Computer Security Symposium 2018
22 - 25 October 2018

匿名データの安全性指標としての再識別率とその活用方式の提案

小栗 秀暢[†1]

概要：個人情報保護法の改正以降，匿名加工情報に関する制度が定着し，外部に提供する匿名データの安全性や有用性を高める技術が求められている．匿名データに対する安全性検証のコンテスト PWS Cup では，過去のコンテストルールの中で，実際の研究者やアプリケーションを用いた再識別の試みによる安全性の評価方法を採用してきた．再識別を用いた安全性指標は，他の指標に比して対応できるアルゴリズムの多様性において優れている．その反面，過去のコンテストにおいては，全て異なる定義によって安全性を評価しており，その結果として出力された匿名データも異なるものとなった．本稿では，コンテストを通じて得られた知見を通じ，再識別率の持つ安全性指標として必要な要素をプライバシーフレームワークとの比較によって検討し，再識別率と存在否定の妥当性(*Plausible Deniability*)を組み合わせた安全性指標とその活用方法について提案する．

キーワード：プライバシー，匿名化，再識別，匿名加工情報，PWS Cup

1. はじめに

2017 年の改正個人情報保護法の全面施行以降，匿名加工情報によって，情報の第三者提供が簡便な条件で行える制度が整い，一般に定着しつつある．しかし，これら匿名加工情報の作成にあたっては，個人情報データベースの性質を勘案する必要があり，一概に匿名加工情報とするための加工基準を定めることは困難である．

PWS Cup 匿名加工・再識別コンテスト[1,2,3]では，データセットと加工ルールを定めた中で，匿名加工データの安全性と有用性を競う試みがなされている．過去のコンテストの中で一貫して利用されている安全性の検証指標の一つが，他の研究者・アルゴリズムから行われる「再識別」攻撃の成功率（以降 再識別率とする）である．

再識別率は k-匿名性や ε-差分プライバシーなど，処理された手法に依存せずに安全性を検証できる指標であるため，コンテストのような，多様な加工手法が存在する中から最も優れた手法を選び出すのに優れている．しかし，再識別率を用いた安全性評価指標は社会的に利用されていない．

日本における匿名加工情報は，一般人基準からの安全性を確保することが目的となる．その反面，PWS Cup で用いられている再識別率を用いた安全性評価は，プライバシー保護研究者による最も強い攻撃を行うものであり，技術的難易度が高いと考えられている．

また，過去に行われている PWS Cup では，再識別率の定義を毎回異なるものに設定しており，その定式化に関する議論も必要である．

本稿では，匿名データに対する安全性指標としての再識別率に関する定義を明確化し，その安全性をプライバシーフレームワーク要件の観点から整理する．その後，再識別率を実際の現場で簡便に用いるための手法に関して検討し，安全性指標と活用手法の提案を行う．

2. 匿名データの安全性に関する従来研究

2.1 攻撃モデルの区分

個人情報保護法における匿名加工情報は，以下のように定義されている．**「個人情報保護委員会規則で定める基準に従い，個人情報を加工して特定の個人を識別することができないようにするとともに，当該個人情報を復元することができないようにしたものをいう(改正法三十六条)」**

個人情報取扱事業者は，元情報への復元ができないように匿名加工する義務があるが，どの程度の安全性まで加工するかについては多くの手法が存在する．加工の際に提供元の企業自身によっても再識別できないレベルまで加工することを**「提供元基準における安全性」**と呼ぶ．

その際に，多く用いられるのが最大知識攻撃者モデル[4]である．これは，元データの全ての行・列・アルゴリズム等を使用して，匿名データから元データへの復元を試みることで，最も強い攻撃者を想定する．

このような強い攻撃者モデルは，2.2 節にて述べるような，多くの部分知識型の匿名化アルゴリズムと相反する性質を持つ．しかし，部分知識型の攻撃モデルは，その想定外の事象が発生した場合や，複数の手法の組み合わせに対して有効に評価できない課題がある．その点において，最大知識攻撃者モデルは，あらゆるアルゴリズムを同一の条件で評価することができるという利点がある．

2.2 PPDP における攻撃モデル

匿名データを流通させる研究として，プライバシー保護データパブリッシング(PPDP)が存在する．この研究におけるプライバシー侵害攻撃の類例が Fung らによって整理されている[5]．図 1 にその攻撃モデルを示す．

Data Publisher(DP)は，Record Owners(RO)からデータを収集し，Database に蓄積した後，Data Recipient(DR)に提供する．これを Data Publishing と定義する．このよ

†1 株式会社富士通研究所，
FUJITSU LABORATORIES LTD.

