Spread of MAC address randomization studied using locally administered MAC addresses use historic

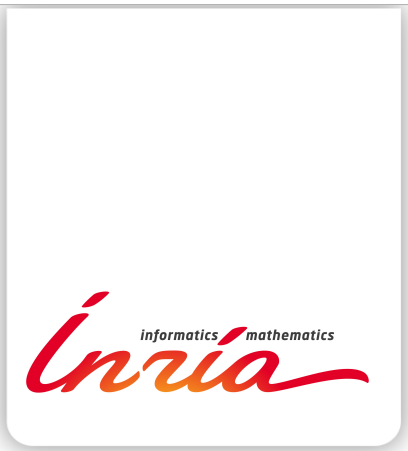
Célestin Matte, Mathieu Cunche

Spread of MAC address randomization studied using locally administered MAC addresses use historic

Célestin Matte, Mathieu Cunche

RESEARCH REPORT
N° 9142
December 2017
Project-Teams Privatics

ISSN 0249-6399 ISRN INRIA/RR--9142--FR+ENG



Spread of MAC address randomization studied using locally administered MAC addresses use historic

Célestin Matte, Mathieu Cunche

Project-Teams Privatics

Research Report n° 9142 — December 2017 — 7 pages

Abstract: In this document, we study the spread of random MAC addresses in Wi-Fi service discovery. To do so, we gather several datasets of probe requests from 2013-2017 and leverage an indicator of such addresses: the Locally Administered bit. We observe a trend of global increase in absolute random addresses use, even though the per-frame count these addresses is still low.

Key-words: Wi-Fi, tracking, privacy, random MAC address, locally administered bit

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Développement du changement aléatoire d'adresse MAC étudié grâce á l'utilisation des adresses MAC localement administrée

Résumé : Dans ce document, nous étudions l'évolution des adresses MAC aléatoires dans la découverte de service Wi-Fi. Pour ce faire, nous rassemblons plusieurs jeux de données récoltés entre 2013 et 2017 et nous servons d'un indicateur de l'utilisation de telles adresses : le bit Localement Administré. Nous observons une tendance globale à l'augmentation du nombre absolu d'adresses aléatoires utilisées, bien que le compte par trame soit toujours faible.

Mots-clés : Wi-Fi, traçage, vie privée, adresse MAC aléatoires, bit d'administration locale

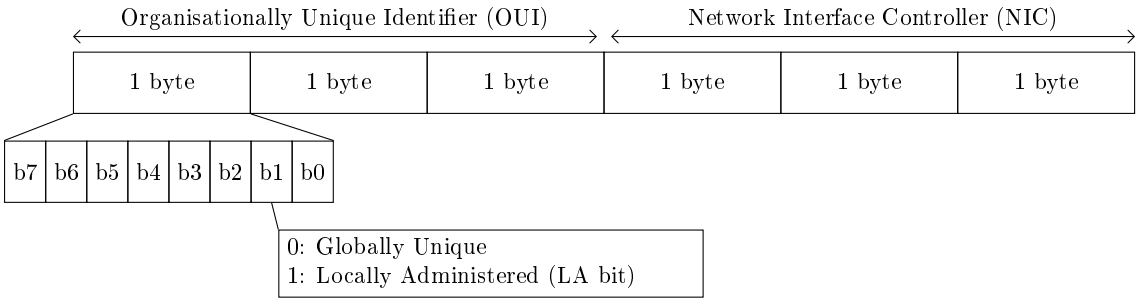


Figure 1: MAC address format.

1 Introduction

As a defense against the possibility of tracking, various vendors of Wi-Fi-enabled devices started implementing MAC address randomization during Wi-Fi service discovery [4]. In this document, we study the spread of such random addresses in the wild.

In order to communicate, devices address each other on the MAC layer using a 6-byte globally-unique identifier called the MAC address. As seen in figure 1, the first three bytes (prefix) of this address are an Organizationally Unique Identifier (OUI) which has to be bought by vendors from the IEEE Registration Authority in order to be used, so as to ensure the global uniqueness of MAC addresses. The last three bytes are the Network Interface Controller (NIC). One very specific bit of the MAC address is the seventh bit of the first byte of the OUI: the Locally Administered bit (LA bit). If set to 1, it indicates that the MAC address has been changed by the administrator of the device and is not guaranteed to be unique. A prefix whose LA bit is set to 1 is called a Company ID (CID), and is automatically purchased by manufacturers along with their OUI equivalent [2]. It is possible for companies to buy private OUIs, which are not publicly tied to the company's name [2].

