Computer Security Symposium 2017
23 - 25 October 2017

移動履歴からの個人特定とそのリスクについて

疋田 敏朗[1]　山口 利恵[1]

概要：近年，様々のモノがインターネットにつながり，自動的に多種多様なデータが取得される IoT 時代となっている．取得される膨大な量のデータはビッグデータと呼ばれ，特に人々の行動を記録したライフログが有名である．これらのセンシングデータを解析することで，人々のライフスタイルを知ることが可能となる．我々は生活行動のセンシングデータを活用した新しい個人認証技術である，「ライフスタイル認証」を提案してきた．本論文では MITHRA プロジェクトによる大規模実証実験のセンシングデータの中から移動履歴に着目して，その解析結果を報告する．
個人情報保護法では匿名加工情報という個人を特定し得ないレベルに加工された履歴を対象に法的条件をクリアすることで第三者提供を許可している．しかしながら，現状の匿名加工情報の条件では個人特定のリスクが高いことが従来から指摘されていた．
そこで我々は独自に取得した移動履歴をもとに，匿名加工を模擬した移動履歴を作成し，その移動履歴を対象に，他のデータとの照合の実験を行った．今回の我々が取得した携帯アプリによる全国のユーザから収集した履歴データでは 99%以上が一意であり，従来知られていた SNS の履歴を用いる方法の他にも，公開情報の組み合わせで個人の履歴が特定される可能性があることが判明した．

Risk of De-Anonymization of MITHRA Trajectories

Toshiro HIKITA[1]　Rie Shigetomi Yamaguchi[1]

Keywords: CSS2017, PWS, Trajetoriy, De-Anonymization

1. はじめに

近年，GPS を始めとする測位デバイスが携帯機器の機能の一部として搭載されるようになり，さらに携帯網の発展により携帯機器が測位した位置情報をセンター上のサーバに送信し，蓄積することが現実的になっている．さらにスマートフォンとそのアプリケーションや各種 IoT デバイスの普及によって，一般のユーザから大量に位置情報を収集・蓄積することができるようになっており，蓄積した位置情報を活用することで従来は困難であった新たなサービスが次々と生まれるようになっている．

また鉄道乗車券の IC カードが進展し，これらの乗車券の利用履歴を用いることで IoT デバイスの移動履歴と同様に利用者の移動履歴を得ることも難しくなくなった．

これらの各種 IoT デバイスや IC カードから得られる移動履歴情報を活用すれば移動履歴の推定ならびに予測を高精度に行うことができると考えられる．

しかしながら，IoT デバイスや IC カードから発生する移動履歴情報はその内容に位置情報やその移動履歴を含むことから，個人を特定しうる情報やセンシティブな情報が含まれることからそのまま利活用することはできない．

すでに，鉄道乗車に関する IC カードの移動履歴が事業者間で提供されたり [1]，個人の移動履歴に関するプライバシ保護に関する懸念が示されるなどの事例も発生している．それらの懸念に対応するためには履歴を匿名化することによって対応を行うことが考えられる．すでに携帯事業者では位置情報の履歴を匿名化してプライバシ保護を行いつつ統計を行う [2] という事業が始まっている．

交通履歴データに関しても同様に移動履歴はプライバシを保護し，個人を特定しないまま移動履歴を利用するため

[1] 東京大学情報理工学系研究科
The University of Tokyo

の変換手法が望まれている．我々は都市開発においては最も必要と考えられる移動者の出発地と目的地の履歴に着目して匿名化手法を検討 [3] した．次に必要になるのは鉄道 IC カードのような移動履歴の匿名化である．

IC カードの利用履歴に関しては数学モデルによるプライバシ検討は行われていたが，大規模な実データを用いた乗降履歴の分布と個人特定性に関する議論は従来行われてこなかった．

本論文では，まず 2 章でまず位置情報と移動履歴に関する匿名化に関して紹介を行う、その上で従来の移動履歴情報の匿名化に関する研究はデータ提供者側のプライバシ保護に着目しており，移動履歴に対してノイズの形でダミーの移動履歴を加えるなどの加工を施すなど，データ利用の観点が存在しないという問題があることを示す．その上で改正個人上保護法における匿名加工のポイントに付いて議論を行う．

我々はセンシングデータを個人認証に利用した，「ライフスタイル認証」を提案してきた．今回はライフスタイル認証の説明を行うとともに，その大規模実証実験である MITHRA プロジェクトの説明を行い，更にその中でも移動履歴に着目したデータ解析結果を報告する．

