

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

近畿大学大学院総合理工学研究科エレクトロニクス系工学専攻

秋山周平

谷口義明

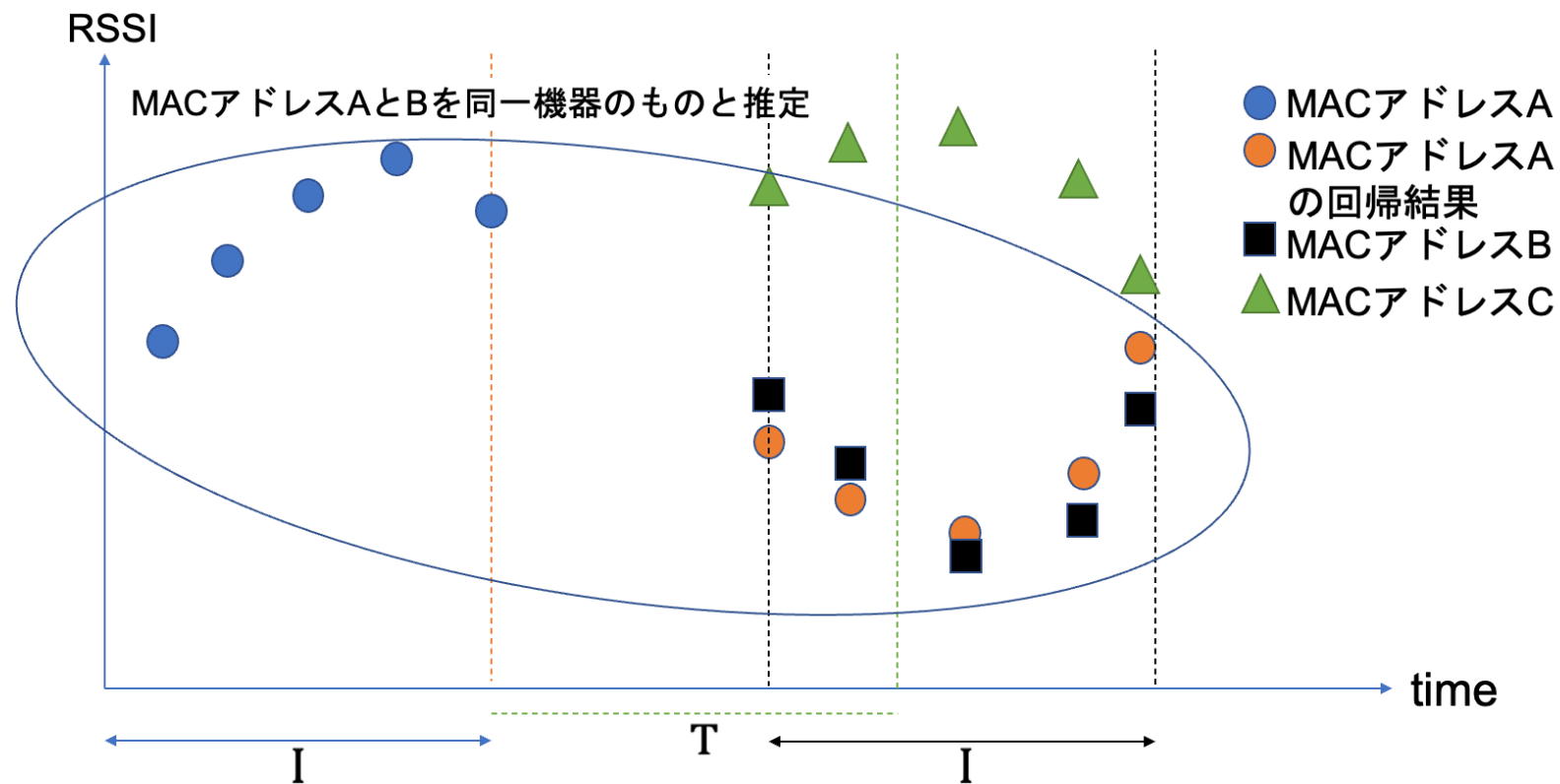
研究背景

- BLE (Bluetooth Low Energy)の利用が増加
 - BLEを利用するスマートフォンアプリやIoT機器が普及
- BLEではMACアドレスがランダム化されることが多い
 - 端末固有のアドレスではなく、ランダムに生成したアドレスを使用
 - アドレスは一定間隔で更新
 - 端末の追跡の困難化
 - 研究において無線機器を追跡したい場合に不都合が生じる

