2018 IEEE Confs on Internet of Things, Green Computing and Communications, Cyber, Physical and Social Computing, Smart Data, Blockchain, Computer and Information Technology, Congress on Cybermatics

Analyzing the Effect of Bluetooth Low Energy (BLE) with Randomized MAC Addresses in IoT Applications

Golnar Kalantar, Arash Mohammadi, S. Nima Sadrieh

***Abstract*—Based on the rapidly trending field of Location-Based Services, Gateways which are important components of the internet of Things (IoT) are designed to rely on Media Access Control (MAC) addresses of devices at the desired venue in order to locate and track them. However, Apple and similarly Android manufacturers have adopted the strategy of MAC address randomization such that their users' privacy would not be compromised in any sort. This paper, pursued the goal of studying the behavior of MAC Address randomization in iOS devices, finding an answer for the following key question: "*What does randomized MAC addresses mean for BLE-based Location Tracking and Analytics?*" and investigating potential solutions to overcome this challenge and to keep track of all devices at a target venue, while respecting users' privacy. As per our studies, two main approaches are suggested, adopting each will enable Gateways to cope with the MAC address randomization strategies of the iOS devices. The first suggestion is an extension of tracking algorithm, relying on the fact that the frequency of MAC Address randomization is low enough for to keep track of the devices present at the venue, based on the first detected MAC address and its associated device, and the second proposed approach is to incorporate other data fields of the advertised Probe Requests from a target device to enhance localization performance.**

***Index Terms*—MAC Randomization, Indoor Localization, Probe Requests, Bluetooth Low Energy (BLE).**

1. Introduction

With the rapid emergence of Internet of Things (IoT) [1], we are more and more surrounded by smart devices with sensing, processing, and communication capabilities [2]. Although numerous benefits are associated with the evolution of the IoT, several new problems are also emerging. Among all detected problems in this field, the focus of this paper is on indoor localization of smart devices which plays a critical role in various pervasive applications such as medicare [3], smart home [4], vehicular networks [5], and smart museums [6] to name a few. In such application domains, Location-Based Services (LBS) are becoming essential for different businesses and, therefore, are considered as one of the main technological innovations in real-world scenarios ranging from authentication, to place recommendations, geosocial networking, opportunistic networks, and vehicular social networks [7].

Traditional localization and tracking techniques rely on fingerprints of signals obtained from cellular, or Wi-Fi infrastructure. Such systems require employment of a large number of access points (APs) and spending significant cost associated with surveying the site in the training phase, which in turn calls for in-depth research on alternative solutions. To this end, exploring Bluetooth technologies and its received signal strength indicator (RSSI) is an attractive alternative [2]. The Bluetooth Low Energy (BLE), referred to as Bluetooth Smart, is considered as the backbone technology for future indoor navigation as BLE-based tracking provides superior and more fine-grained location information because of possessing several special characteristics including high scan rate, short handshake procedure, very low power consumption, and better signal geometry [8]. Management of the LBSs by utilizing BLE technology depends on Gateways which are important components of the IoT vision used to connect BLE devices to the Internet via a local infrastructure or using a cellular connection. In this work, we used Fathom$^{\text{TM}}$ Gateway [9], [10] which is a new, extensible proximity and context management solution, which brings unparalleled control, intelligence and context integrity to large networks of proximity beacon and IoT devices. Furthermore, Fathom Hub is a new BLE gateway device and transforms stand-alone BLE beacons into a managed proximity ecosystem, reporting on the location and status of all BLE beacons in the coverage area. Tracking via Gateways is possible because BLE-enabled devices routinely transmit a packet called "Probe Request (PR)" to search for nearby networks, and these requests contain the unique Media Access Control (MAC) address of the device, which is known as the best device identifier. However, randomized MAC addresses negatively impact the statistics and insights extracted from scanning the BLE devices, via their MAC addresses, in the venue which is the main motivation of this work.

Randomizing MAC address has been adopted very recently in practice to safeguard the privacy of data transmitted over BLE specially as compromises were made to simplify its protocol for target devices with low computation capabilities. As BLE becomes pervasive with its adoption for IoT and proximity sensing services, it becomes imperative to

- *This work was partially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada "EGP 508086 16".*

978-1-5386-7975-3/18/$31.00 ©2018 IEEE
DOI 10.1109/Cybermatics_2018.2018.00039

secure user's privacy. Several counter-measures have been proposed mainly focusing on the modification of the Wi-Fi service discovery process. For instance, Reference [11] proposed a privacy-preserving Wi-Fi discovery process using cryptographic challenge-responses appended to the PRs. Reference [12] proposed to broadcast different probes depending on the context. Beresford and Stajano [13] proposed to change MAC addresses over time. Reference [14] proposed to remove link-layer identifiers altogether. The aforementioned approaches have not been adopted in practice, instead, very recently, a privacy feature has been adopted that uses short-term and changing MAC addresses under certain conditions [15]. This approach is similar in nature to the concept of mix zones [13], [16]. In other words, in order to increase privacy and security of each associated device, very recently, *scanning behavior has changed to use random, locally administrated MAC addresses, i.e., the MAC address used for scanning may not always be the device's real (universal) address*.

