Collecting Identifying Data for Re-Identification of Mobile Devices carried at a Crime Scene using Wi-Fi Routers

Daniel Vogel[1], Markus Krämer[2]

Abstract: Mobile devices such as smartphones constantly probe the surrounding wireless medium for providing services to users, thus leaving traces that, when analysed, can link their presence to a crime. Those passive observations, combined with actively tricking devices to cooperate with foreign networks, can yield an attractive device profile. Though past research explores independent methods for collecting single identifiers of devices, a holistic approach for data seizure and utilization for law enforcement is missing. As devices carried by criminals are not willing to share identifiers with a victim's home network, we combine collection processes for device-identifying data with exploiting viable wireless communication protocols. Using the presented approach allows to build a system supporting law enforcement with evidence that certain mobile devices were present at a crime scene and thus may give clues on further investigations on suspects.

Keywords: Wi-Fi; MAC-Address; Cellular; IMSI; Evidence; Law Enforcement

1 Introduction

More than 100,000 thefts by burglary of a dwelling are registered only in Germany every year. Still, the crime clearance rate is rather low, fluctuating between 16-18%. [Po] Therefore, provisions to reach a higher clearance rate are desirable which may in addition lead to less burglaries happening. We aim at providing strategies for using home routers as alarm equipment and furthermore, as a collector of re-identifying information of burglars' technical equipment. While there are existing technologies for localising Wi-Fi devices via home routers [BP00, AHP17], these are out of the scope of this paper and are assumed to be working and given. Goal of this paper is creating a collection of Device Identifying Data (DID) types collectable via home routers, a classification of their usefulness in crime clearance and as these techniques collide with anti-tracking technologies, collecting workarounds against those techniques to allow a re-identification of devices brought by burglars. The rest of the paper is organized as follows:
Sect. 2 includes a motivation and our use case. Sect. 3 comprises an introduction of DID from the Wi-Fi and the cellular domain while Sect. 4 introduces collection processes for the DID. Related Work is discussed in Sect. 5. The paper will be concluded in Sect. 6.

[1] Universität Bonn, Informatik IV, Friedrich-Hirzebruch-Allee 8, 53115 Bonn, Germany vogel@cs.uni-bonn.de
[2] Universität Bonn, Informatik IV, Friedrich-Hirzebruch-Allee 8, 53115 Bonn, Germany kraemerm@cs.uni-bonn.de

doi:10.18420/inf2022_10

2 Background and Use Case

When breaking into homes, police intelligence shows that burglars often bring communication devices in order to keep in contact with their accomplices. When considering non-organized groups, it can be assumed that amateur burglars might bring their phone just in a lack of awareness.
This presents an detection opportunity as communication devices such as phones transmit signals even when currently not used that can be detected by any nearby receiver that is set up accordingly. The goal is to utilize the commonly existing home router as that receiver as Wi-Fi routers are ubiquitously available in most homes and provide a commonly used networking interface for mobile phones already. This scenario is visualised in figure Fig. 1.

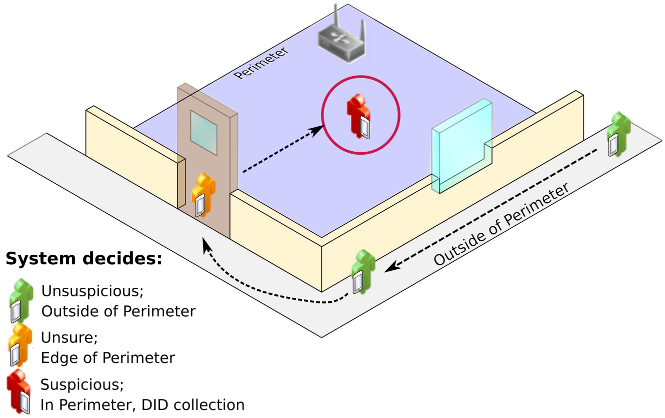

Fig. 1: Use case: Brought devices of different suspicion levels as detectable by the home router.

