MAC アドレスがランダム化された BLE パケットからの同一機器推定手法の改良と評価

Improvement and evaluation of a method for estimating same device by BLE packets from moving devices with randomized MAC addresses

秋山 周平 † 谷口 義明 ‡,§
Shuhei Akiyama Yoshiaki Taniguchi

1. はじめに

近年、IoT やスマートフォンの普及により電力消費の少ない BLE (Bluetooth Low Energy) で通信を行う機器が増加している。また、COCOA (COVID-19 Contact-Confirming Application) などの新型コロナウイルス接触確認アプリの普及により BLE パケットを発信するスマートフォンの数が増加している。BLE では、ネットワークに接続しようとしている機器が発信するパケットに含まれる MAC アドレスは端末利用者のプライバシー向上のためランダム化されることが多い。その際、端末は自身が持つ固有の MAC アドレスの代わりにランダムに生成した MAC アドレスを用いてネットワークに接続する。また、その MAC アドレスは一定間隔で更新される。そのため、MAC アドレスに基づいて端末数推定や端末追跡を行う場合に端末数推定に誤差が生じたり、長時間同一端末のパケットを追跡できないなどの影響が生じる。

MAC アドレスがランダム化された BLE 端末の追跡手法としてアドレスキャリーオーバーアルゴリズム [1] がある。この手法では、BLE パケットから得られる端末固有の値と MAC アドレスのランダム化タイミングが同期されていない脆弱性を利用している。しかし、本アルゴリズムが利用している脆弱性は将来的に改善されることが想定される。これまでに我々は端末固有の値を使用することなく端末の同定を行う手法を検討してきた [2, 3]。[3] で提案した手法では受信したパケットの電波受信強度 RSSI (Received Signal Strength Indicator) の回帰に基づいて端末の同定を行なっていた。しかし、この手法は単純な手法とほぼ変わらない同定精度であり、改善の余地が残されていた。

そこで、本研究では [3] で提案された手法の改良を行う。本研究でもこれまでの検討 [3] と同様に、ノート PC 等で周囲にある BLE 端末からの BLE パケットを長期間キャプチャし、得られた MAC アドレスの同定を行うことを想定する。BLE 端末としては BLE パケットをキャプチャできる機会が最も多いと考えられるスマートフォンを想定する。MAC アドレスの組み合わせに対して最後に受信した時刻とはじめて受信した時刻が一定範囲内で、かつ、電波受信強度 RSSI の変動が類似する MAC アドレスの組を同じ端末が使用する MAC アドレスの候補とする。提案手法では、候補となる MAC アドレスの組み合わせを線形割り当て問題として定式化し、その問題を解き、その結果に基づいて MAC アドレスの同定を行う。本研究でも、これまでの検討と同じく、COCOA パケットを使った実験により、提案手法の有効性を評価する。

本論文では、以降、2 章で関連研究について述べ、3 章で提案手法の説明を行う。その後、4 章でその評価結果を述べ、5 章でまとめと今後の課題を述べる。

† 近畿大学大学院総合理工学研究科, Graduate School of Science and Engineering, Kindai University
‡ 近畿大学情報学部, Faculty of Informatics (KDIX), Kindai University
§ 近畿大学情報学研究所, Cyber Informatics Research Institute, Kindai University

2. 関連研究

MAC アドレスがランダム化された BLE 機器の同定手法としては、アドレスキャリーオーバーアルゴリズムと呼ばれる手法がある [1]。この手法では MAC アドレスとは別の端末固有の情報である識別トークンがアドバタイジングパケットから得られ、その値の更新のタイミングと MAC アドレスランダム化のタイミングが同期されていないことを利用して端末の追跡を行う。しかし、この手法での端末の追跡は MAC アドレスのランダム化タイミングと同期してメッセージのペイロードを更新することで防ぐことが可能であるとされている。また、デバイスによってはアドバタイジングパケットに含まれるカウンタや UUID などのの静的識別子によってアドレスのランダム化方式が無意味になる場合があるとされている [4]。

