表データの最適セル秘匿処理に対するマッチング攻撃とその実証的評価

南 和宏[1,2,3,a)]　阿部 穂日[1,3,b)]

概要：表データのセル秘匿問題は，行計・列計の線形式を内包する表データに対し，与えられた一次秘匿セルの集合の値の保護を拘束条件とし，情報損失を最小化する二次秘匿セルの集合を決定する最適化問題である．この問題に対する既存のアルゴリズムは，一次秘匿セルが取りうる可能値の区間が，与えられたしきい値より大きな幅をもつことを保証し，機密セルの安全性を担保する．しかし，決定論的なセル秘匿アルゴリズムを用いる場合，そのアルゴリズムを取得した攻撃者により，秘匿された表データの安全性が侵害されるリスクが存在する．本論文では，秘匿セルを推測値で補完した表データに同じアルゴリズムを適用し，その秘匿箇所を攻撃対象である表データの秘匿パターンと比較することで，秘匿セルの候補値を絞り込むマッチング攻撃を報告する．評価実験の結果，多数の秘匿セルに対して，可能値の区間幅がしきい値を下回るまで絞り込まれ、安全性が侵害されることが判明した．

キーワード：統計的開示抑制，セル秘匿問題，整数計画問題

Algorithmic Matching Attacks on Optimally Suppressed Tabular Data

Kazuhiro Minami[1,2,3,a)]　Yutaka Abe[1,3,b)]

Abstract: The objective of the cell suppression problem (CSP) is to protect sensitive cell values in tabular data under the presence of linear relations concerning marginal sums. Previous algorithms for solving CSPs ensure that every sensitive cell has enough uncertainty on its values based on the interval width of all possible values. However, we find that every deterministic CSP algorithm is vulnerable to an adversary who possesses the knowledge of that algorithm. We devise a matching attack scheme that narrows down the ranges of sensitive cell values by matching the suppression pattern of an original table with that of each candidate table. Our experiments show that actual ranges of sensitive cell values are significantly narrower than those assumed by the previous CSP algorithms.

Keywords: statistical disclosure control, cell suppression problem, integer linear programming

1. はじめに

2009 年施行の新統計法では，行政機関が作成する公的統計は社会全体で利用される情報基盤として位置付けられ [1]，2019 年 5 月の改正では，公的統計データの二次利用を推進する制度改正が行われた．特に調査票情報（ミクロデータ）の提供については，法第 33 条の 2 が新設され，提供対象が拡大され [2]，オンサイト利用による新たな提供方法が追加された．オンサイト施設は、情報セキュリティが確保されたデータ分析環境を提供し [3]，利用者が施設で作成した分析結果を持ち出す場合，調査票情報に含まれる個人及び組織の機密情報の漏洩を防止するための安全性審査 [4] が実施される．特に、度数表、集計表等の表データは公的統計の分析における基本データ形式であり，表デー

[1] 総合研究大学院大学
The Graduate University for Advanced Studies
[2] 統計数理研究所
The Institute of Statistical Mathematics
[3] 独立行政法人統計センター
National Statistics Center
[a)] kminami@ism.ac.jp
[b)] yabe3@nstac.go.jp

タの秘匿処理は統計開示抑制の研究分野で長年研究が行われてきた．

表データの秘匿処理においては表セル秘匿が主要な方法であり，どのセルを秘匿するかを決定するセル秘匿問題（CSP）[7] として定式化される．CSP は，表の行計・列計に関する線形関係の制約の下，機密セルの値に対して十分な不確定性を保証する秘匿セルの集合を決定する問題である．CSP は NP 困難であることが知られており，多くの研究者が効率的な近似やヒューリスティックに基づく CSP アルゴリズムを開発してきた [8], [9], [10]．特に，Benders 分解 [11] に基づくアルゴリズム [7] は，多くの現実的な表データに対して効率的に最適解を生成できることが知られており，τ-ARGUS[12]，SDCLink[13]，sdcTable[14] 等の統計的開示抑制（SDC）ツールで実装されている．

CSP に関する従来研究 [7], [15] では，秘匿セルの値の不確定性をそのセルが取りうる値の区間（以下、可能区間）の幅で定量化し，その区間幅が与えられたしきい値よりも大きければ，秘匿セルは安全と判断する．秘匿セルの可能区間は，表の行計・列計を制約条件として，セル変数の最小値と最大値を求める 2 つの線形計画問題を解くことで得られる．しかし，決定論的な CSP アルゴリズム（例: [8], [16]）が使用される場合，秘匿セルの可能区間が絞り込まれるリスクが存在する．各秘匿セルを候補値で補完し，秘匿前の元の表の候補の表を作成する．もしこの候補の表に同じ CSP アルゴリズムを適用し，攻撃対象の表とは異なる箇所のセルが秘匿された場合は，表を候補リストから除外できる．