Incorporation of randomized and locally assigned MAC addresses that changes over time negatively impact the performance of the gateways and makes tracking a BLE-enabled device a non-trivial task. In this paper, we provide an answer for the following key question: "*What does randomized MAC addresses mean for BLE-based Location Tracking and Analytics?*". We are particularly interested in analyzing the main features of BLE with MAC randomization, investigate the impact of various critical parameters on its performance, and explore its practical implementation. In brief, the main goals of this work are as follows: (i) Conduct different measurement studies by collecting BLE traffic between different devices and the gateway and monitor packets sent after randomization to discover potential privacy leakages; (ii) Investigate potential solutions to re-identify randomized probes; (iii) Investigate potential methodologies to properly overcome MAC randomization of legitimately associated devices, i.e., to find whether or not it would be possible to link PRs coming from the same associated device using different MAC addresses, and; (iv) Understand the behavior of randomized MAC devices. In this regard, we have conducted comprehensive set of tests based with the goal to collect, analyze, and interpret test results to better understand the behavior of random MAC used by BLE devices; understand/classify the conditions (e.g., sleep mode, connected to Wi-Fi, paired with another device) under which the MAC randomization happens, and; to model the impacts of MAC randomization on LBSs and analytics. To understand the randomization of the BLE MAC address, PRs are captured using a practical gateway from different models and the transmitted MAC address is analyzed in different frames, under various scenarios. For instance, one test was conducted based on enabled cellular voice and data services, but disabled location service. In another test, we turned off the device's cellular data service and observed the behavior in different scenarios, e.g., when the screen is locked/unlocked. Finally, we connected the device to the wireless network and monitor its MAC randomization behavior.

2. Preliminaries

In this section, we briefly introduce and present different key concepts used in the reminder of the paper.

2.1. Bluetooth Low Energy (BLE)

Bluetooth Core specifications (Version 4.0) were successfully extended and merged into specifications of the BLE, marketed as Bluetooth Smart, in 2010, by the incentive of providing users with significantly low power consumption and reduced cost while maintaining a similar communication range. Similar to Bluetooth, the BLE operates in the 2.4 GHz Industrial, Scientific and Medical (ISM) radio band. The BLE operates on 40 channels spaced 2 MHz apart, with center frequencies ranging from 2402 MHz to 2480 MHz. Channels are organized as three advertisement channels and 37 data channels. The BLE uses frequency hopping to avoid interference and to coexist with other devices operating in the ISM band. The three advertisement channels are indexed as 37, 38 and 39.

The mechanism of the BLE advertisement has been considered quite fitting and applicable to the concept of LBS which represents a category of software-level services that uses the data regarding the location of the end-user, and provides a broad variety of services, ranging from shopping to health-care services. Also, many cognition-related studies can be done by datasets of BLE devices, since customer insights can be investigated by studying the customer behaviors. For instance, in marketing applications of LBSs, push notifications are like that virtual "*tap on the shoulder*", taken place in order to complete the in-person shopping experience with the information customer usually gets only while online shopping. Then BLE is a new kind of protocol/standard for devices to send very small packets of data to smartphones and other devices that intercept those Bluetooth signals. The BLE protocol is extremely optimized, battery-wise, as it does not send complicated, big packages of data. In other words, BLE basically sends the data that simply says "*I'm here*", similar to the function of a Lighthouse. With the info retrieved from the data package, we can say that BLE information packages say: "*I'm here, and this is my nature*". To provide a reliable communication channel, the proposed protocol should be based on opportunistic communication windows and rely on frame retransmissions, acknowledgments, and timeout patterns. Without a reliable protocol, a receiver might miss data frames.

2.2. MAC Address Randomization

There are two main modes of communication, and as a result, tracking smartphones. The cellular radio itself is the first and most obvious means of communication, and as a result can be considered as one candidate mode for tracking. The second mode is the 802.11 Wi-Fi protocol which is used by almost all cell phones. Every 802.11 radio on a mobile device possesses a 48-bit link-layer MAC address that is a globally unique identifier for that specific Wi-Fi device. MAC address, being included in every link-layer frame that is sent to or from the device, plays a crucial role for LBS.

A particular type of Wi-Fi packet, referred to as the PR frame, is an especially vulnerable part of Wi-Fi traffic with respect to surveillance, due to continuous broadcasting at a semi-constant rate which makes tracking trivial. The Wi-Fi MAC address in the PR, used to uniquely (anonymously though) identify a device and track the behavioral patterns of the user, is aptitude of providing the business with some insights like the number of devices seen, the number of repeat users, and the number of new users to name a few. To address this privacy issue, different mobile Operating Systems (OSs) have started implementing MAC Address Randomization.

IEEE assigns blocks of addresses to organizations in exchange for a fee in order to guarantee the uniqueness of MAC addresses all over the world. A MAC Address Block Large (MA-L), commonly known as an Organizationally Unique Identifier (OUI), may be purchased and registered with the IEEE, which gives the organization control of and responsibility for all addresses with a particular three-byte prefix. The manufacturer is then free to assign the remaining low-order three bytes (224 distinct addresses) any value they wish when initializing 2 devices, subject to the condition that they do not use the same MAC address twice. In addition to the public, globally unique, and manufacturer assigned MAC address, modern devices frequently use locally assigned addresses. Locally assigned addresses are used in a variety of contexts, including multi-Service Set IDentifier (SSID), configured access points (APs), mobile device-tethered hotspots, and peer-to-peer (P2P) services.

An implication of the IEEE registration system is that it is trivial to look up the manufacturer of a device given its MAC address. Moreover, locally assigned addresses may also be used to create randomized MAC addresses as an additional measure of privacy. Similar to an OUI, a three-byte Company Identifier (CID) prefix can be purchased from the IEEE, with the agreement that assignment from this address space will not be used for globally unique applications. A particularly sensitive privacy issue arises from the manner in which wireless devices identify access points within close proximity. To cope with this privacy concern, both Android and Apple iOS operating systems allow for devices in a disassociated state to use random, locally assigned MAC addresses when performing active scans. Since the MAC address is now random, users gain a measure of anonymity up until they associate with an AP.

