RO-003

# プライバシ情報の安全な流通と利活用を実現するシステムのアーキテクチャと評価

宮川 伸也* 森 拓也† 岡田 勲* 佐治 信之‡
Shinya Miyakawa* Takuya Mori† Isao Okada* Nobuyuki Saji‡

## 1. はじめに

近年、GPS や RFID などのセンシング技術の発達により、その場その時の個々人の位置情報等のプライバシ性の極めて高い情報(以下、プライバシ情報)を得ることができるようになってきた。これらのプライバシ情報をサービスプロバイダ(以降 SP と略す)が有効に活用することによって、パーソナライズされ利便性の高いサービスを提供することができる。しかしながら、個人が特定されたり、知られたくない属性が推定されるという個人側のリスクがあり、一方で SP 側も、漏えいや不正アクセス等のリスクを想定して SP 毎にこれらプライバシ情報の収集や管理を行わなければならない。これらリスクを軽減・解消していくには、法制度の整備や社会的コンセンサスの醸成、プライバシ情報の適度な集中管理と流通の仕組み、プライバシ情報を安全・適切に匿名化する技術など、社会の仕組みと技術の両面からの取り組みが必要である。

位置情報を活用したサービスは、ここ数年で海外では foursquare、Gowalla などに Google、Twitter、Facebook などが加わり、利用者数も着実に増加を続けている。国内でもゲームなどを中心に既に多くのサービスが提供されている。これらサービスでは、位置情報等の収集は個別の許諾で行われており、それら情報の扱いも利用規約上は使途が限定されているものの、実際に位置情報等がどのような形式に処理され利用されるのか、あるいは他のサービスに提供されるのか、などは必ずしも明確でなく、また、今後、個々人の他の情報と組み合わせることにより、個人の意図を超えて個人の情報が精緻に解析される可能性もあるなど、十分な安全性が担保されているとは言い難い状況であるといえる。

各種データの蓄積が暗黙のうちに進み、それが既成事実化し、ユーザの求めに応じて後から第三者提供を停止する方法 (オプトアウト) でしか自己の情報を守れないといった事態を避けるためにも、第一に、プライバシ情報の扱い方や組合せの範囲などを法制度など含めてより深く検討し明確化すべきであり、第二に、それらの情報を保護しつつ安全に交換し、場合に応じて匿名化するなどの技術を確立していくことで、プライバシ情報の拠出に関する個々人の納得性を高めていくことが必要である。

本稿は、後者の技術面におけるアプローチ、アーキテクチャおよび実証実験を通じた評価について述べるものである。

経済産業省の情報大航海プロジェクト(平成 19 年度～21 年度)においては、個々人の承諾を得た上で、各種のプライバシ情報を安全に収集・管理しつつ、そこから行動のパターンを抽出・分析して、それを適切な形に変換して SP に提供する仕組みを構築した。併せて、個人のプライバシ情報を個別事業者が扱うのではなくプラットフォーム(PF)として包括的に扱う仕掛けの実装、PF から提供された匿名化情報を元に SP が提供するサービスの実装と効果の評価、サービスや PF の仕掛けに対する個人の受容度の評価などを行った。実証実験は毎年 3,000 人規模で実施した(NTT ドコモ、NEC によるマイ・ライフ・アシストサービス実証実験、表 1)。

表 1 マイ・ライフ・アシストサービス三か年の開発

| | 初年度 | 二年度 | 三年度 |
|---|---|---|---|
| 実証サービス | 各種お役立ち情報の携帯配信 | ・商店街の商品推薦<br>・改札通過・購買履歴連携 | ・電子チラシ配信<br>・カーナビ連携<br>・各種お役立ち情報配信 |
| プライバシ情報セキュア流通基盤 | 利用者主導開示制御 | 位置情報単純あいまい化 | $k$-匿名性・$\ell$-多様性保証、段階的オプトイン |
| 行動情報分析・活用 | 行動情報の安全な管理 | 行動パターン抽出 | コンテキスト推定 |
| 行動情報ベース推薦 | エリア滞在頻度に基づく推薦 | 複数推薦方式のマルチモード学習 | コンテキスト適応マルチモード学習 |

情報大航海プロジェクトの実績に基づき、種々の課題解決に向けて、以下の３点について述べる。

A) 情報流通制御のアーキテクチャの研究開発と実証
情報大航海プロジェクトでは、情報の秘匿度と提供先の信頼度に応じた開示制御、位置情報における匿名化技術の研究開発と適用を行い、包括的な処理モデルとして提案している。本稿では、そのアーキテクチャおよびポリシーの考え方について述べる。

B) 匿名化技術の研究開発と実証
今後は、位置情報、生体情報、行動センシング情報などの多様なデータが履歴として蓄積され、利用されることが予想される。情報大航海プロジェクトでは、人の行動履歴を活用するサービスにおける匿名性に注目し、従来の $k$-匿名化方式に加えて、多面的知識を持った攻撃者に対しても推定リスクを抑える手法として、位置情報に対する $\ell$-多様性保証の方式を新たに提案している。本稿ではその有用性について述べる。

C) 段階的オプトインの有用性の実証
ユーザが SP 等に提供する情報の詳細度を設定/変更し、開示するプライバシ情報の匿名度の程度を制御することができる段階的オプトイン機能を実現した。本稿ではその仕組み、および実証実験において、ユ

* 日本電気株式会社　サービスプラットフォーム研究所、Service Platforms Research Laboratories, NEC Corporation
† 日本電気株式会社　知的資産Ｒ＆Ｄ企画本部、Intellectual Asset R&D Planning Division, NEC Corporation
‡ 日本電気株式会社　公共・医療ソリューション開発本部、Community and Medical Solutions Development Division, NEC Corporation

ーザが実際に匿名度をどのように変更していったかの傾向についても述べる。

## 2. 関連研究

情報アクセス制御、プライバシ情報等の保護、匿名化等に関しては、これまでに様々な標準化活動や研究が行われている。本研究に関連するものを以下に挙げる。

### 2.1 情報流通制御関連の標準化動向

OASIS の XACML[1] はアクセス制御の記述言語と処理モデルを規定しているもので、基本仕様として広く利用されている。Kantara Initiative (元 Liberty Alliance Project) では位置情報を含むユーザの属性情報を SP と共有する仕組み、及びユーザに都度提供可否を確認する仕組みを ID-WSF [2]、及び ID-SIS[3] で規定している。IETF では、プライバシを考慮して位置情報を扱うための仕様群として Geopriv[4]を定めており、ユーザは SP へ提供する位置情報の精度や粒度をポリシーで制御できる。

