MITIK-LINK: MAC Address Association Tool

Abhishek Kumar Mishra, Fernando Molano Ortiz, Nadjib Achir, Aline Carneiro Viana

Mobility and contact traces from
non-intrusive passive measurements

ANR PRC call

MITIK-LINK
MAC Address Association Tool

Version v1.0

Abhishek Mishra[1], Fernando Molano Ortiz[1], Nadjib Achir[1], Aline Carneiro Viana[1]

[1] *INRIA, France.*

Contents

Copyright

Acknowledgment

This work has been partially funded by the ANR MITIK project, French National Research Agency (ANR), PRC AAPG2019.

1 Introduction

MAC address randomization disrupts the continuity and semantics of probe-requests and breaks the network data collection and analysis process. While this mechanism protects the user's privacy, it impacts the continuity and accuracy of crucial works and strategies relying on MAC addresses as user-device identifiers. To address continuity and accuracy issues, recent research extensively explores MAC address association, which involves linking (associating) randomized MAC addresses emitted by a specific device.

2 MITIK-LINK's principle

To protect user privacy, the WiFi standard strictly recommends mobile devices to periodically change (randomize) the true MAC advertised in their probe-request frames while performing active scanning on available channels. While MAC randomization mechanism protects the user's privacy, it reduces the correlation between probe-requests (with unique MAC addresses) and the corresponding emitting devices, impacting the process of network data collection, the corresponding data analysis, and devices' trajectory reconstruction.

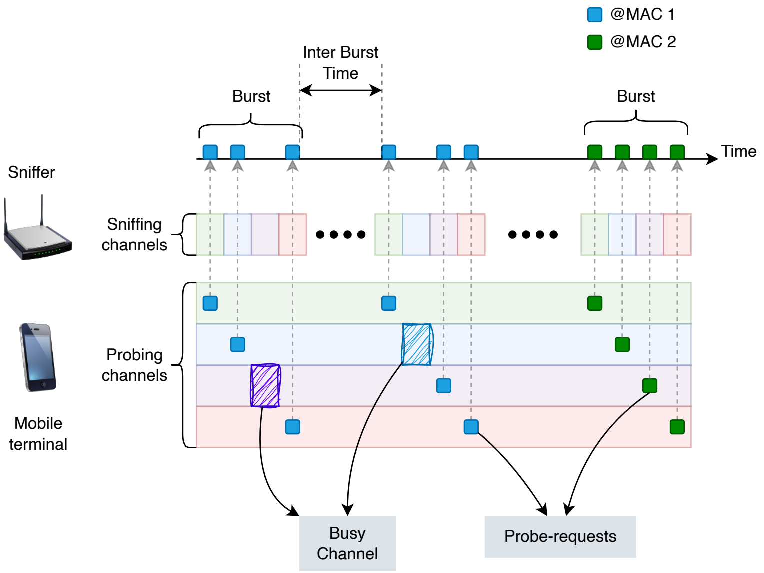

Fig. 1: A device's randomized probe-requests.

As shown in Figure 1, multiple rounds of active scanning contain bursts of probe-requests, captured by the sniffer with the MAC address of individual probes within a burst remaining consistent. The MAC address of a device is likely to change (randomize) in subsequent bursts, known as MAC randomization. The longer it takes for a device to discover a network, the more probes will circulate from the same device, increasing the number of randomized MACs.

`MITIK-LINK` performs the MAC association of randomized MAC addresses used by the same device. This tool models the frame association to resolve MAC conflicts in small intervals. It uses time and frame content-based signatures to resolve and associate MACs inside a conflict. Finally, a logistic regression-based algorithm using the obtained signatures is proposed to associate devices with similar signatures.

3 Bleach framework

The framework **Bleach** takes probe-request trace with randomized MAC addresses as input and yields a dictionary (A) of randomized addresses (M_j) associated with particular devices (U_n).

Diagram of the framework:

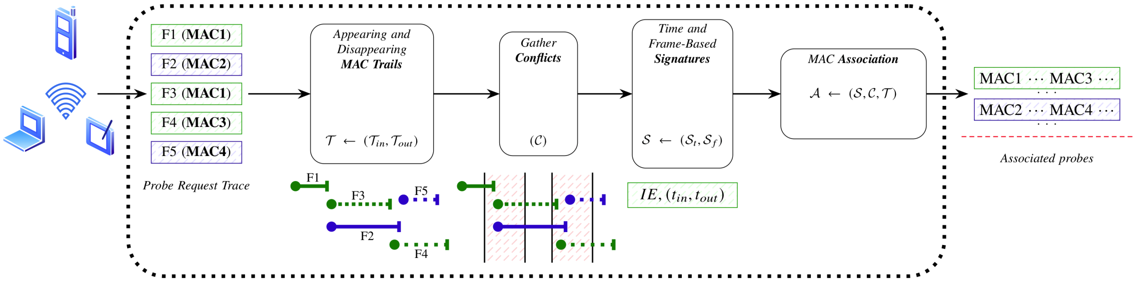
F1 (MAC1)
F2 (MAC2)
F3 (MAC1)
F4 (MAC3)
F5 (MAC4)
Probe Request Trace
Appearing and Disappearing MAC Trails
$\mathcal{T} \leftarrow (\mathcal{T}_{in}, \mathcal{T}_{out})$
Gather Conflicts
$(\mathcal{C})$
Time and Frame-Based Signatures
$\mathcal{S} \leftarrow (\mathcal{S}_t, \mathcal{S}_f)$
$IE, (t_{in}, t_{out})$
MAC Association
$\mathcal{A} \leftarrow (\mathcal{S}, \mathcal{C}, \mathcal{T})$
MAC1 ⋯ MAC3 ⋯
MAC2 ⋯ MAC4 ⋯
Associated probes
F1
F3
F5
F2
F4

