IEICE Communications Express, Vol.13, No.3, 64–67

ComEX

LETTER

A device identification method from BLE advertising packets with randomized MAC addresses based on regression of received signal strength

Shuhei Akiyama[1, a)] **and Yoshiaki Taniguchi**[2, 3, b)]

Abstract In this letter, we propose a device identification method from observed Bluetooth Low Energy (BLE) advertising packets for tracking BLE devices even if their MAC addresses are changed periodically and randomly. In our proposed method, the combination of MAC addresses is formulated as a linear assignment problem. In addition, in a cost function of linear assignment, we combine two types of cost: time-based cost and received signal strength-based cost, which is calculated based on regression of received signal strength. Through experimental evaluations, we confirmed that the accuracy of our proposed method is the highest compared to traditional methods.

Keywords: Bluetooth Low Energy, randomized MAC address, identification, advertising packet, tracking, privacy
Classification: Network

1. Introduction

The number of mobile devices that communicate using Bluetooth Low Energy (BLE) has been increasing. Many BLE devices such as mobile devices enabling Find My [1], lost prevention or tracking tags [2], smartphone applications [3], broadcast BLE advertising packets periodically. By observing MAC addresses of BLE advertising packets, device tracking can be accomplished [4, 5]. For example, by installing multiple monitoring devices within a facility and observing BLE advertising packets, it is possible to know how mobile devices move within the facility. However, MAC address in packets sent by BLE devices are often randomized periodically to improve the privacy of device users.

To track BLE devices even if their MAC addresses are randomly changed, there are some researches to identify whether the observed packets are sent from the same device [6, 7, 8, 9, 10]. In Ref. [9], the authors proposed a device identification method using time difference between the time when the MAC address was first observed and the time when the MAC address was last observed. They formulated the method as a linear assignment problem [11] with the time difference as a cost function. Although they only consider the time difference, we consider that the performance of the method can be improved by also using the received signal strength that can be obtained at the time of packet reception.

In this letter, we propose a device identification method from BLE advertising packets based on regression of received signal strength. Here, the initial study of this work was presented previously [8][1]. Similar to the traditional method, our proposed method also formulates the combination of MAC addresses as a linear assignment problem. On the other hand, our proposed method uses both time difference and difference between the actual received signal strength and the received signal strength estimated by regression in a cost function unlike the traditional method only uses time difference. We evaluate our proposed method using actual data obtained through experiments.

The rest of this letter is organized as follows. In section 2, we explain related work. Next, we propose a device identification method from BLE advertising packets with randomized MAC addresses in section 3. Then, we evaluate our proposed method in section 4. After that, we discuss how to countermeasure against tracking using our proposed method in section 5. Finally, we conclude this letter in section 6.

2. Related work

In Ref. [6], the authors proposed a method, called the address carryover algorithm, for identifying BLE devices with randomized MAC addresses. This method can be used when BLE devices have vulnerability that the timing of the update of an identification token, which is device-specific information separate from the MAC address, is not synchronized with the timing of MAC address randomization. Here, if static identifiers such as counters or UUIDs are included in BLE advertising packets, MAC address randomization is meaningless [12]. On the other hand, in Ref. [7], we proposed a method to identify devices without using device-specific values for devices that do not move. This method performs identification based on the change timing of the MAC address and the proximity of the received signal strength identifier (RSSI).

In Ref. [9], the authors formulated a device identification method as a linear assignment problem. They use time dif-

[1] Graduate School of Science and Engineering, Kindai University, Higashiosaka, Osaka 577-8502, Japan
[2] Faculty of Informatics, Kindai University, Higashiosaka, Osaka 577-8502, Japan
[3] Cyber Informatics Research Institute, Kindai University, Higashiosaka, Osaka 577-8502, Japan
[a)] 2233340433n@kindai.ac.jp
[b)] y-tanigu@info.kindai.ac.jp

DOI: 10.23919/comex.2023XBL0157
Received November 16, 2023
Accepted December 12, 2023
Publicized January 15, 2024
Copyedited March 1, 2024