As soon as the router can eavesdrop packets from a mobile device it evaluates whether the device is located inside a certain perimeter or not. Therefore, three different states exist. Devices localised outside the perimeter are considered as unsuspicious. At the edge of the perimeter it might be unsure whether the device has entered the perimeter or not. Thus, the device will be observed more intensively. In case the the device has surely entered the perimeter, the system alarms the user and collects as much DID as possible.
However, as we assume that mobile phones of burglars are not set up to willingly connect to the victim's home network in order to collect their data, some detection and collection strategy needs to be employed. We present a strategy that looks to collect DID for a brought non-cooperative phone by its sent packets solely through the usage of a home router.
In this paper the following terminology will be used:

- **Target Device (TD)** denotes a mobile device, most commonly a mobile phone, that is brought by a malicious intruder and will be the target for re-identification.
- **Defensive Router (DR)** denotes the Commercial-Off-The-Shelf (COTS) router a home has already installed and is used to collect information on the TD.

Note that devices need to be located inside a defensive perimeter to be considerable as TDs. The federally founded german research project WACHMANN is currently researching exactly that scenario and thus we consider the requirement as satisfactory for the scope of this paper [WA].

2.1 Viable Wireless Technologies

Mobile phones use a number of wireless communication technologies depending on different factors such as brand, vendor, series and more. Since most wireless technologies try to serve different requirements, they vary in both their technical specification, protocols, and transmitted data. When deciding on whether a technology is suitable for our use case, the capabilities of the chosen receiver as well as the capabilities of the technology itself have to be considered. As the home router shall be utilized as the DR, we restrict the technically available wireless technologies. Not all wireless technologies commonly used by mobile phones are supported by COTS routers, such as Bluetooth or cellular. In short, the wireless communication technology Wi-Fi provides both the most promising approaches as well as the highest level of compatibility between existing home routers and the most commonly expected TD, being mobile phones.

Mobile phones, when not connected to a Wi-Fi network, will only use certain frames [IE21]. These frames comprise management frames used for network discovery or general control frames. Since these frames are of limited use for device re-identification as presented in Sect. 3.1, it may prove useful to entice the TD to send more and preferably different frames. Applying enticing strategies accordingly will allow the DR to receive different Wi-Fi frames in certain situations, thus allowing for further information gathering. In Sect. 4 we present different strategies, a DR can employ to receive packets from a TD.

2.2 Criteria for Device Identifying Data

Concerning Wi-Fi packets hovering around, such as probe requests, many data types can be extracted from these. Even more data types can be collected through forcing more and tailored communication with foreign Wi-Fi devices. However, not all of these data types can be considered as DID. While specific DID are given in Sect. 3 the following catalogue of criteria should give an imagination of features needed for data types being DID:

- **Technical traceability:** There must be a technical possibility to trace collected DID. This includes reproducability at other locations at a later point of time.
- **Legal traceability:** Due to the strict data privacy guidelines of the german law BDSG [Re14] and the european GDPR [Ge] the use of DID may be restricted to specific DID or specific use cases.

- **Uniqueness:** DID must be unique or at least collision free in the considered area.
- **Re-identifiability:** The given DID allows a re-identification of wireless devices and is helpful in detecting criminals.
- **Unforgeability:** DID should be hard to forge.

3 Device Identifying Data

Some DID are useful by design and used for exactly that purpose, to identify and address unique devices in a heterogeneous wireless network environment. Other information are not designed specifically to allow identification but support other services. These, however, can be utilized as DID when viewed in combination with each other and in context of the use case.
DID are used to solve two distinct tasks. Firstly, in the context of an ongoing burglary, DID are used to attribute packets to unique devices. This is important as the DR receives packets from potentially many devices in a certain time frame, many of whom are not carried by burglars. Correctly attributing the sent packets from the TD to their original sending device allows for a complete overview on what has been sent by this device during the whole duration the DR was able to observe it. Not only is this information crucial on gathering intelligence on the device and its behaviour, being able to tell individual devices apart by their packets assists packet-based device positioning. Secondly, once sufficient DID were collected for a TD, they can be used to re-identify that TD at a later time. Correctly re-identifying a mobile device and linking it to a crime scene, where its attributing DID were collected, is the main goal to achieve as presented in this paper.