2. 履歴データの匿名化に関する従来の研究

まず一般的な履歴データの匿名化について説明を行う．

個人を直接的かつ一意的に識別する属性，たとえば氏名[*1]，個人番号[*2] などを示し，これを **個体識別属性** と呼ぶ

個人を一意に識別できないとしても複数の属性を組み合わせると個人を一意的に識別できるものもある．たとえば性別，生年月日，住所などが該当する．これらの属性を**疑似識別属性** (Quasi Identifier，以下 **QID**) と呼ぶ．

あるデータ T から個人が特定できないようなデータ T' を生成する変換作業を匿名化と呼ぶ．匿名化の手法としては k– 匿名化 [4] が有名である．k– 匿名化は概念で同一の擬似識別属性に対して，最低でも $n \geq k$ のデータが存在するように，擬似識別属性を曖昧化する．例えば氏名情報のみを削除し，会員番号のみを利用する方法は一般的には仮名化と呼ばれる．仮名化を行っても個体識別属性が残っていると一意に識別できるため，仮名化は厳密な意味での匿名化ではない [5]．また，k– 匿名化の情報では，匿名化として不十分として，データの種別を定量的に計る手法 l-diversity [6] や データの全体の割合傾向を計る手法 t-clossness といった手法も提案されている [7]．

次に匿名化の位置情報への拡張について述べる．位置情報について k– 匿名化を行った例 [8] は 2003 年に Gruteser らによって報告されている．この例では地点をグリッドごとに区切り，それぞれの地点情報をもとに k– 匿名化が行われている．Gkoulalas-Divanis らによるまとめ [9] によれば，k– 匿名化の手法は一般的に今あるデータを中心とした区切り方と地形情報を活用したグリッドベースの区切り方の 2 種類に分けることができると主張している．

k– 匿名化の他の匿名化手法としてはノイズを混入するという手法が挙げられる．文献 [10] [11] では実際の位置情報の他に複数のダミーの位置情報を挿入させることでデータ自体の匿名性を担保する手法について記述されている．またダミーデータの混入手法についてはより高度な手法が提案されている，Niu ら提案 [12] によればダミーデータの配置場所を統計的に検討することで，ダミーユーザの現実的な配置が可能になり，より強固な配置が可能になるとされる．

移動履歴に関してもダミーデータを加えて匿名化するという手法が提案 [13] されている．この手法はランダムにダミーデータを加えた移動履歴情報を生成することで，リアルユーザのデータを秘匿化する．しかしながらダミーを利用する方法では受領した位置情報にダミー情報がかなりの確率で紛れ込むため位置情報の利用者側から見るとデータが使いにくいという問題が発生する．例えば実際の情報に 4 倍のダミーデータを混入した場合，位置情報を的中させることが出来る確率を 20%近くに低下させることができるが，利用者から見ると 1/5 でしか正確なデータが存在しないということになる．これは特にビッグデータ処理を前提とした場合にデータ自体の信頼性がなくなること言うことを意味しているため，データの利用目的によってはこの手法は使えない．

また移動履歴をグリッド化して k– 匿名化する方法はいくつか提案されている山口 [14] の手法では単一のグリッドで k– 匿名化を実施するという手法が提案されており，著者ら [3] は可変グリッドを利用した単体移動履歴の匿名化を提案している．

一方で乗車のような履歴の匿名化については菊池ら [15] が数学的モデルにより，鉄道駅の乗降客数データの分布から類推した移動履歴の匿名性に関する検討を行っている．

しかしながら大規模な実データを用いた乗降履歴の分布と個人特定性に関する議論は従来行われてこなかった．

3. 個人特定性と移動履歴の匿名化要件

本論文の目的は IoT デバイスやスマートフォンの移動履歴を想定した実際の移動履歴を前提にどの程度の履歴データであれば個人特定性が存在するかを検討することである．

本章では上記目的を達成するために必要な手法について我が国の個人情報保護法に照らし合わせて，個人特定性を失わせることについて検討し，その上で移動履歴における個人特定性について検討する．まず評価データについて説

[*1] 厳密には氏名だけでは同姓同名の個人が複数存在する可能性があるが，社会通念では個体識別属性とみなされている

[*2] 各個人に一意に割り当てられている番号，例えば日本でいえばマイナンバー，米国でいえばソーシャルセキュリティーナンバー．

明を行う．その上で評価実験の結果について説明をする．

3.1 個人情報保護法と個人の特定

2017 年 5 月現在における我が国の個人情報保護法においては，個人情報とは個人が特定できるような情報のほかに，『(他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるものを含む。)』という形で他の情報と照合することで個人が特定できる情報もまた個人情報であるとされている．ここで識別とはそれが誰だかわからないが特有の 1 名に分離できるということであり，特定とはそれが固有の 1 名を示すこととされる，

実際に米国においては仮名化されたデータセットの履歴情報と別のサイトの履歴情報との間で順序が一致することを利用して照合を行い，履歴の個人を特定したという例が報告されている**?**.

