UJI Probes: Dataset of Wi-Fi Probe Requests

BRAVENEC, T.; TORRES-SOSPEDRA, J.; GOULD, M.; FRÝZA, T.

2023 IEEE 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN)

eISBN: 979-8-3503-2011-4

DOI: https://doi.org/10.1109/IPIN57070.2023.10332508

Accepted manuscript

UJI Probes: Dataset of Wi-Fi Probe Requests

Tomáš Bravenec*†, Joaquín Torres-Sospedra‡, Michael Gould* and Tomas Fryza†
*Institute of New Imaging Technologies, Universitat Jaume I, Castellón, Spain
†Department of Radio Electronics, Brno University of Technology, Brno, Czech Republic
‡Algoritmi Research Centre, University of Minho, Guimarães, Portugal
bravenec@uji.es – jtorres@algoritmi.uminho.pt – gould@uji.es – fryza@vut.cz

***Abstract*—This paper focuses on the creation of a new, publicly available Wi-Fi probe request dataset. Probe requests belong to the family of management frames used by the 802.11 (Wi-Fi) protocol. As the situation changes year by year, and technology improves probe request studies are necessary to be done on up-to-date data. We provide a month-long probe request capture in an office environment, including work days, weekends, and holidays consisting of over 1 400 000 probe requests. We provide a description of all the important aspects of the dataset. Apart from the raw packet capture we also provide a Radio Map (RM) of the office to ensure the users of the dataset have all the possible information about the environment. To protect privacy, user information in the dataset is anonymized. This anonymization is done in a way that protects the privacy of users while preserving the ability to analyze the dataset to almost the same level as raw data. Furthermore, we showcase several possible use cases for the dataset, like presence detection, temporal Received Signal Strength Indicator (RSSI) stability, and privacy protection evaluation.**

***Keywords*—Dataset, Privacy, Probe Requests, RSSI, Wi-Fi, Wireless Communication, WLAN**

I. Introduction

The growing interest in indoor positioning and indoor navigation also raises questions related to privacy. Unlike Global Navigation Satellite Systems (GNSS), where the whole world uses the same coordinate systems and User Equipments (UEs) can calculate their own location from the received information, indoor location in most cases requires cooperation between the UE and the positioning infrastructure. This creates a window of opportunity, through which the privacy of users can be breached. This is especially true when it comes to Wi-Fi since the management frames used for probing the environment for nearby Access Points (APs) are not encrypted.

These management frames of Wi-Fi have been the center of attention of researchers for years now. Back before the introduction of Media Access Control (MAC), Musa and Eriksson [1] used the real MAC addresses in probe requests for mobility tracking. Then the Sapienza probe request dataset [2], [3] was published including the analysis of social relationships from the probe requests. Before the implementation of MAC address randomization Cunche, Kaafar, and Boreli [4] and Cheng *et al.* [5] showed the vulnerability of probe requests to tracking techniques. Then in 2014, the MAC address randomization was introduced by Apple in the iOS version 8 [6].

Since Apple introduced the randomization of MAC addresses, the research community explored the weaknesses of this privacy-related measure. Martin, Rye, and Beverly [7] explored the ways of revealing the globally unique MAC addresses of devices employing MAC address randomization. Just a year later Freudiger [8] worked on reverse engineering MAC address randomization. In 2016, Di Luzio, Mei, and Stefa [9] analyzed probe requests captured in Italy during political events, trying to infer the origin of people participating in the events. Their results did indeed match the officially published voting reports. There are also other possible ways to analyze probe requests: Matte *et al.* [10] exploited the time difference between subsequent probe requests to identify different devices. In 2021 Fenske *et al.* [11] published a deep look into user privacy protection, which was created as a follow up to a similar study published few years before it by Martin *et al.* [12].

A. Probe Requests

Probe requests belong to the management frames family of the IEEE 802.11 standard [13]. These frames are specific in that they are sent without any encryption and any device with monitoring mode available in their wireless interface driver can capture and read these frames. The primary purpose of probe requests is the detection of nearby Wi-Fi networks. Primarily its purpose was to find a known AP to connect to, but a secondary purpose came with the popularity of Wi-Fi. That is for acquiring a rough location estimate, by matching the Service Set Identifier (SSID) of nearby AP to either public database of APs [14] or private database of the system developer [15], [16].

