

# MAC アドレスがランダム化された BLE 機器の同定手法

## An identification method for BLE devices with randomized MAC addresses

秋山 周平<sup>†</sup>森本 涼也<sup>†</sup>谷口 義明<sup>†,‡</sup>

Shuhei Akiyama

Ryoya Morimoto

Yoshiaki Taniguchi

### 1. はじめに

Wi-Fi や BLE (Bluetooth Low Energy) 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。アドレスがランダム化された端末は、自身が持つ固有の MAC アドレス (グローバル MAC アドレス) の代わりにランダムに生成したアドレス (ローカル MAC アドレス) を用いてネットワークの探索を行う。また、そのアドレスは一定間隔で再生成され更新されるため、長時間同一端末のパケットを追跡することは困難になっている。しかし、端末利用者の行動追跡によって得られるデータの需要が高いため、特に Wi-Fi では、MAC アドレスがランダム化された端末の追跡手法の研究が進んでいる。

一方、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 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 のパケットを発信するスマートフォンの数が増加することも考えられる。BLE パケットには、Wi-Fi での MAC アドレスと同等の端末固有のアドレスが含まれる。今後、BLE 端末やその端末が発信するパケットを利用する研究が進められていくことが予想される。その時、MAC アドレスがランダム化されていると、実験用端末の発したパケットがどれかわからなくなる等、実験の妨げになってしまう可能性がある。例えば、文献 [1] では、スマートフォン上の COCOA の送信した BLE パケットの MAC アドレスとスマートフォンを対応付けるために、スマートフォンの Bluetooth 機能をオンにしたタイミングを使っている。しかしながら、MAC アドレスが再度変わった場合に対応付けが失われてしまうため、MAC アドレスのランダム化が長期間の実験の妨げとなっている。

現在、MAC アドレスがランダム化された BLE 端末の追跡には、アドレスキャリーオーバーアルゴリズム [2] という手法が使われている。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、アドレスキャリーオーバー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。

本研究では、MAC アドレスがランダム化された BLE 端末のアダプティブパケットから端末を同定する手法を提案する。本研究で提案する同定手法ではアドレスキャリーオーバー

アルゴリズムのように端末固有の値を用いることはない。本研究では、ノート PC 等で周囲にある動きの少ない BLE 機器のパケットをキャプチャし、そのデータから同定を行うことを想定する。また、BLE 端末としては、街中で 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。提案手法では、最後に使われなくなった時刻とはじめて使用された時刻が近く、かつ、平均受信電波強度に近い MAC アドレスを同じ端末が使用する MAC アドレスと同定する。本研究では、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。

本論文の以降の構成は以下のとおりである。まず、2 章で関連研究について述べる。3 章で提案する同定手法について述べ、4 章でその評価結果を述べる。最後に 5 章でまとめと今後の課題を述べる。

### 2. 関連研究

MAC アドレスがランダム化された Wi-Fi 機器の同定手法としては、プローブ要求から得られる MAC アドレス、シーケンス番号等を利用して同定、追跡を行う手法 [3] [4] [5] [6] や、攻撃者が RTS フレームを送信することでランダム化されていない MAC アドレスを得てランダム化を無効化する手法 [7] がある。前者の手法では、主にランダム化された MAC アドレスに加えてパケットのサイズやシーケンス番号などの情報と、ランダム化間隔が 10 分以上と追跡するために十分に長いことなどを利用している。後者の手法では IEEE 802.11 の脆弱性を利用している。端末に対して RTS (Request to Send) パケットを送信すると、端末は CTS (Clear to Send) パケットを返信するが、このとき、CTS パケットから端末のグローバル MAC アドレスが得られる。以降、得られたグローバル MAC アドレスに対して RTS パケットを送信することで端末を追跡できる。

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 [2]。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ダプティブ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。

### 3. 提案手法

本章では、提案する BLE 機器の同定手法について説明する。

#### 3.1 手法の概要

本研究では、図 1 のように、複数台の動きの少ない BLE 端末が周囲にある環境においてノート PC などの端末上で 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定を行うことを想定する。BLE 端末としては、COCOA 等のアプ

