電気学会論文誌 C（電子・情報・システム部門誌）
IEEJ Transactions on Electronics, Information and Systems
Vol.135 No.6 pp.660–670　DOI: 10.1541/ieejeiss.135.660

論　文

確率的ダミー生成による統計的な位置情報収集のための プライバシ保護手法の提案

非会員　清　　雄一*a)　非会員　大須賀昭彦*

An Algorithm for Privacy-Preserving Location Data Collection by Probabilistic Dummy Generation

Yuichi Sei*a), Non-member, Akihiko Ohsuga*, Non-member

（2014年10月14日受付，2014年12月19日再受付）

Mobile devices, which can sense their locations by GPS or Wi-Fi, have become popular these days, and we can collect and analyze location information of many users to examine traffic flow, conduct marketing analysis, and so on. However, several users hesitate to provide their accurate location information. Therefore, researches which anonymize user's location information on their devices and send the anonymized information to the data collection server have been proposed. These researches can protect user's privacy and let the data collection server to estimate the distribution of users' locations by a statistical way. However, they need many users to help with the data collection. In our proposed method each user sends several dummy locations to the data collection server and the server can estimate the location distribution with high accuracy. By mathematical analysis and simulations, we prove our proposed method can reduce the estimated errors by approximately from 85% to 99%.

キーワード：位置情報，プライバシ，データマイニング
Keywords: Location information, Privacy, Data mining

1. はじめに

ユーザの位置情報をスマートフォン等のモバイル端末から取得し，多数のユーザの位置情報を収集することによって，ユーザの消費行動と行動記録を結びつけたマーケティング分析等を行うことができる[(1)]。しかし，正確な位置情報を提供することに抵抗のあるユーザも多く存在しているため，プライバシの配慮が必須である。

このような問題に対応するため，位置情報を匿名化して，データマイニングに活かす研究が盛んに行われている。しかしながらその多くは，ユーザの情報を収集するサーバが信頼できるという前提を置いており，ユーザが情報収集サーバを信用しない状況には対応できない。近年では，収集した位置情報を他企業にユーザの許可無く販売を行うことや，情報収集サーバへの攻撃によって情報が漏洩する懸念もあることから，ユーザが情報収集サーバを信頼できない状況を考慮する必要性が高まることが考えられる。

そこで，ユーザ自身のモバイル端末において，ユーザの位置情報を匿名化する手法が近年提案されている[(2)~(4)]。これらの手法は，Negative Survey[(5)]と呼ばれる手法を応用したものであり，ユーザの位置情報を収集するサーバが信頼できない状況においても利用することが可能である。

Negative Survey に基づく手法では各ユーザの正確な位置情報を収集しないが，多くのユーザから情報を収集することにより，どの場所にどのくらいのユーザが存在しているかを，ある程度の精度で推測することができる。しかし，情報収集サーバにおいてユーザの位置情報の分布を精度良く推測するためには数多くのユーザから情報を収集する必要がある。

本論文では，各ユーザは真の位置情報とともに確率的にダミーの位置情報を生成してサーバに通知する手法を提案する。提案手法を用いることにより，既存研究と同じプライバシ保護レベルを維持しつつ，より少ないユーザ数から，高精度な推測を行うことができる。本論文で採用するプラ

a) Correspondence to: Yuichi Sei. E-mail: sei@is.uec.ac.jp
* 電気通信大学大学院情報システム学研究科
〒 182-8585 調布市調布ヶ丘 1-5-1
Graduate School of Information Systems, The University of Electro-Communications
1-5-1, Chofugaoka, Chofu, Tokyo 182-8585, Japan

イバシ指標及び有効性指標の下においては，数学的解析及びシミュレーション結果から，既存研究よりも推測精度を85%から99%程度向上させることができることを示す。

本論文の構成を示す。2章では，本論文が想定しているシナリオ，攻撃モデル及びプライバシモデルを定義する。3章において本論文で採用するプライバシ指標及び有効性指標について記述する。4章では既存研究について述べる。5章において本論文が提案する手法を記述し，6章では提案手法と既存手法の比較を数学的解析及びシミュレーションによって実施する。7章において考察を述べ，8章で本論文のまとめを記す。

2. 想定環境

〈2·1〉 想定シナリオ ユーザは匿名化された位置情報と，個人に紐付けられたくない属性情報をあわせてサーバに情報収集サーバに送信する。情報収集サーバは，どの属性を持った人が，どの場所に何人存在しているかについて推測を行う（図1）。

なお，ユーザごとの位置情報の履歴をデータマイニングに利用するシナリオもいくつかの既存研究では想定されているが，本論文では，各ユーザの移動履歴についてはデータマイニングの対象としない。したがって，各ユーザが時間をずらして複数の位置情報をサーバに送信することになっても，同一ユーザからの情報であることをサーバは認識する必要はない。このようなシナリオの一つとして，参加型環境センシング[6]~[8]が挙げられる。参加型環境センシングでは，各ユーザが様々な場所で位置情報等の環境情報をセンシングし，サーバへ通知する。情報を受け取ったサーバは位置情報ごとの統計的な分析を行うが，同一人物からの情報かどうかについては通常考慮していない。

また，前提として，店舗，公園等の防犯カメラ等から，ある時刻・ある位置にいた人物の顔は判明する可能性があると想定する。通常は，その人物がどのようなアプリケーションを使っているか，どのようなキーワードで検索を行っているかは推測することさえできない。しかし，位置情報を提供することにより，個人に紐付けられたくない属性情報を紐付けられてしまう場合がある。

想定されるシナリオ例を示す。

シナリオ1

都市開発や公衆衛生のために，自治体が，どのような場所にどのような属性を持った人物が存在しているのかを知りたいという状況を考える。ユーザは，情報が匿名化されていれば，社会貢献のために，自分の年齢や性別の情報と，位置情報やその他センシングされた情報をサーバに通知しても良いと想定する。これは上述した参加型環境センシングの一例である。

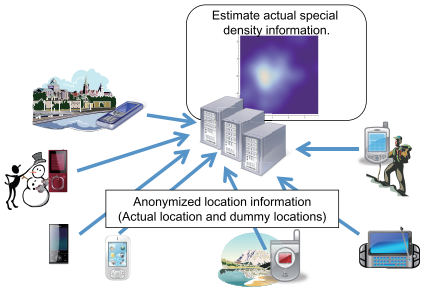


Fig. 1. General model

シナリオ2

あるモバイルアプリケーション開発会社が，開発したアプリケーションがどこで利用されることが多いのかについて知りたいという状況を考える。ユーザは，位置情報の提供がアプリケーションを利用するために必須の情報でないことを認識しているが，匿名化されているのであれば，無料または安価にアプリケーションを利用する見返りとして，サーバに情報を通知しても良いと考えていると想定する。

この場合，「個人に紐付けられたくない属性情報」とは，このモバイルアプリケーションを利用していることそのものである。

シナリオ3

検索エンジンサービスを提供している会社が，どのようなキーワードがどの場所で検索されたかを知りたいという状況を考える。位置情報に関連した検索結果を知りたいユーザは，自分の位置を正確に検索エンジンサービスに提供するが，そうでないユーザは位置情報を提供する必要は無い。しかし，匿名化されているのであれば，キーワード検索時に位置情報を提供しても良いと考えていると想定する。

