位置情報サービスのための 架空情報を用いた位置プライバシー保護手法とそのコスト評価

貴戸　秀年† 　柳沢　豊†† 　佐藤　哲司†,††

† 大阪大学 大学院情報科学研究科　〒 565-0871 吹田市山田丘 1–5
†† 日本電信電話会社 NTT コミュニケーション科学基礎研究所　〒 619-0237 京都府相楽郡精華町光台 2–4
E-mail: †h-kido@ist.osaka-u.ac.jp, ††yutaka@cslab.kecl.ntt.co.jp, †,††satoh.tetsuji@lab.ntt.co.jp

あらまし　近年，GPS 端末などの位置情報を取得するデバイスの普及に伴い，人間の位置情報を利用したさまざまなサービスが提供されつつある．こうしたサービスは便利な反面，位置データ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．そこで筆者らは，位置情報に関するプライバシーを保護する手法として，実際の位置データと同時に架空のデータを混入して送信する手法をこれまでに提案している．本稿ではさらに，この手法を用いたサービスについて，架空の位置データにかかる送受信コストを削減する方法を提案する．実際に運用されているサービスを用いて行った実験の結果，コスト削減手法を用いることによって現実的なサービスが実現可能であることを確認した．

キーワード　時空間 DB，セキュリティ，情報検索，プライバシー，匿名性

A Method to Protect Location Privacy using Dummies and Its Cost Evaluation for Location-based Services

Hidetoshi KIDO†, Yutaka YANAGISAWA††, and Tetsuji SATOH†,††

† Graduate School of Information Science and Technology, Osaka University
1–5 Yamadaoka, Suita, Osaka, 565-0871 Japan
†† NTT Communication Science Laboratories, NTT Corporation
2–4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237 Japan
E-mail: †h-kido@ist.osaka-u.ac.jp, ††yutaka@cslab.kecl.ntt.co.jp, †,††satoh.tetsuji@lab.ntt.co.jp

Abstract　Recently, high accuracy positioning devices enable us to provide various types of location-based services. In these services, the protection of location privacy is one of the most important issues. For protecting our location privacy on such services, we have proposed a communication method for such services based on the idea of the dummy users. In our proposed communication methods, a service client device generates several dummy position data sent to the server with true position data of the user. In this paper, we present a method to reduce the cost for communication between a client and a server. Moreover, we describe the experiment of performance study about our proposed reducing method. As a result of the experiment, we concluded that our method can reduce the communication cost enough to apply our communication method into practical location-based services.

Key words　Spatial and Temporal Database, Security, Information Retrieval, Privacy, Anonymity

1. はじめに

近年，センシング技術の発展に伴い，GPS 端末 [4] 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を利用したさまざまな位置情報サービスが提供され始めている [10]. 例えば，車でドライブに出かけていて昼食をとり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位置情報サービスなどがある．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．

こうした位置情報サービスが普及する一方で，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，位置情報をはじめとしたさまざまな個人情報がサービス提供者から漏洩し，プライバシーが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，位置情報に関するプライバシー，すなわち位置プライバシーをユーザ自身が保護する手段の必要性が高まっている．

そこで筆者らは，サービス提供者に位置情報を送信する際に，送信メッセージに架空の位置情報を大量に混入して送信することにより，真の位置情報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した [12]．架空の位置情報を真の位置情報を混合してサービス提供者に送信した場合，仮にこれらのデータが第三者に漏洩した場合でも，第三者が多数の架空の位置情報を含むデータ群の中から真の位置情報を見つけることは困難である．こうした利点を持つ反面，提案手法では，大量の架空の位置情報が混入されることで送受信にかかるコストが増大するという問題点があった．

そこで本稿では，筆者らが提案した架空情報を用いた位置プライバシー保護手法を適応したサービスにおいて，送受信にかかるコストを削減する手法を提案する．まず送信メッセージに対しては，位置情報の座標の組（x,y）を x の集合，y の集合に分けて送信し，x, y の全組み合わせを全位置情報とすることでコスト削減を図る．一方，返信メッセージに対しては，ユーザから取得した位置情報と無関係な情報を用いてフィルタリングし，返信データを限定してコスト削減を図る．また，これらの提案手法を実際に GeoLink Kyoto [7] に適応して行った送受信コストに関する評価実験についても述べる．実験の結果，提案手法を適応することで送信，受信ともにコストが減少し，従来よりも低い送受信コストで位置プライバシーを保護できるサービスが実現できることが確認された．

以下，2 章では位置プライバシーの概要と匿名性を示す指標である Anonymity Set についてそれぞれ詳細に説明する．3 章では架空情報を用いた位置プライバシー保護手法を説明する．そして，4 章で提案手法の概要を説明し，5 章で提案手法の評価実験と結果について述べる．最後に 6 章でまとめる．

2. 位置プライバシーと Anonymity Set

A.Beresford らは，位置プライバシー（location privacy）を「第三者によってある人の現在，もしくは過去にいた位置を知られることを防ぐ能力」と定義している [2]．また，位置情報を取得できるシステムによって，位置プライバシーを侵害される可能性が非常に高くなったとも述べている．