<sup>†</sup> 近畿大学理工学部情報学科, Faculty of Science and Engineering, Kindai University

<sup>‡</sup> 近畿大学情報学研究所, Cyber Informatics Research Institute, Kindai University



図1 想定環境

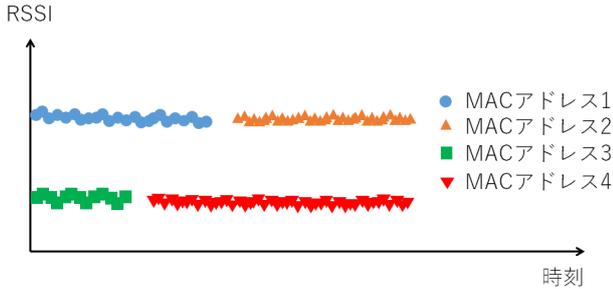


図2 キャプチャしたパケットのMACアドレスとRSSIの分布

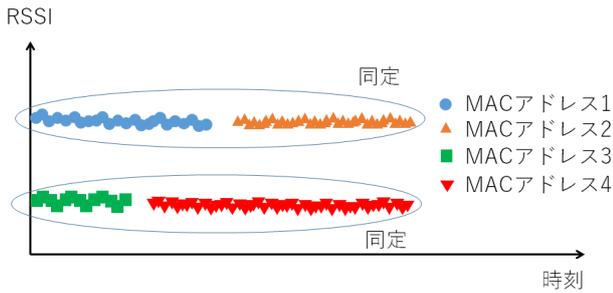


図3 提案手法によるMACアドレスの同定例

リケーションを利用し、定期的にBLEパケットを送信するスマートフォンを想定する。

キャプチャしたパケットからは、BLE端末のMACアドレスやパケット受信時の受信電波強度 (RSSI: Received Signal Strength Indicator) が得られる。図2は受信パケットから得られるRSSIの分布の例である。この例の場合、4つのMACアドレスからのパケットを受信している。提案手法では、MACアドレス1とMACアドレス2の平均受信電波強度が近いことと、MACアドレス1から送信されるパケットを最後に受信してから、MACアドレス2から送信されるパケットを最初に受信するまでの期間が短いことを使って、図3のようにMACアドレス1とMACアドレス2が同じものであると同定する。同様に、MACアドレス3とMACアドレス4が同じであると同定する。

このように、本研究で提案する手法では、パケットの受信時刻、RSSIからランダム化されたMACアドレスの同定を行う。本手法では、データ中の任意のMACアドレスとその他のMACアドレスが、同じ端末によって使用されているアドレスであるかどうかを判定する。ローカルMACアドレスが切り替わると、前に使っていたアドレスは一切使われなくなるた

#### Algorithm 1 提案手法

```

1: for  $1 \leq i \leq N$  do
2:    $C_i \leftarrow \emptyset$ 
3:   for  $i+1 \leq j \leq N$  do
4:     if  $0 \leq t_i - t'_j \leq T$  and  $|\bar{r}_i - \bar{r}_j| \leq R$  then
5:        $C_i \leftarrow C_i \cup \{a_j\}$ 
6:     end if
7:   end for
8:   if  $|C_i| > 0$  then
9:      $a_j \in C_i$  のうち  $a_i$  との正規化した距離が最も小さいものを  $a_i$  と同じアドレスと同定
10:  end if
11: end for

```

MACアドレス	初回受信時刻	最終受信時刻	平均RSSI
AA:AA:AA:AA:AA:AA	10:15:20.100	10:25:25.231	-70 dBm
BB:BB:BB:BB:BB:BB	10:18:34.531	10:35:55.209	-90 dBm
CC:CC:CC:CC:CC:CC	10:25:27.213	10:40:40.339	-72 dBm
DD:DD:DD:DD:DD:DD	10:35:59.090	10:55:20.878	-59 dBm

図4 手法の動作例

め、切り替わる前後の2つのアドレスの受信時刻が重なることはない。そのため、パケットの受信時刻から、同端末が発した可能性のあるアドレスをある程度絞り込むことが可能である。RSSIは、端末間の距離の目安として利用する。端末周囲の環境や障害物等によって値が変化するため正確な距離を得ることはできないが、距離が大きく離れている端末のアドレスを見つけることができる。

