Vol.2017-DPS-171 No.13
Vol.2017-MBL-83 No.13
Vol.2017-ITS-69 No.13
2017/6/1

k-匿名性とl-多様性による プライバシに配慮した車両データ収集・開示手法の検討

岡部 友介[1,a)]　五味 和良[1,b)]　重野 寛[1,c)]

概要：近頃，様々な情報システムから得られる膨大な情報を活用することで，我々の生活をより豊かにするための取り組みが広くなされている．本稿にて提案するシステムは車両から取得できるデータである，目的地/出発地等の走行情報と車種情報のプライバシに配慮した収集・開示手法を検討する．これらの情報は交通流の解析を目的とした交通機関にとって有用である．一方，走行情報や車種の特定によりユーザのプライバシが侵害される恐れがある．この問題に対し，プライバシを保護しつつデータを公開するための技術 Privacy-Preserving Data Publishing の研究が盛んである．代表的なプライバシ保護基準として，k-匿名性や l-多様性が知られている．本システムは位置プライバシを保護するために，k-匿名性のプライバシ基準が保証されるまで情報の空間的，時間的精度を低下させ，少なくとも k 人が含まれる共通の走行情報を持つグループを車車間通信にて形成する．車両はそれらの走行情報を秘密分散法に基づいて安全にアップロードする．また，準同型暗号を用いた秘匿集計により，車種において l-多様性のプライバシ保護を保証した開示を実現する．シミュレーション評価では，プライバシ保護が中央のデータ収集者に依存せず，ユーザ同士で分散的実施される車両データの収集・開示が実際に可能であることを示す．

1. はじめに

インターネットを介してあらゆるモノがつながり，大量の情報を入手できる今，ビッグデータの利活用が注目されている．本稿にて提案するシステムは，車両データとして走行情報と車種情報を収集する．アップロードされたデータに個人を区別することができる識別子が含まれておらず，同じ車両の異なる走行記録が関連付けられない場合，車両の出発地または目的地に関する情報は推測できないため，ドライバー個人のプライバシには影響しない．しかし，大規模な交通流データ解析を目的とした車両データの収集において，これに含まれる位置情報等の分析がユーザのプライバシを侵害する可能性がある．実際に個人の識別子を取り除いたとしても，詳細な位置追跡や出発地/目的地（OD）のペアを使用してドライバーの自宅の位置，またそれらの身元を識別できることが示されている [1][2][3]．そのため，出発地/目的地等のセンシティブな情報を取り扱う場合，ユーザのプライバシを保護したデータ収集・開示が必要となる．

位置プライバシを保護するための一般的な方法として，特定のプライバシ基準が保証されるまで意図的に情報の空間的，時間的精度を低下させる手法が存在する [4]．また，プライバシ保護の度合いを定量的に表すプライバシ基準の代表例として k-匿名性 [5] が挙げられる．あるユーザによって提示されたデータが他の $k-1$ 人のユーザのものと区別することができない，つまりデータの持ち主を k 人以下に絞り込めないとき，そのユーザは k-匿名である．

ユーザからアップロードされた全てのデータが信頼できるデータ収集者のデータベースによって安全に管理される場合，データ収集者がそれらのデータに処理を施すことで k-匿名性のプライバシ基準を満たしたデータ開示を実現することができる．しかし，データベースの管理を担うデータ収集者が必ずしも信用できるとは限らない．また，このようなセンシティブ情報を含むデータベースはハッキング等の標的になる恐れがある．したがって，プライバシ保護がデータ収集者 （中央データベース） に依存せず，ユーザ同士で分散的に実施されるシステムが必要である．

本稿はプライバシに配慮した車両データの収集・開示手法を提案する．車両は走行情報を秘密分散法に基づいて安全にアップロードし，k-匿名性のプライバシ保護を達成する．また，準同型暗号を用いた秘匿集計により，車種において l-多様性 [6] のプライバシ保護を保証した開示を実現する．また，ルクセンブルクの現実的な交通環境を用いた

[1] 慶應義塾大学大学院理工学研究科
[a)] okabe@mos.ics.keio.ac.jp
[b)] gomi@mos.ics.keio.ac.jp
[c)] shigeno@mos.ics.keio.ac.jp

シミュレーション評価を行う．評価ではプライバシ保護手法である k-匿名性に加え，k-匿名性を強化し属性推定を防ぐ l-多様性を車両同士が分散的に実施し，プライバシに配慮した車両データの収集・開示が実際に可能であることを示す．

以下，2 章で関連研究について説明する．3 章では提案システムについて述べ，4 章ではその提案システムの評価と考察を行う．最後に，5 章で結論を述べる．

2. 関連研究

本研究の関連研究として，Förster らによって提案されたシャミァの秘密分散法 [7] を利用した，k-匿名性を保証した車両の OD 情報の開示手法が存在する [8]．当研究は複数の車両が V2X 通信を用いたキー交換により同一の OD 情報を作成することでプライバシ保護を実現している．本稿では車両からより多様な情報を収集し，またプライバシ保護との両立を目指したシステムを提案する．以下，関連技術について説明する．