うな，Data Publishing されたパーソナルデータに対して，Untrusted な Data Recipient から受ける攻撃モデルを定義し，それぞれの攻撃モデルに対して有効な指標を整理している．

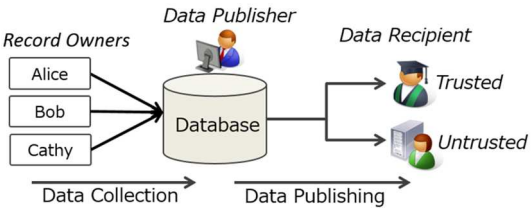

図 1　PPDP における攻撃モデル

1) レコード結合(Record linkage)は，最も多く発生する攻撃モデルである．パーソナルデータ内に存在する識別子，準識別子を用いて，ユーザの一意絞込み（シングルアウト）が可能となる．それによって個人が識別され，公開してはならないセンシティブ情報が漏洩する攻撃モデルである．解決手法として k-匿名性[6]が知られている．

2) 属性結合(Attribute linkage)は，攻撃者が攻撃対象となる個人について詳細に情報を知らない場合でも，その所属している属性によって個人のプライバシーを侵害することができる攻撃である．属性結合による攻撃手法は「同種攻撃」と「背景知識攻撃」が知られている．

同種攻撃は属性値を抽象化した場合でも，その属性に含まれる内容が単一である場合に，個人が知られたくない情報が推定されてしまう攻撃である．解決手段として l-多様性[7]が知られている．

背景知識攻撃は，準識別子とセンシティブ属性の組み合わせ，または属性値の出現数や分布の特性などから，センシティブ属性の値を知られてしまう攻撃である．解決手段として t-近傍性[8]が知られている．

テーブル結合(Table linkage)は，公開された匿名データにおける個人のデータが，何らかの別の方法にて攻撃者に知られていた場合に，個人を再識別できる可能性が高まる攻撃方法である．

4) 確率的攻撃（Probabilistic Attack） は，パーソナルデータにおけるレコードや属性値を用いるのではなく，その公開されたデータの集計値に対して行う．過去に提供したデータについて，その後，時間を置いた後に再度提供したデータとの統計的差異を検証することで，変化した個人を識別する攻撃であり，差分プライバシーとも呼ばれる．

これらの 4 つの攻撃手法における攻撃者知識を整理すると，図 2 のように分類できる．

匿名データ D'を Publish したとき，元となる個人データ D における識別子や属性を知っており，それらの結合による攻撃がレコード結合，属性に加えてセンシティブ属性を一部知っている場合の攻撃が属性結合，レコードを一部分知っている場合の攻撃がテーブル結合，そして，匿名データ D'と同等のアルゴリズムで出力された，異なる匿名データ D"を知っている場合の攻撃が確率的攻撃，と分類することができる．

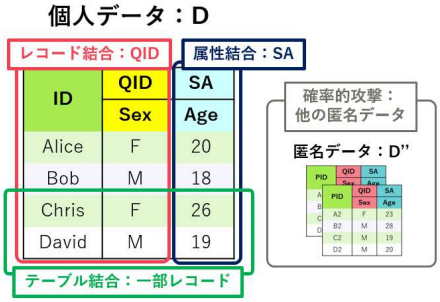

図 2　PPDP における 4 種の攻撃者知識

2.3 プライバシー保護フレームワークからの定義

次にプライバシー保護フレームワークにおいて守るべき指標を検討する．まず，一般的なプライバシーの概念に含まれるものとして，自己情報に対する「侵入攻撃からの自由」と「自己決定権」が存在する．Danezis らはそれらの要素を Hard Privacy と Soft Privacy に区分した[9]．

