様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ６ 月　 ７ 日現在

機関番号：８２６２６
研究種目：研究活動スタート支援
研究期間：2014～2015
課題番号：２６８８００３０

研究課題名（和文）ビッグデータに向けた匿名生体認証の研究

研究課題名（英文）A Study on Anonymous Biometric Authentication for Privacy Protection in the Era of Big Data

研究代表者

村上　隆夫（Murakami, Takao）

国立研究開発法人産業技術総合研究所・情報技術研究部門・研究員

研究者番号：８０５８７９８１

交付決定額（研究期間全体）：（直接経費）　　2,100,000円

研究成果の概要（和文）：本研究課題では，ビッグデータにおけるプライバシー保護技術として研究されている匿名認証や匿名生体認証だけでは防ぎきれない攻撃として，位置情報のトレースを用いた個人識別攻撃に着眼し，そのリスクの明確化と対策に関する研究で成果を挙げた．具体的には，攻撃者が入手可能な学習用トレースは現実には少量という状況を考慮し，テンソル分解やグループスパース正則化を用いた個人毎の遷移行列の学習法を提案した．また，対策として，攻撃成功確率を一定値以下に抑えつつ，グループ領域サイズを最小化する位置情報の曖昧化法を提案した．実データを用いた評価実験を通して，提案手法の有効性を示した．

研究成果の概要（英文）：In this research project, we focused on de-anonymization of mobility traces as a privacy risk that cannot be mitigated using only anonymous authentication (or anonymous biometric authentication), which is widely studied to protect privacy in the era of Big Data. Specifically, we considered the fact that the number of training traces available to an adversary is very small, and proposed a learning method of personalized transition matrices using tensor factorization and group sparsity regularization. As a defense against this kind of attack, we proposed a location obfuscation method that minimizes the region size while keeping the attack success probability less than a required value. We showed the effectiveness of our proposed methods through experimental evaluation using real datasets.

研究分野：情報セキュリティ

キーワード：位置情報プライバシー　個人識別攻撃　マルコフモデル　テンソル分解　グループスパース正則化　匿名化　曖昧化

１．研究開始当初の背景

近年，インターネットの通信速度・通信量の増加や，スマートフォン，タブレット端末，センサーなどのネット接続端末の増加・多種多様化に伴って，データセンターやクラウドに蓄積された多種多量なデータ（ビッグデータ）を利活用するサービスが，マーケティング・医療・防災などの幅広い分野で注目されている．しかし，その一方でビッグデータに含まれるパーソナルデータ（個人に関する情報）が，プライバシーを侵害する可能性が指摘されている．例えば，複数のサービス間でユーザ認証用の ID が関連付けられ，位置情報，購買情報，医療データなどのプライバシー情報と個人が紐づけられた「個人プロファイル」が作成される恐れがある[C. Terence+, Privacy and Big Data, O'reilly, 2012]．

ユーザ ID を紐付ける名寄せによるプライバシー侵害を防止するための研究は幅広く行われている．例えば，ユーザが自身の ID を秘匿したまま，サービスを受ける権限を持つ（例：会員である，成年である）ことのみを証明する「匿名認証」が盛んに研究されており，生体認証との融合も試みられている[Blanton+, ICICS'09]．このような技術を用いることで，ユーザ ID の紐付けを防ぎつつ，強固なユーザ認証を実現することが可能となる．

しかし，ユーザ ID 以外にも個人を識別し得る情報（準識別子）が存在し，これを用いた個人識別のリスクが残る．例えば，ユーザがスマートフォンなどの小型端末から（チェックインやタグ付けなどによって）公開した位置情報のトレース（移動軌跡）が準識別子となっており，これを用いた個人識別のリスクが残る．但し，ユーザが普段から公開する位置情報は一般には非常に少量なため，攻撃者が入手可能な学習用トレースは極端に不足する恐れがある．この現実的な状況下で個人識別リスクがどの程度あるのかについては，従来では明らかにされていなかった．

２．研究の目的

本研究の目的は，攻撃者が入手可能な学習用トレースが非常に少量であるという状況下での個人識別リスクがどの程度なのかを明確にすることである．この目的の実現のため，上記の現実的な状況下で有効な位置情報プライバシーの攻撃法を提案し，その対策法も併せて検討した．

３．研究の方法

（１）攻撃法

まず，位置情報のトレースを用いた個人識別攻撃として，最も盛んに研究されているマルコフモデルに基づく個人識別攻撃法[Shokri+, S&P'11]に着眼した．この攻撃法では，まず人々が移動可能な領域を計 M 個の領域 x_1，…，x_M に分割し，時間についても予め定められた時間間隔（30 分など）で区切って離散化する．次に，ユーザの行動にマルコフ性を仮定し，攻撃対象とするユーザ u_1，…，u_N 毎に，領域 x_i（$1 \leqq i \leqq M$）から領域 x_j（$1 \leqq j \leqq M$）に遷移する確率で構成される遷移行列を学習する．ユーザ ID が削除されたトレース（匿名化トレース）を入手後，遷移行列を用いて，ユーザ u_1，…，u_N のうちどのユーザのものなのかを識別する（図 1）．