一方、端末固有の値を使わない方法として、MAC アドレスが変更されるタイミングを使用し、タイミングが重なる場合には線形割り当て問題 [5] を使用する手法 [6] がある。我々も端末固有の値を使わず端末を同定する手法を提案してきた [3]。我々の手法では受信時刻に加えて RSSI の情報を用いる。しかし、これまでの検討では 1 章で述べたように単純な手法とほぼ変わらない精度しか達成できず、改善の余地が残されていた。本研究では、精度を改善するための手法の改良の検討および評価を行う。

3. 提案手法

3.1 想定環境

本研究では文献 [3] と同様の環境を想定する。図 1 のように複数台の移動する BLE 端末が周囲にある環境においてノート PC などのモニタリング端末上で BLE パケットをキャプチャし、キャプチャしたデータから同一機器を推定することを想定する。BLE 端末としては、COCOA 等のアプリケーションが導入されているなど定期的に BLE パケットをブロードキャスト送信するスマートフォンを想定する。また、BLE 端末が使用する MAC アドレスは一定間隔ごとに変化するものとする。

図 2 は 2 台の BLE 端末が当初 MAC アドレス A、MAC アドレス B を使用し、途中、ほぼ同じタイミングで MAC アドレスが変化した場合にモニタリング端末でキャプチャされたパケットの RSSI の推移の一例である。MAC アドレスの変化以

図 1　想定環境

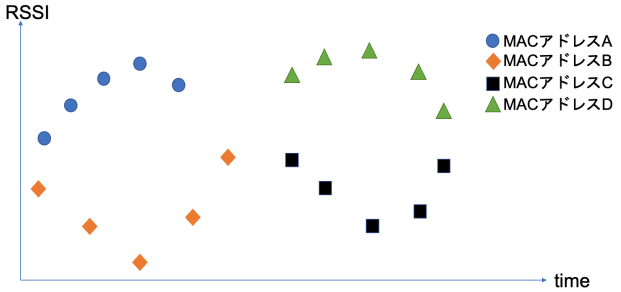

図 2　キャプチャしたパケットの RSSI の推移の例

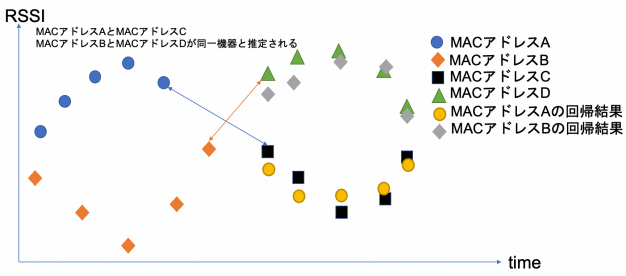

図 3　図 2 における同一機器推定

降は MAC アドレス A、MAC アドレス B を持つパケットは受信されない。BLE 端末が継続的にパケットを送信している場合、その後新たな MAC アドレスを持つパケットを受信する。この時、ほぼ同じタイミングで MAC アドレスが変化し、新たな MAC アドレスとして複数の候補となるアドレスがある場合、どの MAC アドレスがどの以前のアドレスから切り替わったアドレスかは自明ではない。図 2 の例の場合、MAC アドレス A、MAC アドレス B の両方において MAC アドレス C、MAC アドレス D の 2 つの候補がある。どの MAC アドレスがどの MAC アドレスに切り替わったかを推定する手法を本章では提案する。

以降、キャプチャデータから取得した送信元 MAC アドレスを $a_i \in \mathcal{A}$ と表記する。$\mathcal{A}$ は MAC アドレスの集合である。また、それぞれの MAC アドレス a_i の初回受信時刻を t_i、最終受信時刻を t'_i と表記する。さらに、時刻 k に MAC アドレス a_i を持つ端末からのパケットを受信した場合、その時の RSSI を $r_{i,k}$ と表記する。

