位置情報サービス利用における ダミーを用いたユーザ位置曖昧化手法の視認性評価

林田　秀平† 天方　大地† 原　隆浩† Xing Xie††

† 大阪大学 大学院情報科学研究科 マルチメディア工学専攻
†† Microsoft Research Asia
E-mail: †{hayashida.shuhei,amagata.daichi,hara}@ist.osaka-u.ac.jp, ††xingx@microsoft.com

あらまし　位置情報サービス利用時にユーザの位置プライバシを保護するための手法として，ダミーによる位置曖昧化手法がある．この手法では，ユーザが位置情報サービスを利用する際，自身の位置情報と同時に複数のダミーの位置情報をサービスプロバイダに送信することにより，ユーザの位置を曖昧化している．筆者らは既存研究において，ユーザの行動に関して与えられる情報が訪問場所の集合のみの際に位置を曖昧化する手法を提案している．この方法に対し，攻撃者はエンティティ(ユーザまたはダミー) の位置情報群を地図上にマッピングし，目視でユーザの識別を試みることが考えられる．そのため，ユーザの位置プライバシを保護するには，ユーザとダミーのトラジェクトリを見た際に両者を区別できないことが重要である．本稿では，エンティティのトラジェクトリを可視化するシステムを Web 上に構築し，ユーザの識別可能性を検証する視認性評価実験を行った．実験の結果，上述の既存手法は，目視によるユーザの識別に対して頑強性があることを確認した．

キーワード　位置プライバシ，位置情報サービス，ダミー位置情報

1. 序　　論

近年，スマートフォンやタブレット端末などの GPS を搭載したモバイル端末の普及により，ユーザの現在地に対応した情報を提供する位置情報サービスが数多く展開されている．例えば GoogleMap(注1) では，現在地を送信すると，周辺に存在する人気のレストランや観光スポットの検索が行える．

位置情報サービスでは，ユーザは自身の位置情報をサービスプロバイダに送信し，サービスプロバイダは受信した位置情報に対応したサービスを提供する．この際，サービスプロバイダにはユーザの位置情報が蓄積され，その管理はサービスプロバイダに一任されている．そのため，サービスプロバイダが不正アクセスを受けたり，サービスプロバイダ自体に悪意がある場合，ユーザの位置情報が悪用され，自宅，勤務地，訪問場所，および行動パターン等が露見し，プライバシが侵害される可能性がある．文献 [2] によると，約 50 万人のトラジェクトリデータから，あるユーザの位置情報が 4 点特定されると，95 %の確率で個人との紐づけが可能であることが報告されており，わずかな位置情報の流出でもユーザのプライバシが侵害される危険性は高い．そのため，位置情報サービスを有用なものとするには，ユーザの位置プライバシを保護することが重要である．文献 [1] では,「第三者に自身の現在地や過去の位置情報履歴が露見することを防ぐ能力」を位置プライバシと定義しており，本稿で考える位置プライバシは，この定義に従う．

これまでに，位置プライバシの保護を目的とした研究が数多く行われている [3–6, 8, 10, 11]．文献 [5] では，k-匿名性を考慮した手法が提案されている．k-匿名性とは，k 個の要素の集合において，対象となる要素が他の $(k-1)$ 個の要素と区別できない状態を示す．つまり，集合の中から対象となる要素が特定される確率が $\frac{1}{k}$ の状態である．この手法では，ユーザは自身の位置情報をサービスプロバイダに直接送信するのではなく，匿名化サーバに対して位置情報を送信する．匿名化サーバは，管理しているユーザの位置情報の集合から，k 人以上のユーザを包含するような領域を選択し，その領域をサービスプロバイダに送信することにより，ユーザの位置を曖昧化している．文献 [6] では，Mix Zone を用いた手法が提案されている．この手法では，ユーザは第三者サーバに自身がサービス利用を行いたい領域を登録し，この領域以外の部分にサービス利用を禁止する領域 (Mix Zone) を生成する．その後，Mix Zone 内に同時に入ったユーザ同士で ID を入れ替える．これにより，受信した位置情報群の中から，同一ユーザのサービス利用を対応付けることが困難になり，ユーザの連続的なサービス利用の追跡が防止される．これらの手法で用いられている匿名化サーバおよび第三者サーバは，完全に信頼できるものと想定されているが，実環境でこの想定が保証されるとは限らない．

この問題を解決するために，文献 [3, 4, 8, 10, 11] では，偽の位置情報 (ダミー) を用いることで，第三者サーバを使用せずにユーザの位置を曖昧化する手法を提案している．この手法では，図 1 のようにダミーの位置情報を複数生成し，ユーザの位置情報とともにサービスプロバイダに送信する．サービスプロバイダは受信した位置情報それぞれに対応したサービスを提供し，ユーザは自身の位置情報に対応したサービスのみを利用する．サービスプロバイダはユーザとダミーを区別できないため，ユーザの位置情報を識別することが困難になり，ユーザの

(注1)：https://www.google.co.jp/maps

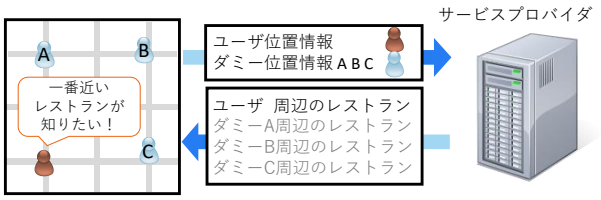


図 1: ダミーを用いた位置曖昧化手法

位置が曖昧化される．また，サービスプロバイダに送信する位置情報の中に必ず自身の位置情報が含まれているため，サービスの質は低下しない．文献 [3] では，ユーザの施設利用を伴う停止時間を考慮し，いくつかの訪問場所で停止しながら移動するユーザを想定したダミーの生成手法が提案されている．しかし，この手法はユーザのトラジェクトリが事前に既知である想定のもとダミーを生成するため，実環境への適用は困難である．そこで文献 [11] では想定を緩和し，ユーザの行動に関する既知の情報が訪問場所の集合のみの条件下でユーザの位置を効果的に曖昧化するためのダミーを生成する手法 (Edge) が提案されている．シミュレーションによる評価実験の結果，文献 [10] で提案されている，同じ想定のもとダミーを生成する手法よりも，Edge の方が効果的にユーザの位置の曖昧化できることが確認されている．

