Wi-Fi Tracking: Fingerprinting Attacks and Counter-Measures

Célestin Matte
Univ Lyon, INSA Lyon, Inria, Privatics team, CITI, France, Région Rhône-Alpes & Inria fundings
Supervised by: Marine Minier, Mathieu Cunche, Franck Rousseau

Insa Lyon, December 7th 2017

Introduction - Physical tracking

[1]

Physical tracking

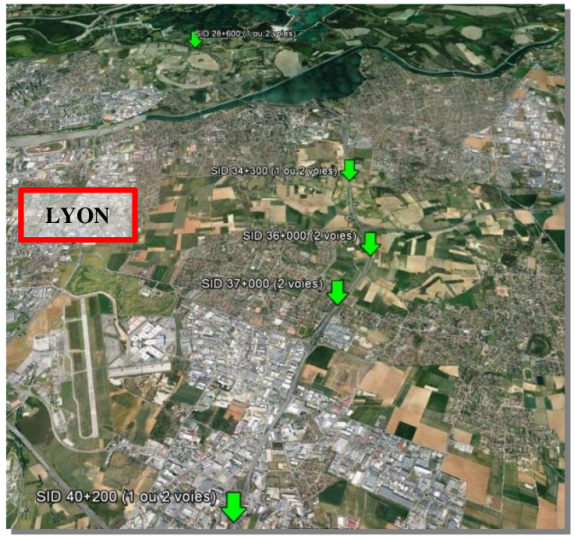
Getting the knowledge of a device's presence and mobility over time

- How: Monitoring Wi-Fi frames emitted by Wi-Fi-enabled devices, e.g. smartphones
- Retailers for analytics, intelligence services or nasty neighbours for spying, employers for monitoring...
- Privacy issue: no consent nor awareness to get sensitive information
- Why now? Spread of ubiquitous computing

[1] Source: `http://www.libelium.com/products/meshlium/smartphone-detection/`

Introduction - Tracking - Example: Lyon's ring road

Fig. (7) - Les balises Bluetooth/Wifi installées sur portique (à gauche) et mât temporaire (à droite)



Tronçon	Longueur
AB	3,2 km
BC	1,2 km
CD	1,6 km
DE	5,9 km

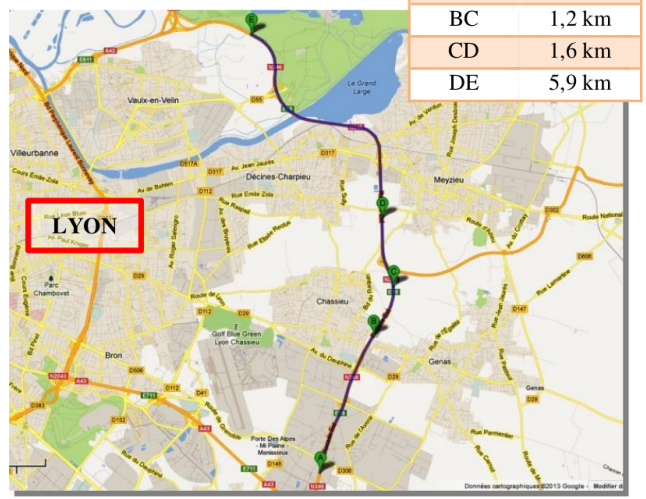


Fig. (1) et (2) - Localisation des points d'identification

- Bluetooth and Wi-Fi-based installation on Lyon's ring road[2][3]
- Goal: real-time travel time estimation
- Problem: lack of information and consent

[2] Atec ITS France, ed. *Evaluations simultanées de différentes technologies innovantes de recueil de données trafic pour le calcul de temps de parcours en temps réel*. 2015.

[3] Guillaume Grolleau. *La captation Bluetooth au service des aménagements urbains*. Ed. by Atec ITS France. 2015.

Introduction - Tracking - Other examples

Le BHV aspire les données de ses clients, mais il est loin d'être le seul

Par Elisa Braun | Mis à jour le 03/08/2017 à 14:58 / Publié le 02/08/2017 à 18:39

2 commentaires

La célèbre enseigne de l'Hôtel de Ville a mis en place un système pour tracer le

Introduction - Tracking - Other examples

Le BHV aspire les données de ses cli

Par Elisa Braun

La célèbre enseigne de

Exclusive: Here's What 3 Big Museums Learn By Tracking Your Phone

By **James O Malley** on 11 Apr 2017 at 12:00PM

At least three of Britain's most popular cultural institutions have been tracking visitors using the wifi on their phones, Gizmodo UK can exclusively reveal. Following a series of Freedom of Information Requests, the National Gallery and Natural History Museums in London, as well as the National Railway Museum in York, have all revealed that they have tested or deployed tracking software - which could conceivably help curators and managers make decisions.

Introduction - Tracking - Other examples

To ensure your journey via Schiphol is as pleasant as possible, we measure the number of passengers to allow us to give you information on expected peaks and waiting times. In the terminal we therefore work with a Wi-Fi and Bluetooth tracking system. You are of course always free to switch off your Wi-Fi and Bluetooth. Schiphol uses sensors to trace Bluetooth and wifi signals. A device (mobile telephone, laptop, etc.) can be identified by its unique 'MAC address'. This MAC address does not link to individual user data and personal information is not compromised.

Le BHV
ses cli

Par Elisa Braun

La célèbre enseigne de

19th Feb 5th Mar' 28th Mar'

Exclusive: Here's What 3 Big Museums Learn By Tracking Your Phone

By **James O Malley** on 11 Apr 2017 at 12:00PM

At least three of Britain's most popular cultural institutions have been tracking visitors using the wifi on their phones, Gizmodo UK can exclusively reveal. Following a series of Freedom of Information Requests, the National Gallery and Natural History Museums in London, as well as the National Railway Museum in York, have all revealed that they have tested or deployed tracking software - which could conceivably help curators and managers make decisions.

Introduction - Tracking - Other examples

Le BHV
ses cli

Par Elisa Braun

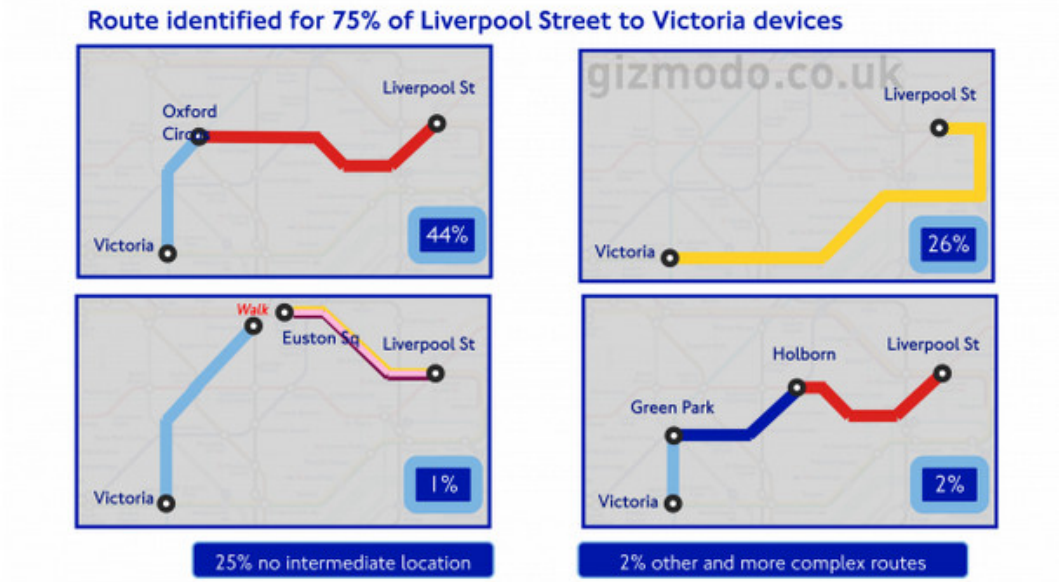
LE FIGARO PREMIUM
› 1€ le premier mois

La célèbre enseigne de

GIZMODO | UK

UK NEWS GADGETS DESIGN WATCH THIS WTF SCIENCE APPLE AN

TRANSPORT



Here's What TfL Learned From Tracking Your Phone On the Tube

By James O Malley on 13 Feb 2017 at 1:24PM

sible, we measure the
on on expected peaks
ith a Wi-Fi and
e to switch off your
tooth and wifi signals.
ed by its unique 'MAC
ser data and personal

Introduction - Tracking - Other examples

Le BHV
ses cli

Par Elisa Braun

LE FIGARO PREMIUM
› 1€ le premier mois

La célèbre enseigne de

GIZMODO | UK

UK NEWS GADGETS DESIGN WATCH THIS WTF SCIENCE APPLE AN

TRANSPORT

Route identified for 75% of Liverpool Street to Victoria devices

Her
Fro
On the Tube

By James O Malley on 13 Feb 2017 at 1:24PM

ssible, we measure the
on on expected peaks
ith a Wi-Fi and
e to switch off your
tooth and wifi signals.
unique 'MAC
and personal

Introduction - Tracking - Other examples

GIZMODO | UK

...sible, we measure the
...on on expected peaks
...-Fi and
...ch off your
...nd wifi signals.
...unique 'MAC
...and personal

Le
s

20 minutes

Lire le journal du
jeudi 23 novembre
TÉLÉCHARGER LE PDF

NEWSLETTER CONNEXION

Recherche (ex : Réforme des retraites, etc.)

#CoupeDavis #SousMarinDisparu #HarcelementSexuel #LigueEuropa Actualité Locales Sport Entertainment Economie

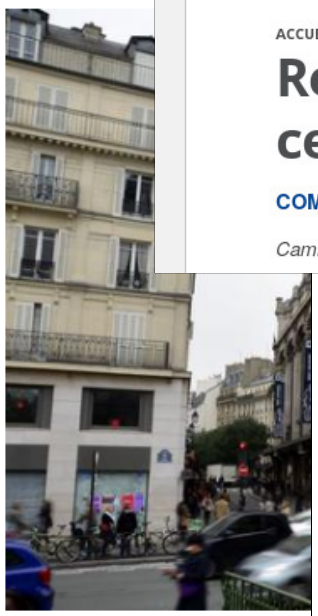
Bordeaux Strasbourg Toulouse Lille Lyon Marseille Montpellier Nantes Nice Paris Rennes

ACCUEIL › RENNES

Rennes : Des capteurs wifi pour suivre les clients du centre-ville

COMMERCE Trente magasins seront équipés dans les jours à venir…

Camille Allain | Publié le 10/02/17 à 07h05 — Mis à jour le 10/02/17 à 07h05



LE FIGARO PREMIUM
› 1€ le premier mois

La célèbre enseigne de

Her
Fro
On the Tube

By **James O Malley** on 13 Feb 2017 at 1:24PM

Introduction - Tracking - Other examples

GIZMODO | UK

sible, we measure the
on on expected peaks
Fi and
ch off your
nd wifi signals.
unique 'MAC
and personal

Le
s

VB NEWS EVENTS RESEARCH Search

MARKETING EXCLUSIVE

Drones overhead in L.A.'s Valley are tracking mobile devices' locations

BARRY LEVINE @XBARRYLEVINE FEBRUARY 23, 2015 6:11 AM

Above: The Adnear drone over LA

Image Credit: Adnear

It was only a matter of time before drones started monitoring signals from mobile devices.

VB Recommendations

Samsung Galaxy S9 and
iterative upgrades, will
cameo at CES

How is Uber still even in
this point?

Microsoft just took a big
AWS and VMware with
offerings

Upcoming Events

BLUEPRINT Mar 5 - 7

LE FIGAR
1€ le p

La célèbre

Introduction - Tracking - Other examples

GIZMODO | UK

ssible, we measure the
on on expected peaks
Fi and
ch off your
nd wifi signals.
unique 'MAC
and personal

Le
s

VB NEWS EVENTS RESEARCH Search

MARKETING EXCLUSIVE

Drones overhead in L.A.'s Valley are tracking mobile devices' locations

BARRY LEVINE @XBARRYLEVINE FEBRUARY 23, 2015 6:11 AM

Although this experiment in the Valley is currently testing location-mapping from drones and is not yet used to send ads, Adnear said it currently has over 530 million user profiles covering various Asian markets for its other location-based campaigns.

Above: The Adnear drone over LA

Image Credit: Adnear

It was only a matter of time before drones started monitoring signals from mobile devices.

VB Recommendations

Samsung Galaxy S9 and
iterative upgrades, will
cameo at CES

How is Uber still even in
this point?

Microsoft just took a big
AWS and VMware with
offerings

Upcoming Events

BLUEPRINT Mar 5 - 7

LE FIGAR
1€ le p

La célèbre

Introduction - Tracking - Other examples

The Intercept_

825

In addition to the GILGAMESH system used by JSOC, the CIA uses a similar NSA platform known as SHENANIGANS. The operation – previously undisclosed – utilizes a pod on aircraft that vacuums up massive amounts of data from any wireless routers, computers, smart phones or other electronic devices that are within range.

One top-secret NSA document provided by Snowden is written by a SHENANIGANS operator who documents his March 2012 deployment to Oman, where the CIA has established a drone base. The operator describes how, from almost four miles in the air, he searched for communications devices believed to be used by Al Qaeda in the Arabian Peninsula in neighboring Yemen.The mission was code named VICTORYDANCE.

"The VICTORYDANCE mission was a great experience," the operator writes. "It was truly a joint interagency effort between CIA and NSA. Flights and targets were coordinated with both CIAers and NSAers. The mission lasted 6 months, during which 43 flights were flown."

VICTORYDANCE, he adds, "mapped the Wi-Fi fingerprint of nearly every major town in Yemen."

Introduction - Tracking - Opposition

M PIXELS | CHRONIQUES DES (R)ÉVOLUTIONS NUMÉRIQUES | VIE EN LI

Le Conseil d'Etat empêche définitivement JCDecaux de pister les téléphones des passants

L'entreprise souhaitait collecter les identifiants des téléphones portables des personnes passant à côté de ses panneaux publicitaires à La Défense. Le Conseil d'Etat le lui a interdit, en confirmant une décision de la CNIL.

LE MONDE | 09.02.2017 à 15h41 • Mis à jour le 09.02.2017 à 16h50

Abonnez vous à partir de 1 € | Réagir | Ajouter | Partager (3 174) | Tweeter

C'est non : JCDecaux ne pourra pas tracer les téléphones des passants à partir de ses panneaux publicitaires. Mercredi 8 février, le Conseil d'Etat a mis un point final à l'affaire qui opposait depuis deux ans l'entreprise de mobilier urbain à la CNIL (Commission nationale de l'informatique et des libertés).

Introduction - Tracking - Opposition

M PIXELS | CHRONIQUES DES (R)ÉVOLUTIONS NUMÉRIQUES | VIE EN LI

Le Conseil d'Etat empêche définitivement JCDecaux de pister

Rennes: Les commerçants reportent la mise en service des capteurs wifi suivant les smartphones

COMMERCE Une demande complémentaire a été adressée à la CNIL...

C.A. | Publié le 09/03/17 à 12h57 — Mis à jour le 09/03/17 à 14h26

LE MONDE | 09.02.2017 à 15h41 • Mis à jour le 09.02.2017 à 16h50

Abonnez vous à partir de 1 € | Réagir | Ajouter | Partager (3 174) | Tweeter

C'est non : JCDecaux ne pourra pas tracer les téléphones des passants à partir de ses panneaux publicitaires. Mercredi 8 février, le Conseil d'Etat a mis un point final à l'affaire qui opposait depuis deux ans l'entreprise de mobilier urbain à la CNIL (Commission nationale de l'informatique et des libertés).

Introduction - Tracking - Opposition

PIXELS

CHRONIQUES
DES (R)ÉVOLUTIONS NUMÉRIQUES

VIE EN LI

Tech

APR 23, 2015 @ 02:37 PM 1,733

The Little Black Book of Bil

FTC Pursues Tech Company It Claims Violated Privacy Policy While Tracking 9 Million Phones

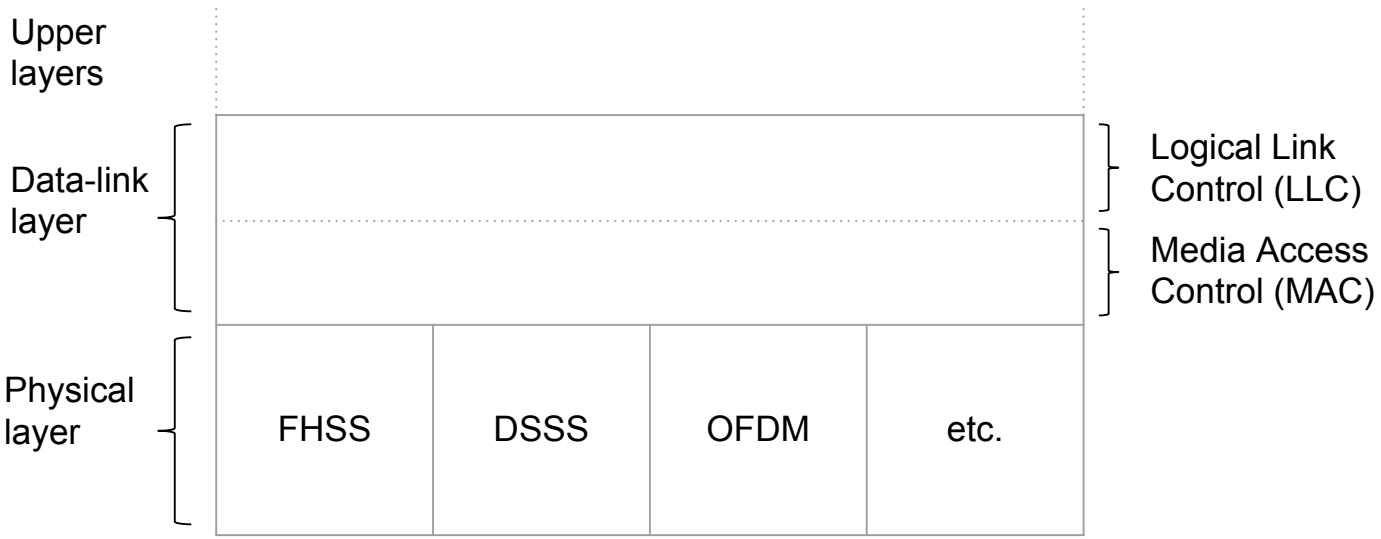
Kate Vinton, FORBES STAFF
FULL BIO

A technology company that helps retailers track consumers through mobile devices agreed Thursday to settle after the FTC issued a complaint about the company's violations to its privacy policy. In the first year after being founded in September 2012, Nomi Technologies allegedly tracked at least 9 million consumer cell phones without giving consumers a promised opt-out option in store, according to the FTC's complaint. Most shoppers would have had no idea they were being tracked at all.

C'est non : JCDecaux ne pourra pas tracer les téléphones des passants à partir de ses panneaux publicitaires. Mercredi 8 février, le Conseil d'Etat a mis un point final à l'affaire qui opposait depuis deux ans l'entreprise de mobilier urbain à la CNIL (Commission nationale de l'informatique et des libertés).