The probe requests consist of a header and an information element. The header is not probe request specific and contains the frame control number (4 for probe requests), MAC addresses of the destination and source devices (with destination address being most often broadcast address - *ff:ff:ff:ff:ff:ff*) and sequence number of the frame. When the broadcast address is used, the packet is targeted to every device in the network, or in the case of wireless communications to every device in the proximity. This is important so every Wi-Fi AP receives the probe request. There can be a lot of information included in

The authors gratefully acknowledge funding from European Union's Horizon 2020 Research and Innovation programme under the Marie Skłodowska Curie grant agreement No. 813278 (A-WEAR: a network for dynamic wearable applications with privacy constraints, http://www.a-wear.eu/) and No. 101023072 (ORIENTATE: Low-cost Reliable Indoor Positioning in Smart Factories, http://orientate.dsi.uminho.pt/). This work does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of its content.

Time of Arrival	Radio Interface Information
Antenna	Channel
RSSI	More Optional Information

Frame Control = 4	Optional
Duration	SSID
Destination Address	Supported Rates
Source Address	Extended Supported Rates
BSS ID	HT Capabilities
Sequence Number	VHT Capabilities
Information Element	Vendor Specific
	More optional Information

Fig. 1. Probe request structure.

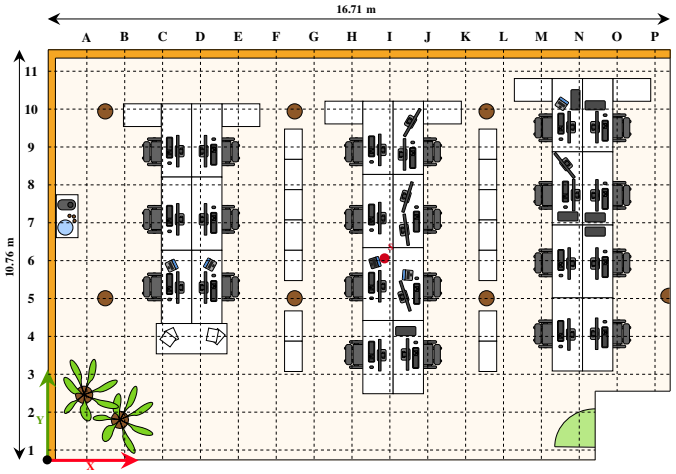

Fig. 2. Floor plan and location of sniffer in the office space of GEOTEC department at UJI, Spain.

the probe requests in the information element section. It can contain SSID, in cases the device is searching for a specific network, otherwise, this field can stay empty. The rest of the fields contain information about the supported transfer speeds, other capabilities of the UE, information about the manufacturer of the wireless interface, and in some cases Wi-Fi Protected Setup (WPS) field.

The probe request frame does not contain information about the transmission time, which means the Time of Arrival (ToA) is taken from the Real Time Clock (RTC) timer of the receiver. Similar to the ToA, but dependent on the implementation of the capture device, the radio information can be extracted from the wireless interface. The most interesting information that can be gathered from the wireless interface is the channel on which the frame was received, to which antenna in case the system has more than one, and the Received Signal Strength Indicator (RSSI) of the captured frame. The structure of the probe request, including the information gathered by the wireless interface of the receiver is visualized in Fig. 1.

B. Existing Probe Request Datasets

There are several publicly available datasets of probe requests. The datasets are described in the following list, including their advantages and disadvantages:

- **Sapienza2013:** is the most commonly known probe request datasets [2], [3]. It was published in 2013 which is before the implementation of MAC randomization algorithms in iOS, Android, or Windows. This means the dataset does not provide the same information as would be found in the real world. The anonymization in this dataset uses pseudonyms for SSIDs.
- **Glimps2015:** was captured at Glimps 2015, a music festival that took place in Ghent, Belgium [17]. There are several issues with this dataset. First, only one probe request is collected per unique MAC address, as well as timing issues due to the capture style. The anonymization techniques used in this dataset prove problematic for analysis as well. The dataset uses hashes for most of the information element fields, however, all of the sequence numbers are set to 0 and all SSIDs are hidden. This makes the analysis of this dataset quite difficult or straight-up impossible as the data of interest are hidden or missing.
- **Nile2021:** was captured during the night in a shopping center and it is the only dataset [18] not applying any anonymization which is perfect for analysis, however, the length of the data capture is just 40 minutes. This is limiting to the analysis.
- **IPIN2021:** is our previous dataset, published as supplementary materials for a case study performed at IPIN 2021 [19], [20]. The dataset was collected over the course of only 4 days and our capture device lacked the capability of storing the radio information as well. The anonymization of this dataset was done using SHA512 overall privacy-sensitive fields, which does not prevent analysis while preserving user privacy.