一方，攻撃者はエンティティ(ユーザまたはダミー) の位置情報群を地図上にマッピングし，目視でユーザの識別を試みることが考えられる．そのため，ユーザの位置プライバシを保護するには，ユーザとダミーのトラジェクトリを見た際に，両者を識別できないことが重要である．そこで本稿では，目視によるユーザの識別に対する頑強性を調査する．具体的には，エンティティのトラジェクトリを可視化するシステムを Web 上に構築し，ユーザの識別可能性を検証する視認性評価実験を行った．エンティティのトラジェクトリ群からユーザの位置情報が判別される割合，およびユーザのトラジェクトリを追跡できた割合を調査した結果，Edge は目視によるユーザの識別に対する頑強性があることを確認した．

本稿の構成. 本稿は以下のように構成される．まず，第 2 章で本稿の想定環境およびユーザの位置プライバシの保護に関する要求について述べる．次に，第 3 章で視認性実験に適用した文献 [11] のダミーを生成する手法の概要を説明し，第 4 章で視認性評価実験について述べる．最後に第 5 章で本稿のまとめについて述べる．

2. 事 前 準 備

2.1 想 定 環 境

位置情報サービスの利用. 本稿では，ユーザがサービスプロバイダに頻繁に位置情報を送信する位置情報サービス (例えば GoogleMap) を利用することを想定する．ユーザは，自身が訪問する予定である訪問場所の集合を事前に入力し，文献 [11] の手法を用いてダミーを生成することで，自身の位置情報を保護する．なお，ユーザが所有するモバイル端末は，地図情報を保持しており，ユーザが歩行可能な道路および訪問場所の情報を全て把握しているものとする．

ユーザの行動モデル. ユーザは，いくつかの訪問場所を訪問しながら徒歩で移動する状況を想定する．ユーザは訪問場所に到着した際，施設の利用を行うために最小で pt_{min}[sec]，最大で pt_{max}[sec] の間停止し，訪問場所に移動する際，最短路もしくは目的地に近づく経路を通って v_{min}[m/sec] から v_{max}[m/sec] の範囲の一定の速度で移動する．なお，ユーザが徒歩に加え，車や電車などを併用しながら移動するハイブリッドな移動モデルについては，本稿の対象外とする．

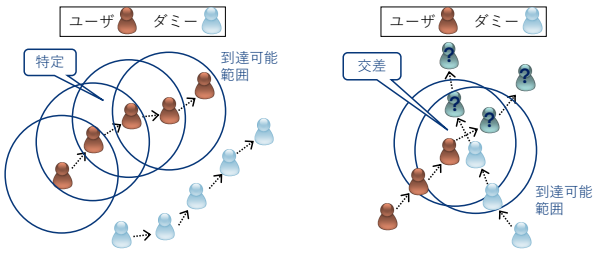
攻撃モデル. 本稿では，悪意のある第三者またはサービスプロバイダが攻撃者であるとし，ユーザの行動モデルに関する情報および地図情報を保持しているものとする．上述のような攻撃者は，特定のユーザの個人情報 (例えば自宅の住所) を取得するために，そのユーザのトラジェクトリの追跡を試みる．具体的には，エンティティのトラジェクトリ群を地図上にマッピングし，目視によりユーザのトラジェクトリの特定を図る．

2.2 ユーザの位置プライバシ保護に関する要求

ユーザの位置プライバシの保護に関する要求として，実環境における行動の制約，追跡可能性，および匿名領域の 3 点について述べる．

実環境における行動の制約. まず，実環境における制約について考える．連続的に位置情報サービスを利用する場合，直前のサービス利用を行った位置からの地理的な到達可能性を考慮する必要がある．つまり，ダミーは，直前のサービス利用時に生成した位置から移動可能な範囲内に生成しなければならない．なお，2 点間の距離を計算する際は，ユークリッド距離ではなく，実際の道路ネットワークを考慮して距離を計算する必要がある．さらに，実環境では海上や高速道路など，歩行するユーザが存在し得ない領域が存在する．このような領域にダミーを生成すると，それがダミーであると容易に特定されてしまう．そのため，ダミーは実際のユーザが存在し得る領域に生成する必要がある．

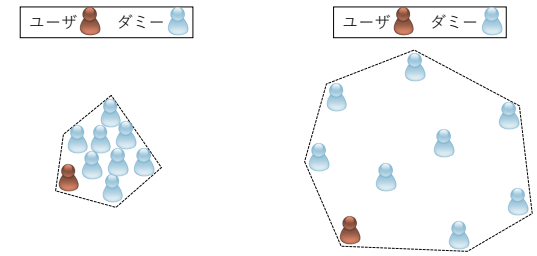
追跡可能性. ユーザが連続して位置情報サービスを利用する場合，ユーザのサービス利用が追跡される可能性を考慮しなければならない．これは，サービスプロバイダに蓄積されている位置情報の履歴から，ユーザのトラジェクトリが推定される危険性があるためである．本稿では，文献 [7] を基に，ある区間内の連続した位置情報の組み合わせにより，ユーザのトラジェクトリが追跡されてしまう可能性を追跡可能性と定義する．追跡可能性は，ユーザの位置が偶発的に特定された時 (例えば，ユーザが無意識に自身の位置を示す写真などをソーシャルネットワーキングサービス (SNS) にアップロードした際) に大きな問題となる．もし，ユーザの位置が追跡されてしまうと，過去 (ともすれば将来) の位置も明らかになり，より多くの情報流出を招く．例えば，図 2a に示すように，ユーザの各地点から到達可能な範囲内に，ユーザのみが包含されている場合，ユーザとダミーのトラジェクトリの識別が容易になってしまう．追跡可能