Hard Privacy とは，自己情報が外部に漏洩しないためにデータを最小化するための方策である．即ち，全てのプレイヤーを信頼しない前提に立ち，システムセキュリティの充実とデータに対する匿名化処理などを施すことで，漏洩する情報を最小にする要素を指す．それに対して Soft Privacy はデータを保持するコントローラーが適切にデータ主体のパーソナルデータを管理し，プライバシーに対する意図が反映されていることを保証する要素である．そのため，システムは信頼モデルで定義される．

また，ElShekeil らは [10]において，プライバシー関連法規やプライバシーフレームワークが採用している基本原理の比較調査を行っている．その比較調査を参考にして，意味の似ているもの基本原理を整理したものが表 1 である．

表 1 では LINDUNN モデル[11]の Hard/Soft の定義を基準とし，Hard Privacy の要件を，非識別性，匿名化 or 仮名化，観察/盗み見，拒絶不能性，情報公開やアクセス権と定義した．また，Soft Privacy には，説明責任，規約の遵守が挙げられているが他と比較し，データ保護，を追加した．

ElShekeil らの報告では更に多くのプライバシー基本原理の比較を行っているが，それらを含め，多くのプライバシーフレームワークは異なるものを原理として利用しており，厳密に守るべき要素は定まっていないと言える．

本稿では，匿名化処理や PPDP と概念が似ているフレームワークとして，LINDDUN モデル[11]を取り上げ，その中で求められるプライバシー保護要素について検討する．

LINDDUN モデルは，脅威分析等で用いられる Data Frame Diagram(DFD)を用いて，システム全体をデータストア，データフロー，処理，データ主体の 4 種類に区分し，

作者		Deng[11]	Kalloniatis[12]	Schaar[13]	M. Rost[14]	M.C. Oetzel[15]	GDPR
タイトル		LINDDUN	PriS	Privacy by Design	New Protection Goals	Privacy targets	
Hard (データ自体の安全性)	非識別性	L-Linkability	Unlinkability	Data minimization	Unlinkability	-	Data Minimization
	匿名(仮名)	I-Identifiability	Anonymity or Pseudonymity		Anonymity, Bindingness	-	
	拒絶不能性	N-Non-repudiation	-	-	-	-	-
	観察/盗み見	D-Detectability	Un-observability	-	Unobservability	-	-
	情報公開やアクセス権	D-Information Disclosure	Identification Authentication	Controllability, Transparency	Transparency, Concealdness, Intervenability Findability,	Access right of data subject, Data subject's right to object, Information right of data subject	Lawfulness consent, fairness, and Transparency
Soft (データ保持企業の安全性)	説明責任	U-Content Unawareness	-	-	Accountability	Accountability	Accountability
	規約の遵守	N-Non-compliance	Authorization	Data Quality,		Data Quality, Processing Legitimacy	Accuracy, Purpose limitation, Storage limitation
	データ保護 (機密性, 完全性, 可用性)	-	Data Protection	Data Confidentiality, Possibility of segregation	Confidentiality, Integrity, Availability	Security of Data	Integrity and confidentiality, Data Protection by Design and by Default

表 1 プライバシーフレームワークが求めている基本原理の比較

それぞれの要素の中で表 1 に示すプライバシー侵害要素が発生する可能性を調査し，対応の優先度を定める，プライバシーリスクアセスメント手法である．これは，欧州ネットワーク・情報セキュリティ機関(ENISA)発行の資料[16]においても「LINDDUN の方法論は、フランスのデータ保護局 CNIL の原理を広く共有しており、より体系的なアプローチを進めている」と記載されている．

この定義において，匿名化技術は公開されたデータベースにおける Hard Privacy に属する技術である．即ち Information Disclosure を除いたプライバシー要素が満たされていることが求められる．

これらの要素を匿名データの安全性モデルに適用すると, L-Linkability はデータ主体の属性や嗜好によるレコード結合攻撃，I-Identifiability は不十分な仮名化（SNS でのハンドルネーム等を含む）によるレコード結合攻撃，D-Detectability はデータ全体のプロパティや意味などからの背景知識攻撃や不十分な *l*-多様性による同種攻撃を想定している．

