Extraction and Accumulation of Identity Attributes from the Internet of Things

Lothar Fritsch[1] Nils Gruschka[2]

Abstract: Internet of Things (IoT) devices with wireless communication provide person-relateable information usable as attributes in digital identities. By scanning and profiling these signals against location and time, identity attributes can be generated and accumulated. This article introduces the concept of harvesting identifiable information from IoT. It summarizes ongoing work that aims at assessing the amount of person-relatable attributes that can get extracted from public IoT signals. We present our experimental data collection in Oslo/Norway and discuss systematic harvesting, our preliminary results, and their implications.

Keywords: IoT; profiling; identification; identity attributes; privacy

1 Introduction

The goal of this article is an assessment of how much measureable person-realteable IoT devices are. Detected devices and their re-identifiability, linkability, tracability and potential for placement in contexts such as private addresses will provide opportunities for collection of personal data and for identification.

Internet of Things (IoT) devices are omnipresent nowadays not only in professional environments like industrial production or smart cities, but also as personal devices. Many household devices are nowadays “smart” devices that are communicating via wireless communication protocols like IEEE 802.11 (commonly called Wi-Fi), Bluetooth, ZigBee or Z-Wave. This applies especially to home automation (like smart bulbs) and home entertainment equipment (like TVs or speakers). Also, outside our home we are surrounded by IoT devices. Modern cars not only offer Bluetooth communications (typically for hands–free phone calls), but also span a Wi-Fi access point. And even when just walking (or using public transport), personal IoT devices follow us. In addition to the smart phone, which more or less everyone uses, many people carry wireless headsets/headphones, fitness trackers or smart watches.

All these devices are constantly transmit signals and it is very easy for an adversary to scan the communication from a safe distance and to analyse the scan results later or even in

[1] Oslo Metropolitan University, Dept. of Computer Science, Oslo, Norway Lothar.Fritsch@oslomet.no
[2] University of Oslo, Dept. of Computer Science, Oslo, Norway nilsgrus@ifi.uio.no

real time. Although nearly all communication is nowadays encrypted and most protocols contain mechanisms to obfuscate the sender, many devices (like TV or headsets) do not make use of these methods and simply broadcast their identifier. Also, research has shown that, even if mechanisms like MAC randomization are active, it is still possible to profile the communications signals and identify the sending device. And this allows in many cases to learn the behaviour of the owner (like "is at home") or even track him with high accuracy.

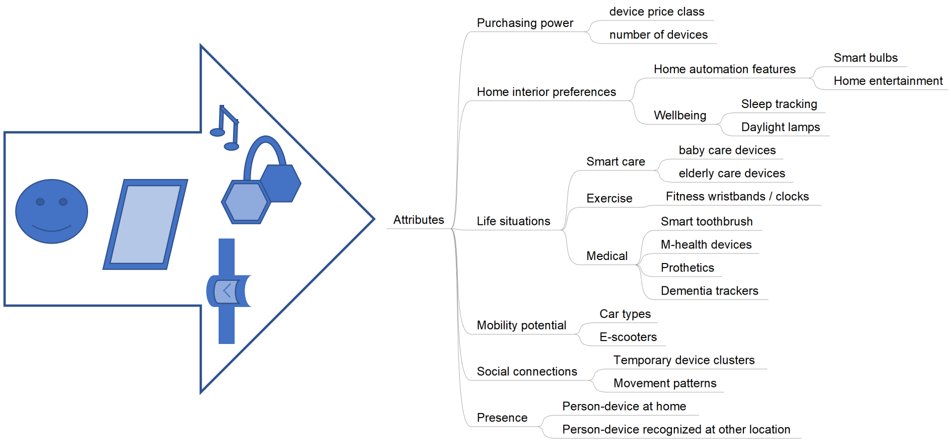

Fig. 1: IoT attributes (examples) available from IoT devices.

In this paper, we will show how personal attributes (examples shown in Fig. 1) can be retrieved from IoT environments of personal devices by scanning their presence and by recognizing what kinds of devices they are. We introduce the concept of IoT-extracted identity attributes, describe several proof-of-concept-measurements, and propose an approach for systematic collection and mapping of such attributes in order to enrich digital partial identities.