本論文では、攻撃対象の秘匿セルを含む表に対して，全ての候補となる表を列挙し、秘匿パターンのマッチングを実施することで各秘匿セルの emph 実質的な有効可能区間を算出するマッチング攻撃の手法を紹介する．我々の攻撃プログラムはセル変数を含む行計・列計の制約条件に関する不定線形方程式を解くことにより，各候補の表を列挙し、秘匿パターンが一致した候補の表の補完値のみを用いて各セルの有効可能区間を更新する反復的処理を行う．候補の表の数は不定線形方程式の解空間を表現する零空間の次元に大きく依存し、実証評価では零空間の次元が低い場合に秘匿セル値が判明するリスクが高いことが判明した．人工的に生成した多数の度数表に対する実験では，このマッチング攻撃は計算時間的に十分実行可能であり，約 46 %から 83 %のセンシティブなセルの安全性が損なわれることがわかった．

2. セル秘匿問題

この章では，統計表の開示リスクを説明し，セル秘匿問題（CSP）を定式化する．

2.1 統計表の開示リスク

まず，調査客体に関する秘密情報を含む統計表の**開示リスク**について説明する．統計表は，年齢・収入・地域等個々の調査客体の変数のレコードで構成されるミクロデータから生成される．統計表には**度数表**（*frequency tables*）と**数量表**（*magnitude tables*）の 2 種類があり，ミクロデータのレコードを変数の特定の組合せに対応する統計表のセルに分類する．例として**図 1** にカテゴリ変数 P（職業) を表頭，カテゴリ変数 M（地域）を表側として個人データから生成した度数表と数量表の組を示す．

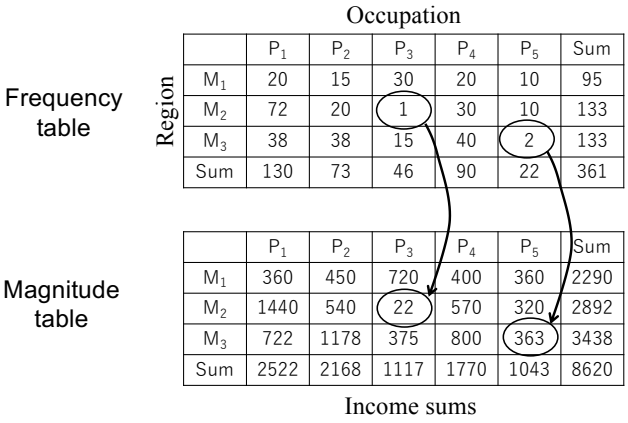


図 1 度数表と数量表における開示シナリオの例

Fig. 1 Disclosure scenarios with example frequency and magnitude tables.

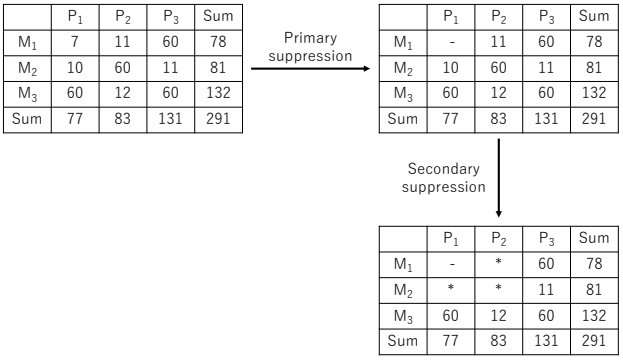
上の度数表には各セルに属する調査客体の数が表示され，下の数量表には各セルの所得の合計が表示されている．ここで，攻撃者はある者が属性 M_2 と P_3 に属することを知っており，その者のレコードが図 1 に含まれているとする．度数表のセル (M_2, P_3) の度数が 1 であることから，攻撃者はその者を一意に特定できる．そのため，攻撃者が数量表も取得していれば，特定された者の正確な所得を知ることができる．すなわち，度数 1 のセルに属する調査客体は，その調査客体の情報の一部を知る攻撃者に秘密情報が開示される重大なリスクがある．

同様に，度数が小さいセルにも開示リスクが存在する．例えば，度数表のセル (M_3, P_5) の度数は 2 であるが，このセルに攻撃者が含まれるとする．攻撃者が地域 M_3，職業 P_5 のもう 1 人の調査客体を特定できる場合，度数表のセル (M_3, P_5) から自分の所得を引くことで，特定した調査客体の所得を知ることができる．一般に，統計表内の調査客体に共謀者が存在する攻撃者をモデルとするため，度数が小さいセルは安全ではないと見なされる．