本章では，本研究においてキーワードとなる位置プライバシーと Anonymity Set について，それぞれ詳細に説明する．

2.1 位置情報サービスと位置プライバシー

位置情報とは GPS などから取得した対象が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サービスのことを，一般に位置情報サービスと呼ぶ．位置情報サービスを提供するシステムの概念図を図 1 に示す．

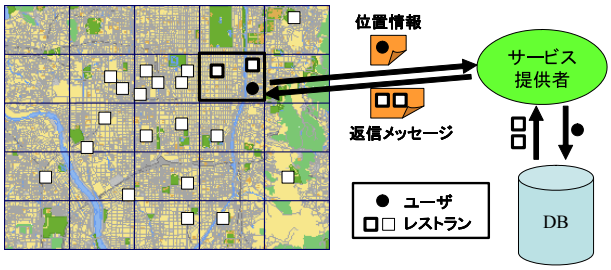

図 1　位置情報サービスの例
Fig. 1　Example of a location-based service

位置情報サービスを利用するユーザは，GPS から取得した位置情報を含む送信メッセージをサービス提供者に送信し，その位置データに応じたサービスを受け取る．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，ユーザが変更・削除することはできない．送信メッセージには位置情報とネットワークアドレスが必ず含まれる．この内，ネットワークアドレスは明らかに対象を特定する情報であるため，これを隠蔽する方法として，Crowds [9] や Onion Routing [5] といったさまざまな研究がなされている．

一方，位置情報にもプライバシーに関わる情報が含まれている．例として，バスの検索システムを挙げる．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．この場合，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まる．これは位置情報によって位置プライバシーが侵害されることを示している．

つまり，位置プライバシーを保護するためには，「ユーザがどこにいるか分からなくする」こと，すなわち位置の匿名性を高めることが必要となる．この位置に関する匿名性を本稿では位置匿名性と呼ぶ．

2.2 位置匿名性を高める性質

位置匿名性の高さを計測する最も簡単な方法は人が位置情報の分布を実際に見て，ユーザの真の位置が分かるかどうかを判断する方法である．しかしこの方法は主観的であるため評価尺度としては適さない．客観的な評価指標を定めるために，まず位置匿名性を高める 3 つの性質に着目する．

- 遍在性——人がいたるところにいる

図 2 の (b) のように人が広がって分布している場合，その中の 1 人を特定しようとしても候補エリアが非常に広く，特定することは非常に困難である．逆に，図 2 の (a) のように人がある程度の範囲に固まっている場合，人のいる区画が少なくなるため比較的特定しやすくなる．つまり，人がいる区画が多くなると位置匿名性は高まる．

- 稠密性——一定の範囲内に人がたくさんいる

図 2 (b) の (1) と (2) の区画を比べた場合，(1) は万一その区画

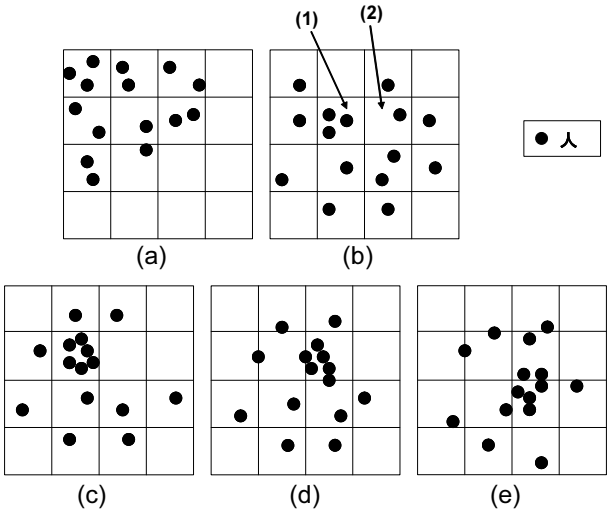

図 2　位置情報の分布の例

Fig. 2　Example of distribution of position information

内にいるという情報が知られたとしても，その候補が 3 人おり，特定には至らない．逆に (2) は候補が 1 人となり特定されてしまう．つまり，同じ位置情報を示す区画内に多くの人がいると位置匿名性は高まる．

- 一様性——人が一様に分布している

図 2 (c) のように人の多い区画と少ない区画が混在する場合，稠密性を考えると人の少ない区画の位置匿名性は人の多い区画よりも低いといえる．また時間が経過とともに人の位置が図 2 (c)，(d)，(e) のように変化した場合，人の多い区画の動きを追跡されやすくなる．つまり，人が偏りなく一様に分布しているときが最も位置匿名性が高まる．

2.3　Anonymity Set と $AS(i)$

2.2 節で述べた位置匿名性を高める性質を数値で評価するために，Anonymity Set という概念を導入する．Anonymity Set は D.Chaum が提唱した，匿名性を評価する指標値である. [3]. A.Pfitzmann らは Anonymity Set の定義を明確化し,「情報によって特定されるそれぞれが区別できない要素全体」とした [8].

