Tracking 802.11 stations without relying on the link layer identifier

Mathy Vanhoef\textsuperscript{†}, Célestin Matte\textsuperscript{‡}, Mathieu Cunche\textsuperscript{‡}, Leonardo S. Cardoso\textsuperscript{‡}, Frank Piessens\textsuperscript{†}

\textsuperscript{†}iMinds-Distrinet, KU Leuven, \textsuperscript{‡}Univ Lyon, INSA Lyon, Inria, CITI, France

IEEE P802E - 14th April 2016
The paper

*Why MAC Address Randomization is not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms*. Accepted at AsiaCCS 2016.
Introduction

- MAC address randomization proposed to prevent tracking
  - Idea of a disposable link-layer identifier
  - Being deployed in major OSes
    - iOS 8, Android 6, Windows 10, Linux kernel 3.18
- Is it enough to prevent tracking?
  - Probe requests contains a lot of other information
  - Can we track devices despite the lack of a stable link-layer identifier?
  - Can we link together probes from the same device based on their content?
  - Can we force a device to reveal its real MAC address?
Introduction

- Attacker capabilities
  - Monitoring wireless channels
  - Injecting 802.11 frames
- Attacker objectives
  - Group together frames belonging to the same device
- Link-Layer identifier is assumed to change periodically
Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Lab</th>
<th>Train-station</th>
<th>Sapienza(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#MAC addr.</td>
<td>500</td>
<td>10 000</td>
<td>160 000</td>
</tr>
<tr>
<td>#Probe Req.</td>
<td>120 000</td>
<td>110 000</td>
<td>8 million</td>
</tr>
<tr>
<td>Time frame</td>
<td>Oct ’15</td>
<td>Oct/Nov ’15</td>
<td>Feb/May ’13</td>
</tr>
<tr>
<td>Location</td>
<td>Lab</td>
<td>Train Station</td>
<td>Rome</td>
</tr>
</tbody>
</table>

Fingerprinting using Information Elements

- Information elements (a.k.a. tagged parameters, or tags)
  - Indicates the support of capabilities
  - Ex. Supported Rates, High Throughput capabilities and Interworking Capabilities
- High diversity in term of values and in term of information elements present in probe requests
  - Idea: Exploit this diversity to fingerprint devices
Fingerprinting using Information Elements

Figure: Example of the HT_Extended_capabilities Information Element
Empirical evaluation using the datasets

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

- Single Information Elements
  - Can provide up to 5.24 bits of entropy
  - Some IE are found in almost all device (Supported rates)
  - Ex. HT capabilities info (Train-station): 4.74 bits of entropy, 90% of devices affected, stable for 95.9% devices

- Global fingerprint based on most common IE
  - Entropy: 7.03 bits (Train-station)
  - Enough to uniquely identify 1 device among 128 (in average)
<table>
<thead>
<tr>
<th>Element</th>
<th>Entropy (bits)</th>
<th>Stability</th>
<th>Affected devices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lab</td>
<td>Station</td>
<td>Sapienza</td>
</tr>
<tr>
<td>HT capabilities info</td>
<td>3.94</td>
<td>4.74</td>
<td>3.35</td>
</tr>
<tr>
<td>Ordered list of tags numbers</td>
<td>4.23</td>
<td>5.24</td>
<td>4.10</td>
</tr>
<tr>
<td>Extended capabilities</td>
<td>2.59</td>
<td>2.57</td>
<td>0.064</td>
</tr>
<tr>
<td>HT A-MPDU parameters</td>
<td>2.59</td>
<td>2.67</td>
<td>2.54</td>
</tr>
<tr>
<td>HT MCS set bitmask</td>
<td>1.49</td>
<td>1.43</td>
<td>1.16</td>
</tr>
<tr>
<td>Supported rates</td>
<td>1.18</td>
<td>2.10</td>
<td>1.36</td>
</tr>
<tr>
<td>Interworking - access net. type</td>
<td>1.08</td>
<td>1.11</td>
<td>0.006</td>
</tr>
<tr>
<td>Extended supported rates</td>
<td>1.00</td>
<td>1.77</td>
<td>0.886</td>
</tr>
<tr>
<td>WPS UUID</td>
<td>0.878</td>
<td>0.788</td>
<td>0.658</td>
</tr>
<tr>
<td>HT extended capabilities</td>
<td>0.654</td>
<td>0.623</td>
<td>0.779</td>
</tr>
<tr>
<td>HT TxBeam Forming Cap.</td>
<td>0.598</td>
<td>0.587</td>
<td>0.712</td>
</tr>
<tr>
<td>HT Antenna Selection Cap.</td>
<td>0.579</td>
<td>0.576</td>
<td>0.711</td>
</tr>
<tr>
<td>Overall</td>
<td>5.48</td>
<td>7.03</td>
<td>5.65</td>
</tr>
</tbody>
</table>
Fingerprinting using Information Elements