〈2·2〉 攻撃モデル 情報収集サーバは，semi-honestであることを想定する。本論文においては，情報収集サーバはプロトコルから逸脱したことは行わないが，受信したデータからユーザの位置情報や属性情報を推測しようとするという攻撃モデルである。また，ユーザは匿名化されており，サーバはユーザを識別できないと想定する。したがって，同じユーザが複数回情報をサーバに通知したとしても，それが同一ユーザからの情報であるのかどうかを攻撃者は判定できないものとする。

当然のことながら，もしユーザの移動履歴をデータマイニングすることを目的とする場合は，各ユーザにIDを付与して情報を収集する必要があるため，攻撃者も，同一ユーザからの情報であるかどうかを判定可能である。しかし，繰り返しになるが，本論文が対象とするシナリオでは，移動履歴のデータマイニングを対象としていないことに注意いただきたい。

また，スマートフォン等のセンシングデバイス自体への攻撃による情報抽出等も想定される[9]。この問題は本論文のスコープ外であるが，マルウェア検知手法等を用いて対処することが考えられる[10]。

〈2·3〉 位置情報の表現 位置情報は，2次元平面上のグリッドセルで表されるものとする。本論文では，「位置情報」と「グリッドセル」を同じ意味で使用する。

なお，対象とするエリアは，ユーザが存在し得る場所に限定する。

〈2･4〉 プライバシモデル　本論文で想定するプライバシモデルを述べる。各ユーザが自分の情報を開示することによってサーバに与える情報をプライバシ情報と定義する。言い換えると，ユーザが匿名データ収集に参加しているかどうかにかかわらず，サーバが当該ユーザに関して推測できるような情報はプライバシ情報とはみなさない。

一般的なアンケート調査を例に挙げて説明する。あるユーザ集団に対し，出身地のアンケート調査を行うことを考える。選択肢として，鳥取県，島根県，その他，があるとする。あるユーザ A の回答が「鳥取県か島根県のいずれか」であり，ユーザ B は未回答であったとする。このとき，その他ほぼ全てのユーザが「鳥取県である」と回答した場合，ユーザ A やユーザ B についての出身地も「高い確率で鳥取県である」と推測することができる。しかしアンケートに回答していないユーザの情報が，その他多くのユーザの回答結果から推測されたとしても，通常はプライバシ情報の漏洩とはみなされないと考えられ，本論文においてはこのようなプライバシモデルを想定する。これは，Evfimievski ら[11] や Kasiviswanathan ら[12] が想定するモデルと同一である。

このモデルに基づいて，プライバシ保護レベルを具体的な数値として表すプライバシ指標は，〈3･1〉において述べる。

3. 指　　標

〈3･1〉 プライバシ指標　多くの論文で採用されているプライバシ指標として，k-匿名性[13] がある。多数のユーザの属性データを格納したデータベースに対し，あるユーザ A について，ユーザ A の秘匿属性以外の全属性を知っている攻撃者が，データベースのどのレコードがユーザ A のものであるか，高々$1/k$ の確信度でしか分からない，という定義の下で利用されていることが多い。

本論文でもこの k-匿名性の指標を採用する。本論文はユーザの位置情報を対象としているため，ユーザの位置情報の候補として，k 個以上のグリッドセルが存在すれば k-匿名性を満たすと定義する。

一方，位置情報の k-匿名性に関するいくつかの既存研究では，ある範囲に k 人以上のユーザがいるように，各ユーザの位置情報を曖昧化することを目指している[14]~[16]。しかし，このアプローチでは，k 人が全く同じ地点に存在している場合，各ユーザの位置情報は全く曖昧化されず，正確な位置情報がサーバに伝わってしまうという問題がある[17]。また，ある範囲に k 人以上のユーザがいるように位置情報を曖昧化する手法を用いる場合，ユーザは自分だけで k-匿名性を実現することはできず，曖昧化されたエリアに本当に k 人存在するかどうかを確認することは困難である。一方，本論文で採用する k-匿名性の指標では，本論文で提案する手法を用いることで，ユーザ自身が k-匿名性を満たす匿名化を行うことが可能であるという利点がある。

なお，k-匿名性に関する多くの研究もそうであるように，攻撃者の事前知識によっては，実質的に k-匿名性が満たされなくなる場合がある。たとえば，あるユーザ U が存在するグリッドセルの候補が，$\{A_1, A_2, A_3, A_4\}$ であることが攻撃者に判明した状況を考える。この時点では，4-匿名性が満たされている。このとき攻撃者が，A_3 には物理的に誰も存在しないことを知っていたとする。この場合，ユーザ U が存在するグリッドセルの候補は $\{A_1, A_2, A_4\}$ の 3 箇所に絞られるため，実質的に 3-匿名性までしか満たされていない。しかしながら，各グリッドセルは一定程度の面積を持っているため，あるグリッドセルに物理的に誰も存在していないことを把握することは，一般に困難である。たとえば，各グリッドセルの範囲を，建物内含めて全て網羅する監視カメラ群を設置することや，多数の人員を投入して物理的に観測することが考えられるが，いずれもコストが大きいと考えられる。

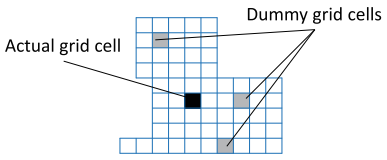


Fig. 2.　Example of anonymization for 4-Anonymity

また，誰か一人でも上記の各グリッドセルに存在している場合は，その人物がユーザ U なのかどうかを，攻撃者は判断することはできない。

なお，攻撃者の事前知識によって k-匿名性が実質的に保護されなくなることは，データベースに対する k-匿名化を行う多くの研究でも同じであることに注意いただきたい。たとえば，あるデータベースにおいて，秘匿属性以外の全属性が全く同じ値を持つ k 個のレコードがあったとする。また，これらの k 個のレコードの秘匿属性は全て異なっているとする。この場合，この k 個のレコードのみに注目すると，k-匿名性が満たされている。しかし攻撃者が，その一つのレコードに該当する人物 X の秘匿属性を知っていたとする。このとき，人物 X 以外の $(k-1)$ 人にとっては，実質的に k-匿名性が満たされておらず，$(k-1)$ 個のレコードのいずれかである，ということが判明してしまう。

本論文では，各グリッドセルが既に一定範囲の面積を持っており，かつ，ユーザが存在する候補として k 個以上のグリッドセルが存在することを保証する。本論文と同じように，複数箇所のダミーを発生させることによってユーザのプライバシを守ろうとする既存研究も数多く存在する[18]~[21]。

k-匿名性を満たした匿名化を行う例を図 2 に示す。線で区切られたエリアは，各グリッドを表す。黒いグリッドセルは，あるユーザの真の位置情報を表しているとする。当該ユーザが情報収集サーバに，黒いグリッドセルまたは 3 つの灰色のグリッドセルのいずれかに存在している，という情報を伝える。このとき 4-匿名性が満たされている。

本論文では，この k の値を「プライバシ保護レベル」と呼ぶ。

〈3･2〉 有効性指標　サーバ側で各グリッドセルに存

在するユーザ数を推測するが，その推測値と真の値との差における，平均二乗誤差を有効性指標とする。多くの既存研究においても，推測値の精度を測る指標として真の値との平均二乗誤差が利用されている[3][22]~[24]。