3.2　同一機器推定手法

提案手法では、同一機器推定を下記の線形割り当て問題として定式化する。この線形割り当て問題を解き、その結果に基づいて、MAC アドレスの同定を行う。

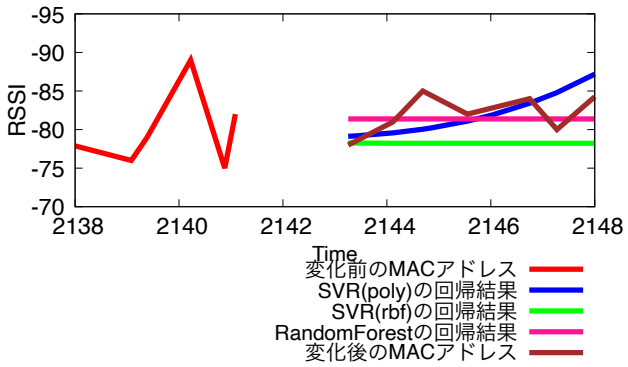

図 4　さまざまな回帰手法による回帰結果の例

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

ここで、$x(a_i, a_j)$ は変数で、MAC アドレス a_i が MAC アドレス a_j に変化したと推定した場合は 1、そうでなければ 0 となる。$c(a_i, a_j)$ は受信時刻と RSSI に基づくコスト関数であり、下記で表される。

$$
c(a_i, a_j) = \begin{cases} \bar{r}_{i,j} & \text{if } t'_i \le t_j \le t'_i + T \\ \infty & \text{otherwise} \end{cases} \tag{2}
$$

なお、T は閾値である。$\bar{r}_{i,j}$ は、MAC アドレス a_j の開始 I 秒間にキャプチャされたパケットから得られる RSSI $r_{j,k}$ （k は $t_j \le k \le t_j + I$ を満たす MAC アドレス a_j のパケット受信時刻）と、時刻 k における MAC アドレス a_i に対応する推定 RSSI $\hat{r}_{i,k}$ との絶対差の平均であり、下記で表される。

$$
\bar{r}_{i,j} = \frac{\sum_k |r_{j,k} - \hat{r}_{i,k}|}{K} \tag{3}
$$

ここで、推定 RSSI $\hat{r}_{i,k}$ は MAC アドレス a_i の末尾 I 秒間のキャプチャパケットから推定するものとする。また、K は MAC アドレス a_j の開始 I 秒間のパケットの総数である。

本稿では、推定のために受信時刻を説明変数、RSSI を目的変数とした回帰を行う。[3] では回帰手法として線形回帰、SVR (rbf カーネル)、ランダムフォレストを用いていたが、本稿の評価においては、回帰手法として、最も単純な回帰方法である線形回帰、非線形回帰手法である SVR (poly カーネル) を使用する。[3] においては非線形な回帰を行うことを目的として SVR (rbf カーネル) とランダムフォレストを用いていた。だが実際は学習データが少ないために殆ど線形的な回帰しか出来なかった。図 4 にその様子を示す。図 4 のように今回は少ない学習データでも非線形な回帰が可能な SVR (poly カーネル) を代わりに用いた。

本手法の動作例を説明する。例えば図 2 の場合、MAC アドレス A、MAC アドレス B の回帰を行うと図 3 のような状態となる。この場合に MAC アドレス A において候補のアドレ