In particular, and upon the release of iOS 8.0, Apple introduced MAC address randomization, continuing with minor but valuable updates to the policy across subsequent iOS releases. The initial assumption is that Apple would use an OUI or CID like other manufacturers and simply randomize the least significant 24 bits of the MAC address. *However, based on our studies, it is observed that the MAC addresses randomly generated by iOS devices do not share any common prefix. In fact, they appear to be 10 completely random, including the 24 OUI bits, except for the local bit which is always set to 1 and the multicast bit which is set to 0. In other words, iOS most likely implements true randomization across the entire MAC address. This is interesting given the fact that the IEEE licenses CID prefixes for a price, meaning that Apple is freely making use of address space that other companies have paid for.*

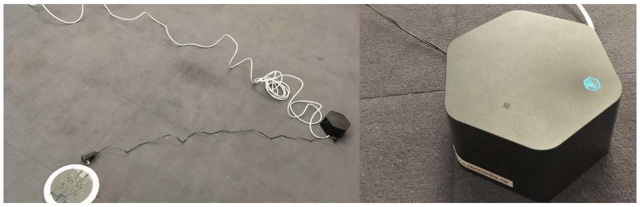


Figure 1. Fathom Hub [9], [10] and its connection to electricity and internet network.

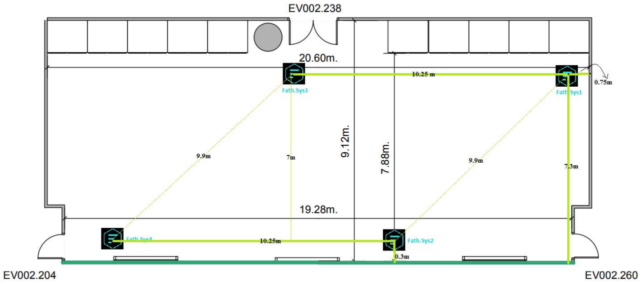


Figure 2. The map of room EV 2.260 and the placement of hubs.

3. Experimental Setup

In this section, we describe the implementation details for setting the experimental result and performing data collection experiments.

3.1. Equipment and Logistics

In the experiments, we used four Fathom Hubs [9], [10] each with 6 antennas, which respectively from the first to the sixth, take one second to look for BLE devices active in the surrounding area. Each hub is capable of detecting BLE

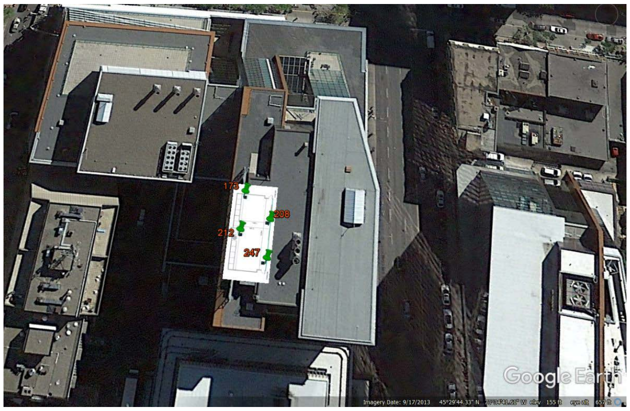


Figure 3. The spots of Fathom Hubs [9], [10] on Google Earth map.

MAC addresses within the radius of 10-15 meters. Fig. 1 illustrates one hub and how it is employed.

There were two main steps to get the hubs set and ready for data collection, i.e., providing network connection for the hubs, and the managing the logistics. In terms of logistics, the first step was to choose an appropriate room/hall, big enough to provide us with a freedom to design the coordination of the hubs. In order to have the hubs perform acceptably accurate, we needed them to be of approximately 10-meter distant from one another, therefore, we needed a place where we could place a quadrilateral with edges and diagonals of at least 10 meters long. Once we found the room which met all our criteria, we tagged the four spots to place each hub at its corresponding spot, consistently for all experiments done throughout this work. The map in Fig. 2 shows the logistics of the venue and hubs' placements. The final step was to manually set the location of the hubs and set the orientations. Fig. 3 illustrates the absolute location of the hubs on Google Earth map. In order to have a clear idea of previous and ongoing work in the field of MAC Address randomization and indoor positioning, Based on our comprehensive literature review, it was noted that *MAC address randomization implemented by Apple is, firstly not sufficiently investigated, and secondly, is severely sensitive to the status of the device*. This point, motivated us to design scenarios of data collection considering different authentic modes of cellphones used by the users in their daily life. For instance, Cellular data being off or on, the cellphone being locked or unlock and active, as will be discussed next.

3.2. Designing Scenarios of Data Collection

At the beginning, considered one iPhone to start with and designed four scenarios, as a result of the rule of product on static or dynamic location of the phone and phone's status which would be idle or actively used by the user. In the consequent data collection sessions, we figured out there is no need to focus on a moving device around the venue, because the movements would only complicate the analysis and *based on the observations, apparently, movement has almost no impact on the randomization*. Therefore, we focused on designing our scenarios on different status of utilities such as Wi-Fi, Cellular Data, Location Services, and also apps which can be active in the background while the phone is idle. Throughout our experiments, we considered the following scenarios:

- *Wi-Fi*: On (not connected to any network), or OFF.
- *Cellular Data*: Off, or On with no app having access to internet, or On with active apps.
- *Location Services*: Off or On.
- *Phone*: Locked with no app open in the background, Locked with an/few app(s) active in the background, or Unlock while the user is normally using it.