2 Background

The potential for identity data extraction from smart environments has been identified in 2006 by the SWAMI project. In its final report, in chapter three, four dark scenarios for sensing environments are summarized that envision data collection by ambient systems [Wr08]. Its vision has now developed into ubiquitous sensing infrastructure. For Android apps, it has been shown that systematic access to smartphone data can provide partial identity information to app ecosystems [MF20; PH10]. Such profiling against IoT devices' publicly broadcast radio signals could deliver similar identity attributes. Possibilities for clandestine identification, location tracking and eavesdropping on individuals carrying such communication terminals, by exploiting functionalities available in the wireless communication protocols and their implementations [An19].

We present a set of attacks that allow an attacker to link a Wi-Fi device to its owner identity. We present two methods that, given an individual of interest, allow identifying the MAC address of its Wi-Fi enabled portable device. Those methods do not require a physical access to the device and can be performed remotely, reducing the risks of being noticed. Finally, we present scenarios in which the knowledge of an individual MAC address could be used for mischief [Cu13]. By profiling device network addresses, it will be possible to distinguish devices, follow them over space and time, and to correlate then to each other. It will be possible to extract which other devices or networks they exchange messages with, too. From a database, it will be possible to look up past observations for the devices.In addition, presence of certain devices will indicate presence of people or their absence.

The classification of mobile consumer devices trough information provided by their wireless communication interfaces has been demonstrated successfully. Using this opportunity, the type of device can get recognized. A demonstrator was presented in [VG13]. By adding device type information, it will be possible to infer specific context attributes, such as purchasing power (device price), specific interests (fitness performance monitoring) or specific infrastructure (such as home automation). Similar insights about the side channel information gained from profiling RFID tags for business intelligence was discussed in [Fr09]. The consideration of location information adds context information [Fr08] that can enrich or even create new attributes.

Many sensing approaches focus on Wi-Fi device profiling and tracking based on passive and active discovery probe requests [CKB14]. Here, frequent probing reveals device characteristics. In particular, passive identification techniques for sensing without the sensed devices involved in the protocol are useful in the context of attribute extraction [Bh19]. When sensing nodes collaborate as a sensing network over a larger geographic area, tracking of mobile devices and profiling of their movements is possible. This feature is used in indoor navigation applications. However, it has been discovered to be able to track recognizable smart cars through their Wi-Fi interfaces, too [BZ19].

Finally, the application of forensic analysis and visualization of devices and their movements can generate insights that may lead to qualified identity attributes added [TLT16]. The mapping of a device to a place of living and to a place of work is one obvious source of direct identity information. Classification of whereabouts from geographical metadata, proximity to other devices—stationary or mobile—and the analysis of stationary equipment in private homes can easily generate attributes.

3 Data collection

As a proof-of-concept, we ran several small data collection campaigns. Their purpose was the demonstration of the availability of identity attributes from IoT devices. In an exploratory series of data collection experiments, we used the WIGLE[3] scanning app for Android

[3] https://www.wigle.net

devices as a data collection sensor. We planned a series of small experiments. Google Earth was used for visualizing geographical plots of the data.

By analyzing scans of wireless device communication, we accumulate IoT-extracted personal identity attributes that enhance our knowledge about the device owner.

3.1 Results

In initial random scanning during daily movements, we learned that several vendor's television sets, and home entertainment equipment (Apple TV, Android TV, local cable TV providers' boxes) were easily identified in abundance. The same holds for various wireless audio speakers, smart light bulbs, an occasional smoke detector or wireless dimmer. Even a smart electric toothbrush was observed. Mobile phones and headsets were met in most places. We collected data in four targeted locations:

1. in a multi-floor, dense urban housing area in central Oslo;
2. in a villa district with free-standing houses with large gardens (Fig. 2);
3. on a highway-crossing pedestrian bridge targeting passing cars (Fig. 3);
4. by a footpath in a urban park that leads to a childcare facility (Fig. 4).

Below, we summarize our findings.

