

移動軌跡データベースにおける位置匿名性

貴戸 秀年¹ 柳沢 豊² 佐藤 哲司^{1,2}

¹ 大阪大学 大学院情報科学研究科

² 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所

本稿では、位置情報を利用するサービスにおいて、ユーザの位置に関する匿名性を確保する技術に関して述べる。近年、センシング技術の発展に伴い、GPS 端末や RFID タグといった位置情報を取得するデバイスが安価に入手できるようになった。それに伴い、移動する物体や人間の位置情報のデータを利用したさまざまなサービスが提供され始めてきている。こうしたサービスが広まる一方で、ユーザが発信した位置情報が第三者に閲覧され、プライバシーを侵害される危険性が高まっている。そこで本稿では、サービス提供者に位置情報を送信する際に、実際の位置情報と同時に架空の位置情報データを混入して送信することで、真のデータがどれであるかを判別することを困難にする手法について述べる。手法の有効性を示すためシミュレーションシステムを作成し、実際に GPS を用いて取得したデータを用いて評価実験を行った。この結果、提案手法がユーザの位置に関する匿名性確保に有効であることを確認した。

Location Anonymity of Moving Objects in Trajectory Database

Hidetoshi KIDO³ Yutaka YANAGISAWA⁴ Tetsuji SATOH^{3,4}

³Graduate School of Information Science and Technology, Osaka University

⁴NTT Communication Science Laboratories, NTT Corporation.

In this paper, we describe the technique to guarantee the location anonymity of people who uses location information services. Recently, we can use high accuracy positioning devices to obtain location information. There are various types of location information services such as GPS. On the other hands, because the location information of a person include deep personal information, security of the location information is one of the most significant problem in the location information systems. Therefore, we propose new methods to secure the location anonymity of people for these services. In our proposed methods, a client system generates a number of dummy position data, moreover the system sends these data with the true position data of the person to the server. Even if another person intercepts the data, the person can not distinguish the true position data from a number of sending data. For evaluation of our methods in an experiment using practical trajectory data, we develop a simulation system. Finally, we observe that our proposed methods secure location anonymity of people.

1 はじめに

近年、センシング技術の発展に伴い、GPS 端末や RFID タグといった位置情報を取得するデバイスが安価に入手できるようになった。それに伴い、移動する物体や人間の位置情報のデータを利用したさまざまなサービスが提供され始めている [4]。例えば、車でドライブに出かけていて昼食を食べたいと思ったときに、車の位置情報を送信して近くのレストランの情報を取得するサービスがある。他にも、バス停でバスを待っている間に携帯電話から位置情報を送信して、現在のバスの位置を取

得するサービスなどもあり、これらの一部はすでに実用化されている。

こうしたサービスが広まる一方で、ユーザが発信した位置情報が第三者に不当に閲覧される危険性が高まっている [1]。一般に、位置情報を利用したサービスを受けるためには、ユーザがサービス提供者に位置情報を送信する必要がある。サービスを受けるために一度送信した位置情報はサービス提供者側に管理が完全に委ねられ、ユーザ側で管理・変更することができない。そのため、サービス提供者のデータ管理の不備により位置情報が外部に漏洩してしまうと、これらのデータが第三者に閲覧

されてしまう。実際、位置情報データに限らずさまざまな個人情報サービス提供者から漏洩し、プライバシーを侵害されるという事件が頻繁に発生している。こうしたことから、位置情報の匿名性をユーザ自身が確保する手段の必要性が高まっている。

本稿では、サービス提供者に位置情報を送信する際に、実際の位置情報と同時に雑音（架空の位置情報データ）を大量に混入して送信することにより、真のデータがどれであるかを判別することを困難にする手法を提案する。提案法では、仮に位置情報が第三者に漏洩した場合でも、第三者が多数の架空のデータを含むデータ群の中から真の位置データを推測することをできないようにする。具体的には、時刻 t の位置データ分布をもとにして、架空の位置情報データが空間上に自然に分布するよう配置するアルゴリズムを用いる。

さらに本稿では、提案手法を評価するために作成したシミュレーションシステムについても述べる。実装したシステム上で、位置情報データを用いて評価実験を行った。評価用のデータとしては、実際に GPS を用いて取得した、奈良市内で営業する人力車のべ 39 台分の移動軌跡データを用いた。評価方法としては、匿名性を評価する指標である Anonymity Set (AS) を用いた。そして実験の結果、提案手法を用いることによりユーザの位置に関する匿名性が確保されることが確認された。また、架空の位置情報データの個数によって匿名性の度合いが変更されることも確認した。

