Computer Security Symposium 2017
23 - 25 October 2017

行動認証への無線LAN情報の活用

平岩 啓[1,2]　満保 雅浩[3]

概要：本論文では，利用が広がる Wi-Fi に着目し，ユーザが生活する中で得た Wi-Fi の情報をもとに行動認証システムを構成する．認証精度を導出する具体的な方法が示されていなかった著者らの先行論文に対して、本論文では，ユーザが記録する Wi-Fi の情報に対してクラスタリング分析とマルコフモデル化を順に行った後に，与えられたユーザのテストデータ系列が生成されたモデルにおいてどのように振る舞うかを数値化して，認証精度の議論につなげる．そして，提案手法の有効性を検証する評価実験として，本人拒否率と他人受入率の評価を行い，パラメータに対して適切な閾値を設定することで，提案手法が行動認証システムの構成に有効であることを示す．

キーワード：無線 LAN，Wi-Fi，個人認証，行動認証，リスクベース認証

How to Utilize Wireless LAN Information for Behavior Authentication

Satoshi Hiraiwa[1,2]　Masahiro Mambo[3]

Abstract: Behavior authentication uses unique behaviors inherent to user and has becoming paid more attention these days. In this paper, we discuss how to effectively use Wi-Fi information for behavior authentication. In our method, Wi-Fi information recorded by the user is analyzed by clustering, and the obtained cluster is modeled using the Markov model. Then, the behavior in the model of the given test data sequence is handled as a parameter to determine authenticity of the user. Through the evaluation experiments, we show that the proposed method is useful for constructing behavior authentication system by setting appropriate thresholds for parameters.

Keywords: wireless LAN, Wi-Fi, user authentication, behavior authentication, Risk-based authentication

1. まえがき

ユーザのライフログなどの行動に関する情報を認証に活用することへの注目が高まりつつある．このような認証は行動認証と呼ばれ，バイオメトリクス認証の一種として，一部の行動に関しては既に様々な製品化の取組みも行われている．例えば，ログイン中のユーザの行動として打鍵やタッチパネル，マウスの動作, GUI とのやりとりなどを用いたものが，行動バイオメトリクスによる継続認証 [1], ユーザ行動認証 [2], 行動 ID[3] などの異なる名称にて複数の企業より製品化されている．また，クレジットカードの支払いデータにユーザの応答などの行動データを加えたリスクベースの適応的認証 (Adaptive Authentication)[4] なども知られている．行動認証では，ユーザの行動から得られる情報をもとに，ユーザを分析，学習することで得られたユーザの癖などのユーザ独自の一意な特徴を本人確認に利用する認証方式である．

一方，近年の情報通信ネットワークの社会への浸透状況として，無線 LAN を介したインターネットへの接続を可能にする公衆無線 LAN が街中に増加している．中でも無線 LAN 通信規格の一つである Wi-Fi は，Wi-Fi 製品の充実や高速通信規格の誕生などを背景に，多くの場所で Wi-Fi の導入が広がっている．株式会社 ICT 総研の調査による

[1] 金沢大学自然科学研究科
Graduate School of Natural Science & Technology, Kanazawa University
[2] 現在、東日本電信電話株式会社に所属
[3] 金沢大学理工研究域
Institute of Science and Technology, Kanazawa University

と，2013 年度公衆無線 LAN の利用者は 1,702 万人にのぼり，Wi-Fi 通信機能が標準装備されたモバイル端末の国内出荷台数は 2016 年度には 5,423 万台に達するという [5].

本論文では，近年利用が広がる Wi-Fi に着目し，Wi-Fi の情報から得られた行動の特徴が，行動認証の要素となるか否かを検討する．具体的には，Wi-Fi のネットワーク識別子である BSSID と，そのネットワークの信号強度である RSSI からユーザの行動パターンを分析し，パターン化した行動をマルコフモデル化することで，ユーザを認証する行動認証手法を提案する．本論文と同様の情報を扱った既存研究 [11][12] では，Wi-Fi の情報が認証に有効であることを示すに留まっているのに対して，本論文では，実際に複数の被験者の Wi-Fi の情報を収集し，認証精度を導出する具体的な方法についても議論することで，提案手法の認証システムとしての有効性を示す．なお，提案手法は，行動に付随する情報をパスワードのように利用することで，ユーザにパスワード管理などの負担を掛けることがないためユーザビリティが高く，また，パスワードの推測も難しいことから，不正ログインを防ぐ効果もあると考えられる．

