On Wi-Fi Tracking and the Pitfalls of MAC Address Randomization

Mathieu Cunche* Célestin Matte

Univ Lyon, INSA Lyon, Inria, CITI, France

Abstract

Wi-Fi has imposed itself as one of the key radio technology in portable devices such as smartphones, tablets, and other wearable devices. Because they periodically scan for nearby access points, Wi-Fi devices act as portable radio beacons emitting short messages, called *probe requests*. The latter contain a unique identifier: the MAC address of the Wi-Fi interface. It can be used to passively track individuals. Owners of such devices are thus exposed to passive tracking in the physical world.

MAC address randomization has been proposed as a way to prevent passive tracking, and is being progressively adopted by the industry. However, the MAC address is not the only piece of information that can be used for tracking. For instance, it has been recently shown that exploiting the content of frames as well as their timing could still lead to tracking despite MAC address randomization.

Keywords: 802.11; Wi-Fi; privacy; tracking; MAC randomization.

Index Terms: K.4.1 [Computers and society]: Public Policy Issues — Privacy;

1 Introduction

Wi-Fi has become a key technology in ubiquitous computing and is today integrated in many computing objects such as smartphones, tablets or wearable devices.

It is especially popular for smartphones because it provides free Internet connectivity in many places thanks to the growing number of hotspots or wireless community networks. Smartphones users thus tend to leave their Wi-Fi interface activated most of the time.

However, by having the Wi-Fi interface of their portable device activated, users are exposed to tracking. Smartphones broadcast Wi-Fi packets including a unique identifier when they scan for available Access Points (APs). These frames are sent in cleartext over the radio channels and can thus be passively recorded by any eavesdropper in range.

The technique of detecting the presence of individuals and tracking their movements using this information is known as *Wi-Fi tracking* and has been adopted by several kinds of entities. Researchers have deployed Wi-Fi tracking systems in cities to study urban mobility [15], while commercial entities are using Wi-Fi tracking to track their customers in their brick and mortars shops [5][2].

Passive tracking represents a privacy threat, especially if collected data is not properly anonymized [6]. In fact, the problem of passive tracking in wireless network has been known for some time and several countermeasures were proposed [10][16][8]. The most practical countermeasure is the use of a disposable random identifiers [10]: instead of using a unique and stable identifier, each wireless device periodically changes its identifier to a new random value. The recent development of Wi-Fi tracking has triggered the adoption of MAC randomization in Wi-Fi hardware by major industry stakeholders.

However, recent research works [7][17][14] have demonstrated that despite the use of a random identifier, Wi-Fi tracking is still possible. Indeed, some counters within Wi-Fi frames are not properly reset when switching for a new identifier [7][17], data elements can be used for fingerprinting [17], and timing patterns can be exploited to track a device over time [14].

This paper presents a comprehensive review of the privacy issues associated with Wi-Fi-based physical tracking. The technical details and applications of Wi-Fi-based physical tracking are first presented. Then, we describe MAC address randomization, the technical measure that is being widely adopted to prevent Wi-Fi tracking. Finally, we provide a list of attacks that can be used to track Wi-Fi devices despite the use of MAC address randomization, including attacks based on the content and the timing of Wi-Fi frames.

2 Wi-Fi and Service Discovery

Wi-Fi, also known as IEEE~802.11 [13], conveys information on radio channels. A key element of Wi-Fi is the service discovery mechanism that enables Wi-Fi stations to discover available APs and their capabilities. A Wi-Fi station can passively discover APs by listening to beacons they broadcast, or it can actively broadcast Probe Request frames to which nearby APs will respond by Probe Responses (see Figure 1:). The active service discovery mechanism is widely used by mobile devices, as it is less energy consuming. As a result, devices that are not associated to an AP periodically broadcast probe requests that contain a unique identifier, the MAC address, in their header (see Figure 2:).