容易照合性を加味した場合は履歴が一意であるということは照合により個人が特定されるリスクを増大させるといえる．このような履歴の個人特定リスクを考慮した場合には k– 匿名化は有用な方法であると言われている．すなわちデータ T が存在した場合にそのリストの全項目を QID として k– 匿名化を行ったデータ T' については，1 つの情報について少なくとも 2 つ以上の列が該当することから一意に特定ができないことが知られている．すなわちデータ T の全項目を QID としてに k– 匿名化による変換を行うことができればその情報について個人を特定することはできないとすることができる．

3.2 匿名加工情報とその要件

平成 27 年に成立し，平成 29 年から施行されている改正個人情報保護法には匿名加工情報という類型が追加された．改正個人情報保護法によれば匿名加工情報は『第三十六条 個人情報取扱事業者は、匿名加工情報（匿名加工情報データベース等を構成するものに限る。以下同じ。）を作成するときは、特定の個人を識別すること及びその作成に用いる個人情報を復元することができないようにするために必要なものとして個人情報保護委員会規則で定める基準に従い、当該個人情報を加工しなければならない。』とされている．

その『個人情報保護委員会規則で定める基準』の考え方について個人情報保護委員会の事務局レポート [16] で示されているおり，「匿名加工情報は、個人情報から作成されるものであり、特定の個人を識別することができず、かつ、元となる個人情報を復元することができない、個人に関する情報である。」とされている．また匿名加工を行うにあたっては「作成の元となる個人情報に含まれている全ての個人識別符号を削除又は他の情報に置き換え，そして特定の個人を識別することとなる記述を削除」する必要があるとされる**?**.

3.3 移動履歴の匿名化のために必要な要件

昨今，Foursquare や Facebook などの SNS への投稿は位置情報を付加することが可能であったり，地点情報を付加してチェックインすることができるため，別の手段で収集された履歴情報や位置情報とデータ処理位行う履歴情報を照合することで個人を特定することも難しくはなくなりつつある．そのため個々の履歴情報に関して，他の情報を用いた場合でも個人を一意特定しうる状態ではないことが必要ということになる．

本研究では上記の事情を鑑みた上で移動履歴の長さが個人特定性にどのような影響をあたえるのかを検討することとする．移動履歴から個人が特定されるケースを列挙すると以下のようなケースが考えられる．

(1) **その情報自体が個人を特定できる情報を含む場合**
(2) **位置情報自体が自宅などを指し示す場合**
(3) **位置情報と時刻の組み合わせにより個人が特定される場合**
(4) **履歴が固有のために個人が特定される場合**

今回はこのうち 4 について検討をすることとする．これは組み合わせ情報が固有であるために個人が特定できることである．履歴情報は同じ ID の履歴をリンクさせることで個人が特定可能であると指摘をされている．

そこで著者らは実際の移動履歴を元にどの程度の履歴長であれば，個人が特定をされるのかについて実際のデータとして，東京大学空間情報科学研究センターの「人の流れプロジェクト」[17] の「2008 年東京都市圏　人の流れデータセット（空間配分版)」(以下，東京地区) と「【空間配分版】平成 12 年京阪神都市圏　人の流れデータセット」(以下，大阪地区) データを利用して検証を行った．

しかしながら，人の流れプロジェクトのデータは大本のデータをパーソントリップ調査のデータに依存しているため，データの記録内容が 1 日分でありデータの習慣に関して判断できない，調査対象が偏るという課題があった．

そこで今回, 著者らは多要素認証の研究として「MITHRA プロジェクト」**?**により，他要素認証技術のためのセンシングデータを独自に収集することとし，そのセンシングデータの中に移動履歴を含めてデータ取得を行った．

4. MITHRA プロジェクトデータの解析

本章では MITHRA プロジェクトデータについて，取得方法と概要について説明を行った上で, MITHRA プロジェクトデータの中で移動履歴データについて解析した結果を報告する

4.1 MITHRA プロジェクトデータ

MITHRA プロジェクトは，東京大学情報理工学系研究科ソーシャル ICT 研究センター次世代個人認証技術講座が実施主体となり，各種センシングデータをもとにした他

表 1 Quadtree による地域分割

222	223	232	233	322	323	332	333
220	221	230	231	320	321	330	331
202	203	212	213	302	303	312	313
200	201	210	211	300	301	310	311
022	023	032	033	122	123	132	133
020	021	030	031	120	121	130	131
002	003	012	013	102	103	112	113
000	001	010	011	100	101	110	111