[1] In this letter, we improved our proposed method to consider both time and received signal strength in a cost function, and conducted new data acquisition experiments and evaluations. The proposed method in Ref. [8] appears as a comparative method in the evaluation in section 4.

ference between the time when the MAC address was first observed and the time when the MAC address was last observed, as a cost function. In Ref. [8], which is the initial study of this work, we also formulate the combination of MAC addresses as a linear assignment problem. Unlike the traditional method, we use RSSI difference between the actual RSSI and the estimated RSSI obtained by regression as a cost function. In our proposed method in this letter, we use both time difference and RSSI difference as a cost function. We conduct comparative evaluation of our proposed method using actual data obtained through experiments.

3. Proposed method

3.1 Assumed environment

In this letter, we assume that BLE advertising packets are captured on a monitoring device such as a laptop PC in an environment where multiple moving BLE devices are in the vicinity. The BLE device is assumed to be a smartphone that broadcasts BLE advertising packets periodically, such as when an application such as Find My [1] is installed. The MAC address used by the BLE device is assumed to be randomly changed at a regular interval. The same device is estimated from the captured data using our proposed method.

Hereafter, the i-th source MAC address obtained from the captured BLE packets is denoted as $a_i \in \mathcal{A}$, where $\mathcal{A}$ is a set of MAC addresses. The first received time of packets with MAC address a_i is denoted as t_i^{first} and the last received time of that is denoted as t_i^{last}. When a packet with MAC address a_i is received at time t, the RSSI is denoted as $r_i(t)$.

3.2 Formulation of the method

In our proposed method, device identification is formulated as the following linear assignment problem. By solving this problem, same devices with different MAC addresses are identified.

$$\begin{aligned} \text{minimize} \quad & \sum_{a_i \in \mathcal{A}} \sum_{a_j \in \mathcal{A}} c(a_i, a_j) x(a_i, a_j) \\ \text{subject to} \quad & \sum_{a_j \in \mathcal{A}} x(a_i, a_j) = 1, \qquad a_i \in \mathcal{A} \\ & x(a_i, a_j) \in \{0, 1\}, \qquad a_i \in \mathcal{A}, a_j \in \mathcal{A} \end{aligned} \tag{1}$$

where function $x(a_i, a_j)$ is defined as one if the MAC address a_i is estimated to have changed to MAC address a_j and zero otherwise.

Function $c(a_i, a_j)$ is a cost function based on the time difference $\tau_{i,j}$ and RSSI difference $\rho_{i,j}$, expressed as follows.

$$c(a_i, a_j) = \begin{cases} \sqrt{\tau_{i,j}^2 + (\alpha \rho_{i,j})^2} & \text{if } t_i^{\text{last}} \le t_j^{\text{first}} \le t_i^{\text{last}} + T \\ \infty & \text{otherwise} \end{cases}. \tag{2}$$

Here, T is a time range to search for a new MAC address when reception from a device with a certain MAC address ends. α is a parameter that determines how much importance is given to the RSSI difference in the cost function. When α is set to zero, our proposed method is similar to the traditional method [9].

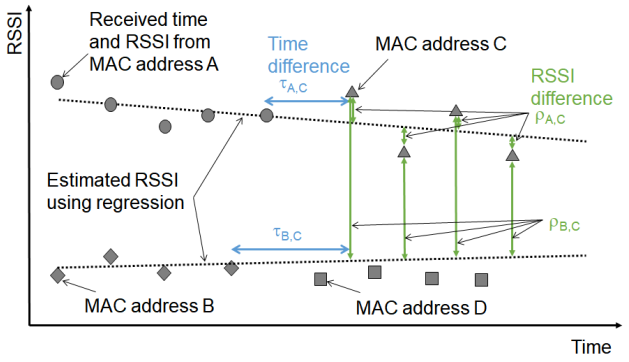

Fig. 1 Example of device identification

In the function, time difference $\tau_{i,j}$ is calculated as follows.

$$\tau_{i,j} = t_j^{\text{first}} - t_i^{\text{last}}. \tag{3}$$

In addition, RSSI difference $\rho_{i,j}$ is calculated as follows.

$$\rho_{i,j} = \frac{1}{|\mathcal{T}_j|} \sum_{t \in \mathcal{T}_j} |r_j(t) - \hat{r}_i(t)|, \tag{4}$$

where $\mathcal{T}_j$ is a set of packet reception time from the device with MAC address a_j between time t_j^{first} and $t_j^{\text{first}} + I$. Here, I is a parameter that determines the duration for evaluating the RSSI difference.