あるグリッドセル L_i に存在する真のユーザ数を V_i とおく。グリッドセル L_i に存在するとサーバ側で推測されたユーザ数を $\widehat{V_i}$ とおく。グリッドセルの総数を D，データを収集したユーザ総数を N とする。このとき，平均二乗誤差 σ^2 を次式で定義する。

$$\sigma^2 = \frac{1}{D}\sum_{i=0}^{D-1}\left(\frac{V_i}{N} - \frac{\widehat{V_i}}{N}\right)^2 \cdots\cdots\cdots\cdots\cdots\cdots(1)$$

4. 関連研究

〈4·1〉 個別ユーザへのサービス提供のためのダミー生成　各グリッドセルに存在するユーザ数を求めるのではなく，各ユーザに対して個別のサービスを提供する際に，プライバシを保護するためにダミーの位置情報を利用する手法が提案されている[19][20][25][26]。これらは，ユーザがロケーションアウェアサービスを利用したい際に，真のグリッドセルだけではなく，ダミーのグリッドセルもサービス提供サーバへ通知し，各グリッドセルに対する応答を受領する。ユーザは，真のグリッドセルに相当する応答のみを採用する，というプライバシ保護手法である。

ダミーを利用するという点ではこれらの研究と本研究は同じであるが，これらの研究では，各グリッドセルに存在するユーザ数を推測することはできない。各グリッドセルに存在するユーザ数を推測したい場合は，本研究の手法を併用することで可能である。

〈4·2〉 Obfuscation アプローチ　真のグリッドセルを含む，k 個のグリッドセルで表現される矩形エリア（Obfuscation エリア）を導出し，この Obfuscation エリアのどこかに存在する，という情報のみをサーバへ通知する手法が提案されている[27]。本論文で提案する手法ではダミーのグリッドセルをランダムに設定するが，Obfuscation アプローチでは，ダミーのグリッドセルを真のグリッドセルと隣接するひとまとまりのエリアとして設定する。この研究は各グリッドセルに存在するユーザ数を導出することを目的とはしていないが，本論文の提案手法を用いることにより，Obfuscation アプローチで収集された情報からも，各グリッドセルに存在するユーザ数を推測することが可能である。

〈4·3〉 秘匿性を考慮した位置情報収集　各ユーザにとって秘匿性が高い場所であれば，位置情報を何ら提供せず，そうでない場合にのみ位置情報を提供する手法も取られている[28][29]。各場所に対し，秘匿性が高い場所としてユーザが設定する割合があらかじめ既知であれば，この手法を利用することで，各場所に存在するユーザ数を推定することが可能であると考えられる。しかし，一般にそのような割合は未知であると考えられ，これらの手法を用いて各場所に存在するユーザ数を高精度に推定することは困難であると考えられる。

〈4·4〉 Negative Survey　ユーザは自分の真のデータをサーバに絶対に報告しない，という制約を設けることで，プライバシを保護しつつ，サーバで真のデータ分布を推測する Negative Survey[5] という手法が提案されている。

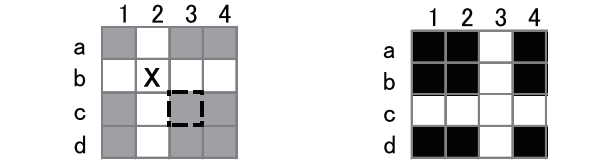
Negative Survey は，真のグリッドセルが L_i であるとき，サーバには L_i 以外のグリッドセルを一つ報告する。ユーザは必ず真実でない情報をサーバに提供するため，プライバシが一定程度守られる。またサーバ側も，ユーザは必ず真実でない情報を提供する，という制約を考慮することで，多くのユーザから情報を収集することにより，真のデータ分布を推測することができる。しかし，数多くのユーザから情報を収集しなければ高い精度で推測することができない。

サーバでの推測精度を向上させるため，Negative Survey を位置情報に応用した研究として，Forrest らの手法がある[4]。あらかじめ対象エリアを $2^n \times 2^n$ セルに分割し，各セルに対して，0〜3 までの数字のみを利用して n 桁からなる一意の ID を振る。各ユーザは，ユーザの真のセル ID と，$2^n \times 2^n$ 個全てのセル ID を比較し，全ての桁においてユーザの真のセル ID とは異なる数字が設定されている ID の集合を作成する。たとえば $n = 2$ であり，ユーザの真のセル ID が 01 である場合，$\{10, 12, 13, 20, 22, 23, 30, 32, 33\}$ の集合が作成される。ユーザはこの集合からランダムに一つを選び，サーバへ報告する。この手法を **NQT 手法**と呼ぶ。

また，収集対象のユーザ属性が複数存在するときに特に有効な Multidimensional Negative Survey という手法が提案されている[2][3]。この手法は位置情報に特化したものではないが，位置情報を緯度と経度の 2 つの属性とみなして適応させることが可能である。この手法を **MDA 手法**と呼ぶ。あらかじめ対象エリアを $x \times y$ セルに分割し，各セルを X 座標と Y 座標のペアで表現する。つまり，各セルは $(1,1), (1,2), \ldots, (x,y)$ のいずれかにより参照される。各ユーザは，ユーザの真のセル座標が (u_x, u_y) である場合，X 座標が u_x でなく，かつ，Y 座標が u_y でないセルの集合を作成する。たとえば $x = 2$，$y = 3$ であり，ユーザの真のセル座標が $(1,2)$ である場合，$\{(2,1), (2,3)\}$ の集合が作成される。ユーザはこの集合からランダムに一つを選び，サーバへ報告する。

この NQT 手法，MDA 手法は，6 章で提案手法と比較評価を行う。

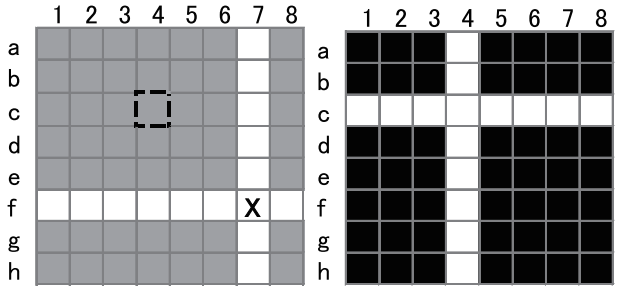
なお，この NQT 手法及び MDA 手法ではいずれも，ユーザは真のグリッドセルの情報をサーバに通知しないが，サーバ側は，ユーザが存在し得るグリッドセルをある程度絞ることができる。ここでは例として，MDA について説明する。まず，4×4 のグリッドセルを対象に匿名化を行う場合について述べる。図 3 (a) において，あるユーザが X で表される $(b, 2)$ のグリッドセルに存在すると想定する。このとき，MDA では，ユーザは図 3 (a) において灰色で示されているグリッドセルのいずれかをサーバに報告する。こ



(a) Candidates of grid cells reported to the server when the user exists at $(b, 2)$)

(b) Grid cells where the user might exist when the user reported $(c, 3)$

Fig. 3. 4×4 grid cells in MDA



(a) Candidates of grid cells reported to the server when the user exists at $(f, 7)$)

(b) Grid cells where the user might exist when the user reported $(c, 4)$

Fig. 4. 8×8 grid cells in MDA

こでは，$(c, 3)$ を報告したと仮定する。$(c, 3)$ を受け取ったサーバは，図 3 (b) において，ユーザが黒いグリッドセルのいずれかに存在している，ということが分かる。何故なら，$(c, 3)$ を報告し得るグリッドセルが，この黒いグリッドセルに限定されるためである。このとき，黒いグリッドセルは 9 箇所あるので，$k = 9$ となる k-匿名性が満たされている。