3.1 Device Identifying Data from Wi-Fi

Since we selected Wi-Fi as the wireless communication technology to focus on, Wi-Fi-based DID will be shortly summarized now. We distinguish between *direct* DID and *indirect* DID. It must be noted that there are more information extractable from Wi-Fi packets than listed here, such as Channel State Information (CSI), or Time of Flight (ToF). We omitted these as they either have uses primarily for device positioning, which is not within the scope of this paper, or do not majorly assist as device identifiers in a way that a COTS router can make use of.

Direct DID are DID that allow device identification directly, meaning that they can be extracted directly from the received packets without the need of further interpretation or context. For our use case we consider Media Access Control (MAC) addresses and Information Elements (IEs). In Wi-Fi, MAC addresses are used for addressing individual devices in a wireless networking scenario and are meant to be a unique device identifier by design [IE21]. Thus, in their originally intended use, they serve as a direct DID for

our purpose as they allow to directly attribute packets containing certain MAC addresses to individual devices. In practice, modern mobile phones tend to utilize MAC address randomization to enhance the privacy of these devices [Va16, Fe21]. As MAC addresses may provide complications and can therefore not solely be relied on for device re-identification, other approaches for DID are considered.
Wi-Fi probe requests and responses contain what are called IEs. IEs contain information used for different uses, ranging from supported rates, requested Service Set Identifiers (SSIDs), up to Wi-Fi features specific information such as Wi-Fi Protected Setup (WPS) data. Research shows that IEs allow to generate fingerprints for packets, which can allow to attribute packets correctly to individual devices, especially when combined with an analysis of sequence numbers. [Va16, Ma17]
Due to fabrication inaccuracies mobile devices could be identified on specific characteristics during Wi-Fi transmissions, such as minor deviations in the used transmission frequencies [US07]. We consider DRs however to not be sufficiently sensitive receivers to utilize physical device profiles as DID.

Indirect DID are not by themselves of great use, often because they are not unique per device, but can provide additional insights which may amplify the usefulness of direct DID in re-identification. We consider information on time, vendor, Operating Systems (OSs), SSIDs, User Equipment (UE) usage, and logical device profiles as indirect DID.
Time information can proof useful in the context of both sketching the process of an ongoing burglary as well as assisting the law enforcement in giving a burglary context between other burglaries. Mobile phones or UEs regularly probe the surrounding wireless medium for available networks, which are identified by their SSIDs. Whether a UE probes using SSIDs or using wildcard probe requests affects both their power efficiency but also their privacy [Da18, KA05]. Using knowledge about SSIDs in combination with services such as WiGLE [bau] allows for further tracking opportunities.
UEs are further shown to send packets differently depending on their usage. For example different mobile phones wait for varying time slots between sending packets based on whether the display is locked or unlocked, or whether their Wi-Fi is enabled or not [Wa13, MCT17, Fr15]. Using this knowledge, it is possible to infer the usage of a UE based on the sent packets and their amount within a time frame.

3.2 Device Identifying Data from the Cellular Domain

From the cellular domain, three different DID types have been identified which strictly speaking are not DID of the mobile phone but identifier of the SIM card of the mobile user. However, this information still is really helpful to get information about the mobile user or the International Mobile Equipment Identity (IMEI) of the used device, as pairs of International Mobile Subscription Identity (IMSI) and IMEI are stored by the cellular provider for up to seven days. This allows law enforcement to study the history of the DID IMEI through the IMSI. Cellular networks also use identifiers such as the TMSI and GUTI, where it is unclear, whether these identifiers are stored in a way that they can be useful to our use case. [ET21]

4 Methods for Data Collection

A holistic approach for data seizure and utilization for law enforcement given our use case (c.f. Sect. 2) is yet missing. To formulate such an approach, we compiled various collection methods for DID presented in past research and enriched this compilation with known attacks on commonly uses Wi-Fi communication protocols. The three main strategies for DID collection are:

1. Passively receiving the packets the TD is sending ordinarily.
2. Actively engaging the TD to increase the amount and variety of packets sent by it.
3. Actively engaging the TD, tricking it to try to connect to a network, the DR is claiming to provide access to.

