The pitfall of address randomization in wireless networks

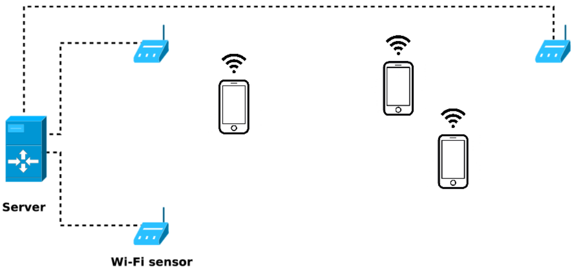
Mathieu Cunche

In collaboration with: Célestin Matte, Mathy Vanhoef, Guillaume Celosia, Franck Rousseau

INSA-Lyon CITI, Inria Privatics

IEEE 802 RCM TIG
17 july 2019, Vienna

Tracking people using radio signals I



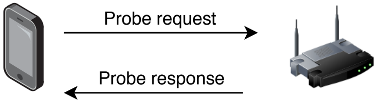
Cyber-physical tracking

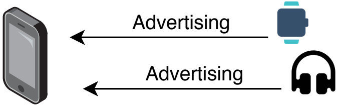
Systems that leverage the ubiquitous digital infrastructure to track individuals in the *physical world*.

- Set of sensors capturing identifiers found in wireless signals ...
- emitted by portable devices (phones, tablets, computers, smartwatches etc.)
- to collect presence and mobility data.

Background: discovery mechanisms

- Discovery protocols in wireless networks
  - Request/Inquiry approach
    - Initiator ask surrounding devices to declare themselves
    - Bluetooth *Inquiries*, Wi-Fi *Probe Requests*
  - Advertising approach
    - Device declare itself by broadcast advertising messages
    - BLE *Advertising Packets*, Wi-Fi *Beacons*

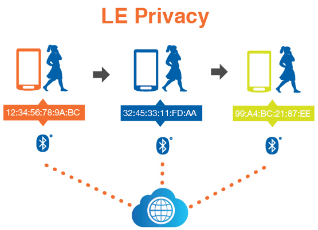




- Wireless-enabled devices broadcast signals
  - Periodically: several pkts/min
  - Packets include a device address

Address randomization

- Address randomization: a simple countermeasure to tracking
  - Tracking is based on the device address in packets
  - Solution: use a random and temporary device address[1]
- Adoption of address randomization
  - Random WiFi address implemented in major systems (iOS, Android, Windows, GNU/Linux)
  - Random BLE address since version 4.2 of Bluetooth



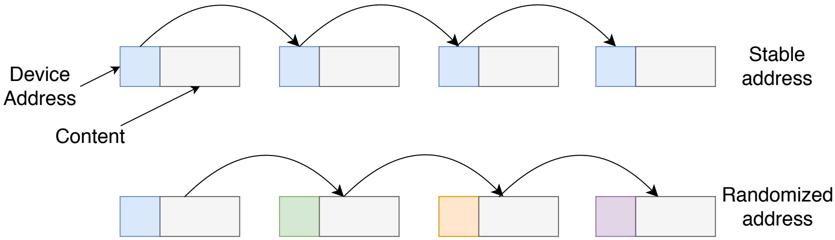
[1] Gruteser and Grunwald, "Enhancing Location Privacy in Wireless LAN Through Disposable Interface Identifiers".

Address randomization is not enough

- We and others have studied implementations

- ... and identified a number of flaws

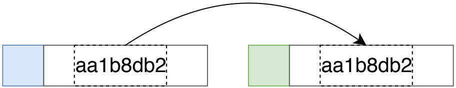
Model



Attacker Model

- Capabilities: can monitor the wireless channel(s)
- Objective: track a device over time
- Success: linking together several packets emitted by the a single device

Secondary Stable Identifiers I



- Stable identifiers: several byte-long fields whose value is constant across frames

$$v_i = v_{i-1} = \dots v_0 = \text{Cst}$$

- Microsoft CDP `Device Hash`
  - a 24-byte identifier found in *Manufacturer Specific* field (BLE)
  - Rotated at a frequency lower than the device address

Time (s)	BD_ADDR	Microsoft CDP Data Device Hash
959.522	37:ee:cb:91:79:0a	db950efc53eff7e427f2a91ae9a67b...
959.719	18:e3:48:43:af:84	db950efc53eff7e427f2a91ae9a67b...
1919.074	2d:39:47:eb:2c:e8	db950efc53eff7e427f2a91ae9a67b...
2879.527	19:fc:04:f1:f3:9a	db950efc53eff7e427f2a91ae9a67b...
3599.189	19:fc:04:f1:f3:9a	4658a402b7da02e09585cb8c4aa1c7...

