Public Wireless Packets Anonymously Hurt You

Abhishek Kumar Mishra*†, Aline Carneiro Viana†, Nadjib Achir†‡, Catuscia Palamidessi†,
* *Institut Polytechnique de Paris*, France
† *Inria*, France
{abhishek.mishra, aline.viana, nadjib.achir, catuscia.palamidessi}@inria.fr
‡ *Université Sorbonne Paris Nord*, France
nadjib.achir@univ-paris13.fr

2021 IEEE 46th Conference on Local Computer Networks (LCN) | 978-1-6654-1886-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/LCN52139.2021.9524956

***Abstract*—With growing privacy concerns over the last decade, two of the most notable wireless technologies – i.e., BLE and WiFi – are being more and more investigated in terms of privacy vulnerabilities. In this paper, we explore this problem, prospect the related consequences, and alert the need for privacy-preserving public packets. We identify key flaws in the current design of public packets like beacons and probe requests. We discuss them as the cause of privacy issues that require the community's attention. We address the flaws in detail and propose solutions that facilitate the devices to protect user privacy. We also give recommendations based on the findings to the standard.**

***Index Terms*—Privacy, wireless, BLE, WiFi, public packets**

I. Introduction

The growth of WiFi and BLE networked devices has brought increasing concerns for user privacy. These concerns vary from the protection of receiver-location to anonymity and traceability in general. Sniffing public wireless traffic such as beacon messages is very straightforward. After early question marks on user privacy due to device linkability by advertised identifiers, steps were taken by the standard, notably MAC address randomization, to address these concerns. But recent works in the literature have raised doubts on the effectiveness of current measures.

Devices that perform *MAC address randomization* can hide the device's identity to some extent. This feature has been the backbone of user-privacy in wireless networks, especially BLE and WiFi. Mac address randomization in mobile devices has been thoroughly studied. Martin et. al claim to effectively defeat randomization for around 96% of android devices [1]. An artificial intelligence-based approach shows that 91% of the WiFi devices could be tracked [2]. Bluetooth Classic (BT) does not randomize the addresses and has already been shown to be de-anonymized [3]. Even MAC address randomization in BLE has been claimed to be defeated specific to apple devices [4] and for generalized devices [5]. 100% device association for a small set of devices on sniffing public-packets in a controlled environment (inside Faraday cage) is claimed in [5].

MAC address randomization is not enough to safeguard user privacy. There have been attacks of fingerprinting devices just using the timestamps of advertised public packets [6]. Further vulnerabilities have been discovered in the information fields of the packets that could reveal private information of the user like language detection on the broadcast WiFi SSIDs and even the sociological aspects of the people like nationality, age, and socioeconomic status [7]. Smartphone Screen ON/OFF State can be classified using WiFi Probe patterns [8]. Similarly, using BLE beacons, [9] show that user profiling, beacon hijacking, presence inference, and even user harassment is possible. Combining both BLE and WiFi's public packets of the same user could lead to more devastating breaches in privacy. Both communities need to jointly come up with improved regulations and recommendations in the standard to ensure user privacy.

Most of the existing related works suggest flaws in randomization [1] [2] [5] [6] or suggest inferring insights [10] [4] [9] from the transmitted public data in the packets. Encryption-based defense of WiFi 5 packets is suggested by [11], but as we discuss later in this paper, such solutions are not feasible due to resource constraints. *To the best of our knowledge, none of current work give a global view on the privacy issues in the design of public wireless packets itself. Current works do not look into counter-measures of timing-based attacks, which are more generic and effective than we would see in the upcoming sections.*

The novelty of our work lies in the investigation of the root causes of growing privacy concerns in public wireless packets. The paper's key contributions are 1) Classification of current attacks in the literature based on methodology 2) Revealing key design flaws in current WiFi and BLE public packets 3) Solutions and recommendations to rectify the flaws we detect in the design.

The rest of the paper is organized as follows. Section II presents the state-of-the-art on privacy concerns that arise from the current design of public packets in BLE and WiFi. Next, we identify problems in the current standard for public packets in Section III and propose respective solutions/recommendations. Finally, in Section IV, we conclude our work with final remarks and look into the future directions.