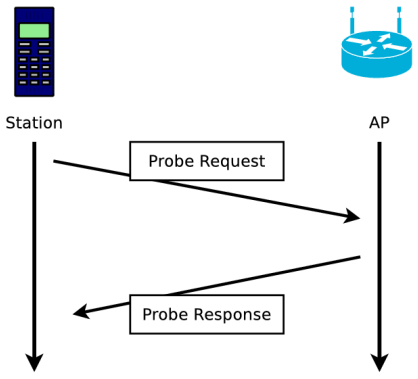

Figure 1: Active service discovery in 802.11. The station broadcasts Probe Requests and Access Points reply with Probe Responses.

The MAC address is an identifier associated to a network interface. It is composed of a 24-bits Organization Unique Identifier (OUI) designating the vendor, followed by a 24-bits

* mathieu.cunche@insa-lyon.fr

Network Interface Controller (NIC). Because the MAC address must be globally unique, each device can be identified by its Wi-Fi MAC address that can serve as an identifier for tracking.

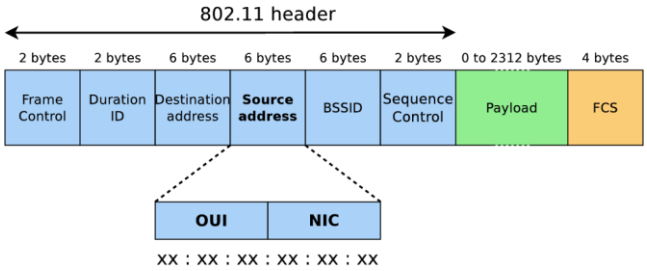

Figure 2: A 802.11 Probe Request frame, including the MAC address of the emitting device in the source field. The MAC address is a 48-bits identifiers composed of a OUI and a NIC.

3 Wi-Fi Based Physical Tracking

Physical tracking is the transposition in the physical world of tracking happening in the digital world, especially on the Web. Through passive collection of unique radio identifiers, third parties collect presence information about any individual. In the case of Wi-Fi, this collection is performed by a set of monitoring nodes, deployed over an area of interest, that forward collected information to a central server [2] (see Figure 3:).

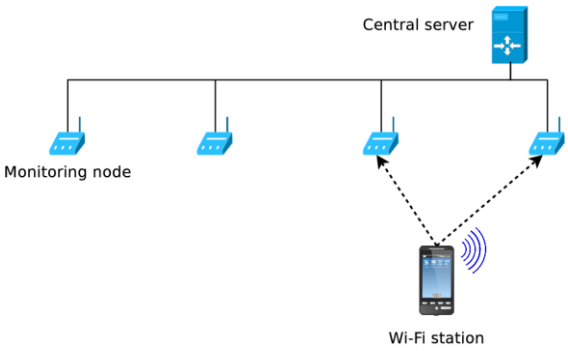

Figure 3: A Wi-Fi tracking system composed of monitoring nodes and a central server. Wi-Fi probes requests broadcast by stations are received by monitoring nodes that forward the presence detection information to the central server.

A first application of this presence data is the implementation of a physical analytics system providing aggregated information on the human activity at a specific location: e.g. number of visitors, duration and frequency of visits in a brick and mortar retail shop.

Data collected by Wi-Fi physical analytics is generally obtained without explicit consent of the user. In addition, a number of Wi-Fi trackers keep presence data in a raw or poorly anonymized format. Consumers are thus exposed to a real privacy threat which calls to technical solutions.

4 MAC Address Randomization

The privacy threat of the unique identifier in Wi-Fi protocols was identified early by Gruteser et al. [5], who suggested replacing this identifier by a temporary one that would be renewed periodically. The recent deployment of physical tracking systems and the popularization of portable devices has triggered the adoption of this technique in several operating systems.

Apple is the first major industrial stakeholder to adopt MAC address randomization, in version 8 of iOS [12]. The feature was subsequently introduced in Android 6.0 [1], Windows 10 [11], and the Linux iwlwifi driver [9]. For all platforms except Windows 10, the MAC address randomization is only applied to probe request frames sent when the station is in a scanning phase. However, Windows 10 also extends the use of random MAC addresses when the device is associated to a network: a random MAC address is generated each time the station connects to a new network [11].