These strategies can be used in combination or be layered into different modes of operation, the DR can employ given its burglary detection certainty, especially when taking privacy into account.

4.1 Passive Data Collection

Passively receiving packets sent by the TD and all other devices nearby means that no further active communication shall be issued by the DR other than what it usually would do to provide its service to nearby devices. Usually this means that the TD would only engage minimally with the DR by occasionally sending Wi-Fi frames, such as probe requests. Even this minimally invasive strategy will yield DID as present in probe requests albeit in a limited supply. Observables are MAC addresses, although likely randomized for probe requests, IEs usable for fingerprinting, physical and logical device profiles, and most indirect DID, depending on how the TD is used and set up. It is very much possible to observe further transmissions by the TD, if it is actually communication e.g. with some nearby network to further collect more information.

4.2 Active Data Collection

As alluded to earlier, receiving more and different packets from the TD assists in creating a more reliable fingerprint. Past research shows, how certain packets like data [AA20] or Ready To Send (RTS) [Wa13] packets can be utilized to force responses like Acknowledgement

(Ack) or Clear To Send (CTS) packets even from non-cooperative devices. Abedi et al. suspect that the required response times as defined by the Wi-Fi standard is not sufficient to check the validity of said packets [AA20].
Certain Wi-Fi features such as *WPS* or *Passpoint* (also called *Hotspot 2.0* in *Wi-Fi Vantage* [Wi20]) may allow for additional communication between two unconnected devices, as these features are specifically designed to aid in establishing connections. A DR could be set up to transmit Passpoint-enabled and WPS-enabled packets aiming to collect information from TD supporting these Wi-Fi features.

4.3 Router-based Attacks for Data Collection

The third approach to gathering more information on the TD focuses on exploiting vulnerabilities of Wi-Fi protocols. While MAC-randomization usually is enabled for sent packets in an unconnected state, a successful connection requires the mobile device to be addressable by the network. As such, mobile devices will either revert to using their original Wi-Fi MAC address or a MAC address that is rolled for this network specifically, when attempting to associate to a network. In the latter case, we call that MAC address a session MAC address and mobile phones do maintain an internal database of network SSIDs and corresponding session MAC addresses.
The primary goal is to trick the TD to try to connect to a network that the DR is advertising, using so called association attacks. These association attacks have been discussed thoroughly as a way to attack MAC randomization, as the internal memory of the TD would directly allow linking the TD to any network it has a database entry for. It is important to note that for most known attacks these will not result in a fully established Wi-Fi connection with a malicious Access Point (AP). As the desired information can however be extracted just from the attempt of a connection establishment, a fully established connection is not actually required.
There are several attacks presented in past research that aim to either hijack an existing Wi-Fi connection or trick a mobile device to connect to an AP, revealing its sensitive information. Once such an association attacks succeeds and the TD tries to establish a connection to the DR, additional DID can be collected to improve both the fingerprinting precision and thus the likelihood of a successful re-identification of the TD. Furthermore, in [OBH17] O'Hanlon et al show, how Extensible Authentication Protocol (EAP) can be exploited to build a Wi-Fi Based IMSI Collector (WBIC). Using this technique, we were able to successfully extract the IMSI of TD in certain scenarios using an SSID that is pre-configured on certain SIM cards used in Germany.

4.4 Resulting Unified Collection Approach

The presented approaches can be unified into a unified collection approach. The DR is set up to monitor Wi-Fi packets passively. Wi-Fi packets that it receives will be used for device fingerprinting using direct and indirect DID. These fingerprints will be used to differentiate between packets from different devices and can be used for any implemented indoor positioning approach to identify TDs. Once, a TD has been identified, meaning a device has been spotted inside the protected perimeter, both active data collection and data collection attacks can be utilized to gather even more information to improve the fingerprint for the TD. This fingerprint may then be used to re-identify the TD.