It is unclear whether MAC address randomization during active scanning should set the LA bit to 1, as, to our knowledge, no document mentions this case explicitly [2]. While all firmware implementing randomization do not respect this behaviour, it constitutes a good indicator of MAC address randomization use among time. Apart from randomization, this bit is only set when users manually change their MAC address, or for rarely-used protocols (e.g. Wi-Fi Direct). We note, however, that some common devices (Nintendo devices), operate in P2P using an address with LA bit set to 1 [3]. Numbers presented later in this section indicate that, at a time when randomization was not integrated by vendors, LA bit use was anecdotal.

As no public study following the evolution of MAC address randomization exists, we compare LA bit use in datasets from different times and locations in table 1. This table lists LA bit use and the fraction of unallocated OUIs in randomized addresses in different datasets.

2 datasets

Used datasets are summarized in figure 2. We distinguish the situation of these datasets into several cases:

- *hotspot*: the recording matches what one would obtain by recording traffic at a public hotspot, i.e., users stay for some time (from minutes to hours) and the turn-over is high,

Name	Time	Place	Situation	MAC addr.	probe requests	Source
`Sapienza`	2013.02 - 2013.05	Rome	mix	160 000	8 000 000	[1]
`Middleware2014`	2014.12	Bordeaux	hotspot	900	140 000	personal
`Lab`	2015.10	Lyon	local AP	1 300	120 000	personal
`Train station`	2015.10 - 2015.11	Lyon	street	9 700	110 000	personal
`Glimps2015`	2015.12	Belgium	local AP	83 000	120 000	[6]
`Belgium`	2016.01 - 2016.02	Belgium	hotspot	3 700	200 000	[6]
`Martin`	2015.01 - 2016.12	Maryland	street	2 600 000	66 000 000	[3]
`Madeira`	2015.12 - 2017.06	Madeira	hotspot	13 000 000	300 000 000	not public

Figure 2: Used datasets

- *local AP*: what one would obtain by recording traffic at a home AP: users stay for a long time (possibly the whole capture), and the turn-over is low,
- *street*: what one would obtain by recording traffic in a street or along a road: most users are seen very briefly and the turn-over is high,
- *mix*: a combination of above cases.[1]

To give more details, we compare the fraction of vendors (according to OUIs) in the different datasets we have access to (in terms of number of non-random MAC addresses, identified by their LA bit) in figure 3. This gives an overview of the distribution of devices' vendors in the different datasets. Apple arrives first in all datasets, as the company manufactures all iOS devices, unlike Google and its Android OS. Large differences of appearance of a given vendor across several datasets can be attributed to their dates. Lower scores in all popular manufacturers for the `Madeira` can be explained by a hardly quantifiable number of random addresses not setting the LA bit to 1, and using already-attributed OUIs[2]. We're unsure about this source of incoherent addresses: it could be explained by some unidentified device models performing MAC address randomization in a yet-unobserved manner, or by an attacker deliberately spamming the systems with fake addresses[3].

3 Methodology

We compute the amount of devices setting their LA bit to 1 while emitting probe request frames. As we cannot reliably group probe requests from devices using MAC address randomization, we cannot obtain an exact per-device count. As a solution to this issue, we use both a per-MAC-address count and a per-probe-request count to compute the results. The former gives information on the actual number of detected MAC addresses, while the latter more accurately matches a per-device count. We could not compute the per-probe requests counts for 2 datasets because we did not have access to them. We must note that the per-MAC-address counts are expectedly highly variable: as devices using randomization use multiple MAC addresses, a small number of devices can drastically increase the amount of received randomized addresses. This is especially true if devices are monitored for a long period. For example, compare both counts for the `Middleware2014` dataset in the result (see next section).

[1] The `Sapienza` contains all these cases in different capture files.

