表データのセル秘匿問題に対する非決定論的手法の有効性評価

阿部　穂日　　総合研究大学院大学　統計科学専攻　博士課程(5年一貫制)4年

【表データのセル秘匿】

表データにおいては，度数が小さいといった危険なセルが攻撃され**調査対象の秘密情報が開示されるリスク**がある．セルの値を秘匿することで表データの秘密情報を保護する場合，まず危険なセルを秘匿(**一次秘匿**)し，行計・列計から一次秘匿セルの値が再計算されないようにいくつかの安全なセルも秘匿(**二次秘匿**)する必要がある．

また，秘匿したセルの値の**可能区間**の幅(取り得る値の上限値と下限値の差)が十分な長さ(あるしきい値以上)を持つようにすることで秘匿セル値の機密性を保護する．

図1　度数表の秘匿の例

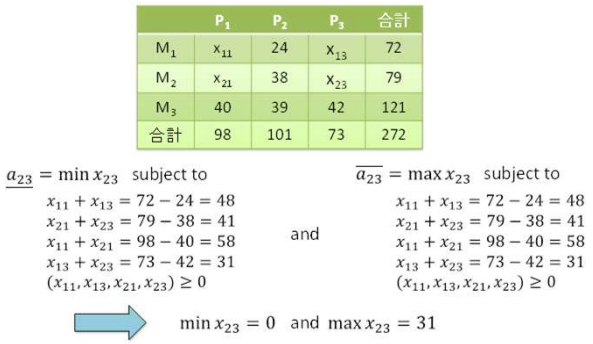
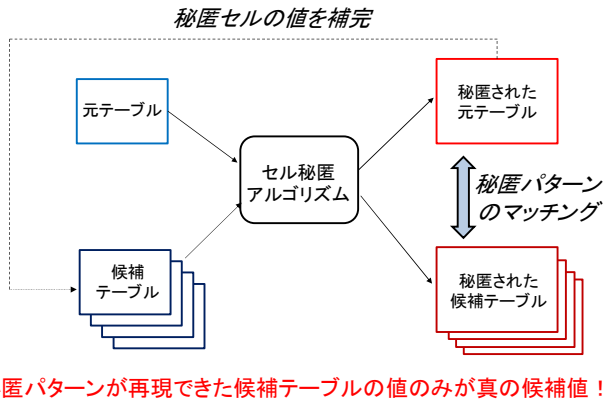


図2　秘匿セルの可能区間の計算

【セル秘匿問題を解くアルゴリズムとマッチング攻撃】

一次秘匿セルが適切に保護されるよう，また二次秘匿による情報損失が最小限となるよう二次秘匿セルを選択する問題を**セル秘匿問題(CSP)**という．

秘匿された表データと共に用いられたCSPアルゴリズムが公開されている場合，可能区間から秘匿セルが取りうる値を代入した候補表を同じアルゴリズムで秘匿し，秘匿パターンをマッチングする攻撃により度数表の秘匿セルにおいて可能区間が絞り込まれ，**機密性が保護されなくなる可能性がある**．



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！

図3　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

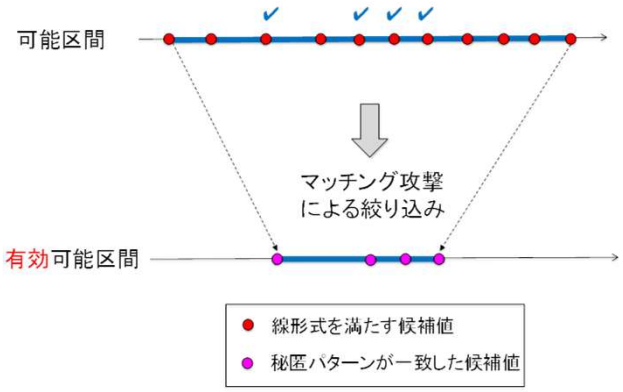


図4　可能区間の絞り込み

【非決定論的CSPアルゴリズムに対するマッチング攻撃】

上記マッチング攻撃は，CSPアルゴリズムが決定論的であり，同じ表からは常に同じ秘匿表が出力されることを利用している．

そこで，一次秘匿後に安全なセルを一つランダムに選び，そのセルも一次秘匿してから二次秘匿を行う単純な**非決定論的CSPアルゴリズム**を考察してみたが，このアルゴリズムに対しても**マッチング攻撃が可能**であることがわかった．

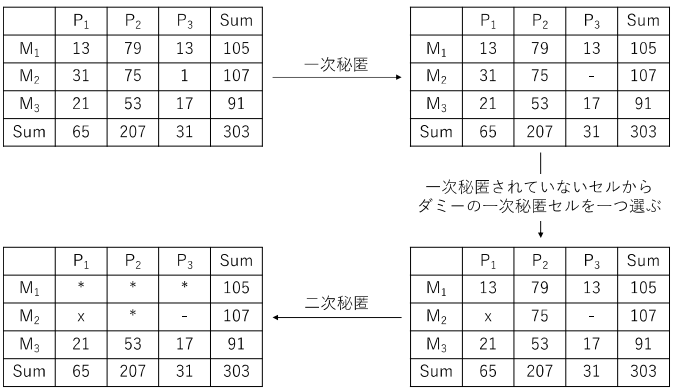


図5　非決定論的CSPアルゴリズムの例

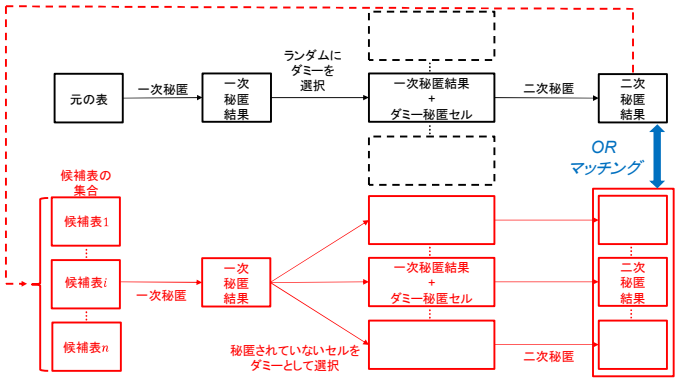


図6　非決定論的CSPアルゴリズムに対する攻撃の例

【非決定論的CSPアルゴリズムに対するマッチング攻撃の実証評価】

6×6の表データを50個乱数生成し，決定論的CSPアルゴリズムと非決定論的CSPアルゴリズムによる秘匿とマッチング攻撃の実験を行ったところ，**安全でない一次秘匿セルの割合**は決定論的CSPアルゴリズムで**82.4%**であったところ，非決定論的CSPアルゴリズムでは**70.4%**となり，非決定論的CSPアルゴリズムでも一次秘匿セルの安全性が侵害されることを確認した．

そのため，**更に安全な非決定論的CSPアルゴリズム**について研究を行っているところ．