Assuming a fingerprint has been generated that is sufficient in re-identifying the TD correctly, the device must first be available for comparison. Though this is not focus of this paper, some directions include direct intervention by law enforcement, cell site analysis (if an IMSI has been collected), or various MAC or SSID tracking techniques. Generated fingerprints could further be used to link crimes with each other, if identical fingerprints have been generated on different occasions.

5 Related Work

From our research it has become clear that Wi-Fi as a technology is presenting viable leads to police investigation that is not well integrated into common investigation mechanisms, primarily because there is no reliable infrastructure to capture these DID. Since the research in the field of digital forensics is vast and we cited many relevant publications, we selected some examples to present, just to highlight further work in nearby use cases.

The analysis of wireless traces for device detection has been discussed e.g. in [Bu20], where they discuss that criminals may set up a system to detect nearby mobile devices carried by law enforcement officers in order to warn themselves against an upcoming police raid. Here too, previously collected DID are utilized to correctly detect mobile devices nearby, though the intention is malicious.

Wi-Fi device identification is also relevant to network administrators to prevent malicious participants to wreak havoc in their network, potentially exploiting an open network, as Yu et al present in [Yu20]. Here, passively collected packets are fed into a deep learning framework to assess identity anomalies and allow device classification.

6 Discussion and conclusions

In this paper we presented a holistic approach in analyzing DID of intruders' mobile phones leading to new ways in crime clearance through digital forensics. We examined Wi-Fi as a viable technology and its receivable packet types. In addition, we featured a catalogue of criteria for classifying data types as DID. Using this catalogue we worked out the DID from the Wi-Fi and cellular domain. Furthermore we presented collection processes for these and workaround for existing problems, such as MAC randomization.

Concerning the practical use an effective countermeasure might be to shut the smartphone down or even do not bring it to burglaries. While this would neutralize our DR, on the other hand the technical possibilities for communication in burglary groups would decrease.

As an effective approach device-free localization [Zh21, Da19] could detect intruders without a need for carrying Wi-Fi devices, but is not able to collect DID then. There is a race between anti-tracking techniques such as MAC randomization which affects crime clearance and countermeasures such that further research is needed in the future to keep a working system. We stress, that the presented approaches must be evaluated clearly with regards to precision, presented criteria, as well as conformity with applicable (privacy) laws in the context of burglaries.

Acknowledgement

The work on this paper has been funded by the German Federal Ministry of Education and Research (BMBF) for the project WACHMANN (WLAN-based recording of characteristics of mobile devices near a crime scene for alarming and tracing property crimes) under grant number 13N15359.

Bibliography

[AA20] Abedi, Ali; Abari, Omid: WiFi Says"Hi!"Back to Strangers! In: Proceedings of the 19th ACM Workshop on Hot Topics in Networks. pp. 132–138, 2020.

[AHP17] Ali, Muhammad Usman; Hur, Soojung; Park, Yongwan: Locali: Calibration-free systematic localization approach for indoor positioning. Sensors, 17(6):1213, 2017.

[bau] bobzilla; arkasha; uhtu: , WiGLE: Wireless Network Mapping. https://wigle.net. Accessed: 2021-05-20.

[BP00] Bahl, Paramvir; Padmanabhan, Venkata N: RADAR: An in-building RF-based user location and tracking system. In: Proceedings IEEE INFOCOM 2000. Conference on computer communications. Nineteenth annual joint conference of the IEEE computer and communications societies (Cat. No. 00CH37064). volume 2. Ieee, pp. 775–784, 2000.

[Bu20] Bug, Steffen: Wenn der Schutzmann nicht mal klingelt - Über das automatische Erkennen von und Warnen vor der Staatsgewalt vor des Straftäters Haustür. 06 2020.

[Da18] Dagelić, Ante; Perković, Toni; Vujatović, Bojan; Čagalj, Mario: SSID oracle attack on undisclosed Wi-Fi preferred network lists. Wireless Communications and Mobile Computing, 2018, 2018.

[Da19] Dang, Xiaochao; Si, Xiong; Hao, Zhanjun; Huang, Yaning: A novel passive indoor localization method by fusion CSI amplitude and phase information. Sensors, 19(4):875, 2019.