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
  - Could be reversed to reveal the original MAC
- Re-identification attack on the datasets
  - UUID derived from the real Wi-Fi MAC address in 75% of the cases

---

Predictable fields

- Predictable fields in 802.11 frames
  - Fields with a content that can change over time
  - Value in a given frame can be predicted from the previous frames

- Example: Sequence Number field
  - Incremented for each frame
  - Not reset when MAC address is changed in iOS
  - Can be used to trivially defeat MAC Randomization

---

3 Julien Freudiger. “How Talkative is Your Mobile Device? An Experimental Study of Wi-Fi Probe Requests”. In: WiSec. 2015.
Predictable scrambler seed

Scrambler seed

- Scrambler in OFDM frames of 802.11 PHY
  - Scrambler used from the SERVICE field to the end
  - Seed contained in the 7 first bits of SERVICE field

Scrambling sequence generated by a Linear Feedback Shift Register (LFSR)

- Seed set the initial state of LFSR
Predictable scrambler seed

- Scrambler seeds can be predictable

  - Bloessl. et al. showed that it is the case for two prototype implementation of 802.11p
  - No specification in the standard on how to generate the seeds
  - Implementation choice taken by the vendor

- What about commodity 802.11 implementations?

---

Predictable scrambler seed

- Study of scrambler seeds in 802.11 commodity hardware
  - Experimental setup
    - 11 Wi-Fi commodity hardware
    - GNU-Radio implementation of 802.11 based on gr-ieee802-11
    - USRP N210
    - Faraday room from FIT CortexLab

---

5 Bastian Bloessl et al. “An IEEE 802.11 a/g/p OFDM Receiver for GNU Radio”. In: SRIF Workshop. 2013.

6 http://www.cortexlab.fr/
Predictable scrambler seed

- Observed behaviors
  - Freewheeling: State of the LFSR at the end of a frame is reused for the next frame
    - Sometime with a constant number of shift of the LFSR
  - Constant seed, or limited to a small set (bug ?)
  - Incremental: seed value is incremented by one at each frame
Active attacks

- Karma attack
  - Fake AP with popular SSID
  - Trigger authentication/association from STA
  - STA switch back to their real MAC when connecting to AP

- Exploiting Hotspot 2.0
  - Enable Wi-Fi roaming
  - STA send ANQP query to AP to retrieve list of available services
  - STA switch back to their real MAC addr. when querying
  - Query also contain predictable counter that could help tracking
Countermeasures

- Information elements in probe requests
  - Are they really needed?
  - Remove them or restrict to a bare minimum

- Scrambler seed and counters
  - Reset to a random value upon MAC addr change
  - Unpredictable scrambler seeds
    - Use a crypto PRNG to generate seeds
    - Chipsets allowing a reset of the seed

- Active attacks
  - Keep random MAC addr. when sending ANQP queries