It is worth mentioning that there was a notable limitation in designing the scenarios, i.e., the hubs scanning iBeacons uniformly in 3 dimensions, and as the venue was in the middle of a crowded building, we had to consider less crowded days, for instance, weekends, in order to reduce the possibility of non-target iBeacons scanned by the hubs.

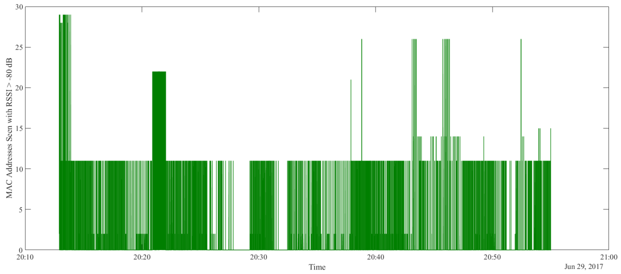


Figure 4. MAC Addresses seen, filtered by RSSI Threshold of -80 dB and more.

3.3. Processing the Collected Data

Our processing was mostly done in MATLAB environment and the implemented processing algorithms consist of the following items:

(i) Distinguishing all the MAC addresses seen during the experiment;
(ii) Following the changes in the MAC addresses over time per antenna;
(iii) Following the changes in the MAC addresses over time per hub, considering the distance from each hub, and;
(iv) Identifying/illustrating events of importance.

Thereafter, the frequency of shift occurred in the string of scanned MAC addresses was compared manually to the designed scenario and figures of MAC address rotation versus time were plotted to verify or reject the hypothesis we had for each experiment.

In order to study the MAC address randomization through the collected datasets, we first detect all the MAC addresses scanned during the data collection, and then we allocate an index to each MAC address (index 0 is always corresponding to "no observation"). Thereafter, we filter the MAC addresses based on the RSSI to make sure the MAC addresses under study are advertised by the target phone, and not randomly scanned from unwanted sources. The Threshold for RSSI filtration is not always the same constant. Based on our experiments, the distance of half a meter and less, is scanned with RSSI of -75 dB or more. One way to find the threshold is to find the indexes corresponding to the MAC addresses seen by all hubs constantly over time. These indices resemble a constant noise added to a target signal. Figs. 4 and 5 are two examples of the trend of MAC Address randomization filtered with two different RSSI thresholds. In this example and based on the log of our data collection corresponding to the Figs. 4 and 5, only MAC addresses at 20:13' and 20:22' were the targets, however, static devices nearby the venue of the data collection, and those passing by the venue randomly cause the random MAC addresses in our dataset.

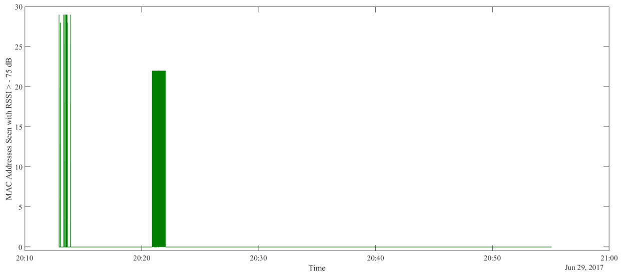


Figure 5. MAC Addresses seen, filtered by RSSI Threshold of -75 dB.

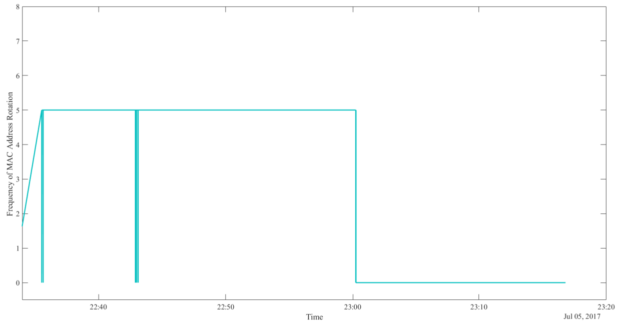


Figure 6. Frequency of MAC address rotation during one data collection.

4. Case studies: Three Sample Sessions

Given the approaches mentioned for studying the datasets collected, the data collection sessions that we will focus on throughout this paper are the followings: (i) E-1: Simplistic Scenario to become acquainted with the behavior of the randomization; (ii) E-2: Experiment of Rotation of Three Phones, and; (iii) E-3: Mitigation of the Randomization Problem with UUID.

4.1. E-1: Simplistic Scenario

This experiment was done using only one hub and one iPhone. One iPhone was placed next to the hub within a distance of 10cm away from it, parallel to antenna 0 in the hub. For approximately the first 10 minutes, Location Services and Cellular Data were off, Wi-Fi was on but not connected to a network. Also, three apps (Locate, Camera, and Settings) were open but the phone was locked while. Then for the second 10 minutes, apps were closed, only Locate remained open. Wi-Fi was also turned off and then, the phone was locked. Afterwards, Wi-Fi was turned off, cellular data was turned on (no app had the access to internet via cellular data), everything else was the same until the end of the session.

Fig. 6 illustrates different aspects of this data collection session. As shown in Fig. 6, the MAC address does not change during the data collection session. Moreover, once the phone has no access to the internet, in other words, while Wi-Fi and Cellular Data are both off, no MAC address is advertised. We believe, during that time, proximity UUID would have been the best option to keep tracking the device. It is worth mentioning that in this scenario we can not comment on the originality of the MAC address observed in this session. Since the MAC address randomization is not too frequent to disable us to continue tracking the device, the MAC address being fake or original, does not play a critical role here.

