移動履歴を保持する際の k-匿名性の考え方

山口 (繁富) 利恵 †　　副田 俊介 ‡

† 独立行政法人 産業技術総合研究所
情報セキュリティ研究センター
〒 101-0021 東京都千代田区外神田 1-18-13
秋葉原ダイビル 1003
rie-shigetomi@aist.go.jp

‡ 独立行政法人 産業技術総合研究所
情報技術研究部門
〒 305-8568　茨城県つくば市梅園 1-1-1
中央第 2
shunsuke.soeda@aist.go.jp

あらまし　ユーザの様々なサービス利用や移動履歴などをもとに，その情報を利用して別のサービスへ展開するためにどのような解析を行うべきかという研究が盛んに行われている．この元となる情報をユーザのヒストリー情報と呼ぶこととする．こういった情報はライフログなどと言われることも多い．こういったヒストリー情報は，ユーザの行動を捕捉することとなるため，ユーザのプライバシに問題があるといわれている．こういった問題を解決するために様々なプライバシ保護のための研究が行われており，k-匿名性という概念が存在する．本研究では，移動履歴情報を元にする情報における k-匿名性の考え方について検討を行った．

A Note for Collectiong User's Location History Information Called Weak-k-anonymity

Rie Shigetomi YAMAGUCHI†　　Shunsuke Soeda‡

†National Institute of Advanced Industrial Science and Technology (AIST)
1-18-13 Sotokanda, Chiyoda-ku, Tokyo 101-0021 Japan
rie-shigetomi@aist.go.jp
‡National Institute of Advanced Industrial Science and Technology(AIST)
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568 Japan
shunsuke.soeda@aist.go.jp

Abstract　There are a lot of appliaction to collect user's location history information. The information collected is usually a combination of some information, plus the location of the user and the time of probing, which could easily breach privacy. Here, we defined the privacy of a vehicle (thus the driver) based on the idea of k-anonymity with ($k < 2$), and see if any user could be identified from others by its path. In this paper, we show that in what condition can privacy be protected.

1　はじめに

ユーザの様々なサービス利用や移動履歴などをもとに，その情報を利用して別のサービスへ展開するためにどのような解析を行うべきかという研究が盛んに行われている．この元となる情報をユーザのヒストリー情報と呼ぶこととする．こういった情報はライフログなどと言われることも多い．こういったヒストリー情報は，ユーザの行動を捕捉することとなるため，ユー

ザのプライバシに問題があるといわれている．こういった問題を解決するために様々なプライバシ保護のための研究が行われており，k-匿名性　という概念が存在する．本論文では，ヒストリー情報の中でも移動履歴について，k-匿名性の概念について整理する．しかし，現状の k-匿名性では，制約がきついため，実際のサービスの展開が難しい．そこで，途中一カ所でも k-匿名性を実現することでもプライバシが保てるのではないかということを考察し，弱 k-匿名性と名付け，検討を行う．

1.1　プローブ情報システム

ここで，プローブカーを例にとり，移動履歴における，k-匿名性 について検討する．プローブ情報システムは，自動車の保持するセンサデータ (プローブデータ) を，インターネット等の汎用的な情報通信基盤を用いて収集して，統計的な処理等を施すことで，交通情報や気象情報，安全運転支援情報等の価値ある情報（プローブ情報）の生成・提供を実現する高度道路交通システム (ITS:Intelligent Transport Systems) の 1 つである [6, 12, 13].

プローブ情報システムは世界中で様々な研究及び開発が行われており，実用化をしている例もある．日本では，2004 年横浜や 2003 年名古屋などにおいて，大規模な実証実験が行われ，プローブ情報としてセンサー情報を収集し，渋滞情報や降雨情報などにおいて実験が行われた [8]．ところが，センサー情報は，情報提供者の位置情報とともに常にサーバへ送られるために，情報提供者のプライバシに問題があるとされている．特に，何らかの他のプライバシ情報との組み合わせとなることで，プライバシのリスクが高くなり，個人を特定しない処理が必要である．登録された個人情報とユーザの位置情報などが含まれたセンサー情報のリンクをはずすことにより特定できないことで実現することも考えられるが，プライバシを保護するためには，これでは十分とはいえない．そのため，センサー情報を詳細に収集するのではなく，サービスを提供する際に必要な分の情報を収集することで解決することができるといわれている．

