道路ネットワークにおける位置情報プライバシ

高木　駿† 曹　洋†† 浅野　泰仁†† 吉川　正俊††

† 京都大学工学部情報学科　〒 606–8501 京都府京都市左京区吉田本町 36-1
†† 京都大学情報学研究科　〒 606–8501 京都府京都市左京区吉田本町 36-1
E-mail: †takagi.shun.45a@st.kyoto-u.ac.jp, ††{yang,asano,yoshikawa}@i.kyoto-u.ac.jp

あらまし　近年，位置情報の保護に，差分プライバシの概念を応用した Geo-I(Geo-Indistinguishability) を用いた方法が注目されている．Geo-I は，摂動法と呼ばれる自分の位置情報に雑音を加える際のプライバシ基準であるが，ユークリッド平面上で定義されており，道路ネットワークを考慮していない．そこで私たちは，それが原因で，近くのレストランを探すなどの近傍検索などの道路ネットワークを用いる LBS（位置情報サービス）に，単純な Geo-I の適用をすると，プライバシ保護の保証の不十分さ，有用性の低下という弱点が生じ得ることを明らかにした．さらに，それらの弱点を解決することのできるグラフ上の摂動法と，その満たすべきプライバシ基準 GGI(Geo-Graph-Indistinguishability) を提案した．次に，具体的なグラフ上の摂動法を提案し, それが GGI を満たすことを証明した．最後に，グラフ上でのプライバシ保護の度合いと有用性の定式化を提案し，日本の二つの都市の道路ネットワーク上で実験を行うことで，そのトレードオフに関してそのメカニズムが Geo-I を満たすあるメカニズムよりも優れていることを示した．

キーワード　位置情報の保護，Geo-Indistinguishability，差分プライバシ

1 はじめに

近年，スマートフォンなどのモバイルデバイスの普及と GPS システムの発達により，LBS（位置情報サービス）が急激に増加している．それらの使用は便利である一方，それらによる正確な位置情報の漏洩が，自宅や，職場などの個人情報の漏洩に繋がり，個人の特定などのプライバシに関する問題を起こしている．それを解決するために，位置情報の保護に関する研究が盛んに行われている．その保護の方法の一つに，この論文で扱う，摂動法と呼ばれる自分の位置情報にランダムな雑音を加える方法がある．摂動法では，ユーザが LBS 提供者に自分の位置情報を送る前に，メカニズムと呼ばれる確率分布を返す関数に従ってランダムな雑音を加えることでプライバシを保護する．そのため，雑音を自身の計算機のみで計算ができ，信頼できるサーバを必要としないが，LBS 提供者は雑音の加わった位置情報を基にサービスを提供する必要があるため，有用性（サービスの質）は雑音の大きさに依存する．有用性を上げるためにはユーザは雑音を小さくする必要があるが，その場合，プライバシ保護の度合いが低下する．つまり，有用性とプライバシの保護の度合いはトレードオフの関係にあり，摂動法の最も重要で，難しい課題がこのトレードオフの良いバランスを実現することである．摂動法の中でも特に有名なのが，プライバシ基準である Geo-I(Geo-Indistinguishability) [1] を用いる方法である．この基準は数学的に厳密なプライバシ基準である差分プライバシという広く受け入れられている強力な概念を位置情報に応用したものであり，この基準を満たすことでプライバシ保護を保証することができる．Andrés ら [1] はこの基準を満たすことでプライバシ保護の保証を得ると同時に，高い有用性を示す摂動法 PL を提案した．また，それは信頼できるサーバが必要なく，攻撃者の前知識に強く，アルゴリズム，計算が簡単であるという利点を持ち，優れたものとして一般的に認められている．

Geo-I が発表されてから，Geo-I の弱点に関して様々な研究がされてきた．Yu ら [2] は Geo-I が最適化攻撃と呼ばれる攻撃に弱いことを示し，Geo-I に Geo-I と相補的な関係にある Expected Inferece Error [3] の概念を取り入れて，弱点を補い合うような枠組みを作った．Chatzikokolakis ら [4] は Geo-I がプライバシの意味的な部分を捉えられていないことに着目し，人口密度や地域情報などの意味的なプライバシを保護できるメカニズムを構築した．Oya ら [5] は，プライバシ保護の度合いを [1] とは別の観点で定式化し，攻撃者がある事前分布を持つ場合には Geo-I の満たすメカニズムのプライバシ保護の度合いがその観点で低いことを示した．

近くのレストランを探すなどの近傍検索の LBS は，道路ネットワークを用いた方が近傍が正確に表されることから，より良いサービスを提供可能である．この研究では，こういった LBS において，単純な Geo-I の道路ネットワークへの適用では，Geo-I が道路ネットワークを考慮せずに平面上で定義されていることが原因で，プライバシ保護の保証の不十分さ，有用性の低下という弱点が生じ得ることを明らかにした．例えば PL を用いると，雑音を加えることによって，海の上などの人のいるはずのない場所を出力する可能性がある．この出力を見ると，その場所が本当の位置でないことは明らかであり，プライバシ保護の度合いが低下している可能性がある．また，近くに橋のない大きな川がある場合，ユーザは対岸を遠いと感じるが，Geo-I はユークリッド平面上で定義されているため近いと判断してしまい，有用性の低い場所を出力する可能性が上がってしまう．

そこで私たちは，グラフ上での摂動法とその満たすべき道路ネットワーク上のプライバシ基準 GGI(Geo-Graph-

Indistinguishability) を提案する．グラフ上での摂動法を用いることで，人がいるはずのない場所に曖昧化することを防ぐことができ，上で述べたプライバシに関する弱点を解決することができる．さらにこの摂動法が差分プライバシに基づく基準 GGI を満たすことで，強力な道路ネットワーク上のプライバシ保護の保証を得ると同時に，道路ネットワーク上に合わせた基準であることから，近傍検索やミーティングポイントの決定などのサービスにおいて，Geo-I を用いた摂動法よりも高い有用性とプライバシ保護の度合いを期待できる．実際に，実験を行うことで，GGI を満たす一つのメカニズム GE を提案し，それが PL と PL をグラフ上のメカニズムへと拡張したメカニズムよりも，その観点で優れていることを示した．