2. 予備知識

2.1 行動認証

行動認証は，日常的な動作，生活をもとにして認証を行うことで，パスワード認証方式のようにパスワードを記憶して管理する必要や，所有物認証のようにカードキーといったものを持ち歩く手間がなく，ユーザビリティが高い認証方式を実現できるという利点がある．一方，人の行動は常に一定ではないため，一つの行動により高い精度で認証できる保障が必ずしもないという欠点を持つ．そのため，行動認証は，複数の要素を用いた複合認証や追加認証を必要とするリスクベース認証などを行うことが，セキュリティを確保する上で必要となる．ここで，複合認証とは，複数の認証作業を行うことで初めてユーザを識別する認証方式であり，高い認証精度を実現できる．リスクベース認証は，最初に行う認証精度が低かった場合，その時に認証してしまうことへのリスクに応じて，追加で別の認証を要求する方式であり，成りすましを防ぐためのセキュリティ対策として利用される．これらの認証精度の向上手法を検討していく上で，まず，着目した行動についての認証精度を求める方法を十分に考察することが望まれる．行動認証のその他の欠点としては，ユーザの行動を分析し，学習するためにはまとまったデータが必要であり，学習データとその収集期間が必要となることが挙げられる．行動認証システムを適切に構成・運用していく上で，この学習に必要となるデータの収集期間を見積もることが求められる．

2.2 Wi-Fi 認証

Wi-Fi の情報を基にした行動認証は Wi-Fi 認証と呼ばれる．Wi-Fi は身近で収集しやすい情報であり，ユーザの位置とも関連するため，ユーザの行動を分析し，行動認証に用いることに適した情報だと考えられる．ユーザの位置情報という観点では，文献 [8] において，GPS による位置情報はユーザ推定が可能な情報であることが示されている．Wi-Fi の情報は，屋内における位置測位が GPS に比べて優れていることから，オフィス内や都市などにおいても Wi-Fi の情報を用いることによりユーザ認証が可能になると期待される．

小林らは文献 [9][10] において，Wi-Fi 認証方式を提案している．彼らは，ユーザは常に一定の行動パターンに従うだけでなく，パターンからずれた行動を行っていることを指摘し，画像処理における 2 値化手法を用いることで，ユーザの行動パターンを分析している．我々は小林らと独立に Wi-Fi 認証に着手しており，小林らが利用していなかった RSSI(受信信号強度) を用いていたため，文献 [11] において，細かなユーザの識別を行うと同時に，小林らの手法 [9][10] との比較を行い，提案手法の有効性を示している．以後，特に明示しない場合，Wi-Fi 認証とは RSSI を考慮した方式に限定して記述を行う．

既存の Wi-Fi 認証には幾つかの課題が残されている．1 つ目に，行動パターンの分析において，各時間帯と Wi-Fi の関係についてしか着目しておらず，通常のパターンから時間がずれた行動に対応できないなど認証精度に問題があった．また，時系列的な行動パターンを考慮していないことも挙げられる．2 つ目に，行動認証に関する課題として，ユーザの行動を分析するために必要な学習データの量についての議論がなされていなかった．3 つ目に，引越しや転勤など，急なユーザの環境の変化に対応することが難しいと考えられる．4 つ目に，Wi-Fi の情報はユーザの位置情報に関わるために，データ収集そのものに対するプライバシーの問題が挙げられる．

本論文では，これらの課題に対して以下の節にて考察する。1 つ目の課題に関して 3 節の提案手法を示し，4 節の評価実験手法を基に評価を行い，その結果を 5 節に示す。また，2 つ目の課題に関して 4.2 節において考察し，3 つ目と 4 つ目の課題に関して 6.2 節において考察する。

3. 提案手法

本論文では，時系列的な行動の順序にも，ユーザ独自の行動パターンが存在すると考え，時系列的なユーザの行動パターンをマルコフモデルによってモデル化した認証方式を提案する．マルコフモデルにより，ユーザの Wi-Fi の情報 (位置情報に相当) を状態として表現し，時系列的な行動の順序を状態遷移確率，さらに，我々が文献 [11] にて示すように RSSI にも着目し，RSSI を各状態における出力記号として表す．マルコフモデルについては文献 [7] を参照されたい．マルコフモデルにおける，状態の様子と状態遷移

確率，出力記号の出力確率の 3 つを認証時のパラメータとして扱い，ユーザ認証の可否に利用する．

さらに，予備的な評価実験として，モデルの学習期間を調査研究する．これにより，既存研究において課題であった環境の変化への対応とプライバシー保護についても考察する．

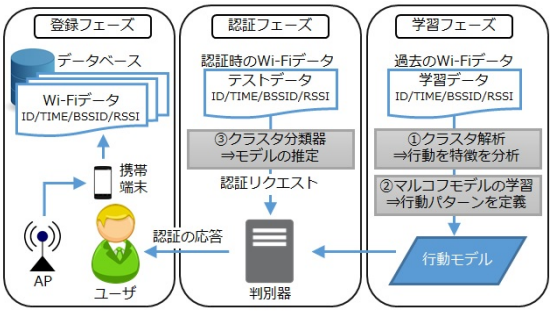

図 1　提案手法の概要図

Fig. 1　Overview of the proposed scheme

3.1　Wi-Fi の取得，登録フェーズ

図 1 の登録フェーズでは，ユーザは，スマートフォンのセンサを用いて，周辺にある Wi-Fi から BSSID とその RSSI を取得する．取得する Wi-Fi は，ユーザがデータ通信を相互に行うアクセスポイント (以降，AP) に限らず，ビーコンをもとに観測した周辺の AP 全てから取得する．また，Wi-Fi の情報は一定時間間隔ごとに取得され，データベースには，BSSID，RSSI に加えて，ユーザ ID と取得時刻を登録する．

