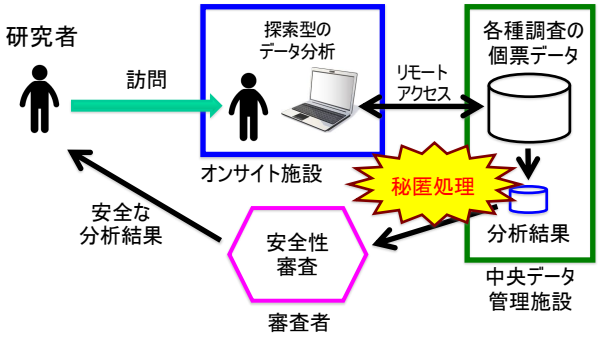
表データの最適セル秘匿処理に対するマッチング攻撃の実証的評価

南　和宏　　データ科学研究系 教授

オンサイト利用の安全性審査



表データからのセル秘匿処理

度数分布表

	P1	P2	P3	合計
M1	20	24	28	72
M2	38	38	3	79
M3	40	39	42	121
合計	98	101	73	272

1次秘匿

	P1	P2	P3	合計
M1	20	24	28	72
M2	38	38	NA	79
M3	40	39	42	121
合計	98	101	73	272

2次秘匿

	P1	P2	P3	合計
M1	NA	24	NA	72
M2	NA	38	NA	79
M3	40	39	42	121
合計	98	101	73	272

最小度数ルール
(e.g., $x_i > 10$) を侵害

2次秘匿処理は秘匿セル数の最小化問題

- 秘匿パターン　$y_i \in \{0,1\} \quad i = 1,\ldots,n$

10	NA
5	80

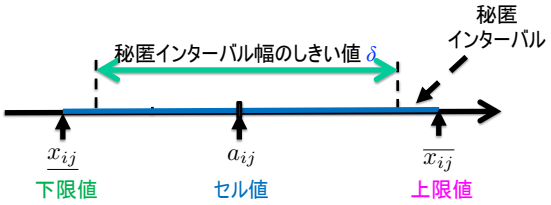
$\leftrightarrow$ (0,1,0,0)

- 目的関数: 秘匿セル数

$$\sum_{i=1}^{n} y_i$$

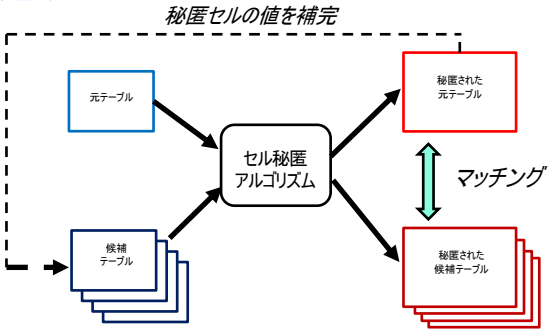
- 拘束条件：各1次秘匿セル値の機密性保護

秘匿インターバルの要件



行計、列計の線形式を満足する可能な値の幅

秘匿パターンのマッチングによる候補テーブルの絞り込み



秘匿パターンが再現できた候補テーブルの値のみが真の候補値！

秘匿セルの候補値の列挙

- 秘匿セルの候補値ベクトルxは下記拘束条件を満たす

$$Ax = b$$

- 行列Aの零空間N(A)は

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

- $Ax = b$ の解の集合Sは

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

評価実験

Q:マッチング攻撃で秘匿インターバルの条件が侵害される一次秘匿セルの割合はどの程度か？

- セル数: 16, 25, 36, 48の2次元の度数分布表をランダムに各50個生成
  - セル値は平均15, 標準偏差10の正規分布から抽出
- Benders分割アルゴリズムで2次秘匿テーブルを作成
  - 度数しきい値: 5
  - 秘匿インターバルのしきい値: 8
- Benders分割アルゴリズムを用いた秘匿パターンマッチング攻撃を実施
  - 度数しきい値: 5
  - 秘匿インターバルのしきい値: 8
- 安全でない1次秘匿処理セルの割合を集計

マッチング攻撃で安全要件（秘匿インターバルの最小幅）が破られた1次秘匿セル数

表セル数	安全でない一次秘匿セル数	一次秘匿セル数	安全でない一次秘匿セルの割合
16	48	104	46%
25	117	170	69%
36	190	230	83%
48	226	271	83%