and analysis for the privacy.

5.2 Preserving IoT Privacy

Various methods were proposed to preserve privacy in IoT, including cyber-attack detection, and sensitive data management.

Cyber-attack Detection. Most IoT devices are designed to execute tasks or feed information back by remote control with low cost and computing power, which leads to the nature that they are easily attacked and result in sensitive data leakage at large scale [31]. Hence, a number of studies preserved IoT privacy via detecting or preventing IoT devices from DDoS attack [32], [33], [34], [35] or unauthorized access [36], [37], [38], [39], [40], a large proportion of them focused on designing secure architectures and protocols. Moreover, Sha *et al.* [41] detected privacy leakage scenarios of IoT by measuring security degree of sensitive information. Then, they proposed attack-defense and fix-distribution mechanisms to reject sensitive information leverage attack based on selected taint tracking and real-time memory modification.

Sensitive Data Management. However, we pay more attention to the management of privacy sensitive data than security guarantee structures. To preserve privacy for fog computing-enhanced IoT by data aggregation, Lu *et al.* [42] presented a scheme named LPDA, which employs the homomorphic Paillier encryption, one-way hash chain techniques, and Chinese Remainder Theorem. The scheme aggregated hybrid IoT devices' data into one, and filtered injected false data at the network edge early. Their security analysis proved that the scheme satisfies differential privacy constraint. Yin *et al.* [43] also used differential privacy techniques to protect location data privacy in Industrial IoT, while maximizing the utility of algorithm and data. Except for differential privacy, Liu *et al.* [44] used a K-anonymity method to preserve data privacy for wearable IoT devices while guaranteeing the data usability. The above methods

preserve the privacy of each individual by aggregating data or clustering devices, which prevents others from profiling one person or device. However, the sensitive attributes of groups are still unprotected, which inspires us to compare between the privacy of individuals and groups in the future work.

6 CONCLUSION AND FUTURE WORK

In this paper, we quantify IoT privacy leakage in mobile networks systematically. We combine systematic analyses with real-world measurements by generating privacy fingerprints from the network traffic and attributing them to a privacy quantification framework. Our quantitative analyses and case studies present that the behaviors and attributes can be associated with the basic information to profile an IoT network entity in both cyberspace and physical space, which leads to high risks.

As the future work, we plan to extend the time duration of our dataset and consider more platforms in our study, while going deep into significant privacy issues. The case studies inspire us to associate the basic information with muti-platform behaviors. For example, proceeding from the application calling list, activities of the same user can be traced in different services. Then, we can aggregate all kinds of privacy sensitive data, e.g., purchase record and vehicle tracks.

REFERENCES

[1] C. V. N. Index, "Global mobile data traffic forecast update, 2017–2022 white paper," *Cisco: San Jose, CA, USA*, 2019.

[2] H. Mohajeri Moghaddam, G. Acar, B. Burgess, A. Mathur, D. Y. Huang, N. Feamster, E. W. Felten, P. Mittal, and A. Narayanan, "Watching you watch: The tracking ecosystem of over-the-top tv streaming devices," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2019, pp. 131–147.

[3] Z. B. Celik, L. Babun, A. K. Sikder, H. Aksu, G. Tan, P. McDaniel, and A. S. Uluagac, "Sensitive information tracking in commodity iot," in *27th {USENIX} Security Symposium ({USENIX} Security 18)*, 2018, pp. 1687–1704.

[4] D. Wood, N. Apthorpe, and N. Feamster, "Cleartext data transmissions in consumer iot medical devices," in *Proceedings of the 2017 Workshop on Internet of Things Security and Privacy*, 2017, pp. 7–12.

[5] O. Arias, J. Wurm, K. Hoang, and Y. Jin, "Privacy and security in internet of things and wearable devices," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 1, no. 2, pp. 99–109, 2015.

[6] A. K. Sikder, G. Petracca, H. Aksu, T. Jaeger, and A. S. Uluagac, "A survey on sensor-based threats to internet-of-things (iot) devices and applications," *arXiv preprint arXiv:1802.02041*, 2018.

[7] N. Apthorpe, D. Reisman, S. Sundaresan, A. Narayanan, and N. Feamster, "Spying on the smart home: Privacy attacks and defenses on encrypted iot traffic," *arXiv preprint arXiv:1708.05044*, 2017.