#### 3.2 手法の詳細

提案する手法では、まず、キャプチャしたパケットから、ユニークなMACアドレス一覧を取得する。ここでMACアドレスを  $a_i$  ( $1 \leq i \leq N$ ) とする。  $N$  はパケットから取得されたユニークMACアドレス数の最大値である。また、各アドレス毎に最初と最後にパケットを受信した時刻 (それぞれ初回受信時刻  $t_i$ 、最終受信時刻  $t'_i$ ) およびパケットのRSSIの平均値  $\bar{r}_i$  を記録しておく。

次に、初回受信時刻と最終受信時刻の差が0秒以上  $T$  秒以下のMACアドレスの組み合わせを探す。条件を満たすアドレスが見つかった場合はそれぞれのMACアドレスから受信されたパケットの平均RSSIの値を比較し、その差が閾値  $R$  以下のものを同一端末候補アドレスとする。

条件を満たすアドレスが見つかった場合、平均RSSIの差と初回受信時刻と最終受信時刻の差をそれぞれ正規化し、その距離が最も小さいものを同一端末とみなす。具体的には2つのアドレス  $a_i, a_j \in C_i$  に対して下記式に基づき距離  $d_{i,j}$  を計算し、距離の最も小さなものを同一端末とみなす。ここで  $C_i$  はアドレス  $a_i$  の同一端末候補アドレスの集合である。

$$d_{i,j} = \sqrt{\left(\frac{t_i - t'_j}{T}\right)^2 + \left(\frac{\bar{r}_i - \bar{r}_j}{R}\right)^2} \quad (1)$$

提案手法を疑似コードとして表記したものを Algorithm 1 に示す。

表1 実験に用いたデータの取得条件

データ番号	方向、状態	距離	アドレス変化	平均 RSSI [dBm]	RSSI の標準偏差 [dBm]
1	後ろ	0.5 m	5 回	-77	2.24
2	前	0.5 m	5 回	-74	1.39
3	右	0.5 m	5 回	-77	2.46
4	左	0.5 m	5 回	-67	2.44
5	後ろ	1.0 m	3 回	-88	1.65
6	左	1.0 m	5 回	-82	1.94
7	前	1.0 m	5 回	-84	6.27
8	右	1.0 m	5 回	-92	3.56
9	後ろ	1.5 m	5 回	-91	1.56
10	前	1.5 m	5 回	-87	2.88
11	右	1.5 m	5 回	-85	1.65
12	左	1.5 m	5 回	-81	1.56
13	後ろ	2.0 m	3 回	-71	3.68
14	左	2.0 m	4 回	-83	0.97
15	前	2.0 m	5 回	-84	1.86
16	右	2.0 m	3 回	-90	1.73
17	前、カバン内	0.5 m	6 回	-74	4.3
18	前、カバン内	1.0 m	5 回	-79	2.65
19	前、カバン内	1.5 m	6 回	-83	2.54
20	前、カバン内	2.0 m	5 回	-82	2.05

以下、提案手法の動作例を示す。図4は、ノートPCで取得されたパケットから得られた各MACアドレスと、その初回受信時刻 $t_i$ 、最終受信時刻 $t_f$ 、平均RSSI $\bar{r}_i$ をまとめたものの例である。今、 $T = 5$ 、 $R = 10$ とする。MACアドレスAA:AA:AA:AA:AA:AAから送信されるパケットの最終受信時刻とMACアドレスCC:CC:CC:CC:CC:CCから送信されるパケットの初回受信時刻の差は約2秒であり、閾値 $T$ より小さい。また、これらのパケットの平均RSSIの差は2dBmであり、閾値 $R$ より小さい。したがって、提案手法では、MACアドレスAA:AA:AA:AA:AA:AAとMACアドレスCC:CC:CC:CC:CC:CCは同じ端末が使用したMACアドレスであると同定する。同様に、MACアドレスBB:BB:BB:BB:BB:BBから送信されるパケットの最終受信時刻とMACアドレスDD:DD:DD:DD:DD:DDから送信されるパケットの初回受信時刻の差は約4秒であり閾値 $T$ より小さい。しかし、平均RSSIの差は約30dBmと閾値 $R$ より大きい。したがって、提案手法ではこれらのMACアドレスは同一端末が使用したMACアドレスと同定しない。