Using a random MAC address instead of the genuine unique MAC address breaks the widely accepted axiom that each interface is associated with a unique identifier. For instance, this can lead to problems such as identifier collisions or resource exhaustion in protocols such as DHCP [3]. However, in the case of service discovery, the use of random MAC addresses will have little impact on other services, since the random MAC address is not used for traffic with an associated AP.

5 Content-based attacks

MAC randomization only ensures that the source field of a Wi-Fi frame cannot be used to track a device over an extended period of time. However, probes requests contain a number of other fields, both in their header and their payload that can be leveraged for tracking.

The first weakness with early implementations of MAC address randomization was identified by Freudiger [7]. He noticed that the sequence number field of probe requests emitted by iOS devices was not reset upon a MAC address change. This means that although the MAC address had been changed, it was still possible to link together consecutive random MAC addresses of a device using the sequence number.

A second issue related to another predictable field has been identified at the physical layer: the scrambler seed used with OFDM follows a predictable sequence that can be leveraged to link random MAC addresses of the same device together [4][17]. This issue has been experimentally confirmed on multiple commodity hardware [17]. Being a physical layer field, the scrambler seed cannot be directly modified by the driver, which makes it difficult to solve the problem only through software modifications.

Another way to defeat MAC address randomization is to use fingerprinting techniques to isolate a single device among a large group. Wi-Fi probe requests include *Information Elements* (IEs) in their payload. IEs are data blocks describing capabilities and features supported by the station. The number of these information elements and their variance is such that they provide enough information to create a fingerprint capable of defeating MAC address randomization [17].

One IEs found in probe request, the WPS IE, is of particular interest to defeat MAC randomization. Indeed, this IE contains a UUID field that is derived from the MAC address, and can be reversed back to the original MAC address of the device [17]. Using this re-identification attack, it has been possible to retrieve MAC addresses in an anonymized dataset [17].

6 Timing-based attacks

Timing information represents a second resource to defeat MAC address randomization. Indeed, when scanning in active service discovery mode, stations tend to follow predictable and identifying temporal patterns. Scans are performed by bursts, during which a station sends probe requests over Wi-Fi channels in a short period of time, typically less than 500ms (see Figure 4:).

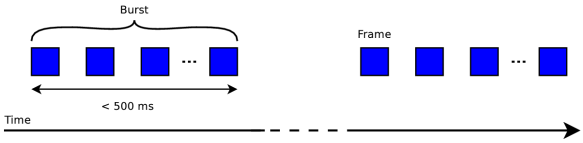

Figure 4: Transmission of Probe Requests over time. A burst is a group of frames transmitted during a time window of less than 500 ms.

By measuring the temporal distribution of probe requests within a burst but also between bursts, it is possible to create a temporal fingerprint as seen in Figure 5: that can help to isolate a device [14]. Using this technique it has been showed [14] that device can be tracked over time even if they change their identity, i.e. their MAC address.

Additionally; the order in which the channels are scanned can be used to further improve this temporal fingerprint [18].

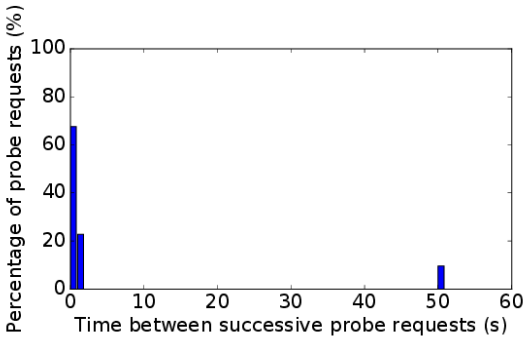

Figure 5: Signature of the probe requests timing of a device.

7 Conclusion

Owners of Wi-Fi enabled portable devices are exposed to passive physical tracking. A common practice adopted to prevent Wi-Fi tracking is the use of a MAC address randomization, especially in scanning phases of service discovery. However this method only remove the unique identifier in Wi-Fi frames and other means of identification and tracking can remain. Indeed, implementation of MAC randomization have left other fields, such as sequence number, untouched that can be exploited by a passive tracker. Worst other data items found in probe requests, such as the WPS UUID, can lead to the re-identification of the original MAC address. Finally, the timing of probe requests broadcast during scanning phases can also be leveraged to track a device despite its change MAC address.