特に位置情報と時刻情報については，ある地域ある時間などを大きく利用することでも，二次利用の際の情報サービスには影響がない可能性が高い．たとえば，2008 年 11 月 21 日 12H09M21S 北緯 35 度 41 分 45 秒 21 東経 139 度 46 分 39 秒 32 という情報までは必要ないことが多いため，2008 年 11 月 21 日 12H00 北緯 35 度 41 分 40 東経 139 度 46 分 30 とすることにより，同じぐらいの時間同じあたりのエリアに，複数台いることによって，個人の車が特定できないようにし，プライバシに問題がないように処理を行ったとしてもサービスを提供するのに十分な情報を収集したということができる．

1.2　k-匿名性

一方，匿名性というのは，元来，人々の感覚や経験則に基づいた手法によって実現されたきた．このような実現方法がプライバシ保護として機能しているか，つまり目的を果たしているかどうかの検証を行うために，匿名性とは何かと評価する手法を利用し匿名性が確保されているかを明確に示す必要がある．ここで重要となる概念が k-匿名性（もしくは k Anonymity）である [11, 10, 4]．k-匿名性は，与えられた情報から個人を特定しようとした結果，その条件にあてはまる人が最低でも k 人いる状態を指す．つまり，同じ行為や資格を複数の者が同じカテゴリに配属されるということにより，どの人であるかがわからないということである．また，あるサービスを受けた人が一人である場合は，匿名性は保証されない．

この概念に基づいた複数の手法が提案されており，その中にグループ署名や匿名認証なども，この概念の上に立つと整理することができる [3, 9]．匿名認証とは，同じカテゴリへ所属している人がグループ内のうち誰であるかがわからないにもかかわらず，認証を行うことができる技術である．つまり，匿名性とは $k > 1$ を満たすことにより実現された技術であり，このような認証技術には，信頼度の仮定の置き方により複数の方式が提案されている．

これは，サービス提供に十分な情報を取得したとしても，あるエリア，時間帯など，どうい

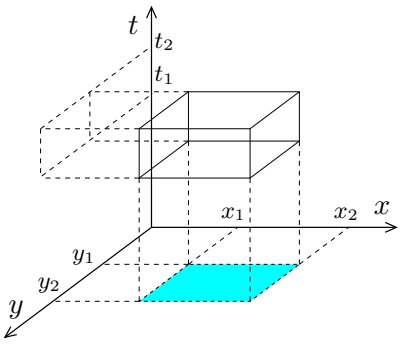

図 1: 3 軸におけるメッシュのモデル

つの種類の行為が複数人いるという仮説にたって研究された技術である．

こういった場合，収集した情報，特に位置情報などを属性情報として「丸め」ることにより匿名性を確保することができる．サービス上において，匿名性を確保するのに必要最低限の「丸め」操作を定義しなければならない．この論文では，各サービスにおいてどの程度の「丸め」が適切かは検討を行わない．

2　モデル

ここでは，情報を「丸め」るとはどういうことであるのかについて，モデルを定義する．特に，時空間を メッシュ状 (Mesh 状) に切り刻むことを提案する．

2.1　定義

ある時間帯およびあるエリアに複数人いるために，エリア (x, y) と時間 t の 3 軸で考える．各軸において図 1 のようにある範囲を定め，メッシュと呼ぶこととする．

同じメッシュの上に複数人が存在したとき，他の人に会うことができたとし，k-匿名性が保たれていると考える．このメッシュが小さいほどセンサー情報としては有効であると考えられるが，その分，他の人には会いにくくなる．

