SimBle: Comment générer des traces réelles Bluetooth conformes aux recommandations de préservation de la vie privée ?

Abhishek K Mishra, Aline Carneiro Viana, Nadjib Achir

SimBle: Comment générer des traces réelles Bluetooth conformes aux recommandations de préservation de la vie privée ?

Abhishek K. Mishra[1,2] et Aline C. Viana[1] et Nadjib Achir[1,3]

[1] *INRIA Saclay, Batiment Alan Turing Campus de l'Ecole Polytechnique, 91120 Palaiseau, France*
[2] *Ecole Polytechnique, Palaiseau, France*
[3] *Université Sorbonne Paris Nord, Paris, France*

Dans cet article, nous présentons le premier outil de simulation du protocole Bluetooth capable de générer des traces compatibles avec les recommandations du standard sur la préservation de la vie privée. En effet, les outils de simulations existants, tels que NS-3, sont dépourvus de toutes fonctionnalités de préservation de la vie privée telle que la randomisation d'adresses MAC. Dans cet article, nous proposons une solution pour intégrer la randomisation d'adresse MAC dans NS-3 afin d'émuler et de générer des traces de simulation Bluetooth conformes aux traces du monde réel. De plus, étant donné que le temps d'exécution des simulations croît de façon exponentielle avec le nombre d'appareils, nous introduisons également une optimisation pour linéariser le temps de collecte des paquets publics. Nous validons notre proposition par une étude de cas d'une stratégie d'association d'adresses MAC de la littérature. Des scénarios réalistes d'appareils et de mobilité ont été évalués.

Mots-clefs : Bluetooth, randomisation d'adresses MAC, NS-3, préservation de la vie privée

1 Introduction

Human-to-human interaction has been a prevalent paradigm over the Internet for quite some time now. Different wired and wireless communication technologies are used to serve the purpose. The Internet of Things (IoT) is expected to connect billions of low-end devices to the Internet. The total count of products and businesses that use IoT technologies has increased to about 25 percent in the period(2014-2019), and the number of connected devices is projected to reach 43 billion by 2023.Growth of networked devices has brought increasing concerns for user-privacy. These concerns vary from protection of receiver-location to anonymity and traceability in general. Devices that perform *MAC address randomization* can hide a device's identity to some extent. This feature has been the backbone of user-privacy in wireless networks especially BLE and WiFi. MAC address randomization in mobile devices has thoroughly been studied[MMD$^+$17]. [MMD$^+$17] claim to effectively defeat randomization for around 96 percent of android devices. [UCF$^+$20] use artificial intelligence to show that 91 percent of the WiFi devices could be tracked. MAC address randomization in BLE has been claimed to be defeated for Apple devices[MAB$^+$19] and for generalized devices[JVAF21]. [JVAF21] claim to get 100% device association for small set of devices on sniffing public-packets in a controlled environment (inside Faraday cage). All these solutions are either evaluated on manually anonymized non-random traces or controlled environments with a small set of devices in a particular environment. If the results are similar to what have been claimed i.e close to 100 percent in realistic environments, immense threats to user-privacy are posed.

Amidst raising privacy intrusion findings in BLE, there has been an absence of frameworks to test these suggestions in scalable real-world conditions. Present works validate their effectiveness using various trace-collection techniques like test-beds, active user-participation and controlled experimentation. But these methodologies lack real-world factors like scalability, extensive mobility, *ground truth* in large-scale

scenarios, and variability of considered devices. *Ground truth* here refers to the knowledge of a set of randomized MAC addresses that are emitted from a particular device. If we manage to capture the device heterogeneity in the population and emulate a standard's privacy provisions in the *simulation*, we can evaluate any privacy-intrusion solution using generated traces. This equips us to propose adjustments in the standard to guarantee the user's privacy. This paper introduces *SimBle*, which is capable of generating real-world BLE traces with *ground truth* for large scale scenarios in the NS-3(https://www.nsnam.org) framework. This work is part of the project ANR MITIK[†].

2 SimBle Framework

The framework of *SimBle* generates real-world privacy preserving traces and uses them to test BLE standard's privacy. The framework can be divided into three major components: 1) Formalisation 2) Simulation and 3) Evaluation. *Formalisation* handles two main issues: capturing the *device heterogeneity* in the population and emulating a particular standard's privacy provisions. *Simulation* takes care of deploying real-world nodes belonging to different manufacturers, incorporating device mobility and most importantly optimising the run-time of the sniffer-capture to practical bounds. *Evaluation* collects per-sniffer traces, extracts the *ground truth* information and subsequently evaluates standard's privacy-intruding solutions. We detail *Formalisation* in this section while we describe *Simulation* and *Evaluation* in Section 3.1 and 3.2 respectively. *SimBle* accepts any user-specific parameters like those related to mobility and address randomization. In this paper, we just pick suitable mobility profile and randomization behaviour for the evaluation of framework in Section 3.2.