[8] Z. Ling, J. Luo, Y. Xu, C. Gao, K. Wu, and X. Fu, "Security vulnerabilities of internet of things: A case study of the smart plug system," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1899–1909, 2017.

[9] D. Kozlov, J. Veijalainen, and Y. Ali, "Security and privacy threats in iot architectures," in *Proceedings of the 7th International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and . . . , 2012, pp. 256–262.

[10] K. Zhao and L. Ge, "A survey on the internet of things security," in *2013 Ninth international conference on computational intelligence and security*. IEEE, 2013, pp. 663–667.

[11] R. Mahmoud, T. Yousuf, F. Aloul, and I. Zualkernan, "Internet of things (iot) security: Current status, challenges and prospective measures," in *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*. IEEE, 2015, pp. 336–341.

[12] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," *Computer Networks*, vol. 57, no. 10, pp. 2266–2279, 2013.

[13] P. Porambage, M. Ylianttila, C. Schmitt, P. Kumar, A. Gurtov, and A. V. Vasilakos, "The quest for privacy in the internet of things," *IEEE Cloud Computing*, vol. 3, no. 2, pp. 36–45, 2016.

[14] F. Loi, A. Sivanathan, H. H. Gharakheili, A. Radford, and V. Sivaraman, "Systematically evaluating security and privacy for consumer iot devices," in *Proceedings of the 2017 Workshop on Internet of Things Security and Privacy*. ACM, 2017, pp. 1–6.

[15] G. Chu, N. Apthorpe, and N. Feamster, "Security and privacy analyses of internet of things toys," *CoRR*, 2018.

[16] N. Xia, H. H. Song, Y. Liao, M. Iliofotou, A. Nucci, Z.-L. Zhang, and A. Kuzmanovic, "Mosaic: Quantifying privacy leakage in mobile networks," in *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4. ACM, 2013, pp. 279–290.

[17] Z. A. Almusaylim and N. Zaman, "A review on smart home present state and challenges: linked to context-awareness internet of things (iot)," *Wireless networks*, vol. 25, no. 6, pp. 3193–3204, 2019.

[18] N. Apthorpe, D. Y. Huang, D. Reisman, A. Narayanan, and N. Feamster, "Keeping the smart home private with smart (er) iot traffic shaping," *Proceedings on Privacy Enhancing Technologies*, vol. 2019, no. 3, pp. 128–148, 2019.

[19] M. H. Mazhar and Z. Shafiq, "Characterizing smart home iot traffic in the wild," in *2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, 2020, pp. 203–215.

[20] K. Dziubinski and M. Bandai, "Your neighbor knows what you're doing: Defending smart home iot device traffic from privacy lan attacks," in *Workshops of the International Conference on Advanced Information Networking and Applications*. Springer, 2020, pp. 526–534.

[21] X. Feng, Q. Li, H. Wang, and L. Sun, "Acquisitional rule-based engine for discovering internet-of-things devices," in *27th {USENIX} Security Symposium ({USENIX} Security 18)*, 2018, pp. 327–341.

[22] X. Shu, J. Zhang, D. D. Yao, and W.-C. Feng, "Fast detection of transformed data leaks," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 3, pp. 528–542, 2016.

[23] J. H. Ziegeldorf, O. G. Morchon, and K. Wehrle, "Privacy in the internet of things: threats and challenges," *Security and Communication Networks*, vol. 7, no. 12, pp. 2728–2742, 2014.

[24] G. Alpár, L. Batina, L. Batten, V. Moonsamy, A. Krasnova, A. Guellier, and I. Natgunanathan, "New directions in iot privacy using attribute-based authentication: Position paper," 2016.

[25] B. Krishnamurthy and C. E. Wills, "Privacy leakage in mobile online social networks," in *Proceedings of the 3rd Wonference on Online social networks*. USENIX Association, 2010, pp. 4–4.

[26] Z. Yang, M. Yang, Y. Zhang, G. Gu, P. Ning, and X. S. Wang, "Appintent: Analyzing sensitive data transmission in android for privacy leakage detection," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. ACM, 2013, pp. 1043–1054.

[27] A. K. Das, P. H. Pathak, C.-N. Chuah, and P. Mohapatra, "Uncovering privacy leakage in ble network traffic of wearable fitness trackers," in *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications*. ACM, 2016, pp. 99–104.