しかしながら、ID-WSF、ID-SIS については SP の信頼度や情報の重要性に応じて属性情報の開示範囲を動的に制御することは想定されていない。また Geopriv のポリシーでは提供先の SP を指定する必要があるため、新たな SP の動的な追加に即時に対応することは想定されていない。

### 2.2 開示制御技術

ユーザ主導によるプライバシ情報の開示制御に関してもさまざまな研究がなされてきた。

情報収集エージェント技術[5]はプライバシ情報を安全に保護した上でネット上の情報を収集できる技術であり、ユーザがプライバシ情報の提供を限定できるという特徴を持つ。

視聴情報等保護対策技術[6]は通信ネットワーク利用放送において、ユーザが指定した開示設定に基づきプライバシ情報の属性情報などを保護して放送局へ提示するための技術である。プライバシ情報を管理する事業者と情報利用者へ提供する事業者を分離するアーキテクチャによりプライバシ情報が一箇所に集中することを防ぐという特徴を持つ。

匿名 P2P ネットワーク基盤[7]は P2P ネットワークにおいて公開可能なプライバシ情報を共有できる技術であり、ユーザが指定した公開可能なデータのみを P2P ネットワークで共有し、オニオンルーティングにより送受信者の情報を隠ぺいできる。

これらの技術は、ユーザが主導的に開示するプライバシ情報を取捨選択することを可能にするが、開示したデータは情報利用者の信頼度とは無関係に提供される。

### 2.3 匿名化

*k*-匿名性は、ユーザが特定されないことを保証する最も基本的で重要な指標である[8][9]。位置情報の場合、ある位置に *k* 人のユーザがいるとき、その位置は *k*-匿名性を満たすことになる。文献[10][11]には、位置情報が *k*-匿名性を満たすために、位置情報の空間と時間を匿名化する手法が提案されている。文献[12]では、移動軌跡には自宅、職場、閲覧者がユーザを見かける場所等のユーザを特定できる場所が含まれているということを前提にして、移動パターンから得られる組合せによって個人を特定できる属性 (準識別子) を汎化する方法が提案されている。これらの研究は、個人を特定する位置情報を処理しており、個人にとっての秘密性・機密性の高い (センシティブな) 位置には着目していない。

*k*-匿名性を満たしていたとしても、*k* 人以上のユーザのセンシティブ属性のパターンが少なければ、センシティブ属性を推測できる場合がある。これを解決するために提案されたのが、ℓ-多様性である[13]。ℓ-多様性は、同じ準識別子を持つユーザのセンシティブ属性のパターンが ℓ 通りあることを保証する指標である。ℓ-多様性を満たすための匿名化手法は、センシティブ属性の性質に依存する。これまでの位置情報の ℓ-多様性に関する研究では、センシティブ属性は、位置情報そのもの[16][17]、位置情報から得られる建物の名前[14]、移動軌跡に付与された診断結果[15]等、多岐にわたっている。これらの研究では、データの公開者、誰が閲覧するのか、ユーザのよく行く場所を閲覧者が知っていることが前提となっている。

文献[18]は、情報大航海プロジェクトの共通技術開発の枠組みで行われた個人情報匿名化基盤の開発に関するものであり、国際標準化活動、オープンソース化などが進められている。この基盤には、本稿の成果の一部である位置に関する *k*-匿名化の機能も組み込まれている。

文献[19]では、長時間撮影され蓄積された背景画像上に、その都度画像認識された人物を棒状にするなどして重ねることで、画像処理に誤りがあってもプライバシが保護される「変身カメラ」など、示唆に富んだ研究が行われている。加えて、公共空間などで個人が情報を収集される環境にあることを知りながら、ある環境下に自主的に入る行為によってその時点で情報提供に同意したとする「環境オプトイン方式」が提案されている。

## 3. プライバシ情報流通プラットフォーム

### 3.1 アプローチ

パーソナライズされ利便性の高いサービスを享受するために、プライバシ情報を SP に提供することは一般的に行われているが、特定されないはずの個人が特定されたり、知られたくない属性が推定されることは避けなければならない。そのためには、プライバシ情報の利活用において、提供のしやすさ(安心感、信頼性)、管理のしやすさ(安全性、標準化されている)、利用のしやすさ(安全性、情報の扱いやすさ)を実現する必要がある。

これまでは、個々人が複数の SP にその都度プライバシ情報等の様々な情報を提供し、サービスを享受してきた。このようなサービスの形態を SP 直接接続モデルと呼ぶことにする(図 1)。

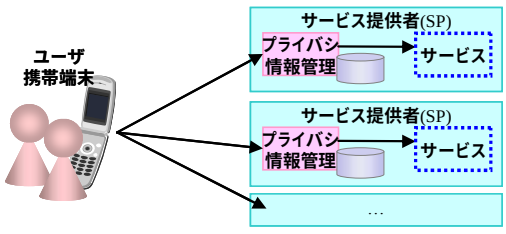


図 1 SP 直接接続モデル

このモデルでは、ユーザは SP 毎にどのような情報を提供したのかなどきめ細かく管理しなければならない、SP は SP 毎にプライバシ情報を管理したり加工したりする仕掛けを用意しなければならないなど、課題が多い。

そこで、本プロジェクトでは、プラットフォームがユーザのプライバシ情報を一元的に収集・管理し、ユーザの開示設定に応じて SP 側に情報を適切な形式に変換して提供することとした。このようなサービスの形態を PF 信頼型仲介モデルと呼ぶことにする(図 2)。

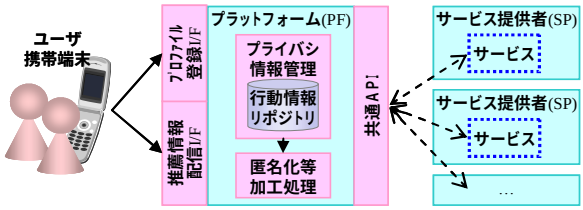


図 2 PF 信頼型仲介モデル

このモデルでは、ユーザは信頼できる PF を選択すればよい。その PF にプライバシ情報の管理を一括して委託し、サービス毎の利用状況は PF で確認する。一方、SP では、プライバシ情報を管理、処理する仕掛けを用意する必要がなく、使用毎破棄を原則とした利用が可能となる。さらにユーザ側・SP 側のインタフェースを標準化することで、様々なデバイス(ユーザ機器)やサービスとの相互運用性を確保することができる。