2 関 連 研 究

この章では，差分プライバシとその一般的なデータへの拡張，Geo-I と位置情報のプライバシ保護に関する研究を紹介する．

2.1 ϵ-差分プライバシ

この節では Geo-I の枠組みである Dwork [6] の差分プライバシを説明する．差分プライバシはデータベースにおける個人のデータのプライバシ保護を目的としたプライバシ基準である．あるデータベースから母集団に関する統計的な知識を提供する目的で統計量が公開されたとする．実はこのとき，個人の情報を知りたい攻撃者にこの統計量を見られると，個人の情報が高い精度で推測されてしまう恐れがある．例えば，ある疾患に関する調査のデータベースがあったときに，”個人が疾患にかかっているかどうか”という情報は個人が知られたくない情報である．そこで，”全国の 30 代で疾患を持つ人の人数”・”東京都在住で疾患を持つ人の人数”・”東京都在住の 30 代で疾患を持つ人の人数”という三つのクエリを発行したとする．このとき，東京都在住の 30 代の A さんが疾患を持つかどうかを攻撃者は推測できてしまうのだろうか．それは，攻撃者の事前知識や，データベースの内容に依って，できる可能性もあるし，できない可能性もあるとしか言えなく，プライバシが保護されている保証はない状態なのである．そこで，Dwork は差分プライバシと呼ばれるデータベースから個人のプライバシを保護するために満たされるべき基準を提案し，今やその基準はあらゆる攻撃者やデータベースに対してプライバシの保護ができる，強力な基準であることが広く認められている．

2.1.1 定　　義

ここではその定義を述べる．そのためにまず，記法を導入する．X をレコードのドメイン，レコード $x \in X$ を個人の情報を含むデータ，レコードの集合をデータベース $D = \{x_i\}_{i=1}^{n}$ とする．データベース $D \in \mathcal{D}$ とクエリ $q \in Q$ を受け取り，クエリの応答値 $q(D)$ にランダムな雑音を加えた応答値 $y \in Y$ を返す計算機構を m と置く．ここで取りうるデータベースの集合を $\mathcal{D}$，クエリ応答値に雑音を加えた結果得られる値の集合を Y，取りうるクエリの集合を Q とした．2 つの同じサイズのデータベース D, D' において，同一でないレコードの数を $d_h(D, D')$(ハミング距離) で表し，$d_h(D, D') = 1$ のとき，D, D' は隣接しているという．このとき，$\epsilon \in \mathbb{R}^+$ について差分プライバシは以下のように定義される．

定義 1. ϵ-差分プライバシ　クエリ $q \in Q$ において，隣接している任意のデータベースの組 $D, D' \in \mathcal{D}$，および任意の出力の部分集合 $S \subseteq Y$ について，

$$\frac{Pr(m(q, D) \in S)}{Pr(m(q, D') \in S)} \leqq \mathrm{e}^{\epsilon} \tag{1}$$

ならば，計算機構 m の従う確率分布は ϵ-差分プライバシを満たすという．

直感的には，m が従う確率分布が ϵ-差分プライバシを満たすとき，$m(D)$ を観測されたとしても，隣接するデータベースに対する出力 $m(D')$ が似ていることが保証されているため，ある一つのレコード，すなわち個人のレコードが何であるかを推測できないことを表している．従って，データベースの統計量を ϵ-差分プライバシを満たす確率分布によってランダムな雑音を加えることによって，個人のプライバシを保護しつつ有用性の高い出力をすることができるのである．

2.1.2 一般的なデータへの拡張

Chatzikokolakis ら [7] はデータベース上でのみ定義されていた差分プライバシを一般的なデータへの拡張を行った．差分プライバシにおいて，データベース間のハミング距離 d_h を用いて以下が導出される．

$$\frac{Pr(m(q, D) \in S)}{Pr(m(q, D') \in S)} \leqq \mathrm{e}^{\epsilon d_h(D, D')} \tag{2}$$

[7] はこの不等式が表す概念を，データベースからデータへの一般化を行った．ここで，あるドメイン $\mathcal{Z}$ 上の確率分布を与えるあるドメイン $\mathcal{X}$ 上のメカニズム $K : \mathcal{X} \to \mathcal{P}(\mathcal{Z})$ を考える．$\mathcal{P}(\mathcal{Z})$ は，$\mathcal{Z}$ 上の確率分布を表す．このとき，$\mathcal{X}$ 上の”距離”を表す $d_{\mathcal{X}}$ を用いて，ある $\epsilon \in \mathbb{R}^+$ について，差分プライバシを一般化したものである $\epsilon d_{\mathcal{X}}$-privacy は以下のように定義される．

定義 2. メカニズム K が任意の $x, x' \in \mathcal{X}, Z \subseteq \mathcal{Z}$ について，以下の不等式

$$\frac{K(x)(Z)}{K(x')(Z)} \leqq \epsilon d_{\mathcal{X}}(x, x') \tag{3}$$

を満たすとき，K は $\epsilon d_{\mathcal{X}}$*-privacy* を満たすという．

この定義は二つのデータベースが似ているほど，生成される分布は似ているべきであるという差分プライバシの概念を一般化したものと言える．これに従えば，差分プライバシの場合 ((2) 式) は $x \in \mathcal{X}$ はデータベースを，$d_{\mathcal{X}}$ はハミング距離 d_h を表している．[7] ではメカニズム K が $\epsilon d_{\mathcal{X}}$-privacy を満たすとき，一般的に，以下の二つの性質を満たすことを示している．

a） hiding function

一つ目の性質は任意の関数 hiding function ϕ: $\mathcal{X} \to \mathcal{X}$ の概念を用いて示される．一般的に，任意の攻撃者が任意の $z \sim K(x)$,$z' \sim K(\phi(x))$ のそれぞれを観測したときに得る事後分布間の multiplicative distance は以下の不等式を満たす．

$$d_{\mathcal{P}}(p(x|z), p(x|z')) \leqq 2\epsilon d_{\mathcal{X}}(\phi) \tag{4}$$

