DEMO: Passive Identification of WiFi Devices in Real-Time

Niruth Bogahawatta
nbog5307@sydney.edu.au
The University of Sydney

Gerry How
yhow4819@sydney.edu.au
The University of Sydney

Yasiru Karunanayake
yasiru.karunanayaka@sydney.edu.au
The University of Sydney

Suranga Seneviratne
suranga.seneviratne@sydney.edu.au
The University of Sydney

Kanchana Thilakarathna
kanchana.thilakarathna@sydney.edu.au
The University of Sydney

Salil Kahere
salil.kanhere@unsw.edu.au
University of New South Wales

Rahat Masood
rahat.masood@unsw.edu.au
University of New South Wales

Aruna Seneviratne
a.seneviratne@unsw.edu.au
University of New South Wales

ABSTRACT

WiFi has emerged as the standard method for local connectivity across various devices, including smart assistants, IoT devices, smart TVs, and AR/VR devices. Identifying WiFi devices passively in neighbourhoods has implications for law enforcement, urban planning, and socio-economic analysis. In this demo paper, we introduce a novel approach to constructing WiFi device-type signatures using Information Element (IE) attributes from wildcard WiFi probe requests. Our method accurately identifies device types even when dealing with randomized MAC addresses, requires minimal training as opposed to machine learning methods, and operate in real time.

CCS CONCEPTS

• **Networks** → **WiFi**; **Wireless local area networks**; **Device Identification**.

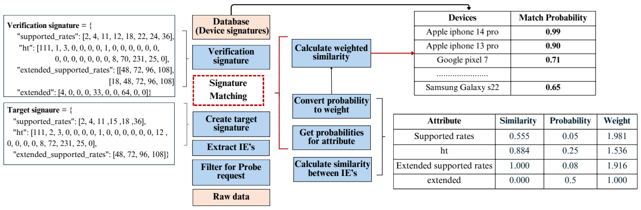
ACM Reference Format:
Niruth Bogahawatta, Gerry How, Yasiru Karunanayake, Suranga Seneviratne, Kanchana Thilakarathna, Salil Kahere, Rahat Masood, and Aruna Seneviratne. 2024. DEMO: Passive Identification of WiFi Devices in Real-Time. In *ACM SIGCOMM 2024 Poster and Demo (SIGCOMM Poster Demo 24), August 4–8, 2024, Sydney, NSW, Australia.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3672202.3673716

1 INTRODUCTION

WiFi adoption has transformed daily life, providing compatibility with various devices and fueling the rise of smart homes and the Internet of Things (IoT). This widespread integration creates opportunities for intelligence gathering and improves public safety, law enforcement, and urban design applications. In scenarios where law enforcement monitors buildings and tracks devices, passive identification emerges as a superior method to active approaches. While cellular triangulation and direct network connection demand complex processes and cooperation from carriers or network administrators, passive identification operates without active communication, mitigating risks, and enabling scalable bulk device identification. Early passive WiFi device identification utilized WiFi sensing and initially relied on MAC addresses for identification, but their effectiveness declined due to MAC address randomization [4]. Subsequent methods focused on analyzing WiFi devices' probe requests, employing techniques like profiling based on probing behaviour and device fingerprinting using Inter Arrival Time (IAT) [2, 3]. Recent advancements integrated machine learning (ML) methods such as neural networks and a mix of supervised and unsupervised learning [5, 6]. However, these methods face challenges, including the labour-intensive process of collecting training data.

We introduce a passive WiFi device-type identification method that uses Information Elements (IE) from probe request frame bodies and ensures resistance to MAC address randomization. In contrast to current ML methods, our classification method calculates a similarity score through weighted average probability by constructing device signatures with minimal WiFi probe request observations. Unlike complex temporal pattern-dependent ML approaches, our method is simple, easily implementable, and scalable. It achieves a 99% F1 score compared to a neural network-based method's 89%, our method also achieves a 92% F1 score with just one training data point per device type. Furthermore, we provide a publicly available WiFi device-type signature database containing 50 widely-used WiFi devices, including smartphones, IoT devices, and laptops.