Function $\hat{r}_i(t)$ is an estimated RSSI function corresponding to MAC address a_i at time t. The estimated RSSI function $\hat{r}_i(t)$ is obtained using regression from the captured packets between time $t_i^{\text{last}} - I$ and t_i^{last}. We note here that we tried several regression methods, however there was no significant difference in performance. Therefore, we simply use linear regression for estimation in this letter.

An example operation of our proposed method is shown in Fig. 1. In the figure, MAC addresses A and B are simultaneously changed to other MAC addresses C and D. The RSSI estimations for MAC addresses A and B are shown as dotted lines. In this case, the cost between MAC addresses A and C is smaller than the cost between MAC addresses A and D. Similarly, the cost between MAC addresses B and D is smaller than the cost between MAC addresses B and C. Therefore, MAC addresses A and C and MAC addresses B and D are estimated as the same device.

4. Experimental evaluations

4.1 Packet capture experiments

To verify the performance of our proposed method, packet capture data was first obtained through experiments. In the experiment, Raspberry Pi 4 (OS: Raspbian 10 Buster) was used as a monitoring device, and moto g7 power XT1955-7 (OS: Android 9) was used as a smartphone. In the smartphone, nRF Connect for Mobile Version 4.26.1 was used to broadcast BLE advertising packets. Here, because it is difficult to obtain data on MAC address changes while moving, in this letter we collected packet capture data when devices use fixed MAC addresses. In the evaluation, the MAC addresses were changed in a pseudo manner.

In the experiment, a smartphone user crossed the vicinity of the monitoring device and the monitoring device cap-

tured BLE advertising packets sent from the smartphone. We obtained packet capture data for a total of 20 patterns, considering 10 walking routes and 2 holding patterns: holding the smartphone in the hand and putting it in the pocket.

In this letter, we simulate a change in MAC address by changing the MAC address after a random timing in the captured data. Here, the start and last I seconds of the capture data are excluded from the timing of MAC address change. In addition, by combining randomly selected M data from the 20 captured data, we reproduce the captured data when BLE advertising packets from M BLE devices are received. Here, to adjust the difficulty of device identification, we introduce a new parameter, the address change interval D. In the evaluation, the M captured data are shifted and combined so that MAC address changes occur at random timing during interval D. For example, if $D = 0$ is used, all MAC addresses change at the same time. On the other hand, $D = 300$ corresponds to a situation where each MAC address changes once every 300 seconds.

4.2 Evaluation conditions

We implemented our proposed method using python. The linear regression were implemented using *scikit-learn*, which is a library for machine learning. In addition, a numerical analysis library *scipy* was used to solve the linear assignment problem. For the parameters of our proposed method, we used $T = 6$ and $I = 5$ [8]. In addition, based on the preliminary experimental evaluations, we determined parameter α so that the width of distribution of time difference and the width of distribution of RSSI difference overlapped by 90%. As a result, $\alpha = 0.14$ was used.

For comparison purpose, we also conducted experiments using two comparative methods: the time-based method (similar to the traditional method [9]) and the RSSI-based method [8]. In the time-based method, time difference $\tau_{i,j}$ is used as the cost function (2). On the other hand, in the RSSI-based method, RSSI difference $\rho_{i,j}$ is used as the cost function.

As the evaluation metric, we used the identification accuracy. The identification accuracy is defined as the number of times a correct identification was made divided by the number of times an address change occurred, multiplied by 100. In the following sections, all results are average of 3000 evaluations.