II. Privacy concerns from public packets

Threats to user privacy from public wireless packets cover a wide range of private aspects. Next, we classify the existing attacks and concerns on the basis of methodology into three broad umbrellas. Finally, we end this section by identifying the key design flaws in currently advertised public packets.

A. Attack methodologies

1) Timing-based attacks: These kinds of attacks rely upon the temporal information that could be extracted from the observed public packet sequence from a device [6]. The adversary's aim here is to extract metrics that are characteristic of a device and remain consistent over a period of time irrespective of address randomization. Examples of such metrics are Inter-frame space (IFS), inter-burst duration, etc.

2) Frame-field attacks: Here, an attacker learns the information fields in the frames that are generally sent in the clear to classify and subsequently fingerprint devices [11]. There are flags, client capability information, Manufacturer names, frame-type, etc., which possess the potential of being part of a fingerprint.

3) Inference attacks: In the case of inference attacks, an adversary observes the activity of a user along with packets it sends over a period of time to infer private information [7] [8]. Preferred network list(PNL) in the WiFi probe requests, variation in two attempts of probes, etc., are some of the information that helps the attacker to learn regarding the user under threat.

B. Key design flaws

We briefly list the following key flaws in the design and implementation of public packets.

1) *Ineffective address randomization* - MAC address randomization, if implemented effectively, can prevent user tracking to some extent. The current implementation of randomization is not adaptive to the user surroundings and is predictive.
2) *Uniform timing parameters* - Parameters specific to the timing of advertised public packets are uniform distributed across the device population. Most of them are manufacturer-specific and even vary within a brand. This makes users fall prey to fingerprinting and profiling.
3) *Inadequate privacy measures in clear-text packets* - Majority of current public packet fields are sent in plaintext. They lack necessary privacy measures to prevent inference, even though many of these fields contain potentially private-intruding information.

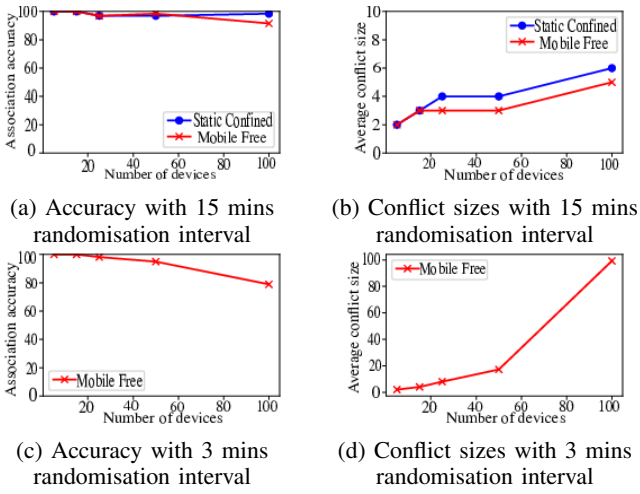
In the following section, we go for each of the above flaws in detail.

III. Proposed Solution/Recommendations

Among public packets, in this paper, we focus on BLE beacons and WiFi probe requests as they are the most privacy-revealing. We next address the design flaws concerning current standard provisions that we identify in the previous section and propose solutions to them, respectively.

A. Choice of randomization interval

Timing-based attacks on BLE MAC address randomization take benefits of the current interval after which a device changes the public identifier. The more frequently we perform the randomization, the more probable it is for a higher number of devices in the population to change their MAC addresses around the same time. We denote this number by *Conflict size*. Higher is this value on average, the more difficult it is to associate MAC addresses from the same device for the adversary.