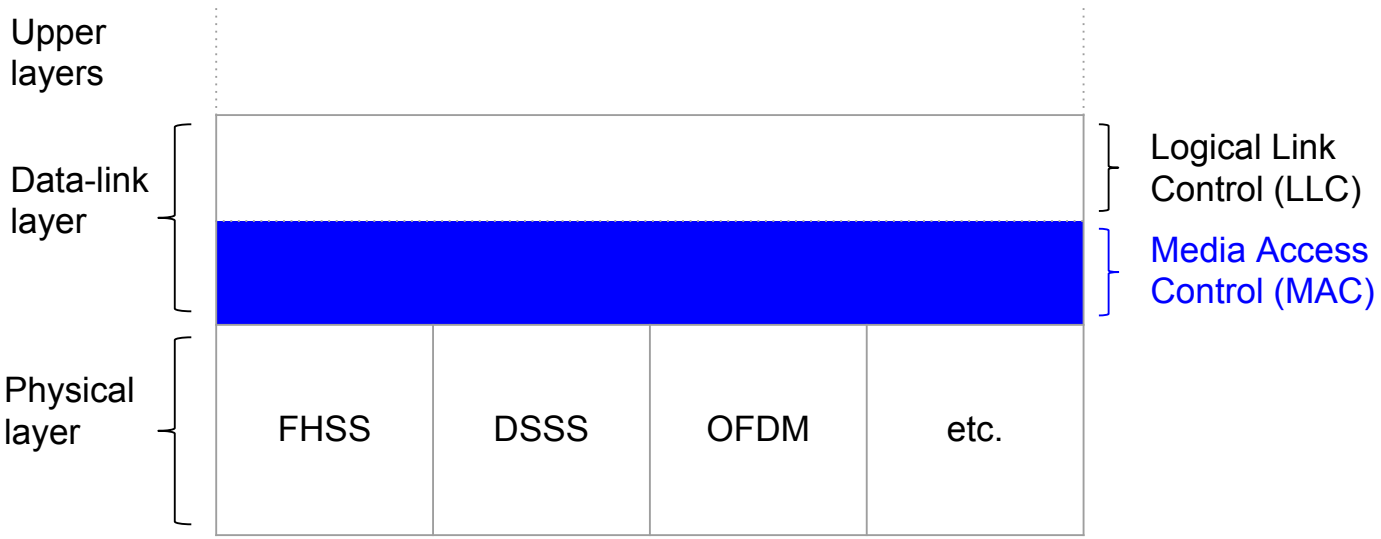
Introduction - Wi-Fi

- IEEE 802.11 Wi-Fi: set of protocols
- Short-range wireless networks
- PDU: On this layer, messages are called **frames**



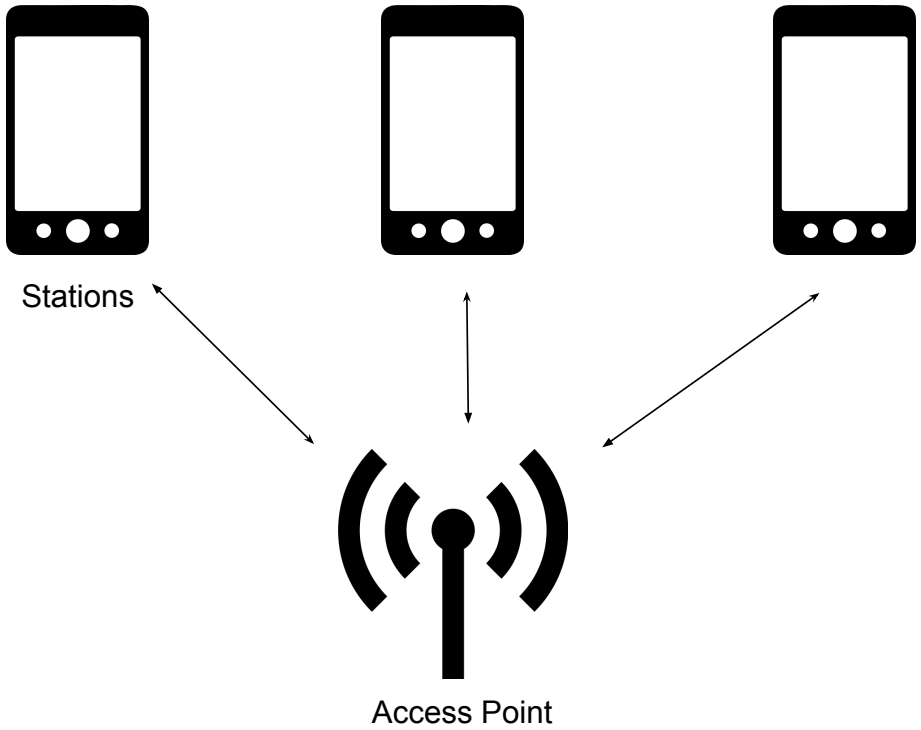
Introduction - Wi-Fi

- IEEE 802.11 Wi-Fi: set of protocols
- Short-range wireless networks
- PDU: On this layer, messages are called **frames**



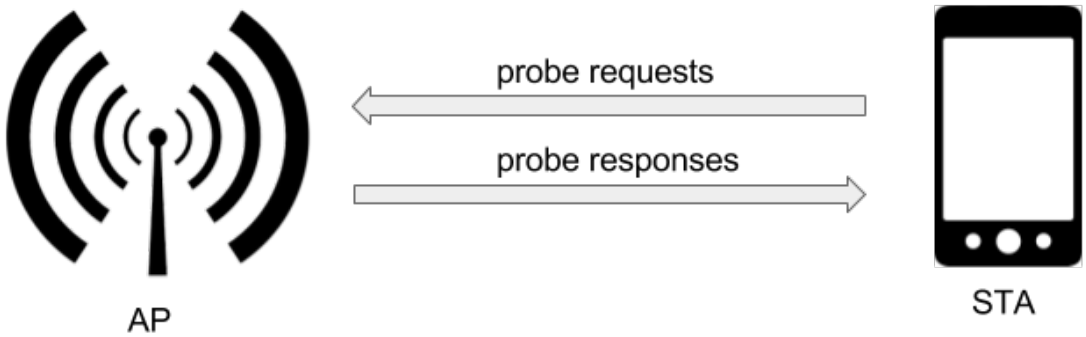
Introduction - Wi-Fi service discovery 1/2

Wi-Fi infrastructure mode is asymmetric: Access Point (AP, "server") and stations ("clients")



Introduction - Wi-Fi service discovery 2/2

- Stations discover APs by sending **probe request** frames
- Sent in groups called **bursts** (< 100 ms)
- These frames are sent several times per minute[4]
- Even unassociated devices emit these frames

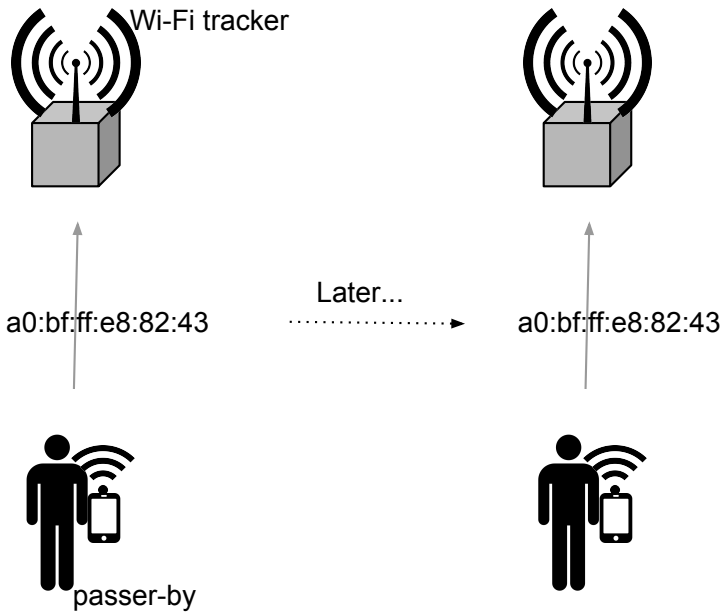


MAC address

- MAC layer's globally unique identifier
- a 6-byte serial number. Ex: `ef:4b:48:ab:42:37`

[4] Julien Freudiger. "How Talkative is Your Mobile Device? An Experimental Study of Wi-Fi Probe Requests". In: *ACM WiSec.* 2015.

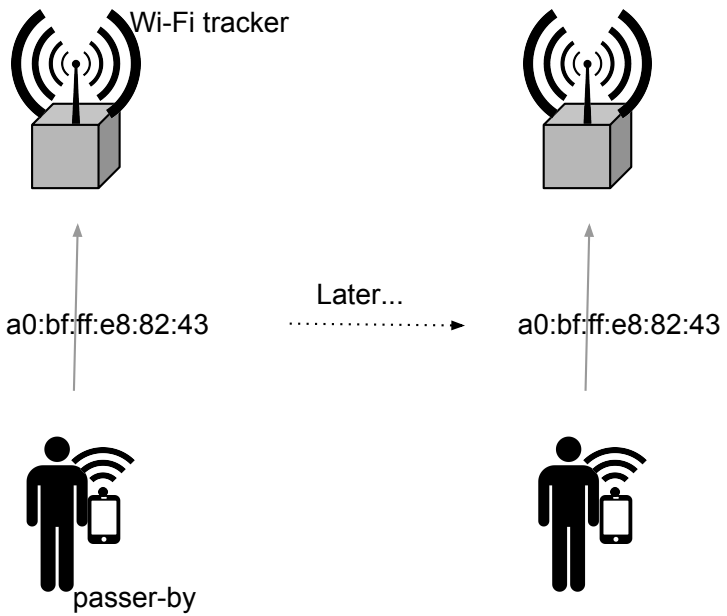
Introduction - Device tracking



Random MAC addresses - Introduction

MAC address randomization

The idea: frequently changing the MAC address identifier to a different randomly generated address[5]

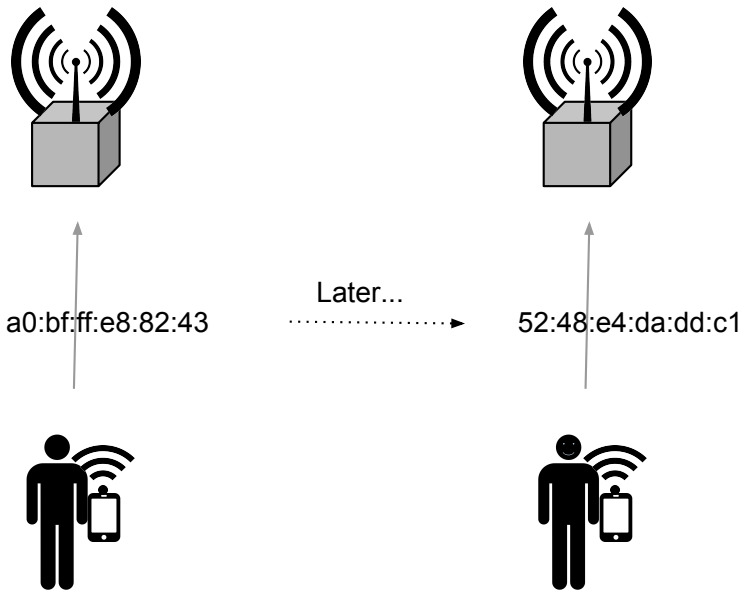


[5] Marco Gruteser and Dirk Grunwald. "Enhancing Location Privacy in Wireless LAN Through Disposable Interface Identifiers: A Quantitative Analysis". In: *Mobile Networks and Applications* 10.3 (2005)

Random MAC addresses - Introduction

MAC address randomization

The idea: frequently changing the MAC address identifier to a different randomly generated address[5]



[5] Marco Gruteser and Dirk Grunwald. "Enhancing Location Privacy in Wireless LAN Through Disposable Interface Identifiers: A Quantitative Analysis". In: *Mobile Networks and Applications* 10.3 (2005)

Random MAC addresses - Implementations

- First core implementations: 2014
- Various implementations:
  - iOS since iOS 8
  - Windows since Windows 10
  - Android since Android 6.0
  - Linux since kernel 3.18
- No standard specifying random MAC addresses
- Examples of differences:
  - Random addresses used only during service discovery or not
  - Full address changed, or only the last 3 bytes?
  - Frequency of change
  - etc.
- Requires support from various components: firmware, driver, software
- Tests on various devices revealed many shortcomings in current implementations[6]

[6] Julien Freudiger. "How Talkative is Your Mobile Device? An Experimental Study of Wi-Fi Probe Requests". In: *ACM WiSec.* 2015.

Random MAC addresses - Case study - Nexus 6P

- End of 2015, manufactured by Huawei and developed by Google
- Android 6.0, Broadcom chipset for Wi-Fi
- Monitored multiple channels, according to several use cases

Positive points	Negative points
• Random MAC address • Changed on every burst • Android "random" OUI	• Biased PRNG: reused addresses • Contiguous sequence numbers • Actual MAC address leaked under certain conditions • Regular timing patterns • Plenty of Information Elements

Random MAC addresses - Case studies summary

- 6 recent devices tested
- Many flaws identified
  - All devices are affected by at least 2 of these flaws
  - Some flaws affects all devices

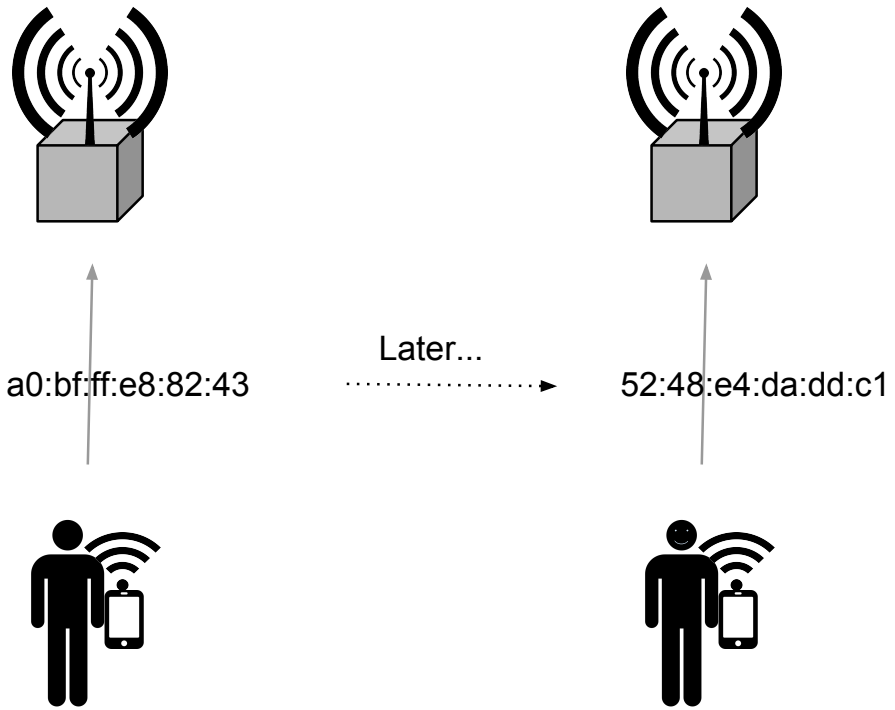
	Nexus 6P	Nexus 5X	OnePlus 3	iPad 2	iPhone 6	iPhone 7
Random MAC address	✓	✓	✓	✓	✗	✓
Information elements	✗	✗	✗	✗	✗	✗
Random sequence numbers	✗	✗	✗	(✓)	(✓)	(✓)
No SSID in random probes	✗	✗	✗	✗	✗	
Global address not leaked	✗	✗	✓	✗	✗	
Address changed each burst	✓	✓	✓	✗		✗
No reused addresses	✗	✗	✗			
Random OUI	(✓)	(✓)	(✓)	✓		✓
No regular timing pattern	✗	✗	✗	✗	✗	

Random MAC addresses - Case studies summary

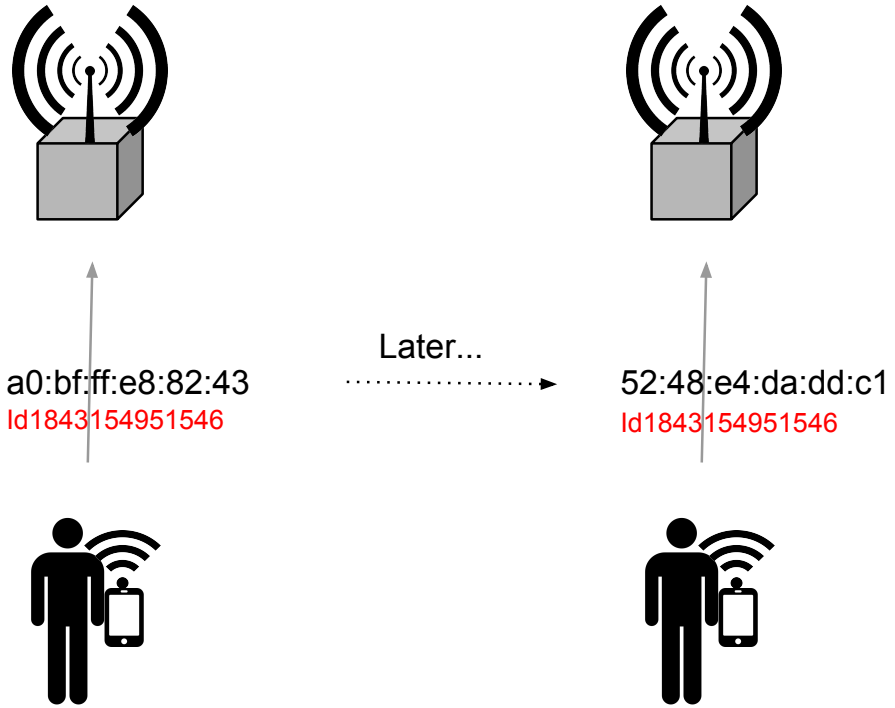
- 6 recent devices tested
- Many flaws identified
  - All devices are affected by at least 2 of these flaws
  - Some flaws affects all devices

	Nexus 6P	Nexus 5X	OnePlus 3	iPad 2	iPhone 6	iPhone 7
Random MAC address	✓	✓	✓	✓	✗	✓
Information elements	✗	✗	✗	✗	✗	✗
Random sequence numbers	✗	✗	✗	(✓)	(✓)	(✓)
No SSID in random probes	✗	✗	✗	✗	✗	
Global address not leaked	✗	✗	✓	✗	✗	
Address changed each burst	✓	✓	✓	✗		✗
No reused addresses	✗	✗	✗			
Random OUI	(✓)	(✓)	(✓)	✓		✓
No regular timing pattern	✗	✗	✗	✗	✗	

Random MAC addresses - Example of failure



Random MAC addresses - Example of failure



- MAC address randomization fails if other identifiers exist
- This thesis focused on this kind of issues