4.3 Evaluation results

Figure 2 shows the identification accuracy when the number of devices M is changed. As shown in the figure, it can be seen that the identification accuracy decreases as the number of devices increases. In addition, the identification accuracy of our proposed is the highest among three methods. Furthermore, the difference in identification accuracy between methods widens as the number of devices increases. For example, when the number of devices is $M = 20$, the identification accuracy is about 92% for our proposed method, about 90% for the RSSI-based method, and about 88% for the time-based method.

Figure 3 shows the identification accuracy when the address change interval D is changed. As shown in the figure,

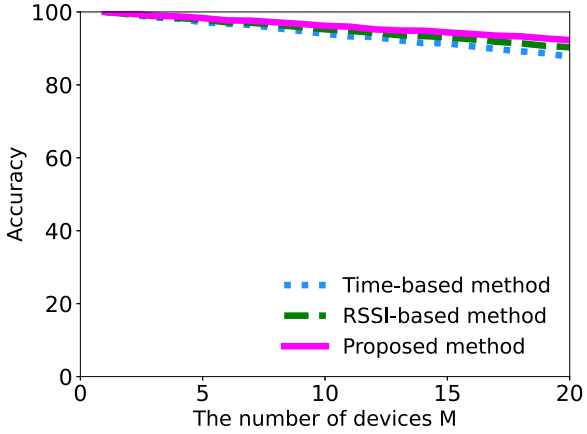

Fig. 2 Accuracy by changing the number of devices ($D = 300$).

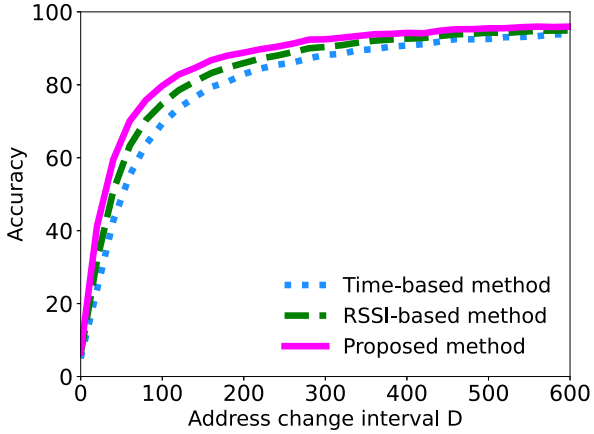

Fig. 3 Accuracy by changing address change interval D ($M = 20$).

it can be seen that the identification accuracy increases as the address change interval increases. This is because increasing the address change interval D reduces the possibility that multiple devices change their MAC addresses almost at the same time. In addition, as shown in Fig. 3, the identification accuracy of our proposed is the highest among three methods. Furthermore, the difference in identification accuracy between methods widens as the address change interval decreases except $D = 0$.

These results indicate that using both time difference and RSSI difference as a cost function is more effective than using each alone. In addition, our proposed method can achieve higher accuracy than other methods when identification becomes difficult, such as when the number of devices increases or the address change interval decreases. However, the evaluation in this letter was conducted in a small-scale environment, and future evaluations in a large-scale environment is necessary.

5. Discussion on privacy protection

Our proposed method can be used for tracking BLE devices. However, in some cases, tracking BLE devices is considered as an invasion of privacy. Therefore, in this section, we discuss what measures can be taken in applications that use BLE advertising packets to reduce the device identification performance of our proposed method.

First, many identification methods, not just our proposed method, perform identification by taking advantage of the fact that the previously used MAC address is no longer used after the MAC address changes. For example, in Fig. 1, MAC addresses A and B are not used after MAC addresses changed to C and D. Therefore, one way to make identification difficult is to use old and new MAC addresses together for a certain period of time, if the application allows it.

Next, as shown in Fig. 3, if the MAC addresses of the devices change simultaneously at short duration, the identification accuracy decreases in all methods. Therefore, for example, in a large-scale application, if the address change timing can be changed so that the MAC addresses change synchronously, the identification accuracy decreases.

