

時系列データの類似度検索を用いたユーザ認証の検討

平岩 啓¹ 満保 雅浩²

概要: 公衆無線 LAN の増加とともに、生活のほとんどの場において無線 LAN 通信が利用できる環境が整いつつある。また、近年ライフログへの注目が集まっており、ライフログの活用が盛んになっている。しかし、ライフログにはノイズや異常値が多く含まれることが考えられる。本論文では、ノイズに強い順序保存照合を時系列データの解析に用いることで、無線 LAN から得られた情報を時系列データとして扱い、ユーザ認証へ利用することを検討する。具体的にはスマートフォン端末上で取得した、電波強度を含む無線 LAN 情報の時系列データにおいて、類似するパターン行動を検索することで、認証への活用の可能性を検討する。

キーワード: 無線 LAN, Wi-Fi, 認証, リスクベース認証, 順序保存照合

User authentication using similarity search of the time-series data

SATOSHI HIRAIWA¹ MASAHIRO MANBO²

Abstract: Wireless LAN communication is available in most places with the increase of the public wireless LAN. In addition, life log has received much attention and its utilization has become popular. However, life log contains many noises and outliers. In this paper, we examine whether user authentication is possible from time-series data obtained from wireless LAN by using order-preserving matching in the data analysis.

Keywords: wireless LAN, Wi-Fi, authentication, Risk-based authentication, Order-preserving matching

1. はじめに

ライフログとは生活の中で得られる多種多様な情報を記録したデータ、あるいはその記録する技術のことである。近年、そのライフログに注目が集まり、その市場規模も拡大している。2011 年度の市場規模は 10 億円超を見込み、2015 年度には 52 億円にも及ぶという [1]。ライフログは GPS の位置情報や Web の閲覧履歴、購買履歴、さらにはジャイロセンサーによって測定したユーザーの細かな動きまで多種に及ぶ。こうしたライフログを活用することで各ユーザーは、自己に適したサービスの提供を受けること

が可能となっている。

現代において個人認証を必要とする場面は多い。代表的な認証方式として予め設定した ID とパスワードを正しく答えられるかで本人か否かの識別を行う知識認証方式（パスワード方式）がある。この方式はシステム導入のコストが低く、多くの場所で利用されている一方で、問題点もある。1 つに ID とパスワードの設定、管理がユーザーの負担となることだ。ユーザーは予め ID とパスワードを設定し管理する。その際にパスワードが推測されることや漏洩することへの対策として、複雑な文字列を設定し記憶する必要があり、ユーザビリティを下げる原因となっている。

これらの問題に対して注目されるのが、ライフログなど行動情報を活用した複合認証である。複合認証では、複数の認証作業を行うことで初めてユーザーが判別される。特徴として、認証精度が高くない認証方式を複数組み合わせることで、高い認証精度を実現できる。また、リスクベース

¹ 金沢大学 大学院 自然科学研究科, 石川県
Kanazawa University Graduate School, Kakuma, Kanazawa,
Isikawa, 920-0911 Japan

² 金沢大学 理工研究域, 石川県
Kanazawa University, Kakuma, Kanazawa, Isikawa, 920-
0911 Japan

認証とは認証におけるリスクの程度に応じて、追加の(複数の)認証を要求する方式である。複合認証により、パスワードの安全性が低い場合でも追加の認証作業により、正しく認証が可能になる。

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。NTTdocomo は、2015 年 2 月時点で全国 121,900 エリアに 152,100 個ものアクセスポイントを設置し、基本無料の Wi-Fi サービスを行っている [2]。同様に、自宅や職場、学校、ショッピングセンターなど生活するほとんどの場においても無線 LAN アクセスポイントが設置され、スマートフォンでも容易に Wi-Fi 通信をすることが可能である。

複合認証に用いる行動情報に、無線 LAN から得られる情報に着目した研究が行われている [3][4]。無線 LAN 情報は普通の生活においても自動的に取得可能であるため、パスワードのように設定管理する必要がなく、ユーザビリティが高い認証方式が実現出来る利点がある。

本論文では、無線 LAN を認証に利用するにあたって、ノイズに強い順序保存照合を新たに活用可能かを検討する。電波強度などの無線 LAN 情報には、ノイズが多く含まれている為、認証精度の向上に役立てる。本論文の無線 LAN 情報とは、Wi-Fi 規格通信を提供するアクセスポイント(以後 AP と呼ぶ)から発信されるビーコン信号のことであり、実際にユーザーがデータ通信で利用する AP に限らず周囲に点在する全ての AP の情報を用いる。具体的には、スマートフォン端末に実装した実験用アプリケーション(以後アプリと呼ぶ)により AP の SSID と BSSID 及びビーコン信号の電波強度を時系列とともに記録し、その時系列データをパターン照合によって個人識別に用いる。