2 SIGNATURE CREATION AND MATCHING



```
Verification signature = {
    "supported_rates": [2, 4, 11, 12, 18, 22, 24, 36],
    "ht": [111, 1, 3, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
            0, 0, 0, 0, 0, 0, 0, 8, 70, 231, 25, 0],
    "extended_supported_rates": [[48, 72, 96, 108],
                                  [18, 48, 72, 96, 108]
    "extended": [4, 0, 0, 0, 33, 0, 0, 64, 0, 0]}
```

```
Target signaure = {
  "supported_rates": [2, 4, 11 ,15 ,18 ,36],
  "ht": [111, 2, 3, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 12 ,
0, 0, 0, 0, 8, 72, 231, 25, 0],
  "extended_supported_rates": [48, 72, 96, 108]}
```

Devices	Match Probability
Apple iphone 14 pro	**0.99**
Apple iphone 13 pro	**0.90**
Google pixel 7	**0.71**
....	
Samsung Galaxy s22	**0.65**

Attribute	Similarity	Probability	Weight
Supported rates	0.555	0.05	1.981
ht	0.884	0.25	1.536
Extended supported rates	1.000	0.08	1.916
extended	0.000	0.5	1.000

Figure 1: Overview of signature creation and matching

Probe requests are frames used by devices to scan for nearby WiFi networks. These requests contain IEs that provide details about the device's capabilities, including supported rates and security protocols. For instance, the absence of specific IEs within a probe request can signal a device's limitations or its lack of support for

SIGCOMM Poster Demo 24, August 4–8, 2024, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0717-9/24/08.
https://doi.org/10.1145/3672202.3673716

certain WiFi standards or encryption methods. By capturing and analyzing probe requests, we can distinguish devices within a WiFi network based on the distinctive and reliable characteristics of each device signature. To create device-type signatures, we extract all IEs from captured probe requests. These extracted IEs are then organized into dictionaries, wherein attribute names are linked to their respective values, as illustrated in Fig 1.

To determine the similarity between the target signature and the verification signature, we evaluate attribute similarity by examining each attribute individually. For attributes that consist of unordered values, such as "supported rates" and "High Efficiency (HE) ", we use Jaccard similarity. For attributes having a specific order, such as the values of the "High throughput (HT)", we use the Hamming distance. Moreover, certain attributes hold greater significance within the signature. Therefore, we have implemented a weighting system that allows us to compute a weighted average of similarities. We derive these weights through a data-driven approach, i.e., by analyzing the probability of occurrence of attribute value from a large signature sample. The probabilities are calculated as: $p = p(s_1 a_n) \times p(s_2 a_n)$. Here s_1 and s_2 represent two signatures, and a_n denotes the attribute value for the n^{th} attribute. $p(s_1 a_n)$ signifies the probability of observing the n^{th} attribute value for s_1.

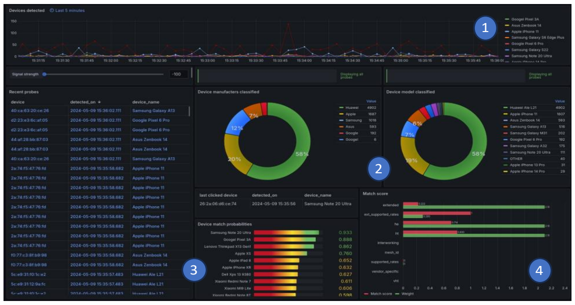
We convert probability values to weights using t_1, defined as $T_1(p) = \frac{1}{\alpha_1 - \beta_1(1-p)} - \frac{1}{\alpha_1}$, with constants $\alpha_1 = 1 - \frac{2}{\lambda}$ and $\beta_1 = \frac{\lambda^2 - 4\lambda + 4}{\lambda^2 - \lambda}$. This transformation prioritizes rare events with higher weights and common events with lower weights. Post-transformation, we compute the weighted average $S = \frac{\sum_{i=0}^{N} s_i \gamma_i l_i}{\sum_{i=0}^{N} \gamma_i l_i}$, where s_i is the attribute similarity, γ_i is its weight, and l_i denotes its importance. S ranges from 0 to 1, with values closer to 1 indicating a stronger match. This computation compares the target signature with every verification signature in our database. If the device isn't in our database, we identify the closest match based on probability, suggesting matches for devices of similar types.