筆者らはこの Anonymity Set の定義を位置情報に関して拡張し，この要素全体の集合を返す関数 $AS(i)$ を定義する．まず，いくつかの記号を定義する．

a　：　対象
A　：　対象の集合，$A = \{a_1, a_2, \ldots, a_n\}$
i　：　A を限定する情報
I　：　i の集合
$|A|$　：　A の要素数
$\hat{A}$　：　A のべき集合　(2^A)

情報 i は任意の対象の集合 $\hat{A}$ の属する集団を限定する情報を示す “文章” を表している．例えば日本人全体を全体集合とした場合，i に「$\hat{A}$ は関西人である」を与えると日本人全体が関西人全体の集合に限定される．また i に,「$\hat{A}$ は 20 歳である」を代入すると 20 歳の日本人全体の集合に限定される．

これらの記号を基に，関数 $AS(i)$ を以下のように定義する．

$$AS(i) = 2^A = \hat{A} \quad (AS : I \to \hat{A})$$
$$|AS(i)| = |\hat{A}|$$

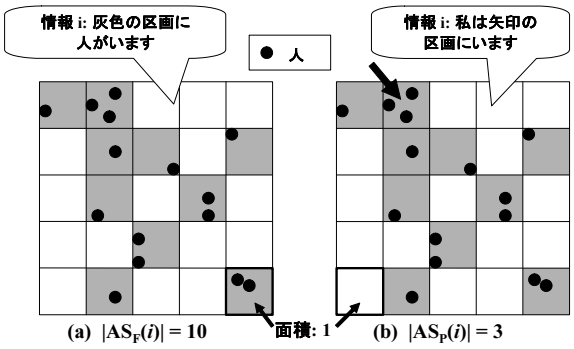

図 3　$AS(i)$ の例

Fig. 3　Example of $AS(i)$

次に，この関数 $AS(i)$ を指標として用いるため，返す要素の種類によって次の 2 つの関数に拡張する．

- $AS_F(i)$

情報 i によって特定される場所の集合 α_F を返す関数である．α_F は i によって限定された区画の集合を示す．$|AS_F(i)|$ は α_F の要素数，つまり α_F の面積の合計値を表す．$AS_F(i)$ の定義を以下に示す．r_j は区画を示す．

$$\begin{aligned} A_F &= \{r_1, r_2, \ldots, r_m\} \subset A \quad (\forall r_j \in A_F) \\ |A_F| &= |\{r_1, r_2, \ldots, r_m\}| = m \\ AS_F(i) &= \alpha_F (\in \hat{A}_F) \quad (AS_F : I \to \hat{A}_F) \\ |AS_F(i)| &= |\alpha_F| \end{aligned}$$

例えば図 3 の (a) の場合,「灰色の区画に人がいる」という情報 i によって求まるのは，1 つの区画の面積を 1 とする場合，$|AS_F(i)| = 10$ である．

- $AS_P(i)$

情報 i によって特定される人の集合 α_P を返す関数である．α_P は i によって限定された人の集合を示す．$|AS_P(i)|$ は α_P の要素数，つまり α_P の人数を表す．$AS_P(i)$ の定義を以下に示す．p_j は人を示す．

$$\begin{aligned} A_P &= \{p_1, p_2, \ldots, p_m\} \subset A \quad (\forall p_j \in A_P) \\ |A_P| &= |\{p_1, p_2, \ldots, p_m\}| = m \\ AS_P(i) &= \alpha_P (\in \hat{A}_P) \quad (AS_P : I \to \hat{A}_P) \\ |AS_P(i)| &= |\alpha_P| \end{aligned}$$

例えば図 3 の (b) の場合,「矢印で示した区画に私はいる」という情報 i によって求まるのは $|AS_P(i)| = 3$ となる．

これらの関数を位置匿名性の評価に利用する．$AS(i)$ を適応する全体集合を位置情報を取得できるエリア全体とし，そのエリアを適当な区画に分割する．そして，位置情報は各区画レベルの精度で取得できるとする．情報 i を各位置データとしたときの $|AS_F(i)|$ の値を **F** とすると，**F** は人がいる区画の面積の合計値を示す．また，情報 i を特定の区画を示す情報としたときの $|AS_P(i)|$ の値を **P** とすると，**P** は区画内にいる人数を示す．この **F**，**P** を用いて，2.2 節で示した位置匿名性を高める性質は次のように表現できる．