次に，8×8 のグリッドセルを対象に匿名化を行う場合について説明する。図 4 (a) において，あるユーザが X で表される $(f, 7)$ のグリッドセルに存在すると想定する。このとき，MDA では，ユーザは図 4 (a) において灰色で示されているグリッドセルのいずれかをサーバに報告する。ここでは，$(c, 4)$ を報告したと仮定する。$(c, 4)$ を受け取ったサーバは，図 4 (b) において，ユーザが黒いグリッドセルのいずれかに存在している，ということが分かる。何故なら，$(c, 4)$ を報告し得るグリッドセルが，この黒いグリッドセルに限定されるためである。このとき，黒いグリッドセルは 49 箇所あるので，$k = 49$ となる k-匿名性が満たされている。

NQT についても同様に，サーバ側は受け取った情報から，ユーザが存在し得るグリッドセルを限定することができる。この数が k であるとき，k-匿名性が満たされている。

5. 提案手法

〈5·1〉 概　　要　ユーザは，「この中のどこかに存在している」という**存在可能性集合**をサーバへ報告する。報告する存在可能性集合には，実際に当該ユーザが存在しているグリッドセル L_i を含め，k 個のグリッドセルを設定する。十分な数のユーザから存在可能性集合を収集できたら，サーバ側において，各グリッドセルに何人のユーザが存在していたかについて推測を行う。

以下では，ユーザプロトコル及びサーバプロトコルを順番に記載し，何人のユーザから情報を集めれば有効性がどの程度の値になるかの期待値の算出方法を次に述べる。最後に，計算量のオーダーについて記述する。

〈5·2〉 ユーザプロトコル　あらかじめ各グリッドセルの緯度及び経度のそれぞれの座標と，グリッドセルの総数 D，及びプライバシ保護レベル k を決定しておき，ユーザ側とサーバ側で共有しておく†。ユーザ u の真のグリッドセルが L_i であったとする。ダミーのグリッドセルをランダムに $k-1$ 個生成し，これら k 個のグリッドセルの ID を，存在可能性集合 R_u としてサーバへ通知する。

例として，$D = 4, k = 2$ の場合を考える。あるユーザ u の真のグリッドセルを L_1 とする。このとき，ユーザ u の存在可能性集合は，$\{1, 0\}, \{1, 2\}, \{1, 3\}$ のいずれかがランダムに生成される。

ユーザプロトコルを Algorithm 1 に示す。ここで，$rand(D)$ は，0 以上 $D-1$ 以下の整数をランダムに返す関数である。

Algorithm 1 User protocol of user u

Input: # of grid cells D, Privacy level k, User u's actual grid cell ID i
Output: Possible Exsiting Set
1: Creates empty set R_u
/*Adds actual grid cell ID*/
2: $R_u \Leftarrow R_u \cup \{i\}$
/*Generates and adds dummy grid cell IDs*/
3: **while** $|R| < k$ **do**
4: $r \Leftarrow rand(D)$
5: $R_u \Leftarrow R_u \cup \{r\}$
6: **end while**
7: **return** R_u

〈5·3〉 サーバプロトコル　各グリッドセル L_i に存在するユーザ数を推測する。L_i に存在した真のユーザ数を V_i とおき，L_i に存在するとサーバ側で推測するユーザ数を $\widehat{V}_i$ とおく。また，存在可能性集合を報告したユーザ数を N とおく。

存在可能性集合 R_j $(j = 0, \ldots, N-1)$ の集合 S のうち，L_i が含まれている存在可能性集合の数を W_i とおく。つまり，

$$W_i = |\{R_j | L_i \in R_j \wedge R_j \in S\}| \cdots\cdots\cdots\cdots\cdots (2)$$

と表される。

グリッドセル L_{i_1} にユーザが存在していた場合，確率 1 で L_{i_1} が当該ユーザの存在可能性集合に含まれ，

$$P_E = \frac{k-1}{D-1} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (3)$$

† 議論を簡単にするため，全ユーザのプライバシ保護レベルが同一であると仮定するが，7 章で述べるように，各ユーザで異なるプライバシ保護レベルを設定できるよう拡張することも可能である。

の確率で，$i_1 \neq i_2$ となるグリッドセル L_{i_2} が存在可能性集合に含まれる。

したがって，確率通り配分されるとすると，以下の式が成立する。

$$W_i = N - \sum_{j \neq i}(1 - P_E)\widehat{V_j} \cdots\cdots\cdots\cdots\cdots\cdots (4)$$

各 W_i $(i = 0, \ldots, D-1)$ に対してこの式を構築することにより，$\widehat{V_j}$ $(j = 0, \ldots, D-1)$ を変数とした，D 元一次連立方程式が構築される。これを解くことにより，各 $\widehat{V_j}$ が求まる。このとき，各 $\widehat{V_j}$ は V_j の最尤推定量であり，かつ不偏推定量となっている。

例として，$D = 4$，$k = 2$ であり，また (2) 式より，$W_0 = 35, W_1 = 50, W_2 = 80, W_3 = 35$ である場合を考える。このとき，(3) 式より，$P_E = 1/3$ である。したがって，(4) 式より，以下の連立方程式が生成される。

$$\begin{cases} 35 = 100 - 2/3 \times (\widehat{V_1} + \widehat{V_2} + \widehat{V_3}) \\ 50 = 100 - 2/3 \times (\widehat{V_0} + \widehat{V_2} + \widehat{V_3}) \\ 80 = 100 - 2/3 \times (\widehat{V_0} + \widehat{V_1} + \widehat{V_3}) \\ 35 = 100 - 2/3 \times (\widehat{V_0} + \widehat{V_1} + \widehat{V_2}) \end{cases} \cdots\cdots\cdots (5)$$

この連立方程式を解くことにより，$\widehat{V_0} = 2.5, \widehat{V_1} = 25, \widehat{V_2} = 70, \widehat{V_3} = 2.5$ が得られる。

サーバプロトコルを Algorithm 2 に示す。

Algorithm 2 Server protocol

Input: # of grid cells D, Possible Existing Sets S
Output: Distribution of user's grid cells

```
Creates Array W, V̂
/*Calculates each W_i*/
for i = 0, ..., D − 1 do
    W_i ⇐ result calculated by Eq. 2
end for
/*Calculates P_E*/
P_E ⇐ result calculated by Eq. 3
/*Creates and solves the linear system of equations*/
Creates Eq. 4 for all j = 0, ..., D − 1
for j = 0, ..., D − 1 do
    V̂_j ⇐ each result calculated by the linear system of equations
end for
return V̂
```

〈5·4〉 有効性の期待値の算出　何人のユーザから情報を収集する必要があるかを決定するために，有効性の期待値を算出する必要がある。

本論文では有効性指標として平均二乗誤差を用いているが，平均二乗誤差を定義する (1) 式より，その期待値は次式で表すことができる。

$$E[\sigma^2] = \frac{1}{D}\sum_{i=0}^{D-1} E\left[\left(\frac{V_i}{N} - \frac{\widehat{V_i}}{N}\right)^2\right] \cdots\cdots\cdots (6)$$

また，$\widehat{V_i}$ は V_i の不偏推定量であることから，$E[\widehat{V_i}] = V_i$ である。また，$E[(E[\widehat{V_i}] - \widehat{V_i})^2]$ は $\widehat{V_i}$ の分散を表している。したがって，$\widehat{V_i}$ の分散を $Var(\widehat{V_i})$ と表現すると，

$$E[\sigma^2] = \frac{1}{N^2}\frac{1}{D}\sum_{i=0}^{D-1} Var(\widehat{V_i}) \cdots\cdots\cdots\cdots\cdots\cdots (7)$$