2.1 秘密分散法を用いた k-匿名性の分散的実施

本節では，Förster らによって提案された OD 情報開示手法について説明する．当研究はシャミァの秘密分散法を利用し，データの k-匿名性を保証している．秘密分散法とは，複数のユーザによって共有される共通の秘密（s）を分割保存する技術である．また，秘密を分割したものを $share$ と呼ぶ．当研究では，複数の車両に共有される同一の OD 情報に関するキー，$trip\_key$ を共通の秘密（s）としている．

2.1.1 秘密分散法

秘密分散法について説明する．車両は $f(0) = s$ となるランダム係数の $k-1$ 次多項式 $f(x)$ を構築する．

$$a_i := h(i \parallel s)\ for\ i \in [1, k] \tag{1}$$

式 1 のランダム係数 a_i を用いて，

$$f(x) = s + \sum_{i=1}^{k-1} a_i x^i \tag{2}$$

式 2 を得る．同じ精度レベルの走行情報を持つ車両が共通の多項式 $f(x)$ を構築する．次に，各車両は秘密（s）を分散させる $share$ を計算する．$share$ とは，$k-1$ 次多項式 $f(x)$ 上の座標の一点 $(x_r, f(x_r))$ である．秘密（s）を共有する n $(n > k)$ 台の車両はそれぞれランダムに x_r を選び，自身の $share$ を計算する．x_r は車両同士での値の重複を防ぐために大きな範囲から選択される．n 台の車両はそれぞれ計算した $share$ を匿名ネットワークを用いて収集者のデータベースへアップロードする．このとき，分割数 n のうち任意の k 個の $share$ のアップロードで秘密（s）が判明する．

2.1.2 k-匿名性を保証した走行情報の開示

車両は作成した OD 情報を対応する $trip\_key$ を用いて暗号化する．OD 情報を $trip\_data$ とし，車両は収集者のデータベースにアクセスし，$ID(trip\_key)$，$share$，$ENC_{trip\_key}(trip\_data)$ の 3 つの情報をアップロードする．秘密分散法を用いた k 個の $share$ のアップロードにより $trip\_key$ が再構築される．$trip\_key$ が再構築されることにより，同じ $trip\_key$ で暗号化された k 個の $trip\_data$ を復号することができる．同じ $trip\_key$ で復号される OD 情報は自身を除く他の $k-1$ と区別されない．ゆえに，k-匿名性を保証した OD 情報の開示が可能となる．

2.2 Privacy-Preserving Data Publishing

Privacy-Preserving Data Publishing（PPDP）はプライバシを保護しつつデータを公開するための技術の総称である [9]．PPDP に用いられる手法の代表的な例として，匿名化を用いた手法が挙げられる．匿名化を用いた手法では明確なプライバシ基準の下にデータを一般化するため，元データと照らし合わせられない限り，特定の個人に関する情報を同定することは困難である．一般的なプライバシ基準として k-匿名性，l-多様性が知られている．

次に，本稿にて用いる PPDP 関連用語を定義する．データベースを構成するテーブルを「データテーブル」と呼び，その行を「タプル」，列を「フィールド」とする．また，各フィールドを「属性」と呼ぶ．属性は保持するデータ値の内容を表す．個人一人分もしくはグループ分に相当するデータの一纏まりを「データセット」と呼び，1 タプルに 1 データセットが格納されているものとする．

属性のうち，名前や固有 ID 等，それ単体で個人を一意に特定できるものを「正識別子」，単体では特定されないが，性別や年齢など，正識別子でない他の属性と組み合わせることで間接的に個人を識別できるものを「準識別子」と呼ぶ．

正識別子，準識別子以外で，データ解析者にとって重要な項目であるためにマスキング処理等を行わない属性を「注目属性」と呼ぶ．また，一般化によって形成される準識別子の組み合わせが同一であるタプルのグループを「q*-block」と呼ぶ．

2.3 k-匿名性

k-匿名性とは，データテーブル中の各タプルにおいて，そのタプルの持つ準識別子の組み合わせと同一の準識別子の組み合わせを持つタプルが自身を含め k 個以上存在する状態である [5]．表 1 は 3-匿名性のプライバシ保護を受けたテーブルの例である．ここで v1〜v3，v4〜v7 の 2 グループはそれぞれ準識別子の組み合わせが一致しているため，それぞれ q*-block である．また各 q*-block 内には少なくとも 3 つのタプルが存在するので，3-匿名性を満たす．

Vol.2017-DPS-171 No.13
Vol.2017-MBL-83 No.13
Vol.2017-ITS-69 No.13
2017/6/1

表 1　車種を注目属性，出発時刻および出発地を準識別子としたテーブル（$k = 3$）

	出発時刻	出発地	車種	
v1	15:*	C	セダン	Alice
v2	15:*	C	セダン	
v3	15:*	C	セダン	
v4	16:*	D	ワゴン	
v5	16:*	D	ワゴン	Bob
v6	16:*	D	バン	
v7	16:*	D	バン	