## 4. 評価

### 4.1 実験内容

提案手法の同定精度を確かめる実験を行った。実験は屋内で実施した。実験では、スマートフォンとしてmoto g<sup>7</sup> power XT1955-7 (OS: Android 9)を、BLEパケットのキャプチャを行うためにRaspberry Pi 4 (OS: Raspbian 10 Buster)を使用した。Raspberry Pi 4には、文献[1]で我々が開発したBLEパケットキャプチャプログラムを導入した。スマートフォンにはCOCOA Version 1.2.3をインストールした。COCOAは、GoogleとAppleが開発したExposure Notification System

(ENS) [8]を利用している。ENSでは200から270ミリ秒毎にBLEアドバタイジングパケットを送信する。送信されるBLEアドバタイジングパケットに含まれるMACアドレスは10分から20分毎に変更される。

実験は以下のような手順で実施した。まず、事前にBluetoothをオフにした状態のスマートフォンをRaspberry Pi 4から一定距離離れたところに設置した。その後、スマートフォンのBluetoothをオンにし、Raspberry Pi 4を使って1時間BLEパケットのキャプチャを行った。スマートフォンの位置をRaspberry Pi 4に対して前後左右の方向に距離を50cmから2mの範囲でさまざまに変えて実験を行った。一部の実験では、スマートフォンをカバンの中に入れた状態でパケットキャプチャを行った。本稿では、このようにして20個のデータを取得した。各データの取得条件を表1に示す。表中、アドレス変化は、1時間の間にMACアドレスが変化した回数を表す。なお、平均RSSIは、小数点以下を四捨五入している。

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、これらの20個のデータの中から $M$ 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。なお、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。そのため、データを重ね合わせる際には、各データの開始時刻に0~600秒の間でランダムに遅延を挿入して重ね合わせた。

本研究では、端末の台数が $M$ 台の場合のデータを100通り生成し、それぞれのデータに対して提案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。それぞれのデータに

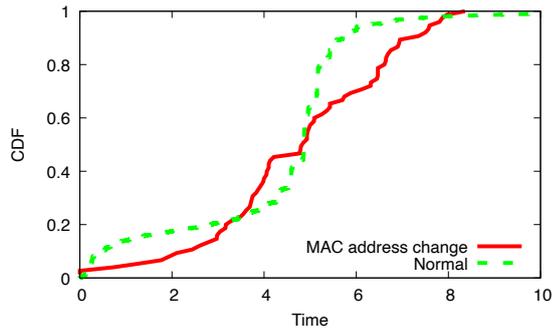


図5 パケット間や MAC アドレス間における時間差の累積分布関数

おける精度は、以下の二通りの式によって求めた。

$$\text{個別精度} = \frac{\text{データ内で正しく同定できた回数}}{\text{データ内でアドレス変化が起きた回数}} \times 100 \quad (2)$$

$$\text{累計精度} = \frac{\text{全て正しく同定できた端末数}}{\text{データセット内の全端末数}} \times 100 \quad (3)$$

#### 4.2 取得データの基本特性の調査

まず、MAC アドレスが切り替わるタイミングでパケットの送信間隔に遅延が発生するかどうかを確認するためにパケット受信間隔に関する基本特性の調査を行った。取得した 20 種類のデータ全てを用いて、MAC アドレスが変化する場合の BLE アドバタイジングパケットの受信間隔の累積分布関数 (Cumulative Distribution Function: CDF) と MAC アドレスが変化しない場合の受信間隔の累積分布関数を調査した。

結果を図 5 に示す。図に示されるように、MAC アドレスが変化しない場合の受信間隔は 5 秒付近で急激に変化しており、5 秒付近でピークとなることがわかる。一方、MAC アドレスが変化する場合の受信間隔の累積分布関数はおおむね直線状であり、おおよそ 0 秒から 8 秒の間でまんべんなく分布していることがわかる。ここで、BLE アドバタイジングパケットの受信間隔は、送信間隔 (200 から 270 ミリ秒 [8]) と比較して大きな値となっている。BLE アドバタイジングパケットは比較の様々なシステムで利用される 2.4 GHz 帯を利用して送信されており、干渉によるパケットロスの影響により受信間隔が大きくなるものと考えられる。