2.2 セル秘匿問題の概要

統計表の各セルにチェック基準 [5], [6], [15] を適用することで判別された安全ではないセルの値を秘匿すること

で，統計表の秘密情報を保護することを考える．例えば，必要な度数の下限値を設定する**最小度数ルール**を度数表に適用し，各セルの度数が最小度数ルールで設定される閾値より小さい場合にそのセルを秘匿する．これを**一次秘匿**（*primary suppression*）といい，一次秘匿で秘匿されたセルを**一次秘匿セル**という．ただし，一般に統計表は行計・列計を含むため，行計・列計から秘匿されていない各セルの値を引くことで，一次秘匿セルの値は簡単に復元される．そのため，いくつかの安全なセルも秘匿する**二次秘匿**を行うことで，攻撃者が一次秘匿セルの値を再計算することを防止する．二次秘匿で秘匿されたセルを**二次秘匿セル**（*secondary suppression*）という．この 2 段階の手順を**図 2** の例で説明する．



	P1	P2	P3	Sum
M1	7	11	60	78
M2	10	60	11	81
M3	60	12	60	132
Sum	77	83	131	291

	P1	P2	P3	Sum
M1	-	11	60	78
M2	10	60	11	81
M3	60	12	60	132
Sum	77	83	131	291

	P1	P2	P3	Sum
M1	-	*	60	78
M2	*	*	11	81
M3	60	12	60	132
Sum	77	83	131	291

図 2　度数表におけるセンシティブなセルを保護する 2 段階手順
Fig. 2　Two-step procedure for protecting sensitive cells in a frequency table.

ここでは，閾値 10 の最小度数ルールを用いて安全ではないセルを決定する．この例では，セル (M_1, P_1) の度数が 10 未満であることから安全ではなく，一次秘匿でこのセル値を秘匿する．しかし，M_1 の行計又は P_1 の列計から秘匿されていないセルの値を引くことで，一次秘匿セルの値が 7 であることは簡単に再計算できる．そのため，安全なセル (M_1, P_2)，(M_2, P_1) 及び (M_2, P_2) を二次秘匿し，一次秘匿セルの再計算を防止する．なお，ここでは識別のため一次秘匿セルの記号（$-$）と二次秘匿セルの記号（$*$）を変えているが，本来は同じ記号を用いる必要があることに注意されたい．

この例ではセル (M_1, P_2)，(M_2, P_1) 及び (M_2, P_2) を二次秘匿セルとして選択したが，この選択が一次秘匿セルを適切に保護するかどうかは明らかではない．また，一次秘匿セルの保護と同時に，二次秘匿による情報損失を最小限に抑える必要がある．そのため，**セル秘匿問題**を以下に定義する．

定義 1 (セル秘匿問題（CSP）)．二次秘匿された統計表 T と，T の一次秘匿セルの集合 S に対し，各一次秘匿セル $p \in S$ の値が適切に保護されるという制約の下で，情報損失を最小限とする二次秘匿セルの集合 S' を決定する問題をセル秘匿問題（CSP）という．

2.3　秘匿された表の安全性

二次秘匿セルの選択が一次秘匿セルの値を適切に保護していることを判定するためには，各一次秘匿セルが行計・列計に関する線形関係の制約下で取りうる値が十分な不確実性を持つことを確認する必要がある．r 行 c 列の 2 次元表を考えても一般性を失わない．T に含まれるセル (i, j) の値を a_{ij} で表す．式 (1)，式 (2) 及び**図 3** に表 T の線形関係を示す．

$$\text{任意の } i \text{ 行に対し，} \quad \sum_{j=1}^{c} a_{ij} = a_{i(c+1)}, \tag{1}$$

$$\text{任意の } j \text{ 列に対し，} \quad \sum_{i=1}^{r} a_{ij} = a_{(r+1)c}. \tag{2}$$

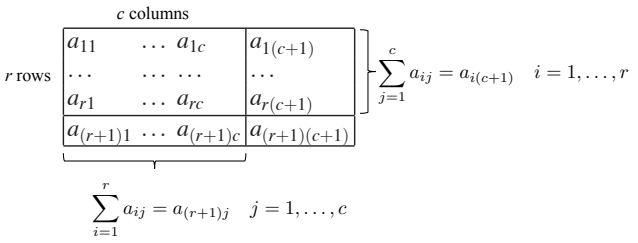


図 3　行計・列計に関するセル値の線形関係
Fig. 3　Linear relationships among cell values concerning marginal sums.