ここで $d_{\mathcal{X}}(\phi)$ を $sup_{x\in\mathcal{X}} d_{\mathcal{X}}(x, \phi(x))$ と定義する．この式は，ある似ている二つの入力に対するメカニズムの出力を観測したとき，攻撃者の二つの事後分布が似ていることを保証しており，本当のデータ x が何であるかについてあまり情報を得ることができないことを表している．ここで，攻撃者を仮定していないため，$d_{\mathcal{X}}$-privacy を満たすメカニズムは任意の攻撃者に対してある程度のプライバシ保護を保証していることがわかる．

b） informed attacker

もう一つの性質は攻撃者の持つ事前分布と事後分布の距離を用いて示される．この距離を測ることで攻撃者が出力を観測したときに，どれだけデータについて知ることができるかを測ることができる．ここで保護したいデータ $x \in \mathcal{X}$ がある集合 $N \subseteq \mathcal{X}$ に含まれていると知っている攻撃者 informed attacker を仮定する．このとき，その攻撃者の持つ任意の $\mathcal{X}$ 上の事前分布 π と任意の $z \sim K(x)$ を観測したその事後分布に対して，全ての $N \in \mathcal{N}$ について以下の不等式が成り立つ．

$$d_{\mathcal{P}}(\pi_{|N}, p_{|N}(x|z)) \leqq \epsilon d_{\mathcal{X}}(N) \tag{5}$$

ここで $\pi_{|N}(x) = \pi(x|N), p_{|N}(x|z) = p(x|z, N)$，$d_{\mathcal{X}}(N) = max_{x,x'\in N} d_{\mathcal{X}}(x, x')$ とする．これは，$d_{\mathcal{X}}(N)$ が小さいほど事前分布と事後分布が似ている，つまり攻撃者が x について詳しく知っているほど新たな情報を得られないというを表している．

2.2 Geo-Indistinguishability

ここでは，差分プライバシの概念を応用した位置情報のプライバシ基準 Geo-I を説明する．$\mathcal{X}$ をユーザの存在しうる位置の集合，$\mathcal{Z}$ をメカニズムによって曖昧化した結果得られる位置の集合とする．直感的には，メカニズム K が Geo-I を満たすとは，任意の似ている位置 $x, x' \in \mathcal{X}$ について確率分布 $K(x), K(x')$ が似ているということを表している．これはユーザが確率分布 $K(x)$ に従って位置 $z \in \mathcal{Z}$ を出力したときに，攻撃者が z を観測しても $K(x)$ から出力されたのか，$K(x')$ から出力されたのか見分けがつかない状態を表し，そうすることによってユーザの位置情報 x を保護している．$\epsilon \in \mathbb{R}^+$ について Geo-I は以下のように定義される．

定義 3. *(ϵ-Geo-Indistinguishability)* メカニズム K が任意の $x, x' \in \mathcal{X}$ について以下の不等式

$$d_{\mathcal{P}}(K(x), K(x')) \leqq \epsilon d(x, x') \tag{6}$$

を満たすとき，K は *ϵ-Geo-Indistinguishability* を満たすという．

$d(x, x')$ はユークリッド距離を表す．この定義は $\forall x, x' \in \mathcal{X}, Z \subseteq \mathcal{Z}, \frac{K(x)(Z)}{K(x')(Z)} \leqq \mathrm{e}^{\epsilon d(x,x')}$ と書き換えることができ，Geo-I は (3) 式 ($\epsilon d_{\mathcal{X}}$-privacy の定義式) において，データを位置情報，距離 $d_{\mathcal{X}}$ をユークリッド距離 d にしたものであることがわかる．

Geo-I は $\epsilon d_{\mathcal{X}}$-privacy の一つであるから，$\epsilon d_{\mathcal{X}}$-privacy の持つプライバシ保護の保証，つまり，hiding function と informed ataccker に関する保証（2.1.2 節）を持つ．2.1.2 節で説明したように，これらは任意の攻撃者に対する，ある程度のプライバシ保護の保証を示している．

この論文 [1] が発表されてから，様々な応用が研究されてきた．To ら [8] はプライバシを保護した時空間クラウドソーシングサービスのフレームワークを Geo-I を使って構築した．Bordenabe ら [9] は位置情報を簡略化し，位置情報を減らすことで最適化の計算を可能にし，Geo-I を満たすメカニズムの中で 2.4 節で述べる有用性の観点で最適なものを導出した．他にも様々な応用の研究がされており，位置情報のプライバシ保護の研究の中で最も注目されていると言える．

2.3 複数の位置情報の曖昧化

ここで複数の位置情報プライバシを保護する場合を考える．[1] でも述べられているように，このプライバシの保護の方法として，一つ一つの位置情報に Geo-I を適用する単純な方法が考えられる．しかし，この方法は，差分プライバシに由来する合成定理と呼ばれる定理によって，プライバシ保護の度合いが低下してしまう．例えば相関性のある k 個の位置情報を ϵ-Geo-I を満たすメカニズムで曖昧化する場合，そのプライバシ保護の保証は $k\epsilon$-Geo-I となり，k 倍低下する．これは差分プライバシの分野でも未解決問題であり，これを解決するための方法は盛んに研究されている．

ここではそれに関する新しい研究の一部を紹介する．Kairouz ら [10] は合成定理をナイーブな場合だとプライバシの保護の度合いが $O(k)$ 倍まで低下していたのを $O(k^{1/2})$ 倍まで抑えるメカニズムを提案した．Chatzikokolakis ら [11] は Geo-I を用いて，相関を持つ位置情報に対応する方法を提案した．Xiao ら [12] は移動しているユーザが位置情報を送信する際の差分プライバシに基づくメカニズムを提案した．

2.4 有用性，プライバシ保護の度合いの定式化

Shokri ら [3], [13] はメカニズムを用いた位置情報を保護する方法で曖昧化した際の，有用性とプライバシ保護の度合いの定式化，攻撃者のモデル化を行った．

2.4.1 有 用 性

LBS がユーザに提供するサービスの質（有用性）SQL(Service Quality Loss) はユーザの曖昧化した位置とサービスの種類に依存する．つまり，ユーザが自分の位置を受けるサービスの観点で似ている位置に曖昧化したなら受けるサービスの質は良いし，似ていない位置に曖昧化したなら受けるサービスの質は悪くなると考えられる．このことからメカニズム K による有用性は以下のように定式化することができる．

$$SQL(\pi_u, K, d_q) = \sum_{r,r'} \pi_u(r) K(r)(r') d_q(r, r') \tag{7}$$

π_u はユーザがいる位置の確率分布を表し，ユーザの事前分布と呼ぶ．$d_q(r, r')$ は使用される LBS によって異なる r と r' の非類似度を表し，(7) 式はメカニズム K で曖昧化した位置 r' と本当の位置 r の非類似度 $d_q(r, r')$ の期待値を表している．Shokri らは $d_q(r, r')$ として，r と r' のユークリッド距離を用いている．