- 遍在性——**F**

F は人がいる区画の面積の合計値を示すため，遍在性と対応す

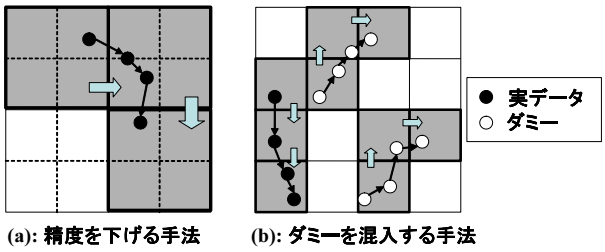

図 4　各位置プライバシー保護手法の特徴

Fig. 4　Feature of two techniques to protect location privacy

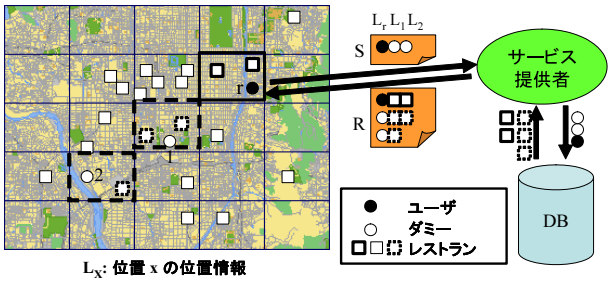

図 5　提案手法を用いた匿名位置情報サービスの例

Fig. 5　An anonymous location-based service using a proposal technique

る．つまり，**F** の値が大きくなると，位置匿名性が高まる．図 2 の (a) と (b) を比べると，(a) は $\mathbf{F} = 8$ で (b) は $\mathbf{F} = 12$ となるので，(b) のほうが位置匿名性が高い．

- 稠密性——**P**

P はその区画内にいる人数を示すため，稠密性と対応する．つまり，**P** の値が大きくなると，位置匿名性が高まる．図 2 (b) の (1) と (2) の区画を比べると (1) は $\mathbf{P} = 3$ で (2) は $\mathbf{P} = 1$ となるので，(1) のほうが位置匿名性が高い．

- 一様性——**P** の分散の値（$\mathrm{Var}(\mathbf{P})$ と表記する）

P の分散の値が小さくなると，各区画の **P** の値のばらつきが小さくなる．これは一様性と対応する．図 2 の (b) と (c) を比べると，(b) は $\mathrm{Var}(\mathbf{P}) = 143/256$ で (c) は $\mathrm{Var}(\mathbf{P}) = 495/256$ となるので，(b) のほうが位置匿名性が高い．

3.　架空情報を用いた位置プライバシー保護手法

位置プライバシーを保護する手法の 1 つとして，位置情報の精度を下げる手法がある [6], [11]．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する．例を図 4 の (a) に示す．図 4 の (a) では，図で点線で示している最も小さい区画の位置情報を入手しているにもかかわらず，実際に送信される位置データは「灰色の区画にいる」という情報のみ保持している．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高めることができる．しかしこの手法では，時間が経過すると図 4 の (a) の通り対象の位置が大まかな軌跡となって現れる．そのため，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．そこで，筆者らはこの問題点に対応した架空情報を用いた位置プライバシー保護手法を提案した [12]．本章ではこの手法の概要を 3.1 節で述べる．そして，3.2 節で追跡可能性に対する対策について述べる．

3.1　手法の概要

筆者らは文献 [12] の中で，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（以下ダミーと記述する）を送信する手法を提案した．この手法を用いた位置情報サービスの例を図 5 に示す．

図 5 では，ユーザは GPS を搭載した無線通信デバイスを所持している．ユーザはまず，GPS を用いて自分の今いる位置 r を取得する．次に，このデバイスが位置 r に応じて位置 1，2 を示すダミーを発生させ，送信メッセージ **S** を作成する．そして **S** をサービス提供者に送信する．サービス提供者はメッセージ **S** を受け取った場合，位置データ $\mathrm{L_R}, \mathrm{L}_1, \mathrm{L}_2$ に対応するサービスのデータをデータベースから取り出し，それぞれの位置データに結合させる．このときサービス提供者はどの位置データが実際の位置を示しているかは判断できない．このあと，返信メッセージ **R** を作成し，ユーザに送信する．最後に，ユーザは自身で覚えていた実際の位置 r に対応するデータのみを取得してサービスを完了する．

この手法では，サービス提供者は実際の位置情報とダミーを区別することができない．つまり，ユーザは自分の実際の位置を知っているが，サービス提供者は実際の位置の候補しか知ることができない．したがって，ユーザはサービス提供者に位置を知られずにサービスを受けることができる．

架空情報を用いた位置プライバシー保護手法を適用した例を図 4 の (b) に示す．図 4 の (b) でも (a) と同様，自分の位置データを灰色の区画で送信している．提案手法では取得した位置データとダミー 2 個を従来手法より小さい区画で送信している．発生させたダミーは実データとは異なる動きをするため，たとえ動きが把握されたとしてもそのデータが真であるかどうかは判断することができない．また，実際には他のユーザの位置データ，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，それらの位置データと混ざりあって，さらに匿名性が高まることになる．

3.2　追跡可能性とダミー発生アルゴリズム

追跡可能性とは時間間隔の短い複数の位置情報を連結して見ることによって対象となる物体の軌跡が判断できてしまう性質のことである．例えば，道案内サービスのように継続的に位置データを送る必要があるサービスを受ける場合，時間経過による追跡可能性を防ぐ必要がある．仮に人が徒歩で移動している場合，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．これを回避するためには，ダミーが実データと比較して大きく異なる挙動をしないことが重要となる．このことを実現し，位置匿名性を下げないための 2 つのアルゴリズムを説明する．