#### 4.3 パラメータを変えた場合の提案手法の精度評価

続いて、提案手法の同定精度の評価を行った。端末台数  $M = \{5, 10, 15, 20\}$  の各データセットに対して、提案手法のパラメータ  $T$  と  $R$  を  $1 \leq T \leq 20$ 、 $1 \leq R \leq 20$  の範囲で変化させた場合の個別精度を図 6 ~ 図 9 に、累計精度を図 10 ~ 図 13 に示す。図より、まず、個別精度の方が累計精度よりも高くなることわかる。これは個別精度では 1 回のアドレス変化に対する同定の成功率を評価しているのに対して、累計精度は 3 ~ 6 回のアドレス変化に対する同定が全て成功するかを評価しているためである。

次にパラメータ  $R$  を変化させた場合の結果を考察する。図 6(a)、図 10(a) のように端末の台数が少ない場合 ( $M = 5$ )、 $R$  を変化させても精度の大きな変化は見られない。一方、端末の台数が多い場合 ( $10 \leq M$ ) は、 $R = 1$  もしくは  $R = 3$  など閾値  $R$  が小さいときに精度が最大となり、 $R$  を大きくすると精

度が徐々に下がる傾向にあることがわかる。閾値  $R$  が小さいほど精度が高くなる理由は、本研究では静止したスマートフォンのパケットを対象としているからであると考えられる。静止した状態であるためパケットの RSSI はアドレスが変わっても大きく変化せず、それゆえ閾値  $R$  を小さくするほど精度が高くなる。周囲の電波伝搬環境が変動するような環境、スマートフォンが移動するような環境では、 $R$  が小さい場合に精度が下がると考えられる。そのような環境での評価と手法の改良は今後の課題とする。

一方、 $T$  を変化させた場合は、どのデータセットでも、 $T$  を大きくするほど精度が高くなり、 $T$  を一定以上の値に設定すると精度が最大となった。これは、MAC アドレスが切り替わってから再びパケットが受信されるまでに遅延があるため、 $T$  を一定以上の値に設定する必要があるためである。図 6 ~ 図 9 に示される個別精度の結果を見ると、いずれの場合でも  $T$  が 3 ~ 5 の場合に個別精度が最大であった。また、図 10 ~ 図 13 に示される累計精度の結果を見ると、いずれの場合でも  $T$  が 8 ~ 9 の場合に累計精度が最大であった。このように似たような  $T$  の値の時に精度が最大となるのは、本研究では、同一のスマートフォンから送信されたパケットを同一の端末 (Raspberry Pi 4) を使って受信する状況で実験評価を実施しており、実験データ内で、MAC アドレスが切り替わってから再びパケットが受信されるタイミングが似たような値となったためと考えられる。また、図 5 に示されるように、MAC アドレスが変化する場合の BLE アドバタイジングパケットの受信間隔はおおよそ 8 秒以内であるため、この範囲内で精度が最大になったと考えられる。なお、提案手法では、同定する候補となるアドレスが複数ある場合は、もっとも距離の近いアドレスを同一の端末からのものと同定する。今回の実験では受信電波強度の変動が少なかったため、 $T$  を一定以上に大きくしても、正解となるアドレスが同定先として選択され、精度に影響を与えなかったものと考えられる。使用する端末の種類を増やした場合の評価や手法の改良は今後の課題とする。

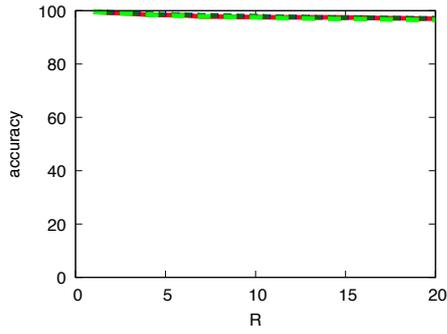
#### 4.4 端末台数を変えた場合の提案手法の精度評価

端末台数  $M$  を変えた場合の、精度の平均値と最大値を図 14 に示す。図に示されるように、端末台数が増えるほど平均精度、最大精度共に精度を維持あるいは精度が低下していることがわかる。これは端末台数が増えれば増えるほど同定候補のアドレスが増加し、正しく同定しにくくなるためである。しかしながら、端末台数 20 台の場合でも、提案手法を用いることで最大で 100% の個別精度、90% の累計精度で MAC アドレスを同定できることがわかる。