以下、まず 2 章では位置情報の匿名性と AS についての詳細な説明を述べ、3 章で具体的な提案手法を説明する。4 章では実装した評価システムを利用した実験と結果について説明し、最後に 5 章でまとめる。

2 位置匿名性

匿名性が確保されているとは、自分が誰であるかを他人から特定できない状態にあることを示す。位置匿名性とは位置情報に関する匿名性で、位置匿名性が確保されているとは自分の位置が他人に知られていないことを示す。

本章では、対象とする位置情報サービスの例と位置情報に対する匿名性について説明し、位置匿名性が確保されていることを測るための指標値である Anonymity Set について説明する。

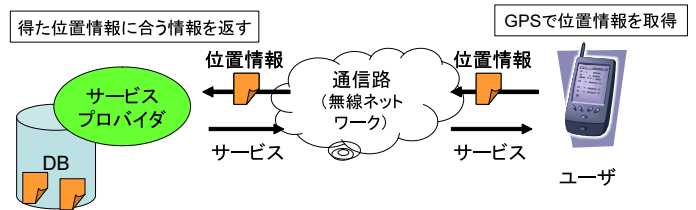


図 1: 位置情報サービスの例

2.1 位置情報サービスと位置匿名性

位置情報とは GPS などから取得した対象の現在存在する位置の座標のことである。この位置情報を利用した典型的なサービスの概念図を図 1 に示す。

このサービスは、サービスを受けるユーザ、位置情報を取得する GPS、位置情報をサービス提供者に送信する無線の通信路、そしてサービス提供者のデータベースで構成される。ユーザはまず、GPS から取得した位置情報のデータを通信路を通してサービス提供者に送る。そしてサービス提供者は送られた位置情報をデータベースの中のデータと照合し、提供するサービスや返送するデータ群を生成する。最後にサービス提供者はデータをユーザに送信してサービスが完了する。レストラン検索サービスの場合、提供するサービスは近隣のレストランの場所、開店時間、メニュー、値段などの情報を送信することである。したがって、サービス提供者はユーザから位置情報のみを取得し、前に示した情報を送信することになる。

このようなサービスにおいて、送信された位置情報はサービス提供者に管理が委ねられる。仮にサービス提供者の管理の不備によりデータが第三者に閲覧されると、ユーザのプライバシーが侵害される恐れがある。例として、バスの検索システムを考えてみる。あるユーザが病院に通院するために家と病院で毎回このサービスを利用するケースを想定する。するとデータベースに家と病院の位置情報が記録されることになる。もしユーザの情報が蓄積されて分析された場合、このユーザが通院する病院が特定される可能性が高い。ユーザの位置情報をそのままサービス提供者に送信してしまうと、このようにしてユーザの位置匿名性が損なわれる可能性がある。

2.2 Anonymity Set

文献 [3] によれば、Anonymity は「ある集合の中に存在して、その集合内で特定されない状態」と定義し、

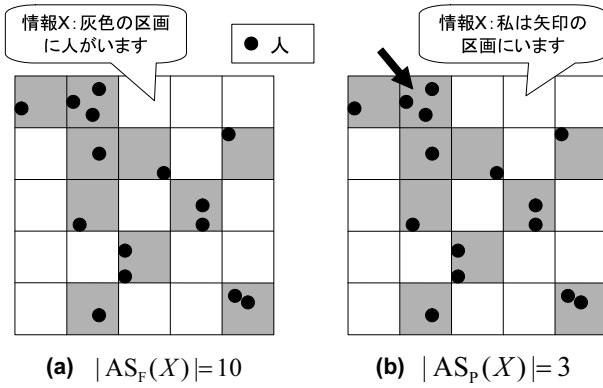


図 2: $AS(X)$ の例

Anonymity Set を「ある情報によって特定される対象の集合」と定義している．これに基づき本稿では、位置情報に関する Anonymity Set を「ある情報 X が与えられた時、 X によって特定される物体あるいは情報 $a_1, a_2, \dots, a_n$ の集合 A 」と定義し、これを $AS(X) = \{a_1, a_2, \dots, a_n\}$ と表記する．情報 X は「対象 a の属する集団を限定する情報」で、たとえば、「 a は関西人である」や「 a は 20 歳である」というような情報である．そこで、 $AS(X)$ は情報 X から集合 A への関数 ($AS: X \rightarrow A$) であり、 $|AS(X)|$ が集合の要素数となる．この $AS(X)$ からは、返す要素の種類によって次の 2 つの関数が考えられる．