LINDDUN モデルは，システムの脆弱性を調査するモデルであるため，データの一部が外部に提供されているようなテーブル結合や確率的攻撃の脅威は検討されていない．逆に PPDP における攻撃モデルには N-Non Repudiation に類するモデルが存在しない．

Non Repudiation については[17]に詳述されている．「例えば匿名のオンライン投票システムや内部通報システムで発生する問題であり，その人が何者で，何を実行したかについて，攻撃者は証拠をいくつも集めることが出来る．攻撃された人には，もっともらしくそれを否認できる(*Plausible Deniability*)システムが必要となる」(著者訳)と記されている．*Plausible Deniability* をあえて日本語にするならば「存在否定の妥当性」となる．

2.4 本章のまとめ

まず，法律によって求めていると考えられる安全性は，提供元基準をベースとした最大知識攻撃者モデルである．その安全性は，最大知識であることから，PPDP で定義されている攻撃者知識について全てをカバーする．

一方 LINDDUN モデルでは，属性と仮名 ID によるレコード結合（L,I）と，背景知識等による属性結合攻撃（D）が定義されているが，N-Non Repudiation は，その範囲に含まれない脅威として定義されている．

LINDDUN モデルは，システム設計時のリスクアセスメントとしてのフレームワークであるため，システム要求事項としてのプライバシー保護について述べている．この脅威の観点は匿名加工情報や，悉皆性のある非識別加工情報や匿名加工医療情報の問題でも考慮するべきである．

一方，再識別率を用いた安全性指標は．複数の手法を組み合わせて作成された匿名データであっても，同一に評価できる利点がある．しかし，その評価方法や定義について過去に多くの議論がされている．

次節では，過去の PWS Cup で議論された課題を，上記の本章で調査したプライバシー保護要件の面から検討する．

3. 再識別攻撃から守られるプライバシー

3.1 PWS Cup における再識別の定義

2015 年から開始している，PWS Cup 匿名加工・再識別コンテストの取り組みでは，知識として元のデータベースと同じレベルの知識を持ち，かつ，プライバシー・セキュリティの研究者という,最も強い攻撃者を設定することで，匿名データの安全性基準の定量化を目指した．

匿名データの安全性指標の課題として，複数の匿名加工アルゴリズムが混在したデータに対して，一律に利用でき

る安全性指標が存在しないとされている．しかし，匿名データに対して，最大知識攻撃者を仮定した強い攻撃者による安全性検証手法は，一律に利用できる点が優れている．現実的には，最も強力な再識別アルゴリズムを提案する必要があるため，非常にコストがかかる．

過去の PWS Cup において採用された再識別の定義は毎回異なるが，それらをまとめたものを図 3 に示す．

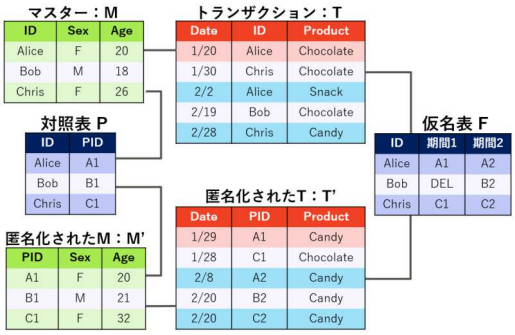

図 3　PWS Cup で利用されたデータセット概要

まず，パーソナルデータには大きく分類して，特定の個人 1 名を 1 レコードとして表現し，その重複を認めないマスター型のデータ（M）と，マスター型データと紐づけて利用し，複数回の個人の出現によって，その行動履歴を表現するトランザクション型データ（T）である．匿名化処理された M と T をそれぞれ M’，T’と表記する．

また，M と T を接続する識別子（ID）は基本的に 1 ユーザ 1 ID であり，M から M’を作成する際に，ID と仮名 ID を紐づける対照表 P が必要となる．