以上のようにプラットフォームセントリックなプライバシ情報管理の手段を提供することによって、ユーザによるプライバシ情報の提供や、SP によるプライバシ情報利用の障壁を下げ、多くの SP の参入、様々な新サービスの展開を促進することを目指す。

いずれのモデルにおいても、SP(および PF)がどの程度信頼できるのか、その度合いの指標化と第三者による認定といった手当が必要になると考えられる。

## 3.2 アーキテクチャ

PF を実現するためのアーキテクチャを図 3に示す。

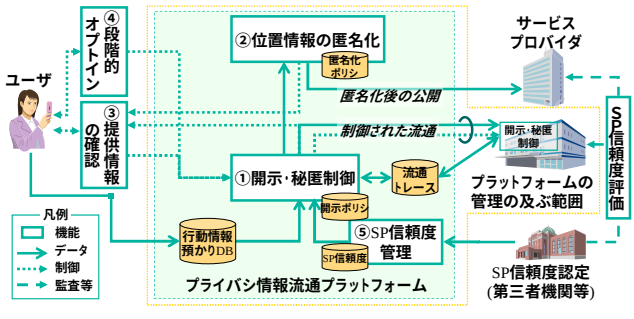


図 3 プラットフォームのアーキテクチャ

本アーキテクチャの特徴は、PF の管理範囲内と管理範囲外の両方において、プライバシ情報を安全に流通するための機能群を備えていることである。

①開示・秘匿制御
管理範囲内においてユーザ主導でプライバシ情報の開示・制御が行えるように、情報のプライバシ性と、SP の信頼度に基づいてユーザ指定の開示ポリシーによって開示・秘匿を制御する機能。

②位置情報の匿名化
位置情報から個人が特定されないように位置情報を匿名化する機能。管理範囲外を含めたプライバシ情報の広範な流通・利用を可能にする。詳細は4章で論じる。

③提供情報の確認
管理範囲内において、提供したプライバシ情報、提供先、利用状況 (参照回数等) をユーザが確認するための機能。

④段階的オプトイン
SP 等に提供する情報の詳細度を設定/変更し、開示するプライバシ情報の匿名度の程度をリアルタイムに制御することができる機能。詳細は5章で論じる。

⑤SP 信頼度管理
SP の信頼度を保管し、必要に応じて参照可能とする機能。外部の第三者機関である SP 信頼度認定により、SP のプライバシ情報の取り扱いに関する管理レベルを評価し、SP に付与する形態を基本とする。PF の信頼度認定も SP と同じ枠組みで行ない、ユーザは PF と SP の、PF は SP の信頼度の評価にあたり、SP 信頼度認定を信頼するものとする。

## 3.3 開示ポリシーと匿名化ポリシー

PF はプライバシ情報を適切かつ広範囲に流通できるように、そのままの情報であればどの範囲まで開示するか、広く流通するためにプライバシ性を除くとしたらどの程度まで除くかを、それぞれ、開示ポリシーと匿名化ポリシーと呼ばれるポリシーでユーザにより設定可能とする。開示・秘匿制御機能により、開示ポリシーに基づいて秘匿されたプライバシ情報は、ユーザが許可する SP のみが開示できるようにする。このモデルでは、SP が他の SP に情報を流通する形態においても同じ枠組みで処理できる。

なお、これらのポリシーはユーザが任意に設定することもできるが、ユーザの受容性を高めるためにポリシー設定を簡略化し、開示ポリシーと匿名化ポリシーの両方について、それぞれ 3 段階の典型的なポリシーを定義した。

### 3.3.1 開示ポリシー

開示ポリシーは情報のプライバシ性と SP の信頼度に基づき規定される。プライバシ性については、プライバシレベルとして、情報が保持するプライバシ性の高低に基づき指標化する。プライバシ性は、個人の感情・感覚に依存するため、レベルを設けてプライバシ情報に対して属性を割り当てることとした(表 2)。

表 2 プライバシレベル

| レベル | 説明 | 例 |
|---|---|---|
| 128 (高) | 単独で個人を特定できる情報、または、漏えいした際に単独で個人に不利益を与える情報 | 携帯メールアドレス、携帯電話番号、位置情報(ピンポイント)、等 |
| 64 (中) | 二つ以上組み合わせることによって、レベル128と同一の効果を発揮する情報 | 生年月日、郵便番号、自宅最寄駅、位置情報(250m四方)、等 |
| 32 (低) | レベル128もしくはレベル64の情報と組み合せることによって、個人の特徴がより明確になる情報 | 興味のある事柄、好きな食べ物、性別、位置情報(1km四方)、等 |
| 0 (なし) | 他の情報と組み合せても個人の特徴がいっさい明確にならない情報、又は匿名化されたプライバシ情報 | 位置情報(2km四方)、匿名化されたプライバシ情報、等 |

また、SP の信頼度は SP 信頼度認定により付与される。信頼度については、トラストレベルとして SP のプライバシ情報の取り扱いに関する管理レベルに基づき指標化す

る。便宜的に、PF 提供者、及び任意の SP の信頼性を示す指標も定義する(表 3)。

表 3 トラストレベル

| レベル | サービス（機能コンポーネント） |
|---|---|
| 128 | プラットフォームに割り当てられるトラストレベル<br>(利用者はプラットフォーム提供者を信頼するものとする) |
| 64 | 比較的高い信頼性を持つサービスプロバイダに割り当てられるトラストレベル |
| 32 | 比較的低い信頼性を持つサービスプロバイダに割り当てられるトラストレベル |
| 0 | 上記に当てはまらないサービスプロバイダに割り当てられるトラストレベル |

開示ポリシーは、どのプライバシレベルのプライバシ情報がどのトラストレベルのサービスで利用可能かを記述するポリシーである。本提案では、図 4の通り積極型、バランス型、消極型の 3 通りの開示ポリシーを定義した。

積極型

| プライバシレベル＼トラストレベル | 128 | 64 | 32 | 0 |
|---|---|---|---|---|
| 128 | ○ | ○ | × | × |
| 64 | ○ | ○ | ○ | × |
| 32 | ○ | ○ | ○ | ○ |
| 0 | ○ | ○ | ○ | ○ |