つまり，移動体の情報収集を行う上で情報の粒度をどのように設定するかはプライバシ保護を考えるときの重要な要素の 1 つとなる．時間・

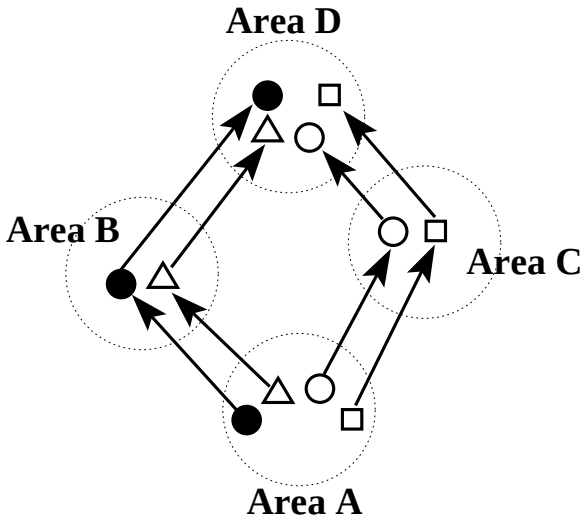

図 2: 一般的な k-匿名性

位置の正確な値が判明すればそれだけ個々人の経路が他の人の経路と区別し易くなり，1 つの車両に関する情報の抽出・分析が行える可能性が高まるためである．

そこで，メッシュを定義し，エリア (x, y) において常に $x_2 - x_1 = y_2 - y_1$ とし，空間メッシュと呼ぶ．また，時間 t については，$t_2 - t_1$ 間を時間メッシュと呼ぶ．

3　移動履歴における k-匿名性の考え方

一般的に，移動履歴における k-匿名性を満たすとは，複数台もしくは複数人が同じ道のりを同時刻に通り，どちらがどう動いたのかわからないため匿名性がある，ということである．

ここで，まったく同じように通ることはできないため，2 章で定めるように情報は「丸め」ることとする．また，●，△，○，及び □が人だとし，移動をしているとする．車などの他の移動体であっても同じである．同一の図のなかで，それぞれ 3 ステップするということで，時間上の情報を丸め，各ステップをステップ 0，1，2 とする．空間上では，Area を A 〜I まであるとして，ある Area に存在することで情報を丸める．

3.1 一般的な k-匿名性

たとえば，$k > 1$ だとすると，2 人以上が図 2 のような移動を行う場合，ステップ 0 において Area A から，ステップ 1 で Area B をとおり，ステップ 2 で Area D にたどり着く人が●と△の二人いるため，k-匿名性を満たしている．また，同様に○と□もステップ 1 において Area C を二人で通っているため，k-匿名性を満たしている．

しかし，図 3 のような場合，図 2 と同様に，Area A から，Area B をとおり，で Area D にたどり着く人が●と△の二人いるため，k-匿名性を満たしている．しかし，○と□は，ステップ 0 及びステップ 1 では，Area A 及び Area B において 2 人いるが，ステップ 3 においては，○は Area B，□は，Area D にいるため，ステップ 2 までの情報のみが k-匿名性を満たしている，ということができる．

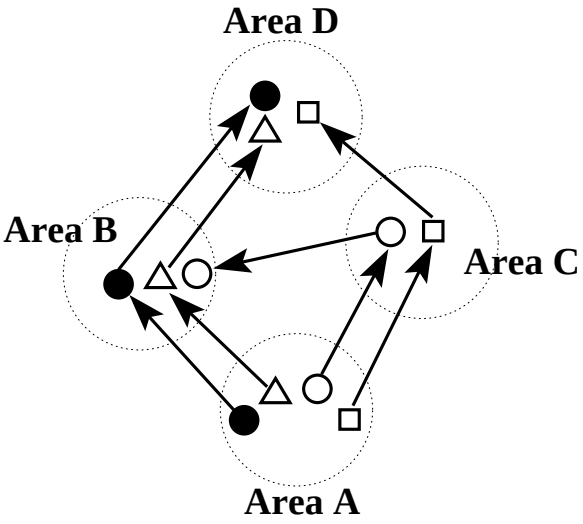

図 3: 状況 2

同様に，図 4 のような場合においても，ステップ 0 での Area F，ステップ 1 での Area G までは，2 人までいるが，ステップ 3 において，○は Area H に移動し，●も Area I に移動することにより，2 人以上いないため，ステップ 2 からステップ 3 の移動については，k-匿名性を満たしていない．