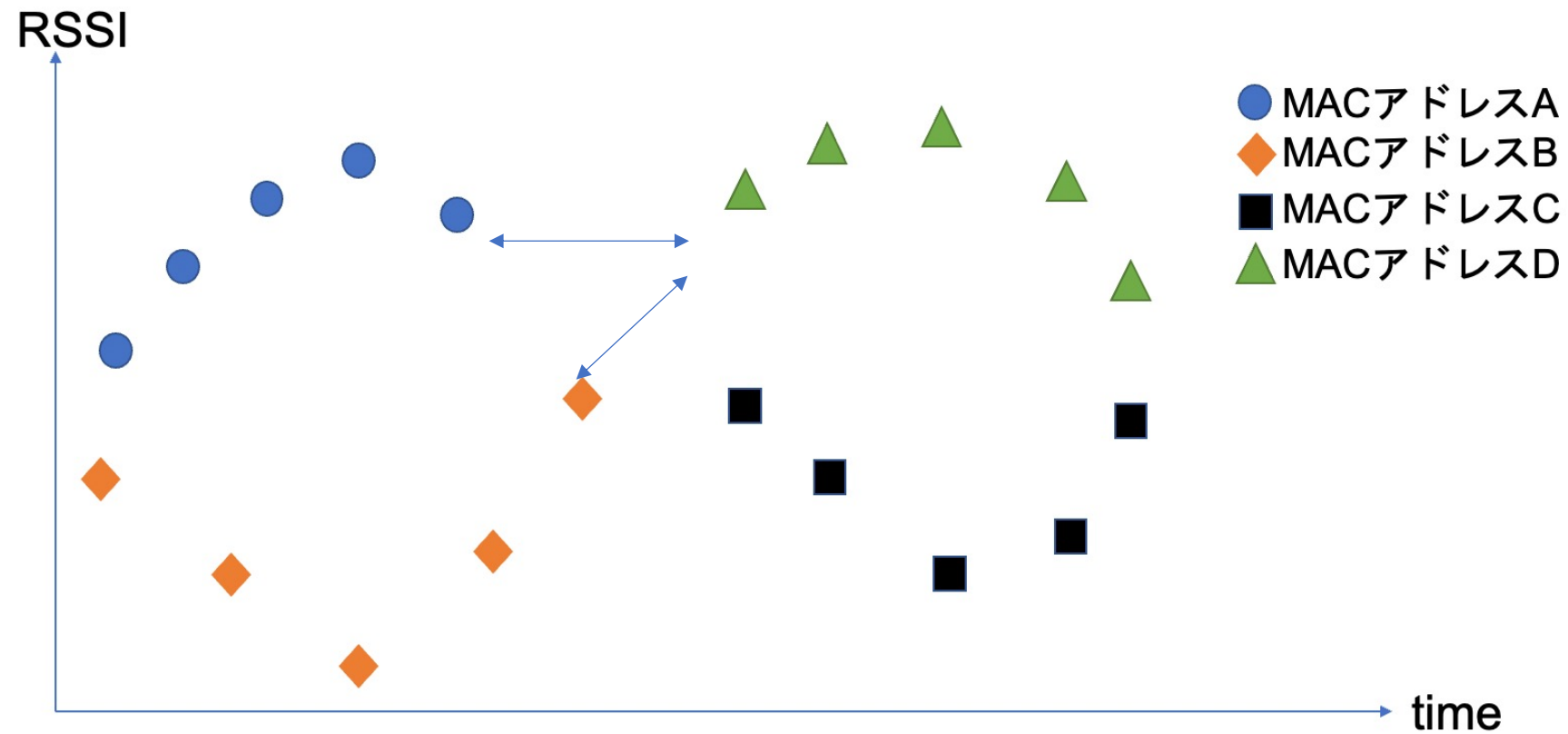
 **BLE機器のMACアドレス同定手法が必要**

既存手法

1. MACアドレスの末尾付近の packets に対して
受信時間を説明変数、RSSIを目的変数にして回帰
2. 回帰結果に類似するMACアドレスを同一機器として
推定



既存手法



変化後に同じMACアドレスを割り当ててしまう場合がある

研究の目的

- 同時に複数のMACアドレスが変化する場合の同定精度の改良
 - 線形割当問題の適用
- 提案手法の評価



線形割当問題

- ある集合と集合を組み合わせた時の総コストを最大(最小)にする問題
- ハンガリアン法やscipyで解ける

例:集合A{1,3},とB{0,2}があったとき、その積をコストとするときの最小の値を求めよ

$$1*0+3*2=6$$

$$1*2+3*0=2 \quad \bigcirc$$

となり、(1,2),(3,0)がコストが最小となる組み合わせとなる

線形割当問題

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。

$$\text{minimize } \sum_{a_i \in A} \sum_{a_j \in A} c(a_i, a_j) x(a_i, a_j) \quad a_i \in A \text{ (MACアドレスの集合)}$$

$$\text{subject to } \sum_{a_i \in A} x(a_i, a_j) = 1 \quad x(a_i, a_j) \in \{0,1\}, \quad a_i \in A, a_j \in A$$

$$\text{コスト関数 } c(a_i, a_j) = \begin{cases} \bar{r}_{i,j} & \text{if } t'_i \leq t_j \leq t'_i + T \\ \infty & \text{otherwise} \end{cases}$$

$t'_i : a_i$ の最終受信時刻
 $t_j : a_j$ の初回受信時刻

$$\text{回帰値との差 } \bar{r}_{i,j} = \frac{\sum_{k=1}^K |r_{j,k} - \hat{r}_{i,k}|}{K}$$