2. 予備知識

2.1 先行研究

無線 LAN 情報を認証に用いた研究に文献 [5][6] がある。文献 [5][6] では、ユーザは時間や曜日に依存した規則的な行動パターン [7] を所持していると考え、複数の日の中からユーザの特徴となるデータを学習データとして認証に用いている。学習データは、複数の日のうちで同じ時間帯に記録され、その記録回数があるしきい値以上である無線 LAN 情報が抽出されたものである。その学習データと実際のユーザの 1 時間ごとの行動とを検証して、ユーザの認証を行う。検証では、同じ AP の BSSID を持ち、その信号強度の差がある範囲以内であれば、その BSSID は一致しているとする。学習データのデータ件数に対して、一致した BSSID の割合からユーザが同一か否かを判断している。文献 [5][6] ではある程度の認証精度が示されているが、3 つの問題点がある。

(1) 認証に用いる時間が長い

認証では 1 時間から 24 時間の無線 LAN 情報を必要と

するために、スマートフォンが盗難にあった直後では成りすましが可能となってしまう。

(2) 検証時や学習データの作成時に、しきい値が必要
しきい値によって認証結果が変化するため、適切なしきい値の設定が必要となっている。

(3) 時間に依存した行動しか認証できない
文献 [] では同じ時刻どうして比較検証を行うため、行動のパターンが一致する場合であっても、時間が異なれば異なるユーザと判断してしまう。

2.2 順序保存照合問題

本論文で用いるパターン照合の 1 つである順序保存照合について説明する。ただし、本論文の時系列データとはすべて一定時間ごとに取得された数値を並べたものとする。順序保存照合は、パターン照合における検索したい“パターン”を順位で表現して、検索対象である“テキスト”内で同じ順位で表現される系列を似ていると判断して、パターン照合を行う手法である。順序保存照合の特徴として以下が挙げられる。

- シフト、ノイズ、異常値に強い
- 思いがけないものを似ていると判断する
- しきい値がいない

3. 提案手法

3 章では、無線 LAN から得られる情報をもとに個人認証を行う手法について述べる。ユーザが認証時に得た無線 LAN 情報をテストデータ、テストデータの系列を順位にして表したものをテストパターン、それまでの生活の中で蓄積された無線 LAN 情報をテキストとする。認証では、テストパターンをテキスト中から検索していき、順序保存照合を行うことで、ユーザが同一か否かを判断する。

3.1 テキストの作成

ユーザは規則的な行動パターンに基づき行動していることが知られている [7]。行動パターンはユーザによって異なっており、行動パターンがユーザの特徴となっている。しかし、ユーザは常にまったく同じ行動をとるものではなく、行動パターンには“揺らぎ”が存在する。“揺らぎ”とは普段活動しない場所での活動など、行動パターンに含まれないイレギュラーな行動のことである。正しくユーザを識別するためにも、揺らぎを除去したテキストの作成を行う。無線 LAN 情報は行動パターンの規則性のために、取得回数が高いものほどユーザの行動に沿った特徴的なデータであると考えられる。一方で、取得回数が低いものはユーザーの行動パターンに合致した行動とは言えない。そのため、テキストとなる無線 LAN 情報の中から文献 [4] のアドレスの選定と同じ手法を用いて、取得回数から BSSID の選択を行った。

3.2 テストパターンの作成

テストデータに必要な時間が短ければ短いほど、認証の利便性が高いと考える。多くの場所で Wi-Fi 通信が利用できる環境が整ってきているため、認証時のユーザのスマートフォンから取得されたテストデータには 1 度に複数の AP が含まれていると予測される。本論文では、テストデータに含まれる複数の信号強度を一定時間ごとに取得された時系列データのように扱うことで、パターン照合を実現する。

テストパターンの生成の流れ

- (1) テストデータとテキストの共通の BSSID を検索
テストデータがテキストと類似した系列かどうかを検証する。
- (2) 共通に持つ BSSID の信号強度の大きさを順位付け
順位を求めてそれぞれの BSSID に順位をタグ付ける。

具体的には、あるタイミングで得られた無線 LAN 情報に 5 つの異なる AP があり、その内 4 つの BSSID がテキストにも含まれる場合、共通して所持する 4 つの BSSID に信号強度の大きさから順位を求める。信号強度の大きさが同じであれば、同率順位とする。ただし、テストデータの BSSID がテキストに含まれない場合や 1 つのみの場合は、順位付けが出来ないため認証失敗となる。

3.3 順序保存照合について

順序保存照合はパターンと近似的に似ているもの検索するパターン照合である。しかし、パターンの長さが長ければテキスト中に一致するパターンが存在する可能性が低くなる。そのため、順序が 1 位から数えて何位まで一致しているのかも重要な要素となる。