バランス型

| プライバシレベル＼トラストレベル | 128 | 64 | 32 | 0 |
|---|---|---|---|---|
| 128 | ○ | × | × | × |
| 64 | ○ | ○ | × | × |
| 32 | ○ | ○ | ○ | × |
| 0 | ○ | ○ | ○ | ○ |

消極型

| プライバシレベル＼トラストレベル | 128 | 64 | 32 | 0 |
|---|---|---|---|---|
| 128 | ○ | × | × | × |
| 64 | ○ | × | × | × |
| 32 | ○ | ○ | × | × |
| 0 | ○ | ○ | ○ | ○ |

○：開示、×：非開示

| 意味づけ |
|---|
| プラットフォーム(トラストレベル=128)は常に信用する |
| 開示ポリシによる制御の範囲(開示・秘匿の対象) |
| 匿名化プライバシ情報(プライバシレベル=0)は広く流通 |
| 匿名化プライバシ情報(プライバシレベル=0)に対する要件は匿名化ポリシで制御 |

図 4 開示ポリシー

ここで、プラットフォーム(トラストレベル=128)は全てのプライバシレベルのプライバシ情報を取り扱い可能とする。また、匿名化されたプライバシ情報などプライバシ性のない情報(プライバシレベル=0)は PF の管理範囲外のサービスでも利用可能とする。

### 3.3.2 匿名化ポリシー

プライバシ情報は単体より組合せの方がプライバシ性が高くなる傾向がある。位置情報を例にとって考えてみると、単体の位置情報よりも複数の位置情報から構成される位置動線の方が、ユーザを特定・追跡できる可能性が増加し、プライバシ性は高くなると考えられる。一方、緯度・経度等によって表現された詳細な位置情報をエリア等に一般化した場合、特定・追跡の可能性は低下するため、プライバシ性は低くなると考えられる。

PF がプライバシ情報を匿名化する際には、組み合わせる位置情報の数に留意している。そのために、表 4 に示される匿名化ポリシーを用い、ユーザが匿名性の高さを指定できるようにしている。匿名化ポリシーでは、より多くの位置情報を組み合わせてもユーザを特定できないことを「匿名性が高い」としている。

表 4 匿名化ポリシー

| 匿名性 | 定義 |
|---|---|
| 高 | すべての位置情報を組み合わせてもユーザを特定できない、加えて自宅･勤務先を500m四方以上にぼかす |
| 中 | 任意の二つの位置情報を組み合わせてもユーザを特定できない、加えて自宅･勤務先を500m四方以上にぼかす |
| 低 | 一つの位置情報からユーザを特定できない、加えて自宅･勤務先を250m四方以上にぼかす |

さらに、プライバシ性が高い場所 (自宅・勤務先) については、かならず 250～500m 四方にぼやかしている。

## 4. プライバシ情報の匿名化技術

ユーザの情報を公開するにあたってユーザのプライバシを保護するためには、ユーザが意図した相手に許可した範囲で情報を提供することが重要であり、そのための方法として、前章で述べたような情報を種別毎に開示する範囲を制御する方法と、開示する相手に合わせて情報のプライバシ性を軽減する方法が考えられる。匿名化は、公開する情報を個人が特定されないように操作することであり、後者を実現するための方法の一つである。本章では、特に位置情報に関して、プライバシ保護の必要性、種別に応じて導入した匿名化手法を説明する。

### 4.1 位置活用サービスにおけるプライバシ保護

近年、店舗検索、ゲーム、ユーザ同士の情報共有等、さまざまなシーンで位置情報が使われる機会が増えている。SP がサービスを提供する際には、ユーザの現在地のみを使う場合もあれば、定期的に蓄積した位置情報を分析し得られた行動傾向を使う場合もある。前者の場合、ユーザの現在地の周辺店舗やその場所から目的地までの経路を配信する等が考えられる。後者の場合、ユーザの行動傾向からユーザが頻繁に訪れる店舗の情報や行動範囲における交通情報を配信する等が考えられる。

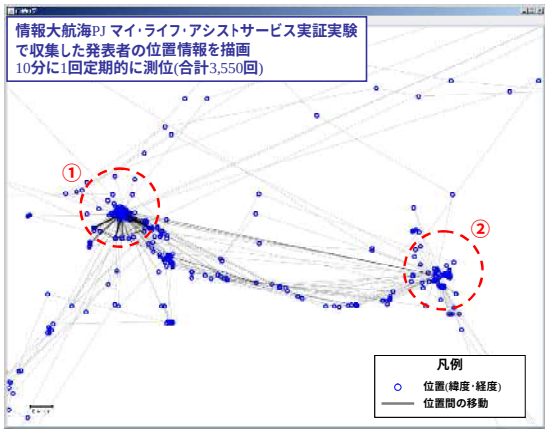


図 5 蓄積された位置情報のプライバシ性

ユーザが提供する位置情報は、プライバシ性が非常に高いことがある。図 5は、情報大航海 PJ において 10 分置きに約 3,500 回計測した筆者の一人の位置情報を描画した図である。点が GPS で計測した緯度・経度によって表された位置であり、点と点を結ぶ線がある位置から別の位置に移動したことを表す。点が密集している場所は長時間滞在した場所や頻繁に訪れた場所であり、自宅や勤務地、趣味・嗜好を表す場所であることが多い。ユーザの自宅や勤務地を知る閲覧者がこのデータを閲覧したとき、ユーザがどこに行くのかを知ってしまう。つまり、頻繁に位置情報を提供することがユーザのプライバシ侵害の可能性を高めることを示唆している。ユーザが意図せずに位置情報を公開する可能性があるならば、位置情報もプライバシ情報として確実に保護すべきである。

匿名化は、ユーザが情報を提供するにあたって個人のプライバシを守るひとつの手段であり、情報が誰の情報であるのか、あるユーザがどのような属性を持っているのかを知られることを防ぐ。ユーザの情報は、組合せに

よって個人を特定できる属性(準識別子)と、他人に知られたくない属性(センシティブ属性)から構成されるとする。複数のユーザの情報を集約したデータセットに含まれるすべての情報について、準識別子が同じ他の情報が $k$-1 個以上存在する場合、そのデータセットは $k$-匿名性を満たす。$k$-匿名化は、データセットを $k$-匿名性を満たすように処理することである。位置情報のプライバシを保護するためには、この $k$-匿名化がひとつの有効な手段である。