表 2 MITHRA プロジェクト移動履歴データ (データはダミーデータ)

ID	日時	位置情報	種別	精度
12345	201701281605	313200312310	network	41.905
12345	201701281631	313200312310	network	41.905
12345	201701451731	313200312301	network	41.905
12345	201701461805	313200312301	network	30.0
12345	201701471835	313200312301	network	30.0
12345	201701471901	313200312301	network	50.0
12345	201702151936	313200312312	network	20.687
12345	201702172001	313200312312	network	20.687
12345	201702202035	313200312312	network	20.687

要素認証技術を実現するための研究を行っているプロジェクトである．

本研究を行うために，スマートフォン (iOS,Android) 向けのアプリケーションを作成し，実験協力者の同意の元日常生活での WiFi,Bluetooth, 位置情報のデータ収集を行った．調査内容詳細については鈴木らの発表?に詳しい．

本実験では 2 月から 4 月までの 3 ヶ月分，ユーザ総数 57046 名 (MITHRA アプリに関しては 16027 名) の参加者を得た．移動履歴の一時解析はこのデータの中から，2 月と 3 月分の 2 ヶ月分のデータの一部 6607 名のデータを利用することする．

表 2 に，MITHRA プロジェクトの位置履歴データの例を引用する．このデータ定義については実際のものだが，データ自体に関しては定義に合わせて著者が作成したダミーデータとなっている．

位置履歴情報に関しては数十 m の精度で取得することが可能であるが今回はより大規模な領域での移動を把握することを目的に，Quadtree を用いて 12 桁に変換したデータを利用することとした．なお大本のデータに関しては緯度経度で記録されているのは言うまでもない．

Quadtree による区域分割はそれぞれのビットごとに 0,1,2,3 の 4 値でエンコードを行う．地域分割の例を表 1 に示す．南西:0，南東:1，北西 2:，北東 3 の順で分割を行っている．

Quadtree は地球上の緯度経度を元にした変換であるため東京付近であれば経度方向については $32811/2^{12} = 8.1km$ 緯度方向に関しては $40000/2^{12} = 9.7km$ となり，$8.1km$ x $9.7km$ の矩形となる．都心部においては非常に大きい値であるが，地方部においては十分に小さい．

4.2 MITHRA プロジェクトデータの取得状況

今回の位置履歴情報に関するデータについて整理を行う．

まずは取得したデータの実験日と有効なデータ数について検討を行う．図 1 に実験日と送信データ数をグラフ化している．プロジェクトへの参加者はインターネット上で募集を行ったほか，東京都内並びに大阪府の商業施設でのデモを通して実験参加者を募っている．そのため実験開始時には 1100 名弱であったデータ送信者数は 3 月末の時点では 3800 名超となっている．

このうち精度が 60m を上回る位置情報データのみを抽出したものを高精度データと呼ぶことにする．高精度データは 3500 名超であり，全データと比較すると 300 名程度の差がある．なお，今回解析に使ったデータは 6607 名のデータであるが，有効な位置情報を送信できなかったものや，データ送信を行っていないものがあり，日単位では 6 割程度の稼働率となっている．

次に時間帯ごとの送信数を図 2 に示す．iOS に関しては移動が発生しない時間帯にはデータを送信しないという特性があるため深夜〜未明にかけて送信量が削減されるという事象が見受けられる．Android に関してはバックグラウンドで常時送信を行っている．サーバ側で把握しているデータとしては 3 月末の時点で Android が 1500 超，iOS

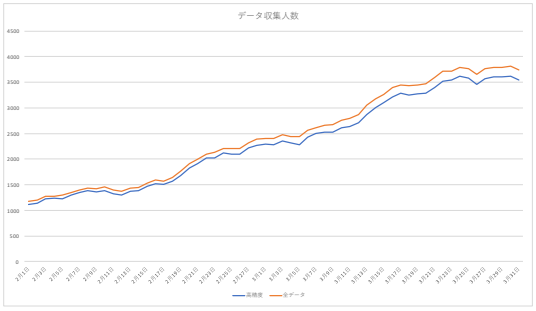

図 1　実験日と送信データ数

が 5000 超となっている

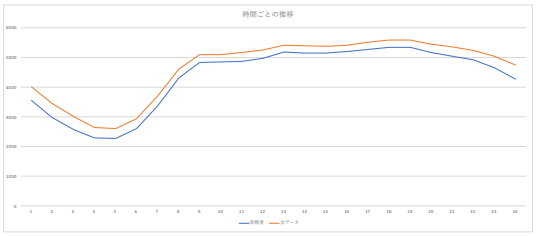

図 2　時間ごとの推移

5000 名超の月単位での移動データは十万人単位の人の流れデータよりは少ないが，充分なデータ量となっているものと考えられる．なお，4 月分まで含めた全参加者数は 16000 名超であり，今後こちらの解析を進めることでより特性が明らかになるものと考えられる．