In this work, we have decided to create a new probe request dataset without the previous shortcomings. This dataset covers the time of 1 month and uses hashing for anonymization. This preserves user privacy and makes it impossible to match the data to real people while allowing full analysis on the same scale as non-anonymized data.

C. Outline

The paper is divided into several sections, in Section II we discuss the technology used for the creation of the dataset. Following is the Section III describing the captured data. Some of the use case examples are presented in Section IV. The very important discussion about ethics is in Section V. And finally, the conclusions and brief look at our future work are in Section VI.

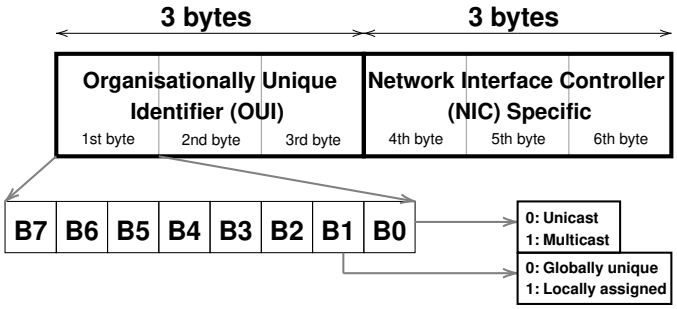

Fig. 3. Structure of MAC address with the functional bits.

II. Dataset Creation

The dataset was collected over the time span of 1 month, specifically during the month of March 2023, in our office at University Jaume I, Spain. For collection, the ESP32-based sniffer from our previous works [20]–[22] was used. The firmware for the ESP32 and all of its variants are publicly available from GitLab repository [23].

The sniffer captures raw packets and stores them in a standardized packet capture file, compatible with network analysis tools like Wireshark, Python package scapy etc. The radio information is not part of the official 802.11 frames, as such we implemented saving of the Radiotap headers [24] in conjunction with the 8o2.11 frames. This ensures compatibility with any packet analysis tools while providing us with additional information about the radio properties of the captured frame as well as preserving all captured information in the frame.

III. Data description

As mentioned before, the dataset was collected in the office of the GEOTEC department of University Jaume I, Spain. The office has a rectangular shape with dimensions of $16.71\,\mathrm{m}$ by $10.76\,\mathrm{m}$. The office is of open-space style with a typical occupancy of 14-20 people. The floor plan of the office is depicted in Fig. 2.

The probe requests were collected during the month of March, to capture regular work weeks and also a local holiday, Magdalena 2023, during which the university was mostly closed. Magdalena 2023 started on Saturday 11$^{\text{th}}$ March, 2023 and ended a week later on Sunday 19$^{\text{th}}$ March, 2023. The reduced occupancy of the office is clearly visible from Fig. 4, where the holiday of Magdalena are visualized in light green color.

There are some noticeable trends present in the dataset, that are visible in Fig. 4. One of these trends is the noticeable constant transmission of probe requests. During the entire collection time, probe requests were being sent at all times, be it during the day, night, workday, or weekends. The explanation for the probe requests captured at night can be all-in-one computers using Wi-Fi instead of a wired connection, phones, or Internet of Things (IoT) devices used for experiments in the office that can be transmitting probe requests at all times. The second noticeable thing is a peak in the transmitted probe requests happening every day around 07:00. This is due to the scheduled reboot of a Wi-Fi access point present in the office, and devices searching for a network to connect to after being disconnected from it. Another noticeable thing is a short time period at night of the Sunday 26$^{\text{th}}$ March, 2023 with no captured probe requests. That is due to the switch to the Summer Time when the time changed from 02:00 to 03:00.