### 4.2 位置情報の匿名化

SP がユーザの現在地情報のみを使う場合、その情報を提供することによって侵害されるプライバシを保護すべきである。位置情報はユーザの保有する携帯電話の個体識別番号等の属性 (識別子) や他のセンシティブな属性と合わせて用いられることがあるため、次のような課題に留意する必要がある。

- 閲覧者が対象ユーザの居場所を知っている場合、位置情報のデータと照合することによって、ユーザの識別子やセンシティブ属性をさらに知ってしまう
- 閲覧者が対象ユーザの識別子を知っている場合、そのユーザの居場所をさらに知ってしまう

このような課題を解決するために、位置情報に対して $k$-匿名化を適用した。これまでの位置情報の $k$-匿名化と同様に、位置情報に対して一般化、もしくは、切落しを行っている。一般化は、位置情報をより広域にぼやかすことによって、ぼやかされた位置のどこにユーザがいるのか分からないようにする。切落しは、ユーザの一部ないし全部の位置情報を公開しないことによって、ユーザがその位置に居たことを分からないようにする。例えば、一般化によって位置情報を $k$-匿名化した例を図 6に表す。

ユーザが密集している場所はより小さなエリアになり、ユーザがまばらな場所は大きなエリアになる。

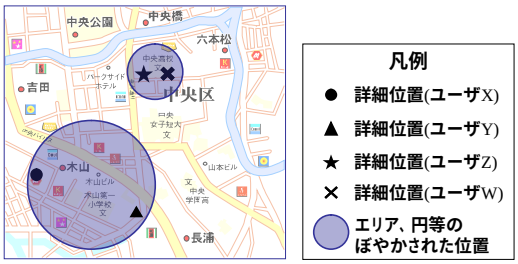


図 6 位置情報の $k$-匿名化

### 4.3 行動範囲の匿名化

SP がユーザの位置情報を蓄積する場合、それによって露わになるユーザの属性を保護すべきである。ここでは、SP は、蓄積された位置情報からユーザが頻繁に訪れる場所や長時間滞在する場所から構成される行動範囲を抽出することを想定する§。行動範囲のように、複数の位置情報から構成される情報の場合、それぞれの位置情報に対して $k$-匿名化を行ったとしても次のような問題がある。

(1) 近くにいるユーザが似たような行動をする場合、訪れる場所を知られてしまう

たとえば、図 7のように、ユーザ A の自宅と病院の位置(点)とその間の移動(点線)が、それぞれ $k$-匿名化($k$≧2)によって、円と実線に表すように匿名化されたとする。左の円にユーザ A が居住していることを知る閲覧者は、ユーザ A の自宅の正確な位置を確認できなくても、ユーザ A が病院に通院していることを知ることができる。

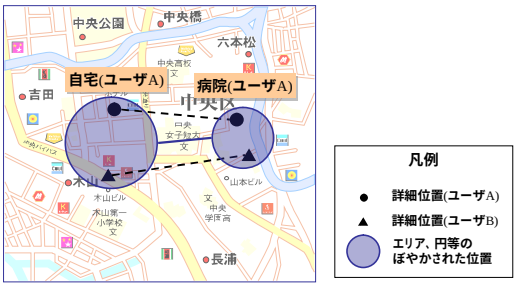


図 7 位置情報の $k$-匿名化の課題(1)

(2) ユーザの複数の居場所が知られている場合、他の場所を知られてしまう

たとえば、図 8のように、ユーザ X の自宅、勤務先、病院の位置(点)とその間の移動(点線)がそれぞれ $k$-匿名化(k≧2)によって円と実線に表すように匿名化されたとする。ユーザ X の自宅と勤務先を知る閲覧者は、自宅と勤務先が合致するユーザ X を特定でき、病院に通院していることを知ることができる。

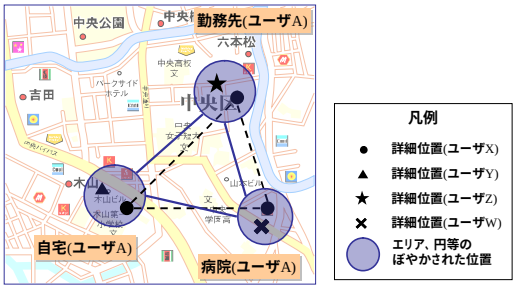


図 8 位置情報の $k$-匿名化の課題(2)

上記の課題を解決するために、本稿では、従来の $\ell$-多様性[13]を拡張した「複数の位置からなる行動範囲に対する $\ell$-多様性」を提案する。従来の $\ell$-多様性では、同じ準識別子を持つデータセットのセンシティブ属性が $\ell$ 通り以上であることを保証するが、本方式では、この概念を行動範囲に拡張する。ある位置 loc を準識別子として、行動範囲に loc を含み、loc 以外の位置が互いに異なるユーザが $\ell$ 人以上いる場合、loc は $\ell$-多様性を満たすと定義する。

例えば、図 7においてユーザ A の自宅を loc としたとき、自宅周辺にいたことがあり、病院周辺以外に行ったことがあるユーザが $\ell$ 人以上いれば $\ell$-多様性を満たすことになる。図 9は、$\ell$≧2 としたときに $\ell$-多様性を満たすように loc をぼやかした例である。病院以外の場所に行く他ユーザを含むように自宅周辺(左円)がぼやかされるため、ユーザ A の自宅を知る閲覧者であっても、ユーザが通院することを確信できない。

§位置情報を解析して得られる他の情報 (移動軌跡等) のプライバシ保護については今後の課題である。

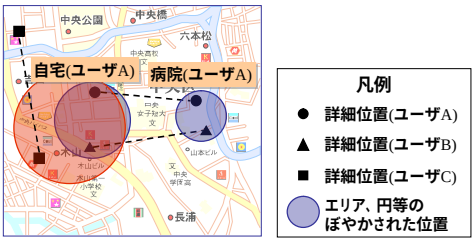


図 9 行動範囲の ℓ-多様化 (1)