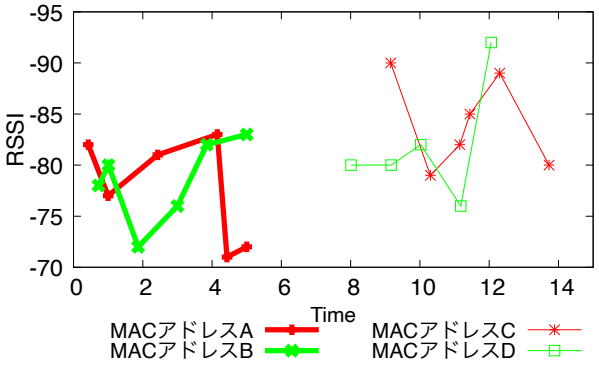

図 5　MAC アドレスが変化するタイミング

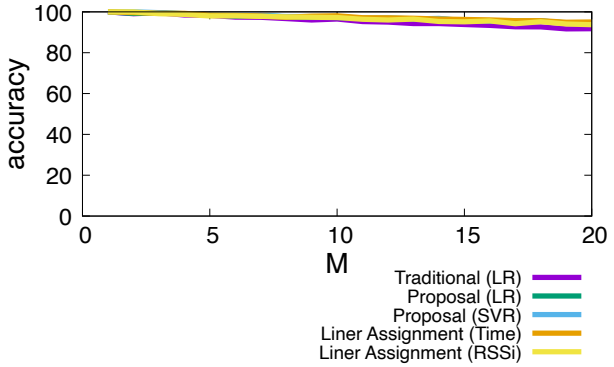

図 6　使用端末数 M を変化させたときの精度

スとなる MAC アドレス C、D においてコスト関数を計算するとコストが低いのは MAC アドレス C となる、同様に MAC アドレス B においてもコスト関数を計算するとコストが低いのは MAC アドレス D となる。そのため図 2、3 においては MAC アドレス A と C、MAC アドレス B と D が同一端末として推定される。

4.　評価

今回は [3] で用いたデータを用いて評価を行なった。このデータは以下のような実験で用意した。まず事前に Bluetooth をオフにした状態のスマートフォンを輪ゴムでプラレールに縛り付けた。その後、スマートフォンの Bluetooth をオンにして一周約 5210 mm のトラックを時計回りに走行させ、その状態で 1 時間 BLE パケットのキャプチャを行った。条件を変えてデータを取得するため、モニタリング端末の位置を 20 箇所変えて合計 20 回のデータ取得実験を行った。

以上のようにして得られた 20 回分のキャプチャデータから COCOA の UUID や Bluetooth をオンにしたタイミング、MAC アドレスが切り替わった場合は、そのタイミングや RSSI を元にスマートフォンの COCOA が使用した MAC アドレスの系列およびその MAC アドレスに関連するパケットキャプチャログを手動で抽出した。なお、本実験では周囲に実験用スマートフォン以外のスマートフォンのない状況で実験を実施したため、手動で MAC アドレスを同定できた。

本報告ではこのようにして得られたデータを重ね合わせることにより、仮想的に複数台の端末が存在する場合のデータを生成する。図 5 は 2 回分のデータを重ね合わせた場合の例である。この例では、2 台の MAC アドレスがほぼ同じタイミングで変化している。

提案手法における回帰処理は機械学習用のライブラリ scikit-learn を用いて実装した。また、線形割り当て問題を解くのに、数値解析ライブラリである scipy を用いた。今回は提案手法と比較評価をするため、下記の手法を用いた場合の評価を行った。

- 既存手法：
  我々の過去の提案手法 [3]。回帰手法として線形回帰を用いた場合を用いる。
- 時間差を用いた線形割当手法（時間差手法）：
  提案手法では $\bar{r}_{i,j}$ をコスト関数として線形割り当て問題を解いていた。これに対して文献 [6] と同様、MAC アドレス間の時間差 $t_j - t'_i$ をコスト関数として線形割り当て問題を解く手法。
- RSSI を用いた線形割当手法（平均手法）：
  末尾 I 秒間の受信パケットと開始 I 秒間の受信パケットの RSSI の平均値の差をコスト関数として線形割り当て問題を解く手法。

今回の評価では同じケースでのデータを 100 通り生成し、それぞれのデータに対して手法を適用した場合の同定精度を求め、その平均値を取得した。同定精度は下記式 (4) によって求めた。

$$\text{精度} = \frac{\text{データ内で正しく同定できた回数}}{\text{データ内でアドレス変化が起きた回数}} \times 100 \quad (4)$$