- $AS_F(X)$
情報 X によって特定される場所 F の集合を与える関数である．つまり、その場所の面積の合計値を $|AS_F(X)|$ で表すものとする．例えば図 2 の (a) の場合「灰色の区画に人がいる」という情報によって求まるのは 1 つの区画の面積を 1 として $|AS_F(X)| = 10$ となる．
- $AS_P(X)$
情報 X によって特定される人 P の集合を与える関数である．つまり、その人数を $|AS_P(X)|$ で表すものとする．例えば図 2 の (b) の場合「矢印で示した区画に私はいる」という情報によって求まるのは $|AS_P(X)| = 3$ となる．

これらの指標はいずれも値が大きくなるほど、対象が限定されにくくなる．例えばあるオブジェクト a があって、 a に関する情報 X を第三者に知られたとき、その第三者が X から推測可能な a を含む集合を $AS(X)$ とすると、 $|AS(X)|$ が多ければ多いほど、 X という情報から a を特定することが困難になる．逆に、 $|AS(X)| = 1$

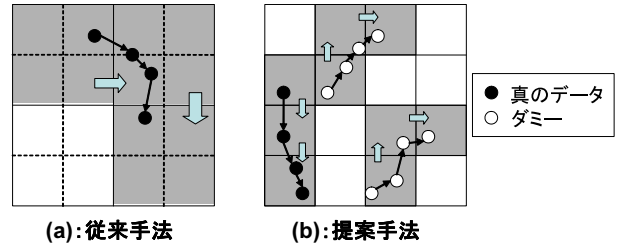


図 3: 従来手法と提案手法

であれば、 X によって a が特定されてしまう場合があるので、 X は a の匿名性を完全に損なう情報であるといえる．

3 実現手法

位置情報の匿名性を確保する従来の手法としては、位置情報の精度を下げるという手法がある [2][5]．この手法では、自分の位置情報を実際に GPS で得た情報よりも精度を下げて送信している．この手法の例を図 3 の (a) に示す．図 3 の (a) では、実際に送信される位置情報は「灰色の区画にいる」という情報である．このように位置情報の精度を落とすことで、自分の位置が特定されにくくなり、位置匿名性をある程度確保することができる．しかしこの手法では、時間が経過すると図 3 の (a) の通り対象の位置が大まかな軌跡となって現れるため、時間が経過するにしたがって動きが把握されやすいという問題点がある．本章では、まずこの問題点に対応した提案手法の概要を 3.1 で述べる．さらに、3.2、3.3 で具体的な手法を時間経過を考慮する場合としない場合に分けて説明する．

3.1 アプローチ

従来手法では、時間が経過するにしたがって動きが把握されやすくなっていくという問題点がある．これを改善するためには、自分と区別がつかない異なる位置情報データが存在すればよい．そこで提案手法では、サービス提供者に位置情報を送信する際に、実際の位置情報と同時に雑音（架空の位置情報データ、以下ダミーと記述する）を送信するという手法を用いる．手法について述べる前に、前提条件について説明する．

1. サービスを利用するユーザは無数に存在する．

2. ユーザが位置情報を複数サービス提供者に送る時、サービス提供者はそれらの複数の位置情報が同じユーザから送られてきたかどうかについては判定できない。
3. サービス提供者に送られた全ての位置情報は、誰でも閲覧することができる。
4. ユーザの位置情報は、サーバに送られる時点では x,y,z という精度の高い位置情報ではなく、「どの区画にいるか」という情報だけを送る。
5. ユーザが送った全ての位置情報に対して、サービス提供者はそれぞれサービスを提供する、ユーザはそこから、ユーザが真に必要なものだけを取り出す。つまりダミーを多く送れば送るほど、不必要なサービスを大量に受けることになる。