[ET21] ETSI: . TS 123 003 V16.5.0 Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; 5G; Numbering, addressing and identification (3GPP TS 23.003 version 16.5.0 Release 16), January 2021.

[Fe21] Fenske, Ellis; Brown, Dane; Martin, Jeremy; Mayberry, Travis; Ryan, Peter; Rye, Erik C: Three Years Later: A Study of MAC Address Randomization In Mobile Devices And When It Succeeds. Proc. Priv. Enhancing Technol., 2021(3):164–181, 2021.

[Fr15] Freudiger, Julien: How talkative is your mobile device? An experimental study of Wi-Fi probe requests. In: Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks. pp. 1–6, 2015.

[Ge] General Data Protection Regulation. https://gdpr-info.eu/. Accessed: 2022-06-01.

[IE21] IEEE: IEEE Standard for Information Technology–Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks–Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Std 802.11-2020 (Revision of IEEE Std 802.11-2016), pp. 1–4379, 2021.

[KA05] KARMA Attacks Radioed Machines Automatically. https://web.archive.org/web/20210401071739/http://theta44.org/karma/index.html, 2005. Accessed: 2021-04-01.

[Ma17] Martin, Jeremy; Mayberry, Travis; Donahue, Collin; Foppe, Lucas; Brown, Lamont; Riggins, Chadwick; Rye, Erik C.; Brown, Dane: , A Study of MAC Address Randomization in Mobile Devices and When it Fails, 2017.

[MCT17] Matte, Célestin; Cunche, Mathieu; Toubiana, Vincent: Does disabling Wi-Fi prevent my smartphone from sending Wi-Fi frames? PhD thesis, Inria-Research Centre Grenoble–Rhône-Alpes; INSA Lyon, 2017.

[OBH17] O'Hanlon, P.; Borgaonkar, R.; Hirschi, L.: Mobile Subscriber WiFi Privacy. In: 2017 IEEE Security and Privacy Workshops (SPW). pp. 169–178, 2017.

[Po] Polizeiliche Kriminalstatistik. Edited: 2021-02-20.

[Re14] Recht, G.: Bundesdatenschutzgesetz (BDSG). G. Recht, 2014. ISBN:978-1-500-10002-5.

[US07] Ureten, Oktay; Serinken, Nur: Wireless security through RF fingerprinting. Canadian Journal of Electrical and Computer Engineering, 32(1):27–33, 2007.

[Va16] Vanhoef, Mathy; Matte, Célestin; Cunche, Mathieu; Cardoso, Leonardo; Piessens, Frank: Why MAC Address Randomization is not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms. In: ACM AsiaCCS. Proc of ACM AsiaCCS, Xi'an, China, May 2016.

[WA] WACHMANN: WLAN-basierte Aufzeichnung von Charakteristiken tatortnaher mobiler Endgeräte zur Alarmierung und Nachverfolgung von Eigentumskriminalität.

[Wa13] Wang, Wei; Joshi, Raj; Kulkarni, Aditya; Leong, Wai Kay; Leong, Ben: Feasibility study of mobile phone WiFi detection in aerial search and rescue operations. In: Proceedings of the 4th Asia-Pacific workshop on systems. pp. 1–6, 2013.

[Wi20] Wi-Fi Alliance: . Wi-Fi CERTIFIED Passpoint Technology Overview, December 2020.

[Yu20] Yu, Lingjing; Luo, Bo; Ma, Jun; Zhou, Zhaoyu; Liu, Qingyun: You Are What You Broadcast: Identification of Mobile and IoT Devices from (Public) WiFi. In: 29th USENIX Security Symposium (USENIX Security 20). USENIX Association, pp. 55–72, August 2020.

[Zh21] Zhang, Yunwei; Wang, Weigang; Xu, Chendong; Qin, Jie; Yu, Shujuan; Zhang, Yun: SICD: Novel single-access-point indoor localization based on CSI-MIMO with dimensionality reduction. Sensors, 21(4):1325, 2021.