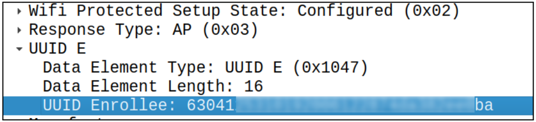
Secondary Stable Identifiers II

- Service UUID in BLE frames
  - A 128 bits UUID including the device MAC address

```
00000020-5749-5448-0037-0024e4659b58
```

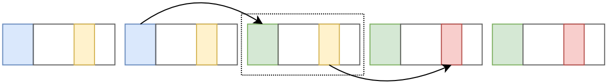
MAC address of a Nokia/Withings Steel HR smartwatch → `00:24:e4:65:9b:58`

- WPS UUID in Wi-Fi frames
  - A 128 bits UUID derived from the MAC address



Synchronization issues

- All identifiers must be rotated together with the device address
  - Those change must be synchronized ...
  - Otherwise the identifier can be used to trivially link two consecutive addresses



  - Ex.: Bad synchronization of *Nearby Id* in Apple Handoff

Time (s)	BD_ADDR	Apple Handoff Data		
		Cnt	Data	Nearby Id
899.885	43:26:33:d5:78:61	-	-	10050b1060c708
899.990	43:26:33:d5:78:61	-	-	10050b1060c708
900.091	6d:01:ff:0a:52:84	-	-	10050b1060c708
900.203	6d:01:ff:0a:52:84	-	-	10050b109d88fb
900.354	6d:01:ff:0a:52:84	-	-	10050b109d88fb

- Need for a cross-layer mechanism for identifier rotation

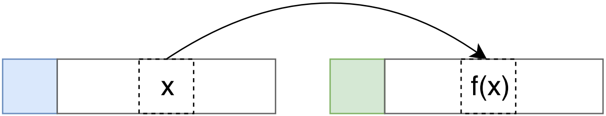
Predictable fields I

- Predictable field: a fields whose value can be computed from the previous occurrences(s)

$$v_i = f(v_{i-1}, \ldots, v_{i-k})$$

  - In general, it only depends on the previous value

$$v_i = f(v_{i-1}),\ f(x) = \begin{cases} x + 1 \\ x + c \end{cases}$$



Predictable fields II

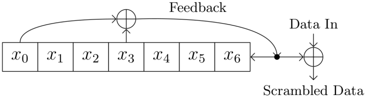
- 802.11 Sequence numbers
  - Found in all 802.11 frames, incremented at each frame
  - Was not reset on address change[2]

324	2.922240000	2a:21:fd:74:38:aa	Broadcast	Probe Request, SN=1035	SSID=Broadcast
328	2.923264000	2a:21:fd:74:38:aa	Broadcast	Probe Request, SN=1034	SSID=Broadcast
331	2.923264000	2a:21:fd:74:38:aa	Broadcast	Probe Request, SN=1035	SSID=Broadcast
338	2.995396000	2a:21:fd:74:38:aa	Broadcast	Probe Request, SN=1039	SSID=Broadcast
538	4.896581000	Apple_74:16:d4	Broadcast	Probe Request, SN=1040	SSID=Broadcast
539	4.896585000	Apple_74:16:d4	Broadcast	Probe Request, SN=1042	SSID=Broadcast
541	4.915017000	Apple_74:16:d4	Broadcast	Probe Request, SN=1043	SSID=Broadcast

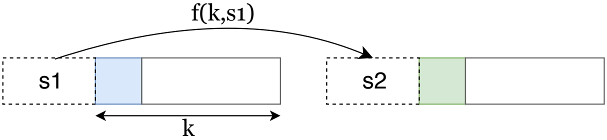
Figure 7: Illustration of randomized iOS 8.1.3 MAC addresses.

Predictable fields III

- 802.11 scrambler seed (PHY layer)[3]
  - Frame scrambled using an LFSR



  - Scrambler seed: state of the LFSR at the beginning of frame.
    - Seed transmitted as part of PHY frame

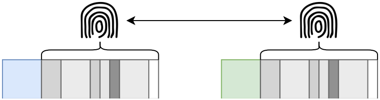


  - *Free Wheeling* mode: LFSR is never reset
  - Seed value depends on previous frame: seed value and length (number of step in the LFSR)

[2] Freudiger, "How talkative is your mobile device?"
[3] Vanhoef et al., "Why MAC Address Randomization is Not Enough".