Fig. 2: Bleach framework.

The stages of the framework are described below:

1. We transform the input probe-request trace into a set of MAC address trails. Each MAC trail can be viewed as an instance of the appearance or the disappearance of a MAC address in the sniffing zone. This reduces the problem of MAC association to that of correctly associating each disappearing MAC trail from a device with an appearing trail from the same device.

2. We separate the trails into disjoint subsets comprising **conflicts** (C). The conflict denotes the set from which a disappearing MAC trail could be possibly associated, with any of the appearing MAC trails present in the dataset, within a period $(T_c)^{\tau_i}$ from the end of the disappearing trail. We identify this period as the conflict period. As a result, the right value of the conflict period allows us to consider all potential associations while making the decision of linking the MAC address trail pairs.

3. Conflicts are either caused by devices that change their MAC addresses or by their entry/exit from the sniffing range. Any address association framework, in essence, has to resolve conflicts to perform correct assignments between the disappearing and appearing MAC from individual devices. After obtaining conflicts of MAC address changes and a generic formulation of the MAC association problem, we take a step further toward the association itself. We need to obtain effective signatures for resolving conflicting MAC address trails.

4. We define and extract the time and frame-based signatures (S_t), (S_f) from the collected MAC trails. We consider two types of signatures:

   - **Time-based signatures**, which utilize the information from the temporal behavior of received probe-request frames;
   - **Frame-based signatures**, which use the control field information present in the captured frame itself to form effective signatures that have the potential of discriminating a device from the rest of the population.

5. We introduce a novel MAC association algorithm capable of resolving the conflicts observed in the input dataset accurately. It uses extracted signatures (S) to fingerprint and differentiates randomized MACs in each conflict duration in order to finally associate them.

This repository contains the code to match random MAC address coming from the same WiFi devices.

4 How to use the tool

To assess the MAC association capabilities of `MITIK-LINK`, we use ***probe requests*** *(a type of WiFi management frame that does not contain user application data)*. These probe requests are captured by a privacy-preserving WiFi Sniffer tool (`MITIK-SENS`).

Once the probe requests are captured and stored in `pcap` files, we extract all available frames to a text file format (`csv`) files for further analysis. It is important to note that `MITIK-LINK` requires all the information elements within the probe requests captured by the sniffer tool.

For usage, run:

```
$ python3 main.py
```

5 Link for the tool

The tool and running instructions are available on the following link:
`https://gitlab.inria.fr/mitik/mac-association/mitik-link`

6 License

This code has been developed within the ANR MITIK project and is partially related to the funded PhD Thesis titled "Revealing and exploiting privacy vulnerabilities in users' public wireless packets" for research purposes. It is released under the license GNU General Public License v3.0 or later. While you are welcome to explore and utilize it for academic or research purposes, we cannot guarantee ongoing support or updates. Use of this code is at your own discretion, and we encourage you to exercise caution and discretion in its adaptation. Terms and conditions to use this software are detailed in the GitLab text of the tool license in `https://gitlab.inria.fr/mitik/mac-association/mitik-link/-/blob/main/LICENSE?ref_type=heads`.

References

[1] Abhishek Kumar Mishra. “Revealing and exploiting privacy vulnerabilities in users’ public wireless packets”. PhD thesis. Institut Polytechnique de Paris, 2023.

[2] Abhishek Kumar Mishra, Aline Carneiro Viana, and Nadjib Achir. “Introducing benchmarks for evaluating user-privacy vulnerability in WiFi”. In: *VTC2023-Spring*. Florence, Italy, June 2023.

[3] Abhishek Kumar Mishra, Aline Carneiro Viana, and Nadjib Achir. “Bleach: From WiFi probe-request signatures to MAC association”. In: *Ad Hoc Networks* 164 (2024), p. 103623.

[4] Abhishek Kumar Mishra et al. “Public Wireless Packets Anonymously Hurt You”. In: *IEEE 46th Conference on Local Computer Networks (LCN)*. 2021, pp. 649–652.

[5] Fernando Molano Ortiz et al. “Collecte de traces WiFi publiques: de la protection de la vie privée à l’analyse de trajectoires”. In: *CoRes 2024 - 9èmes Rencontres Francophones sur la Conception de Protocoles, l’Évaluation de Performance et l’Expérimentation des Réseaux de Communication*. May 2024, pp. 1–4.