である。

また，(4) 式を全ての i に対して構築してできる連立方程式は，式変形を行うことにより，次式で表される方程式と等価である。

$$M \cdot \vec{V}^\tau = \vec{W}^\tau, \text{ where}$$

$$M = \begin{pmatrix} 0 & 1-P_E & \cdots & 1-P_E \\ 1-P_E & 0 & \cdots & 1-P_E \\ \vdots & \vdots & \vdots & \vdots \\ 1-P_E & 1-P_E & \cdots & 0 \end{pmatrix}, \cdots\cdots (8)$$

$$\vec{V} = (\widehat{V_0}, \widehat{V_1}, \ldots, \widehat{V_{D-1}}),$$

$$\vec{W} = (N - W_0, N - W_1, \ldots, N - W_{D-1})$$

したがって，

$$\vec{V}^\tau = M^{-1} \cdot \vec{W}^\tau \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (9)$$

が成り立つ。

ここで，一般に確率変数 X 及び Y に対して，共分散を $Cov(X, Y)$ とおくと，

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y) \cdots\cdots\cdots (10)$$

が成り立つ（a, b は定数）。M^{-1} の i 行 j 列を $M^{-1}_{i,j}$ と表す。(9) 式より $\widehat{V_i} = \sum_j M^{-1}_{i,j}(N - W_j)$ であるから，(8), (10) 式より

$$Var(\widehat{V_i}) = \sum_j (M^{-1}_{i,j})^2 Var(N - W_j) + \sum_{\substack{j,l \\ (j \neq l)}} M^{-1}_{i,j} M^{-1}_{i,l} Cov(N - W_j, N - W_l) \cdots\cdots\cdots\cdots\cdots (11)$$

である。ここで，任意の存在可能性集合に L_j が含まれる確率を P_j とする。また，任意の存在可能性集合に L_j が含まれているとき，$j \neq l$ である L_l もその集合に含まれている確率を $P_{j,l}$ とする。任意の存在可能性集合に L_j が含まれているとき，$j \neq l$ である L_l がその集合に含まれていない確率を $P_{j,\bar{l}}$ とする。このとき，Var 及び Cov は以下のように表すことができる。

$$\begin{aligned} Var(N - W_j) &= E[W_j^2] - E[W_j]^2 \\ &= \sum_{s=0}^{N} P_j^s (1 - P_j)^{N-s} {}_N C_s \cdot s^2 - P_j^2 \\ &= NP_j(1 - P_j), \end{aligned}$$

$$
\begin{aligned}
Cov(N-W_j, N-W_l) &= E[W_jW_l] - E[W_j]E[W_l] \\
&= \sum_{s=0}^{N}\sum_{t=0}^{N-s}\sum_{u=0}^{N-s-t}\{P_{j,l}^{s}P_{j,\bar{l}}^{t}P_{l,\bar{j}}^{u} \\
&\quad \cdot(1-P_{j,l}-P_{j,\bar{l}}-P_{l,\bar{j}})^{N-s-t-u} \\
&\quad \cdot {}_NC_s \cdot {}_{N-s}C_t \cdot {}_{N-s-t}C_u \cdot (s+t)(s+u)\} \\
&\quad -NP_jNP_l \\
&= N(P_{j,l}+(P_{j,l}+P_{j,\bar{l}})(P_{j,l}+P_{l,\bar{j}})(N-1)) \\
&\quad -NP_jNP_l \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (12)
\end{aligned}
$$

あらかじめ V_i の分布がある程度予想できる場合は，それに応じて P_j，P_l，$P_{j,l}$，$P_{j,\bar{l}}$，$P_{l,\bar{j}}$ を計算しておき，平均二乗誤差の期待値を算出する。一方，V_i の分布が予想できない場合は，それぞれ j や l の値によらず一様分布であると想定することになる。この場合，

$$
\begin{aligned}
&P_j = P_l = \frac{k}{D} \\
&P_{j,l} = \frac{k}{D}\cdot\frac{k-1}{D-1} \qquad\qquad \cdots\cdots\cdots\cdots (13) \\
&P_{j,\bar{l}} = P_{l,\bar{j}} = \frac{k}{D}\cdot\left(1-\frac{k-1}{D-1}\right)
\end{aligned}
$$

となる。ここで，P_j 及び P_l は D 個の要素の中から k 個の要素を選択する状況において，ある特定の要素がそれに含まれる確率を表している。$P_{j,l}$ は，同じ状況において，ある特定の 2 つの要素がそれに含まれる確率を表している。$P_{j,\bar{l}}$ 及び $P_{l,\bar{j}}$ は，同じ状況において，ある特定の要素がそれに含まれていて，かつ，別のある特定の要素がそれに含まれていない確率を表している。

このように設定することで本来の平均二乗誤差と乖離が大きくなると，期待値をあらかじめ計算することはできない。しかし，6 章におけるシミュレーション評価において示すように，平均二乗誤差の値は V_i の値の分布にほとんど依存しない。また，Negative Survey や，その一般化された形である Randomized Response の提案を行っている既存研究においても，多くのシミュレーション結果から，平均二乗誤差の値はデータにほとんど依存しないことが報告されている [(2) (3) (22)]。

(13) 式のように仮定する場合，

$$
\begin{aligned}
E[\sigma^2] = \frac{1}{N}\Bigg[&\frac{k}{D}\left(1-\frac{k}{D}\right)\sum_j (M_{i,j}^{-1})^2 \\
&-\frac{k(D-k)}{(D-1)D^2}\sum_{j,l,j\neq l} M_{i,j}^{-1}M_{i,l}^{-1}\Bigg] \cdots\cdots (14)
\end{aligned}
$$

と表される。

ここで，逆行列は，

$$
M^{-1} = \frac{1}{D-k}\begin{pmatrix} 2-D & 1 & \cdots \\ 1 & 2-D & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \cdots\cdots\cdots (15)
$$

となる。

したがって，

$$
\sum_j (M_{i,j}^{-1})^2 = \left(\frac{1}{D-k}\right)^2\{(2-D)^2+D-1\} \qquad \cdots\cdots\cdots\cdots (16)
$$

である。

また，

$$
\begin{aligned}
&\sum_{j,l,j\neq l} M_{i,j}^{-1}M_{i,l}^{-1} = \\
&\qquad \left(\frac{1}{D-k}\right)^2\cdot\{2(2-D)(D-1)+(D-1)(D-2)\} \\
&\qquad\qquad\qquad\qquad \cdots\cdots\cdots\cdots (17)
\end{aligned}
$$

である。

結果，(14)，(16)，(17) 式より，平均二乗誤差の期待値は次の式で表される。

$$
E[\sigma^2] = \frac{k(D-1)^2}{N(D-k)D^2} \cdots\cdots\cdots\cdots\cdots\cdots (18)
$$

〈5·5〉 計 算 量 ノードプロトコルは，$k-1$ 個のダミーのグリッドセルを生成するため，計算量は $O(k)$ である。Algorithm 1 からも分かるように，ノードはごく単純な計算を行うだけで良い。

サーバプロトコルでは，D 元一次連立方程式を解く部分が最も計算量が大きい。D 元一次連立方程式は行列演算で解くことが可能である。ガウスの消去法を用いた場合，計算量のオーダーは $O(D^3)$ となる。反復法を用いたガウス＝ザイデル法を利用した場合，反復回数を r とすると計算量のオーダーは $O(rD^2)$ である。