ここで，ユーザがある時刻 i に取得する Wi-Fi のデータを S_i とすると，S_i は要素として，その時刻で観測した BSSID $bssid_k$ とその RSSI $rssi_k$ を持ち，

$$S_i = \left\{ \begin{array}{cccc} bssid_1 & bssid_2 & ... \\ rssi_1 & rssi_2 & ... \end{array} \right\}$$

のように表す．また，ユーザがある期間 T にわたり取得した Wi-Fi のデータを学習データ L とすると，L は長さ T の Wi-Fi のデータ S_i の状態系列であり，

$$L = S_1 S_2 ... S_i ... S_T$$

のように表す．

3.2　モデルの学習フェーズ

本論文におけるマルコフモデルは，以下の 4 つのパラメータによって構成される．

- $Q = \{q_1, q_2, ..., q_n\}$：状態の有限集合
- $V = \{v_1, v_2, ..., v_m\}$：出力記号の有限集合
- a_{ij}：状態 q_i から状態 q_j への遷移確率
- b_{jk}：状態 q_j で v_k を出力する確率

図 1 の学習フェーズでは，与えられた学習データ L に対して，凝集型階層的クラスタリング [6] を行うことで，ユーザの特徴となる行動パターンを分析し，得られたクラスタを要素とする状態集合 Q を求める．さらに，別の学習データ $L' = S'_1 S'_2 ... S'_i ... S'_{T'}$ の状態系列がどのように状態の遷移を行っているか，かつ，各 S'_i が状態集合 Q における何れの状態に分類されるのかを求めることで，遷移確率 a_{ij} と各状態における出力記号の確率 b_{jk} を求める．出力記号は RSSI の数値とする．本節では，それぞれの解析手法について述べる．

3.2.1　クラスタリングによる状態集合

本節では，ユーザの事前に蓄えられた Wi-Fi の BSSID を学習データとして，Wi-Fi の BSSID と新たに提案する補助パラメータ TF をもとに凝集型階層的クラスタリングを行う手法について述べる．凝集型階層的クラスタリングについては文献 [6] を参照されたい．

(1) クラスタリングの学習データ L_C

学習データとして，BSSID の長さ T の状態系列

$$L_C = S_1 S_2 ... S_i ... S_T$$

が与えられたとき，各状態 S_i を初期クラスタとして，凝集型階層的クラスタリングを行う．

(2) 補助パラメータ：TF

クラスタ間距離を計算する準備として，補助パラメータの TF(Term Frequency; 出現頻度) を考える．TF TF_k は，学習データ L_C 中において，ある BSSID が出現する (記録されている) 回数を表し，

$$TF_{bssid_k} = \sum_{i=1}^{T} C(bssid_k \in S_i)$$

で定義する．ここで，$C(bssid_k \in S_i)$ は状態 S_i 中における $bssid_k$ が記録されている回数を表す．TF が高い BSSID ほど，ユーザはその BSSID の Wi-Fi 環境に多く (長く) 位置していたことになり，ユーザの行動パターンを分析する上で重要なパラメータとなる．一方で，T F の値が低い BSSID はユーザの行動パターンとして偶発的な要因に起因していることが多いと考えられる．そのため，(1) の初期クラスタの状態で，クラスタ内の全ての BSSID の TF が，学習データの件数 T の 0.3% 未満であれば，そのクラスタを削除する．

(3) 重み付きクラスタ間距離

クラスタ間距離は，2 つの集合の類似度を表す Jaccard 係数 [6] をもとに計算する．ただし，本提案手法では，パラメータ TF を用いて，要素ごとに重み w をつけてクラスタ間距離を求める．これにより，ユーザの行動パターンの重要性に即したクラスタ間距離の計算ができ，TF の大きい BSSID がクラスタの中心となるようなクラスタとなる．

クラスタ S_i, S_j のクラスタ間距離の計算方法を示す．このとき，全 $bssid$ 数 $n = |S_i \cup S_j|$ であり，TF の総和 $Sum.TF$ と平均値 $Ave.TF$ は，

$$Sum.TF = \sum_{k=1}^{n} TF_k \ , \ Ave.TF = \frac{Sum.TF}{n}$$

となる．ここで各 $bssid_k$ の重み w_k は

$$w_k = TF_k$$

とする．つまり，クラスタ S_i, S_j におけるクラスタ間距離 $Distance$ は，

$$Distance = 1 - \frac{(S_i \cap S_j \text{に属する bssid の重み w の総和})}{Sum.TF}$$

となる．また，本論文では，全てのクラスタ対の距離が“0.9”よりも大きくなるまでクラスタリングを繰り返し，クラスタの併合時には，クラスタの要素数を重みとした群平均法により，他のクラスタとの距離を更新する．

3.2.2 マルコフ過程によるモデル学習

本節では，ユーザが取得する Wi-Fi の BSSID と RSSI 値を学習データとして，単純マルコフモデルをどのようにモデル化するのかについて述べる．

(1) 状態集合 Q と出力記号集合 V

状態の有限集合 Q の各状態 q_i は 3.2.1 節のクラスタリングによって得られたクラスタとなる．また，出力記号の有限集合 V の各記号 v_i は，RSSI の値となる．