In *Formalisation*, we focus on modelling and abstracting various BLE device specific variables. We extensively deployed real sniffers to collect and analyse BLE public-packet traces in parks, controlled environments, stores and indoor scenarios, to develop *formalisation*. One major challenge is the difference in implementation of the address randomization schemes by different manufacturers. Moreover, vendors have a range of devices supporting various releases. We find a property to classify the device into various groups where the behavior is similar irrespective of manufacturer. The property used is the *frequency of transmitting advertising packets*, F_{Char}, which is characteristic of a device with a maximum variation of 10ms. Based on this property which we take normally distributed, we classify BLE devices into the four *device classes*: 1) *Frequent Emitters*: They represent a highly active device like earbuds with F_{Char} having mean 50 ms and standard deviation 10 ms. We observe from the experiments that these kinds of devices also change their randomized MAC address more often. 2) *Moderate Emitters*: From our experimentation, most smartphones, especially iPhones, fall into this category and have F_{Char} with mean 300 ms and standard deviation 25 ms. 3) *Semi-Moderate Emitters*: This class again mainly includes phones and have F_{Char} with mean 500 ms and standard deviation 25 ms. 4) *Low Emitters*: Smartwatches generally fall in this category with F_{Char} being high having mean 2 s and standard deviation 500 ms. MAC address randomization is the core of BLE privacy and hence we propose to emulate it inside the simulation nodes using the *PrivacyManager* that we propose in Figure 1. GENERATE and RESOLVE modules in the *PrivacyManager* generate and resolve BLE private addresses using IRK (Identity resolving keys). UPDATE and CHECKVALIDATION manage the task of updating MAC after *randomization interval* and finding out if address resolution is needed. We detail the design and implementation of *PrivacyManager* in [MVA21].

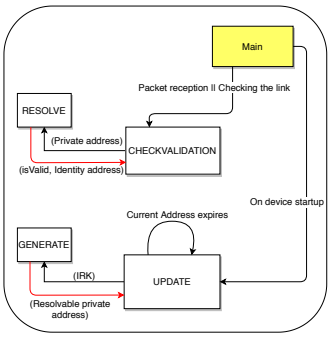
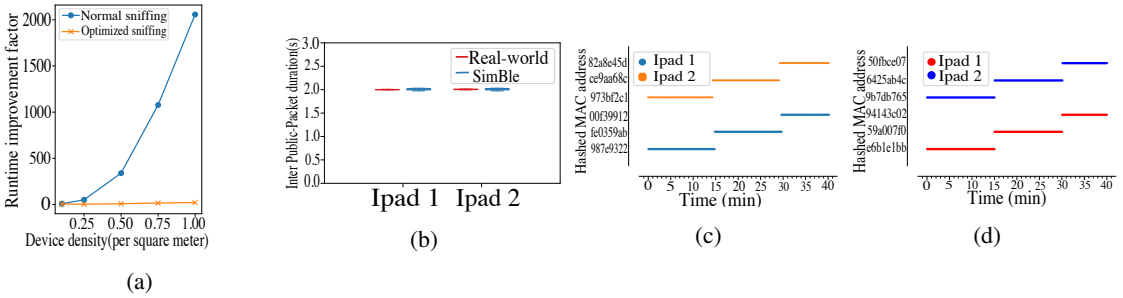


Fig. 1: *PrivacyManager* in *SimBle*

3 Emulating real-world BLE passive-sniffing using *SimBle*

In this section, we first design a simulation environment to emulate real-world conditions and validate it by comparing it to the behaviour of real hardwares. Then, we do the *evaluation* of the robustness of BLE

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



(a) (b) (c) (d)

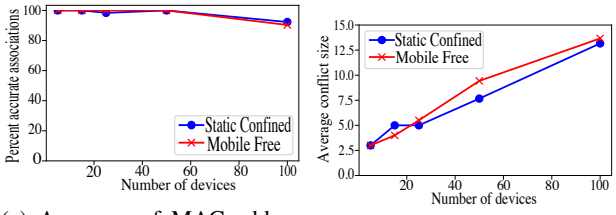
Fig. 2: a) Performance gain in run-time with optimized sniffing in *SimBle*. b) Real-world vs *SimBle* in inter public packet times. MAC address randomization in c) real-world devices, and in d) real-world devices emulated in *SimBle*

privacy provisions by doing a case-study of MAC address association strategy.

3.1 Simulation & Validation

The simulation time explodes with a large number of devices due to the number of simulation events increasing when handling the inter-node public packets although we are interested in the full processing of broadcast packets only at the sniffer. We address this problem by filtering and optimizing the handling of broadcast packets at nodes. We validate the improvement brought by *SimBle* in terms of run-time by increasing the number of devices up to 100 in a 100 square meters grid around the sniffer. We observe in Figure 2a that the optimized sniffing in *SimBle* removes the issue of exponential simulation-times and improves it 2000-folds at high device density of 1 device per square meter.