また、信号強度の大きさは AP からの距離に二乗して反比例している [8]。そのため、数値が大きいほど単純にユーザと AP との距離が近いと推測できる。

4. 実験

提案手法をもとに評価実験を行った。評価実験では、ユーザの識別能力について以下の観点から評価する。

- (1) 本人と他人のパターン照合の差異
- (2) 一致する順序の長さ

無線 LAN 情報の取得には、大学生に協力をしてもらった。被験者が普段使用し、持ち歩くスマートフォンにアプリをインストールし、普段と同じ生活のライフログを取得する。この実験中に取得されたライフログデータを以下に示す。

- 被験者人数：5 人
- ライフログ取得期間：約 2 ヶ月
- 取得データ平均件数：98,754 件/人

アプリは 10 分間隔でデータの取得を行い、取得されたデータは実験用に設置した web サーバーへと送られ、データベースに保存される。アプリが何も取得できなかった場合は SSID, BSSID を null、信号強度を 0dBm として記録し扱う。ユーザの環境によって、取得されるデータ量に差異があるが特に考慮しない。またスマートフォン端末の性能、通信環境によってはデータ取得が出来なかったり、サーバーへデータ送信中にパケットが落ちてしまうケースもあったが、これらは実用上も起こりうるため、データ欠損したままのデータを評価実験に用いた。

4.1 順序保存照合の可用性

テストデータをランダムにサンプリングし、テキストと順序保存照合を行う。その際に、順序同型がテキスト中にいくつ見られるかを評価する。同じ時刻の各被験者のテストデータを用いて、ある被験者のテキストに対して、同じユーザのテストパターンとのパターン照合と異なるユーザのテストパターンとのパターン照合を行う。異なるユーザのテストパターンでの一致件数は平均化する。

実験には試行を 3 回繰り返し、各被験者のパターン照合結果を求めた。

4.2 異なるパターンの長さ

3.3 で述べたように、順序保存照合においてパターンの長さが異なれば、テキストで検出される可能性は低くなっていく。4.2 では 4.1 と同じ実験を行い、その際に照合するパターンの長さを変えることで、一致件数に変化があるかを検証した。

5. 結果と考察

4.1, 4.2 の結果を図 1, 2, 3, 4, 5 に示す。図中の“本人”と“他人”とは、同じユーザのテストパターンとテキストをパターン照合させた結果を本人、異なるユーザどうしてパターン照合を行った結果を他人としている。また、複数の試行回数の平均値をとったものを“平均”、最大となった値または最小となった値を“最大”、“最小”として表している。

5.1 順序保存照合の可用性

図 1 順序保存照合のパターンがテキスト中に一致した本人平均と他人平均を比較すると、概ね本人平均が高い結果を示した。しかし、図 2, 4 ではパターンの長さを“3”とした時、他人最大が突出して高い結果となっている。また、本人平均は高い結果を示すが、同時に本人最小は図 5 以外で他人最大のほうが大きい結果を示している。以上のことから、順序保存照合はテストデータとなるパターンによって大きく認証精度が変化してしまうことが考えられる。

5.2 異なるパターンの長さ

テストパターンの長さが小さいほど、本人と他人ともにパターンの一致件数が増える結果となった。パターンの長さを4にすると、全ての被験者の結果で本人平均が他人最大を上回る結果となっており、適切なパターンの長さの設定が必要であることが分かる。

5.3 考察

順序保存照合は、時系列データにおける近似したパターンを検索する手法である。

本論文中で、時系列データとして扱った信号強度はユーザの行動に合わせて数値が変化だけでなく、ユーザが停留している状況であってもスマートフォンの動作環境や周囲の環境等によってノイズを多く含んでいると考えられる。また、複数のAPごとの信号強度には数値的な差があまりなく、ノイズにより順位もその都度変わってしまった。

先行研究では、信号強度を含む無線LAN情報を匿名化などせずに使用していたが、無線LAN情報を含むライフログはユーザの生活を表すものであり、プライバシーを保護する必要がある。信号強度の数値を曖昧化する手法として、本論文で用いた順序保存照合は有効であると考えられる。

今後の課題には、よりユーザーの特徴を抽出した上での順序保存照合の検討が挙げられる。

6. おわりに

無線LAN情報を用いた認証において、ノイズに強い順序保存照合を適用し、新たに活用可能かを検討した。活用可能性は見込まれるものの、十分な個人識別精度結果は得られておらず、試行回数を増やし精度の改善が今後の課題である。