2.4　k-匿名化では防げない攻撃

k-匿名化によっても防ぎきれないような攻撃として，同種攻撃と背景知識攻撃が存在する [10]．2 つの攻撃はそれぞれ攻撃の対象が属する q*-block が特定された場合において生じる．

同種攻撃は匿名化後の表のタプルグループの注目属性値が同種である場合に対象者の注目属性値を一意に特定する攻撃である．表 1 にて v1〜v3 の q*-block に注目すると，注目属性にあたる車種の値は全てセダンであることが確認できる．従って，例えば攻撃者が Alice の車種を知ろうとした場合，Alice の準識別子を知っていれば，該当するタプルが複数存在していたとしても，攻撃者に Alice の車種が暴かれる．

背景知識攻撃は同種攻撃ほど一意に対象者の注目属性値を特定できないが，対象者についての背景知識を利用して注目属性値の候補を絞り込む攻撃である．表 1 の v4〜v7 の q*-block に注目する．q*-block の注目属性に複数の値が含まれている場合でも，v5 である Bob の車種が暴かれる可能性がある．例えば，「Bob はバンに乗っている確率が非常に低い」という背景知識があると仮定すると，Bob のタプルを含む q*-block からバンを除いたときにワゴンのみが残り，Bob の車種が高い確率でワゴンであることが推測される．

2.5　l-多様性

先に述べたような攻撃によるプライバシの侵害を考慮することで，k-匿名性を強化し，属性推定も防ぐプライバシ基準として l-多様性が提案されている．

l-多様性とは，データテーブル内の全 q*-block において，その注目属性の値が少なくとも l 種以上ある状態を保つというものである [6]．

3.　提案

本章では，k-匿名性と l-多様性によるプライバシに配慮した車両データ収集・開示手法を提案する．

3.1　前提

本稿にて提案する車両データ収集システムに参加する車両はすべて V2V 通信および V2I 通信に対応した V2X 通信機器を搭載している．また，車両はこの V2X 通信機器を用いてインターネットにアクセスし，車両データを中央のデータベースにアップロードする．

3.2　システムモデル

提案手法のシステムモデルを図 1 に示す．図 1 は参加車両同士が V2V 通信にて情報交換を行い匿名化等のプライバシ保護を実施し，セルラ通信を用いて中央のデータベースにデータをアップロードする．また，交通機関等の公開データ利用者は収集されたデータを照会することができる．

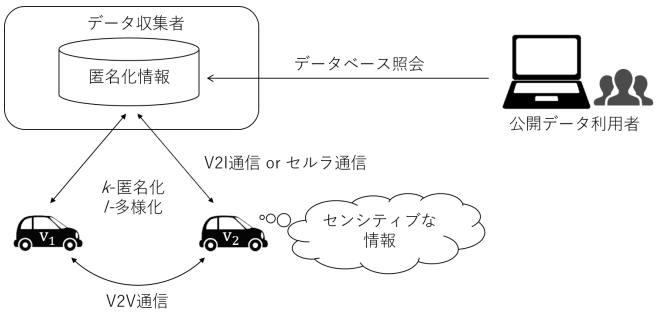

図 1　提案手法のシステムモデル

3.3　プライバシに配慮した車両データの収集・開示の概要

参加者はデータベースに各々の走行に関する暗号化されたデータをアップロードする．このとき，OD に関する位置・時刻情報は異なる精度レベルのキーで暗号化され，複数の精度レベルに変換された走行情報のコピーがアップロードされる．同じ走行情報（同じ OD，時間帯，精度レベル）を作ったユーザ間で共通のキーを使用する．また，キーは秘密分散法によって分割されアップロードされる．少なくともその k 個の *share* がアップロードされたときにキーは再構築され，対応する走行情報が復号される．データベースから得られる各走行情報の精度は少なくとも k 個の走行情報に適用される．これによって参加者は k-匿名性のプライバシ保護を受ける．また，ユーザの車種情報の特定を防止するために，準同型暗号を用いた秘匿集計により車種において l-多様性のプライバシ保護を実施する．

3.4　走行情報の曖昧化

走行情報とは，出発地，目的地，出発時刻，到着時刻の 4 つの属性で構成されるデータセットである．本節にて説明する走行情報の曖昧化の目的は，ユーザの走行情報における k-匿名性のプライバシ保証の実施である．ユーザは自身を含め，k 人と区別されない走行情報を持つとき k-匿名である．また，ここでは精度レベルというパラメータを用