We can also observe that some of our colleagues went to the office during some of the days of Magdalena and during the first weekend since the beginning of the monitoring. This explains the lower, but still detectable, increase in captured probe requests.

A. MAC addresses and SSIDs

Another important factor in the analysis of probe requests is MAC addresses of devices. In case of devices not using randomization of MAC addresses, such UE is very easy to track. Randomized MAC address is also very easily identifiable, due to the 2nd least significant bit **B1** in the first byte of the MAC address. When this bit is set, the MAC address was randomized by the network controller of the UE. The least significant bit of the first byte **B0** distinguishes individual devices and device groups. The structure of the MAC address is presented in Fig. 3. Considering this limitation of the 2 least significant bits, the 2nd digit of locally assigned MAC address in hexadecimal format has only four options: 2 (0010), 6 (0110), a (1010) or E (1110). From the captured probe requests, about $35\,\%$ used randomized MAC addresses, which is presented in Fig. 5.

The second very important parameter is the Preferred Network List (PNL), which is the list of network SSIDs the UE often connects to. This can be leaked by the device sending probe requests targeted for specific networks. In this dataset, $19\,\%$ of the captured probe requests contained an SSID, which is presented in Fig. 5. Even though this seems like a fairly low number, the dataset contains 2030 unique SSIDs.

B. Information Elements

This optional section of probe requests is very useful for fingerprinting, as it presents us with a set of capabilities and supported functions that are not changing over time for a single UE. The information ranges from supported data rates, capabilities related to certain Wi-Fi standards (High Throughput (HT) capabilities -802.11n, Very High Throughput (VHT) capabilities - 802.11ac, and High Efficiency (HE) capabilities - 802.11ax), vendor-specific elements (some devices had several of them) to WPS fields which can contain many user identifying information ranging from the device manufacturer, all the way to device name (and since many devices are named by users - *Julia's iPhone*, this can leak the name of the user). WPS fields also come with Universally Unique IDentifier-Enrollee (UUID-E), also a unique identifier that can compromise the anonymity while using randomized MAC addresses as it does not change over time. The occurrences of each field are shown in Table I.

C. Radio Information

The ESP32 sniffer was configured to gather probe requests in all channel scan. Due to the lack of support of ESP32 for