3.1.1 Urban housing area

We collected both Wi-Fi and Bluetooth information. However, in this article we focus on the collected Bluetooth data. In the multi-floor housing area located central in Oslo, a very large number of devices was found. Mobile phones, headsets, electric scooters from competing scooter pools, vehicle entertainment systems, television sets and set-top boxes were present in abundance. Wireless speakers and other audio equipment frequently was detected. Also, an occasional smart lamp was found in the sample. However, due to the density of housing, the mapping of devices to apartments required triangulation, advances to specific windows and tours around the next corner in order to observe weakening signals. Passing traffic—including public busses filled with passengers with smartphones, headsets and other devices—polluted the measurements. An effort targeting passing cars in this environment was given up due too many passing pedestrians.

3.1.2 Villa district

As a complementary approach to the dense urban center, we approached a villa area in Oslo's Holmenkollen district where spacious gardens surround houses. Due to fencing, we moved at a distance to the buildings. We noticed the presence of more contemporary cars, of smart home devices, televisions, fitness wristbands, wireless audio equipment and phones and tablet devices. The geography made mapping of the devices to houses relatively easy by traversing the property on roadside. We noticed a house that had devices named after rooms, as shown in Fig. 2. We saw a fitness wristband in a house where the lights were on, which possibly indicated the presence of a tenant. In the villa district, the mapping of wireless devices to property should be a task of low effort, due to large spaces in-between houses, low traffic and low number of passersby. It should be a relatively easy task to harvest network addresses, SSIDs and device fingerprints for named persons through their addresses in this environment.

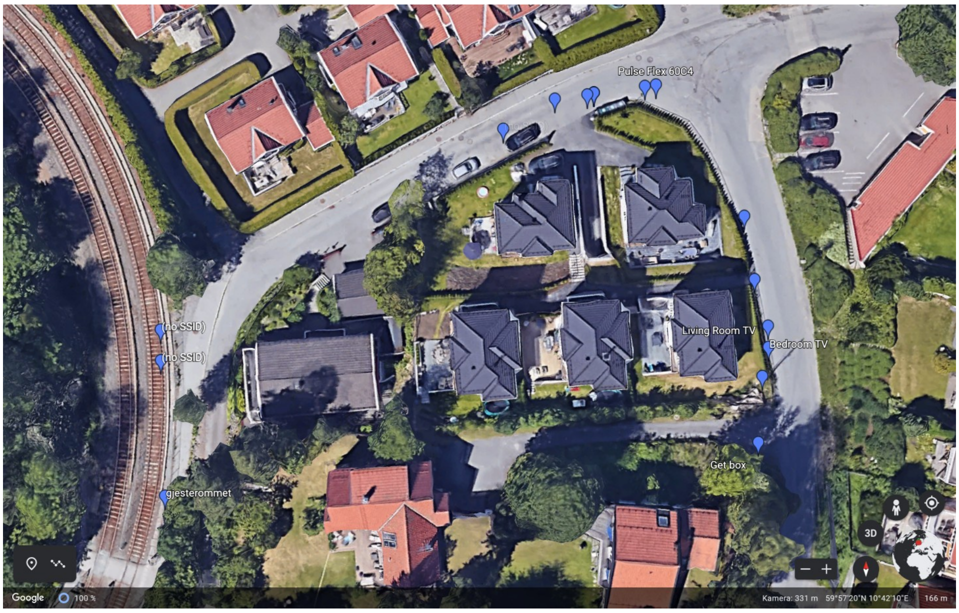

Fig. 2: A housing area with devices named Living Room TV, Bedroom TV and Guestroom, together with a pulse clock indicating presence of a person.