(2) 状態 q_i における主キー

主キーは，その状態 (クラスタ) において最も TF が大きい BSSID とし，その状態において，最もユーザの行動パターンを端的に表す BSSID となる．マルコフモデルにおける各状態の出力記号は，主キーとなる BSSID の RSSI の値とする．

(3) 状態遷移確率の学習データ L_M

学習データとして BSSID と RSSI の長さ T の状態系列

$$L_M = S'_1 S'_2 ... S'_i ... S'_{T'}, t = 1, 2, ..., T'$$

が与えられたとき，この状態系列が，状態集合 Q においてどのように状態遷移をしているのかを分析する．

(4) 学習データ L_M の状態系列のクラスタ推定

本論文では，あるデータ S がどのクラスタに分類されるのかを判断することをクラスタ推定と呼び，クラスタ推定をもとに学習データ L_M の各状態 S'_i が，マルコフモデルの状態集合 Q においてどの状態に分類されるのかを決定する．さらに，決定し得られた学習データ L_M の，状態集合 Q における状態遷移をもとに，マルコフモデルの状態遷移確率と出力記号の出力確率を学習する．

本提案手法では，学習データ L_M の各状態 S'_i に対して，3.2.1 節のクラスタリング同様に全クラスタとのクラスタ間距離を計算し，最もクラスタ間距離が小さくなるクラスタ対のクラスタに，S'_i を分類する．このときにクラスタの主キーとなる BSSID を学習データ S'_i が持っていれば，その RSSI を対応する状態における出力記号系列とする．

ただし，学習データ L_M の状態系列のクラスタ推定は，逐次処理とし，時刻 1 から T' までの学習データに対して，順に処理を行う．そのため，学習データ L_M は，補助パラメータ TF を持たない．S'_i における各 BSSID の TF は，クラスタ間距離を計算する際，クラスタ対と共通して持つ BSSID は同じ TF の値をとり，共通して持たない BSSID の TF は $Ave.TF$ として扱う．また，S'_i の最短クラスタ間距離が 1 となった場合は，S'_i を新たにマルコフモデルの状態の有限集合 Q に加える．

3.3 ユーザの認証フェーズ

本節では，判別器でのモデルとテストデータの比較処理について述べる．ここでテストデータを AD とすると AD は，認証リクエストを送信するタイミングから直近の TL 時間における Wi-Fi の取得データ S_i の履歴であり，

$$AD = S_1 S_2 ... S_i ... S_{TL}$$

のような状態系列で表される．図 1 の認証フェーズでは，判別器は入力として，3.2 節までで学習したマルコフモデルの状態集合 Q と状態遷移確率行列 A，状態の出力記号確率行列 B，さらにユーザからテストデータ AD を受け取り，テストデータの S_i ごとに以下の 3 つの値を求める．

(1) クラスタ推定の精度 $CE(i)$

マルコフモデルの学習フェーズと同様にテストデータ AD の状態系列のクラスタ推定を行う．テストデータ中の状態 S_i が分類されるクラスタとの，クラスタ間距離を $CE(i)$ とする．

(2) 状態系列 AD の遷移確率 $TP(i)$

状態 S_{i-1}, S_i などを持つテストデータ AD に対してクラスタ推定を行い，例えば，テストデータの状態 S_{i-1} の属するクラスタの状態 S'_{i-1} やテストデータの状態 S_i の属するクラスタの状態 S'_i などが得られる。これらクラスタの状態の遷移確率を学習したマルコフモデルの状態遷移確率行列 A から抽出し，テストデータでの状態 S_{i-1} から S_i への遷移確率 $TP(i)$ とする．

(3) 各状態における出力確率 $OP(i)$

状態 S_i の主キーにおける RSSI を，出力記号確率行列 B における確率とで照会し，得られた確率を出力確率 $OP(i)$ とする．

テストデータの S_i における 3 つの値 $CE(i), TP(i), OP(i)$ の算出を，テストデータ AD の

長さ TL だけ行い，以下の式 (1),(2),(3) のように値を統合する．

$$CE = \frac{\sum_{i=1}^{TL}(1 - CE(i))}{TL} \quad (1)$$

$$TP = \prod_{i=2}^{TL} TP(i) \quad (2)$$

$$OP = \frac{\sum_{i=1}^{TL} OP(i)}{TL} \quad (3)$$

(2) 式は状態遷移の同時確率分布を表し，(1), (3) 式は各状態における平均を表している．上記の 3 つの値 CE, TP, OP を一致度と呼び，判別器では，この 3 つの一致度をもとにしてユーザの認証の可否を決定する．