(a) ユーザの移動可能な範囲内にユーザのみが包含される場合，ユーザのトラジェクトリを容易に追跡できる．

(b) ユーザとダミーのトラジェクトリが交差すると，追跡可能性が低下する．

図 2: 追跡可能性



(a) 位置曖昧度が低い状態

(b) 位置曖昧度が高い状態

図 3: 匿名領域

性を低下させるためには，図 2b に示すように，ユーザのトラジェクトリと交差するようにダミーのトラジェクトリを設計する方法が有効であり，ユーザおよびダミーの位置が同じ到達可能な範囲内に存在すると，ユーザのトラジェクトリを追跡することは困難である．

匿名領域. ユーザの位置プライバシを保護するためには，ユーザの位置がどの程度の大きさの領域に曖昧化されているのかも重要である．そこで，文献 [4] に基づき，ユーザの位置がどの程度匿名化されているかを示す指標である匿名領域を定義する．

定義 1 (匿名領域). 現在のサービス利用時点におけるユーザおよびダミーの位置が与えられた時，匿名領域を全てのエンティティで形成される凸包領域と定義する．

図 3a のようにユーザ付近に密集してダミーを配置した場合よりも，図 3b のように広い範囲にダミーを配置した方が位置曖昧度が高い．つまり，匿名領域は図 3b の方が図 3a よりも大きい．なお，匿名領域の適切な大きさはサービスを利用する状況に依存するため，ユーザ (もしくはアプリケーション) が匿名領域の大きさを指定するものと想定する．

3. 文献 [11] の紹介

本章では，文献 [11] において提案されているダミー生成手法 (Edge) について紹介する．

ユーザの行動に関する既知情報. Edge では，ユーザの行動に関する既知の情報は，ユーザの訪問場所 vp_i の集合 $VP = \{vp_1, vp_2, \cdots, vp_n\}$ のみである．つまり，ユーザはサービスの利用を行う前に VP を入力すると想定している．なお，ユーザは入力した全ての vp_i に必ず 1 度ずつ訪問するものとし，ダミーの初期生成は，ユーザが最初の訪問場所に到着した際に行うものとする．

ダミーのトラジェクトリの生成方法の概要. 連続的な位置情報サービスの利用を考慮すると，ユーザのトラジェクトリが追跡されないことが重要である．Edge は，追跡可能性を低下させるためのダミーのトラジェクトリを生成し，その後，匿名領域拡大のためのダミーのトラジェクトリを生成する．

ここで，ユーザの行動プランに関する入力が訪問場所の集合のみという条件下では，ユーザの移動先が既知ではない．そのため Edge では，巡回セールスマン問題を用いて，ユーザの訪問場所の訪問順序の候補をリスト構造として取得し，各訪問順序の候補の移動距離に基いて，ユーザの移動先の候補を推定する．例えば，ユーザが n 個の訪問場所を入力するものとすると，リスト (訪問順序の候補) のあるエントリ (訪問順序) は，$vp_1 \rightarrow \cdots \rightarrow vp_{n-1} \rightarrow vp_n$ と表せる．これは，ユーザが vp_1, ..., vp_{n-1}, vp_n の順番で訪問場所を訪問することを意味する．各エントリは移動距離を保持しており，リストは移動距離の昇順でソートされているものとする．この時，リストの上位 α 個のエントリ，つまり，移動距離が短い上位 α 個のエントリを，訪問場所の訪問順序の候補として選択する．その後，推定した訪問順序の候補に基いて，ユーザと交差を行うダミーのトラジェクトリを α 個生成する．ここで，$\langle vp_{i+1}^{\alpha'}, vp_{i+2}^{\alpha'} \rangle$ を，ユーザが i 番目の訪問場所に到着した際の，α' 番目のエントリの $(i+1)$ 番目および $(i+2)$ 番目の訪問場所の順序対とする．Edge は，α 個のダミーのトラジェクトリを，それぞれ vp_{i+1}^1, ..., $vp_{i+1}^{\alpha'}$, ..., および vp_{i+1}^{α} を訪問するように生成する．これは，直感的には，ユーザの次の移動先が α 個の訪問場所の候補に含まれる場合，α 個のダミーのうち，1 つのダミーはユーザのトラジェクトリと交差できるというアイデアに基づく．

また，追跡可能性を低下させるためには，ユーザとダミーのトラジェクトリの交差が発生する間隔が短いことが重要である．これは，交差の間隔が長いと，ユーザのトラジェクトリが追跡される時間が大きくなるためである．そこで Edge は，ユーザの次の訪問場所の候補 (α 個) のそれぞれの次の訪問場所である vp_{i+2}^1, ..., $vp_{i+2}^{\alpha'}$, ..., および vp_{i+2}^{α} を訪問するようにダミーのトラジェクトリを生成する．つまり，ユーザのトラジェクトリと交差を行うダミーのトラジェクトリを 2α 個生成する．これにより，複数の訪問順序の候補への対応および追跡可能性の十分な低下を図る．なお，α はユーザが入力するものとする．α が大きいほど，推定するユーザの移動先の候補の数が増加するため，ユーザとダミーのトラジェクトリが交差する可能性が大きくなる．