3.1.3 Highway pedestrian bridge

To evaluate the observability of car equipment, we sampled Oslo's ring road ("Ring 3") from a pedestrian bridge. Cars pass at 60km/h in two lanes in both directions. We were able

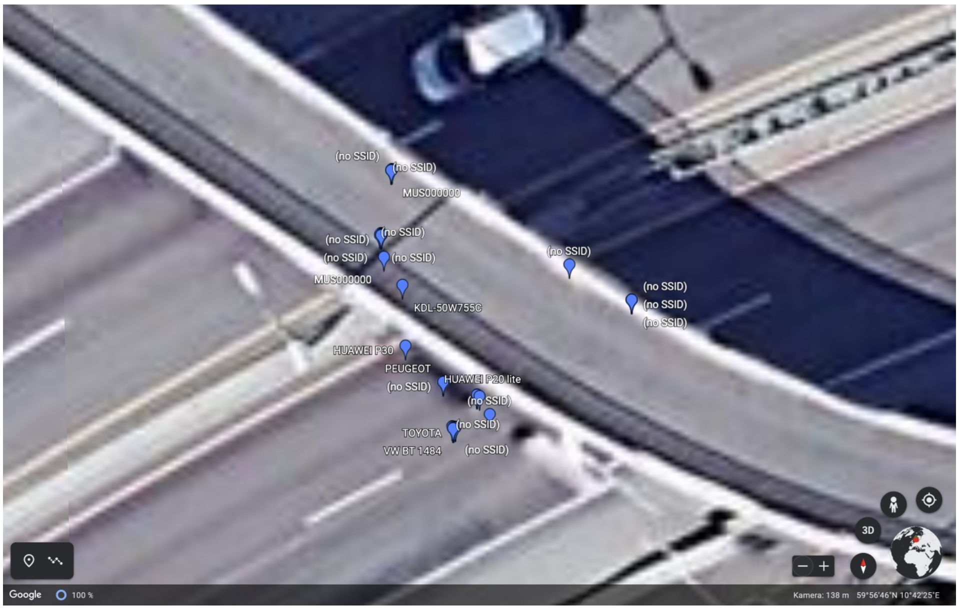

Fig. 3: Captured driving cars at 60-80 km/h from bridge over highway (Bluetooth): Toyota, Peugeot and Volkswagen. Other devices are likely hands-free devices and headsets.

to capture many wireless/hands–free devices, some of them named after the car make. There were even Wi-Fi networkss with car brand names in their SSID. In addition to the cars, we picked up various smartphones and tablets (see Fig. 4). It should therefore be relatively easy to track the car communication interfaces and the driver/passenger phones with a database.

3.1.4 Footpath through park to childcare facility

A last measurement was a stationary placement of a probe in a window next to a footpath on the edge of a public park in Oslo. The path leads to a childcare facility, and leads, in addition, to a bus stop in the area. It is well-used by pedestrians and bicyclists at daytime. Over an interval of three days, passing devices were captured. We sampled phones, headphones, fitness wristbands, smart speakers, GPS navigators, an occasional electric scooter and other devices (see Fig. 4). Interestingly, by looking up the brand names from our log file in online shops, we were able to assess the monetary value of the devices, as shown in Tab. 1. We detected well-known products on a market price scale ranging from 70€ to 500€ sales value from persons passing by our position.