(a) Accuracy with 15 mins randomisation interval

(b) Conflict sizes with 15 mins randomisation interval

(c) Accuracy with 3 mins randomisation interval

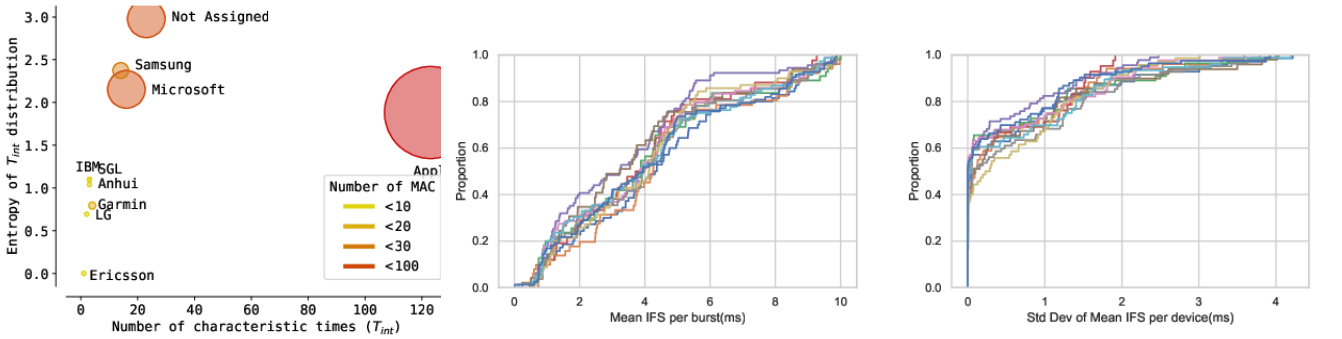
(d) Conflict sizes with 3 mins randomisation interval

Fig. 1: Performance of MAC association strategy [5] with varying BLE randomization intervals

WiFi, Linux, iOS, and Windows have quite different MAC randomization schemes [6]. Linux lets the driver or firmware generate per-burst random MAC addresses. In iOS, randomization is limited to probing and only happens when the device is unassociated and in sleep mode. Windows 10 changes the MAC address when the device connects or disconnects from a network and when it restarts. As timing attacks in WiFi also defeat randomization by up to 75% [6], the standard needs to have a consensus on making randomization mandatory for the manufacturers while also specifying lower duration, preferably every few bursts.

In BLE, the standard currently recommends keeping the random MAC address for at least 15 minutes [12]. We look at the performance of the only generic BLE timing attack in the literature [5] with respect to the size of the randomization interval. We use *SimBle* [13], a framework to generate large-scale real-world BLE traces for the evaluation. This framework provides the *ground truth* information regarding the random MAC addresses generated from the device, which is needed to deduce the accuracy. Also, devices could mimic the behavior of various device classes like smartphones, smartwatches, etc., in terms of generating beacons. We reduce the *randomization interval* of the device population to 3 minutes and evaluate the performance against the standard 15 minutes duration. We pick 3 minutes as the optimal lower value, as a much smaller interval will cause longer connection times in the real devices. We use two mobility profiles, *Static-Confined* where BLE devices are stationary within the sniffing range of a sniffer, and *Mobile-Free* where devices have the freedom to leave and enter the sniffing range. We vary the number of devices inside *SimBle* up to 100.

We observe in Figure 1 that indeed accuracy decreases to



(a) Discriminatory level of the advertising timing for each device Manufacturer in BLE

(b) CDF of mean IFS per burst for WiFi probe requests

(c) CDF of standard deviation in Mean IFS per burst per device for WiFi probe requests

Fig. 2: Behavior of *Weak identifiers* (Better seen in color)

a minimum of around 91% to 78% with *conflict size* growing to from 6 to 97 when decreasing the randomization interval. Based on this observation, we recommended lowering the BLE *randomization interval* while caring for slightly increased connection times as a consequence. Thus, we can optimize the current IRK (Identity resolving key) [12] exchange, for instance, in BLE, to allow devices to change address frequently without compromising performance.

B. Discriminating power of weak identifiers

Weak identifiers are the features deduced from the advertised frames which are capable of discriminating a subset of device-population. The majority of the works rely on *timing-based signatures* to differentiate two devices that change their MAC at the same time, as it works as a fingerprint per device [6] [5]. Inter-frame space(IFS) in WiFi and inter-beacon interval in BLE are the most promising weak identifiers that comprise timing-based signatures.

Inter-beacon interval in BLE consists of a constant part plus a pseudo-random value in the range $[0, 10]\ ms$. [14, p. 2751] The device regenerates the random value every burst, but the constant part seldom changes and could be estimated [5]. We call in this paper this weak identifier as the *characteristic time* or T_{int}. We collect a highly dense dataset that contains more than 2500 MAC addresses in less than 10 seconds to evaluate the vulnerability of the current BLE against this identifier. A powerful weak identifier should be as uniform over a set of values as large as possible. The Shannon entropy is a direct measure of the uniformity of the distribution.