[2] The large number of such addresses using OUIs of small companies not manufacturing any mobile devices tends to make us believe that these are not genuine OUIs.

[3] The attacker hypothesis seems more credible when we consider the fact that almost no MAC addresses in this dataset (0.05%) use both an unregistered OUI and a LA bit not set to 1.

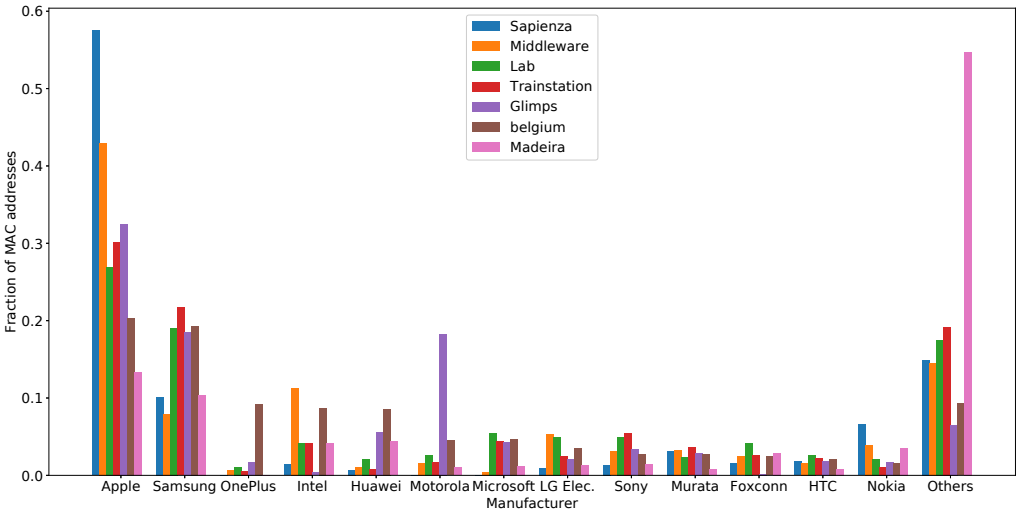


Figure 3: Fraction of non-random MAC addresses belonging to most-spread manufacturers. Represented manufacturers are those for which at least one dataset had 3% of its MAC addresses belonging to it.

4 Results

Results are presented in table 1.

This table indicates a clear evolution of LA bit use among time. Anecdotal in 2013, it impacted an important portion of probe requests in the beginning of 2016. Surprisingly, a very high amount of randomized addresses in most datasets use an unallocated OUI. Vendors are supposed to use registered OUIs when setting local addresses [2][4]. In fact, the equivalent CID of an OUI is automatically bought. This might indicate that most of these received probe requests come from implementations of MAC address randomization not respecting the 802.11 norm regarding MAC address use, such as manually-installed randomization applications. An example of this would be a device manually configured using the `macchanger` tool with its `-b` option ("burned-in-address" option, i.e., do-not-set-LA-bit option). The per-probe request counts indicate that these random addresses actually account for a very small fraction of devices, except in the `Glimps2015` dataset. This exception is not really surprising, considering the fact that devices are seen for a short time in this dataset.

In the `Martin` dataset, a progression of several Android-related CIDs used for MAC address randomization raises the fraction of registered OUI to 4.4%. However, the release of the iOS 10 OS around the same period (2016-09) also plays in favor of the former trend of unallocated OUI use for randomization, as iOS 10 devices use unregistered OUIs when using random addresses [3]. In 2012, Musa et al. remarked a similar but slightly different trend: in a 9-month public street deployment, 14% of the observed unique MAC addresses (over 60 000) used an unallocated OUI, whose LA bit is not set [5]. We did not observe this trend in any of the datasets we have access to: this ratio barely reaches 1% in one of them, and lies between 0.0 and 0.1% in the other ones.

[4] This document, however, does not explicitly mention systematic randomization of addresses during active scanning. To our knowledge, no official document does so.