4.1　端末台数を変えた場合の比較評価

まず、評価に使用するデータ数（端末数）M を 1〜20 の範囲で変えた場合の精度評価を行った。本研究では、端末の台数が M 台の場合の評価を行うために、取得した 20 種類のデータの中から M 個のデータをランダムに選び、さらにそれらのデータを重ね合わせたデータを使って検証を行った。

ここで、それぞれのデータは Bluetooth をオンにしたタイミングから計測したデータであるため、開始時刻をあわせてデータを重ね合わせると、特に 1 回目の MAC アドレス変更タイミングが過度に重なる場合がある。そのため、本研究では、各データの開始時刻に 0〜600 秒の間のランダムな遅延を挿入して重ね合わせた。また、閾値は [3] を参考に、$R = 15$、$T = 6$、$I = 5$ を用いている。[3] では閾値 R と I をそれぞれ $1 \leq R \leq 20$、$1 \leq I \leq 20$ の範囲で変化させた結果を元に $I = 15$ とした。しかし、実際の環境を考慮すると、過去 15 秒間の行動データから回帰曲線を推測するのは難しいと思われるので今回は $I = 5$ に設定している。

比較評価結果を図 6 に示す。図より、端末数 M が増えるほど精度が下がることがわかる。$M = 20$ 時点では精度が高い順に、時間差手法が約 95%、提案手法で線形回帰と SVR を用いた場合が約 94%、平均手法が約 93%、既存手法が約 91% となっている。このことから線形割り当て問題の考え方を用いることによって、同定の精度を向上させることが可能だということがわかる。また、提案手法で精度に差がない理由としては [3] でも述べられている通り使用データに似通ったデータが多く、また回帰に使用可能なパケット数が少ないことが挙げら

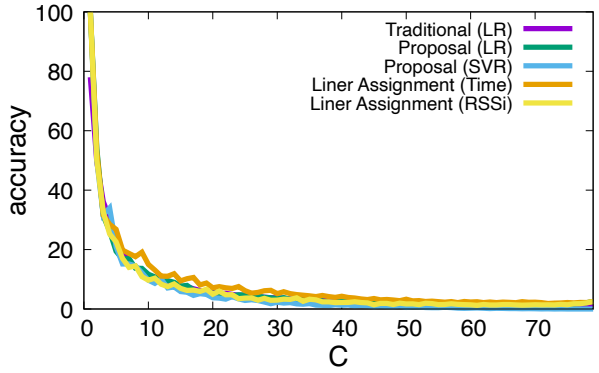

図 7　部分キャプチャデータ数 C を変化させた場合の精度 ($D = 0$)

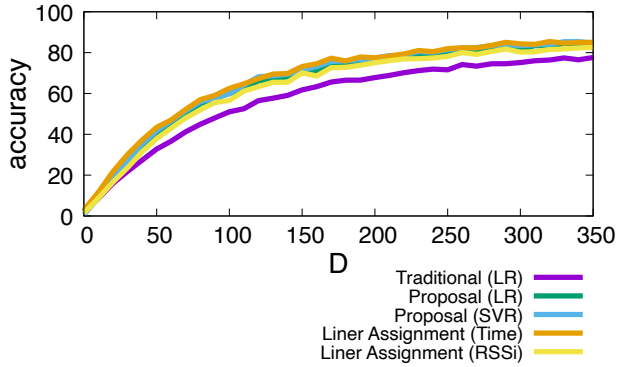

図 8　遅延時間 D を変化させた場合の精度（$C = 50$）

れる。また、時間差手法より精度が低い理由も同様に回帰曲線の再現度の低さが原因と考えられる。

4.2　MAC アドレスの変化タイミングを近接した場合の比較評価

ここまでの評価では最大で 20 台分の端末に相当するデータを意図的にタイミングをずらして重ね合わせ評価していたため、比較的同定をしやすい条件であった。一方、本節では [3] 同様、多くの MAC アドレスがほぼ同時に変化した場合を想定して評価を行う。