4.2. E-2: Three Phones Rotation Experiment

The plan for running the experiment, with three iPhones placed by first three hubs, and shifting them clockwise every 10 mins, until all phones have been placed by each hub once. The configuration was repeated with the following status of utilities:

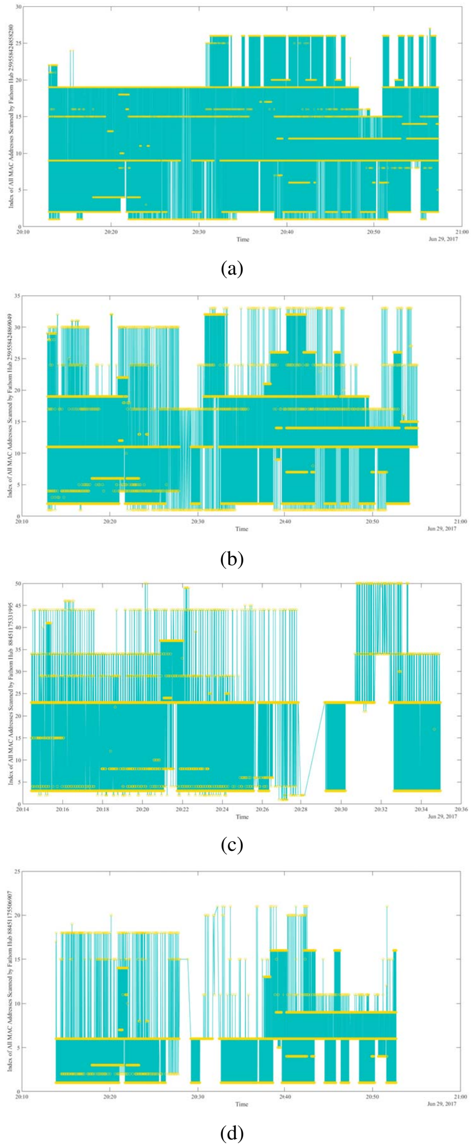
- *Cellular Data*: On (all apps are off).
- *Location Services*: Off.
- *Wi-Fi*: On (but not connected).
- “Locate” running and advertising beacons.

Figs. 7(a)-(d) illustrate all the MAC addresses observed by the four hubs during this session, before RSSI filtration. Figs. 8(a)-(d) illustrate the observed MAC addresses after applying the RSSI filtration.

From Figs. 8(a)-(d), two key points are clearly comprehensible. The first point is the fact that the *frequency of MAC Address randomization is low enough for Fathom's algorithm to keep tracking the devices*. The second point is that needless to figure out whether or not the detected MAC Address is original, the tracking can resume, which is a win-win situation for both LBS Providers and their clients, due to no threat to end-users' privacy. As an example, we followed two of the MAC address indexes in this session, “$59 : f3 : 0f : 9d : 78 : ce$” is trackable from Fathom hub $Fath.Sys.2$ to $Fath.Sys.3$ and “$6f : 6f : 5d : 63 : 72 : 04$” is trackable from Fathom Hub $Fath.Sys.4$ to $Fath.Sys.1$.

4.3. E-3. Mitigation of the Randomization Problem with UUID

After reaching the conclusion that the frequency of MAC address randomization is low enough to make it traceable, we asked our android developer to design an app inspired by the “nRF Connect”, a well-known Android app in this field, to capture information packages sent out by the iPhone advertising beacons, to double check the result we achieved from investigating Fathom datasets. *The developed app is capable of scanning the MAC addresses and UUIDs advertised by the nearby devices, as well as logging them in a word file which is transferable via any social media app on the phone*. Considering the capabilities of our application, we held another session to test our app, while the status of the iPhone utilities was as close as possible to an authentic and real-life situation, for a typical end-user. The Wi-Fi was on and not connected to any Access Point, Cellular Data was on, as well as the Location Services, the phone was locked and on idle mode. The session started with the MAC address captured as “$79 : 16 : 5B : FA : 9C : 95$”. Figure 9



(a)

(b)

(c)

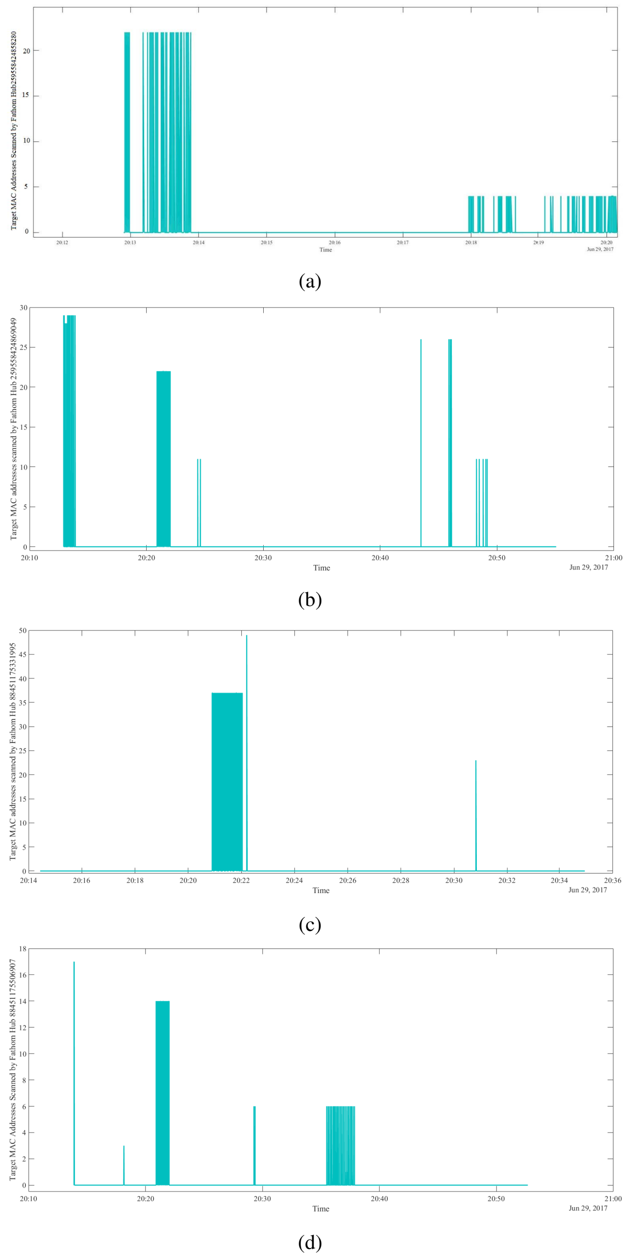
(d)