2.4.2 プライバシ保護の度合い

プライバシの保護の度合い (Location Privacy) はメカニズムによって曖昧化された位置を攻撃者が見たときに，どれくらいの精度でユーザの本当の位置を予測できるかで測ることができる．よって，以下のように定式化することができる．

$$LP(\pi_u, K, h, d_q) = \sum_{\hat{r},r',r} \pi_u(r)K(r)(r')h(r')(\hat{r})d_q(\hat{r}, r)$$

h は攻撃者の予測関数であり，$h(r')(\hat{r})$ は攻撃者が r' を観測したときにユーザの位置を $\hat{r}$ と予測する確率を示す．つまり LP は攻撃者が予測するユーザの位置 $\hat{r}$ と，ユーザの本当の位置 r の非類似度 $d_q(\hat{r}, r)$ の期待値を表す．有用性と同様に，Shokri らは d_q としてユークリッド距離を用いている．

2.4.3 攻撃者モデル

攻撃者は曖昧化された位置を観測して，本当の位置を推測しようとする．このとき，攻撃者はユーザの位置の事前知識を表す確率分布 π_a を持ち，ユーザの使ったメカニズム K を知っているとする．攻撃者は曖昧化された位置を観測したときに，最適化攻撃 [13] を使ってユーザの位置を推測する．最適化攻撃では，攻撃者は以下の線形計画問題を解いて予測関数 h を導出し，この予測関数 h に基づきユーザの位置を推測する．

$$\begin{aligned} \underset{h}{\text{minimize}} \quad & \sum_{\hat{r},r',r} \pi_a(r)K(r)(r')h(r')(\hat{r})d_p(r, \hat{r}) \\ \text{subject to} \quad & \sum_{\hat{r}} h(r')(\hat{r}) = 1, \forall r', \\ & \sum_{\hat{r}} h(r')(\hat{r}) \geqq 0, \forall r', \hat{r} \end{aligned} \tag{8}$$

表 1　記号の意味

記号	意味
LBS	位置情報サービス (location based service)
u, a	ユーザ，攻撃者
X	二次元平面上の座標の集合
$\mathcal{G}$	重み付き無向連結グラフ (V, E)
V	二次元平面上のグラフの頂点集合
E	辺集合．重みは頂点間の最短の道路上の距離
$\pi_u(r)$	ユーザ u が r にいる確率
$\pi_a(r)$	攻撃者 a が持つユーザが r にいる確率
$K(r)(r')$	メカニズム．位置 r を r' に曖昧化する確率
$d(x, x')$	x と x' のユークリッド距離
$d_s(v, v')$	頂点 v, v' 間のグラフ上の最短距離
$SQL(\pi_u, K, d_q)$	曖昧化した位置と u の位置の d_q の期待値
$LP(\pi_a, K, h, d_p)$	a が推測する位置と実際の位置の d_q の期待値

3　道路ネットワークにおける Geo-Indistinguishability の弱点

この章では 1 章で述べた，道路ネットワークを考慮することで生じる Geo-I の弱点について詳しく述べる．

ここで，道路ネットワークを無向連結グラフ $\mathcal{G} = (V, E)$ として定義する．V は平面上の頂点集合，E は重み付きの辺集合である．$e = (v_1, v_2, w_e) \in E$ において，重み w_e は頂点 v_1, v_2 間の道路の最短の距離を表している．

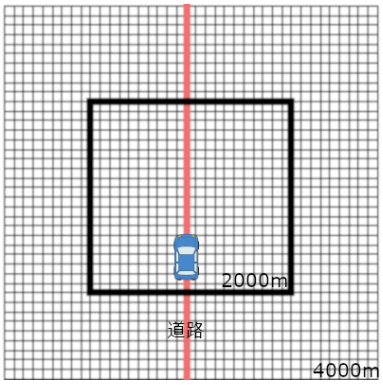

図 1　仮想的なマップ．

3.1 プライバシ保護の保証の不十分さ

ここでは，攻撃者が道路ネットワークを考慮することで，ユーザの位置情報の予測の精度を上げることができる可能性があることを実験を行うことで示す．ここで，Geo-I を満たすメカニズムとして PL [1] を使用する．これは平面ラプラス分布を利用したもので，$PL_\epsilon(x)(z) = \frac{\epsilon^2}{2\pi} \mathrm{e}^{-\epsilon d(x,z)}$ と表される．メカニズムによるプライバシ保護の度合いとして LP（2.4.2 節）を用いることとする．この LP として，[1], [3] で用いられたメトリックにユークリッド距離を適用した以下の LP_e を用いる．

$$\begin{aligned} LP_e(\pi_u, K, h) &= LP(\pi_u, K, h, d) \\ &= \sum_{\hat{r},r',r} \pi_u(r)K(r)(r')h(r')(\hat{r})d(\hat{r}, r) \end{aligned} \tag{9}$$

これはユーザの事前分布に対する，メカニズム K を用いたときの，ユーザの本当の位置と予測関数 h を用いる攻撃者の予測する位置の間のユークリッド距離の期待値を表している．ここからは，実際に PL_ϵ を用いて曖昧化を行った場合，攻撃者が道路ネットワークを考慮すると，LP_e を下げることができる可能性があることを実験を行うことで示す．実験は格子点に座標を持つ 100m*100m で分割されている 4000m*4000m の仮想的な空間上 (図 1) で行なった．

中央の赤線を道路とし，ユーザの事前分布を内側にある黒枠内の一様分布とする．道路ネットワークを考慮しない攻撃者は黒枠内の一様分布を事前分布として持ち，道路ネットワークを考慮する攻撃者はユーザは黒枠内の赤線上の一様分布を事前分布としてを持つとする．このときの ϵ に対する各攻撃者に対する LP を図 2 に示した．攻撃者は道路情報を考慮することにより，LP_e を下げる，つまり予測の精度をかなり上げることができていることがわかる．これは PL が平面上で定義されたメカニズムであり，道路以外の人の存在し得ない場所に曖昧化してしまうために，道路情報を持つ攻撃者がそれによって曖昧化された位置をみると，精度の高い予測ができる可能性があることを表している．