また，各セルの値は非負，すなわち以下を仮定する．

$$\text{任意の } (i, j) \text{ に対し，} a_{ij} \geq 0 \tag{3}$$

T の一部のセルを秘匿し，秘匿された表 T' を得るとする．T' の秘匿されたセル値の不確実性を定量化するため，T' の各セル (i, j) に対するセル変数 x_{ij} を導入する．また，T' と同サイズのバイナリ行列として秘匿パターン Y の概念を導入する．Y は，T' のどのセルが秘匿されているかを以下のように表す．

$$y_{ij} = \begin{cases} 1 & T' \text{ のセル } (i, j) \text{ が秘匿されている} \\ 0 & \text{その他} \end{cases} \tag{4}$$

以下の式 (5)-(8) の条件の下で x_{ij} を最小化する**整数計画問題**（*ILP*）を解くことで，各一次秘匿セル (i, j) が取りうる値の下限値 $\underline{x_{ij}}$ を得ることができる．

$$\text{任意の } i \text{ 行に対し}, \sum_{j=1}^{c} x_{ij} = x_{i(c+1)}, \tag{5}$$

$$\text{任意の } j \text{ 列に対し}, \sum_{i=1}^{r} x_{ij} = x_{(r+1)c}, \tag{6}$$

$$\text{任意の } (i,j) \text{ に対し}, y_{ij} = 0 \text{ ならば } x_{ij} = a_{ij}, \tag{7}$$

$$\text{任意の } (i,j) \text{ に対し}, x_{ij} \geq 0. \tag{8}$$

同様に，式 (5)-(8) の条件の下で x_{ij} を最大化する ILP を解くことで，各一次秘匿セル (i,j) が取りうる値の上限値 $\overline{x_{ij}}$ を得ることができる．

次に，表 T' の各秘匿セルの**可能区間**を以下に定義する．

定義 2 (可能区間（feasibility interval）)．秘匿パターン Y で秘匿された表 T' に対し，秘匿されたセル (i,j) に対する区間 $[\underline{x_{ij}}, \overline{x_{ij}}]$ を x_{ij} の可能区間という．$w_{ij} = \overline{x_{ij}} - \underline{x_{ij}}$ を可能区間の幅という．

さらに，秘匿された表 T' の**安全性**を以下に定義する．

定義 3 (秘匿された表の安全性)．秘匿パターン Y で秘匿された表 T'，一次秘匿セルの集合 P 及び与えられた閾値 δ に対し，全ての一次秘匿セル $(i,j) \in P$ の可能区間の幅 w_{ij} が以下を満たすとき，T' は安全であるという．

$$w_{ij} > \delta \tag{9}$$

2.4 CSP の定式化

定義 3 から明らかに，表内のセルを秘匿すればするほど，各一次秘匿セルの不確実性は大きくなり，秘匿された表はより安全になる．しかし，秘匿に伴い情報損失が発生するため，秘匿セルが多すぎる表はデータ分析の役に立たない．したがって，二次秘匿による情報損失を最小限に抑える必要がある．定義 3 の条件を守りつつ情報損失を最小にするため，以下のとおり CSP を定式化する．

定義 4 (CSP の定式化)．表 T，一次秘匿セルの集合 P 及び可能区間の閾値 δ に対し，CSP は以下を満たす秘匿パターン Y を決定する．

$$\sum_{i}^{r}\sum_{j}^{c} y_{ij} \text{が最小である}, \tag{10}$$

$$\text{任意の一次秘匿セル } x_{ij} \text{ について}, w_{ij} > \delta. \tag{11}$$

ここでは，簡単のため式 (10) のとおり秘匿されたセルの数で情報損失を測定する．しかしながら，次章で紹介するマッチング攻撃は，他の方法で情報損失を測定する場合にも対応する．

3. マッチング攻撃

この章では，同じ CSP アルゴリズムから生成された秘匿パターンと，攻撃対象である表の秘匿パターンを比較するマッチング攻撃について説明する．この攻撃の結果，有効可能区間の幅は定義 2 で定義された幅よりも小さくなる可能性があり，定義 3 で定義された安全性が破られる可能性がある．ここで示すマッチング攻撃は，定義 3 にある閾値 δ が，表内のセル値とは無関係に決定される度数表にのみ適用可能であることに注意されたい．

3.1 マッチング攻撃の概要

決定論的な CSP アルゴリズムによって作成された二次秘匿表に対するマッチング攻撃について説明する．この攻撃は，決定論的アルゴリズムの場合，同じ表からは常に同じ秘匿表が出力されることを利用する．**図 4** にマッチング攻撃の概念図を示す．

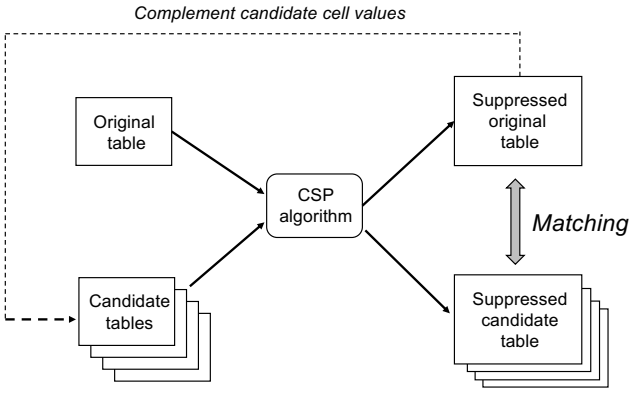


図 4 マッチング攻撃の概念図
Fig. 4 Conceptual scheme of the matching attack.