3 DEMONSTRATION

Our demonstration setup includes three ALFA AWUS036NHA[1] external WiFi antennas, which are connected to a laptop. To capture data, we developed Python scripts leveraging `t-shark` [1], which allowed us to gather data across all 14 WiFi channels and save them as `.pcap` files. Data collection occurs in intervals of 5 seconds.

We developed a real-time dashboard using Grafana[2] that deploys our signature-based classification method and provides an enhanced visualization of the incoming probes as a live feed.

As illustrated in Fig 2, we can monitor the influx of probes alongside the categorization of probes into various devices - ①. An interactive panel enables us to delve deeper into the breakdown of a probe request across all devices stored in the database - ②. Additionally, users can interact to obtain a granular breakdown of the probabilities of how closely a probe resembles a specific device - ③. Summaries of the calculation of the probe frames are also accessible on this page - ④. Furthermore, we provide a dedicated page for visualizing data when monitoring probe requests from multiple sites, as shown in Fig 3. This page features a heatmap illustrating the density of probe activities, accompanied by site-specific breakdowns detailing the types of devices detected. Users can customize their heatmap view across two subsequent pages by using filtering options such as manufacturers or specific device models.



Recent probes

device	detected_on	device_name
40:ca:63:20:ce:26	2024-05-09 15:36:02.111	Samsung Galaxy A13
d2:23:e3:6c:af:05	2024-05-09 15:36:02.111	Google Pixel 6 Pro
d2:23:e3:6c:af:05	2024-05-09 15:36:02.111	Google Pixel 6 Pro
44:af:28:bb:87:03	2024-05-09 15:36:02.111	Asus Zenbook 14
44:af:28:bb:87:03	2024-05-09 15:36:02.111	Asus Zenbook 14
40:ca:63:20:ce:26	2024-05-09 15:36:02.111	Samsung Galaxy A13
2a:74:f5:47:76:fd	2024-05-09 15:35:58.682	Apple iPhone 11
2a:74:f5:47:76:fd	2024-05-09 15:35:58.682	Apple iPhone 11
2a:74:f5:47:76:fd	2024-05-09 15:35:58.682	Apple iPhone 11
2a:74:f5:47:76:fd	2024-05-09 15:35:58.682	Apple iPhone 11
2a:74:f5:47:76:fd	2024-05-09 15:35:58.682	Apple iPhone 11
2a:74:f5:47:76:fd	2024-05-09 15:35:58.682	Apple iPhone 11
2a:74:f5:47:76:fd	2024-05-09 15:35:58.682	Apple iPhone 11
2a:74:f5:47:76:fd	2024-05-09 15:35:58.682	Apple iPhone 11
2a:74:f5:47:76:fd	2024-05-09 15:35:58.682	Apple iPhone 11
f0:77:c3:8f:b9:98	2024-05-09 15:35:58.682	Asus Zenbook 14
f0:77:c3:8f:b9:98	2024-05-09 15:35:58.682	Asus Zenbook 14
5c:e9:31:f0:1c:e2	2024-05-09 15:35:57.483	Huawei Ale L21
5c:e9:31:12:9a:fc	2024-05-09 15:35:57.483	Huawei Ale L21
5c:e9:31:f0:1c:e2	2024-05-09 15:35:57.483	Huawei Ale L21

	Value
Huawei	4902
Apple	1687
Samsung	1018
Asus	593
Google	192
Googel	6

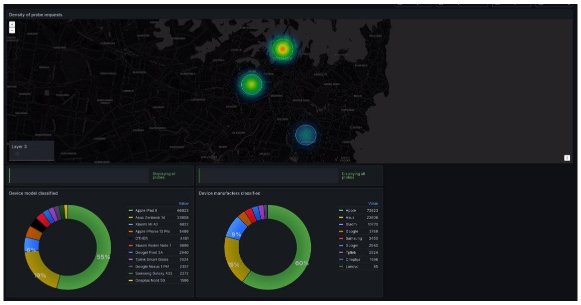
	Value
Huawei Ale L21	4902
Apple iPhone 11	1607
Asus Zenbook 14	593
Samsung Galaxy A13	516
Samsung Galaxy M31	202
Google Pixel 6 Pro	192
Samsung Galaxy A32	175
Samsung Note 20 Ultra	111
OTHER	40
Apple iPhone 13 Pro	31
Apple iPhone 14 Pro	29

last clicked device	detected_on	device_name
26:2a:06:d6:ce:74	2024-05-09 15:35:56	Samsung Note 20 Ultra