さらに，Edge は，ユーザを特定するヒントの低減および匿名領域の拡大のため，β 個のダミーのトラジェクトリを生成する．ここで，ユーザのトラジェクトリと交差させるために，上述の方法を用いて 2α 個のダミーのトラジェクトリを生成し，匿名領域の拡大のためのみに，β 個のダミーのトラジェクトリを生成する手法を考える．この手法を適用してダミーのトラジェク

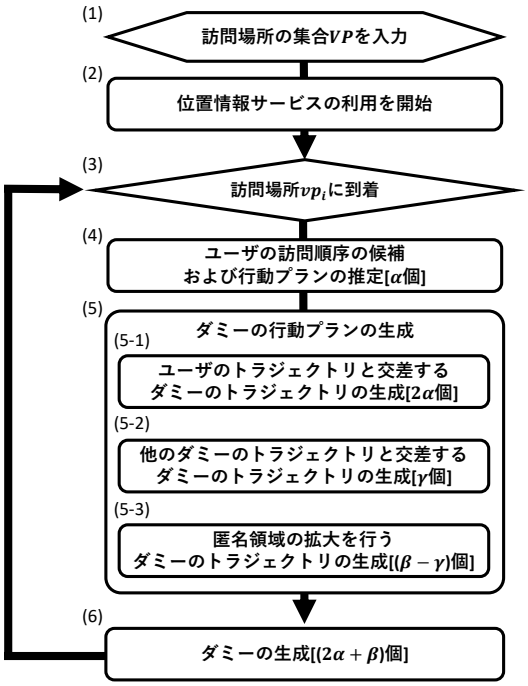


図 4: Edge におけるダミーの生成までの主な手順

トリを生成した場合，交差の対象となるエンティティがユーザのみであるため，それがユーザを特定するヒントになる．これは，あるエンティティが他のエンティティと頻繁に交差をしていると特定できた場合 (注2)，その情報を基にユーザと判断される可能性があるためである．これを防ぐため，Edge は，β 個のダミーのトラジェクトリの一部を他のダミーのトラジェクトリと交差を行うように生成する．なお，β もユーザが指定できるパラメータであり，β が増えるほど，ダミーのトラジェクトリ同士の交差が増えたり，エンティティが広い範囲に分布する．

以上の方法を用いて，ユーザが訪問場所に到着する度に，$m(=2\alpha+\beta)$ 個のダミーのトラジェクトリを逐次的に生成することで，ユーザの移動状況に柔軟に対応し，効果的に位置を曖昧化する．

要約. 図 4 に，Edge においてダミーを生成するまでの主な手順を示す．ユーザが，(1) 訪問場所の集合 VP を入力し，(2) 位置情報サービスの利用を開始する．(3) ユーザが訪問場所 vp_i に到着すると，(4) 始点が最初の訪問場所である巡回セールスマン問題の解に基づき，ユーザの訪問順序の候補および行動プランを α 個推定する．その後，(5) 推定したユーザの行動プランを基にダミーの行動プランの生成する．具体的には，初めに，ユーザのトラジェクトリと交差を行うダミーのトラジェクトリを 2α 個生成し (5-1)，次に，別のダミーのトラジェクトリと交差が発生するダミーのトラジェクトリを γ 個生成する (5-2)．最後に，匿名領域が拡大するように，残りの $(\beta-\gamma)$ 個のダミーのトラジェクトリを生成する (5-3)．そして，(6) 生成したダミーの行動プランに従って，$(2\alpha+\beta)$ 個のダミーを生成する．これらの操作をユーザのサービス利用が終了するまで繰り返し行う．

(注2)：この想定は，各エンティティの交差回数を正確に数えることができる場合にのみ有効だが，実際はエンティティのトラジェクトリ同士が交差するため，容易ではない．

図 5: 判別実験のシステムのインタフェース．各エンティティの位置情報が青色の丸印 (ノード) により表示される．

4. 視認性評価実験

本章では Edge において，目視によるユーザの位置情報の識別に対する頑強性を調査した結果を紹介する．各エンティティのトラジェクトリを可視化するシステムを Web 上に構築し，エンティティのトラジェクトリ群からユーザまたはダミーの位置情報をどの程度判別できるかを検証する判別実験，およびユーザのトラジェクトリをどの程度追跡できるかを検証する追跡実験を実施した．

4.1 実験概要

判別実験. 判別実験では，エンティティのトラジェクトリ群から目視によってどの程度ユーザまたはダミーを判別できるかを検証した．図 5 に，実験で使用したシステムのインタフェースを示す．各青色の丸印 (ノード) は，ある時点における各エンティティの位置情報を示す．各被験者には，エンティティのノードの遷移を見てもらい，ある時点においてユーザまたはダミーであると自信を持って判断できるノードが存在していた場合，そのノードをクリックしてもらった．なお，自信を持って判断しているかを確認するため，判断した根拠をフリーワードで記載してもらった．被験者が円滑に実験できるようにするため，実験画面には，各エンティティのノードの再生，巻き戻し，および再生する速度の調整ができる機能を用意し，被験者にはこれらの機能を用いながら実験を実施してもらった．

追跡実験. 追跡実験では，ユーザのトラジェクトリを目視によってどの程度追跡できるかを検証した．図 6 に，追跡実験の手順を示す．図 6a のように，ユーザのノードだけを別の色 (橙色) で表示することでユーザの位置情報を明示し，実験を開始する．

(a) ユーザのノードだけを別の色 (橙色) で表示することでユーザの位置を明示し，実験を開始する．

(b) 次のサービス利用時点において生成される各エンティティのノードが全て同じ色で表示される．各被験者は，このノード群の中からユーザの位置情報を推測する．

図 6: 追跡実験の手順

表 1: 各パラメータの設定 (太字はデフォルト値)