6. 評　　価

〈6·1〉 評価対象 4 章の関連研究で示した NQT 及び MDA を比較対象とする。NQT 及び MDA は，位置情報を収集する対象範囲のグリッドセル数に依存して k の値が固定されるため，k の値を調整できない。したがって，評価において k の値を変動させる場合，NQT 及び MDA についてはその範囲内で取り得る値においてのみ評価を行った。表 1 に実験環境を示す。

〈6·2〉 評価内容 平均二乗誤差の期待値を算出した数学的解析及び，シミュレーションによる評価を行った。シミュレーション評価については，オープンソースの移動体シミュレータ Siafu [(30)] を利用した。Siafu はコンテキスト情報を考慮したユーザの移動についてのシミュレータとして，多くの研究で利用されている。地図データは，Siafu シミュレータに同梱されている Leimen の地図を，大きさが 4.2 km × 4.2 km になるように修正して利用した。また，

Table 1. Simulation environment

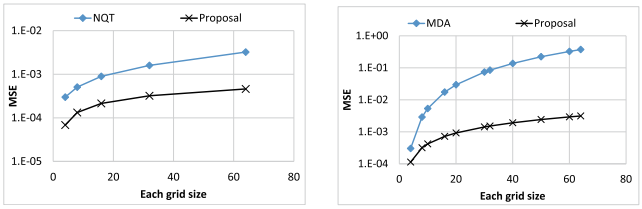
OS	Windows 7 Professional 64 bit
CPU	Intel Core i7-3712QM CPU @ (2 CPUs)
RAM	8.00 GB

ユーザ数は 96,000 人，グリッド数は 16×16=256 に設定した。ユーザは自宅から会社まで徒歩または車で移動するものとし，仕事を開始する平均時間を朝 7 時に設定した。朝 5 時からシミュレーションを開始し，7 時の時点におけるユーザの位置情報を評価に利用した。

また，グリッドセルの対象範囲は，MDA と NQT に関しては矩形である必要がある。特に NQT に関しては正方形，つまり，緯度におけるグリッドセル数と経度におけるグリッド数が同じである必要がある。提案手法は対象範囲がどのような形でも対応可能であるが，これら既存手法との比較を行うため，緯度と経度のグリッド数が同一である環境を設定した。

〈6·3〉 評価結果 まず，平均二乗誤差の期待値を算出した数学的解析結果を示す。

緯度及び経度のグリッドセル数を変えて，平均二乗誤差の期待値を算出した結果を，図 5 に示す。ユーザ数は 10,000 人に設定した。



(a) Comparison with NQT (b) Comparison with MDA

Fig. 5. Relationship between the number of grids and mean squared error

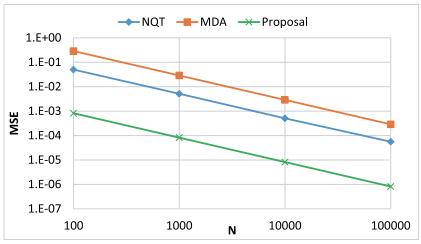


Fig. 6. Relationships between the number of users and mean squared error

図 5 (a) は NQT と提案手法の比較評価を行い，図 5 (b) は MDA と提案手法の比較評価を行っている。なお，図中の MSE は平均二乗誤差（Mean Squared Error）を表している。また，図の横軸は緯度及び経度それぞれのグリッドセル数を表している。たとえば横軸が 50 である場合，合計で 2,500 のグリッドセルが存在することになる。図より，NQT や MDA に対し，提案手法の平均二乗誤差が大きく削減されていることが分かる。

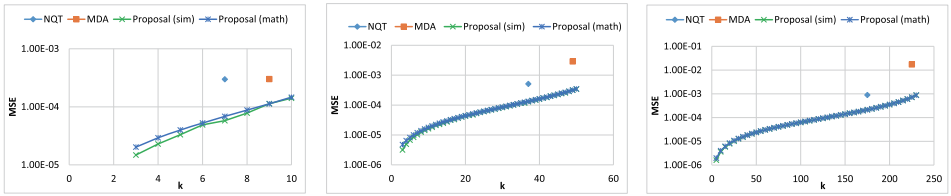
次に，ユーザ数を変えて評価した結果を図 6 に示す。

いずれの手法においても，ユーザ数と平均二乗誤差が反比例の関係にあることが分かる。提案手法においては，(18) 式よりこのような関係になることは明らかである。

次に，ランダムに生成した位置情報を用いてシミュレーション評価を行った。シミュレーションにおいては，各グリッドセルにユーザが存在する密度の分布をガンマ分布，正規分布，一様分布それぞれでパラメータを変えながら実験を行った。各人工データに対してシミュレーションを 50 回繰り返し，その平均値を取った。ガンマ分布における結果を図 7 に示す。なお，他いずれの分布においてもほとんど同じ結果となった。

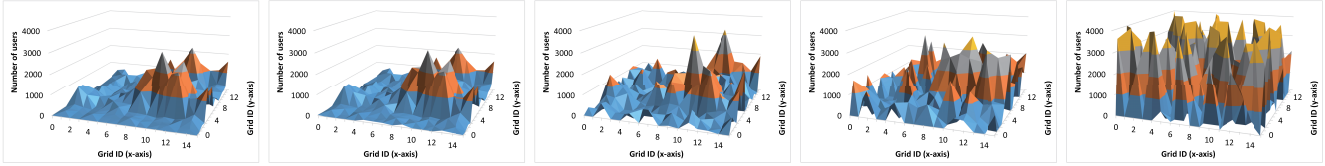
Proposal (sim) は提案手法のシミュレーション結果を，Proposal (math) は，(18) 式を用いて算出した平均二乗誤差の期待値を表す。図より，k の値が大きいほど，平均二乗誤差が大きくなっていることが分かる。また，いずれの k 及びいずれのグリッドサイズにおいても，提案手法が NQT，MDA よりも平均二乗誤差を削減できていることが分かる。さらに，(18) 式が，実際に導出された平均二乗誤差の値とほぼ一致していることも図から読み取れる。

次に，Siafu のシミュレーションデータを用いて評価を行った。Siafu によって生成された生データの分布及び，各手法を用いて推測した結果を図 8 に示す。Siafu を用いて 10 回シミュレーションを行ったうち，ランダムに 1 つを抽



(a) 4x4 grid cells (b) 8x8 grid cells (c) 16x16 grid cells

Fig. 7. For gamma distribution with different k



(a) Original data set (b) Estimated result (Proposal (k=10)) (c) Estimated result (Proposal (k=175)) (d) Estimated result (NQT) (e) Estimated result (MDA)