IE - Fingerprinting using Information Elements

Information Elements (IE, tagged parameters, tags)

- Fields of all management frames (including probe requests)
- Indicate the support of capabilities

```
▼Tag: HT Capabilities (802.11n D1.10)
  Tag Number: HT Capabilities (802.11n D1.10) (45)
  Tag length: 26
 ▼HT Capabilities Info: 0x100c
    .... .... .... ...0 = HT LDPC coding capability: Transmitter does not support receiving LDPC coded packets
    .... .... .... ..0. = HT Support channel width: Transmitter only supports 20MHz operation
    .... .... .... 11.. = HT SM Power Save: SM Power Save disabled (0x0003)
    .... .... ...0 .... = HT Green Field: Transmitter is not able to receive PPDUs with Green Field (GF) preamble
    .... .... ..0. .... = HT Short GI for 20MHz: Not supported
    .... .... .0.. .... = HT Short GI for 40MHz: Not supported
    .... .... 0... .... = HT Tx STBC: Not supported
    .... ..00 .... .... = HT Rx STBC: No Rx STBC support (0x0000)
    .... .0.. .... .... = HT Delayed Block ACK: Transmitter does not support HT-Delayed BlockAck
    .... 0... .... .... = HT Max A-MSDU length: 3839 bytes
    ...1 .... .... .... = HT DSSS/CCK mode in 40MHz: Will/Can use DSSS/CCK in 40 MHz
    ..0. .... .... .... = HT PSMP Support: Won't/Can't support PSMP operation
    .0.. .... .... .... = HT Forty MHz Intolerant: Use of 40 MHz transmissions unrestricted/allowed
    0... .... .... .... = HT L-SIG TXOP Protection support: Not supported
 ▼A-MPDU Parameters: 0x19
    .... ..01 = Maximum Rx A-MPDU Length: 0x01 (16383[Bytes])
    ...1 10.. = MPDU Density: 8 [usec] (0x06)
    000. .... = Reserved: 0x00
 ▶Rx Supported Modulation and Coding Scheme Set: MCS Set
 ▶HT Extended Capabilities: 0x0000
 ▶Transmit Beam Forming (TxBF) Capabilities: 0x0000
 ▶Antenna Selection (ASEL) Capabilities: 0x00
```

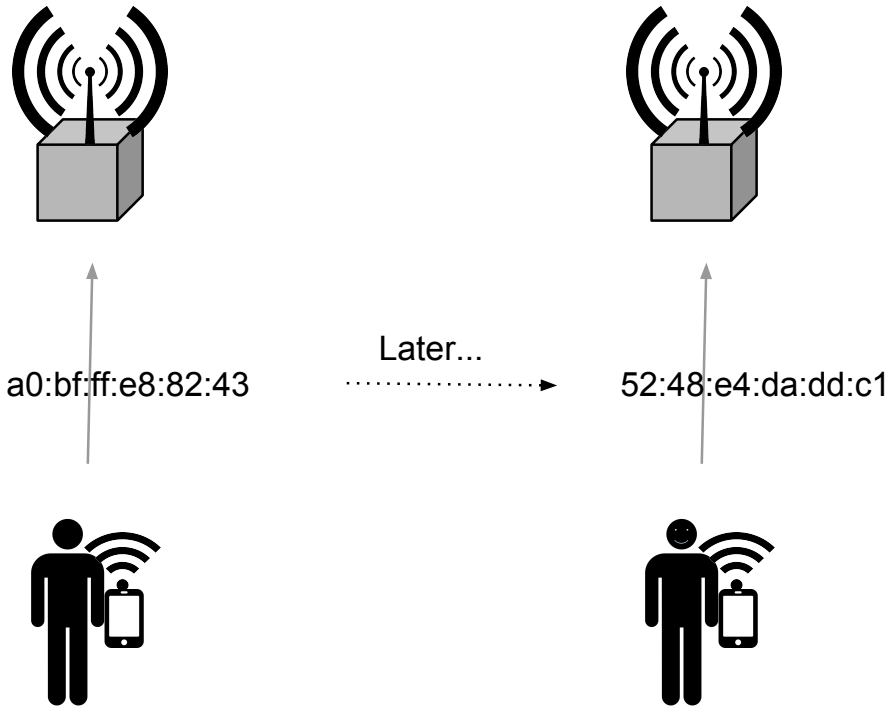
IE - Fingerprinting using Information Elements

Information Elements (IE, tagged parameters, tags)

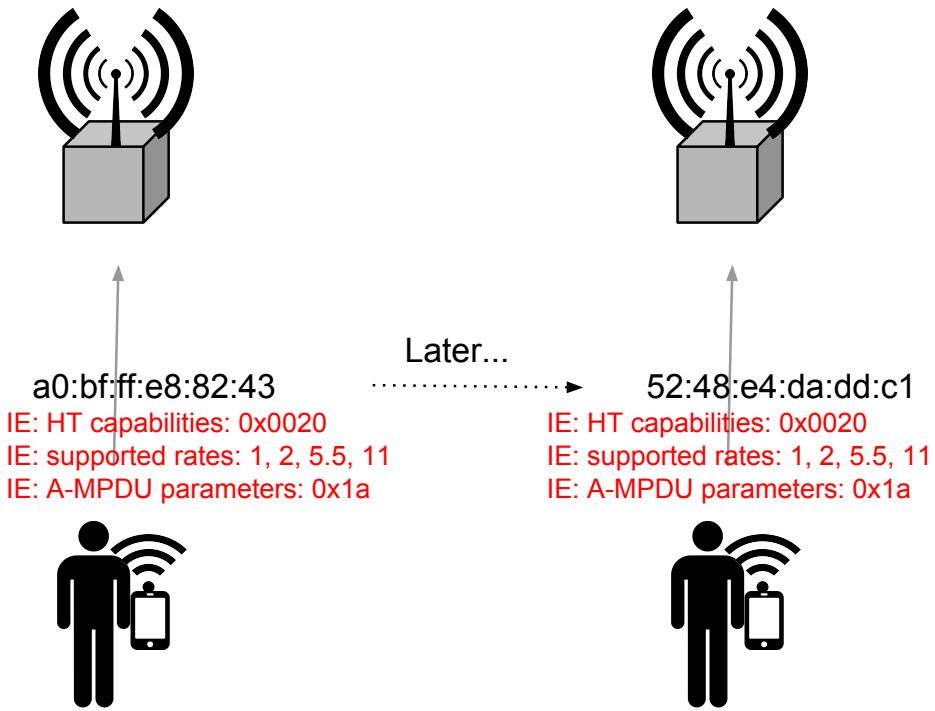
- Fields of all management frames (including probe requests)
- Indicate the support of capabilities

```
▼Tag: HT Capabilities (802.11n D1.10)
  Tag Number: HT Capabilities (802.11n D1.10) (45)
  Tag length: 26
 ▼HT Capabilities Info: 0x100c
    .... .... .... ...0 = HT LDPC coding capability: Transmitter does not support receiving LDPC coded packets
    .... .... .... ..0. = HT Support channel width: Transmitter only supports 20MHz operation
    .... .... .... 11.. = HT SM Power Save: SM Power Save disabled (0x0003)
    .... .... ...0 .... = HT Green Field: Transmitter is not able to receive PPDUs with Green Field (GF) preamble
    .... .... ..0. .... = HT Short GI for 20MHz: Not supported
    .... .... .0.. .... = HT Short GI for 40MHz: Not supported
    .... .... 0... .... = HT Tx STBC: Not supported
    .... ..00 .... .... = HT Rx STBC: No Rx STBC support (0x0000)
    .... .0.. .... .... = HT Delayed Block ACK: Transmitter does not support HT-Delayed BlockAck
    .... 0... .... .... = HT Max A-MSDU length: 3839 bytes
    ...1 .... .... .... = HT DSSS/CCK mode in 40MHz: Will/Can use DSSS/CCK in 40 MHz
    ..0. .... .... .... = HT PSMP Support: Won't/Can't support PSMP operation
    .0.. .... .... .... = HT Forty MHz Intolerant: Use of 40 MHz transmissions unrestricted/allowed
    0... .... .... .... = HT L-SIG TXOP Protection support: Not supported
 ▼A-MPDU Parameters: 0x19
    .... ..01 = Maximum Rx A-MPDU Length: 0x01 (16383[Bytes])
    ...1 10.. = MPDU Density: 8 [usec] (0x06)
    000. .... = Reserved: 0x00
 ▶Rx Supported Modulation and Coding Scheme Set: MCS Set
 ▶HT Extended Capabilities: 0x0000
 ▶Transmit Beam Forming (TxBF) Capabilities: 0x0000
 ▶Antenna Selection (ASEL) Capabilities: 0x00
```

IE - Introduction



IE - Introduction



IE - Introduction - Identifiers/fingerprinting (historics)

- IEs have a high diversity in term of values
- Idea: Exploit this diversity to fingerprint devices

Fingerprint

A set of information used to identify or classify a target

- Old technique. Ex: radio-frequency fingerprinting during WWII
- Wi-Fi-based fingerprinting examples:
  - Franklin et al., 2006[7]: timing of probe requests
  - Pang et al., 2007[8]: implicit identifiers
  - Neumann et al., 2012[9]: many features: transmission time, timing...

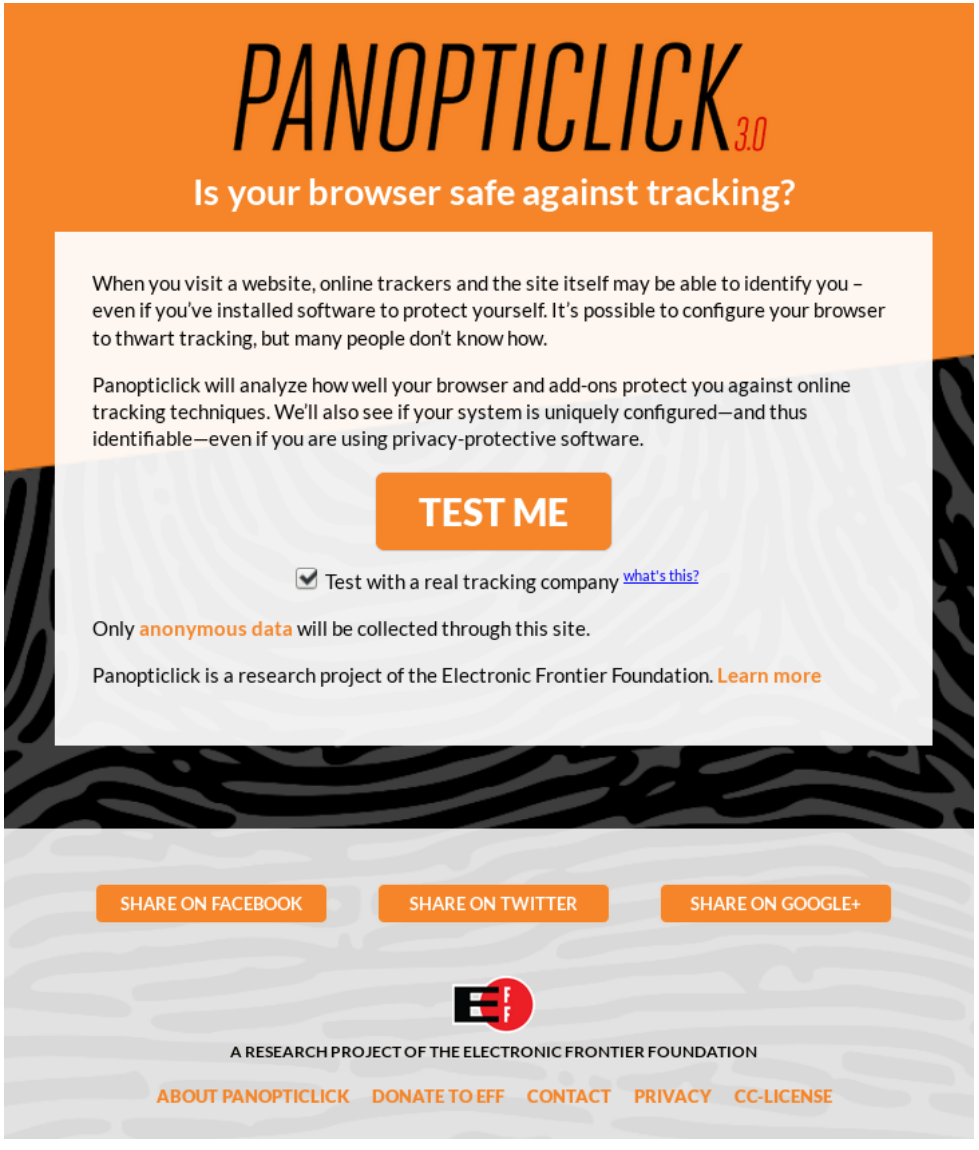
[7] Jason Franklin et al. "Passive Data Link Layer 802.11 Wireless Device Driver Fingerprinting.". In: *USENIX Security*. 2006.

[8] Jeffrey Pang et al. "802.11 user fingerprinting". In: *MobiCom*. 2007.

[9] Christoph Neumann, Olivier Heen, and Stéphane Onno. "An empirical study of passive 802.11 device fingerprinting". In: *ICDCSW*. IEEE. 2012.

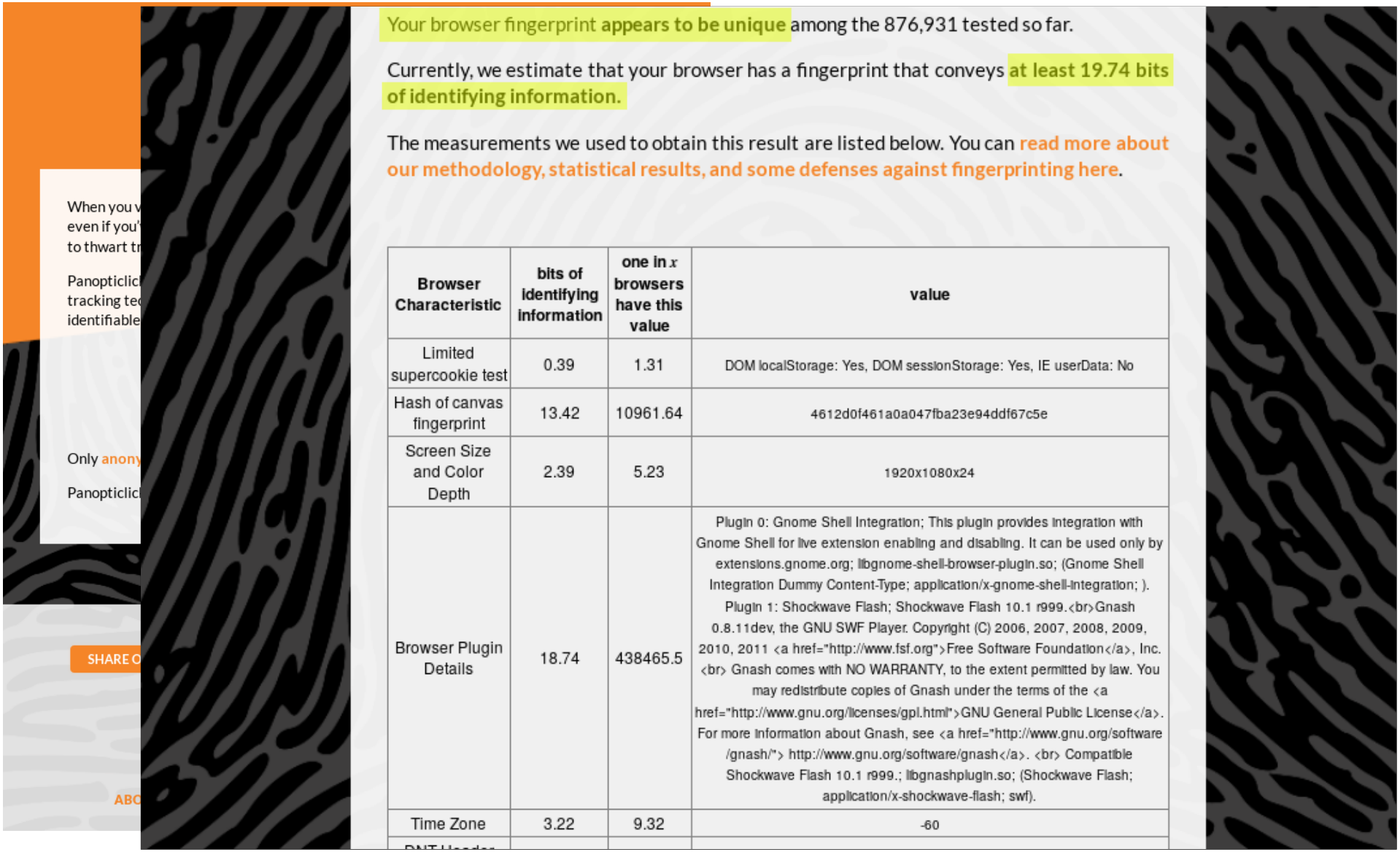
IE - Introduction - Inspiration

Panopticlick: browser fingerprinting: `https://panopticlick.eff.org/`



IE - Introduction - Inspiration

Panopticlick: browser fingerprinting: `https://panopticlick.eff.org/`



Your browser fingerprint **appears to be unique** among the 876,931 tested so far.

Currently, we estimate that your browser has a fingerprint that conveys **at least 19.74 bits of identifying information.**

The measurements we used to obtain this result are listed below. You can **read more about our methodology, statistical results, and some defenses against fingerprinting here**.