Fingerprinting I

- Fingerprint: set of stable fields that can constitute an identifier

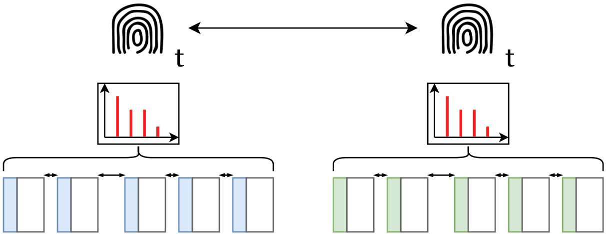


  - Ex: fields describing device capabilities and status
    - Up to 7 bits of entropy in Wi-Fi frames

```
▼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
```

Time-based Fingerprinting I

- Fingerprint: temporal features of packets[4]
  - Device use the same address during a period of time
  - Inter-arrival times statistics: avg, min/max, mean, distribution ...



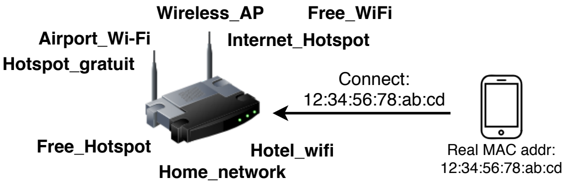
[4] Matte et al., "Defeating MAC Address Randomization Through Timing Attacks".

Active attacks I

Attacker Model

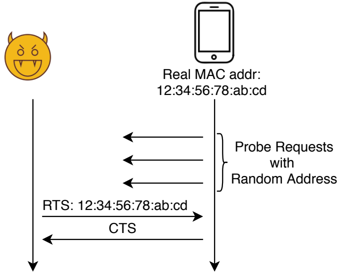
- Capabilities: can capture, replay and forge packets
- Objective: obtain real identity or force to reveal presence

- Revisited Karma Attack[5]
  - Karma attack: fake access point(s) with popular SSIDs
  - Device switch to real MAC address when connecting to AP
  - Attack: set up Karma AP and wait for devices to reveal their MAC addr.



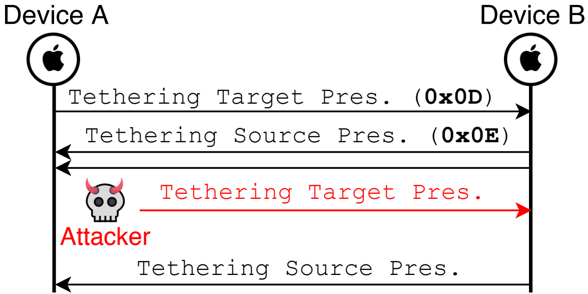
Active attacks II

- Replay Control Frame Attack[6]
  - Request to Send/Clear to Send message
  - Device switch to real MAC address when connecting to AP
  - Pre-requisite: attacker knows target real MAC address
  - Attack: send RTS frame to the target MAC; he will respond if it is nearby



Active attacks III

- Replay of a Tethering Target Presence
  - Apple Instant Hotspot feature: automatically share data connectivity with *friendly*[7] devices
    - Initiation of protocol: Tethering Target Presence → Tethering Source Presence
    - Messages include encrypted identifiers for mutual recognition
  - Attack: replay Tethering Target Presence to test presence of *friendly* device



[5] Vanhoef et al., "Why MAC Address Randomization is Not Enough".
[6] Martin, Mayberry, et al., "A Study of MAC Address Randomization in Mobile Devices and When it Fails".
[7] Associated to same iCloud account

Technical countermeasures

Which Countermeasures ?

Technical countermeasures

Which Countermeasures ?

- Identifiers
  - Remove them or rotate them with device address
- Predictable fields
  - Reset to random value when rotating device address
- Content-based fingerprinting
  - Reduce content to bare minimum
- Timing-based fingerprinting
  - Introduce randomness in timings
- Replay attacks
  - Timestamps and authentication

Lessons learned

Why are those attacks possible ?

[8] http://www.ieee802.org/11/Reports/rcmtig_update.htm

Lessons learned

Why are those attacks possible ?

- Bugs
  - New mechanisms integrated in already complex systems

[8] http://www.ieee802.org/11/Reports/rcmtig_update.htm

Lessons learned

Why are those attacks possible ?

- Bugs
  - New mechanisms integrated in already complex systems
- Lack of specifications
  - Still no specification for address randomization in Wi-Fi
    - Work in progress: IEEE 802.11 Randomized and changing MAC address TIG[8]

[8] http://www.ieee802.org/11/Reports/rcmtig_update.htm

Lessons learned

Why are those attacks possible ?