Figure 7. MAC Addresses scanned by Fathom Hub: (a) Fath.Sys.1. (b) Fath.Sys.2. (c) Fath.Sys.3. (d) Fath.Sys.4.

shows the moments of MAC Address rotation throughout this session of 45 minutes. The information regarding the devices farther away was shown in grey, to show they are not the target of the study. Also, as the time-stamps show, the scan happens with the frequency of more than 1 Hz. that is why we used the Null iBeacon (all digits of UUID set to zero) in "Locate" app for advertisement, so that we would not lose track of our target phone.

5. Findings, and Suggestions

While processing the collected datasets, we faced some interesting points. In advance to processing, we had the assumption that the frequency of randomization is quite high and that it would be an issue for tracking a detected device. However, *we found out the frequency of randomization (in case of happening) is at least 10 minutes, which gives the tracking algorithm enough time to keep a record of other attributes associated with the target MAC address and at the moment of MAC address rotation, the new fake address can replace the previous one*.



(a)

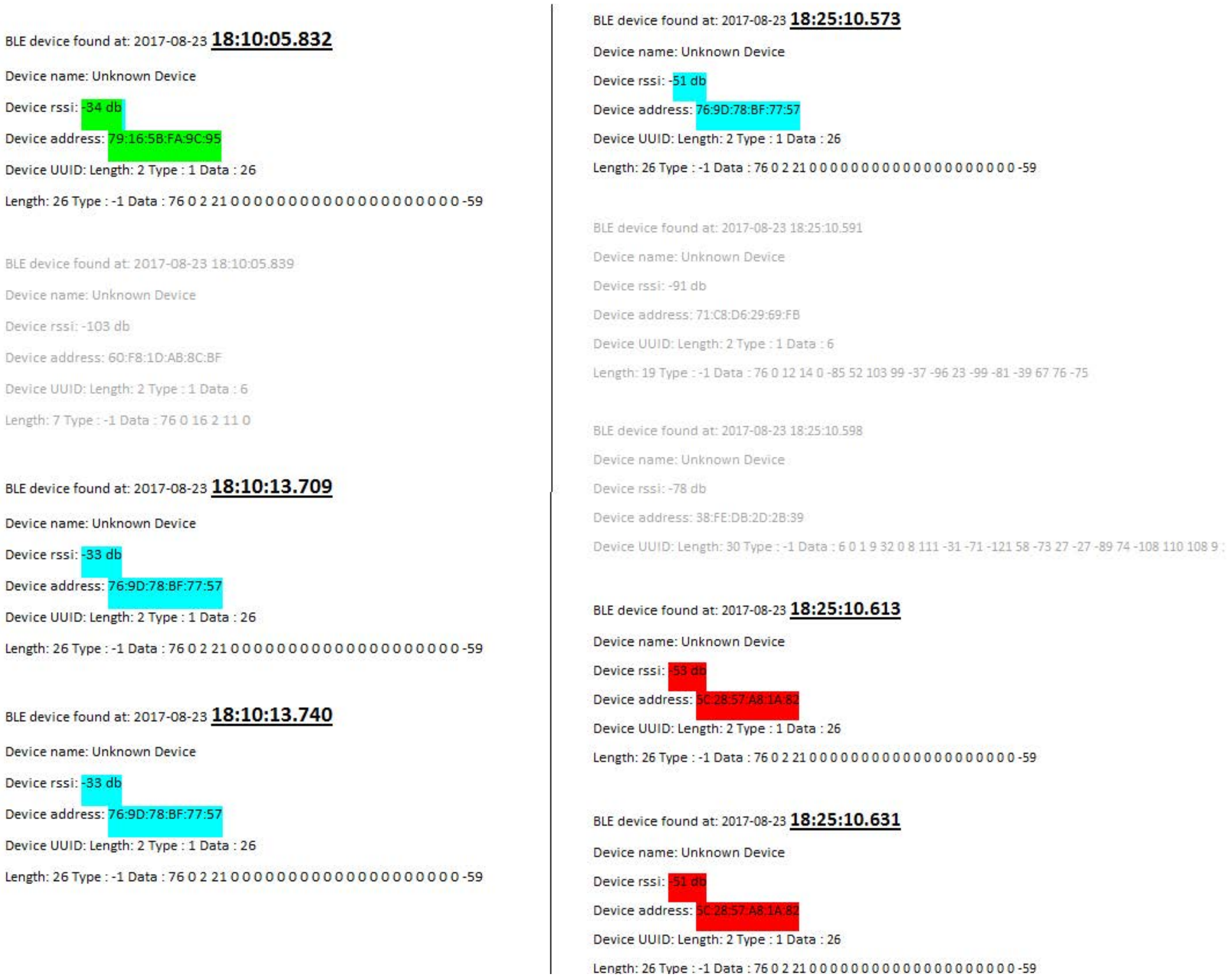
(b)

(c)

(d)

Figure 8. Target MAC Addresses scanned by Fathom Hub: (a) Fath.Sys.1. (b) Fath.Sys.2. (c) Fath.Sys.3. (d) Fath.Sys.4.