[28] N. Cheng, X. O. Wang, W. Cheng, P. Mohapatra, and A. Seneviratne, "Characterizing privacy leakage of public wifi networks for users on travel," in *2013 Proceedings IEEE INFOCOM*. IEEE, 2013, pp. 2769–2777.

[29] H. Li, H. Zhu, S. Du, X. Liang, and X. S. Shen, "Privacy leakage of location sharing in mobile social networks: Attacks and defense," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 646–660, 2016.

[30] Y. Liang, Z. Cai, Q. Han, and Y. Li, "Location privacy leakage through sensory data," *Security and Communication Networks*, vol. 2017, 2017.

[31] I. Andrea, C. Chrysostomou, and G. Hadjichristofi, "Internet of things: Security vulnerabilities and challenges," in *2015 IEEE Symposium on Computers and Communication (ISCC)*. IEEE, 2015, pp. 180–187.

[32] C. Zhang and R. Green, "Communication security in internet of thing: preventive measure and avoid ddos attack over iot network," in *Proceedings of the 18th Symposium on Communications & Networking*. Society for Computer Simulation International, 2015, pp. 8–15.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2020.3038639, IEEE Internet of Things Journal

11

[33] M. De Donno, N. Dragoni, A. Giaretta, and M. Mazzara, "Antibiotic: protecting iot devices against ddos attacks," in *International Conference in Software Engineering for Defence Applications*. Springer, 2016, pp. 59–72.

[34] R. Doshi, N. Apthorpe, and N. Feamster, "Machine learning ddos detection for consumer internet of things devices," in *2018 IEEE Security and Privacy Workshops (SPW)*. IEEE, 2018, pp. 29–35.

[35] A. Mehmood, M. Mukherjee, S. H. Ahmed, H. Song, and K. M. Malik, "Nbc-maids: Naïve bayesian classification technique in multi-agent system-enriched ids for securing iot against ddos attacks," *The Journal of Supercomputing*, vol. 74, no. 10, pp. 5156–5170, 2018.

[36] S. Cirani, M. Picone, P. Gonizzi, L. Veltri, and G. Ferrari, "Iot-oas: An oauth-based authorization service architecture for secure services in iot scenarios," *IEEE sensors journal*, vol. 15, no. 2, pp. 1224–1234, 2014.

[37] S. R. Moosavi, T. N. Gia, A.-M. Rahmani, E. Nigussie, S. Virtanen, J. Isoaho, and H. Tenhunen, "Sea: a secure and efficient authentication and authorization architecture for iot-based healthcare using smart gateways," *Procedia Computer Science*, vol. 52, pp. 452–459, 2015.

[38] A. Tewari and B. Gupta, "Cryptanalysis of a novel ultralightweight mutual authentication protocol for iot devices using rfid tags," *The Journal of Supercomputing*, vol. 73, no. 3, pp. 1085–1102, 2017.

[39] A. Ouaddah, A. A. Elkalam, and A. A. Ouahman, "Towards a novel privacy-preserving access control model based on blockchain technology in iot," in *Europe and MENA Cooperation Advances in Information and Communication Technologies*. Springer, 2017, pp. 523–533.

[40] L. Luo, Y. Zhang, B. Pearson, Z. Ling, H. Yu, and X. Fu, "On the security and data integrity of low-cost sensor networks for air quality monitoring," *Sensors*, vol. 18, no. 12, p. 4451, 2018.

[41] L. Sha, F. Xiao, W. Chen, and J. Sun, "Iiot-sidefender: Detecting and defense against the sensitive information leakage in industry iot," *World Wide Web*, vol. 21, no. 1, pp. 59–88, 2018.

[42] R. Lu, K. Heung, A. H. Lashkari, and A. A. Ghorbani, "A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced iot," *IEEE Access*, vol. 5, pp. 3302–3312, 2017.

[43] C. Yin, J. Xi, R. Sun, and J. Wang, "Location privacy protection based on differential privacy strategy for big data in industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3628–3636, 2017.

[44] F. Liu and T. Li, "A clustering k-anonymity privacy-preserving method for wearable iot devices," *Security and Communication Networks*, vol. 2018, 2018.