Vol.2011-DPS-146 No.12
Vol.2011-CSEC-52 No.12
2011/3/10

統計的開示制御を考慮したセキュアマッチングプロトコル

千田浩司[†1] 寺田雅之[†2] 山口高康[†2]
五十嵐大[†1] 濱田浩気[†1] 高橋克巳[†1]

セキュアマッチングプロトコルとは，複数の主体が各々持つデータを互いに提供せず照合可能とし，照合結果やその集計値のみを出力する暗号プロトコルであり，各組織が保有するパーソナル情報 (個人に関連する情報の総称) を統合利用する際のセキュリティ対策として有効である．しかし既存方式は照合結果やその集計値から生じる個人識別やプライバシ侵害のリスクについては考慮されていない．そこで本稿では，統計理論やデータ工学等の分野で研究されている，データ公開において個人識別やプライバシ侵害のリスク低減を図る統計的開示制御手法をセキュアマッチングプロトコルに適用することを試みる．具体的には，統計的開示制御手法を適用したクロス集計結果を効率良く求めるセキュアマッチングプロトコルを提案する．

A Secure Matching Protocol with Statistical Disclosure Control

Koji Chida,[†1] Masayuki Terada,[†2]
Takayasu Yamaguchi,[†2] Dai Ikarashi,[†1]
Koki Hamada[†1] and Katsumi Takahashi[†1]

The secure matching protocol is a cryptographic protocol that can acquire the matching result or its aggregation value of two or more data sets without disclosing them to one another. Applications of the protocol include the secure integration of personal information shared to multiple organizations. However, the existing secure matching protocols have little attention to the risk caused by the disclosure of a matching result or its aggregation value, such as the individual identification and the privacy violation. In this paper, as a countermeasure against the foregoing risk, we explore a privacy-enhancing secure matching protocol applied to the statistical disclosure control (SDC) method which has been studied in the fields of statistic theory and data engineering. More precisely, we propose a modified secure matching protocol that outputs the cross tabular data applied to a SCD method.

1. はじめに

ICT の発達に伴い各組織が保有するパーソナル情報の統合利用の有用性が注目されてきている．1) によれば，パーソナル情報とは個人に関連する情報の総称であり，法が対象としている個人情報のみならず，単独では一般に特定の個人を識別できない情報 (健康情報や生活行動情報等のいわゆるライフログ情報) も含むとしている．パーソナル情報の統合利用は，各組織が共通して持つ各個人に一意に対応するデータ (これを**照合キー**と呼ぶ) を用いてパーソナル情報を照合することで，組織を跨ったパーソナル情報の連結が可能となる．これにより，より高度な統計分析やデータマイニングが可能となり，社会統計や医療研究，パーソナライゼーションサービス等に資することが期待される．しかしパーソナル情報を別の組織に提供する際は，個人情報保護やプライバシ保護の観点から特に慎重な配慮が求められる．連結したパーソナル情報の漏洩は個人に深刻な被害を与えかねず，またプライバシに関わる情報が第三者に渡ること自体を嫌う個人への配慮も重要な課題であろう．

筆者らはこれまで，互いにデータを提供せず照合結果やその集計値を取得できる**セキュアマッチング**[2),3)] と呼ばれる (プライベートマッチング，秘匿共通集合計算 (secure set-intersection) とも呼ばれる) 暗号プロトコルの応用研究を行ってきた[4),5)]．セキュアマッチングプロトコルの基本方式は，n 主体 P_i $(i = 1, \ldots, n)$ がそれぞれデータ集合 S_i を入力として，互いに S_i を提供せず共通集合 $S_1 \cap S_2 \cap \cdots \cap S_n$ またはその要素数を出力する．現在まで様々なセキュアマッチングプロトコルの応用が提案されており，例えば，共通集合の要素数の閾値判定[3)]，和集合の要素数計算[2)] やその閾値判定[6)]，そして集合が互いに疎かどうかの判定[7),8)] 等がある．筆者らは 4) において，各主体が持つ照合キーとパーソナル情報の組からなる集合から，同一の照合キーで連結したパーソナル情報のクロス集計結果のみを出力するセキュアマッチングプロトコルの応用を提案した．また 5) において同様に，連結したパーソナル情報の暗号文のみを出力する手法を提案した．当該暗号文を**秘匿計算** (secure computation)[9),10)] の入力とすることで，必要最小限の統計量や各種分析結果のみを出力することも可能となる．なお秘匿計算は現在まで様々な演算における手法が提案されている．具体的には 11) 等を参照されたい．

†1 日本電信電話 (株) 情報流通プラットフォーム研究所
NTT Information Sharing Platform Laboratories
†2 (株)NTT ドコモ 先進技術研究所
NTT DOCOMO Research Laboratories

Vol.2011-DPS-146 No.12
Vol.2011-CSEC-52 No.12
2011/3/10

前述のとおり，セキュアマッチングプロトコルは，各組織が保有するパーソナル情報を互いに提供せず照合結果や意中の統計量を得ることができる．しかし既存のセキュアマッチングプロトコルは，出力となる照合結果や統計量等からの個人識別やプライバシ侵害のリスクについては考慮されていない．そこで本稿では，統計理論やデータ工学等の分野で研究されている，データ公開において個人識別やプライバシ侵害のリスク低減を図る**統計的開示制御** (statistical disclosure control)/統計的開示制限 (statistical disclosure limitation) 手法[12),13)] に着目し，同手法を適用した照合結果や統計量を出力するセキュアマッチングプロトコルの実現可能性について考察する．特に複数組織のパーソナル情報のクロス集計に統計的開示制御を適用した結果を出力する，効率の良いセキュアマッチングプロトコルを提案する．なお本稿では 14) に倣い統計的開示制御および統計的開示制限を同義語とみなし，代表して前者を用いる．

以降，2 節で代表的な統計的開示制御手法とその関連話題について触れる．3 節では提案方式が用いる既存技術について説明する．そして 4 節で，統計的開示制御手法を適用したクロス集計結果を効率良く求めるセキュアマッチングプロトコルを提案する．5 節では関連研究について紹介する．最後に 6 節で本稿をまとめる．

2. 統計的開示制御

2.1 準　　備

各主体が保有する個々のパーソナル情報は以下の変数からなるものとする[⋆1]．

正識別子 (formal identifier)　個人を一意に識別できる変数．例えば氏名，住所のような変数の組み合わせは無視できない確率で正識別子となる[14)]．

準識別子 (quasi-identifier)　間接的に個人を識別できる変数．性別や年齢のような変数は間接的に個人の識別に用いることができる[15)]．

センシティブ変数 (sensitive variable)　正識別子または準識別子以外で，個人のプライバシに関するもの等，他人にむやみに知られたくない変数．以降，センシティブ変数の値を**センシティブデータ**と呼ぶ場合がある．

ある個人の正識別子，準識別子およびセンシティブ変数の値の少なくとも一つからなるパーソナル情報を**レコード**と呼び，レコードの集合からなるデータを**ミクロデータ**と呼ぶ．

⋆1 本稿の用語の定義は 14)，15) に基づくが，14)，15) からも分かるように，一般にはパーソナル情報に限ったものではなく，世帯や事業所・企業等も含めたより広義の「個体 (individual)」の情報における用語である．

正識別子を含まないミクロデータを**匿名化データ**と呼ぶ．また匿名化データにおける一つ以上の変数の集計値の集合からなるデータを**集計表**と呼ぶ．なお二変数からなる集計表は一般に行と列からなる表形式で与えられ，行と列が交わった場所を**セル**と呼ぶ．一変数や三変数以上の集計表についても同様にセルを定義することができる．またクロス集計結果は二変数以上からなる集計表である．

2.2 代表的な手法

統計的開示制御は，匿名化データや集計表の公開によって生じる個人識別やプライバシ侵害のリスクを低減させるために，匿名化データや集計表を加工する技術の総称であると捉えることができる．以下に代表的な統計的開示制御手法を紹介する．各手法の詳細は 12)，13)，15)，16) 等を参照されたい．

大域的再符号化 (global recoding)　変数の項目 (変域) をより一般化されたものに置き換える．

局所再符号化 (local recoding)　一部のレコードの値をより一般化されたものに置き換える．

上位/下位符号化 (top/bottom coding)　変数の値に上限/下限値を設定し，これを超える値を直接開示しない．

局所秘匿 (local suppression)　特定のセルやレコードの値を秘匿する．所定の基準に満たないセルの値を秘匿することを一次秘匿 (primary suppression) と呼ぶ．またセルの値の小計から一次秘匿した値を推定されることを防ぐために，基準を満たすいくつかのセルの値も秘匿することを二次秘匿 (secondary suppression) と呼ぶ．

ミクロアグリゲーション (microaggregation)　複数のレコードをグループ化し，グループ毎にレコードの値を代表値に置き換える．

丸め法 (rounding)　切り上げ，切り捨て等，一定の基準に従ってセルやレコードの値を丸める．最も近い定数の倍数に丸める Conventional Rounding の他，いくつかの所定の処理からランダムに選択する Random Rounding や，小計が等しくなるように丸める Controlled Rounding がある．

ランダム撹乱 (random perturbation)　一定の分布に従ったノイズをセルやレコードの値に混入させる．値が量的データであればノイズ付加等により元の値を歪め，質的データであれば確率的に他の値に置き換える方法がある．

PRAM(post randomization method)　マルコフ推移確率行列に基づき，確率的にレコードの値を置き換える．

Vol.2011-DPS-146 No.12
Vol.2011-CSEC-52 No.12
2011/3/10

スワッピング (swapping) 異なるレコード間で同一変数の値を入れ替える．類似したレコード間で入れ替えるランクスワッピング (rank swapping) がある．

2.3 リスク/有用性評価

統計的開示制御に関連して，匿名化データや集計表の公開によるリスクや有用性を評価する指標がいくつか知られている．17) によれば，匿名化データや集計表を公開するリスクは**身元開示** (identity disclosure) および**属性開示** (attribute disclosure) であるとし，これらを合わせて**開示リスク** (disclosure risk) と呼ぶ．身元開示リスクは，レコードが正識別子と対応付くリスクである．属性開示リスクは，匿名化データや統計量から特定個人のセンシティブデータが知られるリスクであり，プライバシ侵害のリスクと捉えることができる．また，身元開示に関して基本的な概念は**母集団一意** (population unique) の概念とされる[15]．ある個人の母集団において，一つ以上の変数からなる値が一意となる個人が母集団一意である．同様に k 意となる個人は母集団 k 意である．また**標本一意** (sample unique) とは，母集団から抽出された標本において，一つ以上の変数からなる値が一意となる個人を指す．同様に k 意となる個人は標本 k 意である．母集団一意であれば標本一意だが，その逆は一般に成り立たない．実際，開示リスクの評価における基本的な問題は，母集団一意数の推定問題とされる．

以下に，匿名化データや集計表の公開によるリスクや有用性を評価する代表的な指標を紹介する．

(n,k) **ルール** セルの値の k % 以上を n 個のレコードの値の和で占めた場合，対象となるセルを公開すべきでないとするルール[18]．占有ルール (dominance rule) とも呼ばれる．

(p,q) **ルール** 集計表公開前におけるレコードの値の推定確率を q 以内としたとき，公開後に p 以内となる場合，対象となるセルを公開すべきでないとするルール[18]．事前事後ルール (prior-posterior rule) とも呼ばれる．特に $q = 100\%$ のとき，P パーセントルールと呼ばれる．

k-匿名性 (k-anonymity)[19] 匿名化データにおいて同一の準識別子 (の組) の値が k 以上存在するかどうかによって身元開示リスクを測る指標．

ℓ-多様性 (ℓ-diversity)[20] 匿名化データにおいて同一の準識別子 (の組) の値を持つレコード全体について，異なるセンシティブデータが ℓ 通り以上存在するかどうかによって属性開示リスクを測る指標．

Individual Risk 母集団一意性に基づき，一つ以上の変数からなる値の頻度から身元開示リスクを測る指標[21]．

レコードリンケージ (record linkage) 元の匿名化データと加工後のデータを照合することで身元開示リスクを測る指標[22]．

Differential Privacy[23] 「ある個人のレコードが含まれていてもそうでなくても出力が変化しない」ことを一定の基準の下で保証するための指標．統計量に Laplase ノイズを付加して基準を満たす手法が提案されている．

情報損失 (information loss) 元の匿名化データと加工後のデータから得られる統計量の差分を計算することで有用性を測る指標[24]．

3. 既存技術

本節では，筆者らが 5) で提案したセキュアマッチングプロトコルの応用方式 (以降，CIHT11 プロトコルと呼ぶ) について紹介し，簡単な考察を述べる．提案方式は CIHT11 プロトコルの利用を前提とする．なお 5) では k-匿名性を考慮したセキュアマッチングプロトコルについて触れているが，具体的な方法は与えていない．

主体 P_i が持つデータ集合 S_i は，照合キーとパーソナル情報の組からなる集合とする．E を閾値準同型暗号関数とする．特に閾値は n, すなわち復号鍵は全ての P_i $(i = 1, \ldots, n)$ に分散されているものとする．このとき CIHT11 プロトコルは，P_i の入力を S_i として，同一の照合キーでパーソナル情報の暗号文を連結した p 個のデータ

$$((E(a_{1,k}), \ldots, E(a_{m,k})) \quad (k = 1, \ldots, p) \tag{1}$$

を出力する．ここで $a_{j,k}$ は何れかの P_i が保有するパーソナル情報の値を表し，以降 $a_{j,k}$ を値に取る変数を便宜上 j とする．CIHT11 プロトコルの特徴は，連結したパーソナル情報の秘匿性を保つことに加え，照合キーを他の主体に一切提供すること無く，同一の照合キーを持つパーソナル情報の連結が可能なことである．

式 (1) における任意の暗号文は，何れの主体も復号できず，平文を推定することは困難となる．さらに (1) を例えば 25) で提案されているような秘匿計算の入力とすることで，連結したパーソナル情報から得られる統計量や分析結果のみを求めることも可能となる．また，式 (1) を得るプロトコルが安全であれば，出力である式 (1) および入力のサイズから得られる情報を除き，各主体は他の主体の入力に関する有意な情報を得られない．出力および入力のサイズから得られる他の主体の入力に関する情報は次のようなものである．

(1) 連結したパーソナル情報の暗号文の組の数 $(= p)$
(2) 各主体が保有するパーソナル情報の変数
(3) 各パーソナル情報のサイズ

Vol.2011-DPS-146 No.12
Vol.2011-CSEC-52 No.12
2011/3/10

(4) 集合 S_i の要素数

上記のうち，(2)〜(4) は特に問題ないか単純な対応が可能である．(2) はデータを利用する上で必須の情報といえる (変数の値では無いことに注意)．(3) は変数毎に値を固定長とすれば良い．(4) は組織の機密情報の観点から知られたくないかもしれない．その場合，統計量等に影響しない形でダミーデータを挿入する対策や，標本のみを利用するといった対策が挙げられる．最後に (1) については，開示リスクに係わる問題となり得る．具体的には，定数 k について $p < k$ であれば，式 (1) を復号して得られるデータは明らかに k-匿名性を満たさないことになる．ただし主体 P_i からみれば，自身が保有するパーソナル情報のうち，どの個人のパーソナル情報が結合したかは分からないため，一般にはリスクが低いといえる．また統計的開示制御の観点からは，そもそも連結可能なレコードの総数が 2.3 節で述べた指標を下回るような例外的なケースでしか問題とならない．したがって本稿では議論の対象外とする．

4. 提案方式

前節で述べた CIHT11 プロトコルに基づき，統計的開示制御を適用した集計表を求める方式を提案する．提案方式は以下の構成要素からなる．

(A) CIHT11 プロトコルの出力である，式 (1) で与えられる暗号文の組から，加法性を持つ閾値準同型暗号関数を利用して，効率良く各セルの集計値の暗号文を計算する．

(B) 既存の秘匿計算を用いて，集計値を秘匿したまま，2.2 節で述べた統計的開示制御を適用する．

(C) (B) の結果を復元し，統計的開示制御を適用した集計表を得る．

上記の手続き (A) において，加法性を持つ閾値準同型暗号関数とは，任意の暗号文 $E(a)$, $E(b)$ および適当な演算子 $\boxplus$ について $E(a+b) = E(a) \boxplus E(b)$ が成り立つような閾値準同型暗号関数であり，Paillier 暗号は $E(a+b) = E(a)E(b)$ が成り立つ．以降，Paillier 暗号の利用を前提とする．なお手続き (A) および (B) は，対象となる演算を論理回路で表現することで，論理回路のサイズに比例する計算量および通信量で秘匿計算が実現できる (例えば 10)，25)，26))．しかし実際には，比例定数が大きいため，大規模なデータに対しては適用困難となる．特にセルの個数は，実用上は一般に式 (1) で与えられるデータ数 p よりも十分小さいと考えられるため，手続き (A) の効率化が大きな課題となる．

手続き (B) については，以下の統計的開示制御が適用可能であることを示す．

(B1) 局所秘匿．具体的には，各集計値を秘匿したまま閾値 t 以上かどうか判定し，t 未満と判定された集計値は復号しない (一次秘匿)．

(B2) 丸め法．具体的には，非負整数からなる各集計値を秘匿したまま最も近い定数 (例えば 5) の倍数に置き換える (Conventional Rounding)．

(B3) ランダム撹乱．具体的には，ノイズを暗号化した状態で発生させ，集計値を秘匿したままノイズを付加する．

提案方式は，式 (1) における $a_{j,k}$ は全て二値 (0 または 1) と仮定する．以下では話を簡単にするため，式 (1) において $m = 2$ として，手続き (A) および (B1) の具体的な処理について説明する．

先ず変数 $j = 1$ の集計値の暗号文を求める．これは手続き (A) となり，単純に $E(a_{1,k})$ $(k = 1, \ldots, p)$ から $E(\sum_{k=1}^{p} a_{1,k}) = \prod_{k=1}^{p} E(a_{1,k})$ を計算する．

次に手続き (B1) として，*Bit-Decomposition* プロトコル (例えば 27))，および 25) で提案されている論理回路の秘匿計算を組み合わせて実行し，$\sum_{k=1}^{p} a_{1,k}$ を秘匿したまま $t \leq \sum_{k=1}^{p} a_{1,k}$ および $t \leq p - \sum_{k=1}^{p} a_{1,k}$ の真偽を判定する．前者の判定式は $a_{1,k} = 1$ となる集計値に対してのものであり，後者は $a_{1,k} = 0$ となる集計値に対してのものである．Bit-Decomposition プロトコルは，整数 a の暗号文 $E(a)$ から，a を秘匿したまま，a のビット毎の暗号文 $E(a_0), E(a_1), \ldots$ $(a_0, a_1, \ldots$ は a の各ビットとする) を計算できる暗号プロトコルである．また 25) の秘匿計算は，ビット毎の暗号文を入力として，各ビットを秘匿したまま，論理回路の演算結果を求めることができ，この場合は大小比較回路とする．

上記の真偽判定において，何れかが偽となれば，$E(\sum_{k=1}^{p} a_{1,k})$ は復号せず処理を終了する．何れも真であれば，全ての暗号文 $E(a_{1,k})$ を復号する．

次に $E(a_{2,k})$ について，$a_{1,k}$ が 0 であるものと 1 であるものに分類し，分類された暗号文の集合毎に，上記手続き (A) および (B1) を行なう．すると (A) で二変数 $j = (1, 2)$ の集計値の暗号文を求め，(B1) で各集計値の閾値判定が行われる．

次に (B2) の具体的な実現方法について述べる．これは手続き (A) によって得られる特定の集計値 a の暗号文 $E(a)$ を考えれば十分であり，二分木探索を用いて効率良く求めることができる．すなわち，a の上限値を m，定数を c として，a に最も近い c の倍数を得るために，$m/2$ に最も近い c の倍数からはじめて，二分木探索によって a 以下の最大の c の倍数 c' を求める．そして a について c' と $c' + c$ の近い方の値を出力するような論理回路を構成し，(B1) 同様，$E(a)$ を入力として Bit-Decomposition プロトコルおよび 25) の秘匿計算を実行すれば良い．

最後に (B3) の具体的な実現方法について述べる．(B2) 同様，手続き (A) によって

Vol.2011-DPS-146 No.12
Vol.2011-CSEC-52 No.12
2011/3/10

得られる特定の集計値 a の暗号文 $E(a)$ を考えれば十分である．ノイズ付加は，各主体 P_i が生成したノイズ r_i を他の主体に秘匿したまま，それらの平均 $r = \sum_{i=1}^{n} r_i/n$ をノイズとして a に付加する．すなわち，各主体 P_i は $E(r_i/n)$ を計算して互いに送信し，$E(a+r) = E(a)\prod_{i=1}^{n} E(r_i/n)$ を求めれば良い．

5. 関 連 研 究

4 節では，加法性を持つ閾値準同型暗号関数を利用して，集計表におけるセルの値のような統計量に対して統計的開示制御を適用できる効率の良いプロトコルを提案した．一方で，ミクロデータから得られる統計量や分析結果をユーザの問合せに応じて返す，統計データベース (statistical database) の開示リスクについても研究が進められている．複数回の問合せによって蓄積される統計量や分析結果の開示リスクは自明でなく，15) での例を挙げると，ある統計データベースに 50 歳の人のレコードが一つだけ含まれていることが分かっており，51 歳以上の人の平均年収と 50 歳以上の人の平均年収を問合せることで，50 歳の人の年収が特定されてしまう．

統計データベースにおける開示リスクの対策は，クエリ監査 (query auditing)[28),29)] とクエリ推論制御 (query inference control) に大別できる．クエリ監査はいくつかの問合せを拒否し，クエリ推論制御は統計量や分析結果にノイズを付加する．クエリ推論制御に関する最近の結果として，Differential Privacy が注目を集めている．

筆者らもこれまで統計的開示制御の指標に関するいくつかの検討を行ってきた．30)，31) ではそれぞれ，維持-置換撹乱 (retention-replacement perturbation)[32)] と呼ばれるランダム撹乱を k-匿名性および ℓ-多様性の指標に基づき評価可能であることを示した．特に維持-置換撹乱は再構築法 (reconstruction method) と呼ばれる手法により，高い精度で統計量を抽出できる特徴を持つ．ランダム撹乱は変数毎に独立の操作であり，セキュアマッチングプロトコルの実行前に処理できるという利点がある．また 33) では，Differential Privacy の不完全性を指摘するとともに，匿名化データにおいても同様の指標が適用できることを示した．これも同様にランダム撹乱を変数毎に独立の操作として行うため，セキュアマッチングプロトコルの実行前に処理できる．

6. ま と め

各組織が保有するパーソナル情報を安全に統合利用する技術要素としてセキュアマッチングプロトコルを挙げ，セキュアマッチングプロトコルの出力から個人識別やプライバシ侵害といった開示リスクがあることを指摘し，その対策について論じた．特にセキュアマッチングプロトコルによって得られるクロス集計結果の出力制御を，統計理論やデータ工学等の分野で研究されている統計的開示制御手法に基づいて実現可能であることを示した．

参 考 文 献

1) 経済産業省，パーソナル情報研究会報告書-個人と連結可能な情報の保護と利用のために-，http://www.meti.go.jp/report/data/g81110aj.html, 2008 年 11 月.
2) R. Agrawal, A. V. Evfimievski, and R. Srikant, Information sharing across private databases, ACM SIGMOD 2003, pp. 86–97, 2003.
3) M. J. Freedman, K. Nissim, and B. Pinkas, Efficient private matching and set intersection, Eurocrypt 2004, LNCS 3027, pp. 1–19, Springer-Verlag, 2004.
4) 千田 浩司, 寺田 雅之, 山口 高康, 五十嵐 大, 濱田 浩気, セキュアマッチングを用いた組織間クロス分析, コンピュータセキュリティシンポジウム 2010 (CSS2010), 2010.
5) 千田 浩司, 五十嵐 大, 濱田 浩気, 高橋 克巳, 匿名等結合プロトコルとその応用, 2011 年暗号と情報セキュリティシンポジウム (SCIS2011), 2011.
6) L. Kissner and D. X. Song, Privacy-preserving set operations, Crypto 2005, LNCS 3621, pp. 241–257, Springer-Verlag, 2005.
7) S. Hohenberger and S. A. Weis, Honest-verifier private disjointness testing without random oracles, PET 2006, LNCS 4258, pp. 277–294, Springer-Verlag, 2006.
8) A. Kiayias and A. Mitrofanova, Testing disjointness and private datasets, FC 2005, LNCS 3570, pp. 109–124, Springer-Verlag, 2005.
9) A. C. Yao, Protocols for secure computations (extended abstract), FOCS '82, pp. 160–164, IEEE Press, 1982.
10) A. C. Yao, How to generate and exchange secrets (extended abstract), FOCS '86, pp. 162–167, IEEE Press, 1986.
11) Y. Lindell and B. Pinkas, Secure multiparty computation for privacy-preserving data mining, The Journal of Privacy and Confidentiality, Vol. 1, No. 1, pp. 59–98, 2009.
12) L. Willenborg and T. de Waal, Statistical disclosure control in practice, Lecture Notes in Statistics, Vol. 111, Springer, 1996.
13) L. Willenborg and T. de Waal, Elements of statistical disclosure control, Lecture Notes in Statistics, Vol. 155, Springer, 2001.
14) 独立行政法人 統計センター 研究センター, 統計データ開示抑制に関する用語集 改訂版 (対訳) 2005 年 8 月, http://www.nstac.go.jp/services/pdf/skk-yogosyu2.pdf, 2006 年 6 月.
15) 竹村 彰通, 個票開示問題の研究の現状と課題, 統計数理, 第 51 巻, 第 2 号, pp. 241–260, 統計数理研究所, 2003.

Vol.2011-DPS-146 No.12
Vol.2011-CSEC-52 No.12
2011/3/10

16) A. Hundepool, J. Domingo-Ferrer, L. Franconi, S. Giessing, Rainer Lenz, J. Naylor, E. S. Nordholt, G. Seri, and P.-P. De Wolf, Handbook on statistical disclosure control (version 1.2), http://neon.vb.cbs.nl/casc/handbook.htm, Jan. 2010.
17) D. Lambert, Measures of disclosure risk and harm, Journal of Official Statistics, Vol. 9, No. 2, pp. 313–331, 1993.
18) 瀧 淳弘, 集計表におけるセル秘匿問題とその研究動向, 統計数理, 第 51 巻, 第 2 号, pp. 337–350, 統計数理研究所, 2003.
19) L. Sweeney, k-anonymity: a model for protecting privacy, Int'l Journal on Uncertainty, Fuzziness and Knowledge-based Systems, Vol. 10, No. 5, pp. 557–570, 2002.
20) A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, ℓ-diversity: privacy beyond k-anonymity, ACM Trans. on Knowledge Discovery from Data (TKDD), Vol. 1, No. 1, ACM Press, 2007.
21) R. Benedetti and L. Franconi, Statistical and technological solutions for controlled data dissemination, New Techniques and Technologies for Statistics 1998 (NTTS'98), Vol. 1, pp. 225–232, 1998.
22) J. Domingo-Ferrer and V. Torra, A quantitative comparison of disclosure control methods for microdata, Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies, pp. 113–134, Elsevier Science, 2001.
23) C. Dwork, Differential privacy, ICALP 2006, LNCS 4052, pp. 1–12, Springer-Verlag, 2006.
24) J. Domingo-Ferrer and J. M. Mateo-Sanz, An empirical comparison of SDC methods for continuous microdata in terms of information loss and disclosure risk, Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, 2001.
25) R. Cramer, I. Damgård, and J. B. Nielsen, Multiparty computation from threshold homomorphic encryption, Eurocrypt 2001, LNCS 2045, pp. 280–300, Springer-Verlag, 2001.
26) B. Schoenmakers and P. Tuyls, Practical two-party computation based on the conditional gate, Asiacrypt 2004, LNCS 3329, pp. 119–136, Springer-Verlag, 2004.
27) T. Nishide and K. Ohta, Multiparty computation for interval, equality, and comparison without bit-decomposition protocol, PKC 2007, LNCS 4450, pp. 343–360, Springer-Verlag, 2007.
28) D. Dobkin, A. Jones, and R. Lipton, Secure databases: protection against user influence, ACM Trans. on Databases Systems, Vol. 4, No. 1, 1979.
29) S. Reiss, Security in databases: a combinatorial study, Journal of ACM, Vol. 26, No. 1, 1979.
30) 五十嵐 大, 千田 浩司, 高橋 克巳, k-匿名性の確率的指標への拡張とその応用例, コンピュータセキュリティシンポジウム 2009 (CSS2009), 2009.
31) 五十嵐 大, 千田 浩司, 高橋 克巳, $P\ell$-匿名性: 属性推定に対する再構築法のプライバシーの定量化, コンピュータセキュリティシンポジウム 2010 (CSS2010), 2010.
32) R. Agrawal, R. Srikant, and D. Thomas, Privacy preserving OLAP, ACM SIGMOD 2005, pp. 251–262, 2005.
33) 五十嵐 大, 千田 浩司, 高橋 克巳, Differential Privacy の数理的解析, コンピュータセキュリティシンポジウム 2010 (CSS2010), 2010.