As illustrated in this paper, protection against Wi-Fi tracking cannot solely rely on the use of random link identifier. More particularly it is necessary to consider other layer of the stack to ensure that no other identifying information is leaked.

References

[1] Android 6.0 changes. Retrieved from https://developer.android.com/about/versions/marshmallow/android-6.0-changes.html, 2015.

[2] Aerohive - Euclid Analytics. Aerohive and Euclid Retail Analytics. Retrieved from, `http://docs.aerohive.com/pdfs/Aerohive-Solution_Brief-Retail-Analytics.pdf` (Accessed 2016-10-06), 2013.

[3] Carlos J. Bernardos, Juan Carlos Zúñiga, and Piers O'Hanlon. Wi-Fi internet connectivity and privacy: hiding your tracks on the wireless internet. In *IEEE CSCN*, 2015.

[4] B. Bloessl, C. Sommer, F. Dressler, and D. Eckhoff. The scrambler attack: A robust physical layer attack on location privacy in vehicular networks. In *ICNC*, 2015.

[5] Brian Fung. How stores use your phoneâ€™s WiFi to track your shopping habits. https://www.washingtonpost.com/news/the-switch/wp/2013/10/19/how-stores-use-your-phones-wifi-to-track-your-shopping-habits/, October 2013.

[6] Levent Demir, Mathieu Cunche, and Cédric Lauradoux. Analysing the privacy policies of wi-fi trackers. In *Proceedings of the 2014 workshop on physical analytics*, pages 39–44. ACM, 2014.

[7] Julien 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*, page 8. ACM, 2015.

[8] Ben Greenstein, Damon McCoy, Jeffrey Pang, Tadayoshi Kohno, Srinivasan Seshan, and David Wetherall. Improving wireless privacy with an identifier-free link layer protocol. In *Proceedings of the 6th international conference on Mobile systems, applications, and services*, pages 40–53. ACM, 2008.

[9] Emmanuel Grumbach. iwlwifi: mvm: support random MAC address for scanning. Linux commit `effd05ac479b`.

[10] Marco Gruteser and Dirk Grunwald. Enhancing location privacy in wireless lan through disposable interface identifiers: a quantitative analysis. *Mobile Networks and Applications*, 10(3):315–325, 2005.

[11] Christian Huitema. Experience with MAC address randomization in Windows 10. In *93th Internet Engineering Task Force Meeting (IETF)*, July 2015.

[12] Lee Hutchinson. iOS 8 to stymie trackers and marketers with MAC address randomization. `http://arstechnica.com/apple/2014/06/ios8-to-stymie-trackers-and-marketers-with-mac-address-randomization/`, June 2014.

[13] IEEE Std 802.11-2012. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, 2012.

[14] Célestin Matte, Mathieu Cunche, Franck Rousseau, and Mathy Vanhoef. Defeating mac address randomization through timing attacks. In *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, WiSec '16, pages 15–20, New York, NY, USA, 2016. ACM.

[15] A. B. M. Musa and Jakob Eriksson. Tracking Unmodified Smartphones Using Wi-fi Monitors. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, SenSys '12, pages 281–294, New York, NY, USA, 2012. ACM.

[16] Jeffrey Pang, Ben Greenstein, Srinivasan Seshan, and David Wetherall. Tryst: The Case for Confidential Service Discovery. In *HotNets*, volume 2, page 1, 2007.

[17] Mathy Vanhoef, Célestin Matte, Mathieu Cunche, Leonardo S Cardoso, and Frank Piessens. Why mac address randomization is not enough: An analysis of wi-fi network discovery mechanisms. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, pages 413–424. ACM, 2016.

[18] Otto Waltari and Jussi Kangasharju. The wireless shark: Identifying wifi devices based on probe fingerprints. In *Proceedings of the First Workshop on Mobile Data*, pages 1–6. ACM, 2016.