Device match probabilities

Samsung Note 20 Ultra	0.933
Googel Pixel 3A	0.888
Lenovo Thinkpad X13 Gen1	0.862
Apple XS	0.760
Apple iPad 8	0.652
Apple iPhone XR	0.632
Dell Xps 13 9380	0.627
Xiaomi Redmi Note 7	0.611
Xiaomi MI9 Lite	0.606
Xiaomi Redmi Note 8T	0.598

Figure 2: Dashboard: Device type identification



	Value
Apple iPad 8	66823
Asus Zenbook 14	23608
Xiaomi MI A2	6822
Apple iPhone 13 Pro	5486
OTHER	4461
Xiaomi Redmi Note 7	3696
Googel Pixel 3A	2940
Tplink Smart Globe	2524
Google Nexus 1 Ph1	2357
Samsung Galaxy A32	2272
Oneplus Nord 5G	1596

	Value
Apple	73823
Asus	23608
Xiaomi	10770
Google	3789
Samsung	3450
Googel	2940
Tplink	2524
Oneplus	1596
Lenovo	85

Figure 3: Dashboard: Device site heatmap

We showcase our method's functionality with two scenarios:

Scenario 1: Known Device within Our Database. Given a known device for which we possess a verification signature, we capture the data and display its classification probability on our dashboard. In this scenario, we anticipate a near-perfect or significantly high match probability. To evaluate the ground truth, we use the MAC address or a random MAC address associated with the device.

Scenario 2: Device Not Present in Our Database. When presented with a device that does not have a verification signature in our database, we capture the trace data and exhibit its classification probability on our dashboard. We expect a mismatch in this scenario as the device is not recognized. Subsequently, we incorporate the device's signature into our database and re-initiate the classification process. Upon subsequent capture of the device's trace data, we anticipate a near-perfect or significantly high match probability.

4 CONCLUSION

This paper introduces a novel method to accurately and passively identify WiFi device types within a defined area using a signature-based approach. We analyze a dataset covering 50 device types, demonstrating high accuracy in passive identification. Our method outperforms neural network-based approaches with comparable data resources. Additionally, we present a real-time device classification dashboard to enhance practical usability.

5 ACKNOWLEDGEMENT

This research is funded by the DSP Research Agreement [Agreement No: 11384], between UNSW and the Department of Defence.

[1] AWUS036NHA is a high-gain directional antenna manufactured by ALFA Network Inc. Compatible with IEEE 802.11 b/g/n wireless standards.
[2] https://grafana.com/

REFERENCES

[1] [n. d.]. TShark - The Network Protocol Analyzer. https://www.wireshark.org/docs/man-pages/tshark.html.

[2] Sandhya Aneja, Nagender Aneja, and Md Shohidul Islam. 2018. IoT device fingerprint using deep learning. In *2018 IEEE international conference on internet of things and intelligence system (IOTAIS)*. IEEE, 174–179.

[3] Prashant Baral, Ning Yang, and Ning Weng. 2023. IoT Device Identification Using Device Fingerprint and Deep Learning. (2023).

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

[5] Xiaolin Gu, Wenjia Wu, Xiaodan Gu, Zhen Ling, Ming Yang, and Aibo Song. 2020. Probe request based device identification attack and defense. *Sensors* 20, 16 (2020), 4620.

[6] Jiajie Tan and S.-H. Gary Chan. 2021. Efficient Association of Wi-Fi Probe Requests under MAC Address Randomization. In *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*. 1–10. https://doi.org/10.1109/INFOCOM42981.2021.9488769