ASIMOV: a Fully Passive WiFi Device Tracking

Rafael Hengen Ribeiro, Bruno Bastos Rodrigues, Christian Killer,
Lenz Baumann, Muriel Figueredo Franco, Eder John Scheid, Burkhard Stiller
Communication Systems Group CSG, Department of Informatics IfI, University of Zürich UZH
Binzmühlestrasse 14, CH—8050 Zürich, Switzerland
[ribeiro|rodrigues|killer|franco|scheid|stiller]@ifi.uzh.ch, lenz.baumann2@uzh.ch

***Abstract*—Mobile devices expose information about their hardware and manufacturer while searching for available WiFi networks via a Media Access Control (MAC) protocol. Thus, to protect the users' privacy and prevent MAC address tracking, manufacturers typically provide anonymity through MAC randomization techniques by randomly and periodically modifying the MAC address. This paper presents the ASIMOV tracking approach, which shows through the correlation of randomized information concerning the displacement of devices in space-time dimensions that it is possible to gain insights into identifiable device information. The proposed system is entirely passive and uses a combined Received Signal Strength Indicator (RSSI) value-based localization and the Information Elements (IE) transmitted in every IEEE 802.11 probe request frame.**

I. Introduction

The analysis of wireless signals emitted by portable devices, such as smartphones, laptops, and tablets, enables the extraction of positional data from those devices passively, *i.e.*, devices can be tracked even if not directly connected to an Access Point (AP) of an IEEE 802.11-compatible wireless network. Tracking devices and their behavior are key aspects of strategic business planning by tracking people's movement and behavior. For instance, business owners can use wireless sensing to understand customer behavior and arrange products corresponding to the customers' interest points inside stores. Also, wireless sensing is an enabler for contact tracing apps used during pandemics [1].

Many different devices carried around, such as telephones, laptops, and watches, expose information about their hardware and manufacturer while probing on their own for available wireless networks via the Media Access Control (MAC) protocol. This is technically necessary to be online and attached to a wireless network actively. As a central piece of information, the MAC address can link and connect different information since it allows for a unique identification of every device participating in wireless data transmission. On the contrary, to protect the users' privacy and prevent MAC address tracking, manufacturers typically provide anonymity through MAC randomization techniques [2].

ASIMOV circumvents "difficulties" introduced by the MAC randomization concerning tracking targets. The approach taken uses the estimated localization and Information Elements (IE) to determine whether traffic captured originates from the same device passively or not, thus, enabling the tracking of devices even when they use a MAC address randomization. For this, the prototype ASIMOV was implemented. Unlike previous de-anonymization approaches, such as *NiFi [3]* and *Wobly [4]*, ASIMOV does not rely solely on specific data fields that are not assumed to be stable over time or universally equal from device to device. On the contrary, ASIMOV uses a combined Received Signal Strength Indicator (RSSI) value-based localization and the IE transmitted in every IEEE 802.11 probe request frame. ASIMOV is entirely passive and can determine how many devices are present and track overtime a single device in the area covered.

Annex to ISBN 978-3-903176-39-3 © 2021 IFIP

Three limited experiments were conducted — Corona prevented a planned for larger-scale, public experiment — to demonstrate the feasibility of the proposed solution. In two of these, ASIMOV located devices at different positions in space within the covered area with high accuracy. The third experiment, conducted during a smaller real-life event, evidences that ASIMOV is very practical for counting the number of devices and localizing them in space over time, correcting the bias introduced by MAC randomization.

The main contributions of this paper include (*i*) a methodology to distinguish MAC-randomized devices as well as a tracking tool in Section II and (*ii*) an evaluation based on three experiments that comprehend the localization and device counting capability in Section III. The summary and future work are part of Section IV.