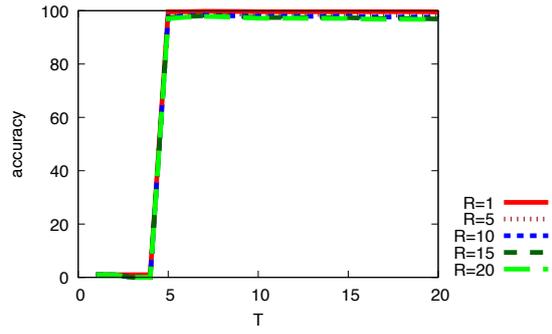
### 5. まとめと今後の課題

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 を用いて同定する手法を提案した。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 90% 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、パケットの受信時刻や RSSI 以外に同定に利用できる情報がないかの検討、端末が移動することを想

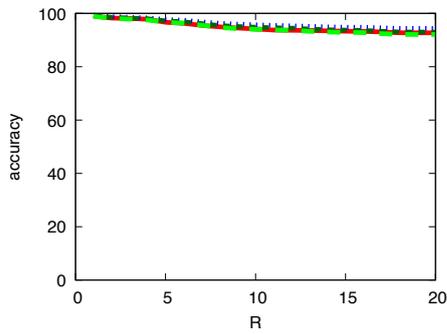


(a)  $R$  を変化させた場合の個別精度

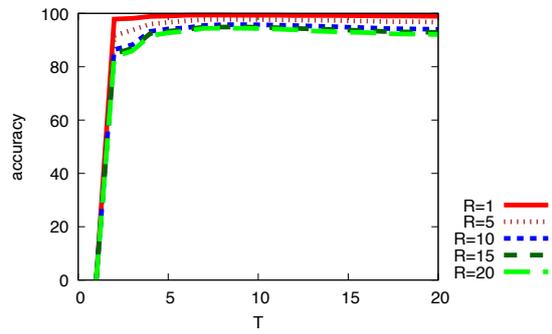


(b)  $T$  を変化させた場合の個別精度

図6 端末台数  $M = 5$  のときの個別精度

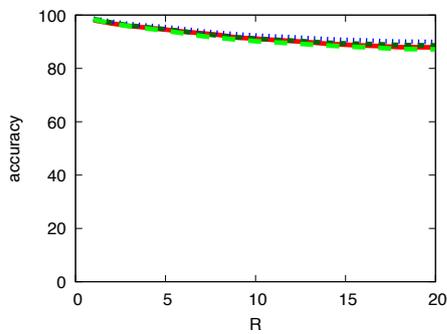


(a)  $R$  を変化させた場合の個別精度

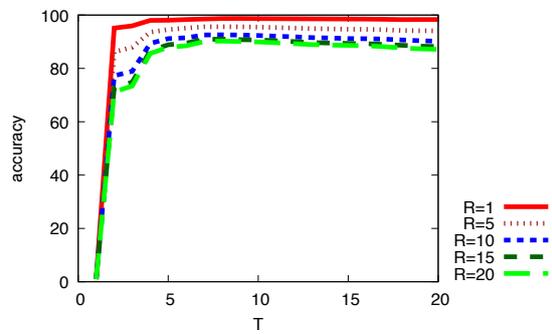


(b)  $T$  を変化させた場合の個別精度

図7 端末台数  $M = 10$  のときの個別精度

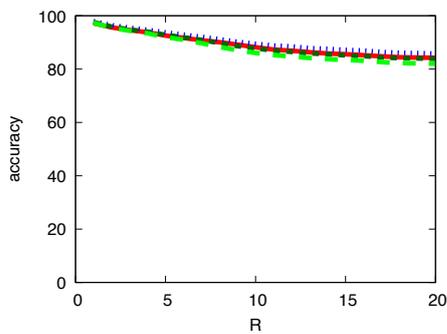


(a)  $R$  を変化させた場合の個別精度

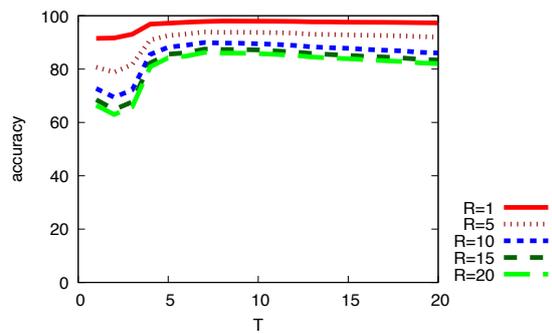


(b)  $T$  を変化させた場合の個別精度

図8 端末台数  $M = 15$  のときの個別精度

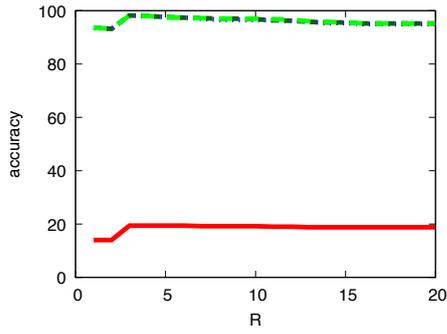


(a)  $R$  を変化させた場合の個別精度

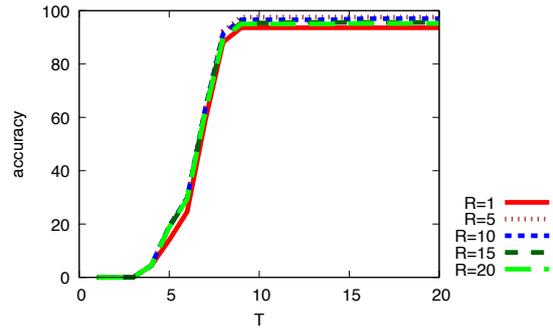


(b)  $T$  を変化させた場合の個別精度