4.3　ユーザの地理的分布

次にユーザ位置の分布について確認をする．

ここでは一時間ごとのユーザ位置の頻度を取って記録した．今回注意するべきは iOS では移動しない場合の位置情報に関してはデータがないということである．

図 3 にグリッドごとのユーザ数を示す．このグラフは両対数グラフとなっていることに注意が必要である．X 軸は 8kmx9km のグリッドごとに存在するユーザの数である．一時間ごとの集計なので 1 日滞在すれば頻度は 24 であり，1 週間であれば 168 である．

最頻値は 1 であり，なんと 1123 エリアもあった，ユーザー数が最も多かったエリアは 3.6 万，次が 2.0 万であり，東京都心部の 4 グリッドがすべて 1.3 万超となっている．

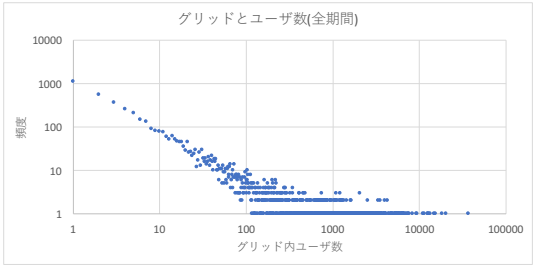

図 3　グリッドごとのユーザー数

滞在頻度が高いグリッドを 3 に示す．東京の山手線付近のグリッドが上位を占め，次に大阪の梅田と新大阪近辺の頻度が高くなっている．その後は札幌の繁華街が上位を占めるという形になっている．

今回の実験に関して商業施設でのデモが東京，大阪に集中したこと，他アプリケーションからの流入が人口が多い地点に集中したことなどが考えられる．

表 3　滞在頻度が高いグリッド

エリア	人数	備考
313200312310	36579	本郷・東京駅
313200312301	20260	新宿
313200312132	18246	銀座・豊洲
313200312312	15483	上野・荒川区
313200020303	14746	大阪梅田
313200312123	13285	東京渋谷
313200312303	12830	池袋
313200020321	12783	新大阪
313200312121	12560	東京目黒区
313200312103	11484	東京大田区
313223021022	10959	札幌すすきの

5.　実移動データを用いた移動履歴データと個人特定性の検討

つぎに 1 日単位の移動履歴について 1 時間ごとの移動位置を記録し，履歴が重複するユーザが存在するかどうかを確認する．今回は 3/21 のデータを利用した．当日のアクティブユーザ数は 3394 である．

まず実際のトリップ情報から移動履歴を抜き出す手法を説明し，その上で移動履歴データの分布について検討を行う．さらに移動履歴と個人特定性の関係について考察を行う．

5.1　実験用の移動履歴データの生成

本節では移動履歴データの生成方法を示す．

移動データから一時間ごとの位置を抜き出し，その地点を記録する．一時間前と同じ地点であれば履歴は伸ばさない．

抜き出したデータを各ユーザごとに整理を行う．

ここで駅数を履歴に関与する駅の数，履歴超を移動区間の数と定義する．たとえば [本郷三丁目],[高尾山口],[渋谷] の 3 駅が履歴にある場合は移動履歴における駅数は 3 であり，履歴長は 2 となる．今後は履歴長を中心に特定性との関係を見ていく．

このように整理した各ユーザーの 1 日分のデータを元にその移動履歴が他のユーザの 1 日の移動履歴と重複しているかどうか？について検討する．

すなわち履歴において個人特定性があるとは，その履歴が全体の中で単独で存在することであり，同じ履歴が複数存在すれば個人特定性はないこととなる．

上記の目的を達成するために移動履歴のうち，移動履歴が一意であるものを抽出する．ここでは一日分のデータを蓄積し，時間については考慮に入れないこととした．また，履歴は履歴自体でのマッチングとし，部分的な履歴が一致するパターン，すなわち [A] → [B] と [A] → [B] → [C] についてはそれぞれを一意とし，重複していないこととする．

今回は移動履歴を表??のように管理する．この例の場合，総数は 5 であり，パターン数は [東京] → [新宿]，[新宿] → [渋谷]，[代々木] → [渋谷] の 3，一意な履歴は [代々木] → [渋谷] の 1 となる．この一意の履歴が個人特定性のある履歴ということになる．