Browser Characteristic	bits of identifying information	one in x browsers have this value	value
Limited supercookie test	0.39	1.31	DOM localStorage: Yes, DOM sessionStorage: Yes, IE userData: No
Hash of canvas fingerprint	13.42	10961.64	4612d0f461a0a047fba23e94ddf67c5e
Screen Size and Color Depth	2.39	5.23	1920x1080x24
Browser Plugin Details	18.74	438465.5	Plugin 0: Gnome Shell Integration; This plugin provides integration with Gnome Shell for live extension enabling and disabling. It can be used only by extensions.gnome.org; libgnome-shell-browser-plugin.so; (Gnome Shell Integration Dummy Content-Type; application/x-gnome-shell-integration;). Plugin 1: Shockwave Flash; Shockwave Flash 10.1 r999. Gnash 0.8.11dev, the GNU SWF Player. Copyright (C) 2006, 2007, 2008, 2009, 2010, 2011 Free Software Foundation, Inc. Gnash comes with NO WARRANTY, to the extent permitted by law. You may redistribute copies of Gnash under the terms of the GNU General Public License. For more information about Gnash, see http://www.gnu.org/software/gnash. Compatible Shockwave Flash 10.1 r999.; libgnashplugin.so; (Shockwave Flash; application/x-shockwave-flash; swf).
Time Zone	3.22	9.32	-60

IE - Methodology

- Empirical evaluation using datasets:

Dataset	Lab	Train station	Sapienza
#MAC addr.	500	10 000	160 000
#Probe Req.	120 000	110 000	8 million
Time frame	Oct '15	Oct/Nov '15	Feb/May '13
Location	Lab	Train Station	Rome

- Considered metrics
  - Fraction of affected devices
  - Stability over time
  - Entropy: amount of identifying information

$$H_i = -\sum_{j \in E_i} f_{i,j} * \log_2 f_{i,j}$$

where $f_{i,j}$ is the frequency of the value j for the element i in the dataset

- n bits of entropy means one can identify 1 device among 2^n on average

- Compute entropy of:
  - individual IEs
  - global fingerprint (using most popular IEs)

IE - Fingerprinting using Information Elements: Results

Element	Entropy (bits)			Stability			Affected devices		
	Lab	Station	Sapienza	Lab	Station	Sapienza	Lab	Station	Sapienza
HT capabilities info	3.94	4.74	3.35	96.0%	95.9%	99.6%	90.9%	90.0%	81.1%
Ordered list of tags numbers	4.23	5.24	4.10	93.6%	94.2%	91.2%	100%	100%	100%
Extended capabilities	2.59	2.57	0.064	98.5%	99.4%	99.9%	55.4%	51.3%	0.6%
HT A-MPDU parameters	2.59	2.67	2.54	97.8%	99.1%	99.7%	90.9%	90.0%	81.1%
HT MCS set bitmask	1.49	1.43	1.16	97.6%	99.0%	99.9%	90.9%	90.0%	81.1%
Supported rates	1.18	2.10	1.36	98.2%	95.9%	99.8%	100%	99.9%	100%
Interworking - access net. type	1.08	1.11	0.006	99.6%	99.6%	100.0%	47.5%	46.1%	0.04%
Extended supported rates	1.00	1.77	0.886	98.0%	96.3%	99.4%	99.1%	72.6%	99.7%
WPS UUID	0.878	0.788	0.658	98.2%	99.2%	99.6%	8.4%	5.5%	3.6%
HT extended capabilities	0.654	0.623	0.779	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT TxBeam Forming Cap.	0.598	0.587	0.712	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT Antenna Selection Cap.	0.579	0.576	0.711	98.0%	98.9%	99.9%	90.9%	90.0%	81.1%
Overall	5.48	7.03	5.65	92.5%	90.7%	88.8%	-	-	-

- Individual Information Element

IE - Fingerprinting using Information Elements: Results

Element	Entropy (bits)			Stability			Affected devices		
	Lab	Station	Sapienza	Lab	Station	Sapienza	Lab	Station	Sapienza
HT capabilities info	3.94	4.74	3.35	96.0%	95.9%	99.6%	90.9%	90.0%	81.1%
Ordered list of tags numbers	4.23	5.24	4.10	93.6%	94.2%	91.2%	100%	100%	100%
Extended capabilities	2.59	2.57	0.064	98.5%	99.4%	99.9%	55.4%	51.3%	0.6%
HT A-MPDU parameters	2.59	2.67	2.54	97.8%	99.1%	99.7%	90.9%	90.0%	81.1%
HT MCS set bitmask	1.49	1.43	1.16	97.6%	99.0%	99.9%	90.9%	90.0%	81.1%
Supported rates	1.18	2.10	1.36	98.2%	95.9%	99.8%	100%	99.9%	100%
Interworking - access net. type	1.08	1.11	0.006	99.6%	99.6%	100.0%	47.5%	46.1%	0.04%
Extended supported rates	1.00	1.77	0.886	98.0%	96.3%	99.4%	99.1%	72.6%	99.7%
WPS UUID	0.878	0.788	0.658	98.2%	99.2%	99.6%	8.4%	5.5%	3.6%
HT extended capabilities	0.654	0.623	0.779	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT TxBeam Forming Cap.	0.598	0.587	0.712	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT Antenna Selection Cap.	0.579	0.576	0.711	98.0%	98.9%	99.9%	90.9%	90.0%	81.1%
Overall	5.48	7.03	5.65	92.5%	90.7%	88.8%	-	-	-

- Individual Information Element
  - Can provide up to 5.24 bits of entropy

IE - Fingerprinting using Information Elements: Results

Element	Entropy (bits)			Stability			Affected devices		
	Lab	Station	Sapienza	Lab	Station	Sapienza	Lab	Station	Sapienza
HT capabilities info	3.94	4.74	3.35	96.0%	95.9%	99.6%	90.9%	90.0%	81.1%
Ordered list of tags numbers	4.23	5.24	4.10	93.6%	94.2%	91.2%	100%	100%	100%
Extended capabilities	2.59	2.57	0.064	98.5%	99.4%	99.9%	55.4%	51.3%	0.6%
HT A-MPDU parameters	2.59	2.67	2.54	97.8%	99.1%	99.7%	90.9%	90.0%	81.1%
HT MCS set bitmask	1.49	1.43	1.16	97.6%	99.0%	99.9%	90.9%	90.0%	81.1%
Supported rates	1.18	2.10	1.36	98.2%	95.9%	99.8%	100%	99.9%	100%
Interworking - access net. type	1.08	1.11	0.006	99.6%	99.6%	100.0%	47.5%	46.1%	0.04%
Extended supported rates	1.00	1.77	0.886	98.0%	96.3%	99.4%	99.1%	72.6%	99.7%
WPS UUID	0.878	0.788	0.658	98.2%	99.2%	99.6%	8.4%	5.5%	3.6%
HT extended capabilities	0.654	0.623	0.779	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT TxBeam Forming Cap.	0.598	0.587	0.712	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT Antenna Selection Cap.	0.579	0.576	0.711	98.0%	98.9%	99.9%	90.9%	90.0%	81.1%
Overall	5.48	7.03	5.65	92.5%	90.7%	88.8%	-	-	-

- Individual Information Element
  - Can provide up to 5.24 bits of entropy
  - Stable for more than 95% of the devices (no change over time)

IE - Fingerprinting using Information Elements: Results

Element	Entropy (bits)			Stability			Affected devices		
	Lab	Station	Sapienza	Lab	Station	Sapienza	Lab	Station	Sapienza
HT capabilities info	3.94	4.74	3.35	96.0%	95.9%	99.6%	90.9%	90.0%	81.1%
Ordered list of tags numbers	4.23	5.24	4.10	93.6%	94.2%	91.2%	100%	100%	100%
Extended capabilities	2.59	2.57	0.064	98.5%	99.4%	99.9%	55.4%	51.3%	0.6%
HT A-MPDU parameters	2.59	2.67	2.54	97.8%	99.1%	99.7%	90.9%	90.0%	81.1%
HT MCS set bitmask	1.49	1.43	1.16	97.6%	99.0%	99.9%	90.9%	90.0%	81.1%
Supported rates	1.18	2.10	1.36	98.2%	95.9%	99.8%	100%	99.9%	100%
Interworking - access net. type	1.08	1.11	0.006	99.6%	99.6%	100.0%	47.5%	46.1%	0.04%
Extended supported rates	1.00	1.77	0.886	98.0%	96.3%	99.4%	99.1%	72.6%	99.7%
WPS UUID	0.878	0.788	0.658	98.2%	99.2%	99.6%	8.4%	5.5%	3.6%
HT extended capabilities	0.654	0.623	0.779	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT TxBeam Forming Cap.	0.598	0.587	0.712	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT Antenna Selection Cap.	0.579	0.576	0.711	98.0%	98.9%	99.9%	90.9%	90.0%	81.1%
Overall	5.48	7.03	5.65	92.5%	90.7%	88.8%	-	-	-

- Individual Information Element
  - Can provide up to 5.24 bits of entropy
  - Stable for more than 95% of the devices (no change over time)
  - Some IE are found in almost all devices (`Supported rates`)

IE - Fingerprinting using Information Elements: Results

Element	Entropy (bits)			Stability			Affected devices		
	Lab	Station	Sapienza	Lab	Station	Sapienza	Lab	Station	Sapienza
HT capabilities info	3.94	4.74	3.35	96.0%	95.9%	99.6%	90.9%	90.0%	81.1%
Ordered list of tags numbers	4.23	5.24	4.10	93.6%	94.2%	91.2%	100%	100%	100%
Extended capabilities	2.59	2.57	0.064	98.5%	99.4%	99.9%	55.4%	51.3%	0.6%
HT A-MPDU parameters	2.59	2.67	2.54	97.8%	99.1%	99.7%	90.9%	90.0%	81.1%
HT MCS set bitmask	1.49	1.43	1.16	97.6%	99.0%	99.9%	90.9%	90.0%	81.1%
Supported rates	1.18	2.10	1.36	98.2%	95.9%	99.8%	100%	99.9%	100%
Interworking - access net. type	1.08	1.11	0.006	99.6%	99.6%	100.0%	47.5%	46.1%	0.04%
Extended supported rates	1.00	1.77	0.886	98.0%	96.3%	99.4%	99.1%	72.6%	99.7%
WPS UUID	0.878	0.788	0.658	98.2%	99.2%	99.6%	8.4%	5.5%	3.6%
HT extended capabilities	0.654	0.623	0.779	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT TxBeam Forming Cap.	0.598	0.587	0.712	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT Antenna Selection Cap.	0.579	0.576	0.711	98.0%	98.9%	99.9%	90.9%	90.0%	81.1%
Overall	5.48	7.03	5.65	92.5%	90.7%	88.8%	-	-	-

- Individual Information Element
  - Can provide up to 5.24 bits of entropy
  - Stable for more than 95% of the devices (no change over time)
  - Some IE are found in almost all devices (`Supported rates`)
  - Ex. `HT capabilities info` (`Train station` dataset) : 4.74 bits of entropy, stable for 95.9% devices, 90% of devices affected

IE - Fingerprinting using Information Elements: Results

Element	Entropy (bits)			Stability			Affected devices		
	Lab	Station	Sapienza	Lab	Station	Sapienza	Lab	Station	Sapienza
HT capabilities info	3.94	4.74	3.35	96.0%	95.9%	99.6%	90.9%	90.0%	81.1%
Ordered list of tags numbers	4.23	5.24	4.10	93.6%	94.2%	91.2%	100%	100%	100%
Extended capabilities	2.59	2.57	0.064	98.5%	99.4%	99.9%	55.4%	51.3%	0.6%
HT A-MPDU parameters	2.59	2.67	2.54	97.8%	99.1%	99.7%	90.9%	90.0%	81.1%
HT MCS set bitmask	1.49	1.43	1.16	97.6%	99.0%	99.9%	90.9%	90.0%	81.1%
Supported rates	1.18	2.10	1.36	98.2%	95.9%	99.8%	100%	99.9%	100%
Interworking - access net. type	1.08	1.11	0.006	99.6%	99.6%	100.0%	47.5%	46.1%	0.04%
Extended supported rates	1.00	1.77	0.886	98.0%	96.3%	99.4%	99.1%	72.6%	99.7%
WPS UUID	0.878	0.788	0.658	98.2%	99.2%	99.6%	8.4%	5.5%	3.6%
HT extended capabilities	0.654	0.623	0.779	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT TxBeam Forming Cap.	0.598	0.587	0.712	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT Antenna Selection Cap.	0.579	0.576	0.711	98.0%	98.9%	99.9%	90.9%	90.0%	81.1%
Overall	5.48	7.03	5.65	92.5%	90.7%	88.8%	-	-	-

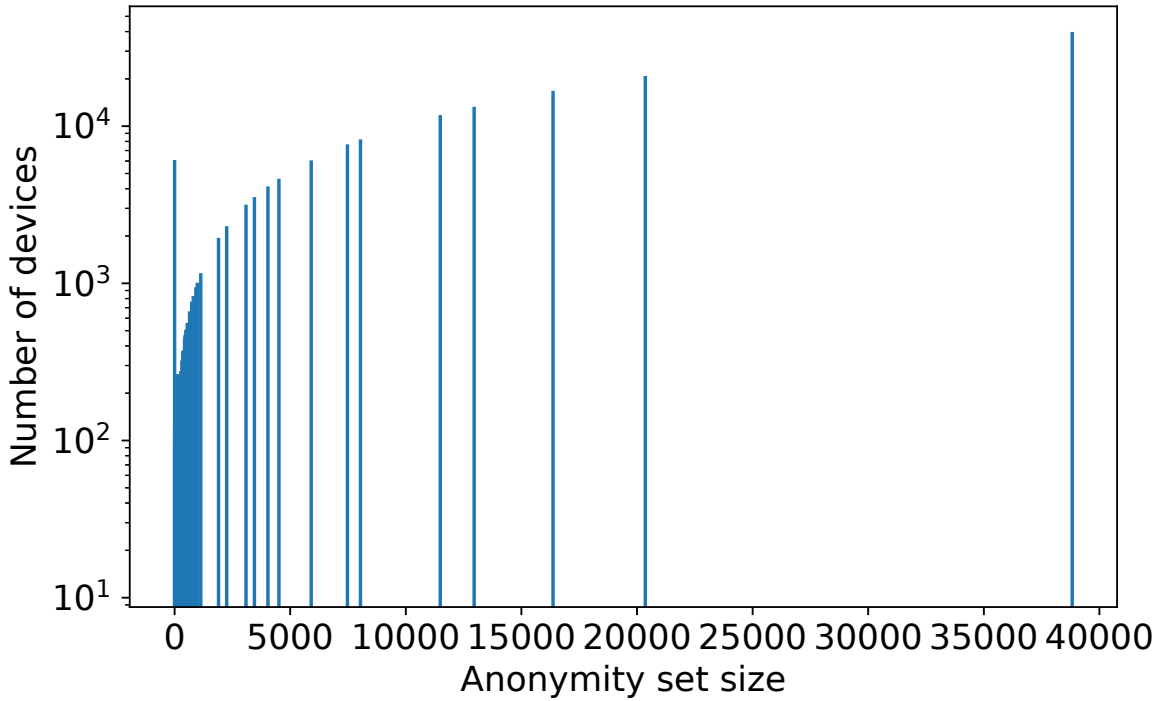
- Individual Information Element
  - Can provide up to 5.24 bits of entropy
  - Stable for more than 95% of the devices (no change over time)
  - Some IE are found in almost all devices (`Supported rates`)
  - Ex. `HT capabilities info` (`Train station` dataset) : 4.74 bits of entropy, stable for 95.9% devices, 90% of devices affected
- Global fingerprint (most common IEs)

IE - Fingerprinting using Information Elements: Results

Element	Entropy (bits)			Stability			Affected devices		
	Lab	Station	Sapienza	Lab	Station	Sapienza	Lab	Station	Sapienza
HT capabilities info	3.94	4.74	3.35	96.0%	95.9%	99.6%	90.9%	90.0%	81.1%
Ordered list of tags numbers	4.23	5.24	4.10	93.6%	94.2%	91.2%	100%	100%	100%
Extended capabilities	2.59	2.57	0.064	98.5%	99.4%	99.9%	55.4%	51.3%	0.6%
HT A-MPDU parameters	2.59	2.67	2.54	97.8%	99.1%	99.7%	90.9%	90.0%	81.1%
HT MCS set bitmask	1.49	1.43	1.16	97.6%	99.0%	99.9%	90.9%	90.0%	81.1%
Supported rates	1.18	2.10	1.36	98.2%	95.9%	99.8%	100%	99.9%	100%
Interworking - access net. type	1.08	1.11	0.006	99.6%	99.6%	100.0%	47.5%	46.1%	0.04%
Extended supported rates	1.00	1.77	0.886	98.0%	96.3%	99.4%	99.1%	72.6%	99.7%
WPS UUID	0.878	0.788	0.658	98.2%	99.2%	99.6%	8.4%	5.5%	3.6%
HT extended capabilities	0.654	0.623	0.779	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT TxBeam Forming Cap.	0.598	0.587	0.712	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT Antenna Selection Cap.	0.579	0.576	0.711	98.0%	98.9%	99.9%	90.9%	90.0%	81.1%
Overall	5.48	7.03	5.65	92.5%	90.7%	88.8%	-	-	-

- Individual Information Element
  - Can provide up to 5.24 bits of entropy
  - Stable for more than 95% of the devices (no change over time)
  - Some IE are found in almost all devices (`Supported rates`)
  - Ex. `HT capabilities info` (`Train station` dataset) : 4.74 bits of entropy, stable for 95.9% devices, 90% of devices affected
- Global fingerprint (most common IEs)
  - Entropy : 7.03 bits (`Train station` dataset)
  - Enough to uniquely identify 1 device among 128 (on average)
  - Can be used to locally track an individual

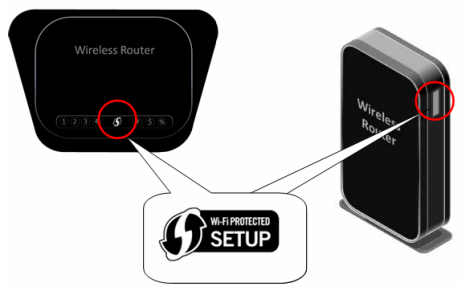
IE - Anonymity sets



- Anonymity sets: devices sharing the same IE fingerprint
- Sapienza dataset:
  - 6000 devices have a unique fingerprint
  - 120 devices share a fingerprint with another device
  - 38 000 devices share a common fingerprint

IE - Wi-Fi Protected Setup (WPS)

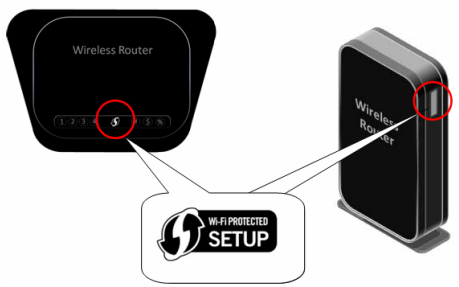
- Information element dedicated to WPS
  - Includes a UUID field
- Universally Unique Identifier UUID
  - A unique identifier *by definition*



[10] P. Leach, M. Mealling, and R. Salz. *A Universally Unique IDentifier (UUID) URN Namespace*. RFC 4122. Internet Engineering Task Force, July 2005.

IE - Wi-Fi Protected Setup (WPS)

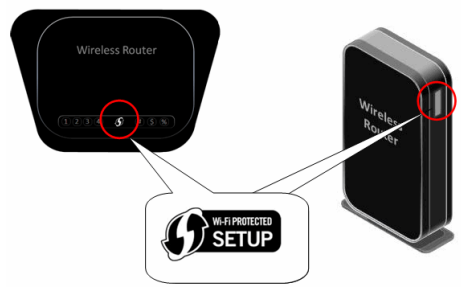
- Information element dedicated to WPS
  - Includes a UUID field
- Universally Unique Identifier UUID
  - A unique identifier *by definition*
  - Generally derived from the MAC address[10]



[10] P. Leach, M. Mealling, and R. Salz. *A Universally Unique IDentifier (UUID) URN Namespace*. RFC 4122. Internet Engineering Task Force, July 2005.