Figure 2a displays the entropy of the characteristic time distributions, over the space of devices, per brand, as a function of the number of characteristic times used. For each brand, a bigger circle signifies multiple devices of the same brand in conflict more probable. The more characteristic times a brand possesses, the more devices could be simultaneously differentiated. Contrary to the intuitive belief that more identifier means more individual privacy, introducing many characteristic times increases identification chances. With the result in Figure 1b we see that, in most cases, the characteristic time is sufficient as a weak identifier, as conflict clusters rarely grow past the 10 devices with current provisions. For such small clusters, [5] already defeat BLE MAC randomization by up to 100% accuracy. Hence, the BLE standard must force the manufacturers to have similar characteristic times to reduce the discriminatory power of this weak identifier.

For WiFi, we also investigate the IFS in a probe-request burst using the Bologna probe-request university dataset that captures probe requests from 3917 MAC addresses. Signatures deduced from the IFS could discriminate up to 75% of the randomized WiFi devices. We look into the root cause of this problem by looking at the mean IFS per burst and the standard deviation of the mean IFS per burst per device with respect to the proportion of devices observed. We see in Figure 2b that mean IFS is almost uniformly distributed in the range $[0, 10]\ ms$ for various traces in the dataset. But at the same time, the standard deviation of mean IFS per burst is less than 1 ms for around 80 percent of devices. This makes mean IFS per burst is susceptible to being used as a fingerprinting solution in WiFi. Again, we recommend the WiFi standard to force manufacturing brands to use similar probe-request bursts in IFS.

C. Confidential packet information fields

To stop the frame field and inference attacks, encryption is the most intuitive solution for limiting the device fingerprinting using public packets. We need the following key properties for encrypting public packets: 1) *Universality*- The solution should be compatible across various wireless standards. 2) *Uncorrelation*- We have to ensure that two different frames from the device across time are not be linked to the same source. 3) *Efficiency*- To save time and energy, a minimum number of exchanges should happen in the control information transfer phase of the proposed protocol. Also, each exchange must happen in realistic bounds to ensure the usability of the control packets. 4) *Conformity*- The structure of the frames must still conform to that of the standard format to hide the presence of security measures from the attacker.

We require an efficient key exchange protocol to establish a

symmetric key for further exchanges, while we need different keys for successive packets from a source to ensure the property of un-correlation. We analyze next the need and the feasibility of encryption to ensure confidentiality in BLE and WiFi, respectively.

BLE beacons contain very weak identifiers like manufacturer names, event types, address types, and flags. We test these identifier's potential to discriminate among the randomized MAC addresses. We observe that the manufacturer information and the advertisement type can resolve less than 5 percent of MAC address conflicts seen in the dataset described in Section III-B. Therefore, we conclude that it is not necessary to pseudonymize any packet fields in the BLE Beacons.

On the contrary, in WiFi, we have a significant number of attacks on user privacy based on inferring from public packets like probe requests. McKinion et. al [15] do a widespread evaluation of existing frameworks for security pit-holes in probe requests. Probe-request-based device identification on IEEE 802.11ac is reduced considerably after the application of stream cipher-based encryption [11]. They test the solution with Two Dell OptiPlex 3600 mini Workstation are used as a client device and AP. We notice that Diffie-hellman key exchange takes $2.4-2.6s$, while transmission of the encrypted packet needs $0.4-0.6s$ [11].

As we already see in Figure 2b, the mean IFS per burst of probe requests in a real dataset is in the order of few milliseconds. The burst duration is around $10ms$ practically, and we see in Section II-A that most of the devices change their mac addresses after every few bursts. Active scans in WiFi are meant for fast re-connection to known networks. The overhead of $500ms$ in sending encrypted probes, even in basic exchanges of a stream cipher, is not realistic. The protocols will get heavier if we introduce more security guarantees to stop replay attacks, for instance. We should also consider the packet losses during the key exchange and the timeouts that it would induce. Moreover, we must support the broadcast probe requests from the client too. Due to these resource constraints, we argue the un-feasibility of classical encryption as a contender for providing confidentiality in WiFi probes.