提案手法を適用した例を図3の(b)に示す。図3の(b)でも(a)と同様、自分の位置情報を灰色の区画で送信している。この手法では位置情報を従来手法より小さい区画で送信しているが、同時にダミーを2個発生させている。このダミーは真のデータとは異なる動きをするため、真のデータとダミーとの見分けがつきにくくなる。実際には他のユーザの位置情報、および他のユーザが発生させるダミーも同時に送信されることになるため、それらの位置情報と混ざり、さらに匿名性が高まることになる。

また、ダミーの個数を増やせば増やすほど位置匿名性が確保できるが、ダミーの個数を増やすことは送受信するデータ量が增大するというデメリットが存在する。具体的には、ダミーの個数を増やすことでサービス提供者に送信する位置情報が増え、遅延など他の問題が発生する可能性が増加する。また不必要なサービスを大量に受けることになるので、受信データが増大する上、不要なデータをフィルタリングする処理にも負担がかかることになる。このような点から、必要なサービスを受けることができる程度にダミーを発生させるなど、ダミーの個数とサービス効率のバランスが重要になる。このサービス効率はサービスの種類によって異なるので、サービスにあわせてダミーの個数を調整することが必要である。

3.2 時間経過を考えない場合

前述のレストラン検索の例のように、ユーザがサービスを利用するのは一時的で、断続的に利用することが考えにくい場合、時間経過を考える必要はない。そこで、

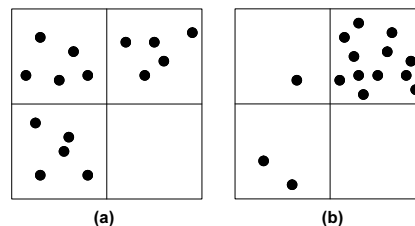


図 4: 位置情報の分布

時間経過を考えない場合において、位置匿名性を向上させるための満たすべき条件を以下に示す。

1. $|AS_F(X)|$ が大きくなるように配置する。つまり、多くの区画に人が存在するようにする。なぜなら、人のいる区画が増えれば増えるほど対象の特定が難しくなり、位置匿名性が向上するからである。
2. それぞれの区画についての $|AS_P(X)|$ もできるだけ大きくする。つまり、1つの区画に入る人の数を可能な限り多くする。一般に、ダミーが多いほど位置匿名性は向上するが、ダミーが多すぎる場合は前述の通り問題点があるため、適切な値に調整することが必要となる。
3. $|AS_P(X)|$ の全体での分散値を小さくする。つまり、全体の区画に分布している人の数を平均化する。 $|AS_P(X)|$ の分散の値が低い場合、ばらつきが小さいと考えられる。つまり、極端に少ない区画が存在する可能性が少ないといえ、位置匿名性が向上する。例を図4に示す。図4の(a)は(b)より $|AS_P(X)|$ の分散の値が低い。図4の(b)では(a)に比べて左上、左下の区画にいる数が極端に少ない。この場合左上、左下の区画では位置匿名性が十分に確保されているとはいえない。

3.3 時間経過を考える場合

時間経過を考える場合、ダミーが真のデータと比較して大きく異なる挙動をしないことも重要である。たとえば人が歩いている場合、人が一定時間で動ける範囲は限られてくる。もし動ける範囲を超えたダミーを発生させた場合、ダミーであると容易に判別されてしまう。

この問題を回避し位置匿名性を確保するためには、時刻ごとの各区画の $|AS_P(X)|$ の値の変化量を小さくする必要がある。もし各区画における人数が急激に変化する

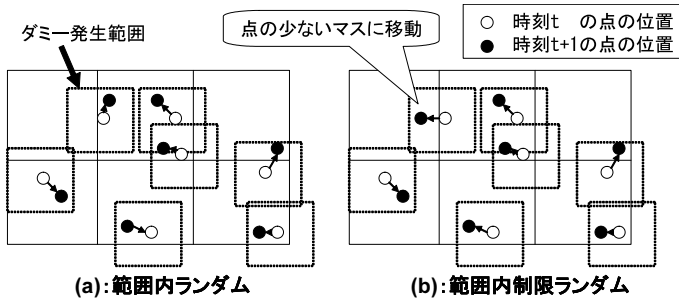


図 5: ダミー発生の 2 つの方法

と、ダミーが真のデータと比較して大きく挙動の異なる動きを行っている可能性が高い．するとダミーであることが判別されやすくなり，位置匿名性が確保できなくなる．時刻ごとの各区画の $|AS_P(X)|$ の変化量が小さいと位置匿名性が向上する．