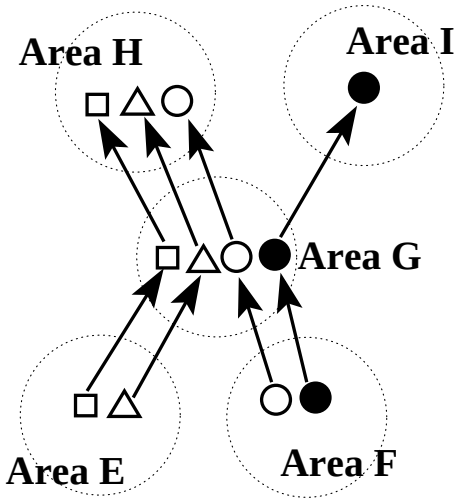

図 4: 状況 3

4 弱 k-匿名性

ここで，我々は，移動履歴における弱 k-匿名性 (Weak-k-anonymity) を提案する．

4.1 移動履歴の特性

移動履歴において，図 2 のような状況だけではなく，図 3 や図 4 ということも存在し，また，そういった情報も何らかの処理をすることにって有用に利用したいという要求がある．

そこで，図 3 や図 4 において，匿名性を保つとは何かについて検討する．

移動履歴において，守らなければならないプライバシとは：

- 移動体が誰であるのかがわからない
- 移動体のスタートとゴールがわからない
- ある程度進んだ後，どの方向へ進んだのか推測ができない

であるといえる．

つまり，n ステップまで（$n > 1$）にどこにいたかがわかるが，$n + 1$ ステップには，どこにいたのかわからない，ということを意味する．この n は有限である．

この n がどのように決めなければならないかは，データの特性に依存しているため，別途の検討とする．また，複数人全員が同一の行動をとっているために，移動の特性がわかってしまうという状況もあり得，l-多様性 (l-Diversity) と呼ばれているが，これも別途の検討とする．

4.2 弱 k-匿名性の検討

$n = 2$ の場合は，2 ステップ前にどこにいたかがわからないということを証明すればよいため，Area A から Area B に二人以上存在すれば，二人の移動経路について推測できないということになる．

ここで，状況 2（図 3）について検討を行う．ステップ 1 から 2 に注目する．○と □が同じ時に Area A から Area C に移動を行っている．また，その後，○は Area B に，□は Area D に移動を行っている．つまり，○と □が Area C において二人存在するため，どちらに行ったかわからない．つまり，2 ステップまでにどこにいたかがわかるが，3 ステップには，どこにいたのか推察することができない，ということを意味する．

また，同様に状況 3 (図 4) においても，Area G において，△，□，○及び ●の 4 人が混在するため，どちらの経路を通ったのか推察することができない．

このように，途中一カ所でも k 人存在することによって経路を推測することができないことを弱 k-匿名性と定義する．

4.3 プローブカーにおける弱 k-匿名性

プローブカーにおいて，弱 k-匿名性を実現するとは，どのようなことであるのかについて考える．

プローブカーのセンサー情報を次のサービスへ二次利用するに当たり，必要となる情報とは，車のスピードやワイパーなどのセンサー情報とともに位置情報および時間が必要となる．この位置情報と時間情報がプライバシと大きな問題となるため，k-匿名性 を満たすようにしなければならない．これは，同じ位置情報と時間情報において，複数人（複数車）存在するような状況にしなければならないのである．

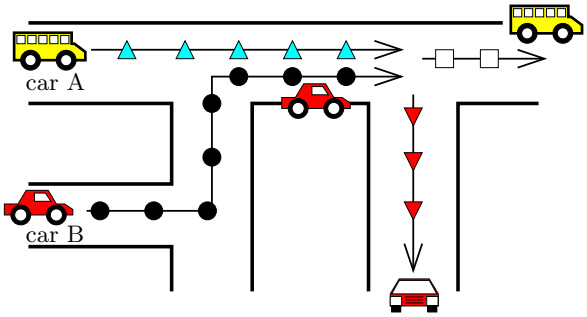

図 5: プローブカーにおける弱 k-匿名性

プローブカーの情報においては，何時何分何秒何々などの点での情報が必要なのではなく，何時何分何というある範囲のある時間帯における，ある程度枠のある情報であれば，有用な情報であろうと考える．そこで，このある範囲のある時間帯に，2 台の車が存在することにより，両者がどちらに進んだのかがわからないようにすることで k-匿名性 を実現するとする．

例として，既に提案された手法の仮定 [14] において，図 5 をもとに再度考える．最初は，Car A と Car B の 2 台の車が 2 カ所の方向から進んできたが，途中で一度同じ方向にすすみ，そののち，2 カ所に離れている．それにより，Car C は，もともとが Car A であるのか Car B であるのかの区別がつかない．