II. ASIMOV Overview

ASIMOV consists of a process to distinguish devices, divided into 1. data gathering and 2. data analysis. Both processes are started by the user, who manages devices and the processes through an intuitive management interface (*cf.* Figure 1). The data gathering process orchestrates components to obtain data from available devices, aggregate these data, and store relevant data. This process starts with Monitor nodes being configured to capture devices' signals within a specified area. Monitor nodes dump Probe Requests (PR) received at the specified Network Interface Card (NIC), extract relevant data fields, such as MAC address, timestamps, RSSI values, and IEs, and return them to the Sync node. Sync nodes aggregate and store data from monitors in a shared database for posterior analysis, interacting with the Interface on one side and with the Monitor nodes on the other side.

The data analysis process is composed out of three steps. In the information retrieving step (**Step A**), the system loads data from the shared database for posterior analysis: all packets are

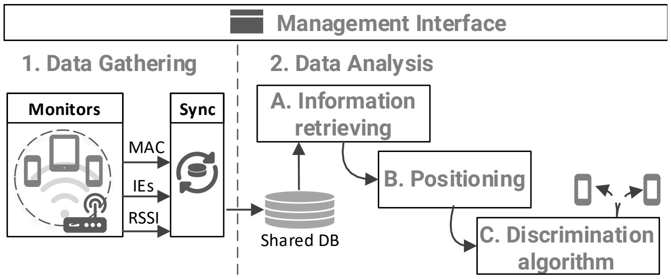

Fig. 1: ASIMOV Architecture and Process to Distinguish MAC-randomized Devices

reassembled, such that for each PR, the RSSI values measured at different monitor nodes are stored. The positioning step (**Step B**) uses RSSI information to localize devices using multilateration, as described further in Subsection II-A. In the discrimination algorithm (*cf.* Figure 1 **Step C**) ASIMOV determines, whether the original device is applying MAC address randomization. The algorithm relies on snapshots of measurements over time, combining knowledge from the current snapshot and the aggregation of past snapshots, to classify devices as “not seen before”, “randomizing” if they are randomizing their MAC addresses, or “non-randomizing,” if devices do not apply MAC randomization at all.

During the first step, the discrimination algorithm checks whether the database contains a combination of MAC address and IEs for that device. If it is true, it is assumed that the device does not apply MAC address randomization. When the database contains equal IEs, but the MAC address is different, ASIMOV uses the position estimated as additional tracking information. ASIMOV then compares this position with previous locations containing these IEs to determine if the device is the same. If these locations are within a predefined distance defined via a threshold, devices are assumed to be the same.

A. Localization

The localization solution is based on the log distance path-loss model for trilateration [5]. All RSSI values received at different APs can be used to localize devices.

The basic equation of the log distance path-loss model sets the finding that a logarithmic function can approximate the decay of a signal over distance. With $RSS(d)$ being the received signal strength at distance d, d_0 a reference distance, n the path-loss coefficient, and $\mathcal{X}_\sigma$ a zero-mean Gaussian random variable, it can be defined as follows:

$$RSS(d) = RSS(d_0) - 10n \log\left(\frac{d}{d_0}\right) + \mathcal{X}_\sigma \tag{1}$$

In practice, the reference distance d_0 is often set to $1\,\mathrm{m}$ and noise is ignored for the calculation, simplifying the model even further. With RSS_C being the received signal strength at $1\,\mathrm{m}$ it can be expressed as follows:

$$RSS(d) = RSS_C - 10n \log(d) \tag{2}$$

RSS_C depends on each device and has to be calibrated. The path-loss coefficient n is a factor depending on the environment. For free-space, it is often chosen at $n = 2$.