パラメータ	値の範囲
r[m]	1,000, 1,250, **1,500**, 1,750, 2,000
α	**2**, 3, 4
β	3, 4, **5**, 6, 7
T[h]	0.5, **1.0**, 1.5, 2.0, 2.5, 3.0
サービス利用間隔 [sec]	**180**
v_{min}[m/sec]	**1.05**
v_{max}[m/sec]	**1.55**
t_{min}[sec]	**600**
t_{max}[sec]	**1,800**

実験を開始すると，次のサービス利用時点において生成される各エンティティのノードが表示される．この際，図 6b のように，全てのエンティティのノードを同じ色 (青色) で表示し，ユーザとダミーのノードの見分けがつかないようにする．その状態で，各被験者には直前までのノードの遷移を基にユーザのノードを推測し，そのノードをクリックしてもらった．もし正しくユーザのノードをクリックできた (推定が正しかった) 場合，ユーザのトラジェクトリを追跡できていると判断し，実験を続行する．誤ってダミーのノードをクリックしてしまう，もしくはユーザのトラジェクトリを最後まで追跡するまで上記の操作を繰り返し行なってもらった．

4.2 設　　定

東京 23 区の地図データおよび FourSquare API(注3) から取得した訪問場所の情報を基に，実際のユーザのサービス利用を再現した．また，各パラメータは表 1 のように定めた．r は要求匿名領域の面積 (簡単化のため，匿名領域の形状は正方形であると想定)r^2[m^2] の正方形の一辺の長さを表し，T はユーザのトラジェクトリの時間の長さを示す．なお，各パラメータのデフォルト値は太字で示している．

ユーザの移動データとして，Foursquare のチェックインデータセットを用いた [9]．ただし，このデータセットには，本稿で想定しない公共交通機関や自動車を用いた移動が含まれていたため，歩行のみによる行動プランに変換したものをユーザのトラジェクトリとして使用した．

(注3)：https://developer.foursquare.com

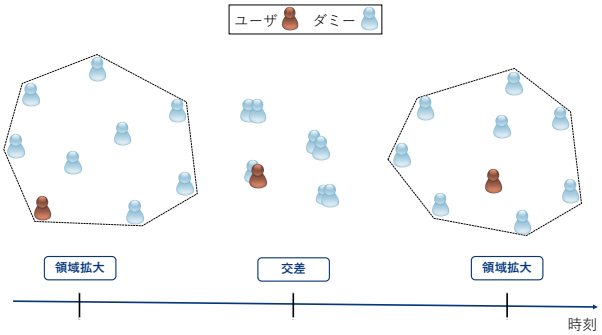


図 7: Dum-P-Cycle におけるダミーのトラジェクトリを生成する方法のイメージ図．ある一定の周期で匿名領域の拡大およびエンティティのトラジェクトリ同士の交差が発生する．

被験者として，同一の研究室に所属する 21 才から 28 才までの学部生および大学院生 23 名に実験に参加してもらった．なお，各被験者の各実験の試行回数の幅は，判別実験が [20, 140] 回であり，追跡実験が [28, 220] 回だった．また，各実験の合計の試行回数はそれぞれ，判別実験が 1,078 回，追跡実験が 2,335 回だった．

4.3 評 価 手 法

各実験において，ダミーを生成する手法として以下の 3 つを適用し，性能を比較した．なお，各パラメータ下において，各手法で生成するダミーの数は同じである．

Dum-P-Cycle [3]. ユーザが，訪問場所の集合，訪問順序，移動経路，移動速度，および停止時間で構成される行動プラン全てを事前に入力し，その行動プランに完全に従って行動すると想定し，ダミーを生成する手法．つまり，ユーザのトラジェクトリを事前に把握できるため，実際のユーザのトラジェクトリを基にダミーのトラジェクトリを生成する．図 7 に，Dum-P-Cycle におけるダミーのトラジェクトリを生成する方法の直感的な例を示す．図 7 のように，実際のユーザのトラジェクトリを基に，エンティティの広い範囲の分布およびエンティティ同士のトラジェクトリの交差が周期的に発生するようにダミーのトラジェクトリを生成する．なお，この手法では，上述の通り正確なユーザの行動プランを事前に入力として与えるため，実環境での適用は困難であることから，純粋な比較対象とはせず，性能評価のための参考手法とする．

E-Dum-P-Cycle（Estimation-based-Dum-P-Cycle）[10]. Edge と同様に，ユーザの行動プランに関する既知の情報はユーザの訪問場所の集合 *VP* のみと想定し，ダミーのトラジェクトリを生成する手法．まず，訪問順序を巡回セールスマン問題の最適解と予測し，予測したユーザに対して Dum-P-Cycle を適用する．その後，毎サービス利用時に，実際のユーザの移動状況と予測したユーザの差異を判断し，その変化に応じて，ダミーのトラジェクトリを動的に修正する．

Edge [11]. 3 章で紹介した手法．

4.4 評価指標

各実験において用いた評価指標について述べる．ただし，各パラメータにおいて実施された実験の数を N，あるデータセットにおいて正しく推定されたユーザのノードの数を TP_i，正しく推定されたダミーのノード数を TN_i，ユーザであると推測されたが実際はダミーだったノードの数を FP_i，ダミーであると推測されたが実際はユーザだったノードの数を FN_i，および生成されたダミーの数を m_i とする．