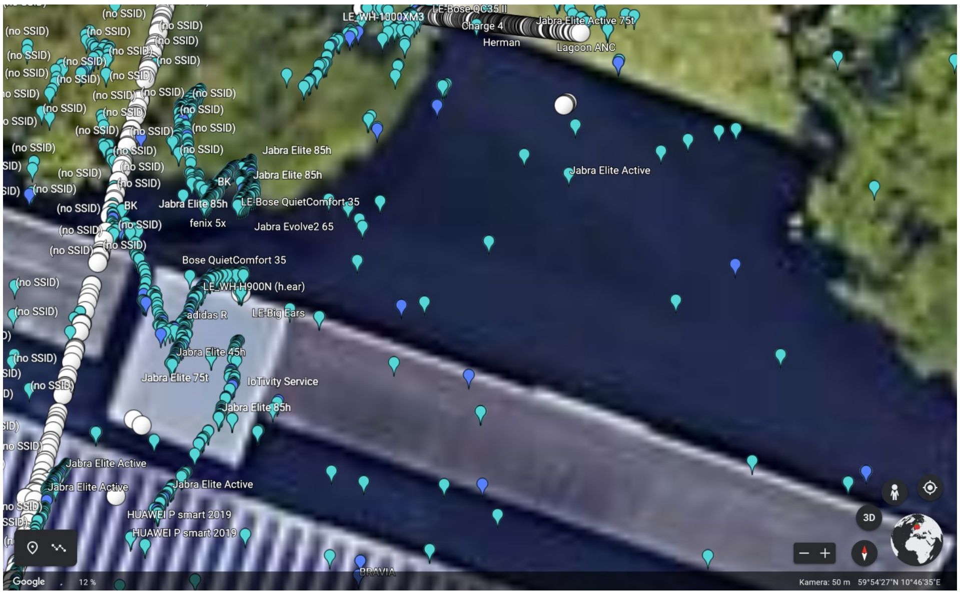

Fig. 4: Data collection about headsets on walkway along a park towards a childcare facility

Garmin Fenix 5x	500€	Bose quite comfort 35	230€
Jabra evolve2 65	200€	Sony WH-H900N h.ear	200€
Jabra Elite 85h	200€	Jabra Elite 75t	140€
Adidas R headphone	130€	Jabra Elite Active (65)	100€
Jabra Elite 45h	70€	Huawei P smart 2019	70€

Tab. 1: Detected devices and their approximate value (online purchase prices in Norway in 2021)

3.2 Issues

We met several issues that require further attention when capturing data.

Discoverable devices and hidden devices Paired and hidden devices are more challenging as sources for identity attributes. They do not advertise device names, do not interact with unpaired devices, and may use network address and device ID obfuscation techniques (see below). Such devices may therefore lack from a measurement sample when comparing or accumulating device status on a timeline.

Changing device identifiers Frequent changes of MAC addresses (as used in Apple devices) and randomness plus encryption as used in Bluetooth do hinder re-identification or might create false positives[4]. However, profiling operating system behavior and individual protocol implementation, devices can still get profiled [BLS19] with certain effort.

Positioning against addresses or points of reference Location is not very precise without further equipment, such as DGPS receivers. As seen in Fig. 4, stationary measurements have a 30–50m inaccuracy in urban space. Radio waves in addition reflect from buildings and other objects. Precise mapping of stationary devices into smaller spaces, such as apartments, will require multiple measurements or triangulation.

Dimensions of measurement It will require several measurements over a time span in order to decide whether a device is static in a location, or whether it is mobile. Differentiation between such devices will require several data points or continuous measurement. Device fingerprinting may be necessary against changing network addresses.

Ethical issues The easy availability of traceable personal devices in private spaces and around moving human beings poses threats to safety, privacy and health. Abusive applications include surveillance, stalking, assault, intrusions, commercial exploitation such as price discrimination, risk for theft, burglary, robbery or personal targeted assault. Andreas Pfitzmann warned against person-specific bombs that explode when certain person's RFID passport walks by [Pf07].

4 Approach for elaborate identity attribute extraction

We propose the following approach for collecting IoT attributes (see Fig. 5). Sensing of wireless communication can be executed by stationary sensors (e.g., mounted at lampposts at popular locations), by mobile sensors (e.g., attached to urban buses) or by a sensor network. A typical instance of the last category would be a large community collecting data using either their smart phones or some special equipment. The WIGLE community mentioned in the previous section is a well-known example for such a crowd-based sensor network.

All these sensors scan for wireless signals and collect header information from the link layer, e.g., Wi-Fi probe requests or Bluetooth advertisements, but potentially also from higher protocol layers. Typical information that can be gathered are network addresses (e.g, MAC or SSID) and device names (e.g., Bluetooth name). As mentioned before, it is not always possible to extract identifiers directly from the scanned data, but in most cases it is possible to profile a signal source and re-identify it, when it appears again at a later point in time. This aforementioned information is enriched with meta-data such as timestamp and the current location of the sensor. Additionally, with a combination of multiple readings from

[4] See Bluetooth LE Privacy, `https://www.bluetooth.com/blog/bluetooth-technology-protecting-your-privacy/`, accessed 2020-02-26.