By applying the equation 1 to the RSSI values measured at the different receivers yields in an approximation of the distance, the sender has to each AP. The sender's position can be estimated by combining these distance estimates with multilateralism. Multilateration combines the multiple distances between a device with an unknown location and multiple spatially separated APs with a known site to estimate the unknown device's location, as expressed in equation 3.

$$(x_i - x_u)^2 + (y_i - y_u)^2 = r_i^2 \text{ for r in 1...3} \tag{3}$$

Mathematically, this corresponds to solving the following non-linear system, with (x_i, y_i, z_i) being the position of the i-th point, (x, y, z) the position of the object, and d_i the distance of the object to the i-th point. This can be simplified for planar problems, leading to the following system in two variables:

$$\begin{aligned}(x - x_1)^2 + (y - y_1)^2 &= d_1^2\\ (x - x_2)^2 + (y - y_2)^2 &= d_2^2\\ (x - x_3)^2 + (y - y_3)^2 &= d_3^2\end{aligned} \tag{4}$$

However, multiple measurements exist for the same position, and more than the minimum of three APs are receiving a signal from the same sender. ASIMOV profits from the excess of data by incorporating as much information as possible into the multilateration process.

III. Evaluation

Three experiments had been conducted to test the localization-based approach of MAC address de-anonymization. Two experiments were focused on localization; the third one was run during a real-life event.

A. Localization Experiments

In **Experiment 1** (conducted in the open field space), only WiFi signals were measured. Four Monitor nodes were placed at the corners of an 8 by 5 m grid, with 48 measurement points. The sender was a Raspberry Pi 3 with Alfa-AWUS036NHA Wifi-Adapter and monitored four Spitz GL-Routers. Experiment 2 (conducted in an urban environment on a concrete balcony) combines the location estimates of RSSI measurements from both WiFi probe requests and Bluetooth-based measurements [6]. Four nodes were placed at corners of a 5 by 5 m grid, with 32 measurement points, including an additional Bluetooth device in place.

For experiment 1, the overall error in meters was 87 m without filtering and 58 meters applying Kalman Filter based smoothing. It results in a per step error of 1.7 m or 1.1 m with the Kalman Filter. Including the maximal possible distance of 10.8 m, this results in an error of 0.15 without filtering and 0.1 using Kalman Filter. The deviation of the unfiltered location estimates ranges from 0.03 to 5 m. A sample of Experiment 1 on position (0, 1) is shown in Figure 2. The true position is shown with the empty big blue circle. The different estimates of the unfiltered (light blue), Kalman Filter based (dark blue), and variance-based approaches (orange, green, violet) are shown as filled dots.

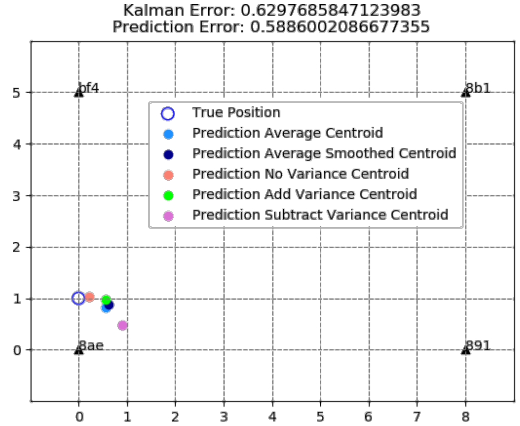

Fig. 2: Sample of Experiment 1 on Position (0, 1)

In **Experiment 2**, the overall error in meters was 72 m without filtering and 68 m applying Kalman Filter based smoothing. It results in a per-step error of 2.2 m or 2.1 m with the Kalman Filter. It implies an error rate of 0.26 without filtering and 0.25 with Kalman Filter. The deviation of unfiltered location estimates ranges from 0.44 to 4.5 m. Bluetooth-based localization [6] in Experiment 2 achieved an overall error of 68 m, yielding a per-step error of 2.1 m. The deviation of the unfiltered location estimates ranges from 0.8 to 3.8 m. It implies an error rate of 0.25.

B. In-Field Experiment

Experiment 3 (conducted during a real-life event on June 2, 2020) counted devices instead of verifying location estimates. Hence, the effectiveness of using the combined approach of IEs and location estimate was evaluated using ASIMOV, too. The setup of this experiment included a Livealytics booth placed at the event with a 4 by 4 m area. Figure 3 details the setup: a cyan square represents the Livealytics booth, and red dots represent monitoring devices. ASIMOV Monitor nodes were placed at the height of 2 m above the floor.