しかし，M と T を接続する識別子（ID）は 1 ユーザ 1 ID とは限らない．T に含まれる行動の推定を避けるため，仮名 ID を定期的，又は非定期的に変更する場合がある．その場合，ID と仮名 ID とその変更周期を示した仮名表 F が必要となる．

3.2 PWS Cup における再識別の定義

まず，M と M’を接続する対照表 P について述べる．P は過去において「連結可能匿名データ」における連結キーと呼ばれていたデータである．本来，匿名データを作成する場合，元に戻すための対応表が存在していることがリスクであるため，厳密に管理，又は削除することが求められる．

PWS Cup 2015 では，M と M’間の連結キーのみを秘匿した場合に，レコード結合がどのレベルまで成功するのかを競った．これは最大知識攻撃者モデルを用いて，レコード結合を想定して，匿名化と再識別を行ったと定義できる．

2016 年のマスターデータ＋トランザクションデータの場合であってもそれは同様で，最終的に守るべきは，一意に示されたマスターデータである．トランザクションデータ中には，マスターデータと一対一に対応する仮名 ID が含まれており，トランザクションデータを属性の一部とみなして，その関係性を推定する，というルールであった．

2017 年においては，マスターデータと仮名 ID は一対多となっており，多数の ID を持つトランザクションと切り離されたマスターデータとの関係性を推定した．

次節では，これらの過去コンテストにおいて議論された課題について検討する．

3.3 山岡匿名化問題（レコード/セルスワップ問題）

山岡匿名化とは，PWS Cup 関連の論文内で利用される用語だが，本質はレコード/セルのスワップ問題である．

安全性指標を最大知識攻撃者からの再識別攻撃と定義した場合，レコードスワップによるかく乱が有効であることが解っている．図 4 はスワップの例を示したものである．

攻撃者は，匿名データの A1 を見て Alice だと推定するが，匿名加工者は Alice と Bob のデータを入れ替えておく．最大知識攻撃者は Alice の特徴量と B1 の特徴量を比較した場合，確実に再識別に失敗する．また，この加工によって全体の統計量や値の相関ルールは変化しない．即ち有用性が最大で匿名化される．

このようなレコードスワップでなくとも，部分的なスワップは匿名化手法として有益な場合があるが，その線引きは困難である．そこで PWS Cup 2016 では，有用性のルールによって加工レベルを縛ることで，無為なスワップ処理を制限することに成功している[18]．

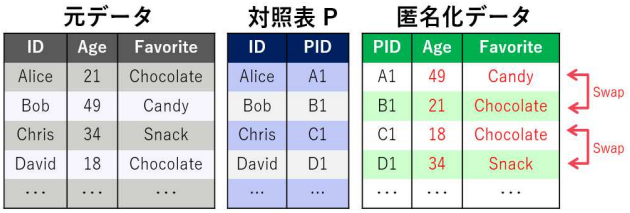

図 4　山岡匿名化（レコードスワップ）の例

本問題は *Plausible Deniability* で説明することが最も解りやすい．即ち，Alice のデータを Bob に入れ替えた所で，外部に出すデータは Alice の特徴量と全く同じものであるため，自分であることを否定できる要素が存在しない，

即ち，レコード結合を防御したが，*Plausible Deniability* の防御が行われていない，と定義できる．

3.4 T→T’間の対応表問題（識別と推定の優先度）

T→T’間の対応表問題は，識別されるリスクと機微な属性が暴露されることのどちらを優先的に対処するべきかという問題である．

例えば図 5 は匿名データを攻撃者が再識別した結果の例である．データ中には購入 2 回，合計金額 €290 の Alice と購入 1 回，合計金額 €3 の Bob が存在し，その両名の識別に成功した．その時，このレコード数，総金額，購入数を有用性の基準として利用できるか，という議論である．

匿名化データ

ID	商品	金額
A1	Chocolate	€ 100
C1	Candy	€ 3
A1	Snack	€ 190
B1	Chocolate	€ 2
…	…	…

攻撃者推定結果

ID	推定	結果	金額	購入数
Alice	A1	✔	€ 290	2
Bob	B1	✔	€ 3	1
Chris	D1	×	-	-
…	…	…	…	…

図 5　属性推定と識別の優先度検討用データ

図 6　PWS Cup データセットの総金額の分布

この匿名データの評価をする際の視点として PPDP における Data Publisher(DP), Record Owners(RO), Data Recipient(DR) の各プレイヤーの要求を考える.