平面上で定義された Geo-I を満たすメカニズムは，道路ネットワークを考慮しない平面上の一定の事前分布を持つ攻撃者に対しては，LP_e は一定の値を取り，強いプライバシ保護の保証を持つ．しかし，この実験で示したように，全ての攻撃者は道路ネットワークを考慮することができ，平面上の点によって

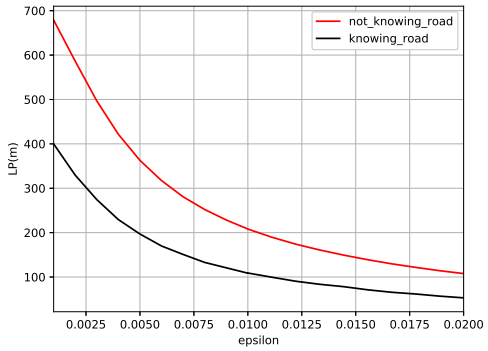

図 2　道路情報を持つ攻撃者と持たない攻撃者の LP.（人工データ）

様々な LP_e（場合によってはかなり低い）を取り得ることがわかる．これは，道路ネットワークを用いた LBS における Geo-I のプライバシ保護の保証の不十分さを表している．

3.2　近傍検索における有用性の低下

近くのレストランを探すなどの近傍検索において，有用性はユークリッド距離ではなく，道路ネットワーク上の最短経路長で測られると考えられる．そのことから，道路ネットワークを用いる LBS における有用性として，メトリックにグラフ上の最短経路長 d_s を用いた SQL_s を提案する．

$$\begin{aligned} SQL_s(\pi_u, K) &= SQL(\pi_u, K, d_s) \\ &= \sum_{v,v'} \pi_u(v) K(v)(v') d_s(v, v') \end{aligned} \tag{10}$$

Geo-I はユークリッド距離を用いて定義されているために，Geo-I を満たすメカニズムは上の有用性が低くなると予測できる．例えば，横断できない大きな川があるとき，対岸に曖昧化されるのは有用性が低いと考えられるが，Geo-I はそれを考慮することはできない．これについては 5 章で詳しく述べる．

4　Geo-Graph-Indistinguishability

本章では前章で述べた Geo-I の弱点を解決することのできる，道路ネットワークを考慮したプライバシ定義である GGI(Geo-Graph-Indistinguishability) の定義を提案し，そのプライバシ保護の保証を説明する．最後に，それを満たすメカニズムを提案する．

4.1　定　　義

グラフ上のメカニズム K は V 上の頂点に対して頂点集合 $W \subseteq V$ 上の確率分布を与える関数 $K : V \to \mathcal{P}(W)$ で与えられる．ここで，$\epsilon \in \mathbb{R}^+$ に対して ϵ-GGI を以下のように定義する．

定義 4. *(ϵ-Geo-Graph-Indistinguishability)* グラフ上のメカニズム K が任意の頂点 $v, v' \in V$ について以下の不等式

$$d_{\mathcal{P}}(K(v), K(v')) \leqq \epsilon d_s(v, v') \tag{11}$$

を満たすとき，K は *ϵ-Geo-Graph-Indistinguishability* を満たすという．

これは $\epsilon d_{\mathcal{X}}$-privacy の概念を道路ネットワークを表すグラフに適用したものであり，(3) 式のデータ $x \in \mathcal{X}$ をグラフの頂点 $v \in V$，距離 $d_{\mathcal{X}}$ を頂点間の最短経路長 d_s にしたものであるから，道路ネットワークに対する $\epsilon d_{\mathcal{X}}$-privacy の自然な適用と考えられる．直感的に，この定義は任意の $v, v' \in V$ に対して v と v' の間の最短経路長 d_s が近いほど，$K(v), K(v')$ が似ていることを意味しており，摂動後の頂点を観測した攻撃者が v から出力されたのか，v' から出力されたのかが見分けがつかないことを保証している．ϵ がその保証の度合いを表しており，ϵ を調整することでユーザは，任意のプライバシーレベルを得ることができる．また，ϵ-GGI は $\epsilon d_{\mathcal{X}}$-privacy の一つであるから，2.1.2 節で説明した $\epsilon d_{\mathcal{X}}$-privacy の持つプライバシ保護の保証 [7] を持つ．

4.2　ϵ の 解 釈

ϵ-GGI を満たすメカニズムが 2.1.2 節で述べたプライバシ保護の保証を持つことがわかったが，結局，ϵ の取る値によって，どの程度のプライバシ保護を保証しているかは理解し難い．そこで，この節では，ϵ の値によるメカニズムのプライバシの保護の度合いを定式化する方法として LP_s を用いることを提案する．LP_s は LP(2.4.2 節) を道路ネットワークに適用したもので，ユーザの事前分布 π_u と攻撃者の予測関数 h に関して，以下のように表される．予測関数 h は, 一例として最適化攻撃（2.4.3 節）によるものが考えられる．

$$\begin{aligned} LP_s(\pi_u, K, h) &= LP(\pi_u, K, h, d_s) \\ &= \sum_{\hat{r}, r', r} \pi_u(r) K(r)(r') h(r')(\hat{r}) d_s(\hat{r}, r) \end{aligned} \tag{12}$$

これは，メカニズム K を用いたときの，ユーザの事前分布 π_u に対する，本当の頂点と予測関数 h を用いる攻撃者の予測する頂点の間の最短経路長の期待値を表している．

さらに，各ユーザが自分のいる位置 r でメカニズム K を用いた場合のプライバシ保護の度合いの定式化 LPr_s を提案する．

$$LPr_s(r, K, h) = \sum_{r', r} K(r)(r') h(r')(\hat{r}) d_s(\hat{r}, r)$$

これは頂点 r でメカニズム K を用いた場合の，ユーザの本当の頂点 r と予測関数 h を用いる攻撃者の予測する頂点の間の最短経路長の期待値を表している．各頂点の各 ϵ に対する LPr_s の値をユーザに与えておくことで，各頂点 r にいるユーザはプライバシ保護の度合いを解釈し，調整することができる．例えば LPr_s が 100 m であるなら，このメカニズムを用いることで，ユーザは平均的に 100 m の範囲で位置情報を保護できることを示す．