- **範囲内移動**

図 6 の (a) に示すとおり，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実データでは

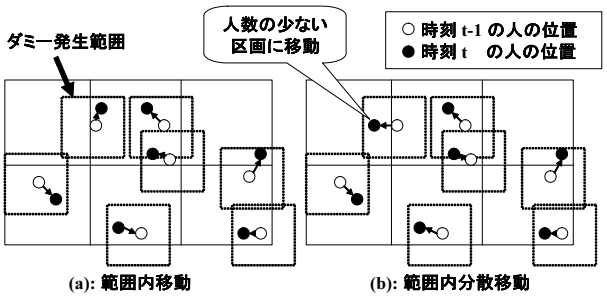

図 6　2 つのダミー発生アルゴリズム
Fig. 6　Two dummy generation algorithms

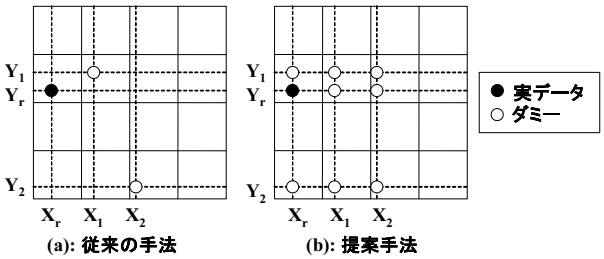

図 7　送信メッセージコスト削減手法
Fig. 7　A cost reduction technique for required message

移りえない位置へ移動することを防ぐ．この方法でダミーが実データと比較して不自然な挙動をしないことが保証できる．表 1 にアルゴリズムの概略を示す．

- **範囲内分散移動**

このアルゴリズムは他人のデータを参照できるという条件の元で適用できる．図 6 の (b) に示すとおり，範囲内移動アルゴリズムに加え，発生させようとす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，一様性を高めるように発生させる．具体的には，ダミーを発生させようとする予定の区画の **P** の値が全区画の **P** の値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．この方法では，範囲内移動アルゴリズムで保証される事項に加え，$\mathrm{Var}(\mathbf{P})$ の値を減少させることができる．表 2 にアルゴリズムの概略を示す．

ここで示した 2 つのアルゴリズムを評価するための評価指標として，時刻ごとの各区画の **P** の値の変化量を定義する．これを $\mathrm{Shift}(\mathbf{P})$ と表記する．仮にある区画で人数が急激に変化したとすると，ダミーが実データと比較して不自然な挙動をしている可能性が高い．このことはダミーが位置匿名性と高める役割を果たしていないことを示す．つまり，$\mathrm{Shift}(\mathbf{P})$ の値が小さい方が位置匿名性が高いといえる．

4. 送受信メッセージのコスト削減手法

3 章で述べた架空情報を用いた位置プライバシー保護手法では，ダミー発生個数が増加すると位置匿名性が増加するが，同時に通信コストも増大する．これを回避するために，架空情報を用いた位置プライバシー保護手法に対応した送受信メッセージのコスト削減手法を提案する．本章ではまず，架空情報を用いた位置プライバシー保護手法に関するコストについて 4.1 節で説明する．そのあと，コスト削減手法を 4.2，4.3 節でクライアントからサーバへの送信とサーバからのデータの受信に分けて提案する．

4.1 架空情報を用いた位置プライバシー保護手法に関するコスト

A.Beresford らは，文献 [2] の中でダミーユーザを混入する手法はリソースを大きく使いすぎる，ということについて言及している．文献 [2] で述べられている問題は，筆者らの提案する手法でも発生しており，プライバシーを保護する効果の副作用としてダミーを送信することによる通信コストの増大が発生する．このことは提案手法の実現に関する障害となる可能性がある．一方，ユーザ側でのダミー発生処理やサービス送信者のサービス発行処理にも処理コストが発生する．ただしこれらの処理にかかるコストはハードウェアの高性能化によって無視できるほど小さいと考えられるため，本稿では考えない．

4.2 送信メッセージコスト削減手法

送信メッセージ **S** は真の位置情報，ダミー，後述する位置情報と無関係なサービスの限定情報，およびその他サービスに必要な情報で構成される．位置 x に関する位置データ L_x は各位置の x 座標（X_x），y 座標（Y_x）の組で構成される．**S** のうち，位置情報以外の情報はダミーの発生とは無関係である．ダミーを発生させた場合，位置情報の個数が多くなり，送信メッセージのコストが増加する．以前の手法では，ダミーが 2 個の場合の **S** の位置情報群は次のように構成される．

$$(X_r, Y_r), (X_1, Y_1), (X_2, Y_2)$$

X_r, Y_r が真の位置データ，$X_i, Y_i (i = 1, 2, 3, \ldots)$ がダミーの位置データを示す．この場合，図 7 の (a) に示すようにダミーが発生する．この方法では，ダミーの数が n に対して位置情報の送信にかかるコストは $O(n)$ である．これは莫大な個数のダミーを発生させると送信メッセージのコストもそれに比例して非常に大きくなる．