Vol.2017-DPS-171 No.13
Vol.2017-MBL-83 No.13
Vol.2017-ITS-69 No.13
2017/6/1

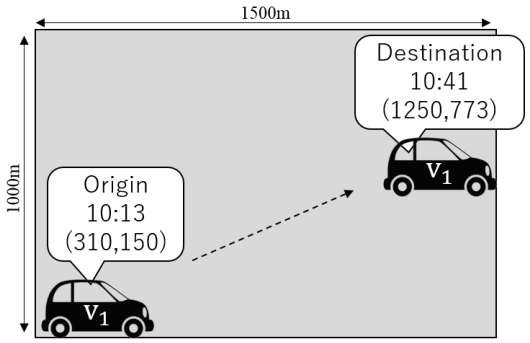

図 2 曖昧化前の走行情報

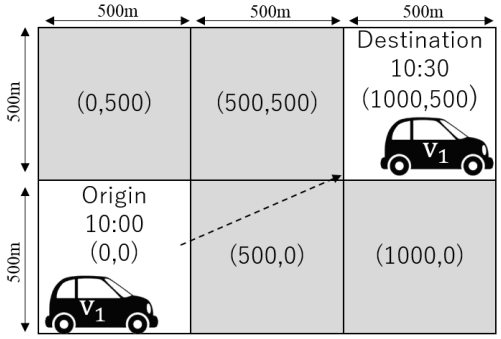

図 3 曖昧化後の走行情報

いて位置・時刻の曖昧化を行う．車両 v_1 の持つ走行情報について，精度レベル（15 分，500m）での曖昧化の例を以下に示す．

位置の曖昧化ではデカルト座標（x，y）を 500m の粗さで切り捨てし，位置情報をピンポイントな 1 点の座標から 1 辺 500m サイズのグリッドに変換する．たとえば，図 2 の車両 v_1 の出発地の位置座標（310，150）は図 3 のグリッド（0，0）内から出発したという情報に変換される．時刻の曖昧化では出発・到着の時刻を 15 分の粗さで切り捨てし，15 分の時間幅を持たせた出発・到着時刻情報に変換する．たとえば，図 2 の車両 v_1 の出発時刻は 10 時 13 分であるが，これを 15 分の時間幅で切捨て，曖昧化すると図 3 の 10 時 00 分という時刻情報に変換される．つまり，10 時 00 分〜10 時 14 分の間に出発したすべての車両が同一の時刻情報を持つことを意味する．本システムでは（15 分，500m），（30 分，1000m），（60 分，1500m）の 3 つの精度レベルで走行情報の曖昧化を行う．

3.5 キー交換

車両が特定の時間に特定の場所にいる場合，共通の時間帯，グリッドに対応するキーを共有する．キーの共有は，車両が他の参加車両と同一の走行情報を作成することを目的として行われる．ある精度レベルに関しての時間帯とグリッド内にいるすべての車両がそのキーを知る．また，異なる精度レベルのキーは独立して同時に確立される．特に，出発地で設定されるキーを *origin_key*，目的地で設定されるキーを *destination_key* とする．キーの属性は，キーを識別する ID である $ID(key)$，精度レベル，時間帯，グリッド，暗号化キー自身である．また，平文 m を暗号化したものを $ENC_{key}(m)$ と表記する．

次に，車両が特定の精度レベルのキーを確立する手順について説明する．

（1） 車両は選択された精度レベルに従って，現在の時刻と位置を曖昧化する．

（2） 車両は時間帯とグリッド内にいる間，V2V 通信を用いて周期的にビーコンをブロードキャストする．他の参加車両との通信範囲に入ると，それまでに取得し，保持しているすべてのキーを交換・受信する．もし，どちらの車両もキーを転送しなかった場合，新しくキーを確立する．

（3） 取得したキーの中から正式なキーを選出する．その手順を以下に示す．

現在の時間帯とグリッドに関する 2 つのキー k_1，k_2 が存在する．キーの ID をそれぞれ $ID(k_1)$，$ID(k_2)$ とする．これらのキーを用いて暗号化したキーレコードを作成する．キー k_2 を用いて k_1 を暗号化したものを $ENC_{k_2}(k_1)$ とする．

$$ID(k_1),\ ID(k_2),\ ENC_{k_2}(k_1)$$

これをデータ収集者が管理する中央のキーサーバにアップロードする．サーバは重複したアップロードを削除する．次に車両はまだローカルにない暗号化されたキーをキーサーバからダウンロードし，所持する暗号キーを使用して復号する．また，まだキーサーバに記録されていないキーについては新たなキーレコードを作製し，キーサーバへとアップロードする．これを一定時間の間繰り返す．最後にキーを辞書順でソートし，1 番目のキーをある時間帯・グリッドにおいての正式なキーとして決定する．

与えられた時間帯・グリッド内に存在するすべての車両が共通の手順でキー転送を行うことで，参加車両すべてが共通の正式なキーを取得することが可能となる．

3.6 *k*-匿名性の保証と *q**-block の形成

走行情報に関して k-匿名性のプライバシ保護を実現する場合，準識別子において区別されないタプルグループ q*-block が形成される．k-匿名性を担保した走行情報の開示手順について説明する．