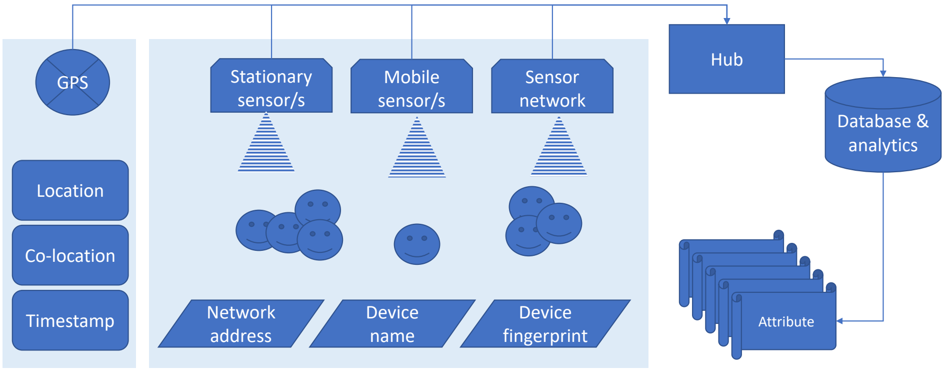

Fig. 5: IoT attribute sensing approach.

the same source, the sender's location can be calculated and stored. However, when storing the signal strength this can also be done later during the analysis phase.

The collected and stored data can then later be analyzed and further attributes can be derived. Here are some typical types of analysis that can be performed.

District profiling Regarding the sum of all collected scan results from a larger area can lead to interesting conclusions in many different aspects. Looking at the number of TVs or other media equipment gives an indication on the number of households or even residents. An analysis of the value of all scanned devices allows derivation of a district's socioeconomic status. In addition, the type of mobile devices can indicate the usage of an area. An example is a large number of fitness trackers on a popular jogging route.

Traffic observation Our experiment has shown that Bluetooth and Wi-Fi devices in cars can be scanned even when the car is driving by (with moderate speed). Nowadays, not all cars are fitted with such wireless equipment, but this will become more and more in the near future. This kind of scan allows for example analysis of traffic density (traffic jams) or characteristics of traffic (types of cars). As it is easily possible to identify a specific car and re-identify it later or at a different location also analysis of traffic flows are possible. Especially the last one is much harder with "traditional" traffic surveillance cameras, as it requires license plate recognition, which is expensive, error prone and does not conform to privacy regulations in many countries.

Home surveillance In sparsely populated areas like villages but also suburbs the location of a wireless device indicates uniquely to which house it "belongs". This allows surveillance of this house regarding the deployed IoT devices, the current status of inhabitants, visitors or trespassers. This information can be used for example as additional authentication factor for smart locks or for triggering a burglar alarm system. However, it can obviously also be used for malicious actions (like described in Section 3.2).

Person tracking The fingerprint of all devices typically carried by a specific person is in most cases unique. Thus, once a link between this device fingerprint and the person is established the person is traceable. This link can be created for example by a treacherous device identifier (e.g., “Mike’s iPhone”) together with other publicly available information or observing entering to which home the devices (and therefore the person) return in the evening. Tracking a person can be used for smart locks in houses or cars or flexible charging of public transport. However, it also poses a huge privacy treat to this person (like described in Section 3.2).

This list of use cases is of course not exhaustive. We plan to examine other possible applications in future work.

5 Conclusion

In this paper, we proposed an approach for identity attribute extraction from personal IoT devices using a network of sensors. We demonstrated the feasibility of data collection, which showed availability of identity attributes directly from device properties, or from their spatial or temporal context. However, data quality is of varying levels and can get greatly improved with additional measures.

In future work, we will research further analysis possibilities using machine learning methods. We plan to look into new applications for attributes extracted from wireless IoT sensing data. Possible use cases might range from additional authentication factors in security models through facility and people management applications up to enhancing consumer surveillance through harvested attributes.

On the other hand, we will carefully analyze possible threats posed by the availability of attributes and the ease of tracing devices and their owners. This may lead to unintended exploitation on all levels of severity, such as surveillance of persons, targeted theft, burglary or kidnapping, assault (e.g. from stalkers) or as targeted liquidations based on device identification, up to weaponization of IoT and the obtained data in acts of cyberwar [FF18].