攻撃者は元の表から CSP アルゴリズムで作成された秘匿表を入手しており，秘匿セルの値を式 1，式 2 及び式 8 の制約を満たす値で補完することで，元の表の候補となる全ての表を生成する．次に攻撃者は，CSP アルゴリズムを各候補表に適用する．攻撃者は CSP アルゴリズムとセキュリティパラメータを知っていると仮定する．ある候補表の秘匿パターンが元の秘匿パターンと一致しない場合，その候補表は候補から除外される．

攻撃の主な考えは，秘匿パターンが一致した全ての候補表のセル値の範囲のみを用いて，秘匿セルの可能区間を再計算することである．全ての候補表から再計算された可能区間を**有効可能区間**という．

3.2 候補表の生成

秘匿表に含まれる秘匿セル変数を含む不定式 1 及び 2 の解を得ることで，候補表を生成する．線形方程式 1 及び 2

を以下のように行列形式で表す．

$$Ax = b \tag{12}$$

ここで，A は係数行列，x は秘匿セル変数の列ベクトル，b は行計・列計の列ベクトルである．

$N(A)$ を行列 A の零空間，すなわち以下を満たすベクトル y の集合とする．

$$N(A) = \{y \in \mathcal{Z}^n \mid Ay = \mathbf{0}\} \tag{13}$$

n はセル変数の数，$\mathcal{Z}$ は整数の集合である．$x_1, x_2 \in \mathcal{Z}^n$ が $Ax = b$ の解，すなわち $Ax_1 = b$ かつ $Ax_2 = b$ である場合，$A(x_1 - x_2) = \mathbf{0}$ となるため，$Ax = b$ の任意の 2 つの解の差は零空間 $N(A)$ に含まれる．したがって，方程式 $Ax = b$ の解は，固定解 v と $N(A)$ の要素の和として表すことができるため，$Ax = b$ の解の集合 S は以下のように表すことができる．

$$S = \{v + y \mid Av = b \wedge y \in N(A)\}. \tag{14}$$

零空間は独立したベクトルの線形結合で表されるため，式 (8) の制約を満たす全ての解を列挙することができる．

3.3 マッチング攻撃のアルゴリズム

まず，有効可能区間を定義する．

定義 5 (有効可能区間)．表 *tbl* とセキュリティパラメータ η を入力すると秘匿表 *stbl* を出力する，CSP を解く関数 f_{CSP} があるとする．秘匿表 T'，関数 f_{CSP} 及びセキュリティパラメータ η が与えられた場合，秘匿セル (i, j) の有効可能区間 $[\check{x_{ij}}, \hat{x_{ij}}]$ は以下のとおり導かれる．

$$\hat{x_{ij}} = \max\{x_{ij}; \mid x \in S \wedge \forall x_{ij} \geq 0 \wedge f_{CSP}(T'(x), \eta) = T'\},$$
$$\check{x_{ij}} = \min\{x_{ij}; \mid x \in S \wedge \forall x_{ij} \geq 0 \wedge f_{CSP}(T'(x), \eta) = T'\}.$$

ここで，S は式 (14) のとおり，$T'(x)$ は秘匿表 T' の秘匿セルを式 (12) のセルベクトル x の値で補完した表である．$\hat{w_{ij}} = \hat{x_{ij}} - \check{x_{ij}}$ を有効可能区間の幅という．

4. マッチング攻撃の評価

乱数生成された表を用いてマッチング攻撃の有効性を評価する実験を行う．可能区間の幅と有効可能区間の幅を比較することで，マッチング攻撃が一次秘匿セルの不確実性をどのように絞り込むかを測定する．

4.1 実験の準備

R 言語によりマッチング攻撃のアルゴリズムを実装した．CSP を解くアルゴリズムとして R 言語で実装したプログラム [13] を使用した．実験用 2 次元度数表の合成は以下のとおり行った．

- 平均 15, 標準偏差 10 のガウス分布から生成した乱数を各セルの値とした．
- 2 次元表のサイズは，行計・列計を除いたサイズを 4×4，5×5，6×6，7×7 とした．
- 同じサイズの表を 50 個ずつ合成した．

セキュリティパラメータは，最小度数ルールの閾値は 5，可能区間の閾値 δ は 8 とした．

CSP アルゴリズムを用いて各合成表 T_i に対し秘匿を行い，二次秘匿表 T_i' を生成した．次に同じセキュリティパラメータを用いて，T_i' に対しマッチング攻撃を実行した．

4.2 一次秘匿セルの安全性

まず，どれくらいの数の一次秘匿セルがマッチング攻撃により安全ではなくなるか，すなわち有効可能区間の幅が閾値 $\delta = 8$ より小さくなるかを検証する．