ここで，上記の式 (1) が乗算であることに起因するゼロ頻度問題について触れる．ゼロ頻度問題 [13] とは，テストデータの状態系列の遷移が学習したモデルにおいては未知の遷移であるとき，その状態遷移確率が“0”であるために，最終的な演算結果も“0”となってしまう問題である．この問題は学習データが少なく，ユーザの行動の学習が十分にできていない場合に起こりやすい．本論文では，ゼロ頻度問題への対策として，事前にモデル学習に必要な期間を調査することで，適切な学習データ量での学習を行う．また，発生した際は，一定値を簡易的に与えることで乗算結果が 0 となるのを防ぐヒューリスティックな手法を適用する．本論文では，ゼロ頻度問題が発生した際のクラスタの数を $N_{Cluster}$ として，

$$\frac{1}{N_{Cluster}}$$

を一定値として与える．

4. 評価実験手法

本節では，まず実験に用いるデータの環境について述べた後，モデル学習に必要なデータ量についての評価実験を行った内容について考察をする．その後，本提案手法の評価実験手法について述べる．

4.1 評価実験の環境

Wi-Fi のデータを取得するためにスマートフォン向けアプリケーションを開発した．アプリケーションは，10 分間隔で周辺の Wi-Fi をキャッチし，SSID，BSSID，RSSI，データの取得日時を，実験用に設置したサーバーへと送信する．ただし，アプリケーションが Wi-Fi をキャッチできなかった場合，SSID，BSSID を“null”，RSSI を“0”として記録し扱う．

また，データの取得には，同一の大学に所属している学生 5 名に協力してもらい，被験者 5 名が普段から使用しているスマートフォンに開発したアプリケーションをインストールし，データを取得する．ただし，被験者のスマートフォンの電源が落ちていることやアプリケーションの起動し忘れなどの要因によって，データの取得が行われない期間や，通信環境によっては取得したデータの送信ができず，データが抜け落ちていることがある．また，ユーザの環境によって，周辺の Wi-Fi の状況も異なり，データの取得件数は同一ではない．

被験者 5 名から取得された Wi-Fi のデータ件数とその取得期間を表 1 に記載する．4.1.2 節で述べたように，データには途中抜け落ちたときがあったものの，実用上も起こりうる問題であるため，データの調整は行わず，欠損したままのデータを実験に用いる．

表 1　データの取得日数と取得したデータ件数
Table 1　Number of data acquisition dates and number of acquired data

被験者	データ取得日数	取得されたデータ件数
1	140	240234
2	128	360225
3	123	353120
4	164	407721
5	132	172412

4.2 学習に必要なデータ量について

評価実験のユーザ認証を行う準備として，学習フェーズで必要となる学習データの量を導出する．

4.2.1 必要な学習データ量の検証方法

文献?で指摘されているように，ユーザの位置情報には決まったパターンがあり，この決まった場所を全てマルコフモデルにおける状態として表されたとき，十分な学習ができたとする．また，ユーザのパターンには限りがあり，決まった場所の数にも上限があると考えられる．そこで，3.2.1 節にて述べたクラスタリングを，学習データ L_C のデータ量 (データの収集期間) を 2，4，8，16，32，64 日の 6 通りに設定し，それぞれ行う．さらに，クラスタリング後に残ったクラスタ数と学習データ L_C の日数から，学習データの日数が 1 日増えるごとのクラスタ数の増加率を求める．行動によるばらつきを抑えるために，各被験者の取得データを設定した学習データ日数で分割し，最大限の試行回数を試す．以降では，平均値をとったものを結果として示す．

4.2.2 学習データ量による違い

結果を図 2 に示す．図 2 では，各被験者でのクラスタリング結果と，被験者 5 名の増加率を平均したものを第 2 軸に示している．学習データ日数が 10 日未満では，急激にクラスタ数が伸びており，その後も学習データ数が増えるほど，クラスタ数は増える傾向にあった．しかし，どの被験者においても，ほぼ同じように学習データが増加するにつれて，クラスタ数の増加率は低下し，クラスタ数も収束

する傾向にある．

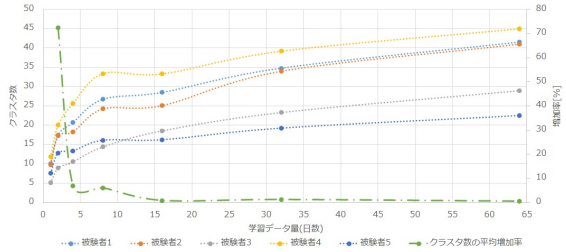

図 2　学習データ量とクラスタ数
Fig. 2　Amount of learning data and the number of clusters

学習データ量 (収集期間) が少ないときでは，ユーザの行動パターンの一部しか観測されていないために，クラスタ数も増加する傾向にあり，学習データ量が 32 日間前後では，ユーザの限りある行動パターンの多くが観測されたと予想されるために，クラスタ数も横ばいになったと考えられる．以上の評価実験により，本論文でのモデル学習に必要な日数は，半月からひと月だと分かった．以降では，学習期間を 32 日間とする．

4.3　評価実験手法

評価実験では，各被験者の取得データに対して，モデルの学習と各一致度 CE, TP, OP の算出を行う．その際に，同一被験者の組み合わせだけでなく，他の被験者による一致度の算出も行い，本提案手法の不正ログインへの耐性も評価する．