評価ではまず、20 種類のキャプチャデータから MAC アドレスの変化するタイミング前後の一定期間分のキャプチャデータを抜き出した。今回のデータでは合計 79 回の MAC アドレスの変化タイミングがあり、79 個の部分キャプチャデータを取得した。その後、MAC アドレスの変化タイミングが揃うように部分キャプチャデータを重ね合わせて評価用データを作成した。その後、使用する部分キャプチャデータ数を C とし、$1 \leq C \leq 79$ の場合の評価用データの精度を求めた。続いて MAC アドレスの変化タイミングをずらした場合の精度を評価するため $C = 50$ と固定し、意図的に MAC アドレスの変化タイミングを $0 \leq D \leq 350$ の範囲で $\pm D$ 秒ずらした場合の精度も求めた。

得られた結果をそれぞれ図 7、図 8 に示す。図 7 より MAC アドレスが変化するタイミングが揃う場合は精度が著しく低下することがわかった。[3] で述べたように、今回のデータは似通ったデータが多いため、変化タイミングを揃えてしまうと同定が非常に困難となり、精度が大きく下がるものと思われる。本実験でも前節と同様に、時間差手法が最大の精度となった。

次に図 8 より、MAC アドレスの変化タイミングに時間差を入れた場合、時間差を大きくするほど精度が向上することがわかる。また、既存手法よりも提案手法やその他の比較手法の方が精度が高くなった。このことから線型割り当て問題の考え方を使うことによって、MAC アドレスが変更されたタイミングか重なった時にもある程度対処することが可能と考えられる。また、前節において提案手法や平均手法、時間差手法の精度が既存手法の精度を上回ったのも同様の理由と思われる。本実験においても、図 7 同様、時間差手法の精度が最大となった。

5.　まとめと今後の課題

本研究では MAC アドレスがランダム化された BLE 端末をその端末が発信したパケットの受信時刻と RSSI に基づいて同定する手法 [3] を改良した手法を提案した。また提案した手法に対する精度評価実験を行った結果、端末台数が 20 台の場合でもパラメータを適切に設定することで 94% の精度で端末を同定することが可能であることを確認した。

今後の課題としては、多様なデータを揃えた上での再検証や MAC アドレスの変化タイミングが重なってしまった場合の対応の検討、より実際の環境に近い条件での精度評価実験を行うことなどが考えられる。

謝辞

本研究の一部は科学研究費（課題番号 19K11934）の補助を受けている。ここに記して謝意を表す。

参考文献

[1] J. K. Becker, D. Li, and D. Starobinski, “Tracking anonymized Bluetooth devices,” in *Proceedings of Privacy Enhancing Technologies*, Jul. 2019, pp. 50–65.

[2] S. Akiyama, R. Morimoto, and Y. Taniguchi, “A study on device identification from BLE advertising packets with randomized MAC addresses,” in *Proceedings of IEEE ICCE Asia 2021*, Nov. 2021, pp. 161–164.

[3] 秋山周平, 谷口義明, “MAC アドレスがランダム化された BLE 機器が移動する場合の同一機器推定手法,” **情報処理学会マルチメディア、分散、協調とモバイルシンポジウム** *DICOMO 2022* **論文集**, pp. 1489–1496, 2022.

[4] G. Celosia and M. Cunche, “Saving private addresses: An analysis of privacy issues in the bluetooth-low-energy advertising mechanism,” in *Proceedings of the 16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2019, pp. 444–453.

[5] S. Martello and P. Toth, “Linear assignment problems,” *North-Holland Mathematics Studies*, vol. 132, pp. 259–282, 1987.

[6] L. Jouans, A. C. Viana, N. Achir, and A. Fladenmuller, “Associating the randomized Bluetooth MAC addresses of a device,” in *Proceedings of IEEE CCNC 2021*, 2021, pp. 1–6.