そこで，新たな手法を提案する．提案手法では **S** の位置情報群を次のように構成する．

$$(X_r, X_1, X_2), (Y_r, Y_1, Y_2)$$

提案手法では位置データを x 座標の集合・y 座標の集合に分けて送信する．各集合には真の位置データを含める．サービス提供者は x 座標・y 座標それぞれの情報の全組み合わせを位置データとして認識し，サービスを提供する．この場合，図 7 の (b) のようにダミーが発生することになる．この手法を用いた場合，n 組の位置データの集合のコストで，真の位置データを含む n^2 個の位置データを送信することができる．つまり，位置データにかかるコストは $O(\log n)$ に抑えられ，従来の手法に比べると大幅に改善される．

4.3 返信メッセージコスト削減手法

返信メッセージ **R** は送信された各位置データ L_j とそれに応じたサービスの内容 D_j で構成される．ダミーが 2 個の場合の **R** の例を以下に示す．

表 1　アルゴリズム：範囲内移動

Table 1　Moving in a Neighborhood (MN) algorithm

```
// Input:  時刻 t-1 のダミーの位置
// Output:  時刻 t のダミーの位置
// random(x,y):  x と y の間の乱数を発生させる関数

struct dummy {
        double x;          // x 座標
        double y;          // y 座標
        double t;          // 時刻
};
void RangeLimit (double m, int n) {
       struct dummy prev[100], next[100];

       (Input の内容を prev[] に代入);
       for (i=1;i<n;i++) {
              next[i]->x = random( (prev[i]->x)-m, (prev[i]->x)+m);
              next[i]->y = random( (prev[i]->y)-m, (prev[i]->y)+m);
              next[i]->t = (prev[i]->t)++;
       }
       (next[] の内容を出力);
}
```

表 2　アルゴリズム：範囲内分散移動

Table 2　Moving in a Limited Neighborhood (MLN) algorithm

```
// Input:  時刻 t-1 のダミーの位置
// Output:  時刻 t のダミーの位置
// random(x,y):  x と y の間の乱数を発生させる関数
// position(x,y):  時刻 t-1 の時の (x,y) を含む区画の位置データの個数を返す関数

struct dummy { (表 1 で定義) };
void RangeNumberLimit (int aveP, double m, int n) {
       struct dummy prev[100], next[100];
       int k = 0;

       (Input の内容を prev[] に代入);
       for (i=1;i<n;i++) {
              next[i]->x = random( (prev[i]->x)-m, (prev[i]->x)+m);
              next[i]->y = random( (prev[i]->y)-m, (prev[i]->y)+m);
              next[i]->t = (prev[i]->t)++;
              if (position(next[i]->x, next[y]->y) > aveP) {
                     if (k<=3) { k++; continue; } else { k=0; }
              }
       }
       (next[] の内容を出力);
}
```

$$\mathbf{R}= ((L_r, D_r), (L_1, D_1), (L_2, D_2))$$

サービスの内容 D_j は名前と位置情報サービスの種類に応じた属性の情報で構成される．一般的な D_j の構造を以下に示す．

$$(name_1, URL_1, address_1), (name_2, URL_2, address_2), \ldots$$

各位置データが持つデータ $(name_j, URL_j, address_j)$ の組の個数はそれぞれ異なる．返信メッセージに含まれるこのデータの組が多すぎる場合，サービス提供者はユーザにデータの個数を減少させるための情報を要求することができる．

この条件において，返信メッセージのコストを減少させる手法について述べる．各位置データごとにサービスのデータが結び付けられているため，位置データにかかるコストを減少することはできない．また，サービス提供者側で不要な情報を省くことはダミーを判別するための手がかりとなるため，この方法は用いることができない．そこで，位置情報と無関係な情報を用いて受信するデータを限定し，コスト削減を図る．

以下に，コストを減少させる 4 つの手法を提案する．

- 範囲の限定

送信する位置情報の精度を上げることで位置情報に一致するサービスの対象を限定する．近隣レストラン検索サービスではこの方法によって返信されるレストランの個数は限定される．ただし，位置匿名性が減少してしまうことが問題点として挙げられる．

- ジャンルの限定

検索サービスの多くは名前や種類などを限定するために，キーワードによってジャンルごとに階層化されていることが多い．こうしたサービスにおいて，万一返信にかかるコストが多すぎる場合，ユーザにデータのジャンルを限定してもらうように指示する．ユーザがジャンルを限定することで，検索で一致するデータの個数を削減することができる．多くのジャンルの情報をもつデータベースを利用したサービスに効果が高いといえる．

- キーワードの指定

位置情報サービスのほとんどはサービスに文字情報を含んでいるか，あるいは検索ワードとして含ませることができる．こうしたサービスにおいて，万一返信にかかるコストが多すぎる場合，ユーザにフリーワードでキーワードを指定させることで，検索で一致するデータの個数を削減することができる．

- 不要データの削除

サービスによっては，ユーザが不要なデータを受け取ってしまうことはよくある．説明のために，ユーザの近隣にあるレストランを検索するサービスを利用する例を挙げる．雑誌などで知った店に行こうとして付近まで来た時にこのサービスを利用したとする．この場合はユーザはその店の正確な位置のみを知りたいため，店の開店時間やメニューの情報は不要となる．このような場合にユーザに不要な情報を指定させ，必要な情報のみを送信することで返信にかかるコストを削減することができる．