RO における匿名データの問題は,「自分の属性」に関する問題である. 特に, QID は周囲の人間が知っている前提の情報だが, SA は「セル」単位で重要性が変化する. 金額や病状が含まれるデータのである場合は, 更にセル単位での優先度が変化する. それは個人のデータの属性結合攻撃に対する防御を求めている

逆に, DR はデータの利用が目的であるため, 元データと比較した有用性(値の変化量の少なさ)が重要である. そのため, セル値ごとに必要性が変化するとは考えにくい.

DP は契約の範囲内で安全性基準に基づいたデータであれば良いため, DR/RO との契約の範囲内で「安全性」を高める努力を行うことが目的である.

この 3 者の要求をまとめると, 安全性の基準は個人の要望が反映されているべきであり, セル単位で顧客が識別されることを防ぐ. 即ち, 個人から見た場合において, 漏洩リスクの高い属性に対して, 個人が識別されないレベルまで, 個人の特徴量を摂動する必要がある. これによって, 個人の *Plausible Deniability* を担保した, と主張することができる.

また, 図 5 で示した例においても, 商品, 金額, 総金額, 総購入数の 4 つのセンシティブな要素が存在するため, どの値を摂動するべきかの優先度設定が必要である. 山岡らは属性の機微度と属性の識別可能性を組み合わせたリスク評価手法[19]を提案しており, そのような手法を用いた上で匿名化アルゴリズムを検討する手法も有効である.

3.5 顧客切り捨て問題（有用性と安全性の理念）

顧客切り捨て問題とは, 有用性と安全性のトレードオフの関係で発生する課題である. 例として図 6 に PWS Cup 2017 で利用したデータセットを, ユーザ毎に顧客の総金額で集計した値の分布を示す. 縦軸は 1 人あたり総金額である. これによると, £ 61,252 の買い物をしたユーザが圧倒的な 1 位であり, データ全体金額の 6.9%を占めている. その後のユーザ分布は完全なロングテール型である.

前節で記した通り, PWS Cup では 1 人の安全性はあくまでも 1 人分でしかないため, このユーザだけを特別な手法で守るモチベーションは無い. しかし, 全体金額のシェアが高いため, この 1 人のデータを残すことで高い有用性を維持することができる. その結果, 多くのプレイヤーが, このユーザを無加工で晒し, 有用性を維持することを選択した. 本来ならば, このような外れ値はトップ/ボトムコーディングによって, 事前に削除されるべきだが, その加工についても, 上位何%まで処理すべきという基準値を示すのは困難である.

これに関しても, ある個人が自分であることを否定できる根拠があるレベルまでトップ/ボトムコーディングと匿名加工手法を組み合わせ, *Plausible Deniability* を確保する手法によって, 安全性基準を定めることができる. 例えば, 個人が再識別されないレベルの摂動量とトップコーディングを組み合わせるアルゴリズムによって実現できる.

4. 再識別の定義を利用した安全性検証

前章 1) 2) 3)の課題について検討した結果, 匿名データに対する安全性の基準は, 再識別率だけでは不十分である. 少なくとも, 個人が再識別される, という概念と *Plausible Deniability* を組み合わせ, 再識別率の課題を補完する形で成立する, 定量的な安全管理指標を提案する.

Plausible Deniability の問題を, 再識別の課題として定義すると, あるユーザ A が, 匿名データに含まれるユーザ A'に確定的に紐づく場合に, そのユーザに対して *Plausible Deniability* が確保されていないと考える. この時, A＝A' ならば識別攻撃に成功, 逆に A≠A'の場合, 濡れぎぬが発生した, と考えることができる.