同じ $m$ 箇所に居場所があるユーザであり、それ以外の居場所が異なるユーザが ℓ 人以上いるとき、$m$ 箇所の位置に対して ℓ-多様性を満たしているとする。$n$ 箇所に居場所があるユーザについて、居場所の任意の組合せに対して ℓ-多様性を満たすことを($n$, ℓ)-多様性を満たすとする。なお、(1, ℓ)-多様性は $k$-匿名性 (k=1) と同義である。($n$, ℓ)-多様化($n$>1)は、($n$-1, ℓ)-多様化を行った後、任意の $n$-1 箇所全ての位置周辺に居場所があり、かつ、残り 1 箇所の位置が異なる ℓ 人の他ユーザが含まれるように $n$-1 箇所の位置を一般化する。

例えば、図 8のユーザ X に着目した場合、まず、任意の 2 箇所について(2, ℓ)-多様化を行う。次に、自宅と勤務先の両方の周辺に居場所があり、かつ、病院には行かないユーザ (ユーザ V) が含まれるように、自宅と勤務先の両方をぼやかす(図 10)。同様の処理を勤務先-病院、病院-自宅に対しても行う。ユーザ X の自宅と勤務先を知る閲覧者であっても、ユーザ X が病院に行くユーザなのか他の場所に行くユーザなのか特定できない。

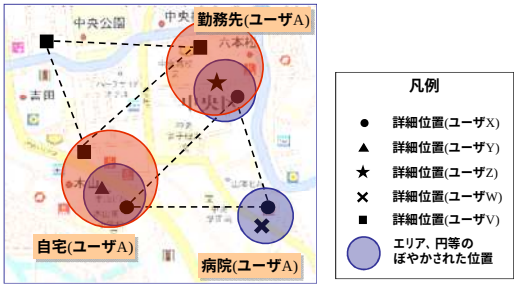


図 10　行動範囲の ℓ-多様化 (2)

## 4.4 匿名化技術の評価

ユーザの行動範囲のグルメ情報を推薦するモデルサービスに、本匿名化手法を適用した。モデルサービスでは、ユーザ位置を携帯電話によって 10〜30 分間隔で定期的に測位し、PF の行動情報リポジトリに蓄積する。行動情報管理は、蓄積された緯度・経度形式の位置を分析して、自宅や勤務地などの複数のセンシティブな位置からなる行動範囲をユーザ毎に生成する。行動範囲の位置の数はユーザ毎に異なる。匿名化エンジンは、位置情報や行動範囲に $k$-匿名化や ℓ-多様化を施して SP に提供し、SP は各位置周辺のグルメ店舗情報をユーザに配信する。

### 4.4.1 ℓ-多様化した行動範囲の有用性

グルメ情報推薦サービスのような場合には、できるだけ多くの位置を含む行動範囲が SP に提供されることが望ましい。個々のユーザの行動範囲の全位置が切り落とされるような匿名化処理がされると、ユーザへの情報配信機会が失われてしまう。そこで、本稿の ℓ-多様化アルゴリズムでは、切り落としを極力抑えてできるだけ多くの位置が残される方法を採用した。

図 11、図 12は、首都圏のユーザ約 3,000 人の行動範囲の多様化を行った結果であり、生成されたエリアの大きさ毎の数、および、切り落とされた位置の数を表す。(∞, 2)-多様化は、ユーザの居場所の数に関わらず、居場所の任意の組合せに対して 2-多様性を満たすことを表す。($n$, ℓ)-多様化の $n$ を増やすと、エリアが大きくなり、切り落とし数も増加する傾向にあるが、ユーザ単位で集計すると、すべての位置が切り落とされたユーザは 2.9%に抑えられている。結果として、多様化による匿名性の保証を行いつつ、できるだけ多くのユーザへの情報配信を可能にする方式であることを示すことができた。

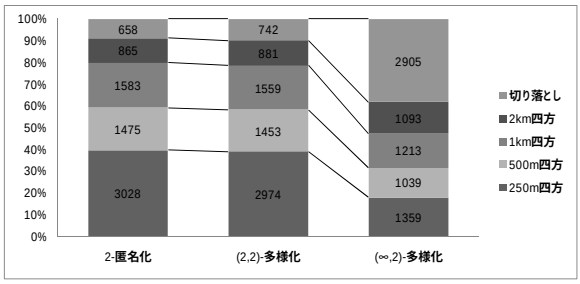


図 11 多様化の結果

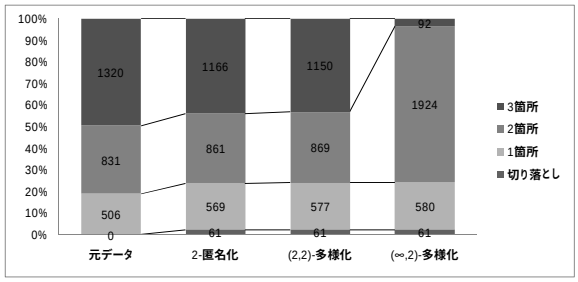


図 12 滞留点数の変化

### 4.4.2 実行時間

図 13は、$k$-匿名化と ℓ-多様化処理の実測値と利用者数が増えた場合の外挿値を示したグラフである。

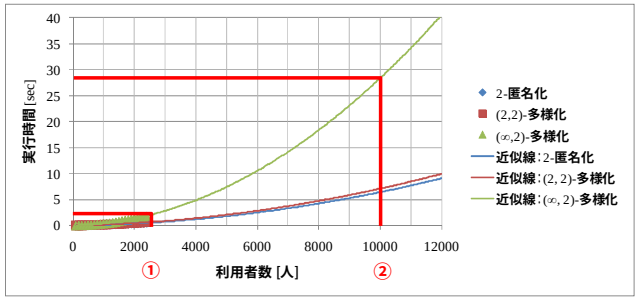


図 13 匿名化の処理時間

$k$-匿名化、ℓ-多様化ともに、実験対象ユーザ(約 3,000 人)に対しては、2 秒程度で処理できている (図 13の①)。10,000 ユーザの場合、最も処理時間がかかる ℓ-多様化は 28 秒程度と算定できる (図 13の②)。

ユーザごとにいくつかの属性が含まれるデータの $k$-匿名化は、属性数が多ければ計算量が多いことが知られている[20]。一方、実現した位置情報の $k$-匿名化は、位置情報をいったん 250m 四方の地域メッシュに変換した後、$k$-

匿名性を満たすまで 2km 四方を上限として拡張する。これにより、位置情報を属性が 1 つしかない情報と同等に扱うことになるため、すべての位置情報の数を $p$、木構造で管理されている地域メッシュの総数を $m$ としたとき、計算量が $O(p \cdot \log m)$に抑えられている。