Fig. 8. Distribution of simulation data generated by Siafu and estimated results

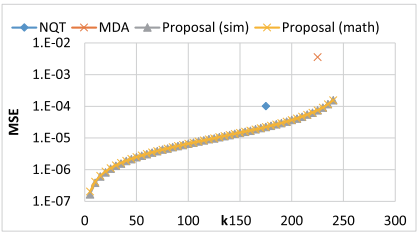


Fig. 9. Relationships between k and mean squared error in Siafu

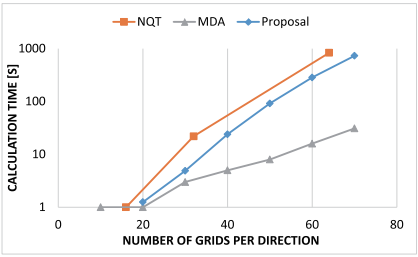


Fig. 10. Calculation time for simulations

出し，その結果を示している。

k-匿名性を行う研究の多くが k の値を 3 から 20 程度に設定しているため[31][32]，ここでは $k = 10$ に設定した結果を載せている（図 8 (b)）。また，NQT の手法を用いると $k = 175$ であるため，この値に設定した場合における提案手法の結果も示している（図 8 (c)）。$k = 10$ と設定した提案手法は，オリジナルのデータ分布をほぼ正確に推測できていることが分かる。k の値が同じ 175 である，Proposal (k=175) の結果と，NQT の結果を比べると，提案手法のほうが生データの分布をよく推測できていることが読み取れる。また，MDA は生データをほとんど推測できていないことが分かる。

Siafu のシミュレーションデータに対し，k を変えながら平均二乗誤差を計測した結果を図 9 に示す。Siafu のシミュレーションを 10 回繰り返し，その平均値を取った。

提案手法が最も小さい平均二乗誤差を実現していること，及び，平均二乗誤差の期待値の算出結果と実際のシミュレーション結果がよく一致していることが分かる。

最後に，人工的に生成したシミュレーションデータに対し，推測に必要な時間を計測した。結果を図 10 に示す。人工データの生成及びシミュレーション実行を 50 回繰り返し，その平均値を取った。

横軸は緯度及び経度におけるグリッド数を表しており，たとえばこの値が 70 のとき，グリッドセルの総数は 4,900 個存在する。いずれの手法もグリッドセル数に対して非線形に計算時間が増加していることが分かる。しかし，グリッドセルの総数が 4,900 個存在するということは，日本の各都道府県を 100 分割したグリッドセルに対しても，1 時間未満で推測可能ということを表しており，少なくともこの範囲内では有効であると考えられる。なお，各グリッドセルにユーザが存在する密度の分布をどのように変えても，推測に要した時間はほぼ同じであった。

7. 考　　察

〈7·1〉 ユーザごとに異なるプライバシ保護レベルの設定　本論文においては，議論を簡単にするために全ユーザで共通のプライバシ保護レベルを設定することを想定している。しかし，ユーザごとにプライバシ保護レベルを変更することも可能である。その手順は以下のとおりである。

- k の値ごとにユーザをグルーピングする。
- 各グループにおいて，提案手法を用いて各グリッドセルに存在するユーザ数を推測する。
- 各グループの計算結果を単純に足し合わせることで，全体の推測値とする。

この手順の有効性を確認するため，各ユーザが k の値を 5 から 15 までの範囲でランダムに設定する状況を考え，追加実験を行った。実験は，6 章で用いた，96,000 ユーザの位置情報を利用した。このときの MSE は，4.0×10^{-7} となった。なお，ユーザ全員が $k = 5$ に設定した場合の MSE は 1.6×10^{-7}，ユーザ全員が $k = 15$ に設定した場合の MSE は 6.1×10^{-7} である。したがって，ユーザ全員が $k = 15$ と設定した場合よりも，ユーザが $k = 5$ から 15 までの間でランダムに設定しているほうが，高い精度で推測できることを確認できた。

k の値がユーザごとに異なる場合における，より精度の高い計算方法は将来課題とする。なお，NQT や MDA では k の値を自由に変更できないため，ここでは比較していない。

〈7·2〉 ダミーのグリッドセルのばらつき　提案手法ではダミーの位置情報として $k-1$ 個のグリッドセルを選択するが，このグリッドセルは，真のグリッドセルに近くまとまっているよりも，広くばらばらに設定されているほうが，直観的なプライバシ保護レベルは高いと考えることもできる。

本論文においては，ダミーのグリッドセルを完全にランダムに選択しているため，ある程度，ダミーのグリッドセルがばらついていることが期待されるが，将来課題として，ばらつきの度合いを測ることができる指標を新たに提案し，提案手法において数学的な解析を行うことを考えている。

8. おわりに

ユーザの位置情報をスマートフォン等のモバイル端末から取得し，多数のユーザの位置情報を収集することによって，ユーザの消費行動と行動記録を結びつけたマーケティング分析等を行うことができる。本論文では，各ユーザが自分の正確な位置情報を知られたくないという状況を考え，k 箇所のグリッドセルのいずれかに存在する，という情報のみをサーバに提供するシナリオを想定した。提案手法では，真のグリッドセルの情報に加え，$k-1$ 個のダミーのグリッドセルをランダムに生成し，存在可能性集合としてサーバへ通知する。サーバでは，当該ユーザがこの k 個の

グリッドセルのうち，どこに存在しているかについて知ることはできない。本論文では，多くのユーザからこの存在可能性集合の情報を収集することで，各グリッドセルに何人のユーザが存在していたかについて小さい誤差で推測できる手法を提案した。また，既存研究では k の値を柔軟に変更することができないという制約があったが，提案手法は任意の値を設定することができる。さらに，その推測誤差の期待値を計算できる式を導出し，シミュレーション結果とほぼ一致していることを示した。提案する手法はシンプルであるが，数学的解析及びシミュレーション評価により，提案手法は既存手法よりもプライバシと推測精度の高いトレードオフを取れることを示した。将来課題として，サーバプロトコルにおける計算速度の向上を考えている。

謝　辞

本研究は JSPS 科研費 24300005, 26330081, 26870201 の助成を受けたものです。

文　献

（1） J. Iio, K. Yoshida, A. Koike, H. Shimizu, Y. Shirai, K. Kuwayama, K. Kuriyama, H. Konami, and S. Takayama: “Location-based Personal Information Log Reveals the Relation between Behavioral Characteristics and Consumption Tendencies”, IPSJ Journal, Vol.52, No.7, pp.2256–2267 (2011) (in Japanese)
飯尾　淳・吉田圭吾・小池亜弥・清水浩行・白井康之・桑山晃一・栗山桂一・小浪宏信・高山隼佑：「属性付き位置情報ログが示す行動特性と消費傾向の関係」，情処学論，Vol.52, No.7, pp.2256–2267 (2011)

（2） M.M. Groat, B. Edwards, J. Horey, W. He, and S. Forrest: “Enhancing privacy in participatory sensing applications with multidimensional data”, Proc. IEEE PerCom, pp.144–152 (2012)

（3） M.M. Groat, B. Edwards, J. Horey, W. He, and S. Forrest: “Application and analysis of multidimensional negative surveys in participatory sensing applications”, Pervasive and Mobile Computing, Vol.9, No.9, pp.372–391 (2013)

（4） S. Forrest and M. Groat: “Reconstructing Spatial Distributions from Anonymized Locations”, Proc. IEEE ICDEW, pp.243–250 (2012)

（5） F. Esponda: “Negative surveys”, Technical report, arXiv.org (2006)

（6） J.A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M.B. Srivastava: “Participatory sensing”, Proc. ACM Sensys Workshop. World Sensor Web, pp.1–5 (2006)

（7） S. Aoki, M. Iwai, and K. Sezaki: “Privacy Protection Method for Environmental Participatory Sensing”, IEICE Trans. Commun., Vol.97, No.1, pp.41–50 (2014) (in Japanese)
青木俊介・岩井将行・瀬崎　薫：「参加型環境センシングを用いた統計情報構築のためのプライバシー保護手法」，信学論，Vol.97, No.1, pp.41–50 (2014)