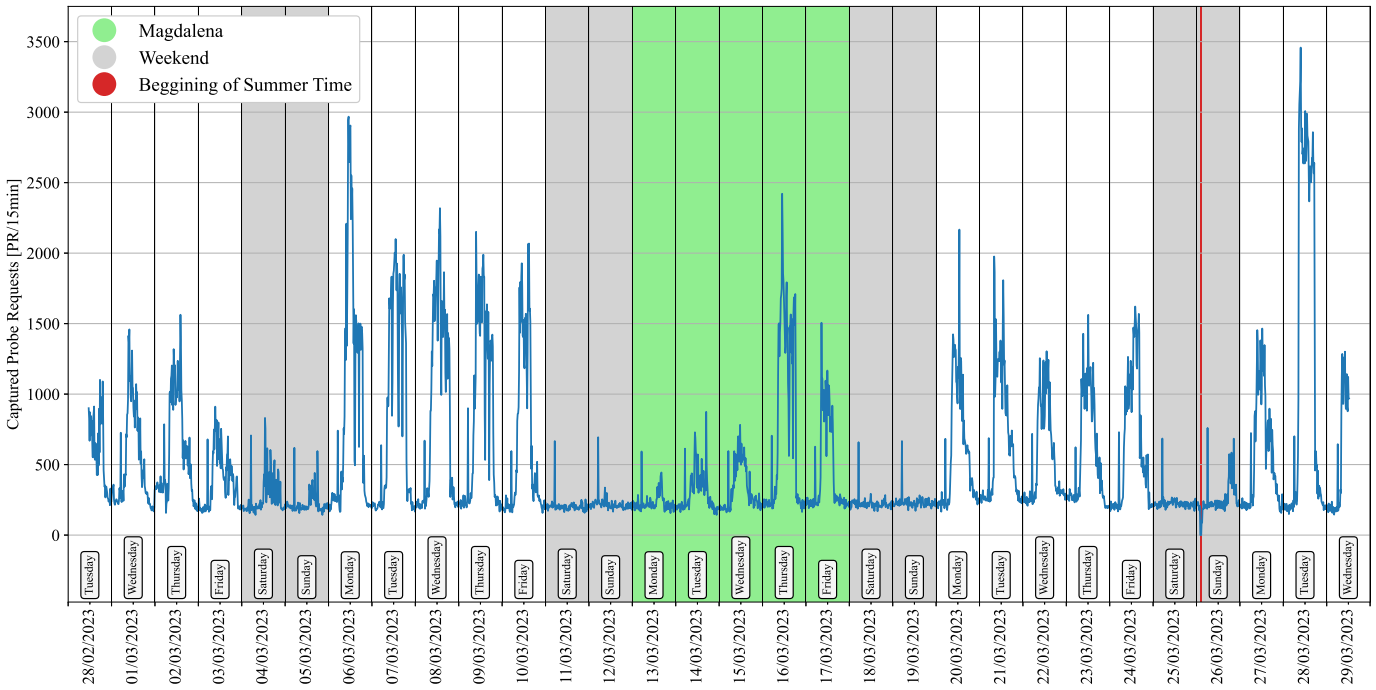

Fig. 4. Density of captured Probe Requests over the course of capture (amount of probe requests grouped in 15-minute clusters).

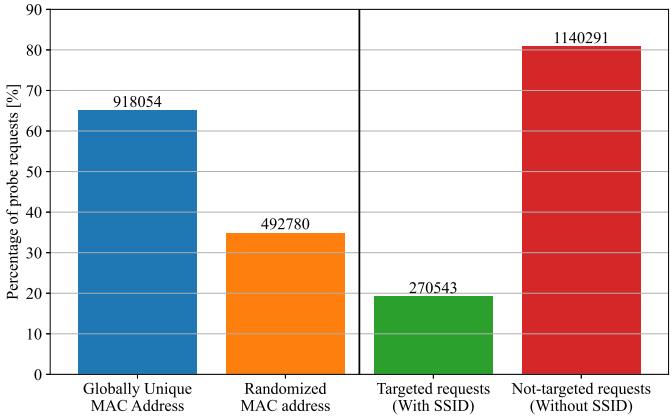

Fig. 5. Split between randomized and globally unique MAC addresses in the captured probe requests and split between targeted and not targeted probe requests.

TABLE I
PROBE REQUEST FIELDS USED TO CREATE DEVICE FINGERPRINT AND FREQUENCY OF OCCURRENCE IN DATA COLLECTED IN OUR LAB

Information Element	Included in Probes	[%]
Supported rates	1 410 832	100.00
Extended Supported rates	1 405 288	99.61
HT Capabilities	1 093 575	77.51
HE Capabilities	394 347	27.95
VHT Capabilities	262 365	18.60
Extended Capabilities	1 182 014	83.78
Vendor Specific elements	315 815	22.39
1 Vendor Specific element	135 973	9.64
2 Vendor Specific elements	110 071	7.80
3 Vendor Specific elements	9607	0.68
4 Vendor Specific elements	257	0.02
5+ Vendor Specific elements	146	0.01
WPS - UUID-E	16 596	1.18
WEP Protected	2	0.00
Total Collected Probe Requests	1 410 834	

the 5 GHz band of Wi-Fi, only probe requests transmitted in the 2.4 GHz band were captured.

The distribution of RSSI values are presented using a histogram in Fig. 6. From the histogram, it is clearly visible that most frames were captured with RSSI in the range of $-100\,\mathrm{dBm}$ to $-40\,\mathrm{dBm}$. Surprisingly, a lot of probe requests had RSSI higher than $-30\,\mathrm{dBm}$: all of these probes were transmitted by a single device that had the wireless interface configured to use higher transmit power than usual. Another point of interest is the high amount of probes captured with the RSSI around $-90\,\mathrm{dBm}$, which are most likely devices from neighboring offices. We have also mapped the radio signal propagation throughout our office for another study. The Radio Map (RM) of the office mapped using another ESP32 microcontroller is in Fig. 7. The capture RM can be used as a rough distance filter, as it can be used for the selection of RSSI threshold.

IV. Usage Examples

There are several ways to use the dataset. In this section, we provide just a few examples of possible use cases.