表 1 は，有効可能区間の幅が閾値 8 より小さくなり，安全ではなくなった一次秘匿セルの比率を示す．**図 5** は安全ではなくなった一次秘匿セルの数を表した棒グラフである．

マッチング攻撃の結果，一次秘匿セルの約 46 %〜83 % が定義 3 の安全性を満たさなくなったことがわかる．また，表のサイズが大きくなると安全ではない一次秘匿セルの数が増加することがわかる．

表 1 マッチング攻撃により安全ではなくなった一次秘匿セルの比
Table 1 The ratio of unsafe primary suppressed cells attacked by the matching algorithm.

#Cells in a Table	#Unsafe Prim. Supp. Cells	#Prim. Supp. Cells	Ratio of Unsafe Cells
16	48	104	0.46
25	117	170	0.69
36	190	230	0.83
49	226	271	0.83

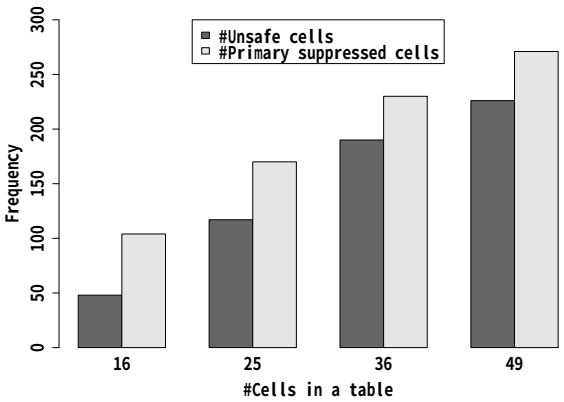


図 5 マッチング攻撃により安全ではなくなった一次秘匿セルの数
Fig. 5 The number of unsafe primary suppressed cells compromised by the matching algorithm.

次に，安全ではない一次秘匿セルの比率が零空間 $N(A)$ の次元に依存することを示す．**表 2** は，零空間の次元と表のサイズによる安全ではない一次秘匿セルの比率のクロス集計表である．零空間の次元が低いほど，表内に安全ではない一次秘匿セルが含まれるリスクが高いことがわかる．

表 2 零空間の次元と表のサイズによる安全ではない一次秘匿セルの比率のクロス集計

Table 2 The ratio of unsafe cells by crossing the dimension of the null space and table size.

#Cells in a Table	Dimension of the Null Space			
	1	2	3	4
16	0.83	0.42	0.17	
25	1.00	0.71	0.57	0.83
36	1.00	0.87	0.78	0.82
49		0.81	0.89	0.73
Total	0.88	0.73	0.75	0.77

図 6〜**図 9** は，秘匿表の零空間の次元による候補表の数の箱ひげ図である．秘匿表の零空間の次元が低い場合，秘匿セル値の可能な組合せの数が少ないため，候補表の数が少なくなることがわかる．

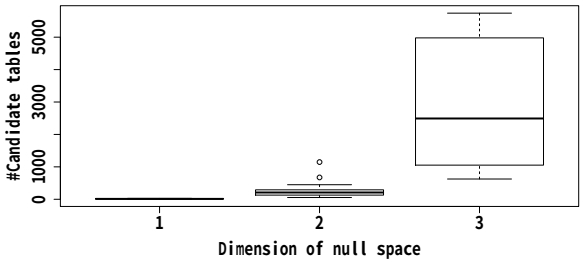


図 6 零空間の次元による候補表の数（4×4 表）

Fig. 6 The number of candidate tables in dependence to the dimension of the null space (4×4 tables).

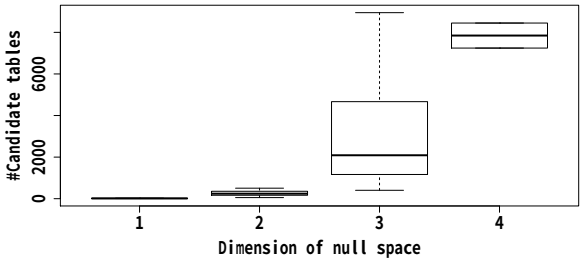


図 7 零空間の次元による候補表の数（5×5 表）

Fig. 7 The number of candidate tables in dependence to the dimension of the null space (5×5 tables).

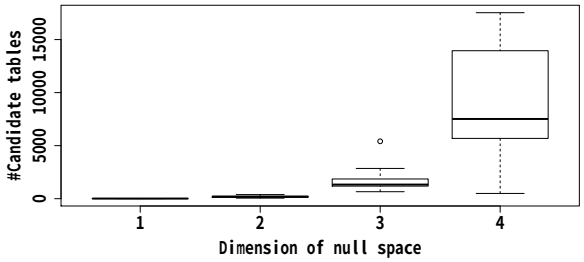


図 8 零空間の次元による候補表の数（6×6 表）

Fig. 8 The number of candidate tables in dependence to the dimension of the null space (6×6 tables).