IE - Wi-Fi Protected Setup (WPS)

- Information element dedicated to WPS
  - Includes a UUID field
- Universally Unique Identifier UUID
  - A unique identifier *by definition*
  - Generally derived from the MAC address[10]
  - Could be reversed to reveal the original MAC



Input: MAC: MAC address of an interface
Returns: 16-byte WPS UUID

$salt \leftarrow$ `0x526480f8c99b4be5a65558ed5f5d6084`
$UUID \leftarrow \text{SHA-1}(MAC, salt)$
$UUID[6] \leftarrow (5 \ll 4) \mid (UUID[6]$ & `0x0f`$)$
$UUID[8] \leftarrow$ `0x80` $\mid (UUID[8]$ & `0x3f`$)$

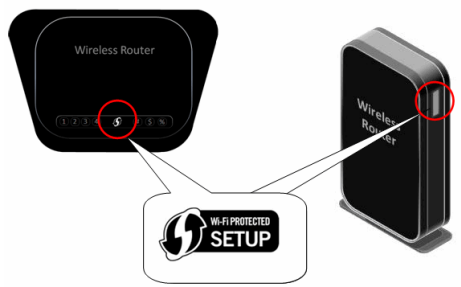
return $UUID[:16]$

---

[10] P. Leach, M. Mealling, and R. Salz. *A Universally Unique IDentifier (UUID) URN Namespace*. RFC 4122. Internet Engineering Task Force, July 2005.

IE - Wi-Fi Protected Setup (WPS)

- Information element dedicated to WPS
  - Includes a UUID field
- Universally Unique Identifier UUID
  - A unique identifier *by definition*
  - Generally derived from the MAC address[10]
  - Could be reversed to reveal the original MAC
- Re-identification attack on the datasets
  - UUID derived from the real Wi-Fi MAC address in 95% of the cases, in a way that can be reversed



Input: MAC: MAC address of an interface
Returns: 16-byte WPS UUID

$salt \leftarrow$ `0x526480f8c99b4be5a65558ed5f5d6084`
$UUID \leftarrow \text{SHA-1}(MAC, salt)$
$UUID[6] \leftarrow (5 \ll 4) \mid (UUID[6]$ & `0x0f`$)$
$UUID[8] \leftarrow$ `0x80` $\mid (UUID[8]$ & `0x3f`$)$

return $UUID[:16]$

---

[10] P. Leach, M. Mealling, and R. Salz. *A Universally Unique IDentifier (UUID) URN Namespace*. RFC 4122. Internet Engineering Task Force, July 2005.

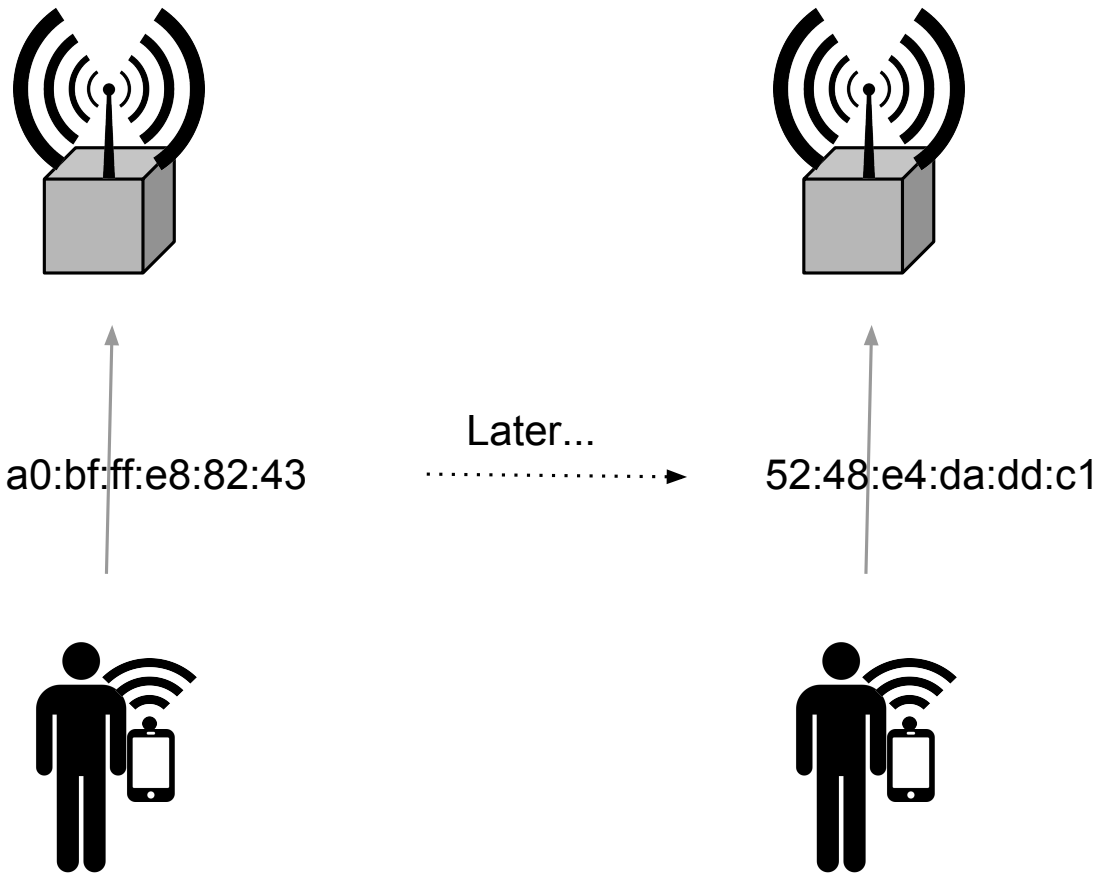
IE - Recommendations

- Problem: Information Elements can be leveraged to defeat MAC address randomization
- They are not needed in probe requests before association
- Our recommendation: Remove them or restrict to a bare minimum
- Solution adopted in Android Oreo based on our work[11]:

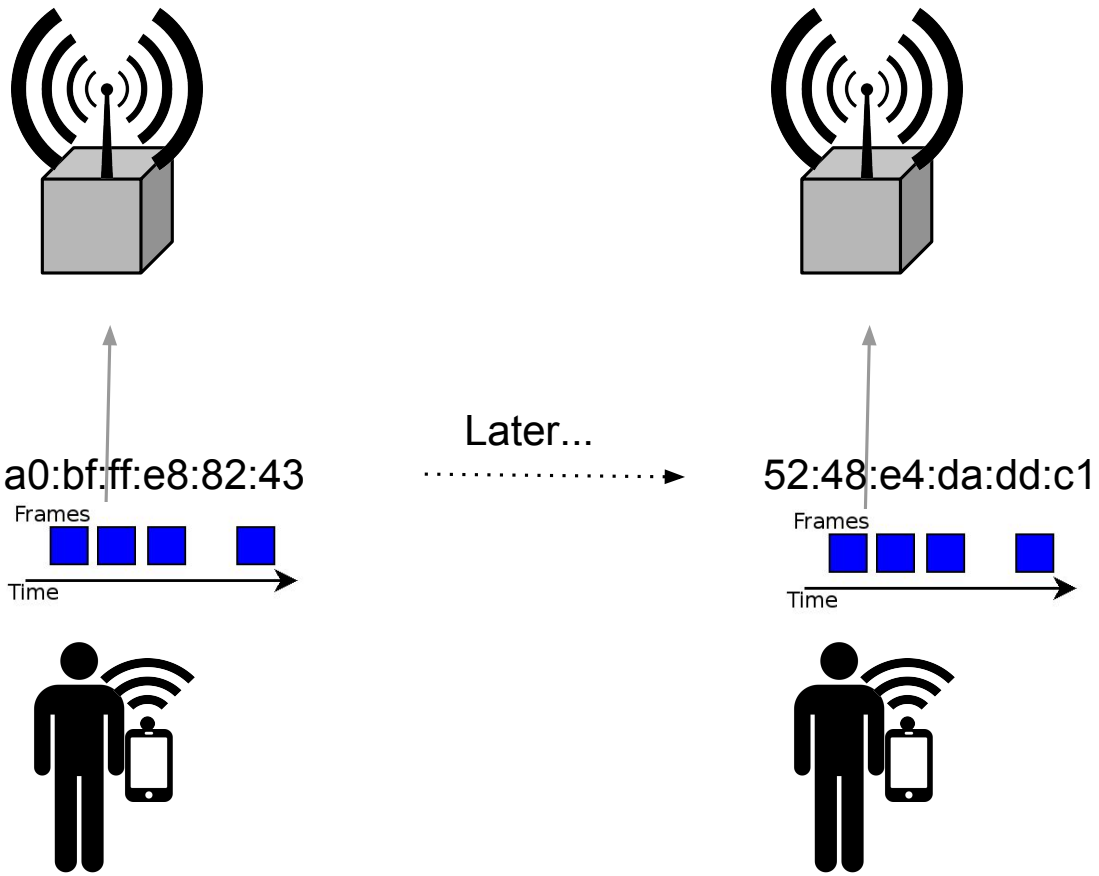
> - Unnecessary Probe Request Information Elements have been removed: Information Elements are limited to the SSID and DS parameter sets.

[11] Giles Hogben. *Changes to Device Identifiers in Android O*. https://android-developers.googleblog.com/2017/04/changes-to-device-identifiers-in.html. 2017.

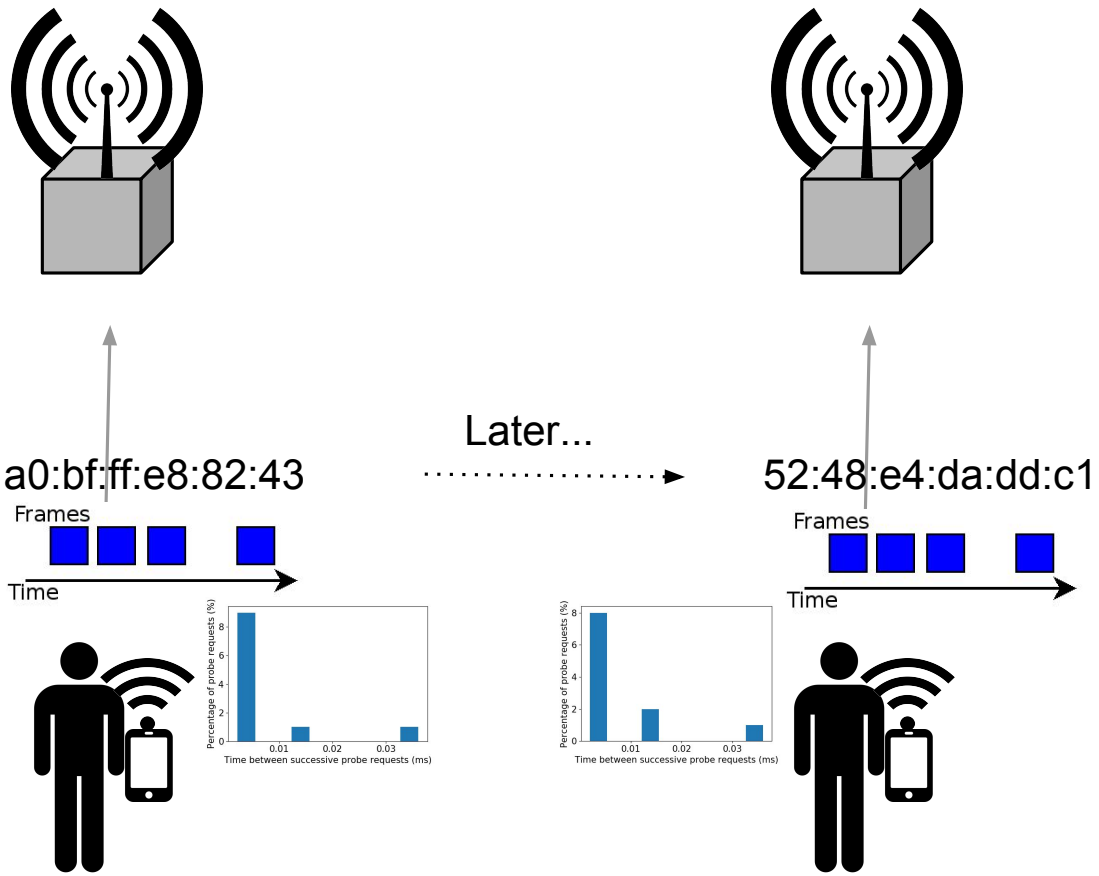
Timing - Introduction



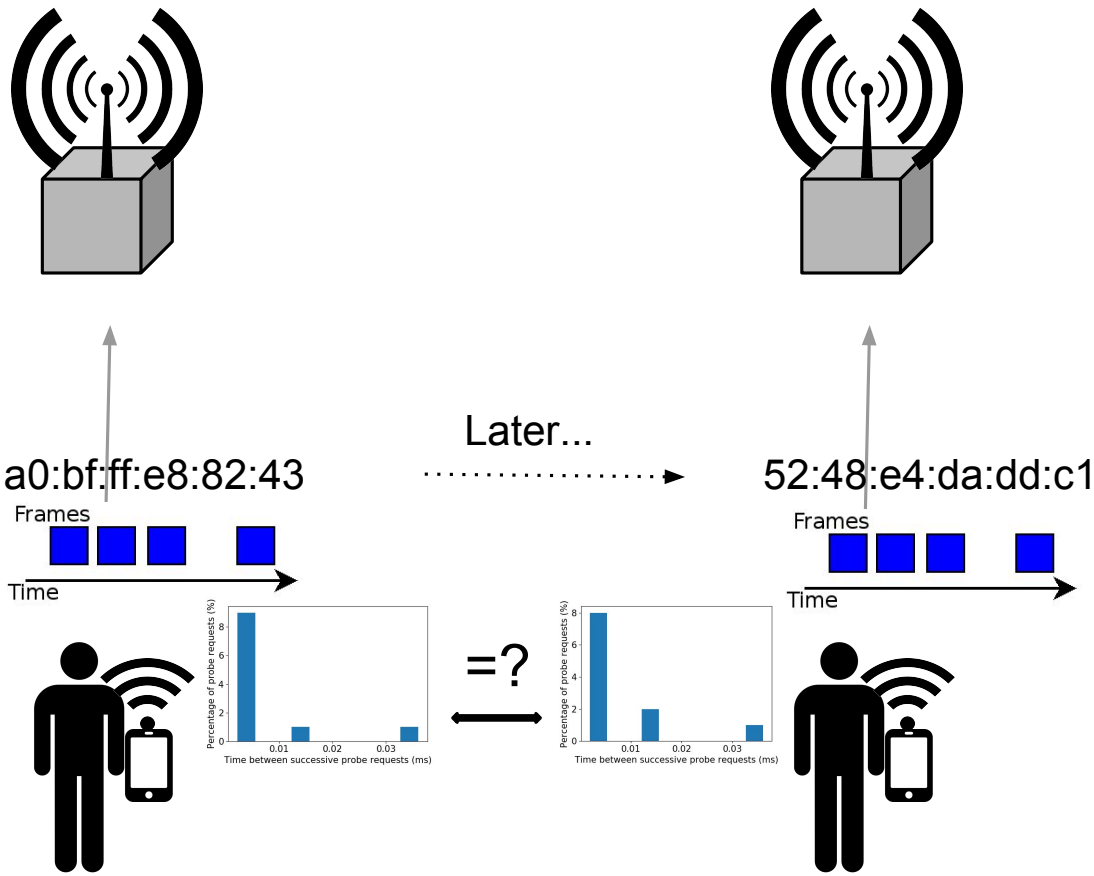
Timing - Introduction



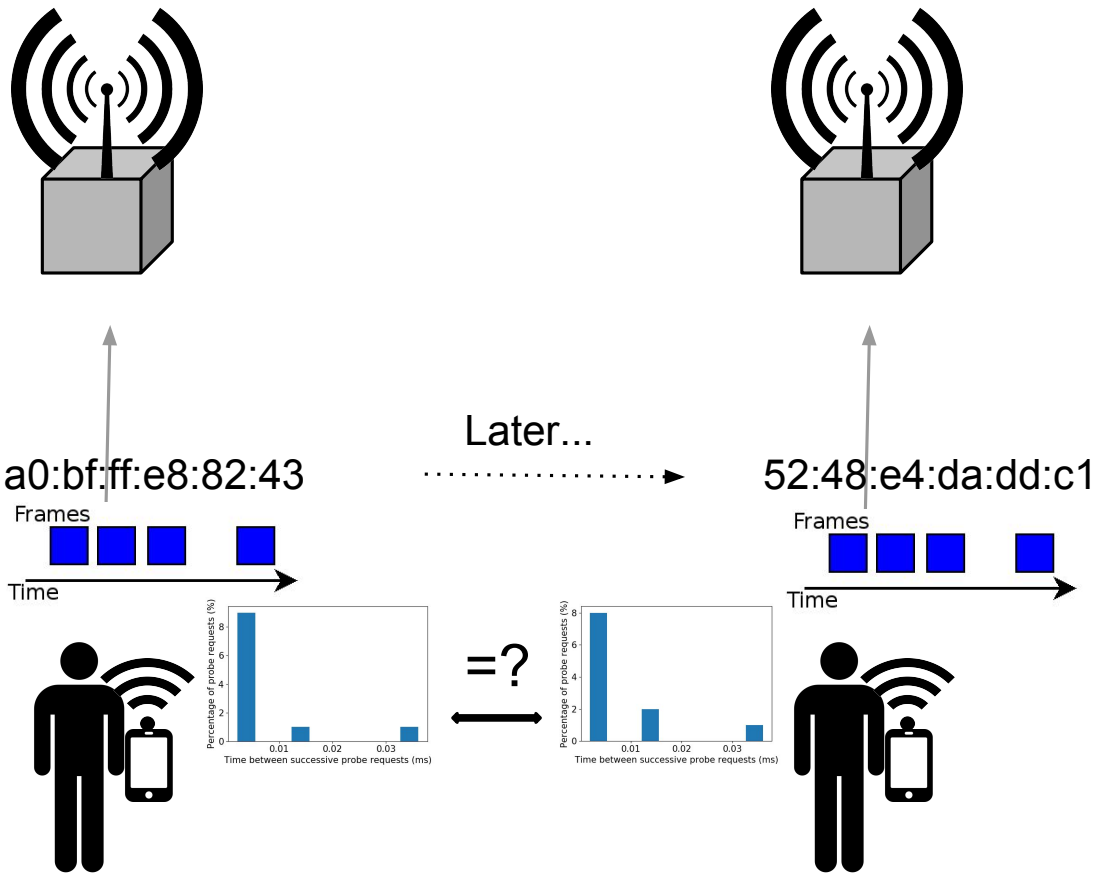
Timing - Introduction



Timing - Introduction



Timing - Introduction



- Idea: defeat randomization's purpose using timing of probe requests
- A given random address is used for a whole burst of probe requests
- More advanced than state of the art because we have a single burst

Timing - Methodology 1/2

- Hypothesis (H_0): "Given two bursts[12] A and B, A and B come from the same device"
  - TP = "Classifier estimates correctly that A and B come from the same device" $\rightarrow$ TPR = rate of TP
  - TN = "Classifier estimates correctly that A and B come from different devices"
  - Overall Success Rate: $OSR = \frac{TPR+TNR}{2}$
- Datasets of different sizes (from 5 minutes to 10 days)