提案手法は，2.1 節で述べた Förster らの情報開示手法

を利用し，k-匿名性のプライバシ保護を実施する．ここで $trip\_key$ を共通の秘密（s）とする．車両は $origin\_key$ と $destination\_key$ を合わせて暗号学的ハッシュ関数を用い，式 3 のような $trip\_key$ を作成する．

$$trip\_key := h(origin\_key \parallel destination\_key) \quad (3)$$

次に車両は 2.1 節と同様にランダム係数の $k-1$ 次多項式式 (式 2) を構築し，自身の $share$ を計算する．車両はこの $trip\_key$ を用いて曖昧化した自身の走行情報 $trip\_data$ を暗号化し収集者のデータベースに $ID(trip\_key)$，$share$，$ENC_{trip\_key}(trip\_data)$ をアップロードする．その後データベースにアップロードされた k 個の $share$ により $trip\_key$ が再構築され，同じ $trip\_key$ で暗号化された k 個の $trip\_data$ が復号される．ゆえに，k-匿名性を保証した走行情報の開示が可能となり，k 個のタプルを持つ q*-block が形成される．

3.7 準同型暗号を用いた注目属性の秘匿集計

本システムでは OD に関する走行情報に加え，注目属性の例として車種情報を収集する．収集する個人に関する情報量が多くなると，そのデータが持つプライバシレベルは高くなり，より強固なプライバシ基準による保護が要求される．たとえば，2.4 節で述べた k-匿名性では防げない攻撃による匿名性喪失の問題を考慮し，対策する必要がある．データテーブル内の q*-block において注目属性の多様性が欠如している場合，同種攻撃や背景知識攻撃により，ユーザの車種情報が推測される可能性がある．したがって，注目属性の開示を行う際には，q*-block において一定の多様性が担保される必要がある．

プライバシ基準 l-多様性は k-匿名性を強化し，属性推定も防ぐ．提案システムはデータ収集者に注目属性の要素を秘匿にした状態で公開判定を行い，l-多様性のプライバシ保護が可能な場合のみ注目属性が開示される．以下，準同型暗号を用いた注目属性の秘匿集計による l-多様性のプライバシ保護の実施方法を説明する．

準同型暗号（Homomorphic Encryption）とは，復号せずに暗号化したまま加算や乗算などの演算を可能とする暗号であり，演算結果も暗号化した状態で出力できる特徴を持つ暗号である [11]．そのため，計算者（収集者）は計算に用いるデータおよび計算結果の内容だけでなく，計算処理の内容も知ることができない．したがって，注目属性の公開判定が中央のデータ収集者に依存せず，車両自身によって行うことが可能となる．以下に手順を示す．

まず，車両は鍵生成アルゴリズムを用いて公開鍵暗号を作成する．与えられた精度レベルに関して，鍵生成アルゴリズムを用い秘密鍵 sk と公開鍵 pk を出力する．次に，各車両は自身の車種情報 Car_tipe を公開鍵 pk を用いて暗号化する．

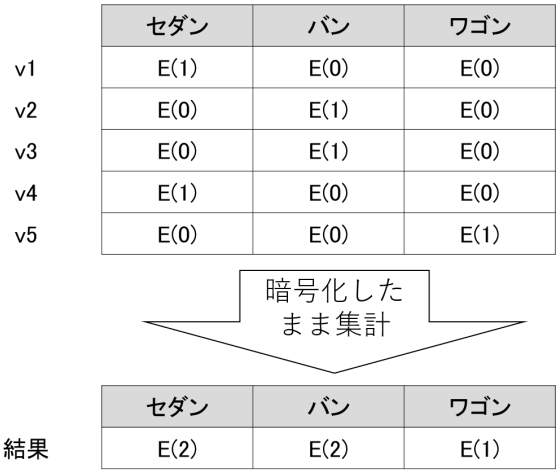

図 4 加法準同型暗号を用いた秘匿集計

$$ENC_{pk}(Car\_tipe) \quad (4)$$

ここで，注目属性の演算プログラムを P とし，以下の情報を収集者データベースにアップロードする．

$$ID(pk),\ ENC_{pk}(Car\_tipe),\ P$$

データ収集者はプログラム P を計算する．ここで，

$$C = ENC_{pk}(Car\_tipe) \quad (5)$$

$$C' = ENC_{pk}(result) \quad (6)$$

とすると，演算プログラム P による車両 k 台の注目属性の演算は次のように表せる．

$$C' = P(C_1, ..., C_k) \quad (7)$$

車両はデータ収集者のデータベースにアクセスし，自身が生成した $ID(pk)$ に対応する演算結果を参照する．このとき車両は準同型暗号の秘密鍵 sk を用いて演算結果を復号する．

$$DEC_{sk}(C') = P(Car\_tipe_1, ..., Car\_tipe_k) = result \quad (8)$$

車両は取得した演算結果 $result$ を照会し，自身が属する q*-block 内に注目属性の多様性が閾値 l 以上であるか確認する．注目属性の多様性が閾値 l 以上である場合，車両は再度車種情報を $trip\_key$ でのみ暗号化しアップロードする．データ収集者は $trip\_key$ を用いて車種情報を復号する．一方，多様性が閾値未満である場合は，要求されるプライバシ基準が満たされない．したがって，車両は注目属性のアップロードを行わない．

3.8 l-多様性を保証した車種情報の開示

ここではプログラム P の処理内容の詳細を示す．プログラム P は加法準同型暗号の特性を利用した電子投票を実施する（図 4）．加法準同型暗号とは，暗号加算のみ可能な暗号方式であり，代表的な方式として Paillier 暗号 [12] が