4.3　GGI を満たすメカニズム

ここでは GGI を満たすメカニズムを提案し，その証明を与える．$\epsilon \in \mathbb{R}^+$，頂点 $v \in V$ が与えられたとき，W 上の確率分布を与えるグラフ上のメカニズム GE を以下のように定義する．

定義 5. *GE_ϵ (Graph-Exponential-Mechanism)*

$$GE(v)(w) = \alpha(v) \mathrm{e}^{-\frac{\epsilon}{2} d_s(v, w)} \tag{13}$$

α は正規化係数で以下のように表される．

$$\alpha(v) = \frac{1}{\sum_{w \in V} \mathrm{e}^{-\frac{\epsilon}{2} d_s(v,w)}} \tag{14}$$

このメカニズムは差分プライバシの指数メカニズム *[14]* に対応している．

定理 1. *GE_ϵ は ϵ-Geo-Graph-Indistinguishability を満たす．*

Proof. 任意の二点 $v, v' \in V$ で以下が成り立つことを示せばよい．

$$d_p(K(v), K(v')) \leqq \epsilon d_s(v, v') \tag{15}$$

つまり，任意の二点 $v, v' \in V$ が生成する分布 $GE(v)$ と $GE(v')$ の比の最大値を調べれば良い．任意の $v, v' \in V, w \in W$ について以下が成り立つ．

$$\frac{GE(v)(w)}{GE(v')(w)} = \frac{\alpha(v)\mathrm{e}^{-\frac{\epsilon}{2} d_s(v,w)}}{\alpha(v')\mathrm{e}^{-\frac{\epsilon}{2} d_s(v',w)}} = \frac{a(v)}{a(v')} \mathrm{e}^{\frac{\epsilon}{2}(d_s(v',w) - d_s(v,w))}$$

これが最大値を取るのは，頂点 $w \in W$ に対して $d_s(v', w) - d_s(v, w)$ が最大値をとるときである．$\mathcal{G}$ は無向連結グラフであるから，任意の頂点を経由でき (三角不等式)，$\forall w \in V, -d_s(v, w) + d_s(v', w) \leqq -d_s(v, w) + d_s(v', v) + d_s(v, w) = d_s(v, v')$ が成り立つ．よって以下の不等式が成り立つ．

$$\frac{GE(v)}{GE(v')} \leqq \frac{\alpha(v)}{\alpha(v')} \mathrm{e}^{\frac{\epsilon}{2} d_s(v,v')} \tag{16}$$

ここで以下が成り立つことを示す．

$$\frac{\alpha(v)}{\alpha(v')} < \mathrm{e}^{\frac{\epsilon}{2} d_s(v,v')} \tag{17}$$

(14) より，任意のグラフ $\mathcal{G}$ の任意の頂点 $v, v' \in V, w \in W$ について，$\sum_{w \in V} \mathrm{e}^{-\frac{\epsilon}{2} d_s(v',w)} - \mathrm{e}^{\frac{\epsilon}{2} d_s(v,v')} \sum_{w \in V} \mathrm{e}^{-\frac{\epsilon}{2} d_s(v,w)} < 0$ を示せば良い．三角不等式より，$\forall w \in W, d(v, w) - d(v, v') \leqq d_s(v', w)$ が成り立つから，$\mathrm{e}^{-\frac{\epsilon}{2} d_s(v',w)} \leqq \mathrm{e}^{-\frac{\epsilon}{2}(d_s(v,w) - d_s(v,v'))}$ が成り立つ．よって，

$$(\text{与式}) \leqq \sum_{w \in V} (\mathrm{e}^{-\frac{\epsilon}{2} d_s(v',w)} - \mathrm{e}^{-\frac{\epsilon}{2} d_s(v',w)}) < 0$$

(16),(17) より，(15) が成り立つ．

□

5 評価実験

この章では，実験を行うことで，GE が Geo-I を満たすあるメカニズムよりも，プライバシ保護の度合いと 3.2 節で定義した有用性の観点で優れていることを示す．まず，プライバシ保護の観点で優れていることを示すために，3.1 節で行った実験を GE を用いて行い，LP_e の値を PL と比較する．次に有用性の観点で優れていることを示すために，日本の二つの都市でグラフを取得し，そのグラフ上で GE と以下で提案する Geo-I を満たすグラフ上のメカニズム PLG で実験を行い，それぞれの SQL_s を比較した．

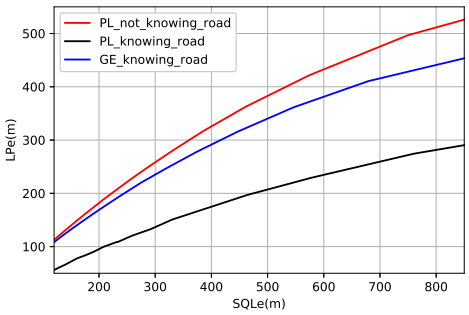

図 3　PL と GE の SQL に対する LP.（人工データ）

5.1 プライバシ保護の保証

ここでは，GE が PL よりプライバシ保護の保証の観点で優れていることを示すために，3.1 節で行なった実験を GE を用いて行い，PL とプライバシの保護の度合い LP_e を比較する．比較を公平にするために，有用性が同じ値のときの LP_e を比較することを考える．ここで，有用性として以下式で表されるユークリッド距離に関する SQL_e を使用する．

$$SQL_e(\pi_u, K) = SQL_e(\pi_u, K, d) = \sum_{r,r'} \pi_u(r) K(r)(r') d(r, r') \tag{18}$$

これはユーザの事前分布に関する，本当の位置と曖昧化された位置の間のユークリッド距離の期待値を表している．GE と PL の SQL_e に対する LP を図 3 に示す．それぞれの SQL_e が等しいときの，道路情報を持つ攻撃者に対する LP_e を比較すると，GE の方が LP_e が高いことが読み取れる．これは GE を用いた場合，攻撃者が攻撃の精度を大きく上げられていないこと示しており，GE が道路情報を持つ攻撃者に対してより強い位置情報の保護ができていることを表している．このことから，プライバシの保護の度合いに関して，GE の方が優れていると言える．