これらの手法ではジャンルやキーワードなど位置情報と無関係な情報をサービス提供者に送信する必要がある．そこで，送信メッセージにこれらの情報を付加する．これらの情報はダミーの数に関わらず一定で，ごく小さい情報なのでコストへの影響は無視できるほど小さい．

5. 評 価 実 験

4 章で示したコスト削減手法を評価するために，GeoLink Kyoto [7] システムのデータベースを用いて評価実験を行った．本章ではその実験の概要と結果を示す．

5.1 実験の概要

架空情報を用いた位置プライバシー保護手法のコストを計測するために以下の評価を行った．

（1） 送信メッセージのコスト削減手法の評価

（2） 返信メッセージのコスト削減手法の評価

各評価を行うために GeoLink Kyoto [7] と呼ばれる，京都市内にあるさまざまなスポットの Web ページを紹介するサービスを用いた．GeoLink サービスの外観を図 8 に示す．GeoLink サービスはラジオボタンで指定したジャンルのスポットを，地図上でクリックした位置から直線距離が近い順に 30 個表示する．このサービスのデータベースには，各スポットの id，名称，Web ページ URL，位置（北緯，東経），住所，ジャンル，備考のそれぞれの情報が格納されている．

評価内容を説明する前に，送信メッセージ **S**，返信メッセージ **R** のコストの基準となる前提条件を説明する．**S** は 4.2 節で示したとおり，位置データ群，サービスの限定情報，およびその他の情報で構成される．位置データ群は実数で，ひとつの数値が 32[bit] で表現されるとする．その他の情報はダミーを発生させても変化しないため一定の値とし，16[Byte] とする．**R** は 4.3 節で示したとおり，送信された各位置データとそれに応じたサービスのデータで構成される．ここでのサービスのデータの内容は送信された位置データの範囲内のすべてのスポットに対するデータベースの全項目とする．1 スポットあたりの総データ量の平均は 121.964[Byte] であった．そこで，位置データやヘッダなどを含めて 128[Byte/1 スポット] と仮定する．送信される位置情報は半径 2km の円の範囲を示すとする．この時，1 個の位置データに対するサービスの個数の平均は 114.47 個，最大は 1,067 個であった．よって，位置データ 1 個に対する平均の返信コストは，

表 3　送信メッセージのコスト比較

Table 3　Cost comparison about required messages

送信位置データ数	メッセージサイズ [Byte]	認識位置データ数	
		従来の手法	提案手法
1	24	1	1
2	32	2	4
3	40	3	9
4	48	4	16
5	56	5	25
10	96	10	100
100	816	100	10000

$$128 \times 114.47 = 14,652.16[\mathrm{Byte}] \simeq 14.7[\mathrm{KByte}] \quad (1)$$

であり，また最大の返信コストは，

$$128 \times 1,067 = 136,576[\mathrm{Byte}] \simeq 137[\mathrm{KByte}] \quad (2)$$

となる．この値はダミーの数が少ない場合は問題ないが，ダミーの数が多くなり，同時接続数が増加した場合，レスポンスに時間がかかる可能性が生じる．

送信メッセージのコスト削減手法の評価では，4.2 節で示した手法の実際の効果を先ほどの規定に従い計算した．返信メッセージのコスト削減手法の評価では，4.3 節で示した手法の実際の効果を GeoLink Kyoto [7] サービスを用いて評価する．評価のために，GeoLink サービスの全データを PostgreSQL のデータベースに格納した．SQL を用いて擬似位置データを作成し，データベースにクエリとして送信することでデータを計測した．実験では，指定した位置から指定した半径以内にあるスポットデータの数を 100 回計測し，平均を取った．また，その際にジャンル，キーワード，位置を指定する個数を SQL を用いて指定した．

5.2 送信メッセージのコストの評価

表 3 は提案手法を用いた場合と用いない場合の送信メッセージにかかるコストをまとめたものである．送信位置データ数は北緯，東経の組の個数を示す．表 3 の通り，位置データの個数がが 1 つ増えるたびにメッセージコストが 8[Byte] づつ増加する．一方，認識位置データ数はサービス提供者が受け取った位置データの個数を示す．提案手法は送信位置データの 2 乗の位置データをサービス提供者が認識することが分かる．このように，提案手法は $O(\log n)$ のコストでダミーを発生させることができることが確認できた．また，10,000 個の位置データを含む送信メッセージでも 816[Byte] とコストも低いので，十分な実用に耐えうる．

5.3 返信メッセージのコストの評価

図 9 はコスト削減手法ごとの 1 サービスにかかる総データ量の平均を示している．x 軸がダミーの数を示し，1 〜 10 個で計測した．制限なしの場合のデータ指定半径は人が歩ける範囲として 2[km] とし，実データとして京都府立体育館付近 2km（半径を制限した場合は 1[km]）のデータが加えてある．グラフの凡例は `no limit` が制限なし，`radius 1/2` はデータ指定半径が 1[km]，`category limit` がジャンルが「食事」，`keyword` はキーワードとして「マクドナルド」を指定，そして `remove URL` は URL データを取り除いたものをそれぞれ示