References

[An19] Andersen, M.: Identification, Location Tracking and Eavesdropping on Individuals by Wireless Local Area Communications, MA thesis, Norwegian University of Science and Technology (NTNU), 2019.

[Bh19] Bhaskar, N.: A survey of techniques in passive identification of wireless personal devices and the implications on user tracking, tech. rep., Department of Computer Science, University of California San Diego, 2019, URL: https://cseweb.ucsd.edu/~nibhaska/papers/RE_paper_19.pdf.

[BLS19] Becker, J. K.; Li, D.; Starobinski, D.: Tracking anonymized bluetooth devices. Proceedings on Privacy Enhancing Technologies 2019/3, pp. 50–65, 2019.

[BZ19] Bruegger, B. P.; Zwingelberg, H.: Location Services can systematically track vehicles with WiFi access points at large scale, tech. rep., Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (ULD), 2019, URL: https://uld-sh.de/LStrack.

[CKB14] Cunche, M.; Kaafar, M.-A.; Boreli, R.: Linking wireless devices using information contained in Wi-Fi probe requests. Pervasive and Mobile Computing 11/, pp. 56–69, 2014, ISSN: 1574-1192, URL: https://www.sciencedirect.com/science/article/pii/S1574119213000618.

[Cu13] Cunche, M.: I know your MAC Address: Targeted tracking of individual using Wi-Fi. In: International Symposium on Research in Grey-Hat Hacking - GreHack. Grenoble, France, Nov. 2013, URL: https://hal.inria.fr/hal-00858324.

[FF18] Fritsch, L.; Fischer-Hübner, S.: Implications of Privacy and Security Research for the Upcoming Battlefield of Things. Journal of Information Warfare 17/4, pp. 72–87, 2018, ISSN: 14453312, 14453347.

[Fr08] Fritsch, L.: Profiling and Location-Based Services (LBS). In (Hildebrandt, M.; Gutwirth, S., eds.): Profiling the European Citizen: Cross-Disciplinary Perspectives. Springer Netherlands, Dordrecht, pp. 147–168, 2008, ISBN: 978-1-4020-6914-7, URL: https://doi.org/10.1007/978-1-4020-6914-7_8.

[Fr09] Fritsch, L.: Business risks from naive use of RFID in tracking, tracing and logistics. In: 5th european Workshop on RFID Systems and Technologies. VDE, pp. 1–7, 2009.

[MF20] Momen, N.; Fritsch, L.: App-generated digital identities extracted through Android permission-based data access - a survey of app privacy. In (Reinhardt, D.; Langweg, H.; Witt, B. C.; Fischer, M., eds.): SICHERHEIT 2020. Gesellschaft für Informatik e.V., Bonn, pp. 15–28, 2020.

[Pf07] Pfitzmann, A.: Personspezifische Bomben mit RFID-Pass, Neues Deutschland, 2007, URL: https://www.neues-deutschland.de/artikel/108709.regierung-baut-personenspezifische-bomben.html, visited on: 02/21/2020.

[PH10] Pfitzmann, A.; Hansen, M.: Anonymity, unlinkability, unobservability, pseudonymity, and identity management-a consolidated proposal for terminology. In: Designing privacy enhancing technologies. Technische Universität Dresden, pp. 1–9, 2010.

[TLT16] Tillekens, A.; Le-Khac, N.-A.; Thi, T. T. P.: A bespoke forensics GIS tool. In: 2016 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, pp. 987–992, 2016.

[VG13] Valeros, V.; García, S.: How bluetooth may jeopardize your privacy. An analysis of peoplebehavioral patterns in the street. Magdeburger Journal zur Sicherheitsforschung 3/, 2013, ISSN: 2192-4260.

[Wr08] Wright, D.; Gutwirth, S.; Friedewald, M.; Vildjiounaite, E.; Punie, Y.: Safeguards in a world of ambient intelligence. Springer Science & Business Media, 2008, ISBN: 978-1-4020-6661-0.