5.2 有　用　性

ここでは，有用性として 3.2 節で定義した道路ネットワークを用いた LBS における有用性 SQL_s を用いて，Geo-I を満たすメカニズムと GGI を満たすメカニズムの有用性を比較する．PL はグラフ上のメカニズムではなく，平面上のメカニズムであるため d_s が定義されていなく，SQL_s が計算できない．そこで，PL の自然な拡張であり，かつ Geo-I を満たすグラフ上のメカニズムとして，下のようなメカニズム PLG を導入する．

$$PLG(v)(w) = \int_{S_w} PL_\epsilon(x_v)(z) dz \tag{19}$$

S_w はグラフの頂点に関する平面上のボロノイ図を考えたときの w のボロノイセル，x_v は頂点 v の座標を表す．$\mathrm{PLG}(v)$ は，直感的には $\mathrm{PL}(x_v)$ に従って頂点 v を平面上の点 z に曖昧化して，z からユークリッド距離で一番近いグラフ上の頂点にリマッピングをすることを考えたときにできる W 上の確率分布を意味している．つまり，$\mathrm{PLG}(v)(w)$ は PL が v を w のボロノイ

セル上に曖昧化する確率を表しており，Geo-I を満たすナイーブなグラフ上のメカニズムと言える．

命題 1. *PLG は ϵ-Geo-Indistinguishability を満たす.*

Proof. 任意の二点 $v, v' \in V$ で以下が成り立つことを示す.

$$d_p(PLG(v), PLG(v')) \leqq \epsilon d(v, v') \tag{20}$$

$\forall v, v' \in V, w \in W$ について，以下を示せば良い.

$$PLG(v)(w) \leqq \mathrm{e}^{\epsilon d(v,v')} PLG(v')(w) \tag{21}$$

PL_ϵ は ϵ-Geo-I を満たすから，任意の $z \in \mathcal{Z}$ で以下を満たす.

$$PL_\epsilon(x_v)(z) \leqq \mathrm{e}^{\epsilon d(x_v, x_{v'})} PL_\epsilon(x_{v'})(z) \tag{22}$$

積分不等式の基礎定理より以下が成り立つ.

$$\begin{aligned}\int_{S_w} PL_\epsilon(x_v)(z)dz &\leqq \int_{S_w} \mathrm{e}^{\epsilon d(x_v, x_{v'})} PL_\epsilon(x_{v'})(z)dz \\ &= \mathrm{e}^{\epsilon d(x_v, x_{v'})} \int_{S_w} PL_\epsilon(x_{v'})(z)dz \end{aligned} \tag{23}$$

(19),(23) より，(21) が成り立つ. □

以上より SQL_s を計算することのできる Geo-I を満たすメカニズム PLG を導出できた．ここから，Geo-I と GGI の有用性を比較するために，PLG と GE の SQL_s を比較する．公平に比較をするために，PLG と GE のプライバシ保護の度合い LP_s（4.2 節）が等しいのときの SQL_s を比較する．PLG の分布は FORTRAN の数値積分ライブラリを用いて近似的に求めた．そのため，確率に誤差があるが，適応 Gaus-Kronrod 積分を用いているため誤差は十分に小さく，理想的な PLG と比べてもプライバシ保護の度合いの低下は無視できる．

比較実験のために使用したデータセットについて説明する．OpenStreetMap[1]を利用して，東京と秋田の二箇所で 4000m*4000m の範囲のグラフを取得した (図 4)．ユーザの事前分布，攻撃者の持つ事前分布を共に中央から最短経路長で 2000m の範囲としたときに，PLG と GE の各 LP_s での SQL_s をプロットするとそれぞれ図 5, 図 6 のようになった．これらを見ると，二つのマップにおいて，PLG よりも GE の方が常に SQL_s が低いことがわかる．つまり，GE の方が有用性が高いことを読み取ることができる．また，東京と秋田の結果を比べると，同じ LP のとき，SQL の差が秋田のグラフの方が大きいことがわかる．これは，秋田のグラフの方がグラフ上の最短経路長とユークリッド距離の差が大きいことが理由の一つに考えられる．

6 議　　論

道路ネットワークの定義より，任意の頂点 v, v' について $d(x_v, x_{v'}) \leqq d_s(v, v')$ が成り立ち，Geo-I を満たす任意のメカニズム K は GGI を満たすことがわかる.

1 : https://openstreetmap.jp/

図 4　取得したグラフ.（左が東京，右が秋田）

$$d_{\mathrm{P}}(K(x_v), K(x_{v'})) \leqq \epsilon d(x_v, x_{v'}) \leqq \epsilon d_s(v, v') \tag{24}$$

逆に，全ての GGI を満たすメカニズムは Geo-I を満すわけではないことがわかる．これは，GGI がユークリッド距離に関して，プライバシ保護の保証を持たないことを意味している．つまり，GGI を満たすメカニズムで位置を曖昧化したとき，ユークリッド距離に関して近い位置でも，攻撃者からは簡単に識別することができてしまう可能性がある．このことを以下で示す．攻撃者がメカニズム K を用いるユーザの位置を識別できる確率は以下のように表すことができる.

$$TP(\pi_u, K, h) = \sum_{\hat{v}, v, w} \pi_u(v) K(v)(w) h(w)(\hat{v}) \delta(\hat{v}, w) \tag{25}$$

$\delta(\hat{v}, w)$ は $\hat{v} = w$ が成り立つときに 1 を，成り立たないときに 0 を返す関数とする．TP は攻撃者が最適化攻撃によって，曖昧化された位置を正しい位置にリマッピングができる確率の期待値を表している．ここで，ユークリッド距離が一定の二つの頂点のみを持つグラフを考える．このグラフの最短経路長が変化するときに，PLG と GE で曖昧化を行った際に TP がどのように変化するかを示すグラフを図 7 に示す．ユークリッド距離が一定であり，ユークリッド距離に関してプライバシ保護の保証を持つ PLG は TP は上がらないが，GE はユークリッド距離に関する保証を持たないため，ユークリッド距離が一定であっても最短経路長が大きくなると，攻撃者の攻撃の精度が大きく上がるのである.