A. Wi-Fi Signal Stability Evaluation

By capturing the RSSI values for each captured probe request, it is possible to analyze the signal from the received

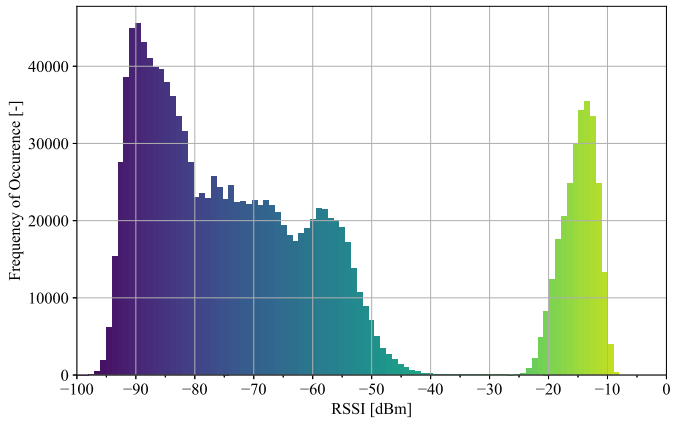

Fig. 6. Frequency of occurrence of RSSI in the captured probe requests.

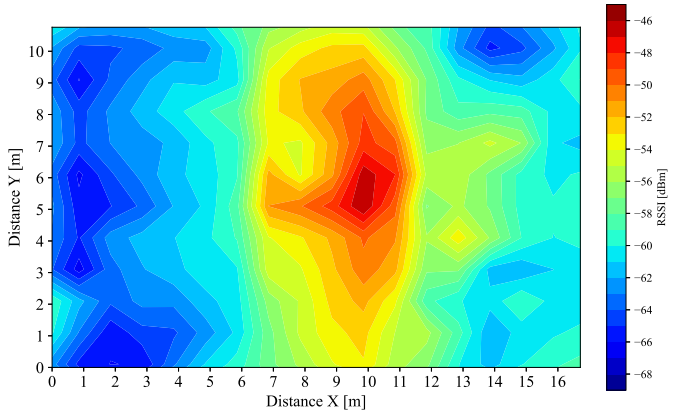

Fig. 7. Radio Map of the RSSI in different locations of the office mapped with ESP32 micro controllers.

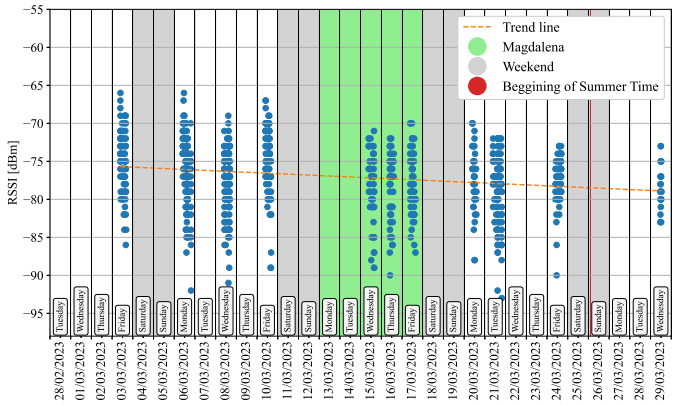

Fig. 8. Example of RSSI long-term stability evaluation for one MAC address from the dataset (*f4:7b:09:1f:5b:72*) including trend line.

signal strength point of view as well, assuming people spend most of the time at their desks. In the example we propose we measure the signal stability over time. As an example, we have randomly selected MAC address of a single device that appears on several days (MAC address of the selected device in the anonymized dataset is *f4:7b:09:1f:5b:72*) and visualized the RSSI of all probe requests over time. The changes over time are visible, and to highlight the change we also included a trend line in a Fig. 8 representing the decrease in RSSI throughout the capture time.

B. Presence Detection & Room Occupancy Estimation

The network traffic can be also used for presence detection. This can be easily seen in Fig. 4, from where we can see when the activity in the proximity of the sniffer increased. We can also see the drops on weekends and holidays. By also using the RM from Fig. 7, it is possible to roughly estimate the distance the UE is from the sniffer. The RM was created by collecting RSSI of probe requests throughout the office in $1\,\mathrm{m}$ grid. This information can be utilized for presence detection, or even room occupancy detection. On a smaller scale, we have estimated the occupancy of rooms on the university campus from captured probe requests [22].