判別実験における評価指標. 判別実験における評価指標として，以下で定義するユーザ検出率およびダミー検出率を用いた．

- **ユーザ検出率.** ユーザ検出率は，以下の式で定義される．

$$\text{ユーザ検出率} = \frac{1}{N}\sum_{i=1}^{N} TP_i$$

つまり，ユーザ検出率は各データセットに対するユーザのノードが判別された数の割合を示す．

- **ダミー検出率.** ダミー検出率は，以下の式で定義される．

$$\text{ダミー検出率} = \frac{1}{N}\sum_{i=1}^{N} \frac{TN_i}{m_i}$$

つまり，ダミー検出率は各データセットのダミーの数に対するダミーのノードが判別された数の割合を示す．

追跡実験における評価指標. 追跡実験における評価指標として，以下で定義する追跡割合を用いた．

- **追跡割合.** あるデータセットにおいて被験者が追跡できたユーザのトラジェクトリの時間の長さを τ_i とすると，追跡割合は以下の式で定義される．

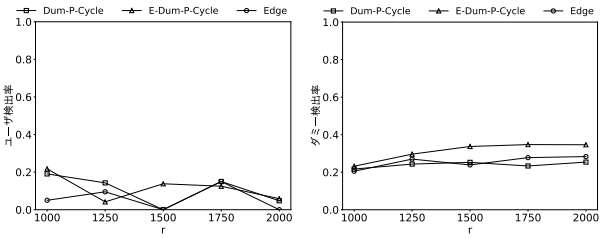
$$\text{追跡割合} = \frac{1}{N}\sum_{i=1}^{N} \frac{\tau_i}{T_i}$$

つまり，追跡割合は各データセットのトラジェクトリの時間の長さに対する被験者が追跡できたトラジェクトリの時間の長さの割合を示す．

4.5 判別実験の実験結果

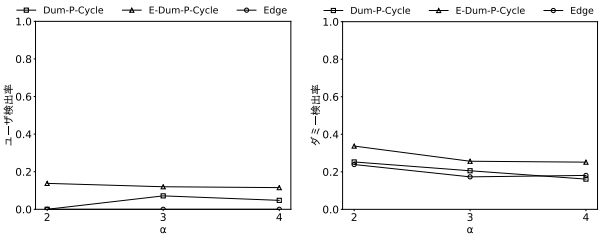
r の影響. 図 8a は，r に対するユーザ検出率の変化を示している．図 8a より，各 r において，各手法とも値は約 0.2 以下であり，ユーザ検出率を小さく抑えられていることが分かる．図 8b は，r に対するダミー検出率の変化を示している．図 8b より，各 r において，Edge および Dum-P-Cycle は E-Dum-P-Cycle よりも値が小さく，Edge はユーザの行動に関する想定が同じ E-Dum-P-Cycle よりもダミー検出率を小さく抑えられていることが分かる．また，Edge および E-Dum-P-Cycle において，r が大きい時の方が小さい時よりも若干値が大きい．r が大きいと，エンティティが広い範囲に分布するようにダミーを配置する必要がある．この時，エンティティ群から孤立するような動きに映るノードが増えることにより，それが理由でダミーであると被験者が推測しやすくなっていた．

α の影響. 図 9a は，α に対するユーザ検出率の変化を示している．図 9a より，各 α において，Edge が最もユーザ検出率が小さいことが分かる．図 9b は，α に対するダミー検出率の変化を示している．各 α において，r の場合と同様に，Edge は E-Dum-P-Cycle よりもダミー検出率を小さくできている．



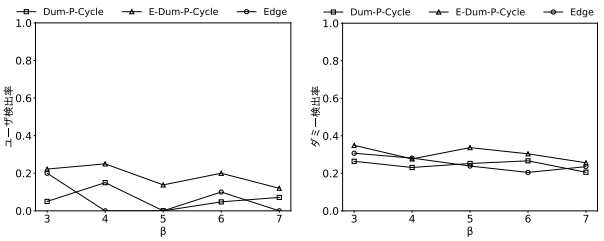
(a) ユーザ検出率　(b) ダミー検出率

図 8: r vs. ユーザ検出率およびダミー検出率



(a) ユーザ検出率　(b) ダミー検出率

図 9: α vs. ユーザ検出率およびダミー検出率

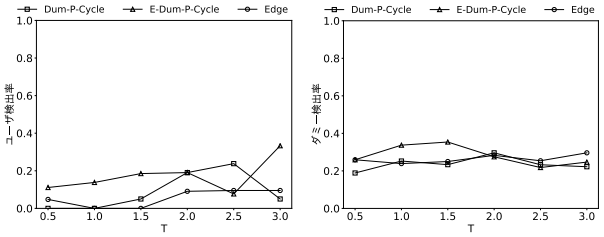


(a) ユーザ検出率　(b) ダミー検出率

図 10: β vs. ユーザ検出率およびダミー検出率

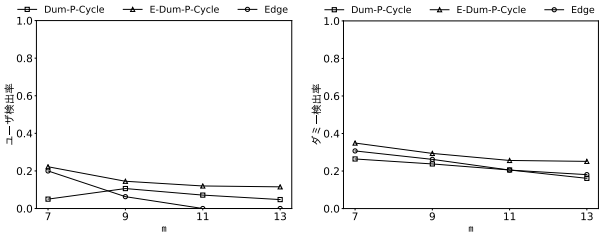
β の影響. 図 10a は，β に対するユーザ検出率の変化を示している．図 10a より，各 β において，Edge は E-Dum-P-Cycle よりもユーザの検出率を小さくできており，ユーザの位置情報の判別が困難であることが分かる．Edge では，他のダミーのトラジェクトリとの交差および匿名領域の拡大をするダミーを β 個生成することにより，ユーザの判別を困難にしている．図 10b は，β に対するダミー検出率の変化を示している．図 10b より，Edge は E-Dum-P-Cycle よりもダミー検出率を小さく抑えられていることが分かる．

T の影響. 図 11a は，T に対するユーザ検出率の変化を示している．図 11a より，各手法において，T が小さい時 ($0.5 \leqq T \leqq 1.5$) に比べ，T が大きい時 ($2.0 \leqq T \leqq 3.0$) の方が全体的にユーザ検出率が大きい．これは，ユーザのトラジェクトリの時間の長さが大きくなるにつれて，ユーザの位置情報を判別するためのヒントが多くなるためである．図 11b は，T に対するダミー検