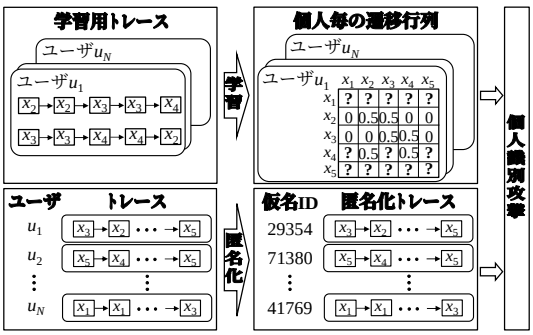

図 1．マルコフモデルに基づく個人識別攻撃

従来では，各ユーザの遷移行列を（最尤推定などを用いて）独立に学習しているが，学習用トレースが少量のとき，遷移行列を正しく学習できない．そこで本研究では，ユーザ毎の集合行列の集合を「3 次元テンソル」と見做し，テンソル分解[Kolda+, SIAM Review, 2009]を用いることで少量の学習データから頑健に遷移行列を推定することを検討した．さらに，位置情報はある種の「グループ構造」（例えば，ユーザが都会エリア内に滞在する確率は一般に高い，Alice は田舎エリアに滞在しやすい，など）を持っていることに着眼し，これを捉えるためにグループスパース正則化（group sparse regularization）を導入することも検討した．

（２）対策法

３（１）の攻撃への対策法として，位置情報の曖昧化[Shokri+, S&P'11]に着眼した．これは，複数の領域を統合してグループ領域を生成し，そのグループ領域を公開することで，具体的な位置情報を不明瞭にする対策である．但し，位置情報の曖昧化によって，データの有用性が大きく損なわれてしまう恐れがあるため，本研究では３（１）の攻撃による攻撃成功確率を一定値以下に抑えつつ，グループ領域サイズを最小化するような最適性を持つ曖昧化法を検討した．

４．研究成果

（１）攻撃法

まず，テンソル分解を用いた遷移行列の学習法を提案した（学会発表②④）．このとき，遷移確率の行き先に対する総和は常に１であり，この制約条件下でのテンソル分解は困難である．そこで，学習用トレースから遷移回数をカウントすることで得られる「遷移回数テンソル」を分解して学習し，その後で遷移回数を（総和が１になるように）遷移確率に

正規化する学習法を提案した（図 2）.

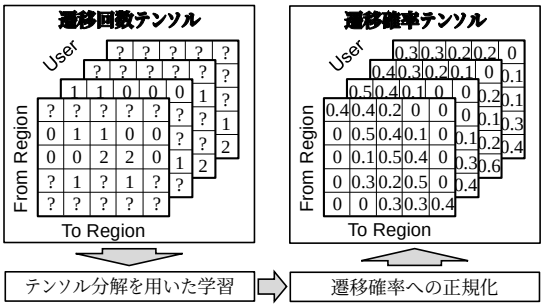

図 2．テンソル分解を用いた遷移行列の学習

次に，位置情報のグループ構造（空間的グループ構造）を捉えるため，グループスパース正則化をテンソル分解に組み込んだ学習法「グループスパーステンソル分解」を提案した（学会発表①）（図 3）．まず，マルコフ・クラスタリング（MCL：Markov Cluster Algorithm）を用いて，領域 x_1，…，x_M を幾つかのグループ（図 3 の例では，R_1，…，R_3）に分割する手法を提案した．次に，グループ化された領域内でテンソル分解のパラメータのスパース性が促進されるような学習法を提案した．具体的には，テンソル分解では通常，パラメータの L_2 ノルムを正規化項とするが，グループスパース性を促進する $L_{1,q}$ ノルム（$1 < q < \infty$）に置き換え，最適化問題をフェンシェルの双対性に基づいて解いた．

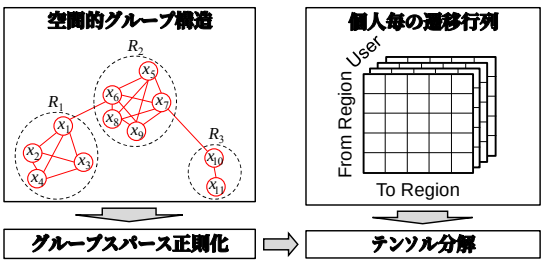

図 3．グループスパーステンソル分解

提案手法の有効性を示すために，Geolife データセット[Y. Zheng+, IEEE Data Engineering Bulletin, 2010]を用いた実験結果を行った．Geolife データセットは，Microsoft Research Asia が 182 名の被験者のトレースを，2007 年から 2012 年にかけて収集したものである．本実験では，各被験者に対して，30 分以上 1 日以下という時間間隔で位置情報を取り出し，10 個の位置情報からなるトレースを 10 個取り出した．但し，そのようなトレースが 10 個も取り出せない被験者が 116 名いたため，残りの 66 名のトレースを実験に用いた（$N = 66$）．各被験者につき，10 トレースのうち 1 つを学習用に，残りの 9 つを評価用に用いた．学習法としては，最尤推定（ML），テンソル分解（TF），グループスパーステンソル分解（GSTF）の 3 つを評価した．各評価用トレースに対して，攻撃者が 66 名の被験者から L （$1 \leqq L \leqq 66$）名の候補者に絞り込んだときの攻撃成功割合（候補者の中に正解が含まれていた割合）を性能として求めた．