GGI の考え方は，ユークリッド距離の代わりにグラフ上の最短経路長に関してプライバシを保証することで，Geo-I と比べて有用性を上げることができるというものである．その考えの根底にはプライバシをどう捉えるかがあり，多くの LBS では，有用性やプライバシをグラフ上の最短経路長で考えるほうが妥当と考えられる．一方で例えば，天気を取得する場合など，広範囲でのプライバシ保護が必要な場合は，プライバシをグラフを考えるより地域，つまりユークリッド距離で考える方が自然であり，GGI は適していないと考えられる．また，この論文では GGI にグラフ上の最短経路長を使用したが，プライバシの捉え方を変えることで，移動にかかる時間で距離を表す時間距離などの他の距離への応用が簡単にできると信じている.

7 おわりに

本研究では，Geo-I が道路ネットワークを考慮できていないことに着目し，道路ネットワークを考慮した新しいプライバシ

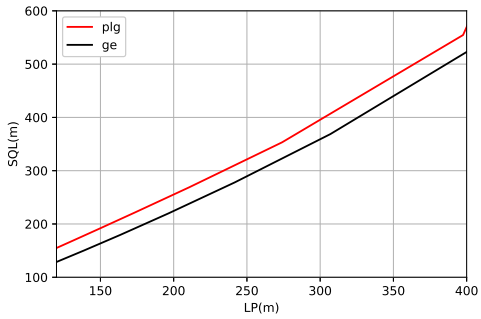

図 5　データ「東京」に対する結果.

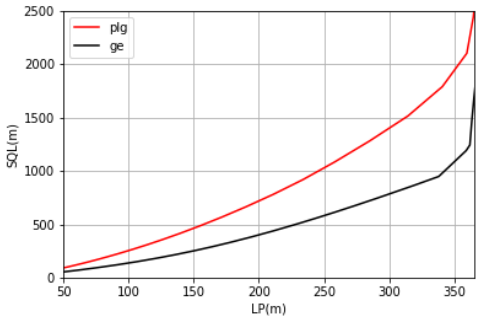

図 6　データ「秋田」に対する結果.

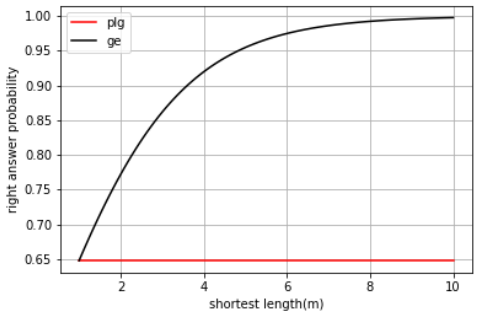

図 7　GE と PLG の TP の変化

定義 GGI を提案し，これを満たすメカニズム GE が，差分プライバシの利点を得ると共に，Andrés らが提案した Geo-I を満たすメカニズム PL よりも，プライバシの観点で優れていることを示し，PL のグラフ上への拡張である PLG よりも，有用性の観点でも優れていることを日本の二つの都市で示した．

しかし，GGI にも考えられる課題が三つある．一つは道路ネットワークを無向グラフで扱っていることである．実世界には一方通行があり，道路ネットワークは有向グラフで表される．このことにより，有用性が低下していると想定される．二つ目が意味的なプライバシを捉えられていないことである．GGI の定義では，グラフ上の距離 r の範囲で ϵr-privacy を保証しているが，今回使用した東京の地図でグラフ上の距離 r で ϵr-privacy が保証されていることと，秋田の地図で同じことが保証されていることは，GGI のプライバシの定義的には同じことであるが，意味的なプライバシでは異なっている．同じ最短経路長 r の範囲であっても，ユーザは頂点や辺（つまり，道路）が少ない方がプライバシを低く感じるのであろう．例えば，秋田の地図を見ると，ある距離 r の範囲には道路の数が少なく，GE を用いる場合，その道路にいることは安易に推測されてしまうだろう．最後が 2.3 節でも述べたように，複数の位置情報を曖昧化した際に相関性がある場合，プライバシが大きく低下してしまうことであり，Geo-I や差分プライバシでも未解決問題である．今後はこれらの課題を解決する方法を考えていきたい．

8　謝　　辞

本研究は JSPS 科研費基盤研究 (S) No. 17H06099, (A) No. 18H04093, (C) No. 18K11314 の助成を受けたものです.

文　　献

[1] K. Chatzikokolakis M. E. Andrs, N. E. Bordenabe and C. Palamidessi. Geo-indistinguishability: Differential privacy for location-based systems.

[2] Calton Pu Lei Yu, Ling Liu. Dynamic differential location privacy with personalized error bounds.

[3] C. Troncoso J.-P. Hubaux R. Shokri, G. Theodorakopoulos and J.-Y. Le Boudec. Protecting location privacy: Optimal strategy against localization attacks.

[4] C. Palamidessi K. Chatzikokolakis and M. Stronati. Constructing elastic distinguishability metrics for location privacy.

[5] C. Troncoso S. Oya and F. Prez-Gonzlez. Is geo-indistinguishability what you are looking for?

[6] C. Dwork. differentiall privacy. *in Automata, Languages and Programming*, pp. 1–12, 2006.

[7] N. E. Bordenabe K. Chatzikokolakis, M. E. Andrs and C. Palamidessi. Broadening the scope of differential privacy using metrics. *Privacy Enhancing Technologies*, pp. 82–102, 2013.

[8] C. Shahabi H. To and L. Xiong. Privacy-preserving online task assignment in spatial crowdsourcing with untrusted server. *in 2018 EEE 34th International Conference on Data Engineering (ICDE)*, pp. 833–844, 2018.

[9] K. Chatzikokolakis N. E. Bordenabe and C. Palamidessi. Optimal geo-indistinguishable mechanisms for location privacy. *in Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, New York, NY, USA*, pp. 251–262, 2014.

[10] S. Oh P. Kairouz and P. Viswanath. The composition theorem for differential privacy. *IEEE Transactions on Information Theory*, 2017.

[11] C. Palamidessi K. Chatzikokolakis and M. Stronati. A predictive differentially-private mechanism for mobility traces. *Privacy Enhancing Technologies*, pp. 21–41, 2014.

[12] Y. Xiao and L. Xiong. Protecting locations with differential privacy under temporal correlations. Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security - CCS ’15, pp. 1298–1309, 2015.

[13] J. L. Boudec R. Shokri, G. Theodorakopoulos and J. Hubaux. Quantifying location privacy.

[14] F. McSherry and K. Talwar. Mechanism design via differential privacy. in 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS’07), pp. 94–103, 2007.