また，ユーザの行動のばらつきによる精度差を減らして正しく評価するために，各被験者の取得データを 2 分割して，$k = 2$ の k-交差検定を行い，各試行の平均値を結果として示す．$k = 2$ とすることで，学習データ L_C, L, M の長さはともに，約 32 日間となり，各被験者ごとに 2 通りの行動モデルを学習する．一方の，テストデータは，テストデータ AD の長さ TL を 6，12，18，36(1 時間，2 時間，3 時間，6 時間) として，TL の長さで，交差検定の一方のデータを分割し，複数のテストデータ AD を用意する．ユーザによって異なるものの，同一被験者同士での認証フェーズの試行回数は，$TL = 6$ で約 3,072 回，$TL = 12$ で約 1,536 回，$TL = 18$ で約 1,024 回，$TL = 36$ で約 512 回である．ただし本稿では，頁数の都合上，最も長さが短く有用であると考えられる $TL = 6$ の場合での結果のみを示す．

5.　評価実験結果

5.1　パラメータによる尤度

各一致度 TP, OP, CE の様子を表 2，3，4 に示す．各表中の，薄く塗りつぶした枠は，対応する学習モデルに対して最も高い一致度を示した結果であり，表中での $E-XX$ は，10^{-XX} を表す．また，本論文中では，ある被験者の学習モデルが，どの被験者のテストデータとの間で，最も高い一致度を示し，学習モデルの被験者本人らしいかを表すものを尤度という．

表 2　一致度 TP の結果 (状態遷移確率)
Table 2　TP (Transition probability)

学習モデル＼テストデータ		テストデータ長：6				
		1	2	3	4	5
被験者	1	5.53E-01	1.96E-02	1.26E-02	6.58E-02	1.72E-03
	2	1.72E-02	5.11E-01	2.83E-03	2.38E-02	9.53E-04
	3	2.46E-01	3.48E-02	2.04E-01	7.09E-02	1.05E-03
	4	6.91E-03	1.53E-02	2.11E-03	1.39E-01	2.23E-02
	5	4.07E-03	9.26E-04	1.75E-03	6.20E-03	5.86E-01

表 3　一致度 OP の結果 (出力記号確率)
Table 3　OP (Output probability)

学習モデル＼テストデータ		テストデータ長：6				
		1	2	3	4	5
被験者	1	4.53E-02	1.24E-02	9.88E-03	1.19E-02	1.13E-02
	2	8.64E-03	6.04E-01	9.22E-03	1.16E-02	8.07E-03
	3	3.11E-02	1.91E-02	6.03E-02	2.37E-02	1.75E-02
	4	8.41E-03	9.85E-03	1.06E-02	7.52E-02	7.47E-03
	5	1.38E-02	1.29E-02	1.15E-02	1.18E-02	7.87E-02

表 4　一致度 CE の結果 (クラスタ推定精度)
Table 4　CE (Cluster Estimation precision)

学習モデル＼テストデータ		テストデータ長：6				
		1	2	3	4	5
被験者	1	4.68E-01	1.09E-01	4.89E-02	1.60E-01	3.56E-02
	2	2.29E-01	6.40E-01	4.51E-02	2.22E-01	7.34E-02
	3	1.58E-01	9.53E-02	6.06E-01	1.52E-01	1.69E-02
	4	5.55E-02	8.53E-02	5.38E-02	5.75E-01	1.45E-02
	5	3.44E-02	4.90E-02	1.24E-02	4.43E-02	7.14E-01

表 2 では，ほとんどの被験者において，本人の尤度が最も大きい値を示した．しかし，一部の被験者の学習モデルにおいては，他の被験者のほうが高い尤度を示す結果となった．表 3 では，本人同士の尤度が最も高い結果を示した．また，被験者 2 では，本人と他人との尤度に大きく差があり，本人らしさを高い精度で示した．他の被験者の尤度では，本人が最も高いものの，他人との差はあまりない結果となった．表 4 では，これまでの表 2，3 で示した結果よりも，本人において高い尤度を示している．このことから，一致度 CE はユーザ認証に最も有効な尺度であると考えられる．

以上のように，どの一致度においても，概ね本人において高い尤度を観測することが出来た．

5.2　各一致度による本人拒否率と他人受入率

ユーザ認証の合否は，算出した一致度に対して閾値を設定することで判断され，一致度が閾値以上であれば認証可となる．本人拒否率 (False Rejection Rate, FRR) とは，ユーザ認証において，本人の情報を照合したにも関わらず，本人であると認識できず，認証拒否されてしまう割合であり，他人受入率 (False Acceptance Rate, FAR) とは，他人

の情報を照合した時に，誤って他人を本人だと誤認してしまい，受け入れてしまう割合である．本人拒否率と他人受入率の両方が低いシステムほど，認証精度が高いとされるが，二つの値はトレードオフの関係にある．

5.1 節で示された各一致度に対して，それぞれの閾値 k_A, k_B, k_C を変えながら，本人拒否率と他人受入率のトレードオフの関係を図 3，4，5 に掲載する．