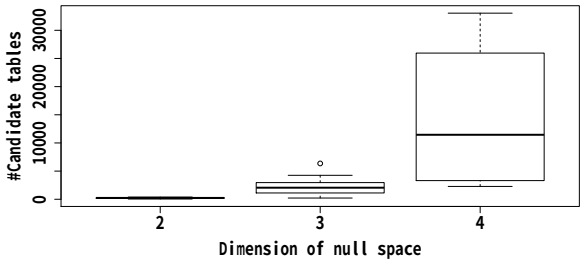


図 9 零空間の次元による候補表の数（7×7 表）

Fig. 9 The number of candidate tables in dependence to the dimension of the null space (7×7 tables).

4.3 安全ではない一次秘匿セルの有効可能区間

次に，マッチング攻撃が安全ではない一次秘匿セルの範囲をどれほど絞り込めるかを示す．**図 10**〜**図 13** は安全ではない一次秘匿セルの有効可能区間の幅の分布を示す．**表 3** に同様の結果を示す．6×6 の表と 7×7 の表で，安全ではない一次秘匿セルの正確な値まで絞り込まれている（$\acute{w}_{ij} = 0$）ことがわかる．これらの結果は，センシティブなセルの値が秘匿されていても，マッチング攻撃により正確な値を推測される重大なリスクがあることを示している．

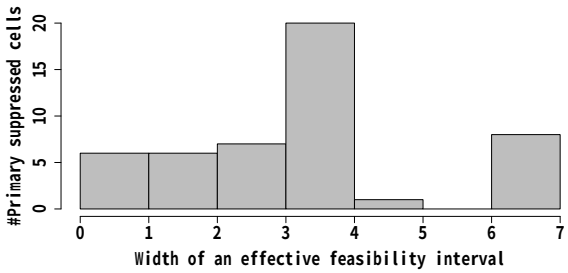


図 10 一次秘匿セルの有効可能区間の幅の分布（4×4 表）

Fig. 10 Distribution of the widths of effective feasibility intervals of primary suppressed cells (4×4 tables).

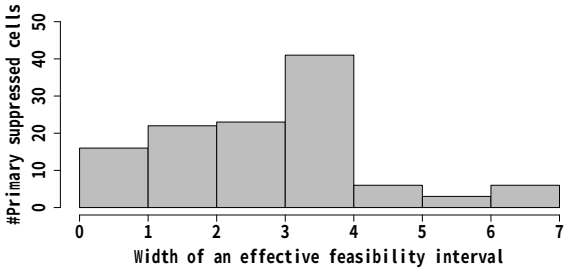


図 11　一次秘匿セルの有効可能区間の幅の分布（5 × 5 表）

Fig. 11　Distribution of the widths of effective feasibility intervals of primary suppressed cells (5 × 5 tables).

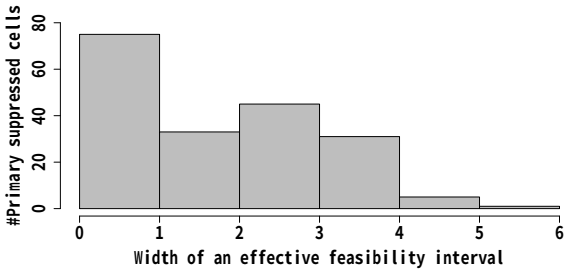


図 12　一次秘匿セルの有効可能区間の幅の分布（6 × 6 表）

Fig. 12　Distribution of the widths of effective feasibility intervals of primary suppressed cells (6 × 6 tables).

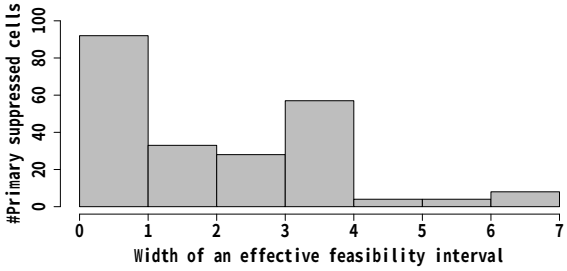


図 13　一次秘匿セルの有効可能区間の幅の分布（7 × 7 表）

Fig. 13　Distribution of the widths of effective feasibility intervals of primary suppressed cells (7 × 7 tables).