また個人特定率を [一意数/総数] と定義し，この例の場合は $1/5 = 20\%$ となる．

[代々木] → [渋谷] は区間としては [新宿] → [渋谷]　に内包されるが，今回はこのような区間の内包は計上せず別件として数えることとした．

検討結果を表 4 に示す．調査の結果重複する履歴は一日全体で僅かに 3 履歴ということになった．

表 4　移動履歴数と履歴長

履歴長	履歴数	一意履歴数	個人特定率
0	1331	1326	99.62
1	472	472	100
2	676	673	99.6
3	384	384	100
4	215	215	100
5	116	116	100
6	83	83	100
7	47	47	100
8	27	27	100
9	21	21	100
10	11	11	100
11	4	4	100
12	6	6	100
20	1	1	100

6. 移動履歴データの匿名加工と個人特定性について

本章では，生成した移動履歴データから履歴の一意性と個人特定性に関して議論を行う．

6.1 移動履歴の匿名加工

移動履歴の匿名加工に関しては，各省庁が所管する匿名加工に関するガイドラインによるところが大きいが，ガイドラインにおいて匿名加工まで具体的にカバーしている例が少ないため今回は事務局レポートの例を参考にする．

実際には事務局レポートは単一の履歴を残しつつ，出発地と目的地前後の詳細な履歴を対象に匿名加工を行うことを意図しているものと考えられる．しかしながら，これでは単一の履歴，すなわち識別可能であるが特定はされない履歴が大量に発生する．

法は再識別に関してを禁止することで実効性をもたせることを意図していると思われるが，匿名加工データになると，再度加工して匿名加工データを生成することが可能であるので，意図せずに特定されるリスクが有るものと考えるのが妥当と思われる．

ここでは MITHRA の移動履歴データを匿名加工情報のガイドライン同等の加工を行い (加工したデータは匿名加工情報ではなく，個人情報である)，その加工した匿名加工情報と同等のデータに対する再識別ができるかどうかを検討することとした．

6.2 移動履歴データの匿名加工と個人特定性について

従来の研究では十万人単位のデータを一日単位に圧縮したとしても過半数の履歴は一意であること．また 3 区間以上の履歴に関しては 96%以上の履歴が一意であるということを示していた．

一方で人の流れプロジェクトの元データであるパーソントリップ調査は調査用紙を配布し回収するという調査方法に起因する調査対象者の偏りもあり，今回はユーザ数が 3000 人程度と従来に比べて遥かに少なく，そのうえ対象地域が東京地区ではなく全国に散っていることから，一意性が高まるであろうことは容易に想定できた．

しかしながら，全移動履歴のうち 3 つを除く全ての履歴が 8km x 9km の領域においても一致しなかったという今回の実験結果に関しては，移動履歴の利活用や匿名加工に関して，注視をすべき状況であると考えられる．

購買や移動に関してはユーザの志向を理解するためにもある程度の履歴長が履歴長が必要であると言われている．しかしながら本件検討結果では 1 日単位ですらほぼすべての履歴は単独であることを示している．複数日時で履歴長が伸びた場合は今回の結果以上に個人が特定されるリスクがあるといえる．

7. 匿名化した移動履歴に対しての照合攻撃

本章では匿名加工を行った移動履歴に対して照合を行う攻撃について検討を行う．

匿名加工に対する攻撃としては，照合方法として LBS(位置情報を利用したサービス) の情報を用いるかどうか．もう一つは一般公開されている情報か SNS のような特定少数に開示されている情報を用いるかの 2 つの考え方がある．

考え方を図に示す．

以下に攻撃方法について検討を行う．

7.1 LBS を用いて，特定少数に開示されている情報を用いる場合

ユーザ A に関しては，LBS を用いた SNS データを用いて攻撃を行うような場合である．

具体例としては Swarm や Facebook，Twitter のデータを用いて攻撃を行うことが考えられる．例えば，Swarm は特定の場所でのチェックインを履歴として表示し，攻撃対象ユーザと友人関係であれば，攻撃対象ユーザの履歴を表示することが可能である．

そのため，攻撃対象ユーザの履歴の中で，大きく移動を行った地点を抽出して，履歴と照合することで，識別された個人を特定することが可能であった．

具体的にはユーザ A がチェックインを行っていた名古屋駅と東京大学，足柄 SA の履歴を組み合わせることで個人の特定が可能であった．

このように LBS のチェックイン履歴は容易に攻撃が可能であり，位置情報を広範囲にぼかしたとしても，履歴長が長く多様性が少ない状況では用意に発見することができるという特徴がある．