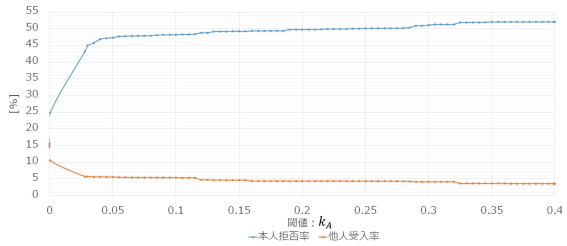

図 3 一致度 TP における本人拒否率と他人受入率
Fig. 3 FRR and FAR for the parameter TP

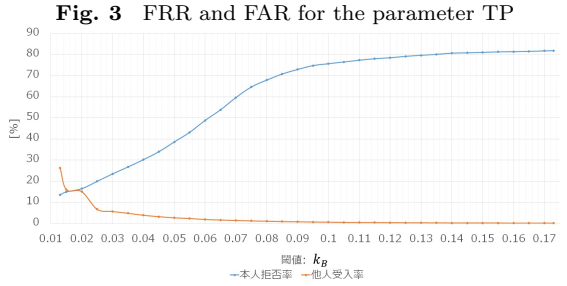

図 4 一致度 OP における本人拒否率と他人受入率
Fig. 4 FRR and FAR for the parameter OP

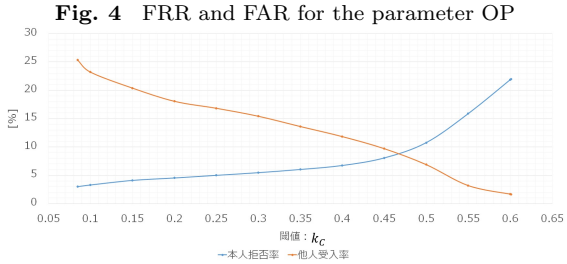

図 5 一致度 CE における本人拒否率と他人受入率
Fig. 5 FRR and FAR for the parameter CE

図 3 では，閾値 k_A がどの値においても，他人受入率が 10%未満でほぼ一定の値を示した．それに対して，本人拒否率は常に 50%近く，非常に高い値を示している．また，一致度 TP に関して，閾値 $k = 5.0E-06$ と小さく設定すると，本人拒否率が約 15%，他人受入率が約 15%と交差する結果になった．図 4 では，閾値 $k_B = 0.02$ 前後で，本人拒否率と他人受入率が交差する結果となった．また，本人拒否率は緩やかに右肩上がりに上昇しており，$k_B = 0.06$ では本人拒否率が約 50%と高い割合になっている．一方，他人受入率は閾値 $k_B = 0.025$ 以降は 10%を下回り，閾値 k_B を大きくした時は，ほとんど 0%を示した．図 5 では，閾値 $k_C = 0.45$ の時に，本人拒否率と他人受入率が交差していることが読み取れる．他人受入率は閾値 k_C が十分に小さいときでも 20%ほどであり，その後は右肩下がりの傾向にある．本人拒否率は，他の一致度に比べて，閾値 k_C を大きく設定しても低い結果となった．

5.3 提案手法としての認証精度

3 つの一致度に対して具体的な閾値を設定し，統合した際の，被験者ごとの認証受入率の結果を，表 5 に示す．ここで，認証受入率とは，ユーザ認証において，本人か他人かに関わらず認証を受け入れる割合である．

各一致度の閾値は，図 3，4，5 にて，本人拒否率と他人受入率が交差したときの閾値を設定した．

表 5 認証受入率 [%]
Table 5 Acceptance rate of authentication[%]

テストデータ ＼ 学習モデル		テストデータ長：6				
		1	2	3	4	5
被験者	1	62.89	7.09	2.05	2.35	0.07
	2	0.61	80.20	1.62	4.37	0.07
	3	14.66	6.86	86.94	19.29	0.00
	4	1.22	2.38	4.73	70.28	0.00
	5	0.10	0.41	0.00	0.02	76.79

表 5 では，本人受入率が高い被験者では，86%を示した．しかし，多くの被験者において他人受入率が 10%未満であるのに対し，被験者 3 の学習モデルにおいては他人受入率が 20%近くを示した場合があった．

6. 考察

6.1 認証精度について

5.1 節の表 2，3，4 の一致度の結果において，ほとんどの被験者では，本人同士のデータから算出された一致度が最も高い値を示しており，提案する手法によりユーザの識別が可能であると考えられる．一方，一部の被験者においては他人のほうが高い結果を示した．これは，被験者 5 名が同じ大学に通う学生であり，さらに被験者 1〜4 は同じ研究室に在籍する学生であったために，平日の日中の行動が類似していた事が原因の一つだと考えられる．しかし，3 つの一致度全てにおいて，他人のほうが高い尤度を示した被験者はおらず，3 つの一致度の組み合わせにより，高い精度でユーザの識別ができることが期待される．

5.3 節の表 5 では，3 つの一致度を統合することで，高い本人受入率を保ちながらも，多くの被験者においては低い他人受入率を示した．したがって，3 つの一致度を組み合わせることと，適切な閾値を設定することにより本提案手法はユーザ認証に利用可能だと考えられる．