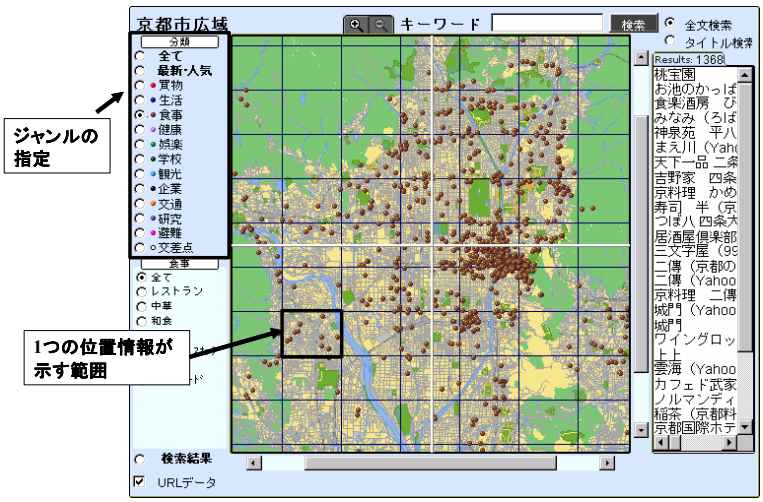

図 8　GeoLink Kyoto サービスにて情報を提示した例（ジャンル: 食事）

Fig. 8　Example of GeoLink Kyoto service (category: meal)

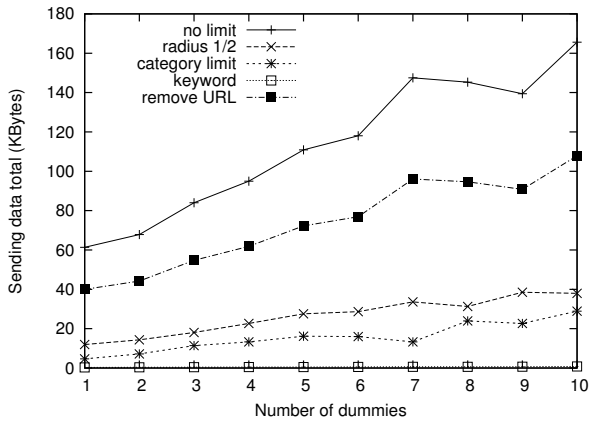

図 9　コスト削減手法ごとの返信メッセージのコストの比較

Fig. 9　Comparison of the communication costs for answered messages

している．グラフから，どの手法も返信メッセージのコストを削減することが読み取れる．特にジャンルや具体的なキーワードを指定した場合，該当するスポット数が激減するためにコストを大きく削減することができる．無線通信の帯域は数年で大きく広がる可能性が高いので，返信メッセージの総コストを数十 [KByte] に抑えることができれば十分に実用に耐えうるシステムを作成することが可能であるといえる．

6. ま と め

本稿では，以前筆者らが提案した架空情報を用いた位置プライバシー保護手法 [12] の概要とコスト面の問題点を説明し，その手法を用いたサービスにおける送受信メッセージの削減手法を提案した．また，実際に運用されているサービスを利用して提案手法の評価実験を行った．その結果，提案手法はコストを減少させることを示した．さらに，コスト面からサービスとして実用可能であることを確認した．

今後は，架空情報を用いた位置プライバシー保護手法 [12] のダミーがより真の位置情報に近い軌跡を描くようにアルゴリズムを改善することが課題である．また，位置プライバシーを評価するより適切な指標も同時に研究する必要がある．

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る．

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] Alastair R. Beresford and Frank Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, Vol. 2, No. 1, pp. 46–55, 2003.

[3] David Chaum. The dining cryptographers problem: Unconditional sender and recipient untraceability. *Journal of Cryptology*, Vol. 1, pp. 65–75, 1988.

[4] I. Getting. The global positioning system. In *IEEE Spectrum*, Vol. 30, pp. 36–47, December 1993.

[5] D. Goldschlag, M. Reed, and P. Syverson. Onion routing for anonymous and private internet connections. *Communications of the ACM (USA)*, Vol. 42, No. 2, pp. 39–41, 1999.

[6] 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.

[7] NTT Open Lab. Geolink kyoto. www.digitalcity.gr.jp/openlab/kyoto/map_guide_j.html.

[8] Andreas Pfitzmann and Marit Kohntopp. Anonymity, unobservability, and pseudeonymity: a proposal for terminology. In *International workshop on Designing privacy enhancing technologies*, pp. 1–9. Springer-Verlag New York, Inc., 2001.

[9] Michael Reiter and Aviel Rubin. Crowds: Anonymity for web transactions. *ACM Transactions on Information and System Security*, Vol. 1, No. 1, pp. 66–92, June 1998.

[10] 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.

[11] 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.

[12] 貴戸秀年, 柳沢豊, 佐藤哲司. 移動軌跡データベースにおける位置匿名性. 情報処理学会研究報告 (2004-DPS-120), pp. 19–24, November 2004.