(a) ユーザ検出率　　(b) ダミー検出率

図 11: T vs. ユーザ検出率およびダミー検出率



(a) ユーザ検出率　　(b) ダミー検出率

図 12: m vs. ユーザ検出率およびダミー検出率

出率の変化を示している．各手法において，ダミー検出率を約 0.4 以下に抑えられていることが分かる．

m の影響. 図 12a は，m に対するユーザ検出率の変化を示している．図 12a より，各手法において，m が大きくなるにつれてユーザ検出率が減少傾向であることが分かる．これは，ダミーの数が増えるにつれてユーザの候補の数が増えることから，ユーザを判別することが困難になるためである．また，$9 \leqq m$ の場合，Edge のユーザ検出率が最も低いことが分かる．Edge は Dum-P-Cycle に比べ，ユーザの行動プランに関する既知の情報が少ないにも関わらず，ユーザの検出率を小さくできる．図 12b は，m に対するダミー検出率の変化を示している．図 12b より，Edge は E-Dum-P-Cycle より値を小さくできていることが分かる．

ここで，各被験者には自信を持ってユーザもしくはダミーであると判断できるノードが存在していた場合，そのノードをクリックしてもらったが，その回答の精度を検証するために，ユーザ正答率およびダミー正答率を定義する．

- **ユーザ正答率.** ユーザ正答率は以下の通り定義される.

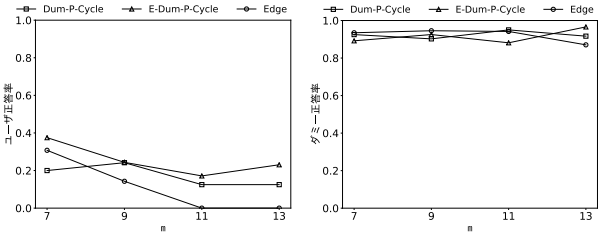
$$\text{ユーザ正答率} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i}$$

ユーザ正答率は被験者がユーザであると推測したノードの数に対して実際にそのノードがユーザだった割合を示す.

- **ダミー正答率.** ダミー正答率は以下の通り定義される.

$$\text{ダミー正答率} = \frac{1}{N}\sum_{i=1}^{N}\frac{TN_i}{TN_i + FN_i}$$

ダミー正答率は被験者がダミーであると推測したノードの数に対して実際にそのノードがダミーだった割合を示す.



(a) ユーザ正答率　　(b) ダミー正答率

図 13: m vs. ユーザ正答率およびダミー正答率

図 13a は，m に対するユーザ正答率の変化を示している．図 13a より，値は大きい場合でも約 0.4 であり，被験者が自信をもってユーザであると判断したにも関わらず，正答率は 50%にも満たない．図 13b は，m に対するダミー正答率の変化を示している．ダミーは，ランダムに選んだ場合でも $\frac{m}{m+1}$(例えば $m = 9$ の場合は 0.9) の割合で正解できるが，値は 1 を満たしておらず，自信をもってダミーであると判断した場合でも，確実にダミーを判別できるわけではない (ユーザと間違える) ことが分かる．

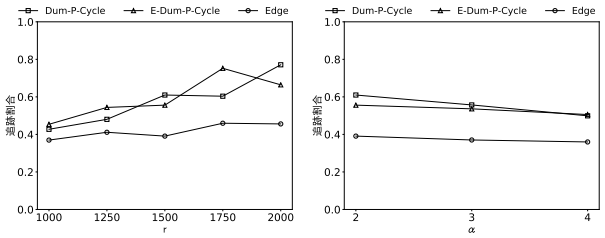
4.6 追跡実験の実験結果

r の影響. 図 14a は，r に対する追跡割合の変化を示している．図 14a より，Edge は追跡割合を最も低下できていることが分かる．Edge では追跡可能性を低下させるために，ユーザの 2α 個の訪問場所で交差が発生するようにダミーのトラジェクトリを生成しているため，ユーザのトラジェクトリと交差が発生する確率が高く，また，交差の間隔も短いため，追跡割合を小さくできる．また，各手法において，r が増加するにつれて追跡割合が増加傾向であることが分かる．r が大きいと，ユーザは大きい匿名領域を必要としていることを意味するが，これを満たすために，ダミーを広い範囲に分布させなければならない．この時，エンティティの距離が大きくなる傾向がある．これにより，ユーザと交差するための時間が大きくなり，追跡割合が大きくなる．

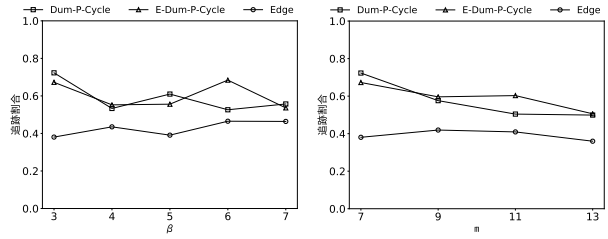
α の影響. 図 14b は，α に対する追跡割合の変化を示している．図 14b より，Edge が最も追跡割合を低下できていることが分かる．また，α が増加するにつれて追跡割合は減少傾向であるが，Edge では，α が大きくなるほどユーザのトラジェクトリと交差する可能性が大きくなるためである．

β の影響. 図 14c は，β に対する追跡割合の変化を示している．図 14c より，r および α の場合と同様に，Edge の追跡割合が最も小さい．なお，Edge では，β が大きくなるに従い，他のダミーのトラジェクトリと交差するダミーおよび匿名領域を拡大するダミーが増加するが，ユーザとの交差には直接的に寄与しないため，β の変化は追跡割合に大きく影響しない．

m の影響. 図 14d は，m に対する追跡割合の変化を示している．m が大きくなるほど追跡割合が減少傾向であることが分かる．これは，ダミーの数が大きくなるにつれてユーザのトラジェクトリと交差するダミーのトラジェクトリの数が多くなることで，