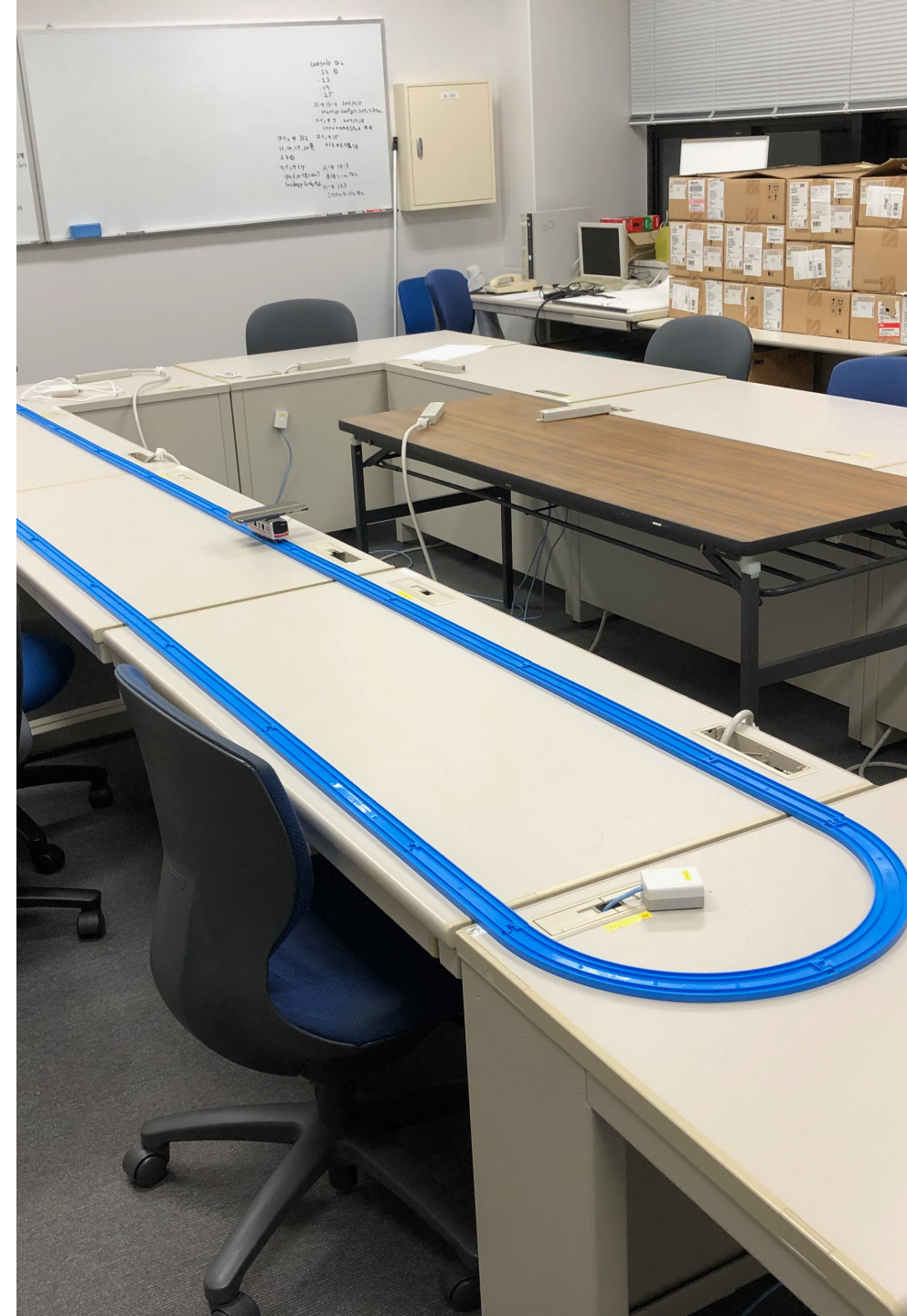
T : MACアドレスの時間差の閾値

同一機器推定を線形割り当て問題として定式化しその問題を解く
→割り当てに基づいてMACアドレスとMACアドレスを紐付ける

評価条件について

データの取得条件

- モニタリング端末としてRaspberry Pi 4を使用
- BLE機器としてスマートフォンのCOCOAを使用
- プラレールにスマホを固定し右図のトラックを走行させた
- 4316mm×418mm
- 一周約5210mm
- Raspberry Pi 4を様々な位置に配置(20パターン)