- Bugs
  - New mechanisms integrated in already complex systems
- Lack of specifications
  - Still no specification for address randomization in Wi-Fi
    - Work in progress: IEEE 802.11 Randomized and changing MAC address TIG[8]
- Specifications: too much freedom given to vendors
  - *"The scrambler should be initialized to any state except all ones when transmitting"* - IEEE 802.11 sec. 15.2.4
  - Some fields are totally free (Vendor/Manufacturer specific)

[8] `http://www.ieee802.org/11/Reports/rcmtig_update.htm`

Lessons learned

Why are those attacks possible ?

- Bugs
  - New mechanisms integrated in already complex systems
- Lack of specifications
  - Still no specification for address randomization in Wi-Fi
    - Work in progress: IEEE 802.11 Randomized and changing MAC address TIG[8]
- Specifications: too much freedom given to vendors
  - *"The scrambler should be initialized to any state except all ones when transmitting"* - IEEE 802.11 sec. 15.2.4
  - Some fields are totally free (Vendor/Manufacturer specific)
- Poor Specifications
  - Privacy is not always considered
  - Interactions with privacy and security researchers could be improved

[8] `http://www.ieee802.org/11/Reports/rcmtig_update.htm`

Manufacturer specific data

- Manufacturer/Vendor Specific Data: fields dedicated to carry custom data
  - Available in BLE and Wi-Fi
  - Up to 32 bytes of data for custom applications
- Used to implement *Proximity Protocols*
  - Custom protocols for close range applications
  - Google Nearby, Apple Continuity, Microsoft CDP ...
  - Activity transfer, pairing, *Instant Hotspot*
- No specification/restriction on their content
- Source of major privacy[9] and security[10][11] issues ...
  - and more is to come[12] ...

[9] Martin, Alpuche, et al., "Handoff All Your Privacy: A Review of Apple's Bluetooth Low Energy Implementation".

[10] Stute et al., "A Billion Open Interfaces for Eve and Mallory: MitM, DoS, and Tracking Attacks on iOS and macOS Through Apple Wireless Direct Link".

[11] Antonioli, Tippenhauer, and Rasmussen, "Nearby Threats: Reversing, Analyzing, and Attacking Google's 'Nearby Connections' on Android".

[12] Celosia and Cunche, "Close Encounters: Privacy Leaks in Apple Bluetooth-Low-Energy Proximity Protocols".

Conclusion

Conclusion

- Address Randomization is hard
  - Complex protocols and a lot of freedom left to vendors

Conclusion

- Address Randomization is hard
  - Complex protocols and a lot of freedom left to vendors
- Wireless networks are affected by other privacy issues
  - Activity inference
  - Inventory attacks
  - Leaks of private data ...

Conclusion

- Address Randomization is hard
  - Complex protocols and a lot of freedom left to vendors
- Wireless networks are affected by other privacy issues
  - Activity inference
  - Inventory attacks
  - Leaks of private data ...
- Issues that are likely to grow ...
  - Growing number of connected objects using wireless communications (IoT, wearables ...)
  - Growing number of the applications and use cases (smarthome, health, V2X, ...)
  - Growing number of number of standards and protocols (LPWAN, 802.11p, Z-Wave, Zigbee, LPD433 ...)

Bibliography

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*. ACM, 2015, p. 8

Mathy Vanhoef et al. "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*. ASIA CCS '16. New York, NY, USA: ACM, 2016, pp. 413–424. ISBN: 978-1-4503-4233-9. (Visited on 08/05/2016)

Jeremy Martin, Travis Mayberry, et al. "A Study of MAC Address Randomization in Mobile Devices and When it Fails". In: *Proceedings on Privacy Enhancing Technologies* (Mar. 2017), pp. 268–286. (Visited on 03/10/2017)

"Saving Private Addresses: An Analysis of Privacy Issues in the Bluetooth-Low-Energy Advertising Mechanism". In: (2019). Under review and embargo due to responsible disclosure

Jeremy Martin, Douglas Alpuche, et al. "Handoff All Your Privacy: A Review of Apple's Bluetooth Low Energy Implementation". In: *arXiv:1904.10600 [cs]* (Apr. 2019). arXiv: 1904.10600. URL: http://arxiv.org/abs/1904.10600 (visited on 05/07/2019)

Contact

Thank you

mathieu.cunche@insa-lyon.fr
http://mathieu.cunche.free.fr
twitter: @Cunchem

SPARTA

INSA IoT chair

SPIE

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON