Computer Security Symposium 2011
19-21 October 2011

属性値を保持する際に効率的な撹乱・再構築法

菊池 亮 †　五十嵐 大 †　千田 浩司 †　濱田 浩気 †

† 日本電信電話株式会社 NTT 情報流通プラットフォーム研究所
180-8585 東京都武蔵野市緑町 3-9-11
{kikuchi.ryo, ikarashi.dai, chida.koji, hamada.koki}@lab.ntt.co.jp

あらまし 近年，データのプライバシーを保護しつつ統計分析が行える技術として，年齢等の属性値に対し撹乱と呼ばれる確率的な変更を施し，その後再構築と呼ばれる統計量のみ抽出する推定を行う撹乱・再構築法が注目されている．しかし撹乱・再構築法の計算量は他方式に比べ優れているものの未だ十分とは言えず，計算量の削減は大きな課題の一つである．一方，分析対象のデータには，様々な理由により値を変更せず保持する属性が存在することがある．そこで本稿では，分析対象のデータに値を保持する属性が含まれる場合，通常の撹乱・再構築法に比べ計算量を削減する手法を提案し，その正当性を示す．

An Efficient Reconstruction Algorithm where Some Attributes Are Conserved

Ryo Kikuchi†　Dai Ikarashi†　Koji Chida†　Koki Hamada†

†NTT Information Sharing Platform Laboratories, NTT Corporation
3-9-11, Midori-cho, Musashino-shi, Tokyo 180-8585 Japan
{kikuchi.ryo, ikarashi.dai, chida.koji, hamada.koki}@lab.ntt.co.jp

Abstract

1　はじめに

1.1　背景

近年，購買履歴や行動履歴等のデータを蓄積・分析し，商品のリコメンドやより良い都市開発等に活かしたいといった要望がある．しかし，このような個人に紐づくデータの利活用はプライバシーの問題があり第三者への提供や分析のアウトソーシング等が難しく，またデータを収集した者にとっても繊細な扱いが要求されるといった問題がある．そんな中，データのプライバシーを保護しつつ統計分析を行う技術であるプライバシー保護統計分析 [1, 6, 9] の研究が注目されている．

プライバシー保護統計分析技術として近年着目されているものに，セキュア関数計算 [6]，k-匿名法 [9, 10]，差分プライバシー法 [3]，撹乱・再構築法 [1, 2] がある．

セキュア関数計算は暗号化や秘密分散を基礎とする技術であり，セキュア関数計算を使用しない場合と全く同等のマイニング結果を得られ，さらに暗号学的な非常に高いレベルのデータ保護を提供できる．しかし，計算効率は必ずしも良いとは言えない．

k-匿名法は，k-匿名性と呼ばれる “全く同じように見えるデータが少なくとも k 個ある” という直感的なプライバシー保護指標，及びこの指標をさらに発展させた ℓ-多様性 [7] や t-近似性 [5] と呼ばれる指標に基づくプライバシー保護処理を行う手法である．しかし，k-匿名法では，統計値の精度を保ちながら k-匿名性等のプライバシー指標を満たすことが難しく，また，プライバシー保護処理はデータ全体に対して一括で行うため，データを一箇所に集める管理者には無防備であるという問題点がある．

差分プライバシー法は近年大きく注目されている分野で，“ある個人がデータに含まれていてもいなくても出力が（それほど）変化しない”

というプライバシー指標に基づくプライバシー保護処理を行う手法である．しかし，差分プライバシー法は対話型データベースという，統計値を得たい分析者がデータベースに対しリクエストを行い，データベースが統計値を分析者に渡すモデルを前提としており，分析者は自由に分析が行えない．また，繰り返し分析を行った際のプライバシー漏洩等未解決の問題も多数残されている．

撹乱・再構築法は，データに対し撹乱と呼ばれる確率的な処理を行うことでプライバシーを保護し，統計値を得る際には再構築と呼ばれるステップを経ることで統計値を復元する手法である．撹乱処理はデータを提供する提供者が自身で行うことが可能であるため，管理者に対してもプライバシーを保護できる．この撹乱・再構築法の中でも，維持置換撹乱と呼ばれる撹乱処理と，クロス集計と呼ばれる統計量を得るための反復ベイズ法という再構築処理を組み合わせた手法は，セキュア関数計算に比べ計算効率が良く，また k-匿名法と比較した際も同等のプライバシー保護度合いを得つつより精度の良い統計値を得られることが実験で示されている [8]．

しかし，撹乱・再構築法においても未だ幾つかの課題が残されており，そのうちの一つが計算量である．撹乱・再構築法の計算量は再構築における計算量が支配的であり，再構築の計算量はデータの取りうる値の広さ，すなわち値域に強く依存している．そのため例えば 1 歳刻みの年齢等データの値域が広い場合，計算量は膨大となるため，既に提案されている幾つかの高速化手法 [4, 8] を用いても，現実的な時間で実行できない場合がある．

1.2　センシティブ属性と撹乱・再構築法

あるデータに対しプライバシー保護の処理を行う際，そのデータの中には "分析に重要で変化させたくない"，もしくは "公開しても問題がない" ために値が変更されない属性がしばしば含まれる．このような値を，k-匿名法の文脈での呼び名に従い本稿ではセンシティブ属性と呼ぶ．

撹乱・再構築法では，筆者らの知る限りこのようなセンシティブ属性を含むデータは議論されておらず，センシティブ属性が存在するか否かに関わらず同様の計算量が必要とされる．前節で述べたとおり，撹乱・再構築法の課題として計算量の低減があることを鑑みるに，センシティブ属性を含むデータに対し処理を高速化し，計算量が削減できることが好ましい．

実際に計算量を削減しようとすると，撹乱処理は各属性ごとに独立して行われるため，センシティブ属性に撹乱を行わないことで自明に計算量を減らすことができる．

しかし，再構築処理では多数の撹乱データを集めた後，互いの属性値を用いた推定を行うため，意図的にセンシティブ属性に対し再構築処理を変更するとその変更の影響が全属性に及び，正確な統計値が得られなくなる可能性がある．そのため再構築処理での高速化は自明ではなかった．撹乱・再構築法の計算量で支配的なのは再構築処理であるので，結果的にセンシティブ属性の有無にかかわらず現状では高速化する前とほぼ同等の計算量が必要とされている．

1.3　本研究の成果

本研究では，クロス集計を得る再構築法である反復ベイズ法について，センシティブ属性が存在した場合に計算量が削減できるようなアルゴリズムを提案する．提案手法は特に，レコードの数が少ない，もしくはセンシティブ属性の値域が広い際に有効な高速化手法である．

さらに本研究では，提案する高速化手法を用いても従来の手法と全く同じ結果が得られることを示す．すなわち，提案手法は計算量を削減しつつ，得られる統計値は高速化する前に比べ全く同等のものを得ることができる．

2　準備

本章では，プライバシー保護処理を行う対象となるテーブルや，統計値であるクロス集計等の説明，及び定義を行う．

本稿において，集合 A，B に対し $|A|$ とは集合 A の要素数を意味し，$A \times B$ は集合の直積を意味する．フォント $\mathbf{x}$ はベクトルを意味する．

$/$ は整数の除算を表し，$\%$ は剰余算を表すこととする．すなわち，ある整数 R_0，R_1 について，R_0/R_1 は R_0 を R_1 で除算しその余りを切り捨てたものを表し，$R_0\%R_1$ は R_0 を R_1 で除算した際の余りを表すものとする．また，ベクトル $\mathbf{x}$，$\mathbf{y}$ について，$\frac{\mathbf{x}}{\mathbf{y}}$ は成分ごとの除算 $(\frac{x_0}{y_0}, \frac{x_1}{y_1}, \ldots)$ を表し，$(\mathbf{x})_q$ はベクトル $\mathbf{x}$ の q 番目の要素 x_q を表し，$\mathbf{x} \cdot \mathbf{y}$ は内積を表すものとする．

2.1　テーブル

本稿では，レコードと属性からなるテーブルをプライバシー保護処理の対象とする．

表 1: テーブルの例

ID	国籍	趣味
1	日本	サッカー
2	中国	野球
3	日本	野球
4	中国	野球

直感的な理解としては，表 1 にあるように，テーブルには年齢，趣味などの属性があり，各行のレコードには各属性に対し属性値が記述されているものである．

次に，テーブル T を定義する．まずテーブルを構成する要素を形式的に以下に示す．

- レコード：Rc_0 から Rc_{N-1} までの N レコード
- 属性：At_0 から At_{n-1} までの n 属性
- 各属性 At_a の値域：$Vl_0^{(a)}$ から $Vl_{M_a-1}^{(a)}$ までの M_a 値

また，属性の組み合わせの数を $M = \prod_{a<n} M_a$ とし，属性の組み合わせの集合を $\mathcal{V} = \{Vl_0^{(0)}, \ldots, Vl_{M_0-1}^{(0)}\} \times \cdots \times \{Vl_0^{(n-1)}, \ldots, Vl_{M_{n-1}-1}^{(n-1)}\}$ とする．以上の要素を用い，テーブル T を以下のように定義する．

定義 2.1. テーブル T とは，$T : \{Rc_0, \ldots, Rc_{N-1}\} \to \mathcal{V}$ を満たす写像である．

2.2 クロス集計

クロス集計とは，あるテーブルに対し同じ属性の組み合わせが幾つあるのかを数え上げたものである．例えば表 1 に対するクロス集計は表 2 のようになる．

表 2: クロス集計の例

	日本∧サッカー	日本∧野球	中国∧サッカー	中国∧野球
レコード数	1	1	0	2

これは，形式的に以下のように定義できる．

まず，表記の簡単のため幾つかの変数を導入する．属性 At_a の属性値に一意に対応する数として $Vl_\ell^{(a)}$ のとき $v_a = \ell$ となる数 v_a を定義し，以降 v_a を属性 At_a のクロス値と呼ぶ．次に，レコードの属性値 $\mathbf{v} \in \mathcal{V}$ と一意に対応する数として，各属性のクロス値がそれぞれ $\mathbf{v} = (v_0, \ldots, v_{n-1})$ のとき

$$p = f(\mathbf{v}) = \sum_{a<n} \left(v_a \prod_{b<a} M_b \right)$$

となる数 p を定義し，以降 p をレコードのクロス値と呼ぶ．p は a 桁目が M_a 進数であるような数となっており，f は $\mathbf{v} \in \mathcal{V}$ から $0 \leq p \leq M-1$ への全単射となっている．また，各属性のクロス値 v_a はレコードのクロス値 p から計算可能であり，特にこれを $d_p^{(a)}$ と書く．

$$d_p^{(a)} = \left(p / \prod_{b<a} M_b \right) \% M_a$$

以上を用い，クロス集計を以下のように定義する．

定義 2.2. 各 $0 \leq p \leq M-1$ において $x_p = \left| T^{-1} \left(\left\{ \mathbf{v} \mid \mathbf{v} \in \mathcal{V} \wedge (\forall a:\ v_a = d_p^{(a)}) \right\} \right) \right|$ であるような $\mathbf{x} = (x_0, x_1, \ldots, x_{M-1})$ を，テーブル T のクロス集計と呼ぶ．

3 撹乱・再構築法

撹乱・再構築法とは，情報提供者がデータに対しプライバシーを保護するための処理を施す撹乱と，撹乱後のデータから統計情報を抽出する再構築の 2 ステップからなる手法である．

撹乱ではデータに対し確率的な変化もしくはノイズ付加といった処理が行われる．これらは非可逆な操作であり，データの情報量を減少させるため，撹乱後のデータから元のデータを復元することはできず，プライバシーが保護される．

再構築では撹乱の逆にあたる推定等を行い，統計値のみを得る．このようなことが可能なのは，撹乱が行われるとレコード個々の情報量は減少するが，全体の統計量としては撹乱処理の期待値に収束するからである．

3.1 撹乱

撹乱･再構築法では，元のデータを確率的に撹乱する．この撹乱をどのように行うのかによって，プライバシーの保護度合いや統計分析の精度が変化する．

まず，属性 At_a に対する任意の撹乱を遷移確率行列 $A^{(a)}$ を用いて表す．テーブル T のうち，属性 At_a のクロス値のみ取り出したものを

$T(\cdot)_a$ とする．つまり，任意のレコード r について $T(r)_a = v_a$ である．次に，属性 At_a の遷移確率行列 $A^{(a)}$ に従い撹乱したテーブルを $T'(\cdot)_a$ とする．

以上を用いて，任意の撹乱は以下のように表現できる．

$$\Pr\left[T'(r)_a = Vl_l^{(a)} \mid T(r)_a = Vl_k^{(a)}\right] = A_{kl}^{(a)}$$

ただし $A^{(a)}$ は確率であるため，$\sum_{k<M_a} A_{kl}^{(a)} = 1$ を満たす．

直感的に，$A_{kl}^{(a)}$ とは属性 At_a の元々のクロス値が l だったものが，撹乱後に k になる確率である．$A^{(a)}$ には何も拘束条件を与えていないため，この表現は任意の撹乱を表している．

複数の属性を持つテーブルを属性ごとに撹乱した場合，撹乱は互いに独立のため，テーブル全体の遷移確率は属性ごとの遷移確率の積となる．これを A と書くことにする．つまり，

$$\begin{aligned} A_{pq} &= \Pr\left[f(T'(r)) = q \mid f(T(r)) = p\right] \\ &= \prod_{a<n} A^{(a)}_{d_p^{(a)} d_q^{(a)}} \end{aligned}$$

となる．

3.2 再構築

撹乱・再構築法では，前節で与えた撹乱を行ったデータに対し再構築という推定を行い統計量を抽出する．

再構築アルゴリズムとしては，クロス集計を計算する逆行列法，反復ベイズ法が提案されている．このうち逆行列法では，属性数が増えると誤差が急激に増大してしまうことが知られている [4] ため，本節では反復ベイズ法のみに注力する．

3.2.1 反復ベイズ法

反復ベイズ法とは，ベイズ推定を繰り返し用いて漸近的にクロス集計を推定する手法である．このアルゴリズムは，属性値が 2 値という条件の元で Aggrawal ら [2] によって提案され，後に五十嵐ら [4] によって 2 値属性から多値属性に拡張された．

$\mathbf{y} = (y_0, \ldots, y_{n-1})$ を，撹乱されたテーブルに対するクロス値とする．反復ベイズ法では，以下のアルゴリズムを初期値 $\mathbf{x}^0 = \mathbf{y}$ として実行する．

$$x_p^{i+1} = \sum_{q<M} y_q \frac{A_{pq} x_p^i}{\displaystyle\sum_{r<M} A_{rq} x_r^i}$$

このアルゴリズムを，終了条件 $\sum_p \left|x_p^{i+1} - x_p^i\right| < \epsilon N$ を満たすまで繰り返し実行し，終了条件を満たしたとき $\mathbf{x}^i$ を出力する．ここで，ϵ はあらかじめ決定された定数である．

この処理を素直に実行すると $\mathbf{x}^i$ から $\mathbf{x}^{i+1}$ を得るためにかかる計算量は $\mathcal{O}(M^3)$ となるが，五十嵐ら [4] はこれを

$$\mathbf{x}^{i+1} = \mathbf{x}^i \cdot \left(A\left(\frac{\mathbf{y}}{\mathbf{x}^i A}\right)^t\right)^t$$

と変形することにより $\mathcal{O}(M^2)$ に，また永井ら [8] は行列 A とベクトルとの計算方法を工夫することにより，最終的に $\mathcal{O}(M \log M)$ の計算量で実行可能であることが知られている．繰り返しの回数を I とおけば，全体の計算量は $\mathcal{O}(IM \log M)$ である．

4 センシティブ属性を含むテーブルの再構築

本章では，センシティブ属性という概念を撹乱・再構築法，特に再構築に反復ベイズ法を用いる手法に適用する．そして，センシティブ属性が存在するテーブルに対し，効率的に適用できる反復ベイズ法アルゴリズムを提案する．

4.1 設定

k-匿名性に合わせ，値を変化させる属性を準識別子と呼ぶこととする．

簡単のため，属性 $At_0, \ldots, At_{n-1}$ の n 個の属性のうち，最初の t 個の属性 $At_0, \ldots, At_{t-1}$ を値を変化させる属性，すなわち準識別子とし，残り $n-t$ 個の $At_t, \ldots, At_{n-1}$ を値を保持する属性，すなわちセンシティブ属性として議論を進める．なお，任意の順序で任意の数のセンシティブ属性と準識別子が存在する場合も同様の議論が可能である．

4.2 提案方式

本節では，センシティブ属性が存在する場合に計算量が削減できるような再構築アルゴリズムを提案する．

直感的には，全体に対して再構築を行うのではなく，センシティブ属性ごとに再構築法を行うことにより，計算量を削減する．

準識別子全体の値域を $M_Q = \prod_{a<t} M_a$，センシティブ属性の値域を $M_S = \prod_{t\le a<M} M_a$，さらにクロス値 p のセンシティブ属性の成分を $p_S = (p/M_Q)M_Q$ とおく．次に，行列 A を M_S 個に分割し，$A_{(p_S)}$ を i 行 j 列目が $A_{(p_S+i)(p_S+j)}$ であるような $M_Q \times M_Q$ 行列と定義する．また，同様に $\mathbf{x}$，$\mathbf{y}$ を M_S 個に分割し，ベクトル $\mathbf{x}_{(p_S)}$ を $(x_{p_S}, \ldots, x_{p_S+M_Q-1})$，ベクトル $\mathbf{y}_{(p_S)}$ を $(y_{p_S}, \ldots, y_{p_S+M_Q-1})$ と定義する．

以上を用いて，提案する再構築法は以下のように表される．

要素数が 0 でない $\mathbf{y}_{(p_S)}$ ごとに

$$\mathbf{x}_{(p_S)}{}^{i+1} = \mathbf{x}_{(p_S)}{}^{i} \cdot \left(A_{(p_S)} \left(\frac{\mathbf{y}_{(p_S)}}{\mathbf{x}_{(p_S)}{}^{i} A_{(p_S)}} \right)^t \right)^t$$

を実行する．要素数が 0 である $\mathbf{y}_{(p_S)}$ については，$\mathbf{x}_{(p_S)} = 0$ とし，$\mathbf{x} = \mathbf{x}_0 || \mathbf{x}_{(p_S)} || \cdots || \mathbf{x}_{M-p_S}$ を出力する．ただし，$||$ はベクトルの結合を表す．

通常の反復ベイズ法の 1 回の計算量が $\mathcal{O}(M \log M) = \mathcal{O}(M_S M_Q \log(M_S M_Q))$ だったのに対し，提案方式では，レコード数を N とすると計算量が $\mathcal{O}(\max(M_S, N) M_Q \log(M_Q))$ となっている．

これは，例えば具体例で何倍とか言う

4.3　提案方式と従来法の等価性

提案した再構築法は，従来の反復ベイズ法と同様の結果が得られることを示す．

定理 4.1. 提案方式は，通常の反復ベイズ法を用いた再構築と全く同じ出力 $\mathbf{x}$ を出力する．

[**証明**]．センシティブ属性 $At_t, \ldots, At_{n-1}$ の属性値は変化しないことから，$t \le a < n$ について，遷移確率行列 $A^{(a)}$ は

$$\forall k \ s.t. \ k \ne l : A^{(a)}_{kl} = 0$$

を満たす．

次に，遷移確率行列 A は

$$A_{pq} = \prod_{a<n} A^{(a)}_{d_p^{(a)} d_q^{(a)}}$$

であるので，$t \le a < n$ について $d_p^{(a)} = d_q^{(a)}$ ならば，$A_{pq} \ne 0$ が成り立つ．

p の定義より，$t \le a < n$ について $d_p^{(a)} = d_q^{(a)}$ ならば $p/M_Q = q/M_Q$ すなわち $p_S = q_S$ を満たすことに注意すると，遷移確率行列 A は以下の拘束条件を満たすことがわかる．

$$\forall k \ s.t. \ M_{S_k} \ne M_{S_l} : A_{kl} = 0.$$

ゆえに，

$$\begin{aligned}
x_p^{i+1} &= \sum_{q<M} y_q \frac{A_{pq} x_p^i}{\sum_{r<M} A_{rq} x_r^i} \\
&= \sum_{q<p_S} y_q \frac{A_{pq} x_p^i}{\sum_{r<M} A_{rq} x_r^i} \\
&\quad + \sum_{p_S \le q < (p_S+M_Q)} y_q \frac{A_{pq} x_p^i}{\sum_{r<M} A_{rq} x_r^i} \\
&\quad + \sum_{(p_S+M_Q) \le q < M} y_q \frac{A_{pq} x_p^i}{\sum_{r<M} A_{rq} x_r^i} \\
&= \sum_{p_S \le q < (p_S+M_Q)} y_q \frac{A_{pq} x_p^i}{\sum_{r<M} A_{rq} x_r^i}
\end{aligned}$$

が成り立つ．また同様に，

$$\begin{aligned}
\sum_{r<M} A_{rq} x_r^i &= \sum_{r<q_S} A_{rq} x_r^i \\
&\quad + \sum_{q_S \le r < (q_S+M_Q)} A_{rq} x_r^i \\
&\quad + \sum_{(q_S+M_Q) \le r < M} A_{rq} x_r^i \\
&= \sum_{q_S \le r < (q_S+M_Q)} A_{rq} x_r^i
\end{aligned}$$

が成り立ち，$p_S \le q < (p_S + M_Q)$ ならば $q_S = p_S$ であるので，結局，以下が成立する．

$$x_p^{i+1} = \sum_{p_S \le q < (p_S+M_Q)} y_q \frac{A_{pq} x_p^i}{\sum_{p_S \le r < (p_S+M_Q)} A_{rq} x_r^i}.$$

この式をベクトルと行列で表すと，

$$\sum_{p_S \le r < (p_S+M_Q)} A_{rq} x_r^i = (\mathbf{x}_{(p_S)} A_{(p_S)})_q$$

となるので，以下が成立する．

$$
\begin{aligned}
x_p^{i+1} &= \sum_{p_S \le q < (p_S+M_Q)} y_q \frac{A_{pq} x_p^i}{\sum_{p_S \le r < (p_S+M_Q)} A_{rq} x_r^i} \\
&= \sum_{p_S \le q < (p_S+M_Q)} y_q \frac{A_{pq} x_p^i}{(\mathbf{x}_{(p_S)} A_{(p_S)})_q} \\
&= x_p^i \sum_{p_S \le q < (p_S+M_Q)} y_q \frac{A_{pq}}{(\mathbf{x}_{(p_S)} A_{(p_S)})_q} \\
&= x_p^i \sum_{p_S \le q < (p_S+M_Q)} \left(\frac{\mathbf{y}_{(p_S)}}{\mathbf{x}_{(p_S)} A_{(p_S)}} \right)_q A_{pq} \\
&= x_p^i \sum_{p_S \le q < (p_S+M_Q)} \left(\left(\frac{\mathbf{y}_{(p_S)}}{\mathbf{x}_{(p_S)} A_{(p_S)}} \right) A_{(p_S)}{}^t \right)_q \\
&= x_p^i \sum_{p_S \le q < (p_S+M_Q)} \left(A_{(p_S)} \left(\frac{\mathbf{y}_{(p_S)}}{\mathbf{x}_{(p_S)} A_{(p_S)}} \right)^t \right)_q^t
\end{aligned}
$$

p_S は p が M_Q 増加するごとに 1 増加することに気をつけると，

$$
\mathbf{x}_{(p_S)}{}^{i+1} = \mathbf{x}_{(p_S)}{}^i \cdot \left(A_{(p_S)} \left(\frac{\mathbf{y}_{(p_S)}}{\mathbf{x}_{(p_S)}{}^i A_{(p_S)}} \right)^t \right)^t
$$

となる．ここで，$\mathbf{y}_{(p_S)}$ の要素が全て 0 であれば自明に $\mathbf{x}_{(p_S)}$ も 0 であるので提案方式と同様の結果となっている □

5 まとめ

本稿では，プライバシー保護処理後にも値が保持されるようなセンシティブ属性の概念を撹乱・再構築法に適用し，特に再構築アルゴリズムである反復ベイズ法について，センシティブ属性を含むテーブルを効率的に処理できるアルゴリズムを提案した．

提案手法はセンシティブ属性の値域が広い際に効果が大きい手法である．

参考文献

[1] Agrawal, R., Srikant, R.: Privacy-preserving data mining. SIGMOD, 2000.

[2] Agrawal, R., Srikant, R., Thomas, D.: Privacy Preserving OLAP. SIGMOD Conference ACM, pp. 251-262, 2005.

[3] Dwork, C.: Differential Privacy. ICALP, 2006.

[4] 五十嵐 大, 千田 浩司, 高橋 克巳: 多値属性に適用可能な効率的プライバシー保護クロス集計. CSS, 2008.

[5] Li, N., Li, T., Venkatasubramanian.: t-closeness: privacy beyond k-anonymity and ℓ-diversity. ICDE, 2007.

[6] Lindell, Y., Pinkas, B.: Privacy Preserving Data Mining. CRYPTO, 2000.

[7] Machanavajjhala, A., Gehrke, J., Kiefer,D., Venkitasubramanian, M.: ℓ-diversity: privacy beyond k-anonymity. ICDE, 2006.

[8] 永井 彰, 五十嵐 大, 濱田 浩気, 松林 達史: クロネッカー積を含む行列積演算の最適化による効率的なプライバシ保護データ公開技術. SCIS, 2010.

[9] Sweeney, L.: k-anonymity: a model for protecting privacy. Int. Jour. of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002.

[10] Samarati, P., Sweeney, L.: Generalizing data to provide anonymity when disclosing information (abstract). PODS, 1998.