実験結果を図 4 に示す．但し，*印がついているものは，9 つの評価用トレースを 1 つに結合したとき（即ち，9 つのトレースに同一の仮名 ID を付与したとき）の性能である．また，点線は候補者をランダムに選んだときの性能である．図 4 から，最尤推定（ML）の性能が最も悪く，また 9 つの評価用トレースを結合することで，逆に性能が劣化していることが分かる．これは，結合されたトレースの中に，学習用トレースにない遷移パターンが多く表れたためである．即ち，最尤推定は学習データ不足問題を抱えている，と言うことができる．これに対して，テンソル分解を用いた提案手法（TF および GSTF）は ML を大幅に上回る性能を示し，9 つのトレースを結合することでさらに性能が上がっている．ここから，提案手法により学習データ不足問題が解決された，と言うことができる．また，グループスパーステンソル分解（GSTF）が最も良い性能を実現しており，位置情報のグループ構造を考慮することで，個人識別の性能がさらに向上することが分かる．以上により，提案手法の有効性が示された．

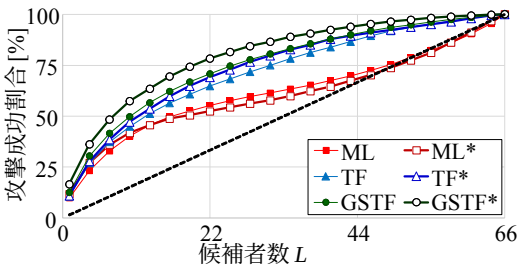

図 4．候補者数 L と攻撃成功割合の関係

尚，実際には学習用トレース内の位置が欠損する状況も考えられるが，この状況下でより頑健に遷移行列を学習する手法も提案し，有効性を示している（学会発表③）．

（2）対策法

4（1）の学習法を用いた攻撃への対策として，グループ領域サイズを最小化する曖昧化法を提案した（学会発表②④）．提案手法は，4（1）の攻撃による攻撃成功確率を要求値以下に抑えつつ，必要なグループ領域サイズを最小化する，という最適化問題を解くものである．最適化問題は，各グループ領域に対して攻撃成功確率を求めることで解ける．評価実験を通して，グループ領域サイズを常に固定する場合と比べて，同じ攻撃成功確率を保ったまま，グループ領域サイズを大幅に削減できることを示した．

尚，この位置情報の曖昧化は，ユーザの端末側で（あるいは位置情報サービス提供者とは別の信頼できる第三者機関[B. Gedik+, IEEE TMC, 2008]において）実行し，曖昧化

されたトレースのみをサービス提供者に渡すことが可能である．匿名認証や匿名生体認証[Blanton+, ICICS'09]も同様に，ユーザの端末側からサービス提供者に対して，ユーザ名を明かすことなく権限を持つことを証明する．従って，これらの対策を併用することで，匿名化と位置情報の曖昧化の両方が施されたトレースのみをサービス提供者に渡すことが可能となり，サービス提供者が保持するトレースの情報漏洩なども考慮した，強固なプライバシー保護が実現できるものと考えている．

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔学会発表〕（計 4 件）

① Takao Murakami, Atsunori Kanemura, and Hideitsu Hino, “Group Sparsity Tensor Factorization for De-anonymization of Mobility Traces,” Proceedings of the 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom'15), pp.621-629, 2015. **[Best Paper Award (278 件の投稿中 2 件が受賞)]**（査読有）

② Takao Murakami and Hajime Watanabe, “Location Prediction Attacks Using Tensor Factorization and Optimal Defenses,” Proceedings of the 1st IEEE International Workshop on Big Data Security and Privacy (BDSP'14), pp.13-21, 2014.（査読有）

③ 村上隆夫,“欠損位置情報の推定を伴うテンソル分解と個人識別攻撃への応用”, コンピュータセキュリティシンポジウム 2015（CSS'15), 2015.

④ 村上隆夫, 渡辺創,“テンソル分解を用いた位置情報プライバシーへの攻撃と対策”,コンピュータセキュリティシンポジウム 2014（CSS'14), 2014.

６．研究組織
(1)研究代表者
村上　隆夫（MURAKAMI, Takao）
産業技術総合研究所・情報技術研究部門・研究員
研究者番号：80587981

(2)研究分担者
なし

(3)連携研究者
なし