位置情報の$(n, \ell)$-多様化は、ユーザの $n$ 箇所の位置情報の任意の組合せ $2^n$について $\ell$-多様性を満たすまで、$n$ 箇所の一部の位置情報を 250m 四方から 2km 四方まで拡張する。それでも $\ell$-多様性を満たさない場合は、拡張した位置情報を行動範囲から削除する(切り落とす)。この方法をナイーブに実装すると、ユーザ数を $q$ としたとき、$\ell$-多様性検査の計算量は $O(q \cdot 2^n)$になってしまう。しかし、図 12に示したように$(n, \ell)$-多様性を満たす行動範囲に含まれる位置の数は高々3 箇所 $(n=3)$ 程度と少ないため、行動範囲の位置の数が多く、$(n, \ell)$-多様性を満たさないことが明らかな場合に、いくつかの位置情報を予め切り落とすことでメッシュの拡張や $\ell$-多様性の検査の処理を省略する最適化を行った。これにより、$n$ の影響を極力抑えられ、実用的な実行時間とすることができた。

## 5. 段階的オプトイン

ユーザにとっては、自身のプライバシ情報は適切な相手・タイミング・内容で開示されなければならず、自分の立ち寄り場所が特定されたり、属性が推定されたりするリスクを抑えつつ、一方で、情報が過剰に非開示にされることで享受できるサービスの質が低下することは避けたい。つまり、開示する情報と提供される情報のバランスをタイミング良くコントロールできるようにすることが重要である。

### 5.1 段階的オプトインの導入と環境オプトイン

ユーザがサービスを享受するために、SP 等に提供するプライバシ情報の詳細度や匿名度を確認し、設定/変更できる機能を実現した。ユーザがサービスの品質を確認しながら、提供する情報を詳細化できるこのような調整機能を「段階的オプトイン」と呼んでいる。

SP 側から見ると、ユーザに不安なくサービスを使ってもらうために、ユーザに開示を求める情報にいくつかのレベルを設けるやり方が考えられる。その場合には、ユーザがそのレベルを自由に選べ、かついつでも変更できなければならない。特に、最近の位置情報サービスで見られるその場所(店舗等) に訪問したことを表明・登録するタイプのサービスにおいてもこの考え方が有効である。このようなサービスでは、ユーザは、得られるサービスの質がかなり高いと期待できるなら、その場その時に限って必要な情報を(通常は開示しないような情報まで)開示する。一方、SP 側も、旬なユーザ情報を得ることにより、効果の高い告知や情報提供を行うことができる。環境オプトイン[19]は、ある環境に入ることで情報提供に同意したとするものであり、これを基本に、段階的オプトインの機能と組み合わせることで、一時的かつ場所依存のプライバシ情報の開示と得られる利便性とのバランスを調整可能とするものとして、環境オプトインを拡張定義して用いている。図 14に環境オプトインのイメージを示す。

筆者らは、このような動的な調整を含むオプトインが重要であると考え、以下の仕組みを実装した。

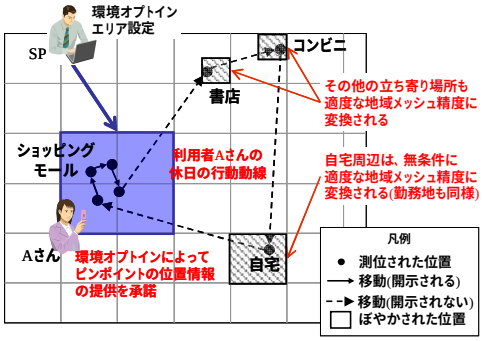


図 14 環境オプトインのイメージ

A) リアルタイムにプライバシ情報の開示/非開示、開示時の匿名化の程度を適切に設定/変更する仕組み
B) ユーザが匿名化ポリシーの変更を判断する際に、提供した情報やサービスの品質を確認できる仕組み
C) 取得情報利用の一時性を保証するために、有効な期間あるいは空間の範囲を限定し、そこから離脱した時に自動的に開示状態等が解除される仕組み

これらの仕組みの有効性を実証するために、実証実験では、匿名化ポリシーの変更を観測し評価を行なった。

### 5.2 段階的オプトインの評価

実証実験では、位置情報を用いた情報配信サービスにおいて匿名化ポリシーを随時、随意に変更できるようにしておき、ユーザには匿名化の度合いの高低(匿名化ポリシーの段階による違い)を表 4の匿名化ポリシーに対応した図 15 のようなイメージによって理解してもらい、ポリシーの設定変更を実測した。

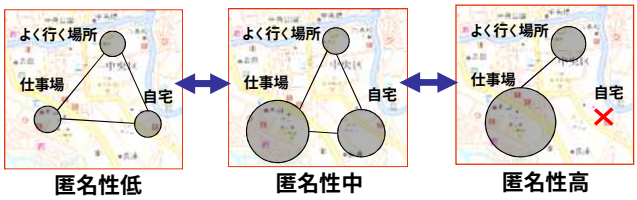


図 15 匿名度のイメージによる理解

実証実験で実際にポリシー変更を行った人数 380 名のうち、1 回のみ変更 203 人、2〜3 回変更 155 人、4 回以上変更 22 人となっている。人数の絶対数は少ないものの、匿名化ポリシーを変更する回数に応じて、一定の傾向が得られたこと、特に回数が多いユーザは、プライバシ情報の開示リスクとサービス有用性を比較して匿名性をその都度調整している、という傾向が観測できた (図 16)。

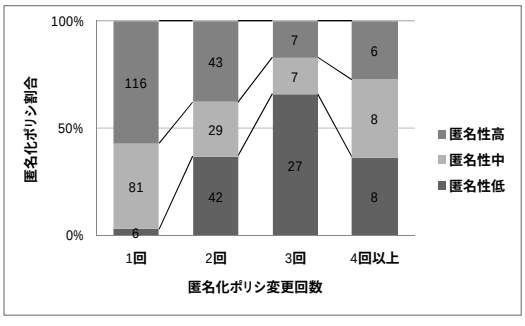


図 16 匿名化ポリシー変更の傾向

- 変更回数 1 回：匿名性高に変更するユーザが 6 割以上 →自身のプライバシ情報保護の観点から、まずは匿名性高への変更に働いたと推測される。
- 変更回数 2～3 回：匿名性を低に戻すユーザが増加 → サービス品質が低下したため、元に戻す方向に流れたと推測される。
- 変更回数 4 回以上：匿名性中～高に変更するユーザが増加 → サービス品質と開示のバランスを考慮した末に各人の基準に落ち着いたか、その都度変更することが習慣化したと推測される。