- それぞれ1時間BLEパケットを計測



比較する手法

平均手法

- 末尾 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

時間差手法

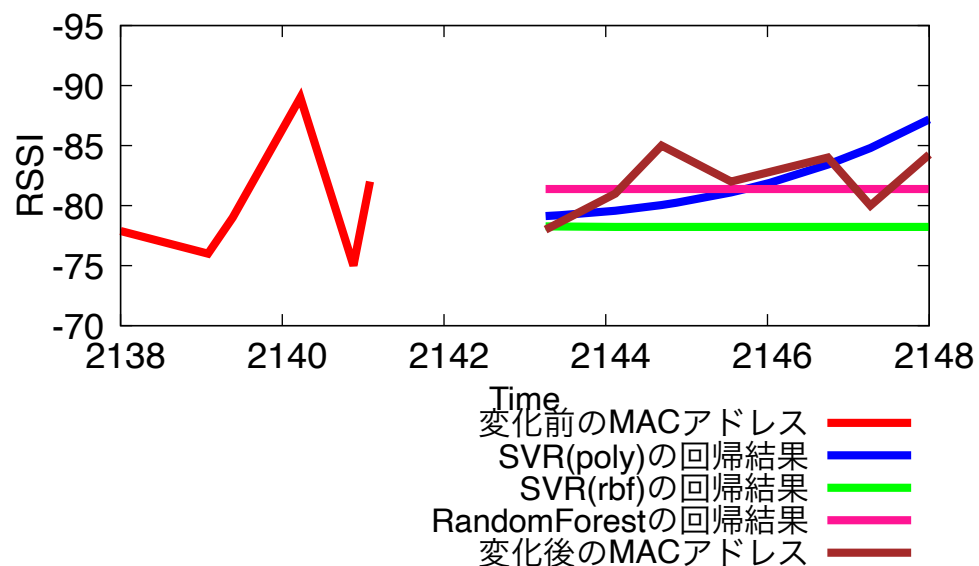
- 文献[6]を参考に考案した手法
- MACアドレスの時間差 $t_j - t'_i$ をコスト関数として線形割り当て問題を解く手法

既存手法

- 我々の過去の提案手法
- 回帰手法として線形回帰を用いる

[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.