When considering the per-probe request count in recent datasets, results depend on the type of dataset. In a *local AP* configuration (`Glimps2015` dataset), the fraction is close to the one obtained with a per-MAC address count, as most devices are detected for a brief amount of time. In a *hotspot* configuration (`Belgium` dataset), devices are detected for a long amount of time, resulting in a lot of address changes. As a consequence, while a lot of different addresses are captured (almost 50% of which are random), it amounts for a small fraction of probe requests (less than 3%). The latter case suggests that the fraction of devices using randomization is actually low in this dataset.

Table 1: Fraction of MAC addresses having a Locally Administered bit set to 1 over the total number of MAC addresses, in different datasets. "LA bit %" columns indicate the fraction of MAC addresses having their LA bit set to 1. The "Unalloc." column indicates fraction of these random addresses also using an unallocated OUI. Results are displayed by MAC addresses counts, and by probe requests count.

Dataset				Results			
				Per MAC addr.		Per probe requests	
Time	Name	MAC addr.	Probe req.	LA bit %	Unalloc.	LA bit %	Unalloc.
13.02-13.05	`Sapienza`	160 000	8 000 000	0.2%	33%	0.2%	13.5%
14.12	`Middleware2014`	900	140 000	47%	99.8%	1.5%	99.8%
15.10	`Lab`	1 300	120 000	14%	100%	1.7%	100%
15.10-15.11	`Train station`	9 700	110 000	23%	97.2%	10.0%	89.1%
15.12	`Glimps2015`	83 000	120 000	66.2%	99.0%	57.7%	98.9%
16.01-16.02	`Belgium`	3 700	200 000	48.8%	99.3%	2.8%	99.3%
15.01-16.12	`Martin`	2 600 000	66 000 000	53.8%	95.5%		
15.12-17.06	`Madeira`	13 000 000	300 000 000	99.8%	99.4%		

5 Conclusion

This study reveals a trend of increasing adoption of MAC address randomization. Most of these random addresses use an unregistered OUI (more than 95% in all recent datasets). While the fraction of detected random addresses is high in recent datasets (2016), a per-frame count of the use of these addresses suggests that the number of devices using MAC address randomization is still pretty low (less than 3%).

As a future work, it would be interesting to compute this trend on a given dataset running over a long period of time, to see the trend in a fixed environment.

References

[1] Marco V. Barbera, Alessandro Epasto, Alessandro Mei, Sokol Kosta, Vasile C. Perta, and Julinda Stefa. CRAWDAD dataset sapienza/probe-requests (v. 2013-09-10). Retrieved 10 November, 2015, from, `http://crawdad.org/sapienza/probe-requests/20130910`, September 2013.

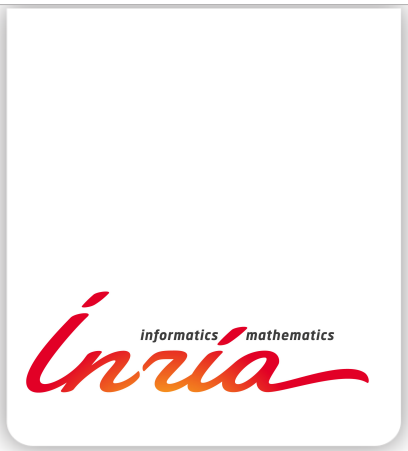
[2] IEEE. Guidelines for use organizationally unique identifier (oui) and company id (cid). `http://standards.ieee.org/develop/regauth/tut/eui.pdf`, consulted on 2017.05.28, 2014.

[3] Jeremy Martin, Travis Mayberry, Collin Donahue, Lucas Foppe, Lamont Brown, Chadwick Riggins, Erik C Rye, and Dane Brown. A study of MAC address randomization in mobile devices and when it fails. *arXiv preprint arXiv:1703.02874*, 2017.

[4] Célestin Matte. *Wi-Fi Tracking: Fingerprinting Attacks and Counter-Measures.* Thesis, Université de Lyon, December 2017.

[5] ABM Musa and Jakob Eriksson. Tracking unmodified smartphones using Wi-Fi monitors. In *Proceedings of the 10th ACM conference on embedded network sensor systems*, pages 281–294. ACM, 2012.

[6] Pieter Robyns, Bram Bonné, Peter Quax, and Wim Lamotte. Noncooperative 802.11 MAC layer fingerprinting and tracking of mobile devices. *Security and Communication Networks*, 2017, 2017.



RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399