Since all the four hubs (which is the minimum number of hubs at any desired venue to ensure the high accuracy of locating and tracking the target devices, according to the instructions given by Dr. Sadrieh) are recording the RSSI associated with every present MAC address, the user's mo-



```
BLE device found at: 2017-08-23 18:10:05.832
Device name: Unknown Device
Device rssi: -34 db
Device address: 79:16:5B:FA:9C:95
Device UUID: Length: 2 Type : 1 Data : 26
Length: 26 Type : -1 Data : 76 0 2 21 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -59

BLE device found at: 2017-08-23 18:10:05.839
Device name: Unknown Device
Device rssi: -103 db
Device address: 60:F8:1D:AB:8C:BF
Device UUID: Length: 2 Type : 1 Data : 6
Length: 7 Type : -1 Data : 76 0 16 2 11 0

BLE device found at: 2017-08-23 18:10:13.709
Device name: Unknown Device
Device rssi: -33 db
Device address: 76:9D:78:BF:77:57
Device UUID: Length: 2 Type : 1 Data : 26
Length: 26 Type : -1 Data : 76 0 2 21 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -59

BLE device found at: 2017-08-23 18:10:13.740
Device name: Unknown Device
Device rssi: -33 db
Device address: 76:9D:78:BF:77:57
Device UUID: Length: 2 Type : 1 Data : 26
Length: 26 Type : -1 Data : 76 0 2 21 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -59
```

```
BLE device found at: 2017-08-23 18:25:10.573
Device name: Unknown Device
Device rssi: -51 db
Device address: 76:9D:78:BF:77:57
Device UUID: Length: 2 Type : 1 Data : 26
Length: 26 Type : -1 Data : 76 0 2 21 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -59

BLE device found at: 2017-08-23 18:25:10.591
Device name: Unknown Device
Device rssi: -91 db
Device address: 71:C8:D6:29:69:FB
Device UUID: Length: 2 Type : 1 Data : 6
Length: 19 Type : -1 Data : 76 0 12 14 0 -85 52 103 99 -37 -96 23 -99 -81 -39 67 76 -75

BLE device found at: 2017-08-23 18:25:10.598
Device name: Unknown Device
Device rssi: -78 db
Device address: 38:FE:DB:2D:2B:39
Device UUID: Length: 30 Type : -1 Data : 6 0 1 9 32 0 8 111 -31 -71 -121 58 -73 27 -27 -89 74 -108 110 108 9 :

BLE device found at: 2017-08-23 18:25:10.613
Device name: Unknown Device
Device rssi: -53 db
Device address: 5C:28:57:A8:1A:82
Device UUID: Length: 2 Type : 1 Data : 26
Length: 26 Type : -1 Data : 76 0 2 21 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -59

BLE device found at: 2017-08-23 18:25:10.631
Device name: Unknown Device
Device rssi: -51 db
Device address: 5C:28:57:A8:1A:82
Device UUID: Length: 2 Type : 1 Data : 26
Length: 26 Type : -1 Data : 76 0 2 21 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -59
```

Figure 9. Snapshots of Dataset collected by our Android application, at the only two moments the MAC address was rotated.

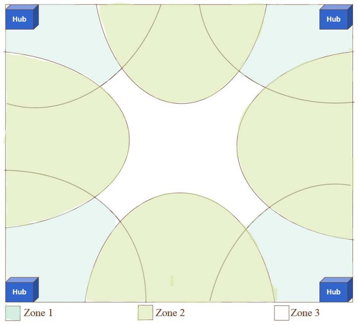


Figure 10. Suggested Zones for hubs.

tion status can be specified using the RSSI (or Angle of Arrival) captured by all four hubs within an ample frequency. Whether the user is static or moving around the venue, and if moving, regardless of what is the velocity/acceleration of their motion, the motion status along with the MAC address can easily help the algorithm not to lose track of the devices present at the venue.

Moreover, inspired by the approach we adopted throughout our scenarios to study the MAC address randomization behavior, LBS companies may deem an approach considering the relative location of target devices, and focus on the hubs nearby in order to have a more accurate positioning result. We suggest a categorized map of venue, with 3 zones as shown in Fig. 10. All the devices detected in the Zone 1 would be tracked with their corresponding hub, and even if they are detected by other hubs, they will be neglected in datasets of remote hubs. All the devices detected in Zone 2, will be monitored by the two corresponding Fathom hubs, and similar to Zone 1, the devices detected in a Zone 2, will be neglected in datasets of remote hubs. Zone 3 will be monitored by all the hubs present at the venue.

Once we observed the aforementioned points, *we figured there might be other fields of information along with the MAC address broadcasted within the probe request frame, which might be helpful*. With this assumption in mind, we asked our android developer to look into the possibilities of capturing other fields of information while the iOS device is advertising. *We studied "nRF Connect" app briefly and as we concluded, without any authorization or permission* *required from the iOS device user, some other useful data fields are contained within the probe request frame. The simplest of them all to take advantage of, is the battery status of the device*.