その他の評価条件

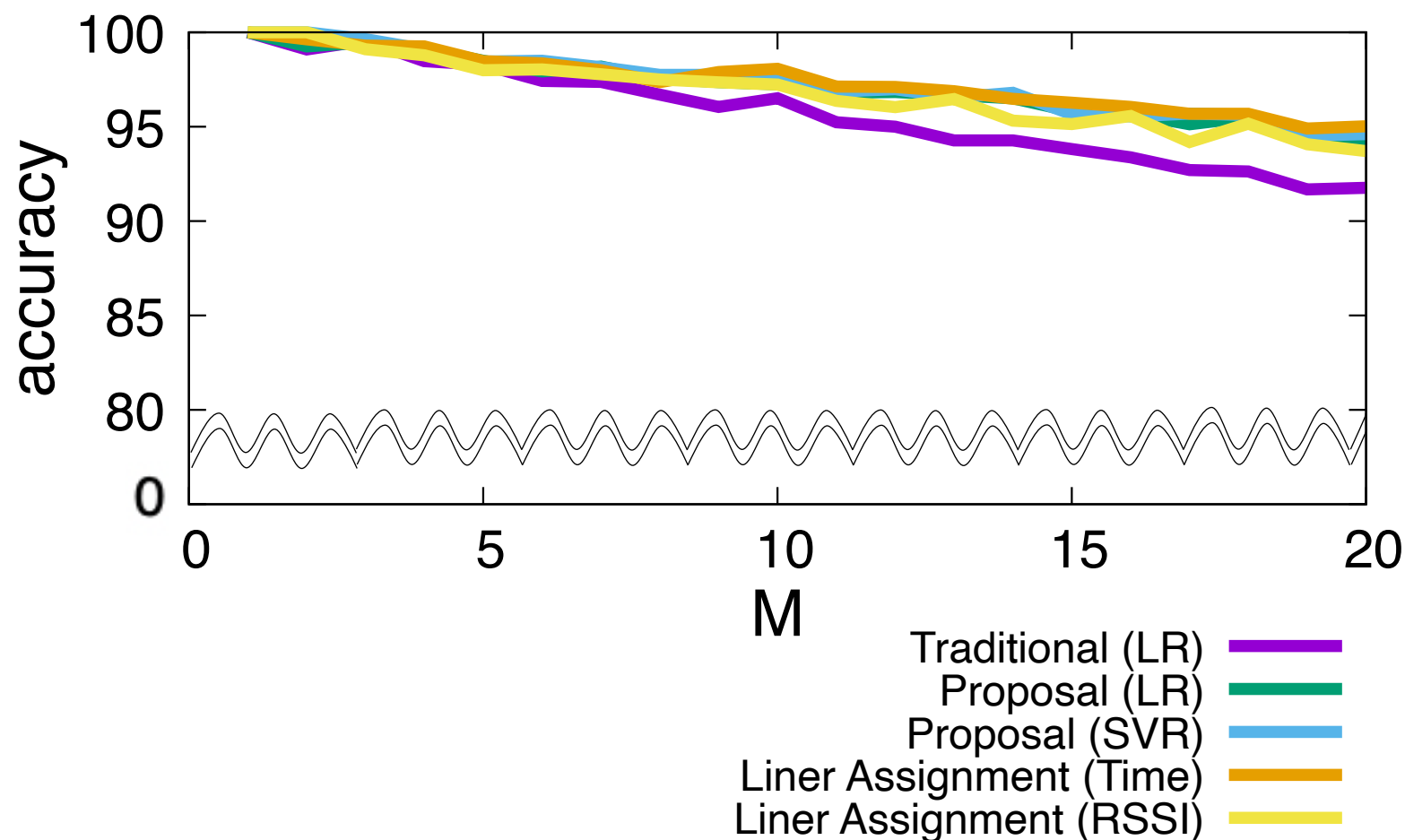


- 回帰手法には線形回帰、SVR (polyカーネル) を使用(scikit-learnで実装)
- 線形割り当て問題を解くのに、数値解析ライブラリである scipy を用いた
- 以下の指標で評価を行った。100回の評価の平均を用いた