そこで，位置匿名性を確保する次の 2 つの方法を提案する．

- 範囲内でランダム

図 5 の (a) のように，各ダミーごとに前回の自分の位置情報を記憶し，その位置から一定範囲内にランダムにダミーを発生させる．発生範囲を限定することでダミーが真のデータでは移りえない位置へ移動することを防ぐ．

- 範囲内で制限付きランダム

基本的には範囲内ランダムと同じだが，図 5 の (b) のように，今いる区画に多くの人が存在する場合に別の区画に移動する確率を上げ，人が分散するように発生させる．具体的には，ダミーを発生させる予定の区画の人数が全体の区画の人数の平均より多い場合は発生をやり直すという処理を有限回繰り返すことで実現する．

4 評価

3 章で示した提案手法を評価するために，シミュレーションシステムを作成し，実際の移動軌跡データを用いて実験した．本章ではその実験の概要と結果を示す．

4.1 実験の概要

まず，実装したシミュレーションシステムについての概要を示す．システムではまず，位置 (x 座標, y 座標) と時刻 t の組み合わせで構成された移動軌跡データを読み込み，ウィンドウに表示する．そして指定した時刻から指定したアルゴリズムでダミーを発生させ，各時刻ごとの $|AS_F(X)|$ や $|AS_P(X)|$ の値や時刻ごとの各区画の $|AS_P(X)|$ の変化量を計測し，表示する．ダミーの個数や発生させる時間，あるいは位置情報の粗さを示す区画の大きさなどは適宜変更できる．

データは実際に GPS を用いて取得した，奈良市内で営業する人力車のべ 39 台分の移動軌跡データを用いた．データは位置情報が経度，緯度で示されているため，システムに合わせて座標を変換して利用した．また，ダミーを発生させる個数は一定とした．つまり，あるユーザが 1 個発生させた場合，他のユーザも 1 個発生させるようにした．それぞれのユーザの位置情報は区画の単位で把握できる．しかし，この情報以外は自分の位置情報同士を関連付ける情報はないとした．つまり，位置以外の情報で追跡されることはない．

実験は時間経過を考える場合と考えない場合に分けて行った．時間経過を考えない場合，区画の数やダミーの発生個数を変えて実験を行い， $|AS_F(X)|$ や $|AS_P(X)|$ の値を計測した．時間経過を考える場合，ダミー発生アルゴリズムを変えて時刻ごとの各区画の $|AS_P(X)|$ の変化や $|AS_P(X)|$ の分散の値を計測した．

4.2 結果

図 6 はダミーの数を 1～10 個に変化させた場合の $|AS_F(X)|$ の値を全体の面積からの割合 (%) で示している．区画の数は 8×8 ， 10×10 ， 12×12 とした．ダミーの数が多くなるにつれて $|AS_F(X)|$ の値が大きくなるが，その度合いは $|AS_F(X)|$ の値が 100% に近づくにしたがって小さくなることが分かった．また， $|AS_F(X)|$ の値が 80% 付近になると位置情報のデータが全体に広がり，特定するのは困難になる．今回の実験では，ダミーの数が区画 8×8 では 3 個， 10×10 では 4 個， 12×12 では 6 個以上にすると $|AS_F(X)|$ の値が 80% を超えた．

一方，図 7 はダミー発生アルゴリズムごとの時刻ごとの各区画の $|AS_P(X)|$ の変化量を示している．この図では，提案した 2 つのアルゴリズムの他に，比較のために完全にランダムにダミーを発生させるアルゴリズムを用いた場合の結果についても合わせて示している．区画の