5.1. Proposed Solutions

We identified two main options to address the problem with iOS MAC address randomization as follows:
a. Continuing the Tracking Algorithm with MAC Address: For now, iOS MAC Address randomization is addressable because firstly, the frequency of randomization is low enough to be trackable, secondly because due to the low frequency of randomization, originality of the MAC address does not play a key role in tracking the devices, and thirdly there are still ways to find the original MAC address through UUID. However, considering the dependency of this approach on MAC address generation policies of the manufacturers, this would not be a robust one.
b. Create a New Unique Fingerprinting Kernel to Identify the Devices Based On More Data Fields of Probe Requests: This approach seems not only to be robust but also futuristic and we highly recommend for LBS companies to shift their focus from merely using MAC address to create a fingerprint data frame for each device in their target venue, based on the MAC address associated with UUID and/or more data fields retrieved from Probe Request frame. The proposed approach, although might take some time to be implemented, is of great value in the sense of the uniqueness of algorithm, robustness against changes in iOS updates in MAC address advertisement, sustainability in the market, and last but not the least, ongoing academic research. Future work of our research group will include working on an extension to our designed app to add the capability of tracking the devices that randomize their MAC address, employing a combination of data fields included in the advertised probe requests.

6. Conclusion

To summarize, the main focus of this paper is to study and analysis the effect of randomized Bluetooth MAC addresses on geospatial statistics and explore the solution to adjust the statistics and remove the effect of randomized MAC addresses. The following tasks were performed in this work to achieve this goal: (i) Conduct comprehensive literature review on Randomized MAC addresses. (ii) Collect real world data and characterize the behavior of randomized MAC addresses. We conducted several data collections with different phones with different settings namely: (a) If a beaconing application is running; (b) If the beaconing application is paired with another device, and (c) different Phone status. We found out that the frequency of randomization is low enough (at least 10 minutes) to remain trackable which is not an obstacle for LBS companies on their way of tracking the present devices at any desired venue. Moreover, there might be other fields of information along with the MAC address broadcasted within the probe request frame, which might be helpful.
performed various scenarios and experiments, and have suggested two potential solutions to address this issue. The

We approached the problem of MAC randomization of iOS devices comprehensively and from different angles, second proposed solution can be an incentive to define a comprehensive project in order to design a kernel of fingerprinting devices by the information in the Bluetooth packets while the device is advertising. This extension can benefit groups working in LBS fields, in terms of robustness of their tracking algorithm, as well as improving the competency of their algorithm in the market.

References

[1] S. Tarkoma and H. Ailisto, "The Internet of Things program: The finnish perspective," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 10-11, Mar. 2013.

[2] Y. Gu and F. Ren, "Energy-Efficient Indoor Localization of Smart Hand-Held Devices Using Bluetooth," *IEEE Access* vol. 3, no. , pp. 1450-1461, 2015.

[3] U. Varshney, *Pervasive Healthcare Computing: EMR/EHR, Wireless and Health Monitoring,* Springer-Verlag, 2009.

[4] X. Li, R. Lu, X. Liang, X. Shen, J. Chen, and X. Lin, "Smart community: An Internet of Things application," *, IEEE Commun. Mag.*, vol. 49, no. 11, pp. 68-75, Nov. 2011.

[5] K. Zheng; H. Wang; H. Li; L. Lei; W. Xiang; J. Qiao; X. Shen, "Energy-Efficient Localization and Tracking of Mobile Devices in Wireless Sensor Networks," *in IEEE Transactions on Vehicular Technology*, In Press, 2017.

[6] S. Alletto et al., "An Indoor Location-Aware System for an IoT-Based Smart Museum," *IEEE Internet of Things Journal*, vol. 3, no. 2, pp. 244-253, April 2016.

[7] P. M. Varela and T. Otsuki Ohtsuki, "Discovering Co-Located Walking Groups of People Using iBeacon Technology," *in IEEE Access*, vol. 4, no. , pp. 6591-6601, 2016.

[8] P. Davidson; R. Piche, "A Survey of Selected Indoor Positioning Methods for Smartphones," in *IEEE Communications Surveys & Tutorials*, In Press, 2017.

[9] https://www.fathomsys.com/

[10] https://www.fathomsys.com/fathom-gimbal-successfully-demonstrate-automated-bluetooth-plant-tracking-cannabis-cultivation-facility/

[11] J. Lindqvist, T. Aura, G. Danezis, T. Koponen, A. Myllyniemi, J. Maki, and M. Roe, "Privacy-preserving 802.11 access-point discovery," *In WiSec*, 2009.

[12] Y.S. Kim, Y. Tian, L.T. Nguyen, and P. Tague, "LAPWiN: Location-Aided Probing for Protecting User Privacy in Wi-Fi Networks," *S&P poster*, 2013. [17] B. Konings, C.

[13] A.R. Beresford and F. Stajano, "Mix zones: User privacy in locationaware services", *In Pervasive Computing and Communications Workshops,* 2004.

[14] B. Greenstein, D. McCoy, J. Pang, T. Kohno, S. Seshan, and D. Wetherall, "Improving wireless privacy with an identifier-free link layer protocol," *MobiSys*, 2008.

[15] J. Freudiger, "How talkative is your mobile device?: an experimental study of Wi-Fi probe requests," *ACM Conference on Security & Privacy in Wireless and Mobile Networks,* 2015.

[16] J. Freudiger, M. Raya, M. Felegyhazi, P. Papadimitratos, and J.P. Hubaux, "Mix-zones for location privacy in vehicular networks," *Workshop on wireless networking for intelligent transportation systems (Win-ITS),* 2007.

[17] B. Bloessl, C. Sommer, F. Dressler, and D. Eckho, "The scrambler attack: A robust physical layer attack on location privacy in vehicular networks," *ICNC,* 2015.

[18] M. Vanhoef, C. Matte, M. Cunche, L. Cardoso, and F. Piessens, "Why MAC Address Randomization is not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms, *ACM AsiaCCS*, 2016.