Unfortunately, the capture of the ground truth of room occupancy in our office is nearly impossible. With up to 30 people present in the office, moving in and out at all times during work hours, the collection of ground truth was not possible with our current capabilities. However, the occupancy can be classified like: *empty, low, medium, high*, or on a scale from *low* to *high* which was used previously in occupation estimation work by Ciftler *et al.* [25].

C. User Privacy Exploitation

Since during the anonymization of the dataset we did not remove any data, but used pseudonyms using hashing for sensitive fields, the dataset allows for analysis of privacy leakage using an up to date data. The ToA of packets was not modified in any way either, which allows for studies into temporal analysis similar to the work done by Matte *et al.* [10].

It can also be used for the analysis of randomized MAC address recurrences or identification of users despite the randomization. The dataset may be used to reveal present vulnerabilities of the probe request mechanism and creation of a new type of analysis.

V. Ethics and Sensitive Information

The output packet capture files created by the ESP32-based probe request sniffer are in the same style as when the packets are captured by network analysis tools like Wireshark. The captured packets are the same as those transmitted by the nearby devices, including any sensitive information that they might contain. This information might contain globally unique MAC addresses, randomized MAC addresses, SSIDs of networks saved in the devices' preferred network list, manufacturer of the device or the wireless network interface, or in some cases even device names.

Following the data capture, we have run an anonymization script to hide any user information. For this purpose, the SHA512 hashing algorithm was used. By employing the hashing algorithm, we replace the original fields with hashes,

which preserves the ability to analyze the data in the same way as data that have not been anonymized. The main reason for anonymization through hashing is to allow the analysis of the probe requests to unveil the vulnerabilities in the management frames of the current implementation of the 802.11 protocol. To also preserve the information about randomized and globally unique MAC addresses, only the last 3 B are hashed. This preserves the least significant bits of the 1st Byte of the MAC address, as well as the Organization Unique Identifier [26]. Thanks to the use of hashing, creating a connection between the real identities of people and the captured probe requests is not possible. The anonymized version of the dataset is publicly available from Zenodo repository [27].

VI. Conclusions & Future Work

In this paper, we have explored existing datasets of Wi-Fi probe requests and presented a new one. As Wi-Fi enabled devices evolve every year, our new dataset provides updated data. Apart from the data itself we also provide firmware for the ESP32 microcontroller, so anyone can collect their own dataset. We present the data-gathering methodology and describe important information about the features of the dataset. Additionally, we have taken into account possible use cases and presented possible outcomes, like temporal RSSI stability, presence detection, room occupancy estimation, and privacy breaches.

In future works, we will dive deeper into the Wi-Fi in indoor positioning applications. We want especially to focus on the exploration of privacy issues and passive user tracking possibilities using the IEEE 802.11 communication protocol.

References

[1] A. Musa and J. Eriksson, "Tracking Unmodified Smartphones Using wi-fi Monitors," in *Proceedings of the 10th ACM conference on embedded network sensor systems*, 2012.

[2] M. V. Barbera, A. Epasto, A. Mei, V. C. Perta, and J. Stefa, "Signals from the Crowd: Uncovering Social Relationships Through Smartphone Probes," in *Proceedings of the 2013 conference on Internet measurement conference*, 2013.

[3] M. V. Barbera, A. Epasto, A. Mei, S. Kosta, V. C. Perta, and J. Stefa, *CRAWDAD sapienza/probe-requests*, 2013. [Online]. Available: https://dx.doi.org/10.15783/C76C7Z.

[4] M. Cunche, M.-A. Kaafar, and R. Boreli, "Linking wireless devices using information contained in Wi-Fi probe requests," *Pervasive and Mobile Computing*, vol. 11, 2014.

[5] 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*, 2013.

[6] L. Hutchinson, *iOS 8 to Stymie Trackers and Marketers with MAC Address Randomization*, Jun. 2014. [Online]. Available: https://arstechnica.com/gadgets/2014/06/ios8-to-stymie-trackers-and-marketers-with-mac-address-randomization/.