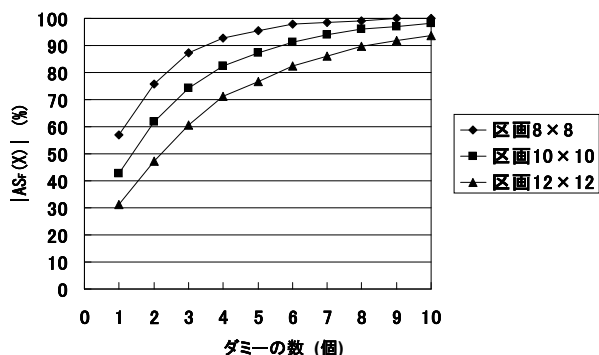


図 6: ダミーの数に対する $|AS_F(X)|$ の値

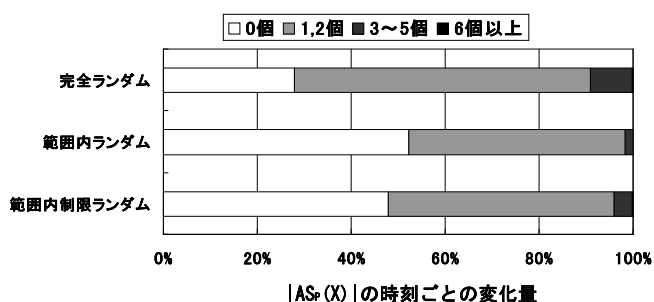


図 7: ダミー発生アルゴリズムに対する $|AS_P(X)|$ の時刻ごとの変化量

数は 10×10 , ダミーの発生個数は 3 個に固定した。提案したアルゴリズムは両方ともランダムのアルゴリズムよりも $|AS_P(X)|$ の時刻ごとの変化量が少なくなることが分かる。このことから、提案手法はランダムのアルゴリズムよりも位置匿名性が向上することが分かった。また、2 つの提案手法では範囲内ランダムのアルゴリズムがわずかに良好な結果となった。これは範囲内制限ランダムのアルゴリズムは点が分散する確率が高くなるので、区画を移動する機会が多くなったためだと考えられる。

表 1 はダミー発生アルゴリズムごとの $|AS_P(X)|$ の分散の値を示している。図 7 と同じ条件で計測を行った。提案したアルゴリズムは両方ともランダムのアルゴリズムよりも $|AS_P(X)|$ の分散の値が少なくなることが分かる。このことから提案手法はランダムのアルゴリズムよりも位置匿名性が向上することが分かった。一方、2 つの提案手法では範囲内制限ランダムのアルゴリズムの方が分散の値が小さくなった。これはアルゴリズムが想定どおりに働いていることを示している。

表 1: ダミー発生アルゴリズムに対する $|AS_P(X)|$ の分散

	ランダム	範囲内ランダム	範囲内制限ランダム
$ AS_P(X) $ の分散	2.43	2.26	2.08

5 まとめ

本論文では、サービス提供者に位置情報を送信する際に、実際の位置情報と同時に架空の位置情報データを送信することにより、位置情報が漏洩した場合でもプライバシーの侵害を困難にする手法について提案した。実際にフィールドで取得した移動軌跡データを用いて提案手法の評価実験を行った。匿名性評価の指標として知られる Anonymity Set の考え方を位置情報に拡張して評価した結果、提案手法が位置情報の匿名性向上に有効であることが確認できた。

今後の課題としては、より前提条件を厳しくした場合の検討がある。本稿では自分の位置情報同士を関連付ける情報はないと仮定して評価した結果について述べたが、追跡可能な場合では別の手法が必要になる。また、他人が悪意をもってダミーを発生させる可能性のある場合についても同様に対策を検討する必要がある。

謝辞

本研究を進めるにあたって多大なご協力を頂いた大阪大学大学院情報科学研究科の村田正幸教授には、この場を借りて厚く御礼申し上げます。

参考文献

- [1] M. Ackerman, T. Darrell, and D. J. Weitzner. Privacy in context. In *Human-Computer Interaction*, Vol. 16, pp. 167–176, 2001.
- [2] Marco Gruteser and Dirk Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the First International Conference on Mobile Systems, Applications, and Services*, pp. 31–42, 2003.
- [3] Andreas Pfitzmann and Marit Koehnopp. Anonymity, unobservability, and pseudonymity: A proposal for terminology. Draft, version 0.17, July 2000.
- [4] Ouri Wolfson, Prasad Sistla, Bo Xu, Jutai Zhou, and Sam Chamberlain. DOMINO: Databases fOr MovINg Objects tracking. In *SIGMOD'99 Conference Proceedings*, pp. 547–549, 1999.
- [5] 中西健一, 高汐一紀, 徳田英幸. 粒度の動的変更による位置匿名性についての考察. マルチメディア, 分散, 協調とモバイル (DICOMO2004) シンポジウム, 2004.