（8） U. Adeel, S. Yang, and J.A. McCann: “Self-Optimizing Citizen-Centric Mobile Urban Sensing Systems”, Proc. International Conference on Autonomic Computing (CAC), pp.161–167, USENIX Association (2014)

（9） M. La Polla, F. Martinelli, and D. Sgandurra: “A Survey on Security for Mobile Devices”, IEEE Communications Surveys & Tutorials, Vol.15, No.1, pp.446–471 (2013)

（10） A. Shabtai, U. Kanonov, Y. Elovici, C. Glezer, and Y. Weiss: ““Andromaly”: a behavioral malware detection framework for android devices”, Journal of Intelligent Information Systems, Vol.38, No.1, pp.161–190 (2011)

（11） A. Evfimievski, J. Gehrke, and R. Srikant: “Limiting privacy breaches in privacy preserving data mining”, Proc. ACM PODS, pp.211–222 (2003)

（12） S.P. Kasiviswanathan, H.K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith: “What Can We Learn Privately ?”, SIAM Journal on Computing, Vol.40, No.3, pp.793–826 (2013)

（13） K. LeFevre, D. DeWitt, and R. Ramakrishnan: “Incognito: Efficient full-domain k-anonymity”, Proc. ACM SIGMOD, pp.49–60 (2005)

（14） O. Abul, F. Bonchi, and M. Nanni: “Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases”, Proc. IEEE ICDE, pp.376–385 (2008)

（15） T. Takahashi, S. Miyakawa, and N. Ito: “Real-time k-anonymization for Trajectory Stream”, DBSJ Journal, Vol.10, No.1, pp.37–42 (2011) (in Japanese)
高橋　翼・宮川伸也・伊東直子：「移動軌跡ストリームに対するリアルタイム k 匿名化手法の提案」，日本データベース学会論文誌，Vol.10, No.1, pp.37–42 (2011)

（16） C.-Y. Chow, M.F. Mokbel, and T. He: “Tinycasper: a privacy-preserving aggregate location monitoring system in wireless sensor networks”, Proc. ACM SIGMOD, pp.1307–1310 (2008)

（17） R. Shokri, C. Troncoso, C. Diaz, J. Freudiger, and J.-P. Hubaux: “Unraveling an old cloak: k-anonymity for location privacy”, Proc. ACM workshop on Privacy in the electronic society, pp.115–118 (2010)

（18） H. Kido, Y. Yanagisawa, and T. Satoh: “An anonymous communication technique using dummies for location-based services”, Proc. International Conference on Pervasive Services (ICPS), pp.88–97, IEEE (2005)

（19） A. Suzuki, M. Iwata, Y. Arase, T. Hara, X. Xie, and S. Nishio: “A user location anonymization method for location based services in a real environment”, Proc. ACM SIGSPATIAL GIS, pp.398–401 (2010)

（20） T.-H. You, W.-C. Peng, and W.-C. Lee: “Protecting Moving Trajectories with Dummies”, Proc. IEEE MDM, pp.278–282 (2007)

（21） R. Kato, M. Iwata, T. Hara, Y. Arase, X. Xie, and S. Nishio: “User Location Anonymization Method for Wide Distribution of Dummies”, Proc. International Conference on Database and Expert Systems Applications (DEXA), pp.259–273 (2013)

（22） Z. Huang and W. Du: “OptRR: Optimizing Randomized Response Schemes for Privacy-Preserving Data Mining”, Proc. IEEE ICDE, pp.705–714 (2008)

（23） H. Xie, L. Kulik, and E. Tanin: “Privacy-aware collection of aggregate spatial data”, Data & Knowledge Engineering, Vol.70, No.6, pp.576–595 (2011)

（24） J. Horey, M.M. Groat, S. Forrest, and F. Esponda: “Anonymous Data Collection in Sensor Networks”, Proc. MobiQuitous, pp.1–8 (2007)

（25） R. Kato, M. Iwata, T. Hara, A. Suzuki, X. Xie, Y. Arase, and S. Nishio: “A dummy-based anonymization method based on user trajectory with pauses”, Proc. ACM SIGSPATIAL GIS, pp.249–258 (2012)

（26） R. Kato, M. Iwata, T. Hara, A. Suzuki, Y. Arase, X. Xie, and S. Nishio: “A Duumy-based Method Based on User Trajectory with Pause for Location Privacy Protection”, IPSJ Journal, Vol.55, No.1, pp.505–518 (2014) (in Japanese)
加藤　諒・岩田麻佑・原　隆浩・鈴木晃祥・荒瀬由紀・シェシン・西尾章治郎：「停止するユーザの移動経路に基づいた位置プライバシ保護のためのダミー生成手法」，情処学論，Vol.55, No.1, pp.505–518 (2014)

（27） B. Agir, T.G. Papaioannou, R. Narendula, K. Aberer, and J.-P. Hubaux: “User-side adaptive protection of location privacy in participatory sensing”, GeoInformatica, pp.1–27 (2013)

（28） B. Hoh, T. Iwuchukwu, Q. Jacobson, D. Work, A.M. Bayen, R. Herring, J.-C. Herrera, M. Gruteser, M. Annavaram, and J. Ban: “Enhancing Privacy and Accuracy in Probe Vehicle-Based Traffic Monitoring via Virtual Trip Lines”, IEEE Trans. on Mobile Computing, Vol.11, No.5, pp.849–864 (2012)

（29） T.T.B. Le and T.K. Dang: “Semantic-Aware obfuscation for location privacy at database level”, Proc. International Conference on Information and Communication Technology, Vol.111-120, pp.111–120 (2013)

（30） M. Martin and P. Nurmi: “A Generic Large Scale Simulator for Ubiquitous Computing”, Proc. MobiQuitous, pp.1–3

(2006)
（31） L. Yao, G. Wu, J. Wang, F. Xia, C. Lin, and G. Wang: “A Clustering K-Anonymity Scheme for Location Privacy Preservation”, IEICE Trans. Information and Systems, Vol.E95-D, No.1, pp.134–142 (2012)
（32） H. Hu, J. Xu, S.T. On, J. Du, and J.K.-Y. Ng: “Privacy-aware location data publishing”, ACM Trans. Database Systems, Vol.35, No.3, pp.1–42 (2010)

清　雄一

（非会員） 1981 年生。2009 年東京大学大学院情報理工学系研究科博士後期課程修了。同年（株）三菱総合研究所入社。同社情報技術研究センター，金融ソリューション本部等に所属。2013 年より電気通信大学助教，現在に至る。分散コンピューティング，セキュリティ，プライバシ保護技術等の研究に従事。情報処理学会，電子情報通信学会，IEEE Computer Society 各会員。

大須賀　昭　彦

（非会員） 1958 年生。1981 年上智大学理工学部数学科卒業。同年 (株) 東芝入社。同社 研究開発センター，ソフトウェア技術センター等に所属。1985〜1989 年 (財) 新世代コンピュータ技術開発機構 (ICOT) 出向。2007 年より，電気通信大学大学院情報システム学研究科 教授。2012 年より，国立情報学研究所 客員教授 兼任。工学博士 (早稲田大学)。主としてソフトウェアのためのフォーマルメソッド，エージェント技術の研究に従事。1986 年度情報処理学会論文賞受賞。IEEE Computer Society Japan Chapter Chair，人工知能学会理事，日本ソフトウェア科学会理事を歴任。情報処理学会，電子情報通信学会，人工知能学会，日本ソフトウェア科学会，IEEE Computer Society 各会員。