[7] J. Martin, E. Rye, and R. Beverly, "Decomposition of MAC Address Structure for Granular Device Inference," in *Proceedings of the 32nd Annual Conference on Computer Security Applications*, 2016.

[8] J. Freudiger, "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*, 2015.

[9] A. Di Luzio, A. Mei, and J. Stefa, "Mind your probes: De-anonymization of large crowds through smartphone WiFi probe requests," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, IEEE, 2016.

[10] C. Matte, M. Cunche, F. Rousseau, and M. Vanhoef, "Defeating MAC Address Randomization Through Timing Attacks," in *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, 2016.

[11] E. Fenske, D. Brown, J. Martin, T. Mayberry, P. Ryan, and E. Rye, "Three Years Later: A Study of MAC Address Randomization in Mobile Devices and When it Succeeds," *Proceedings on Privacy Enhancing Technologies*, vol. 2021, no. 3, 2021.

[12] J. Martin, T. Mayberry, C. Donahue, L. Foppe, L. Brown, C. Riggins, E. C. Rye, and D. Brown, "A Study of MAC Address Randomization in Mobile Devices and When it Fails," *arXiv preprint arXiv:1703.02874*, 2017.

[13] *IEEE 802.11aq-2018 - 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 Amendment 5: Preassociation Discovery*. [Online]. Available: https://standards.ieee.org/standard/802_11aq-2018.html.

[14] WiGLE, *Wigle: Wireless network mapping*, 2023. [Online]. Available: https://wigle.net/.

[15] Google, *Google Developers: Geolocation API*, Google, 2022. [Online]. Available: https://developers.google.com/maps/documentation/geolocation/.

[16] Google, *Google Support: Control Access Point Inclusion in Google's Location Services*, 2022. [Online]. Available: https://support.google.com/maps/answer/1725632?hl=en.

[17] P. Robyns, B. Bonné, P. Quax, and W. Lamotte, *CRAWDAD hasselt/glimps2015*, 2015. [Online]. Available: https://dx.doi.org/10.15783/c7sd19.

[18] B. Abdulrahem, *CRAWDAD nile/probe-requests*, 2021. [Online]. Available: https://dx.doi.org/10.15783/s9mm-z673.

[19] T. Bravenec, J. Torres-Sospedra, M. Gould, and T. Frýza, *Supplementary Materials for "What Your Wearable Devices Revealed About You and Possibilities of Non-Cooperative 802.11 Presence Detection During Your Last IPIN Visit"*, version 1.0, Zenodo, 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6798302.

[20] T. Bravenec, J. Torres-Sospedra, M. Gould, and T. Fryza, "What Your Wearable Devices Revealed About You and Possibilities of Non-Cooperative 802.11 Presence Detection During Your Last IPIN Visit," in *2022 IEEE 12th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2022.

[21] T. Bravenec, J. Torres-Sospedra, M. Gould, and T. Fryza, "Exploration of User Privacy in 802.11 Probe Requests with MAC Address Randomization Using Temporal Pattern Analysis," *arXiv e-prints*, arXiv–2206, 2022.

[22] T. Fryza, T. Bravenec, and Z. Kohl, "Security and Reliability of Room Occupancy Detection Using Probe Requests in Smart Buildings," in *2023 33st International Conference Radioelektronika (RADIOELEKTRONIKA)*, IEEE, 2023.

[23] T. Bravenec, *ESP32 Probe Sniffer*, 2022. [Online]. Available: https://gitlab.com/tbravenec/esp32-probe-sniffer.

[24] *Radiotap*. [Online]. Available: http://www.radiotap.org/.

[25] B. S. Ciftler, S. Dikmese, I. Güvenç, K. Akkaya, and A. Kadri, "Occupancy Counting with Burst and Intermittent Signals in Smart Buildings," *IEEE Internet of Things Journal*, vol. 5, no. 2, 2017.

[26] IEEE, *IEEE Standards - OUI/MA-L*. [Online]. Available: https://standards-oui.ieee.org/.

[27] T. Bravenec, J. Torres-Sospedra, M. Gould, and T. Frýza, *Supplementary Materials for "UJI Probes: Dataset of Wi-Fi Probe Requests"*, 2022. [Online]. Available: https://doi.org/10.5281/zenodo.7801798.