図9 端末台数  $M = 20$  のときの個別精度

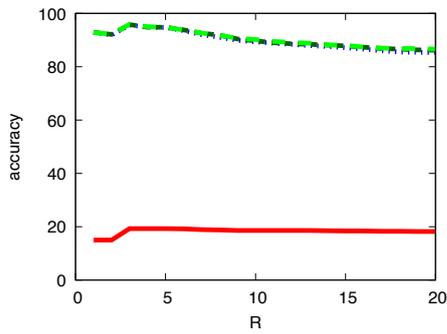


(a)  $R$  を変化させた場合の累計精度

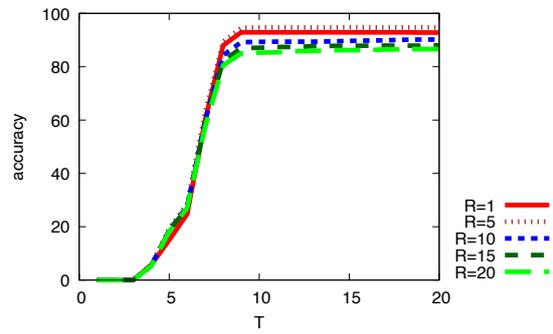


(b)  $T$  を変化させた場合の累計精度

図 10 端末台数  $M = 5$  のときの累計精度

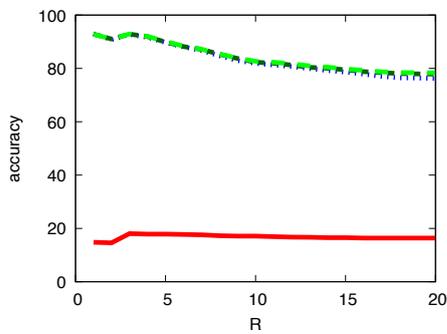


(a)  $R$  を変化させた場合の累計精度

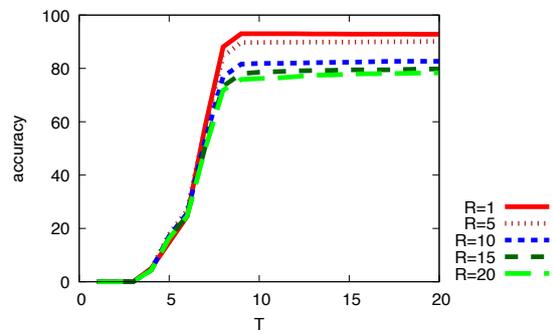


(b)  $T$  を変化させた場合の累計精度

図 11 端末台数  $M = 10$  のときの累計精度

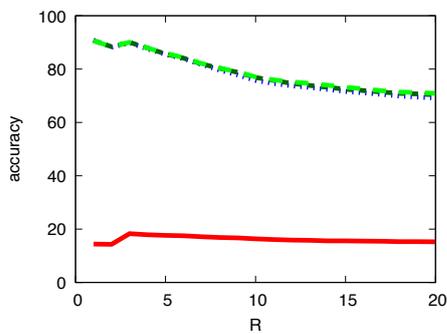


(a)  $R$  を変化させた場合の累計精度

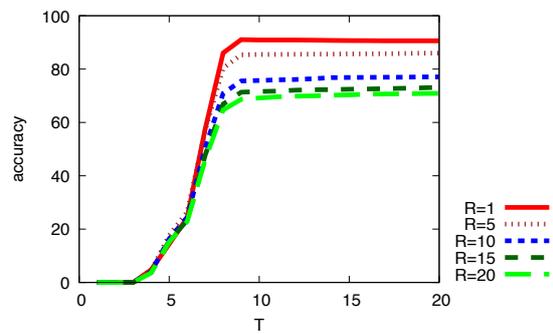


(b)  $T$  を変化させた場合の累計精度

図 12 端末台数  $M = 15$  のときの累計精度



(a)  $R$  を変化させた場合の累計精度



(b)  $T$  を変化させた場合の累計精度

図 13 端末台数  $M = 20$  のときの累計精度

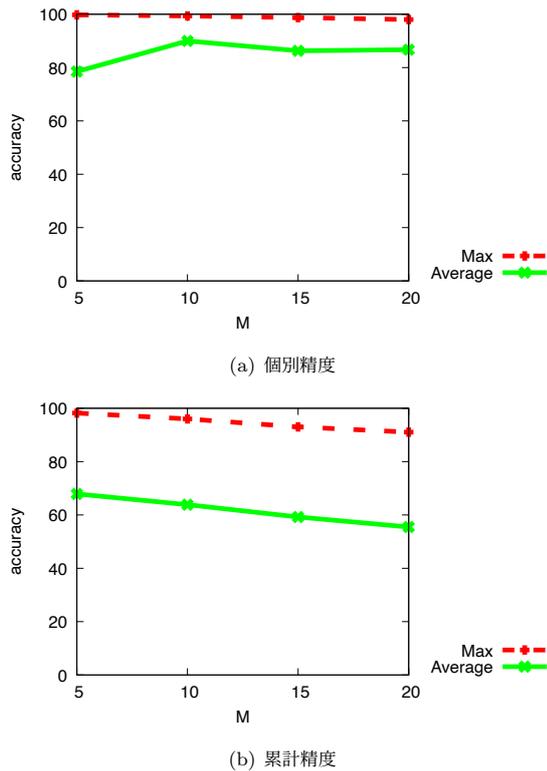


図 14 端末台数を変えた場合の最大精度と平均精度

定した場合の手法の提案、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“オール近大”新型コロナウイルス感染症対策支援プロジェクト、令和2年度近畿大学学内研究助成金(SR08)および科学研究費(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] Y. Taniguchi, T. Mukaida, Y. Ochi, and N. Iguchi, “A BLE-based monitoring system for estimating congestion on university campuses,” in *Proceedings of IEEE LifeTech 2021*, Mar. 2021.
- [2] J. K. Becker, D. Li, and D. Starobinski, “Tracking anonymized Bluetooth devices,” in *Proceedings of Privacy Enhancing Technologies*, Jul. 2019, pp. 50–65.
- [3] 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,” in *Proceedings of ACM AsiaCCS 2016*, May 2016.
- [4] 中田 恵史, 岡本 東, 堀川 三好, “Wi-Fi プローブ要求からの行動分析のための同一端末推定手法,” *情報処理学会第 80 回全国大会*, vol. 3, pp. 149–150, Mar. 2018.
- [5] 古屋 優希, 朝比奈 啓, 豊田 健太郎, 笹瀬 巖, “MAC アドレスがランダム化された Wi-Fi 無線端末の同定のための非匿名化手法の検討,” *信学技報 CS2019-17*, pp. 25–29, Jul. 2019.
- [6] M. Cunche and C. Matte, “On Wi-Fi tracking and the pitfalls of MAC address randomization,” in *Proceedings of IDO 2016*, Nov. 2016.
- [7] 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,” Mar. 2017, <https://arxiv.org/abs/1703.02874>.
- [8] Google LLC and Apple Inc., “Exposure notification Bluetooth® specification.”