In *simulation* component, we introduce *Mobile-free* mobility profile in which devices are mobile and are free to leave and enter the sniffing range around the sniffers. We use the *Nakagmi* path loss model and consider the successful BLE transmission range to be around 20 meters. We deliberately take simulation grid as a square of 2500 square meters and place the sniffer in it's middle. We try to mimic human mobility by using a random-walk mobility model with a speed of 1.5 m/s and direction change after every 2 s. This random-walk by nodes in the area makes them move in and out of the sniffing range during the simulation time. We also have *Static-Confined* mobility profile where the devices are static and are always present in the sniffing zone. We validate the overall *formalisation* by comparing the emission of advertising packets and the MAC address changes in BLE 4.1. We aim to show the similarity in the behaviour of real-hardware, iPad, with that of the same device being emulated in *SimBle*. Figure 2b shows that both iPad and *Simble* have practically same inter-public packets times at the sniffer. Similarly, as seen in Figure 2c and 2d, iPads change their MAC addresses after every 15 minutes both in real-world and inside *SimBle*. This validates the working of *Simulation* component of the framework. Finally, use the generated large-scale BLE traces to evaluate user-privacy using *Evaluation* component of *SimBle* framework in the following.



(a) Accuracy of MAC address associations

(b) Average conflict size

Fig. 3: Evaluating MAC address association [JVAF21] for large-scale scenarios using *Simble*

3.2 Evaluating the robustness of BLE privacy provisions

We generate *ground truth* trace by matching each device's generated private addresses to the *Node ID*, which acts as a unique identifier to the device in simulation time. There are many MAC address association solutions in the literature as discussed in the Section 1. We decide to implement and study performances of

[JVAF21] when using *SimBle*, since to the best of our knowledge, it is the only generic BLE MAC address association strategy currently available in the literature. [JVAF21] is sensitive in run-time to the number of simultaneous MAC changes or the *conflict size* and they specify the practical limits to be around 25. We evaluate it using the traces and the *ground truth* generated in *Static-confined* and *Mobile-Free* mobility profile for nodes with variety of device classes in BLE 5.2. We expect the conflict sizes to rise and hence a decrease in accuracy for a large number of devices. Minimum accuracy that we in Figure 3a, is 89% with 100 devices in the sniffing zone for *Mobile-free* mobility-profile. Conflict sizes increase to a maximum value of 13 as seen in the Figure 3b but it is well within the specified limit.

4 Conclusion

MAC address randomization is the backbone of BLE's provisions to safeguard user-privacy. It is necessary for the standard to develop the randomization specifications enough to fend away the existing attacks on device privacy. *Our results shows that the association strategy [JVAF21] fingerprints close to* 90% *of the devices even in highly dense and mobile scenarios. An adversary could setup multiple sniffers strategically and easily track a particular user device.* The lack of *ground truth* in randomized traces and impracticality of large-scale passive trace collection makes the evaluation of user identification based works almost impossible. *All of the existing works based on device-identification using MAC address must be revisited with the introduction of standard's privacy-provisions like private addresses. SimBle* allows researchers could now generate large-scale traces with devices of their interest and use it to validate their works. We intend to extend this framework for WiFi, as it also lacks a framework to generate privacy preserving real-world traces for large-scale scenarios. *SimBle*'s code and usage is detailed here: https://gitlab.inria.fr/mabhishe/simble.

References

[BLS19] Johannes K Becker, David Li, and David Starobinski. Tracking anonymized bluetooth devices. *Proceedings on Privacy Enhancing Technologies*, 2019(3):50–65, 2019.

[CC19] Guillaume Celosia and Mathieu Cunche. Saving private addresses: an analysis of privacy issues in the bluetooth-low-energy advertising mechanism. In *MOBIQUITOUS*, pages 444–453, 2019.

[JVAF21] Loïc Jouans, Aline Carneiro Viana, Nadjib Achir, and Anne Fladenmuller. Associating the randomized bluetooth mac addresses of a device. In *2021 IEEE 18th Annual Consumer Communications Networking Conference (CCNC)*, pages 1–6, 2021.

[MAB+19] Jeremy Martin, Douglas Alpuche, Kristina Bodeman, Lamont Brown, Ellis Fenske, Lucas Foppe, Travis Mayberry, Erik Rye, Brandon Sipes, and Sam Teplov. Handoff all your privacy–a review of apple's bluetooth low energy continuity protocol. *PoPETs*, 2019(4):34–53, 2019.

[MMD+17] Jeremy Martin, Travis Mayberry, Collin Donahue, Lucas Foppe, Lamont Brown, Chadwick Riggins, Erik C Rye, and Dane Brown. A study of mac address randomization in mobile devices and when it fails. *Proceedings on Privacy Enhancing Technologies*, 2017(4):365–383, 2017.

[MVA21] Abhishek Kumar Mishra, Aline Carneiro Viana, and Nadjib Achir. Simble: Generating privacy preserving real-worldble traces with ground truth. *CoRR*, abs/2101.11728, 2021.

[UCF+20] Marco Uras, Raimondo Cossu, Enrico Ferrara, Ovidiu Bagdasar, Antonio Liotta, and Luigi Atzori. Wifi probes sniffing: an artificial intelligence based approach for mac addresses derandomization. In *2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pages 1–6. IEEE, 2020.