合わせて行った実験終了後のアンケートでは、匿名化ポリシーに関して「必要」、「まあ必要」と答えた被験者が 77.4% (必要・まあ必要・まあ不要・不要の４択形式) という高い値が得られており、環境オプトインのコンセプトに関しても、「位置情報を細めに取られたり、趣味嗜好を聞かれても、お得な情報を得られるなら問題ない」という形で理解され、高い受容度が得られた。

この段階的オプトイン(環境オプトイン)の仕組みを用いて、プライバシ情報の流通において匿名度を可視化すること、かつその度合いを段階的に制御できることが、ユーザの納得性を向上し、不安や懸念を軽減するための有効な手段になることが確認できた。

## 6. まとめ

情報大航海プロジェクトのマイ・ライフ・アシストサービス実証実験を通じて開発を行った、プライバシ情報の安全な流通と利活用のためのアーキテクチャおよび匿名化技術について述べた。

ユーザ情報のプライバシレベルと SP のトラストレベルに基づく開示制御および匿名化ポリシーをベースとしたアーキテクチャを構築した。ただし、PF モデルおよびフレームワーク化による効果の評価等は実証実験の範囲を超えており、今後の課題である。

匿名化技術に関しては、位置情報の履歴に着目し、行動範囲の ℓ-多様性を保証することで、前提知識を備えた攻撃者に対しても推定リスクを抑えることができる新しい匿名化方式を提案した。さらに、従来方式に比較して、データの有用性をほとんど低下させることなく匿名化できることも示した。

また、段階的オプトインの考え方を導入し、ユーザによる匿名化ポリシー変更を観測した。特に変更回数の多いユーザは、情報開示とサービス有用性を比較して匿名性をその都度調整しているという傾向が見られ、ユーザにとって有効な手段になりえることが確認できた。

今後は、複数 PF 間でのプライバシ情報の連携方式、その際に生じる恐れのあるデータの結合による匿名度の低下を抑える方式などに取り組んでいく必要があると考えている。

### 謝辞

本研究は、経済産業省「平成 19～21 年度情報大航海プロジェクト(モデルサービスの開発と実証)」における「マイ・ライフ・アシストサービス」実証実験の一環として実施したものである。関係各位に感謝の意を表する。

### 参考文献

[1] Organization for the Advancement of Structured Information Standards (OASIS), eXtensible Access Control Markup Language (XACML) Version 2.0, http://docs.oasis-open.org/xacml/2.0/ access_control-xacml-2.0-core-spec-os.pdf (2005)

[2] Kantara Initiative, Liberty Alliance ID-WSF 2.0 Specifications including Errata v1.0 Updates, http://projectliberty.org/resource_center/specifications/liberty_alliance_id_wsf_2_0_specifications_including_errata_v1_0_updates (2007)

[3] Kantara Initiative, Liberty Alliance ID-SIS 1.0 Specifications, http://projectliberty.org/resource_center/specifications/liberty_alliance_id_sis_1_0_specifications (2008)

[4] IETF, Geographic Location/Privacy (geopriv), http://datatracker.ietf.org/wg/geopriv/charter/

[5] 豊内 順一, 松田 純一, 情報収集エージェント技術に関する研究開発, 情報通信研究機構(オンライン), 入手先< http://itaku-kenkyu.nict.go.jp/seika/h14/seika/30/30hitachi.pdf >, (参照 2011-04-08)

[6]渡部 優, 間 伸一, 片山 頼明, 視聴情報等保護対策技術, 2006 年電子情報通信学会総合大会 通信ネットワーク利用放送技術 シンポジウム, (2006.3)

[7]國米 仁, 貝沼 達也, 古原 和邦, 個人情報の保護と活用を両立する情報通信プラットフォーム--プライバシー情報の大量漏洩を原理的に防止, 日本セキュリティ・マネジメント学会誌 19(1), 3-14, 2005-09

[8] Aggarwal, C, Philip, Y, Privacy-Preserving Data Mining: Models and Algorithms, Springer Publishing Campany, Incorporated (2008)

[9]Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M., ℓ-diversity: Privacy Beyond *k*-Anonymity, ACM Transactions on Knowledge Discovery from Data, Vol.1, 1 (2007).

[10] Gruteser, M., Grunwald, D., Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking, International Conference On Mobile Systems, Applications And Services, pp.31-42 (2003).

[11] Abul, O., Bonchi, F., Nanni, M., Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases, IEEE 24th International Conference on Data Engineering, pp.376-385 (2008).

[12] Bettini, C., Wang, X., Jajodia, S., Protecting Privacy Against Locationbased Personal Identification, In 2nd VLDB Workshop SDM, pp.185-199 (2005).

[13] Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M., ldiversity: Privacy Beyond *k*-Anonymity, ACM Transactions on Knowledge Discovery from Data, Vol.1, 1 (2007).

[14] Um, J., Jang, M., Jo, K., Chan, J., A New Cloaking Method Supporting both *k*-anonymity and ℓ-diversity for Privacy Protection in Location-Based Service, 2009 IEEE International Symposium on Parallel and Distributed Processing with Applications, pp79-85, (2009).

[15] Fung, B., Cao, M., Desai, B., Xu, H., Privacy protection for RFID data, Symposium on Applied Computing, Proceedings of the 2009 ACM symposium on Applied Computing, pp.1528-1535 (2009).

[16] Moving Objects: How to Hide a MOB in a Crowd?, Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, pp.72-83 (2009).

[17] Terrovitis, M., Mamoulis, N., Privacy Preservation in the Publication of Trajectory, International Conference On Mobile Data Management, pp.65-72 (2008).

[18] 廣田 啓一 他, 情報大航海プロジェクトにおける個人情報匿名化基盤の構築と検証, 第 50 回コンピュータセキュリティ研究発表会 (2010.7)

[19] 美濃 導彦, センサ情報の社会利用のためのコンテンツ化, 総合科学技術会議 科学技術連携施策群 情報の巨大集積化と利活用基盤技術開発連携群の活動 －情報爆発時代におけるイノベーション創出－シンポジウム (2009.12)

[20] A. Meyerson, A., Williams, R., On the complexity of optimal *k*-anonymity. In Proc. of the 23rd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 2004.