[12] Burst = group of frames emitted within less than 100ms

Timing - Methodology 2/2

- Features
  - Inter-frame arrival time, burst length, number of frames
  - Create signatures of groups of frames
- Metrics
  - Euclidian distance: $D_E = \sqrt{\sum_{n}^{i=1} (v_i - w_i)^2}$
  - Cosine distance: $D_c = \frac{\sqrt{\sum_{n}^{i=1} v_i w_i}}{\sqrt{\sum_{n}^{i=1} v_i^2} \sqrt{\sum_{n}^{i=1} w_i^2}}$
- Classifiers: Give signatures to different classifiers
  - Unsupervised: DBSCAN, k-means, Mean shift
  - Supervised: random forests
  - Custom: incremental algorithm: considers frames as a stream

Timing - Results

Dataset TPR/TNR	DBSCAN	k-means	Mean shift	Rand Forest	Inc. alg.	Rand Forest Cross-check
Lab	0.78/0.64	0.11/1.00	0.88/0.12	0.40/0.76	0.46/0.94	0.42/0.75
Lab cut	0.43/0.94	0.18/0.97	0.87/0.08	0.65/0.60	0.30/0.96	0.62/0.45
Lab cut2	0.42/0.96	0.29/0.93	0.88/0.18	0.88/0.29	0.26/0.95	0.89/0.33
Belgium				0.32/0.75	0.45/0.96	0.31/0.77
Belgium cut	0.29/0.88	0.10/0.98	0.82/0.18	0.49/0.68	0.30/0.92	0.58/0.59
Belgium cut2	0.36/0.93	0.14/0.97	0.88/0.15	0.80/0.27	0.35/0.93	0.98/0.06
Avg. OSR	0.66	0.57	0.50	0.57	0.65	0.56

Timing - Results

Dataset TPR/TNR	DBSCAN	k-means	Mean shift	Rand Forest	Inc. alg.	Rand Forest Cross-check
Lab	0.78/0.64	0.11/1.00	0.88/0.12	0.40/0.76	0.46/0.94	0.42/0.75
Lab cut	0.43/0.94	0.18/0.97	0.87/0.08	0.65/0.60	0.30/0.96	0.62/0.45
Lab cut2	0.42/0.96	0.29/0.93	0.88/0.18	0.88/0.29	0.26/0.95	0.89/0.33
Belgium				0.32/0.75	0.45/0.96	0.31/0.77
Belgium cut	0.29/0.88	0.10/0.98	0.82/0.18	0.49/0.68	0.30/0.92	0.58/0.59
Belgium cut2	0.36/0.93	0.14/0.97	0.88/0.15	0.80/0.27	0.35/0.93	0.98/0.06
Avg. OSR	0.66	0.57	0.50	0.57	0.65	0.56

- Up to 66% OSR

Timing - Results

Dataset TPR/TNR	DBSCAN	k-means	Mean shift	Rand Forest	Inc. alg.	Rand Forest Cross-check
Lab	0.78/0.64	0.11/1.00	0.88/0.12	0.40/0.76	0.46/0.94	0.42/0.75
Lab cut	0.43/0.94	0.18/0.97	0.87/0.08	0.65/0.60	0.30/0.96	0.62/0.45
Lab cut2	0.42/0.96	0.29/0.93	0.88/0.18	0.88/0.29	0.26/0.95	0.89/0.33
Belgium				0.32/0.75	0.45/0.96	0.31/0.77
Belgium cut	0.29/0.88	0.10/0.98	0.82/0.18	0.49/0.68	0.30/0.92	0.58/0.59
Belgium cut2	0.36/0.93	0.14/0.97	0.88/0.15	0.80/0.27	0.35/0.93	0.98/0.06
Avg. OSR	0.66	0.57	0.50	0.57	0.65	0.56

- Up to 66% OSR
- Low FPR in some cases

Timing - Results

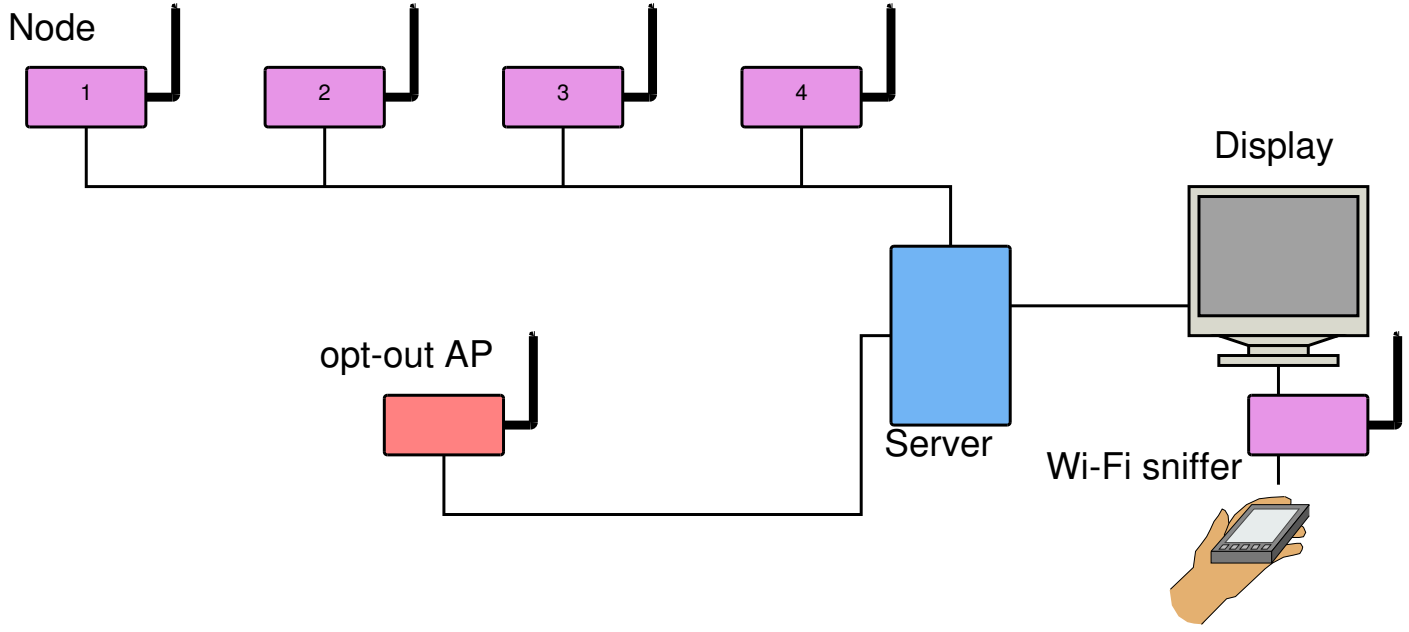
Dataset TPR/TNR	DBSCAN	k-means	Mean shift	Rand Forest	Inc. alg.	Rand Forest Cross-check
Lab	0.78/0.64	0.11/1.00	0.88/0.12	0.40/0.76	0.46/0.94	0.42/0.75
Lab cut	0.43/0.94	0.18/0.97	0.87/0.08	0.65/0.60	0.30/0.96	0.62/0.45
Lab cut2	0.42/0.96	0.29/0.93	0.88/0.18	0.88/0.29	0.26/0.95	0.89/0.33
Belgium				0.32/0.75	0.45/0.96	0.31/0.77
Belgium cut	0.29/0.88	0.10/0.98	0.82/0.18	0.49/0.68	0.30/0.92	0.58/0.59
Belgium cut2	0.36/0.93	0.14/0.97	0.88/0.15	0.80/0.27	0.35/0.93	0.98/0.06
Avg. OSR	0.66	0.57	0.50	0.57	0.65	0.56

- Up to 66% OSR
- Low FPR in some cases
- Slight difference between distances
- Shows that timing must be considered as well as a tool to defeat randomization
- Also interesting: DBSCAN can estimate number of devices for small datasets
- Recommendation: break timing patterns (random noise?)

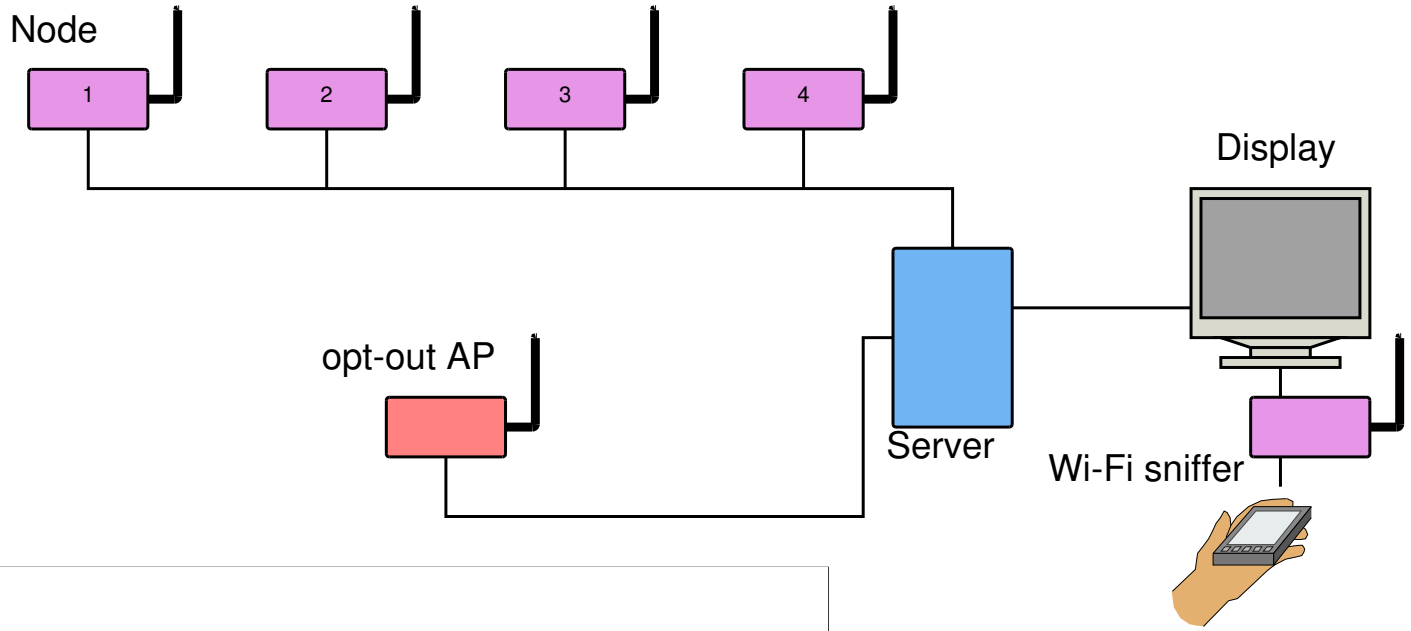
Wombat - Introduction

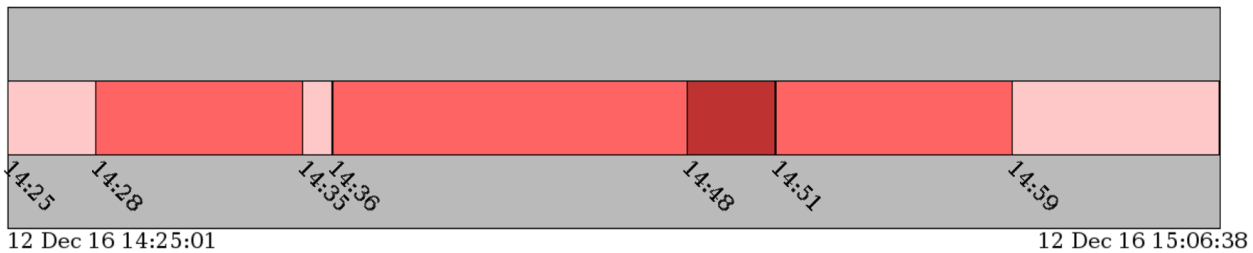
- We needed our own Wi-Fi tracking system to:
  - understand how these systems work
  - test new privacy-enhancing features
  - make demonstrations
- We built the Wombat Wi-Fi Tracking system

Wombat - Architecture and front-end



Wombat - Architecture and front-end





legend: noeud1 noeud3 noeud4

MAC address: 34:23:ba:df:90:ce
Vendor / Manufacturer: Samsung Electro Mechanics co.,LTD.
total number of frames: 2247
visit duration: 0h 41min 37sec
SSIDs: IKEA WiFi, WiFi-Cite-Espace, WiFI Hotel Les Skieurs, hotspot-cite-sciences, grenoble

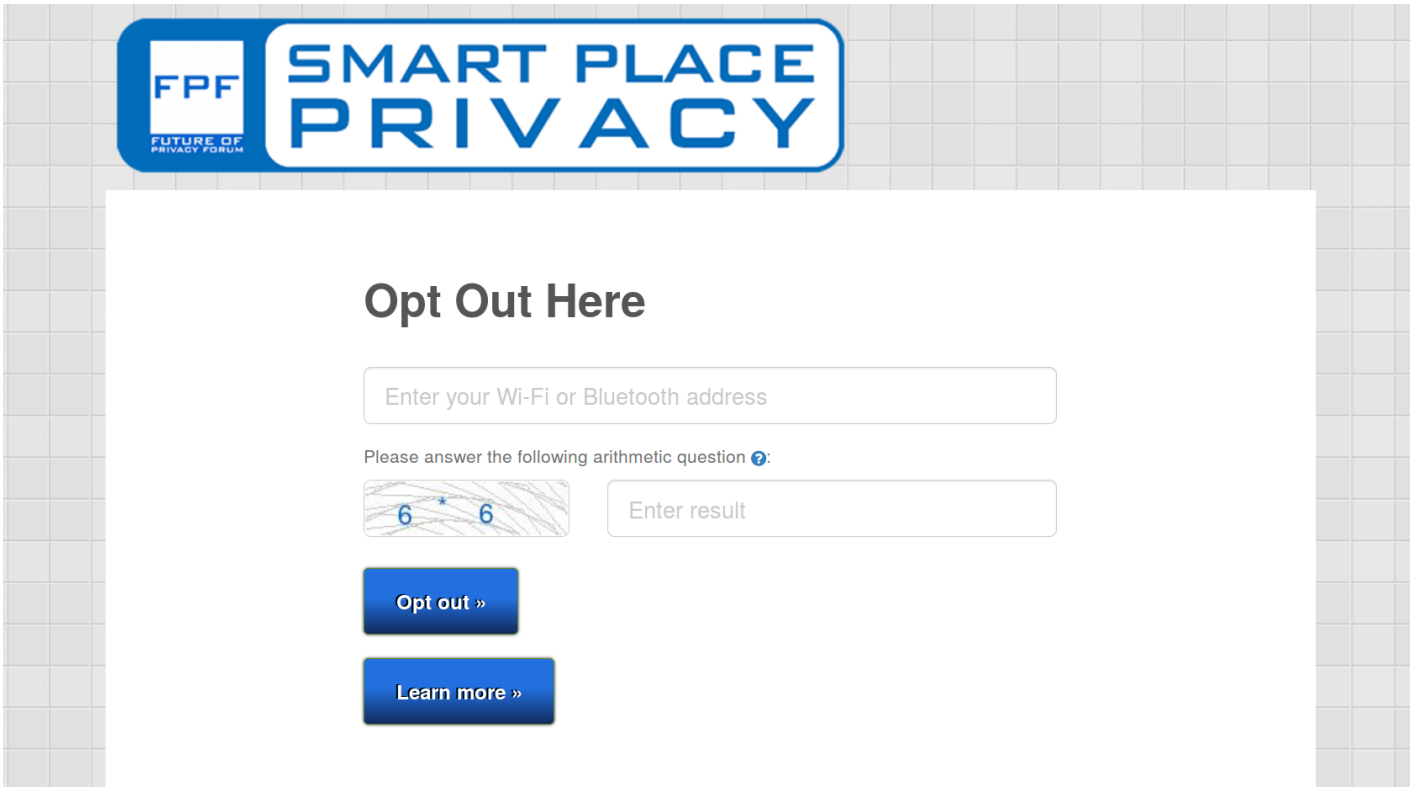
Wombat - Installations

Installed at:

- *Cité des Sciences et de l'Industrie* (Paris), Terra Data exhibition, 2017-04 - 2018-03
- Laboratoire d'Innovation Numérique de la CNIL (LINC)
- Exhibition room @CITI lab
- Used for Arte's TV show X:enius (2017-01)

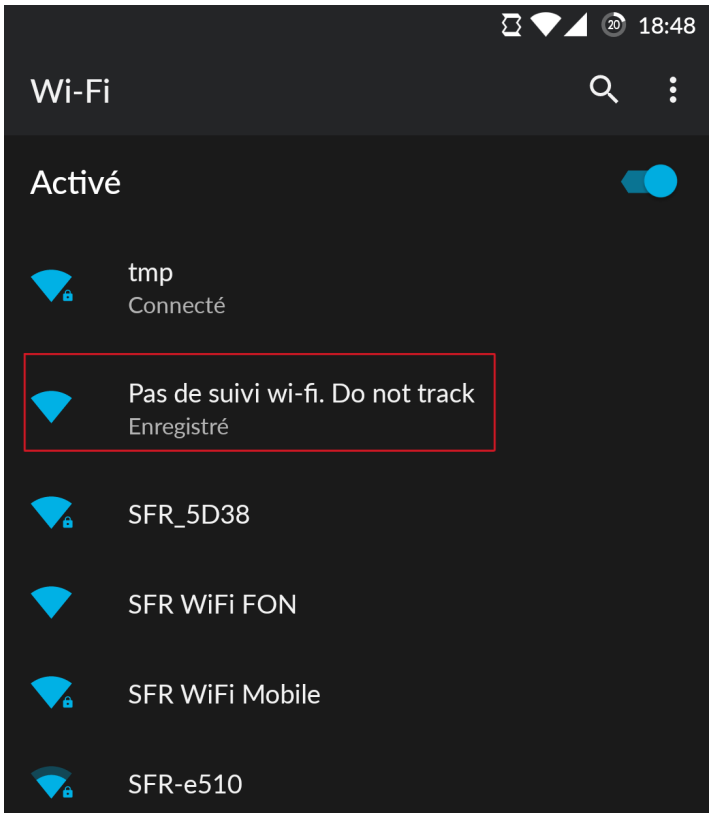
Wombat - Privacy-enhancing feature: Opt-out mechanism 1/2: Current implementations

- Opt-out: exiting the system
- Legislation imposes the possibility to opt out (FTC, CNIL)
- Current mechanisms are hardly useable by common users



Wombat - Privacy-enhancing feature: Opt-out mechanism 2/2: Our proposition

- Our proposition: Wi-Fi-based mechanism
- Using a non-functional AP using a recognizable network name (e.g. “Pas de suivi wi-fi. Do not track”)
- Advantages:
  - No software or hardware modification
  - Simple to use
  - Device will remember the association $\rightarrow$ user action needed only once
  - No memory of blacklisted devices
  - Global if standardized



Conclusion: Discussion

- Wi-Fi tracking: real-world problem, more than a technical issue
- Necessity to work both on the technical aspect and with different actors:
  - regulation entities (e.g. CNIL)
  - standardization bodies (e.g. IEEE)
- Necessity to inform the general public

Conclusion - Impact

- Industry:
  - Qualcomm, Broadcom: chipset bugs reported and acknowledged
  - Google: Changes in Android O to improve randomization in response to our work[13]
  - IEEE: Issues discussed in a plenary session of the 802 Privacy Study Group
- Popularization:
  - General public: Wombat installations:
    - *Cité des Sciences et de l'Industrie*: one-year-long installation
    - CNIL: permanent installation (being prepared)
  - TV:
    - Wombat deployed for a show on Arte
  - General press:
    - 2 articles published in general-public technical journals
    - Tech report about difficulty of disabling Wi-Fi in Android heavily relayed in the press (01.net, UFC-Que Choisir, Hacker News, etc.)