Vol.2017-DPS-171 No.13
Vol.2017-MBL-83 No.13
Vol.2017-ITS-69 No.13
2017/6/1

存在する．車両は *Car_tipe* の各車種フィールドに関して，自身の車種にあたるフィールド値に“1”を書き込み，暗号化する．たとえば，図 4 において車両 v1 の車種はセダンであるため，セダンのフィールド値を“1”とし，他のフィールド値は“0”にする．フィールドごとに秘匿集計を行い．暗号化されたまま結果が出力される．車両は出力結果を復号し，公開判定を行う．図 4 では，3 つのフィールドがそれぞれ 1 以上の値であるため，3-多様性の保証が確認できる．以上によって，*l*-多様性のプライバシ基準を満たした車種情報の開示が可能となる．

4. 評価

本章では，malicious backend provider，active insider attacker，passive insider attacker の 3 つの攻撃者モデルに関して提案システムのセキュリティを評価し，特定のシナリオにおいてシミュレーションを用いて性能評価を行う．

4.1 セキュリティ解析

malicious backend provider は本システム内にて配備されたすべての中央データベースにアクセス可能であり，格納されているデータを自由に閲覧することができると考えられる．一方でこのような攻撃者は V2X 通信を監視・盗聴することは困難である．提案システムは malicious backend provider に対して安全である．キーデータベースと走行情報データベースに完全にアクセスできたとしても，秘密分散法と暗号化を壊す必要がある．また，キーデータベースの削除やデータ改ざん等の妨害が行われた場合，共通鍵確立への影響が懸念される．しかし，これらの試みは走行情報の可用性に影響を与えることはできるが，参加者のプライバシにおいては負の影響を与えることはない．

active insider attacker はシステム破壊を目的として，V2X 通信にユーザとして積極的に参加する．active insider attacker は，秘密分散法に当てはめることなくキーを集め，明らかにすることができる．これは共通のキーを用いて走行情報の暗号化を行う参加者のプライバシを侵害することになる．一方で，この攻撃は攻撃者が出発地や目的地に物理的に存在する場合にのみ有効なため，その行動範囲は非常に限定される．しかし，大規模な active insider attacker が存在した場合，提案システムにおいて深刻な脅威となりうる．

passive insider attacker はシステムに積極的に参加することはないが，近隣で行われる通信を盗聴する．このような方法が提案システムに干渉することは少ない．一般的に passive insider attacker は V2X 通信を盗聴することができるが，キーの転送や交換は暗号化された通信チャネルによって保護されている．

図 5 *LuST* 交通シナリオのマップ

表 2 シミュレーション条件

パラメータ	値
シミュレータ	SUMO
シナリオ	*LuST*
マップ	$155.9km^2$
時間帯	1:00〜3:00
シミュレーション時間	$7200s$
通信範囲	$200m$
車両数	2974 台
車種数	6 種

4.2 シミュレーション環境

今回評価に用いたシミュレータは交通流シミュレータ SUMO[13] である．通信においては車車間の無線接続性を $200m$ と仮定し，SUMO から 1 秒周期で出力される車両のトレースデータを python ベースの実装を用いて解析した．

交通流は交通流シミュレータ SUMO と *LuST* 交通シナリオ [14] を用いて生成した．シミュレーションマップを図 5 に示す．*LuST* 交通シナリオはルクセンブルクの街の 24 時間の現実的な交通流を提供し，約 $156km^2$ の面積をカバーしている．具体的には，通勤や通学，自由時間等のアクティビティ，朝，昼，夜それぞれのラッシュアワーのモビリティが統計データを元に再現されている．また，公共バスを除外した一般車両の 218938 台の走行を考慮している．

4.3 シミュレーション条件

提案システムの評価に用いたシミュレーション条件を表 2 に示す．シミュレータは先に述べた SUMO を用いた．シミュレーション時間は 7200 秒であり，*LuST* 交通シナリオで利用可能な 24 時間の内の深夜帯 1 時〜3 時の 2 時間（7200 秒）における交通流データを使用した．シミュレー

Vol.2017-DPS-171 No.13
Vol.2017-MBL-83 No.13
Vol.2017-ITS-69 No.13
2017/6/1

ションマップの面積は $155.9km^2$ である．サンプル車両はシミュレーション時間内に走行を開始・終了した 2974 台である．これらすべての車両が V2X 通信機器を搭載しているとし，その通信範囲を $200m$ とした．また，本シナリオにて調査する車種情報は *LuST* 交通シナリオにて定義されている（sedan, hatchback, wagon, van, delivery, bus）の 6 種を対象とした．