これにより，弱 k-匿名性を満たしているということができる．

5 周辺研究

k-匿名性とは，ある条件に当てはまる人もしくはものが最低でも k 人（台など）いる状態をさす．これにより，同じ行動をした人，同じ資格を持っている複数の人が同じカテゴリに所属されることにより，その中のどの人であるかがわからないという概念を表す．

この概念は，プライバシ保護および匿名化の分野において，経験的に当たり前のように利用

されてきたが，Sweeney によって，k-匿名性 という概念として 整理された．その概念整理として，複数の論文が提案されている [11, 10, 4]．

複数人該当するカテゴリがあるという前提で，様々な論文が提案されており，匿名認証技術 [9] や，ロケーションプライバシについても同様といえる．つまり，k-匿名性が存在するという前提で論文がかかれている [2, 5]．

k-匿名性を従来のデータベース技術を利用して実現させた論文が複数提案されている．ARGUS や INCOGNITO などがよく知られている [7, 1]．こういった技術は一般のデータについての高速化や汎用性を議論した論文であり，ロケーションプライバシの k-匿名性 について検討した論文とはいえない．

6 おわりに

本論文では，移動体における一般的な k-匿名性の考え方だけではなく，より，情報収集をしたとしてもプライバシを保てると考え，弱 k-匿名性を提案した．ここでは,「丸め」方の適切なやり方については，状況や各データの取り方に依存するため，本稿では検討を行わなかった．また，たとえ，複数人いたとしても全ての人が同じ行動をとるというような場合を検討した l-多様性との関係は改めて整理する必要性がある．

参考文献

[1] L. W. A. Hundepool. Argus for protecting microdata and tables. In *Seminar on New Techniques & Technologies for Statistics*, 1998.

[2] A. R. Beresford and F. Stajano. Mix zones: User privacy in location-awere services. In *2nd IEEE Annual Conference on Pervasive Computing and Communications Workshops*. IEEE, Mar. 2004.

[3] J. Camenisch and A. Lysyanskaya. Efficient non-transferable anonymous multi-show credential system with optional anonymity revocation. In *EUROCRYPT2001*, Vol. 1976 of *LNCS*, pp. pages 331–345. Springer, 2001.

[4] A. Gionis, A. Mazza, and T. Tassa. k-anonymization revisited. In *ICDE2008*, 2008.

[5] J. Krumm. A survey of computational location privacy. In *Workshop on UBICOMP Privacy : Tech-nologies, Users and Policy*, Sep. 2007.

[6] K.Uehara, H.Sunahara, and J.Murai. Problems and tentative solutions in internetcar testing with ipv6. In *Proc. of SAINT2003 IPv6 Workshop*, Jan 2003.

[7] K. Lefevre, D. J. Dewitt, and R. Ramakrishnan. Incognito: efficient full-domain k-anonymity. In *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pp. 49–60, New York, NY, USA, 2005. ACM.

[8] M. Sato, M. Izumi, H. Sunahara, K. Uehara, and J. Murai. Threat analysis and protection methods of personal information in vehicle probing system. In *The Third International Conference on Wireless and Mobile Communications(ICWMC)*, 2007 Mar.

[9] R. Shigetomi, J. Furukawa, A. Otsuka, K. Martin, and H. Imai. A provably secure refreahable partially anonymous token and its applications. In *IEICE Transactions on Fundamentals Special Section on Discrete Mathematics and Its Applications*. IEICE, The Institute of Electronics, Information and Communication Engineers, 2006.

[10] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression,. In *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5), pp. 571–588, 2002.

[11] L. Sweeney. k-anonymity: a model for protecting privacy. In *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10 (5), pp. 557–580, 2002.

[12] T.Ernst, K.Uehara, and K.Mitsuya. Network mobility from the internetcar perspective. In *Proc. of AINA2003*, Mar 2003.

[13] K. Uehara, H. Sunahara, and J. Murai. The internetcar network architecture: Connect vehicles to the internet using ipv6. In *ITST2005*, pp. 187–190, 2005 Jan.

[14] 佐藤, 繁富, 上田, 党, 植原, 砂原, 村井. プローブ情報システムのためのプライバシを考慮した匿名認証方式の提案と評価. 特集：新しい時代を切り拓くモバイル通信と高度交通システム，モバイルセキュリティ. IPSJ, 2009.