[13] Giles Hogben. *Changes to Device Identifiers in Android O*. https://android-developers.googleblog.com/2017/04/changes-to-device-identifiers-in.html. 2017.

Conclusion: publications

- **Peer-reviewed conferences**
  - Célestin Matte, Mathieu Cunche, Franck Rousseau, et al. "Defeating MAC Address Randomization Through Timing Attacks". In: *ACM WiSec*. 2016
  - Célestin Matte and Mathieu Cunche. "DEMO: Panoptiphone: How Unique is Your Wi-Fi Device?". In: *ACM WiSec*. 2016
  - Mathy Vanhoef et al. "Why MAC Address Randomization is not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms". In: *AsiaCCS*. 2016
  - Célestin Matte, Jagdish Prasad Achara, and Mathieu Cunche. "Device-to-identity linking attack using targeted Wi-Fi geolocation spoofing". In: *ACM WiSec*. 2015
- **Technical report**
  - Célestin Matte, Mathieu Cunche, and Vincent Toubiana. *Does disabling Wi-Fi prevent my Android phone from sending Wi-Fi frames?*. Research Report RR-9089. Inria Rhône-Alpes; INSA Lyon, Aug. 2017
- **Articles in general-public technical journals**
  - Célestin Matte and Mathieu Cunche. "Traçage Wi-Fi : applications et contre-mesures". In: *GNU/Linux Magazine France*. Surveillance: Tester les techniques pour mieux se défendre ! HS 84 (May 2016)
  - Célestin Matte. "Fingerprinting de smartphones : votre téléphone est-il traçable ?". In: *MISC - Multi-Systems & Internet Security Cookbook* 81 (Sept. 2015)

Conclusion: Guidelines

Guidelines for MAC address randomization (simplified)[14]:

1. MAC address changed in every burst of probe requests.
2. Probe requests devoid of unnecessary Information Elements.
3. In particular, SSIDs must always be null.
4. Sequence numbers must be randomized or fix.
5. Function generating the random addresses of cryptographic level.
6. Actual address never used for service discovery.
7. Randomize all bytes of the MAC address, while still following MAC address standards.
8. Break timing patterns, e.g., using random delays.

[14] Relayed to the IEEE 802 Privacy Study Group

Conclusion: Guidelines

Guidelines for MAC address randomization (simplified)[14]:

1. MAC address changed in every burst of probe requests.
2. Probe requests devoid of unnecessary Information Elements.
3. In particular, SSIDs must always be null.
4. Sequence numbers must be randomized or fix.
5. Function generating the random addresses of cryptographic level.
6. Actual address never used for service discovery.
7. Randomize all bytes of the MAC address, while still following MAC address standards.
8. Break timing patterns, e.g., using random delays.

[14] Relayed to the IEEE 802 Privacy Study Group

Conclusion: Summary

- MAC address randomization introduced as a countermeasure to Wi-Fi tracking
- No specifications
- Not sufficient because:
  - Content of probe requests frames can be used to form a fingerprint
    - Probe requests contain a lot of Information Elements
    - They bring over 7 bits of entropy
  - Timing of probe requests can be used as well
    - Create signatures of single bursts of probe requests
    - Classify frames with up to 66% accuracy
  - Current implementations (in 2016) possess many shortcomings
- $\rightarrow$ Many ways to defeat MAC address randomization exist

Conclusion: Future works

- Studying possible countermeasures to timing issue (e.g., random noise)
- Building evaluation procedure for implementations of MAC address randomization
- Defining a standard for data control (i.e. opt out procedure) for physical tracking systems

The end

Thanks for your attention!

Backup slide: MAC address format

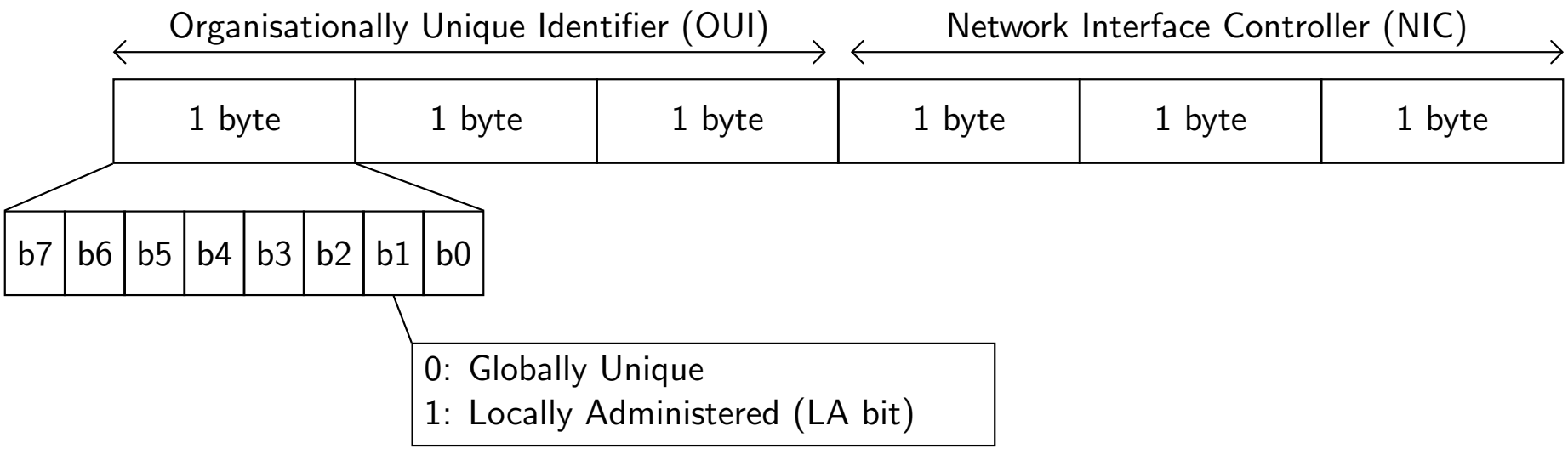


Figure: MAC address format.

Backup slide: Bursts

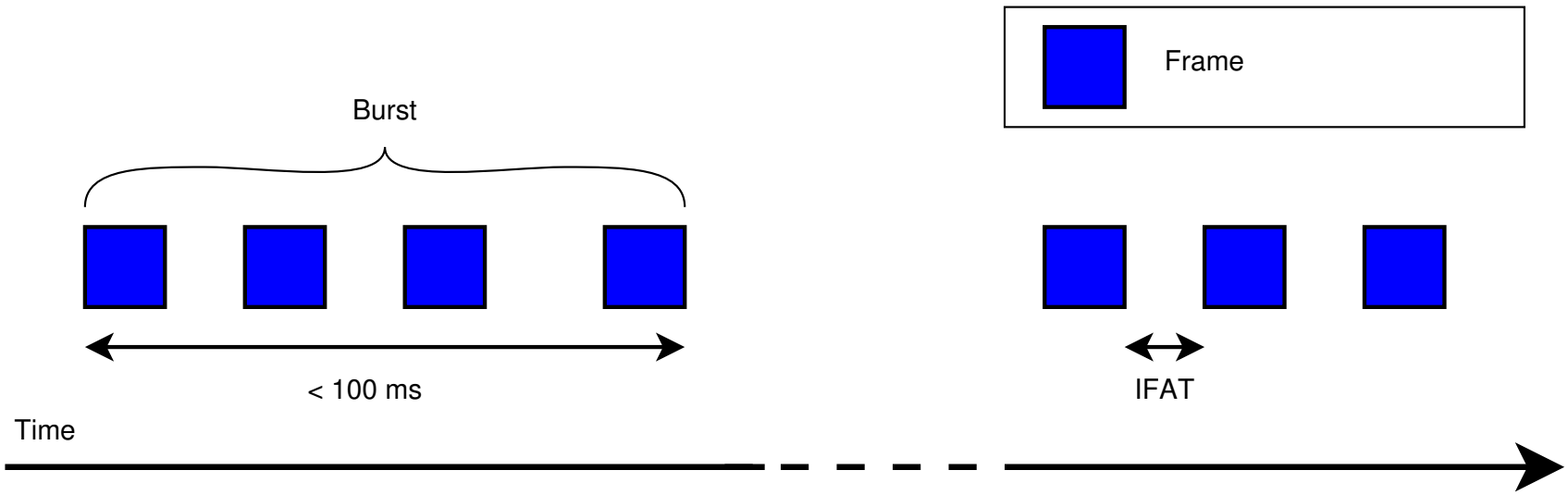


Figure: Transmission sequence of probe request frames with Inter-Frame Arrival Time (IFAT) within a burst, i.e. a group of frames sent by a device within a time window smaller than 100 ms.

Backup slide: MAC address format

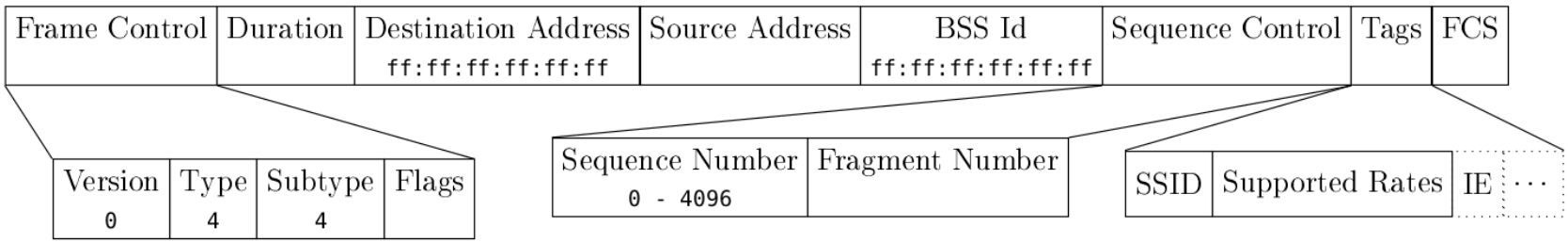


Figure: Probe request frame format.

Backup slide: Entropy

Entropy: Entropy is a measure to quantify the amount of information brought by an element (taking discrete values[15]) in a dataset. The entropy of an element i is computed as follows:

$$H_i = -\sum_{j \in E_i} f_{i,j} * \log_2 f_{i,j} \tag{1}$$

[15] When the element is a continuous variable, variance is used instead.

Backup slide: Used datasets

Name	Time	Place	Situation	MAC addr.	probe requests	Source
`Sapienza`	2013.02 - 2013.05	Rome	mix	160 000	8 000 000	[16]
`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	[17]
`Belgium`	2016.01 - 2016.02	Belgium	hotspot	3 700	200 000	same
`Martin`	2015.01 - 2016.12	Maryland	street	2 600 000	66 000 000	[18]
`Madeira`	2015.12 - 2017.06	Madeira	hotspot	13 000 000	300 000 000	not public

Figure: Used datasets

[16] Marco V. Barbera et al. *CRAWDAD dataset sapienza/probe-requests (v. 2013-09-10)*. Downloaded from http://crawdad.org/sapienza/probe-requests/20130910. Sept. 2013.

[17] Pieter Robyns et al. "Noncooperative 802.11 MAC Layer Fingerprinting and Tracking of Mobile Devices". In: *Security and Communication Networks* (2017).

[18] Jeremy Martin et al. "A Study of MAC Address Randomization in Mobile Devices and When it Fails". In: *arXiv preprint* (2017).

Backup slide: Vendors in datasets

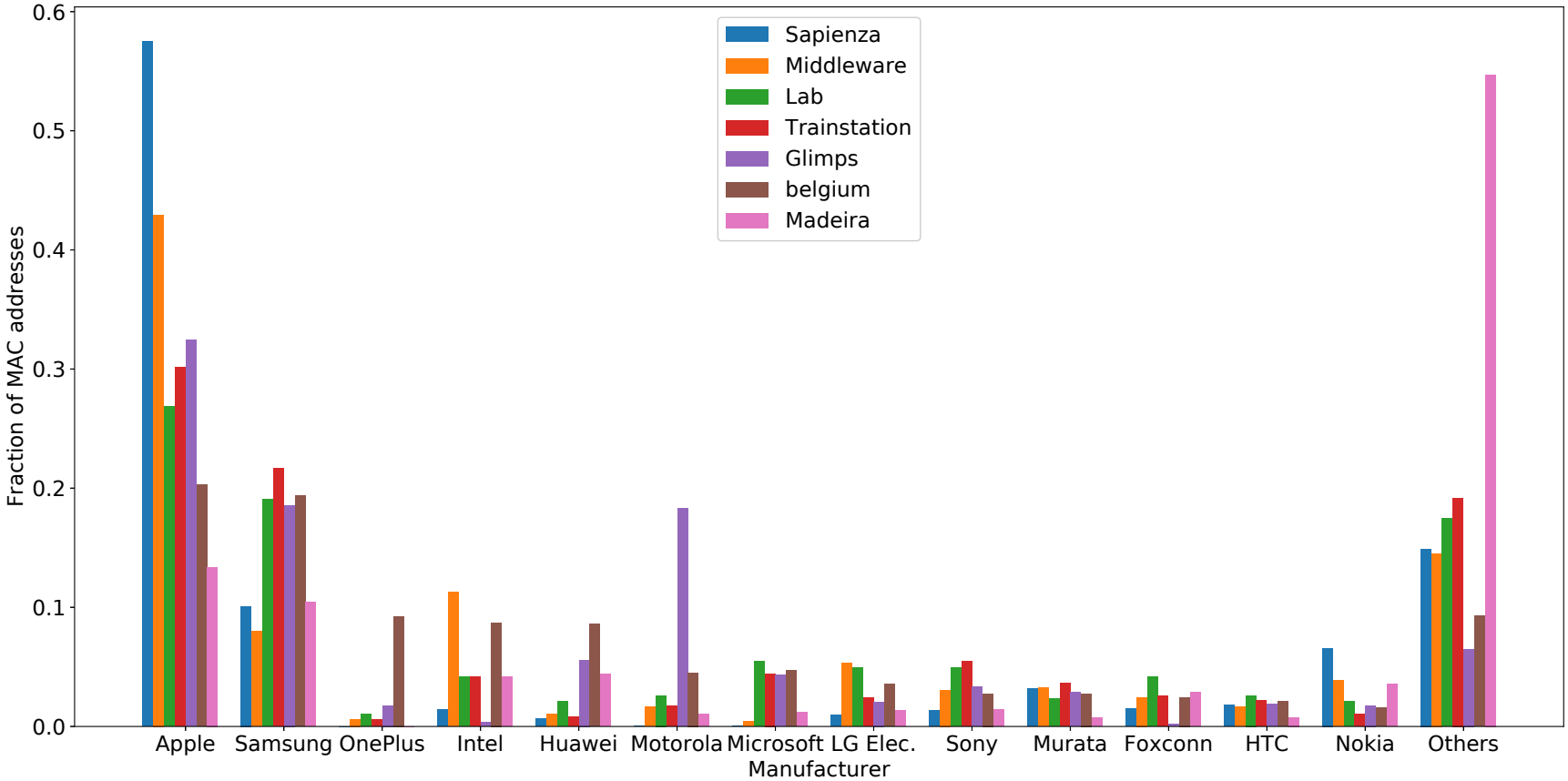


Figure: Fraction of non-random MAC addresses belonging to most-spread manufacturers. Represented manufacturers are those for which at least one dataset had 3% of its MAC addresses belonging to it.