Finally, our proposed method uses RSSI for identification. Therefore, if the transmission power is randomly changed within a range allowed by the application, the identification accuracy using our proposed method may decrease.

6. Conclusion

In this letter, we proposed a device identification method from observed Bluetooth Low Energy (BLE) advertising packets for tracking BLE devices even if their MAC addresses are changed periodically and randomly. In our proposed method, the combination of MAC addresses is formulated as a linear assignment problem. In addition, we used not only time difference but also RSSI difference as a cost function. Through experimental evaluations, we confirmed that the identification accuracy of our proposed method is the highest compared to traditional methods. By using our proposed method, the identification accuracy is 92% when the number of devices is 20. We also discuss privacy protection against our proposed method.

As future research, we plan to evaluate our proposed method in a large-scale environment, where a variety of BLE devices move in different directions and multiple monitoring devices are deployed in the environment.

Acknowledgments

This work was partly supported by JSPS KAKENHI Grant Number 19K11934 and 23K11091. The authors would like to thank Mr. Ryoya Morimoto for his help at the early stage of this research.

References

[1] Apple, “Find My network accessory specification,” Sept. 2020.
[2] Apple, “AirTag,” https://support.apple.com/kb/SP840, accessed Nov. 15, 2023.
[3] Google and Apple, “Exposure Notification Bluetooth specification,” https://blog.google/documents/70/Exposure_Notification_-_Bluetooth_Specification_v1.2.2.pdf, April 2020.
[4] A.S. Ja’afar, K. Suseenthiran, K.M. Saipullah, M.Z.A.A. Aziz, A.W.Y. Khang, and A. Salleh, “Development of real-time monitoring BLE-LoRa positioning system based on RSSI for non-line-of-sight condition,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 2, pp. 972–981, May 2023. DOI: 10.11591/ijeecs.v30.i2.pp972-981
[5] P. Locatelli, M. Perri, D.M.J. Gutierrez, A. Lacava, and F. Cuomo, “Device discovery and tracing in the Bluetooth Low Energy domain,” *Computer Communications*, vol. 202, no. 15, pp. 42–56, March 2023. DOI: 10.1016/j.comcom.2023.02.008
[6] J.K. Becker, D. Li, and D. Starobinski “Tracking anonymized Bluetooth devices,” Proc. Proceedings of Privacy Enhancing Technologies 2019, vol. 2019 no. 3 pp. 50–65, July 2019. DOI: 10.2478/popets-2019-0036
[7] S. Akiyama, R. Morimoto, and Y. Taniguchi, “A study on device identification from BLE advertising packets with randomized MAC addresses,” Proc. IEEE ICCE Asia 2021, pp. 1–4, Nov. 2021. DOI: 10.1109/ICCE-Asia53811.2021.9641870
[8] S. Akiyama and Y. Taniguchi, “Device identification in BLE packets from moving devices with randomized MAC addresses,” Proc. IEEE ICCE Asia 2023, pp. 1–4, Oct. 2023. DOI: 10.1109/icce-asia59966.2023.10326401
[9] L. Jouans, A.C. Viana, N. Achir, and A. Fladenmuller, “Associating the randomized Bluetooth MAC addresses of a device,” Proc. IEEE CCNC 2021, pp. 1–6, Jan. 2021. DOI: 10.1109/CCNC49032.2021.9369628
[10] T. Despres, N. Davis, P. Dutta, and D. Wagner, “DeTagTive: Linking MACs to protect against malicious BLE trackers,” Proc. SNIP2+ 2023, pp. 1–7, Sept. 2023. DOI: 10.1145/3609396.3610544
[11] S. Martello and P. Toth, “Linear assignment problems,” *North-Holland Mathematics Studies*, vol. 132, pp. 259–282, 1987. DOI: 10.1016/S0304-0208(08)73238-9
[12] G. Celosia and M. Cunche, “Saving private addresses: An analysis of privacy issues in the Bluetooth-Low-Energy advertising mechanism,” Proc. MobiQuitous 2019, pp. 444–453, Nov. 2019. DOI: 10.1145/3360774.3360777