5. 結論

この論文は，統計表を秘匿するために使用される CSP アルゴリズムが決定論的であることを利用して，秘匿された表に含まれるセンシティブなセルの値を推測する新しいマッチング攻撃について説明している．このマッチング攻撃アルゴリズムの重要な発想は，秘匿パターンが攻撃対象である秘匿表の秘匿パターンと一致しない候補表を削除することである．このマッチング攻撃が秘匿されたセンシティブなセルの値を絞り込み，不確実性を安全な閾値より小さくすることができることを実験から示す．特に，攻撃対象である統計表の係数行列について，零空間の次元が低い場合，センシティブなセルの値が狭い範囲内で正確に推測される重大なリスクがあることを示す．将来の研究としては，このマッチング攻撃に耐性を持つ，非決定論的 CSP アルゴリズム開発の可能性を検討する予定である．

表 3　有効可能区間の幅による安全でない一次秘匿セルの分布

Table 3　The distribution of unsafe cells with respect to their widths of the effective feasibility intervals.

#Cells in a Table	$\acute{w}_{ij}$	#Unsafe Cells
16	0	6
	1	6
	2	7
	3	20
	4	1
	5	0
	6	4
	7	4
25	0	16
	1	22
	2	23
	3	41
	4	6
	5	3
	6	5
	7	1
36	0	75
	1	33
	2	45
	3	31
	4	5
	5	0
	6	1
	7	0
49	0	92
	1	33
	2	28
	3	57
	4	4
	5	4
	6	5
	7	3

参考文献

[1] 総務省：統計法について (online)，入手先 ⟨http://www.soumu.go.jp/toukei_toukatsu/index/seido/1-1n.htm⟩ (2019.8.21)

[2] ミクロデータ利用ポータルサイト：制度改正について (online)，入手先 ⟨https://www.e-stat.go.jp/microdata/sites/default/files/share/micro/seido-kaisei.pdf⟩ (2019.8.21)

[3] ミクロデータ利用ポータルサイト：調査票情報の利用（オンサイト利用）(online)，入手先 ⟨https://www.e-stat.go.jp/microdata/data-use/on-site⟩ (2019.8.21)

[4] 総務省:調査票情報の提供に関するガイドライン (online), 入手先 ⟨http://www.soumu.go.jp/main_content/000631449.pdf⟩ (2019.8.21)
[5] ミクロデータ利用ポータルサイト:調査票情報のオンサイト利用の手引 (online), 入手先 ⟨https://www.e-stat.go.jp/microdata/sites/default/files/share/data-use/onsite_tebiki.pdf⟩ (2019.8.21)
[6] ミクロデータ利用ポータルサイト:オンサイト利用における分析結果等の提供に関する標準的なチェック内容の解説と例 (online), 入手先 ⟨https://www.e-stat.go.jp/microdata/sites/default/files/share/data-use/onsite_check.pdf⟩ (2019.8.21)
[7] Castro, J.: Recent advances in optimization techniques for statistical tabular data protection, *Eur. J. Oper. Res.*, Vol.216, No.2, pp.257–269 (2012).
[8] Domingo-Ferrer, J. (Eds.): Inference Control in Statistical Databases, From Theory to Practice, Castro, J.: *Network Flows Heuristics for Complementary Cell Suppression: An Empirical Evaluation and Extensions*, pp.59–73, Springer (2002).
[9] Domingo-Ferrer, J. and Torra, V. (Eds.): Privacy in Statistical Databases, Giessing, S.: *Survey on Methods for Tabular Data Protection in ARGUS*, pp.1–13, Springer, (2004).
[10] Smith, J.E., Clark, A.R., Staggemeier, A.T. and Serpell, M.C.: A Genetic Approach to Statistical Disclosure Control, *IEEE Trans. Evol. Comput.*, Vol.16, No.3, pp.431–441 (2012).
[11] Benders, J.F.: Partitioning procedures for solving mixed-variables programming problems, *Numer. Math.*, Vol.4, No.1, pp.238–252 (1962).
[12] Statistics Netherlands: *tau*-ARGUS Homepage (online), available from ⟨http://neon.vb.cbs.nl/casc/tau.htm⟩ (2019.8.21)
[13] Minami, K. and Abe, Y.: Statistical Disclosure Control for Tabular Data in R, *Romanian Statistical Review*, Vol.65, No.4, pp.67–76 (2017).
[14] CRAN: sdcTable: Methods for Statistical Disclosure Control in Tabular Data, available from ⟨https://cran.r-project.org/web/packages/sdcTable/index.html⟩ (2019.8.21)
[15] Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E.S., Spicer, K. and de Wolf, P.P.: Statistical Disclosure Control, Wiley (2012).
[16] Fischétti, M. and González, J.J.S.: Models and Algorithms for Optimizing Cell Suppression in Tabular Data with Linear Constraints, *J. Am. Stat. Assoc.*, Vol.95, pp.916–928 (2000).
[17] van Tilborg, H.C.A. and Jajodia, S. (Eds.): Encyclopedia of Cryptography and Security, *Kerckhoffs' Law*, pp.675–675, Springer (2011)