7.2 LBS を用いて，不特定多数に開示されている情報を用いる場合

次に検討を行うのは，LBS を用いており，なおかつ不特定多数に開示されている情報を用いる場合である．例えば Facebook を利用して投稿を公開設定でチェックインしている場合などが該当する．

また，Swarm と連携している FourSquare からであると公開リストからの攻撃が可能であった．

今回は FourSquare のお気に入りリスト (ユーザはお気に入りに入れた記憶はないので自動的に抽出されたかまたは誤操作である) を利用することにした．

ユーザ B のお気に入りリストには名古屋の百貨店，豊橋の駅ビル，名古屋のオフィスビル，東京の勤務先近くの割烹，埼玉の寿司屋が入っており，これを照合すると履歴を一件に絞り込むことができてしまった．

7.3 LBS は用いずに，特定少数に開示している情報を用いる場合

次に LBS は用いずに特定少数に公開されている情報のみを用いる場合の照合方法について検討する．

この場合は SNS などを利用して勤務先や住所 (市区町村までわかれば十分である) を絞り込むことで，照合作業を行うこととする．

この場合の情報については Facebook や LinkedIn などの情報を用いることができる．

今回は Facebook と LinkedIn の情報を用いて，攻撃を行うこととする．

ユーザ C について特定を行うこととする．このユーザは都内の A 区に在住していることが明らかであり，勤務先は B 区にあることもわかっている．またこのユーザ C は東京大学と関係しており，定期的に本郷キャンパスを訪問していることが想定される．

流石にこの情報ではユーザの特定には至らないが，ここから時間と場所を用いてユーザの行動を特定するか，または履歴として特異な地点を探ることで特定を行うことができる．

ユーザ C が 1 月の SCIS に参加していたことが判明しており，1 月の SCIS 開催期間中に那覇近辺にいた履歴のうち，上記の内容を含む履歴を洗い出すことで特定を行うことが可能であることがわかった．

7.4 不特定多数に公開されている情報から照合を行う場合

最も厳しいシナリオとなる．ある種のソーシャルエンジニアリングのスキルが必要となる．この攻撃は特定個人に関する情報が広く公開されている場合に行いやすく，ほとんど公開されていない場合には行いにくい．

また対象ユーザが特殊な場合には行いやすく，対象ユーザが一般人であればあるほど行いにくい．

今回は 2 パターンで攻撃を行うこととした．

ユーザ D は勤務先の会社の人材募集サイトに氏名が掲載されており，該当の会社に所属することが用意に推察される状態となっていた．また該当の会社は住所を web サイトで公開していた．これにより勤務先と推察されるエリアは容易に推定することが可能であった．

次にユーザ D は学会で発表を行っていることから，学会の開催期間中にそのエリアにいたことが推察されている．今回は SCIS のデータを利用することとした．

また同ユーザは某省庁が開催する協議会の委員であることも Google 検索により判明した．Web サイトを検索すると協議会の開催日程は明らかであった．

上記のデータをまとめると，特定日時に協議会開催場所付近にあったデータでかつ，勤務先の会社付近の頻度が高く，なおかつ SCIS に出席したユーザという検索を行うことができるため，ユーザを特定することができた．

次にユーザ E のケースである．ユーザ E は九州の某大学教員であることが判明している．また，同ユーザは東京の大学において非常勤で授業を受け持っていることがシラバスの検索から明らかになっている (同ユーザが非常勤でコマを持っていることは Web からも明らかであった)．更に大学のサイトからは授業の時限と開講日時の検索も可能となっている．

上記内容をもとに，ユーザ E を検索したところ，特定を行うことが可能であった

8. まとめ

本論文ではライフスタイル認証の大規模実証実験である

MITHRA プロジェクトのセンシングログデータをもとに，その移動履歴に着目して，想定される匿名加工を行った上で，実際の照合リスクについて検討を行った．

個人情報保護法では匿名加工情報という個人を特定し得ないレベルに加工された履歴を対象に法的条件をクリアすることで第三者提供を許可している．しかしながら，現状の匿名加工情報の条件では個人特定のリスクが高いことが従来から指摘されていた．

そこで我々は独自に取得した移動履歴をもとに，匿名加工を模擬した移動履歴を作成し，その移動履歴を対象に，他のデータとの照合の実験を行った．今回の我々が取得した携帯アプリによる全国のユーザから収集した履歴データでは 99%以上が一意であることが明らかであり，位置情報を用いた LBS を使って特定を行うケースの他に，特定少数への開示情報を用いる場合と不特定多数に公開されている情報を用いる場合でも特定が可能であることが明らかとなった．