4.4 評価項目と評価対象

4.4.1 評価項目

LuST 交通シナリオにて以下の項目に関して評価を行う．

(1) 提案システムの V2X 通信ベースのキー確立のアプローチは，同じ精度レベルで曖昧化された走行情報を持つ車両の間での共通の正式なキーの選出に適しているか．

(2) プライバシ基準の実施は，交通機関などのデータ利用者に向けて公開されるデータの開示率にどのように影響するか．

以上の 2 項目に関して，次の対象を用いて評価を行った．

4.4.2 評価対象

先に述べた評価項目関して，提案システムから得た結果と理論最適を比較した．理論最適とは，同じ精度レベルの走行情報を作るべきすべての車両が共通のキーをキー交換により取得し，データ開示を行ったときの結果である．要求される k，l の閾値に対して，k-匿名性と l-多様性のプライバシ基準を満たしたデータの開示率について，異なる 3 つの精度レベル（15 分, $500m$），（30 分, $1000m$），（60 分, $1500m$）にて調査した．また，提案システムから得た結果と理論最適を比較し，データの開示率について，異なる k 値について調査した．

4.5 シミュレーション結果と考察

それぞれの評価項目について，シミュレーションの結果から得られたグラフをもとに考察を行う．図 6 は結果から得られた情報の精度とデータ開示率の関係を示すグラフである．図 7 精度レベルを固定し，結果から得たれた k-匿名性によるプライバシ保護度合とデータ開示率の関係を示すグラフである．

4.5.1 情報の精度とデータ開示率

図 6 は情報の精度とデータ開示率の関係を示すグラフである．また，要求されるプライバシ基準（3-匿名性，3-多様性）について提案システムの結果と理論最適を比較している．このグラフを見てみると，（15 分, $500m$），（30 分, $1000m$），（60 分, $1500m$）の順に，精度レベルが粗くなるほど同じプライバシ基準を満たしたデータ開示率が高くなっている．この結果から分かることは，要求されるプライバシ基準が $k = 3$，$l = 3$ のような固定値のとき，グリッドの面積と時間帯幅が広くなるほど基準を満たしたデータ開示率が高くなるということである．実際に，3-匿名性（$k = 3$）のプライバシ基準が要求された場合，提案システムは（60 分, $1500m$）の精度レベルにおいて 79.89%の開示率を達成した．このとき理論最適の開示率が 79.92%である．また，同精度レベルにおいて k-匿名性を強化する l-多様性のプライバシ基準（$k = 3$，$l = 3$）が要求された場合においても提案システムの開示率は 76.43%であり，このとき理論最適の開示率は 76.56%である．精度レベル（60 分, $1500m$）ではどちらのグラフも提案システムの開示率が理論最適を下回った．一方，他の 2 つの精度レベル（15 分, $500m$），（30 分, $1000m$）では理論最適値と提案システムの値が同値である．このことから，提案システムの V2X 通信ベースのキー確立のアプローチが機能していることが確認できる．粗い精度レベルであるほどグリッド内，時間帯内に属する車両数が増えるため，匿名性の確保は容易になる．一方，グリッドの面積が広くなると車両のモビリティと通信範囲の都合上，グリッド内に存在する車両の中には通信によって共通のキーを取得することが出来ない者が現れる．このとき車両は他の参加車両と同一の走行情報を作ることが出来ない．また，開示率はプライバシ基準によって制約されるが，提案システムは特定の精度レベル，シナリオにおいてプライバシに配慮したデータ開示を実現したと考える．

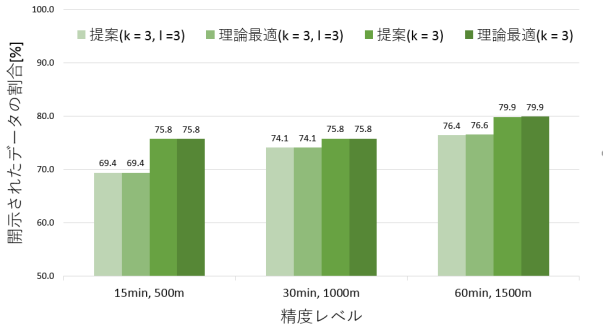

図 6 情報の精度とデータ開示率

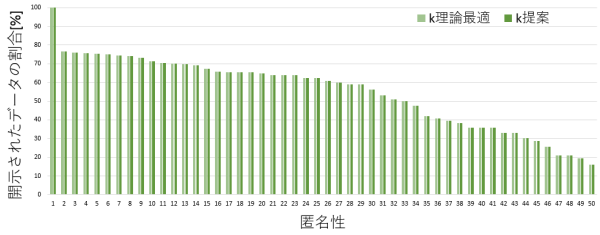

図 7 k-匿名性によるプライバシ保護度合とデータ開示率（30 分, $1000m$）

4.5.2 プライバシ保護度合とデータ開示率

図 7 は k-匿名性によるプライバシの保護度合とデータ開示率の関係を示すグラフである．ここでは精度レベルを（30 分, $1000m$）に固定し k 値を変化させて，プライバシ保

護度合がデータの開示率にどのように影響するのかについて考察する．このグラフを見ると，データ開示率が k 値の増加に伴って減少していることが確認できる．実際に 10-匿名性（$k = 10$）でのデータ開示率が 71.49%であるのに対し，50-匿名性（$k = 50$）でのデータ開示率は 16.07%である．これらの結果より，強固なプライバシ保護，すなわち大きな k 値が要求される場合にデータの開示率が低下することがわかる．一方で，10-匿名性のプライバシ基準を適応した場合でも 70%以上のデータを開示が可能であることを確認した．以上より，提案システムはプライバシに配慮したデータ開示を実現したと言える．