参考文献

- [1] 株式会社矢野経済研究所, “ライフログ市場に関する調査結果 2011”, ライフログ活用ビジネス市場の現状と展望, 2011年11月発刊.
- [2] NTT docomo, “サービスエリア”, https://www.nttdocomo.co.jp/service/data/docomo_wifi/area/index.html.
- [3] 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 Workshop on Information and Communication Security)
- [4] 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.
- [5] 平岩啓, “無線LAN情報を用いた個人識別”, 金沢大学, 平成26年度, 学士学位論文, 2015年2月.
- [6] 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, 2016年.
- [7] 神崎優子, 清水明宏, “ライフログサービスに適した位置情報分析手法の研究”, 高知工科大学, 平成21年度, フロンティアプロジェクト, 学士学位論文, 2010.
- [8] 安達三郎, “電磁波工学”, コロナ社, pp.167-168, 2004年10月初版第17刷発行

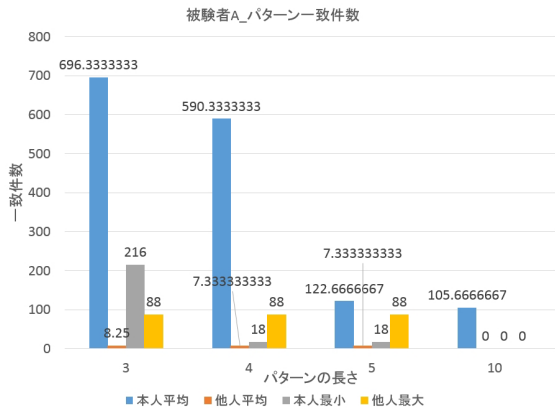


図 1 パターンの一致件数:被験者 A
Fig. 1 number of matched pattern : A

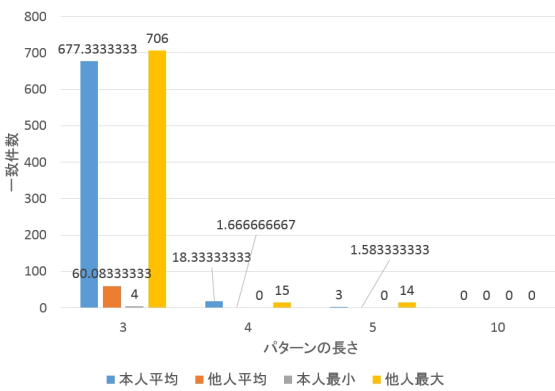


図 2 パターンの一致件数:被験者 B
Fig. 2 number of matched pattern : B

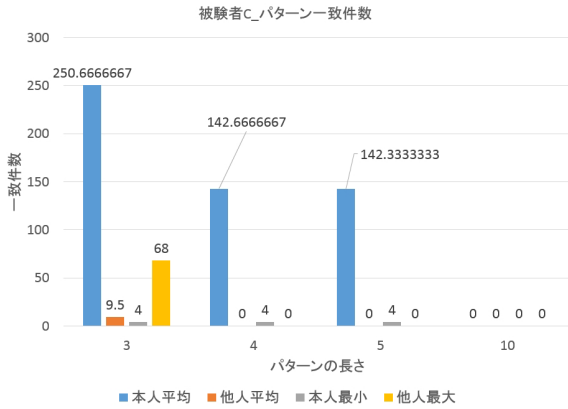


図 3 パターンの一致件数:被験者 C
Fig. 3 number of matched pattern : C

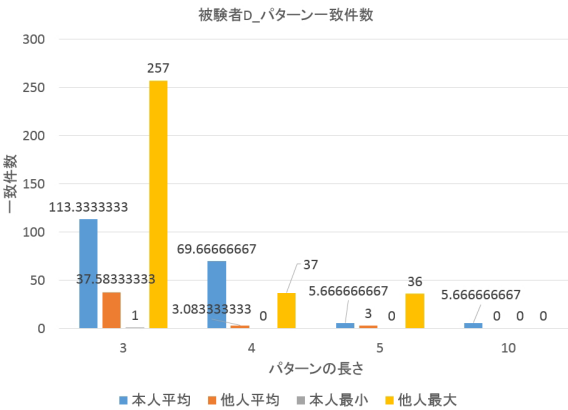


図 4 パターンの一致件数:被験者 D
Fig. 4 number of matched pattern : D

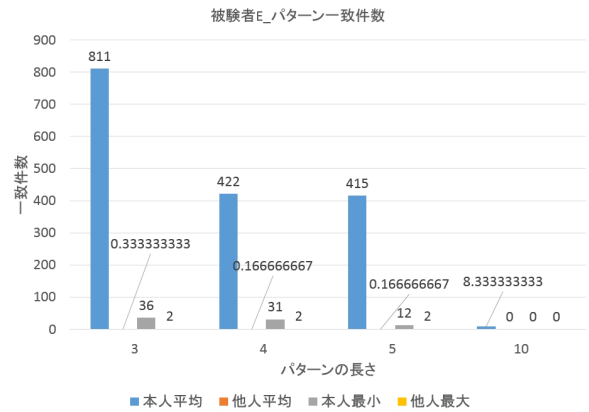


図 5 パターンの一致件数:被験者 E
Fig. 5 number of matched pattern : E