これらのことから，長期間の履歴を用いる場合にはほとんどの履歴が特異な履歴となっており，公開情報や SNS などの特定少数向けの情報と組み合わせることで容易にユーザを特定可能であることがわかった．

今後はユーザを特定しづらい加工方法について検討を行っていく必要がある．

謝辞 本論文の研究は，次世代個人認証技術講座（三菱 UFJ ニコス寄付講座）による．また本実証実験には，株式会社シナジーテック，株式会社小学館，TIS 株式会社，株式会社東京ドーム，凸版印刷株式会社，フェンリル株式会社，株式会社ペイジェント，株式会社 Link-U，含む 13 社の協力を得ている．実証実験協力企業各社ならびに実験に協力いただいた参加者に感謝する次第である．本研究は科研費 (16K12548) の助成を受けたものである，また東京大学空間情報科学研究センターの「人の流れプロジェクト」との共同研究であり，データの整備並びに提供を行っていただいた空間情報科学研究センター各位に感謝する．

参考文献

[1] ”Suica に関するデータの社外への提供についての有識者会議”：“Suica に関するデータの社外への提供について” (2014).

[2] 寺田雅之：“モバイル空間統計：携帯電話ネットワークを活用した人口推計技術とその応用 (ビッグデータ特別セッション)”, pp. 63–66 (2014).

[3] 疋田敏朗, 山口利恵：“階層化符号表現を利用した移動履歴の匿名化手法”, マルチメディア、分散、協調とモバイル (DICOMO2015) シンポジウム 2015 情報処理学会 (2015).

[4] L. Sweeney: “k-anonymity: a model for protecting privacy”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, **10**, 5, pp. 557–570 (2002).

[5] 板倉陽一郎 伊藤孝一 菊池浩明 高木浩光 高橋克巳 中川裕志 疋田敏朗廣田啓一 山口利恵：“「完全な匿名化」幻想を超えて”, 暗号と情報セキュリティシンポジウム 2014 電子情報通信学会 (2014).

[6] A. Machanavajjhala, D. Kifer, J. Gehrke and M. Venkitasubramaniam: “l-diversity: Privacy beyond k-anonymity”, ACM Transactions on Knowledge Discovery from Data (TKDD), **1**, 1, p. 3 (2007).

[7] D. Rebollo-Monedero, J. Forné and J. Domingo-Ferrer: “From t-closeness to PRAM and noise addition via information theory”, Privacy in Statistical DatabasesSpringer, pp. 100–112 (2008).

[8] M. Gruteser and D. Grunwald: “Anonymous usage of location-based services through spatial and temporal cloaking”, Proceedings of the 1st International Conference on Mobile Systems, Applications and Services, MobiSys ’03, New York, NY, USA, ACM, pp. 31–42 (2003).

[9] A. Gkoulalas-Divanis, P. Kalnis and V. S. Verykios: “Providing k-anonymity in location based services”, SIGKDD Explor. Newsl., **12**, 1, pp. 3–10 (2010).

[10] H. Lu, C. S. Jensen and M. L. Yiu: “Pad: privacy-area aware, dummy-based location privacy in mobile services”, Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile AccessACM, pp. 16–23 (2008).

[11] H. Kido, Y. Yanagisawa and T. Satoh: “An anonymous communication technique using dummies for location-based services”, Pervasive Services, 2005. ICPS’05. Proceedings. International Conference onIEEE, pp. 88–97 (2005).

[12] B. Niu, Q. Li, X. Zhu, G. Cao and H. Li: “Achieving k-anonymity in privacy-aware location-based services”, Proc. IEEE INFOCOM (2014).

[13] P. Shankar, V. Ganapathy and L. Iftode: “Privately querying location-based services with sybilquery”, Proceedings of the 11th international conference on Ubiquitous computingACM, pp. 31–40 (2009).

[14] R. S. Yamaguchi, K. Hirota, K. Hamada, K. Takahashi, K. Matsuzaki, J. Sakuma and Y. Shirai: “Applicability of existing anonymization methods to large location history data in urban travel”, Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference onIEEE, pp. 997–1004 (2012).

[15] 菊池浩明, 高橋克巳：“乗降履歴データの安全な匿名化は可能か？”, 暗号と情報セキュリティシンポジウム 2014 電子情報通信学会 (2014).

[16] ”個人情報保護委員会事務局”：“匿名加工情報パーソナルデータの利活用促進と消費者の信頼性確保の両立に向けて” (2017).

[17] Y. Sekimoto, R. Shibasaki, H. Kanasugi, T. Usui and Y. Shimazaki: “Pflow: Reconstructing people flow recycling large-scale social survey data”, IEEE Pervasive Computing, **10**, 4, pp. 0027–35 (2011).