5. おわりに

本稿では k-匿名性と l-多様性によるプライバシに配慮した車両データの収集・開示システムを検討した．そこで，本システムでは走行情報に関して，特定のプライバシ基準が保証されるまで意図的に情報の空間的，時間的精度を低下させ，少なくとも k 人が含まれる共通の走行情報を持つ車両グループを形成した．グループ内の全車両は V2X 通信を用いたキー交換によって，共通の暗号鍵 $trip\_key$ を決定する．このとき，車両は走行情報において他の $k-1$ 台の車両と区別されないことから，k-匿名性のプライバシ保護が保証される．

また，本システムは k-匿名性を強化し車種情報における属性推定を防ぐプライバシ基準，l-多様性を採用している．自身が属するグループ内に少なくとも l 種の車種が存在するとき，車両は l-多様性のプライバシ保護を受ける．本システムでは準同型暗号を用いた演算により，車種情報をデータ収集者に秘匿にした状態で公開判定を行い，l-多様性のプライバシ保護を満たしたデータの開示を可能にした．

ルクセンブルクの現実的な交通シナリオを用いたシミュレーション評価では，データ開示率において提案システムは理論最適に近い値を達成していることが確認された．これは提案システムの V2X 通信ベースのキー交換アプローチが十分に機能していることを示す．データの開示率は要求されるプライバシ基準によって制約されるが，シミュレーション結果は，3-匿名性，3-多様性のプライバシ基準が要求されたケースにおいて，最も粗い精度レベルにてデータ開示率が 76%であることを示した．データの開示率を損なうが，提案システムは走行情報や車種情報に関してプライバシに配慮したデータ収集・開示が実際に可能であることを示す．

謝辞 本研究は JSPS 科研費 16H02811 の助成を受けたものです．

参考文献

[1] Hoh, B., Gruteser, M., Xiong, H. and Alrabady, A.: Enhancing Security and Privacy in Traffic-Monitoring Systems, *IEEE Pervasive Computing*, Vol. 5, No. 4, pp. 38–46 (online), DOI: 10.1109/MPRV.2006.69 (2006).

[2] Krumm, J.: Inference attacks on location tracks, *International Conference on Pervasive Computing*, Springer, pp. 127–143 (2007).

[3] Golle, P. and Partridge, K.: On the anonymity of home/work location pairs, *International Conference on Pervasive Computing*, Springer, pp. 390–397 (2009).

[4] Gruteser, M. and Grunwald, D.: Anonymous usage of location-based services through spatial and temporal cloaking, *Proceedings of the 1st international conference on Mobile systems, applications and services*, ACM, pp. 31–42 (2003).

[5] Sweeney, L.: k-anonymity: A model for protecting privacy, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 10, No. 05, pp. 557–570 (2002).

[6] Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M.: l-diversity: Privacy beyond k-anonymity, *ACM Transactions on Knowledge Discovery from Data (TKDD)*, Vol. 1, No. 1, p. 3 (2007).

[7] Shamir, A.: How to share a secret, *Communications of the ACM*, Vol. 22, No. 11, pp. 612–613 (1979).

[8] Forster, D., Lohr, H. and Kargl, F.: Decentralized enforcement of k-anonymity for location privacy using secret sharing, *2015 IEEE Vehicular Networking Conference (VNC)*, pp. 279–286 (online), DOI: 10.1109/VNC.2015.7385589 (2015).

[9] Benjamin, C., Fung, M., Wang, K., Chen, R. and Yu, P.: Privacy-preserving data publishing: A survey of recent developments, *ACM Computing Surveys*, Vol. 42, No. 4, pp. 141–153 (2010).

[10] 村本俊祐，上土井陽子，若林真一ほか：プライバシ保護データ公開に向けた l-多様化適性の評価，情報処理学会論文誌データベース (TOD)， Vol. 4, No. 2, pp. 126–141 (2011).

[11] Gentry, C. et al.: Fully homomorphic encryption using ideal lattices., *STOC*, Vol. 9, No. 2009, pp. 169–178 (2009).

[12] Paillier, P.: Public-key cryptosystems based on composite degree residuosity classes, *International Conference on the Theory and Applications of Cryptographic Techniques*, Springer, pp. 223–238 (1999).

[13] Krajzewicz, D., Erdmann, J., Behrisch, M. and Bieker, L.: Recent development and applications of SUMO-Simulation of Urban MObility, *International Journal On Advances in Systems and Measurements*, Vol. 5, No. 3&4 (2012).

[14] Codeca, L., Frank, R. and Engel, T.: Luxembourg SUMO Traffic (LuST) Scenario: 24 hours of mobility for vehicular networking research, *2015 IEEE Vehicular Networking Conference (VNC)*, pp. 1–8 (online), DOI: 10.1109/VNC.2015.7385539 (2015).