Backup slide: Entropy table

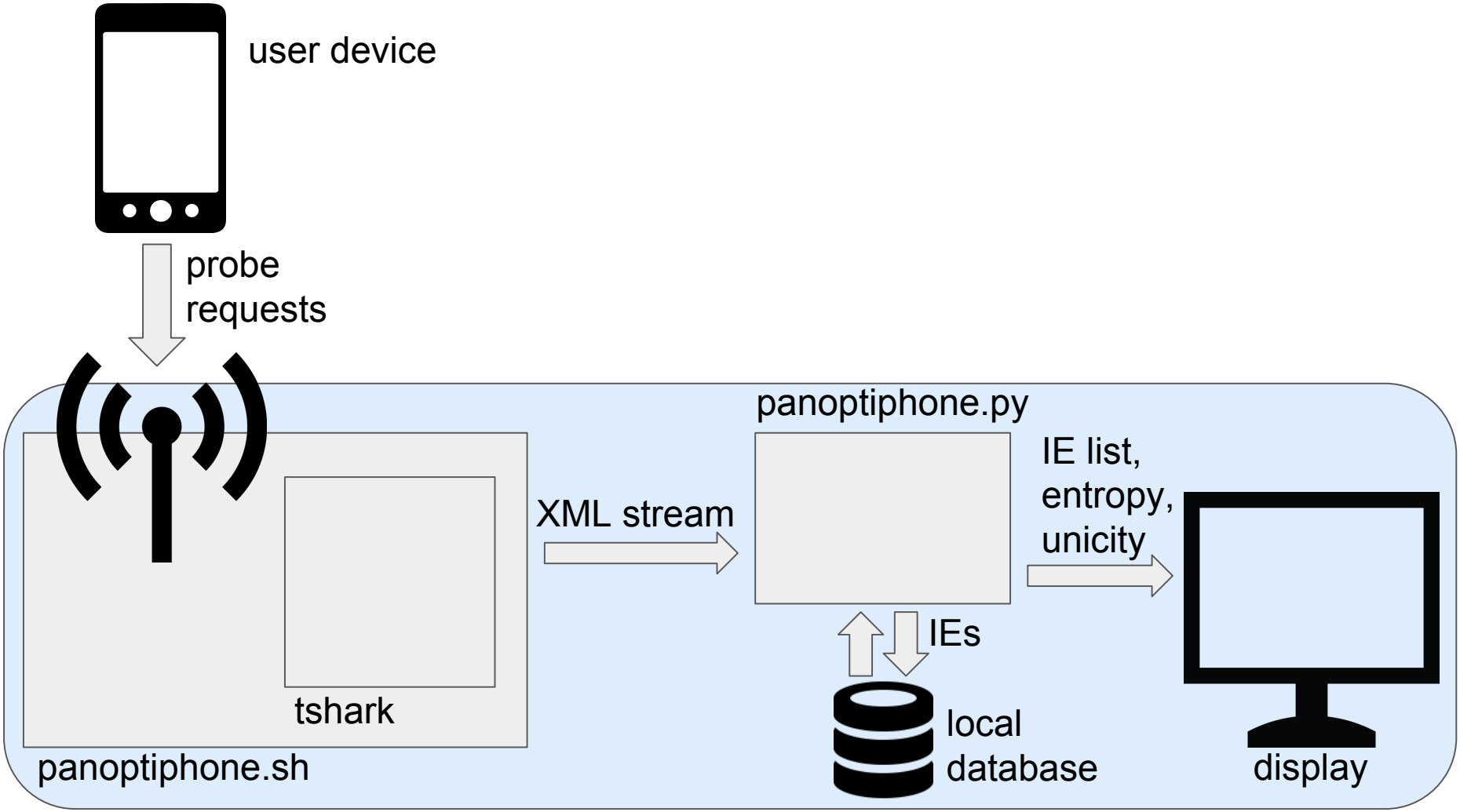
Element	Entropy (bits)			Stability			Affected devices		
	Lab	Station	Sapienza	Lab	Station	Sapienza	Lab	Station	Sapienza
HT capabilities info	3.94	4.74	3.35	96.0%	95.9%	99.6%	90.9%	90.0%	81.1%
Ordered list of tags numbers	4.23	5.24	4.10	93.6%	94.2%	91.2%	100%	100%	100%
Extended capabilities	2.59	2.57	0.064	98.5%	99.4%	99.9%	55.4%	51.3%	0.6%
HT A-MPDU parameters	2.59	2.67	2.54	97.8%	99.1%	99.7%	90.9%	90.0%	81.1%
HT MCS set bitmask	1.49	1.43	1.16	97.6%	99.0%	99.9%	90.9%	90.0%	81.1%
Supported rates	1.18	2.10	1.36	98.2%	95.9%	99.8%	100%	99.9%	100%
Interworking - access net. type	1.08	1.11	0.006	99.6%	99.6%	100.0%	47.5%	46.1%	0.04%
Extended supported rates	1.00	1.77	0.886	98.0%	96.3%	99.4%	99.1%	72.6%	99.7%
WPS UUID	0.878	0.788	0.658	98.2%	99.2%	99.6%	8.4%	5.5%	3.6%
HT extended capabilities	0.654	0.623	0.779	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT TxBeam Forming Cap.	0.598	0.587	0.712	97.8%	98.9%	99.9%	90.9%	90.0%	81.1%
HT Antenna Selection Cap.	0.579	0.576	0.711	98.0%	98.9%	99.9%	90.9%	90.0%	81.1%
Overall	5.48	7.03	5.65	92.5%	90.7%	88.8%	-	-	-

Backup slide: Number of clusters

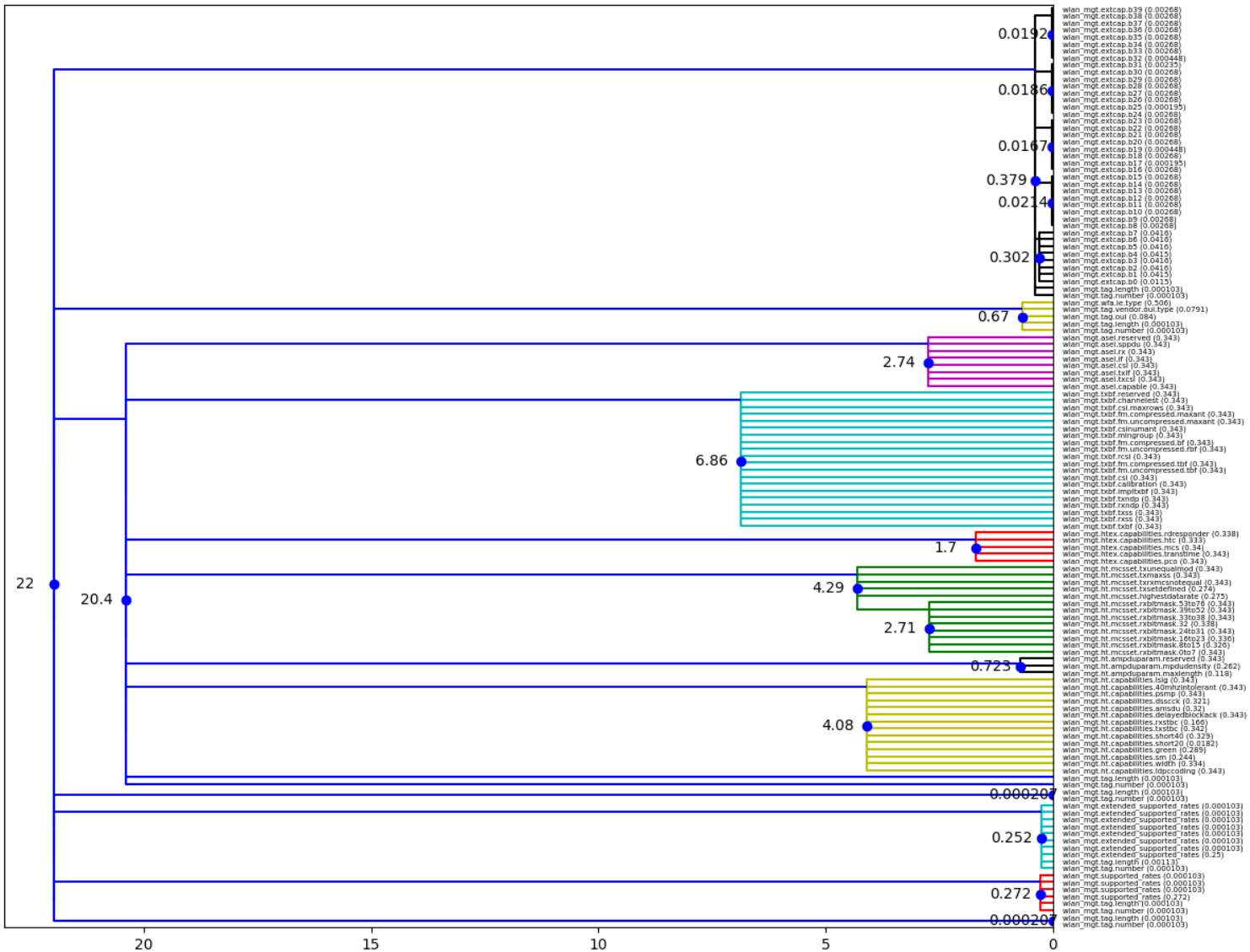
Dataset	DBSCAN	Mean shift	Inc. Algo
Lab	73%	97%	4953%
Lab cut	**11%**	87%	1275%
Lab cut2	**0%**	83%	270%
Belgium	97%		8339%
Belgium cut	38%	90%	2628%
Belgium cut2	35%	79%	674%

Figure: Relative error of the estimated number of clusters.

Backup slide: Panoptiphone



Backup slide: Panoptiphone screenshot



Backup slide: Patterns in sequence numbers and timing

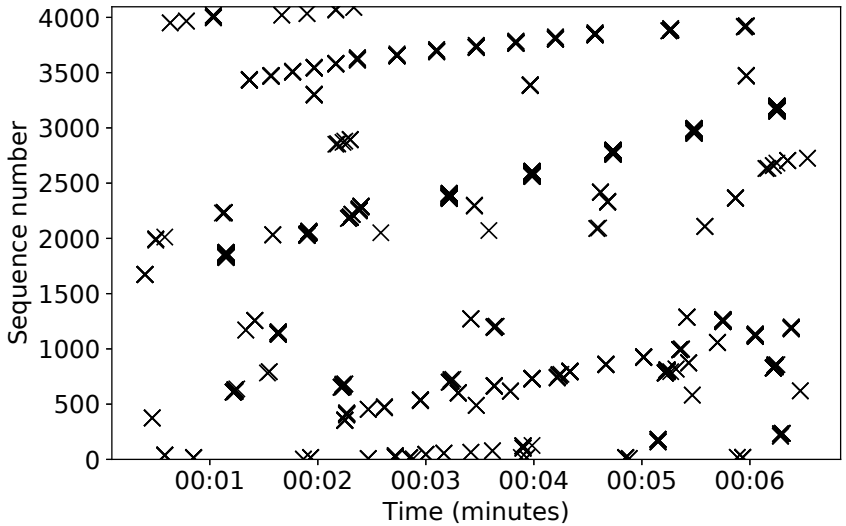


Figure: Sequence numbers wrt. time in part of the Lab dataset.

Backup slide: LA bit

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%		

Table: 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.

Backup slide: Studied devices

Name	OS on release	Release	Developer	Manufacturer	Chipset
Nexus 6P	Android 6.0	2015-09	Google	Huawei	BCM4358 (Broadcom)
Nexus 5X	Android 6.0	2015-09	Google	LG Electronics	QCA6174 (Qualcomm)
OnePlus 3	OxygenOS	2016-06	OnePlus	OnePlus	QCA6174 (Qualcomm)
iPad 2	iOS 4.3	2011-03	Apple	Apple	BCM43291HKUBC (Broadcom)
iPhone 6	iOS 8	2014-09	Apple	Apple	339S0228 (Murata)
iPhone 7	iOS 10.0	2016-09	Apple	Apple	339S00199 (Murata)

Table: Characteristics of the studied devices

Backup slide: Nexus 6P captures

Case Name	Description	Duration
untouched	The phone is turned off, unassociated and not manipulated.	23h
associated	The phone is associated to an access point but not manipulated nor moved	13h30
manipulated	The phone is manipulated every 5-10 minutes, associated to an AP but not moved (except when manipulated). Each time the phone is manipulated, it is turned on, unlocked, and a random app is opened.	4h
moving	The phone is not associated to an AP. It is placed in a person's pocket and not manipulated while the person moves in a small room.	1h40

Table: Description of the different captures of the Nexus 6P.

Backup slide: Filtering approaches for captures

Approach	**Devices**	**Theoretical limitations**
Add specific network to PNL, keep MAC addresses using this SSID at least once	Nexus 6P, Nexus 5X (random uses)	May miss bursts sent using global addresses only, and may increase number of sent probe requests
Keep only probe requests having Google's random OUI `DA:A1:19` or target's global address	OnePlus 3	May include probe requests from other recent Android devices
Faraday cage	iPad 2, iPhone 6	None (requires access to a Faraday cage)
Keep only random addresses, excluding OUIs registered by company not manufacturing the target's model	iPhone 7	May include probe requests from other recent devices of the same manufacturer, or even from different vendors
No filtering	Nexus 5X (untouched)	May include probe requests from other devices
Building a meta-identifier out of Information Elements	None	Collisions may occur, thus mistaking other devices' probe requests for target's ones
Using sequence numbers	None	Difficulty to build a reliable protocol: many sequence numbers are missed, and different devices' sequences may be mixed

Table: Filtering approaches

Backup slide: Address reuse 1

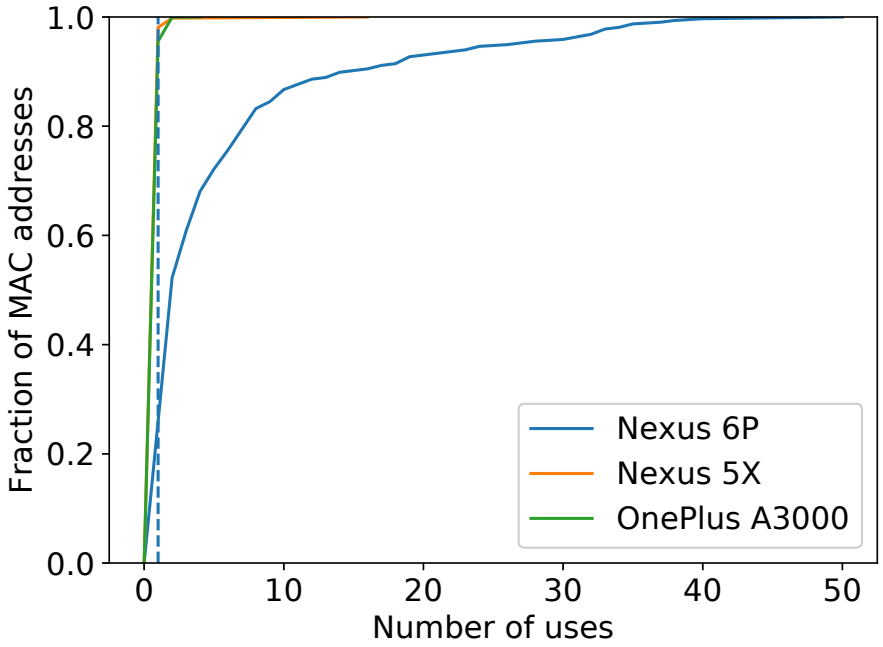


Figure: CDF of the fraction of MAC addresses that are used more than n times.

Backup slide: Address reuse 2

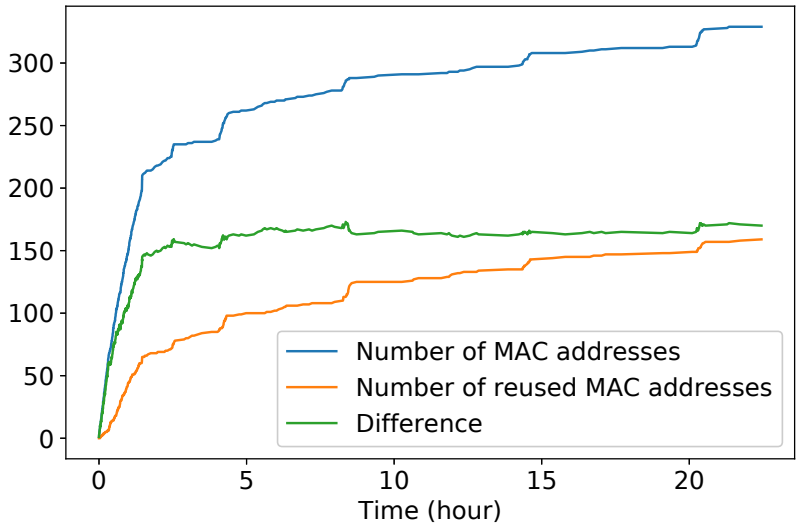


Figure: Evolution of the number of both addresses and reused addresses among time in Nexus 6P's *untouched* case.

Backup slide: Anonymity sets

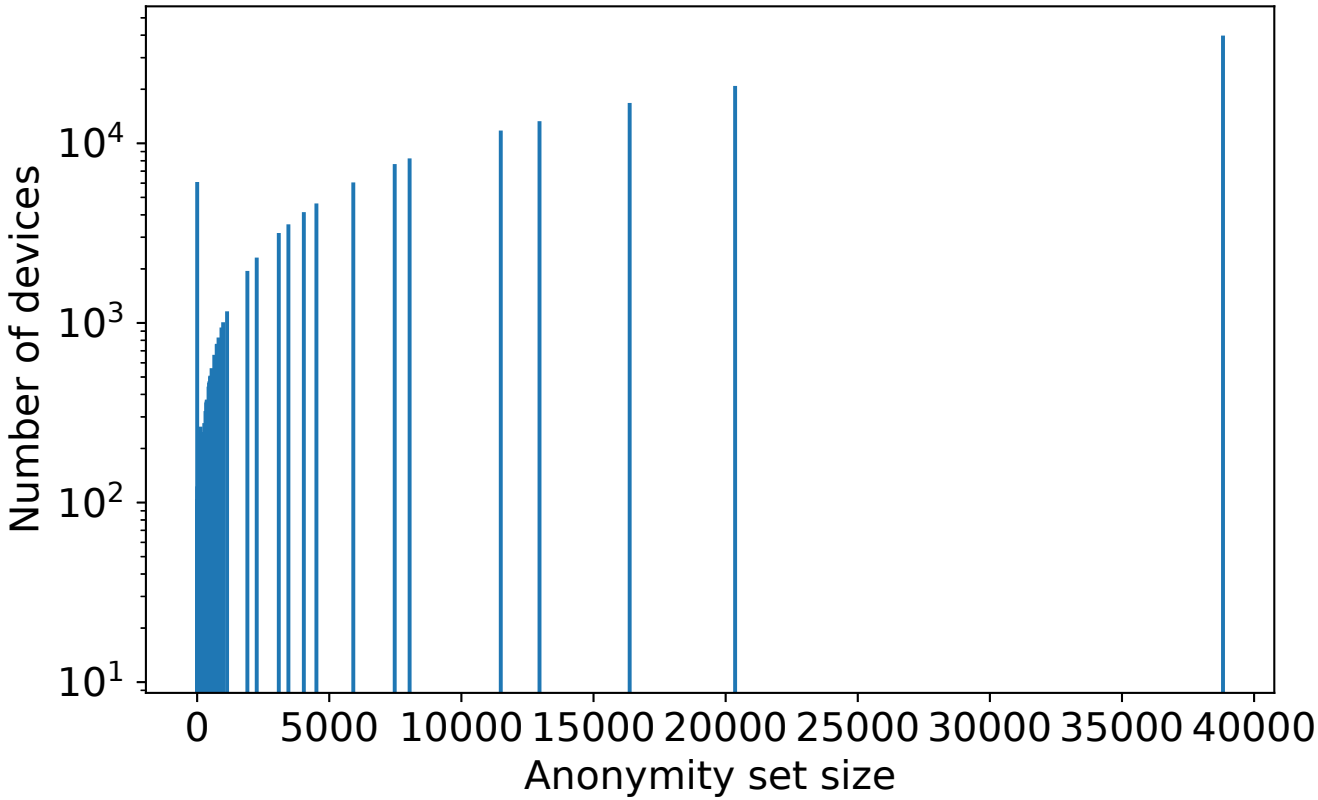


Figure: Number of devices that share the same IE fingerprint with a group (i.e., anonymity set) of various sizes in the Sapienza dataset.

Backup slide: Timing: Incremental algorithm

Algorithm 2: Incremental clustering algorithm

Input: $\mathcal{G}$: bursts, identified by their MAC address
t: distance threshold
d: a distance function

Returns: $\mathcal{C}$: dictionary of clusters

$\mathcal{C} \leftarrow \emptyset$
$\mathcal{D} \leftarrow \emptyset$ // Database of signatures
foreach $\mathcal{B} \in \mathcal{G}$ **do**
  $\mathcal{S} \leftarrow \text{signature}(\mathcal{B})$
  $D \leftarrow \mathcal{D} \cup \mathcal{S}$
foreach $\mathcal{B} \in \mathcal{G}$ **do**
  $d_{min} \leftarrow min(d(\mathcal{S}, \mathcal{S}')$ where $\mathcal{S}' \in \mathcal{D})$
  if $d_{min} < t$ **then**
    $\mathcal{C}[\mathcal{S}'.mac].add(\mathcal{B}.mac)$
  else
    $\mathcal{C}[\mathcal{B}.mac] \leftarrow \mathcal{B}.mac$
return $\mathcal{C}$