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

端末台数M を変更した 場合の評価

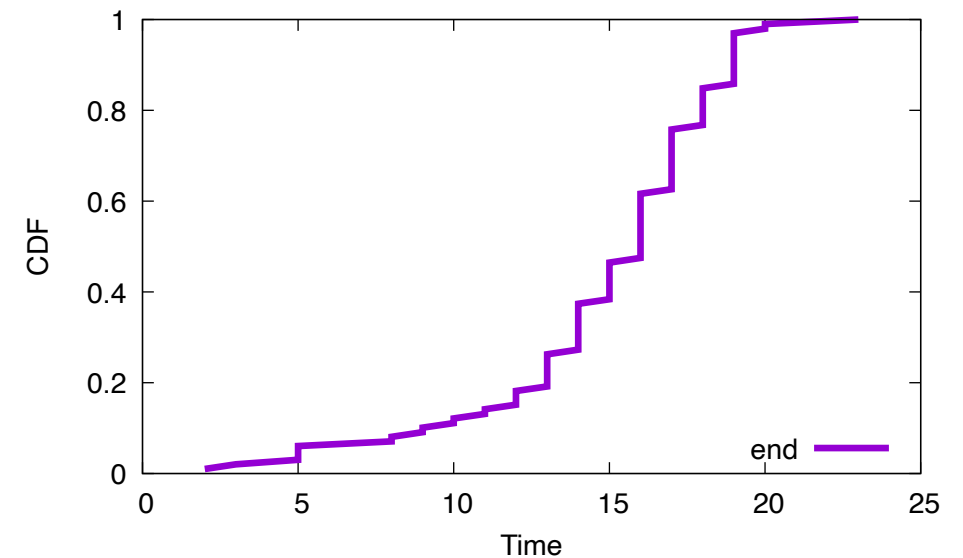
- ◆重ね合わせる端末は20個のデータからランダムで選択
- ◆閾値は $R=15$ (既存手法の閾値)、 $T=6$ 、 $I=15$ に設定



考察

手法間で差が出ていない理由

- 使用したデータがどれも似通ったデータである
 - どれも同一コースを同速度で走行した場合のデータ
- 回帰に使用可能なパケット数が少ない場合がある
 - $I = 15$ における末尾付近のパケット数が最小で 2 個、最大でも 23 個 (右図CDFより)



末尾 $I = 15$ 秒間のパケット数のCDF

まとめと今後の課題

- まとめ
 - MACアドレスがランダム化されたBLE機器の同定手法を提案した
 - 端末台数20台の場合、既存手法の最大精度が91%なのに対して、提案手法を用いることで最大で94%の精度でMACアドレスを同定できた
- 今後の課題
 - 多様なデータを揃えた上での再検証
 - 受信時刻、RSSI以外の同定に利用できる情報を検討
 - 実環境やそれに近い環境での精度評価

- 最初にMACアドレスを同定する方法は
 - Bluetoothをオンにしたタイミングを元にしています
- 同定が必要な理由
 - BLE端末を追跡することによってできる研究がある
Ex(自研究室の混雑度推定やソーシャルディスタンス確保など)
- この研究は目的に合ったものなのか
 - 最終目標としては実環境でのスマートフォンの追跡。今回はその初期検討としては妥当だと思われます。
- アプリが変わるなどしてパケット送信間隔が変わったらどうなるのか
 - 閾値Tに関しては再検討が必要だがそれ以外は大きくは変わらないと思います

- BLEアドバタイズチャンネルはWi-Fiの影響をうけないのか
 - BLEアドバタイズチャンネルはWi-Fiでは使われていないチャンネルなので影響はないがWi-Fi以外の影響を受ける可能性はある
- 一人が複数端末を持っていたらどうなるのか
 - 今回の目的はあくまで端末の推定なので影響はない
- プライバシーの侵害にはならないのか
 - 本手法では端末と人との結びつけができないので影響はないと思われる

- 何故、線形回帰、SVR(rbfカーネル)、ランダムフォレストを使用したのか
 - scikit-learnのチートシートを参考に、回帰を行う上で一番ベーシックな線形回帰、それと高精度を出せそうな非線形な回帰手法であるSVRとランダムフォレストを選びました。
- 回帰手法ごとに精度に差が出ていない理由
 - 今回の評価では端末数が少なく、MACアドレスの変更タイミングがズレているのでどの手法でも高精度を出しやすかったと思われる
- 閾値はどのように決定したか
 - Tは各MACアドレス間の時間差のCDF(累積分布関数)をもとに決定。RとIはデータ数 $M=\{5,10,15,20\}$ においてそれぞれ1~20に変動させたものを参考に決定

- なぜ回帰を全てのパケットに対してではなく各MACアドレスの末尾のパケットに対してのみ行うのか
 - ほんの数秒前の移動情報(RSSIの変動)から端末の移動情報(RSSIの変動)を推測することは可能と思われるがそれを長時間の移動情報(RSSIの変動)から予測するとなると不可能なため