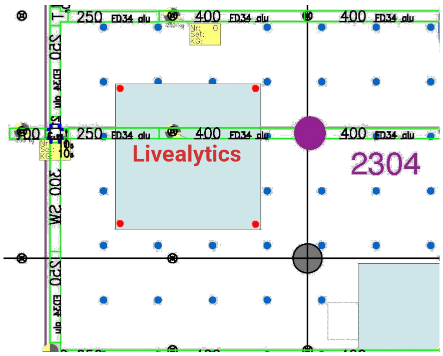

Fig. 3: Map of the Livealytics Booth at the Event

Two external data sources were available, too: a ticketing system and a ceiling camera. The ticketing system counts the number of people entering and the number of people leaving the site. The ceiling camera keeps track of the number of people entering and exiting the LiveAlytics booth area. Thus, it was limited to the booth's exact borders and did not include any person that stood outside of this area.

An overall number of 566 people's first entrances was accounted for, which corresponds to the number of total visitors that entered the site for the entire day. This count, however, does not includes visitors only but also staff and exhibitors. The official number of visitors during the entire day was 360. Additionally, the number of people seen multiple times using MAC address randomization was 370.

IV. Summary and Future Work

ASIMOV applies a combined approach to overcome the trade-off between effectiveness, stability over time and devices, and straightforward deployability. The solution uses the stability and device independence of a location-based approach and combines it with the informative value of a content-based approach. ASIMOV distinguishes two devices from each other, even if both devices are using MAC address randomization, as three experiments were conducted to evaluate this capability. The evaluation demonstrated that RSSI-based localization performs better than anticipated, and it suffices as a proxy to distinguish different devices from each other. Future work includes *(i)* the design of a responsive user interface, *(ii)* the improvement of error handling of monitor nodes in the event of unexpected behavior, and *(iii)* an evaluation of ASIMOV with different hardware types to analyze.

Acknowledgements

This work has been supported by Innosuisse, the Swiss Agency for Innovation Demands under Grant No. 42193.1 IP-ICT, the PasWITS project. Also, the authors would also like to thank Simon Tuck, Livealytics AG, who supported the research and development during those experiments in Corona times of limited public accessibility.

References

[1] A. D. Carli, M. Franco, A. Gassmann, C. Killer, B. Rodrigues, E. Scheid, D. Schoenbaechler, and B. Stiller, "WeTrace – A Privacy-preserving Mobile COVID-19 Tracing Approach and Application," 2020.

[2] J. Martin, T. Mayberry, C. Donahue, L. Foppe, L. Brown, C. Riggins, E. C. Rye, and D. Brown, "A Study of MAC Address Randomization in Mobile Devices and When it Fails," *Proceedings on Privacy Enhancing Technologies*, No. 4, 2017, pp. 365–383.

[3] L. Cheng and J. Wang, "How Can I Guard My AP? Non-Intrusive User Identification for Mobile Devices Using WiFi Signals," in *Proceedings of the 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 91–100.

[4] Y. Li and T. Zhu, "Gait-Based Wi-Fi Signatures for Privacy-Preserving," in *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, ser. ASIA CCS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 571–582.

[5] Yongguang Chen and H. Kobayashi, "Signal strength based indoor geolocation," in *2002 IEEE International Conference on Communications. Conference Proceedings. ICC 2002 (Cat. No.02CH37333)*, Vol. 1, 2002, pp. 436–439 vol.1.

[6] B. Rodrigues, C. Halter, M. Franco, E. J. Scheid, C. Killer, and B. Stiller, "BluePIL: a Bluetooth-based Passive Indoor Localization Method," in *IFIP/IEEE International Symposium on Integrated Network Management (IM 2021)*. Bordeaux, France: IEEE, may 2021, pp. 1–9.