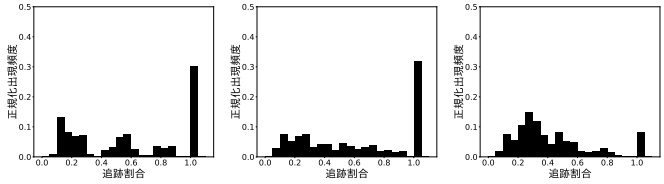


(a) 追跡割合 vs. r　　(b) 追跡割合 vs. α



(c) 追跡割合 vs. β　　(d) 追跡割合 vs. m

図 14: 追跡割合の各パラメータの影響



(a) Dum-P-Cycle　　(b) E-Dum-P-Cycle　　(c) Edge

図 15: 各評価手法の追跡割合のヒストグラム ($m = 9$)

ユーザのトラジェクトリの追跡が困難になるからである．

図 15 は，$m = 9$ における各評価手法の追跡割合のヒストグラムを示している．図 15a より，1.0 の値を除き，Dum-P-Cycle のヒストグラムの形状には凹凸があることが分かる．4.3 節で述べたように，Dum-P-Cycle では領域の拡大およびトラジェクトリ同士の交差の発生を周期的に発生させており，交差が発生するフェーズにおいて，被験者が追跡を失敗しているためである．図 15b より，E-Dum-P-Cycle のヒストグラムの形状は平型になっていることが分かる．E-Dum-P-Cycle では，Dum-P-Cycle で生成したダミーのトラジェクトリを基にユーザの移動状況に応じて適宜ダミーのトラジェクトリを修正しており，交差が発生するフェーズが各ユーザのトラジェクトリによって異なるためである．また，図 15c より，Edge のヒストグラムの形状はなだらかな山型になっていることが分かる．Edge では，2α 個のユーザの訪問場所で交差が発生するようにダミーのトラジェクトリを生成しており，各トラジェクトリの初期の方でも追跡を困難にしていることが分かる．一方，1.0 の値は完全にユーザのトラジェクトリが追跡された割合を示す．ユーザとダミーのトラジェクトリの交差が発生すると，大抵はユーザのトラジェクトリの追跡が困難になるが，偶然ユーザのノードを当てることができたり，直前までのノードの遷移の進行方向に基づく判断により，ユーザの追跡に成功する場合がある．Edge において完全に追跡できた割合は他の 2 つの手法の約 1/3 であり，ユーザとダミーのトラジェクトリの交差を効果的に発生させることで，その割合を低下できることが分かる．

5. おわりに

本稿では，ダミーを用いたユーザ位置曖昧化手法に対して，目視によりユーザのトラジェクトリの識別がどの程度可能かを検証する視認性実験を行った．具体的には，各エンティティのトラジェクトリを可視化するシステムを Web 上に構築し，エンティティのトラジェクトリ群からどの程度ユーザ (またはダミー) の位置情報を判別できるかを検証する判別実験，およびユーザのトラジェクトリをどの程度追跡できるかを検証する追跡実験を実施した．実験の結果，ユーザの行動に関する既知の情報が訪問場所の集合のみの条件下でダミーを生成する手法である Edge は，目視によるユーザの識別に対して頑強性があることを確認した．

謝　辞

本稿の一部は，文部科学省科学研究費補助金・基盤研究 (A)(JP26240013) および挑戦的萌芽研究 (JP16K12429) の研究助成によるものである．評価で利用した東京 23 区の地図データは一般財団法人日本デジタル道路地図協会より貸与されたものである．ここに記して謝意を示す．

文　献

[1] Beresford, A. R. and Stajano, F.: Location privacy in pervasive computing, *PerCom*, No. 1, pp. 46–55 (2003).

[2] De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M. and Blondel, V. D.: Unique in the crowd: The privacy bounds of human mobility, *Scientific Reports*, Vol. 3 (2013).

[3] Kato, R., Iwata, M., Hara, T., Arase, Y., Xie, X. and Nishio, S.: User location anonymization method for wide deistribution of dummies, *DEXA*, pp. 259–273 (2013).

[4] Lu, H., Jensen, C. S. and Yiu, M. L.: Pad: privacy-area aware, dummy-based location privacy in mobile services, *MobiDE*, pp. 16–23 (2008).

[5] Mokbel, M. F., Chow, C.-Y. and Aref, W. G.: The new casper: Query processing for location services without compromising privacy, *VLDB*, pp. 763–774 (2006).

[6] Palanisamy, B. and Liu, L.: MobiMix: Protecting location privacy with mix-zones over road networks, *ICDE*, pp. 494–505 (2011).

[7] Shokri, R., Freudiger, J., Jadliwala, M. and Hubaux, J.-P.: A distortion-based metric for location privacy, *WPES*, pp. 21–30 (2009).

[8] Suzuki, A., Iwata, M., Arase, Y., Hara, T., Xie, X. and Nishio, S.: A user location anonymization method for location based services in a real environment, *GIS*, pp. 398–401 (2010).

[9] Yang, D., Zhang, D., Zheng, V. W. and Yu, Z.: Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 45, No. 1, pp. 129–142 (2015).

[10] 林田秀平，水野聖也，天方大地，原　隆浩，Xing Xie, ：ユーザの移動状況に適応したダミーによる位置曖昧化手法，情報処理学会マルチメディア, 分散協調とモバイルシンポジウム 論文集, pp. 328–335 (2016).

[11] 林田秀平，天方大地，原　隆浩，Xing Xie, ：ユーザの移動先候補の推定に基づく逐次的なダミー生成による位置曖昧化手法，情報処理学会マルチメディア，分散，協調とモバイルシンポジウム　論文集，pp. 332–339 (2017).