しかし，表 5 で示す結果のように，本人受入率は平均で 75%であり，他人受入率が 10%以上の値を示した場合があった．このように，本人と他人とで認証精度には差があるものの，単体で認証が可能なほど十分な精度には達していない．そのため，リスクベース認証のような，一致度の

値次第では，追加認証を求めることで，認証を行うことが求められ，同時に成りすましを防ぐこともできると考えられる．

6.2 学習モデルの更新とプライバシーについて

ユーザが取得する Wi-Fi の情報は，位置情報とほぼ同義であり，蓄積した情報によりプライバシー侵害に繋がる恐れがある．この問題に対して，本論文では，ユーザ ID の更新により問題の解決を検討する．4.2 節で導いたように，ユーザの行動パターンは約一ヶ月で学習可能である．つまり，ひと月でユーザの特定が可能なほどの情報が蓄積されているとも考えられる．そこで，約ひと月のスパンでユーザ ID を変更し，新たな ID でデータの収集を行うことで，それまでに蓄積されたデータから個人が特定されるのを防ぎ，さらに定期的なモデルの更新により，行動パターンの変化にも対応することが可能になると考える．

7. むすび

本論文では，Wi-Fi から予想されるユーザの行動を利用したユーザ認証方式を提案し，提案方式が行動認証の要素となるかを確認するための実験を行った．

ユーザの行動パターンを分析するためにクラスタリングを行い，ユーザが居ることの多い場所での Wi-Fi の様子を解析する手法，さらに，ユーザの時系列的な行動パターンにも注目し，マルコフモデルをもとにユーザの行動をモデル化することを提案した．そして評価実験を通して，学習モデルとテストデータから算出した 3 つの一致度から，本人の尤度や本人拒否率，他人受入率を求めた．その結果から，一致度に対して適切な閾値を設定することで，提案手法が，行動認証を実現する上で有用な構成要素であることを導いた．

今後は，単純マルコフモデルだけでない，より複雑な時間関係を表す手法を導入することにより，時刻と状態の関係性を探す必要があると考えられる．また，本論文では，Wi-Fi の情報を基にした行動認証を提案して評価したが，行動認証に利用可能な情報は様々である．今後は，他の認証要素と組み合わせることで，認証精度の向上やよりユーザビリティが向上されることを目指す．

謝辞 本研究成果の一部は、国立研究開発法人情報通信研究機構 (NICT) の委託研究「Web 媒介型攻撃対策技術の実用化に向けた研究開発」により得られたものです。

参考文献

[1] BehavioSec, Continuous Authentication with Behavioral Biometrics, https://www.behaviosec.com/. 2017 年 3 月 17 日確認.
[2] EZMCOM, Entity User Behavior Authentication - Keystroke Dynamics, https://www.ezmcom.com/products/behavioralauthentication/. 2017 年 3 月 17 日確認.
[3] TeleSign, Behavior ID SDK, Combat Account Takeover with Continuous Protection, https://www.telesign.com/products/behavior-id/. 2017 年 3 月 17 日確認.
[4] RSA, RSA Adaptive Authentication For ecommerce - Risk-based 3D Secure for Credit Card Issuers, https://www.rsa.com/content/dam/rsa/PDF/2016/05/AAeComm-SB-0516.pdf. 2017 年 3 月 17 日確認.
[5] ICT 総研, “2013 年度公衆無線 LAN サービスの利用動向調査”, 2013 年 12 月発刊.
[6] 神嶌敏弘, “クラスタリング Clustering”, www.kamishima.net/archive/clustering.pdf. 2017 年 3 月 17 日確認.
[7] 石井健一郎, 上田修功, “続・わかりやすいパターン認識”, 株式会社オーム社, pp123-132, 平成 26 年 8 月 25 日第 1 版第 1 刷発行.
[8] 船越琢矢, 満保雅浩, 位置情報のユーザ識別への活用, 電子情報通信学会技術研究報告, ISEC2014-68, SITE2014-59, LOIS2014-38, Vol.114, No.319, pp.71-76 (November 22, 2014).
[9] 小林良輔, 山口利恵, “Wi-Fi 履歴情報を活用した複合認証における個人認証手法”, Computer Security Symposium 2015, 3D1-2, October 2015.
[10] 小林良輔, 山口利恵, “A Behavior Authentication Method Using Wi-Fi BSSIDs around Smartphone Carried by a User”, 3rd International Symposium on Computing and Networking (CANDAR) 2015, pp.463-469, IEEE December 2015. (Presented at the associated workshop, the 2nd International Work-shop on Information and Communication Security)
[11] 平岩啓, 満保雅浩, “無線 LAN 情報の認証への応用の検討”, 電子情報通信学会論文誌 D, Vol.J99-D, No.10, pp.1034-1044 (October 2016).
[12] 平岩啓, 満保雅浩, “時系列データの類似度検索手法のユーザ認証への応用”, 電子情報通信学会技術研究報告, ISEC2016-46, SITE2016-46, LOIS2016-34, Vol.116, No.289, pp.33-38 (November 7, 2016).
[13] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, “Introduction to Information Retrieval”, Cambridge University, pp259-260, 2008.