We instead propose the following solutions which could be explored in the future that ensure the practicality and usability of WiFi probe requests: 1) We could add controlled noise to the information fields in the probe requests helps in reducing the effectiveness of fingerprinting attacks. 2) Entries in Preferred network lists (PNLs) advertised by the probe requests can be replaced with pseudo-identifiers, which are agreed upon by each client-AP pair prior.

IV. Final remarks and future steps

Public packets are the backbone of wireless networks as they are essential for ensuring network functionalities and enhancing the user experience. For instance, Beacon and probe requests help the node discover the network connection with very low delays. However, public packets are legal to sniff upon in most geographical areas, which raises user-privacy concerns. Works in literature have already utilized the public packets for device fingerprinting and user profiling. This paper demonstrates that the current design of public packets in WiFi and BLE has significant flaws. We identify these key shortcomings and address them in detail. We observe that both the BLE and the WiFi standard need to act upon the identified flaws in the design. Along with giving solutions to the issues, we also provide recommendations that the standard could enforce upon the manufacturers. Regarding future steps, we plan to test upon some of the suggestions for ensuring confidentiality that we provide in this paper for real-world hardware.

References

[1] 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," *Proceedings on Privacy Enhancing Technologies*, vol. 2017, no. 4, pp. 365–383, 2017.

[2] M. Uras, R. Cossu, E. Ferrara, O. Bagdasar, A. Liotta, and L. Atzori, "Wifi probes sniffing: an artificial intelligence based approach for mac addresses de-randomization," in *2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pp. 1–6, IEEE, 2020.

[3] M. Cominelli, F. Gringoli, P. Patras, M. Lind, and G. Noubir, "Even black cats cannot stay hidden in the dark: Full-band de-anonymization of bluetooth classic devices," in *2020 IEEE Symposium on Security and Privacy (SP)*, pp. 534–548, 2020.

[4] J. Martin, D. Alpuche, K. Bodeman, L. Brown, E. Fenske, L. Foppe, T. Mayberry, E. Rye, B. Sipes, and S. Teplov, "Handoff all your privacy–a review of apple's bluetooth low energy continuity protocol," *PoPETs*, vol. 2019, no. 4, pp. 34–53, 2019.

[5] L. Jouans, A. C. Viana, N. Achir, and A. Fladenmuller, "Associating the randomized bluetooth mac addresses of a device," in *IEEE Annual Consumer Communications & Networking Conference, CCNC 2021, Las Vegas, NV, USA, January 9-12, 2021*, pp. 1–6, 2021.

[6] 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*, pp. 15–20, 2016.

[7] 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*, pp. 265–276, 2013.

[8] S. Jamil, S. Khan, A. Basalamah, and A. Lbath, "Classifying smartphone screen on/off state based on wifi probe patterns," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pp. 301–304, 2016.

[9] C. Kolias, L. Copi, F. Zhang, and A. Stavrou, "Breaking ble beacons for fun but mostly profit," in *Proceedings of the 10th European Workshop on Systems Security*, pp. 1–6, 2017.

[10] M. Cominelli, F. Gringoli, P. Patras, M. Lind, and G. Noubir, "Even black cats cannot stay hidden in the dark: Full-band de-anonymization of bluetooth classic devices," in *2020 IEEE Symposium on Security and Privacy (SP)*, pp. 534–548, 2020.

[11] X. Gu, W. Wu, X. Gu, Z. Ling, M. Yang, and A. Song, "Probe request based device identification attack and defense," *Sensors*, vol. 20, no. 16, p. 4620, 2020.

[12] B. SIG, *Specification of the Bluetooth System, Core v5.2*. 2019-12-31.

[13] A. K. Mishra, A. C. Viana, and N. Achir, "SimBle: Comment générer des traces réelles Bluetooth conformes aux recommandations de préservation de la vie privée ?," in *ALGOTEL 2021 - 23èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications*, (La Rochelle, France), June 2021.

[14] B. SIG, *Specification of the Bluetooth System, Core v5.1*. 2019-01-21.

[15] E. McKinion and A. Lin, "Evaluation of security flaws in the current probe request design and proposed solutions," in *International Conference on Cyber Warfare and Security*, p. 529, Academic Conferences International Limited, 2017.