類似した概念に*(c,t)-Isolation*[20]が存在する. これはクエリ監査攻撃に対応する考え方であり, 匿名データであっても対応できる. 匿名データ中の, ある 1 人のレコード Q の持つ特徴量を q としたとき, その q から最も近い距離に存在するリアルデータ X の特徴量 x_1 との距離を d とした時に, $c*d$ の距離の円の中にリアルデータが t 個以上存在することを求める指標である. 例えば図 9 において, c を 2 と定めたとき, 距離 $2d$ の円の中にリアルデータが 4 個存在する場合 *(2,4)-Isolation* であると定義される.

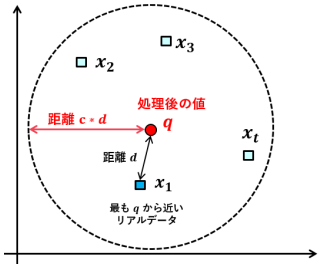

図 7　(c,t)-Isolation の説明図

また，野島らの提案[18]は，このような個人同士の距離の関係性を元情報からの Jaccaad 距離として定義し，加工後の距離が一定以上にならないことを求めることで有用性の基準としたものである．

これらを再識別率という概念で考えるならば，匿名化されたレコード Q_i の特徴量 q_i に最も近い x_t の元情報 X_t において Q_i=X_t である確率が一定以下である，という定義の方が再識別の定義として使用しやすいだろう．

4.1 匿名加工アルゴリズムの特徴

匿名化アルゴリズムの種類について表 2 に示す．まず，匿名化アルゴリズムは一般化とかく乱に区分され，それぞれメリット/デメリットが異なる．そのため，1 つのデータ内においても，属性ごとに異なるアルゴリズムを適用する場合がある．

一般化は，結果データがクラスタ化されていることを確認するため，元データと対照せずに安全性検証が可能である．かく乱は，結果データが復元できないようにノイズを掛けられているため，元データと対照しない限り，そのノイズ量や処理過程が正しいことが検証できない．

表 2 匿名化アルゴリズムの区分

アルゴリズム区分	代表的安全性指標	元情報対照	SA の攻撃者知	得意な属性
一般化	*k*-匿名性 *l*-多様性	不要	想定が必要	カテゴリ型
かく乱	*Pk*-匿名性 ε-D.P.	必要	想定不要	数値型

特に，一般化は数値属性に対して処理が困難であるため，その値域や分析方法を想定しない場合，元データの有用性が大きく失われる場合がある．そのため，数値属性に対しては一般化を無理に適用せずにかく乱を掛ける場合があるが，この両者が混在した場合の安全性の検証方法が存在しない課題がある．

4.2 再識別攻撃アルゴリズムの特徴

前節で示した，複数の匿名化アルゴリズムが混在しているデータを再識別する手法について記す．通常，再識別率を定義する際には，攻撃者の知識量と，その手段や目的を設定する必要がある．例えば，*k*-匿名性などの安全性基準は，攻撃者にはセンシティブ属性やその他の属性（本稿では，この 2 つを OA : Objective Attribute と定義する）が知られていないという前提に立って作成されている．そのため，OA に対して無加工，又は単純な加工を行う処理では，最大知識攻撃者によって情報を復元されてしまう．

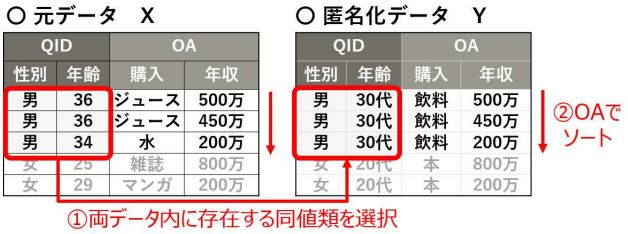

図 8　ソート型攻撃の例

OA を用いた再識別アルゴリズムの例を図 8 に示す．元データ X と匿名データ Y を比較し，同じ同値類を持つ値を確認した上で，X と Y の OA 値をソートし，その順番に行番号を求めることで再識別を行う．その他，OA と QID を交えた再識別手法の例は過去の PWS Cup 論文において多く発表されているが，多くの手法は OA について，何らかの距離関数を用いてソートし，近いレコード同士を結び付ける点において類似している．そのため本稿では，ソート型攻撃の成功率が一定以下になるような定義を示す．

4.3 再識別攻撃と *Plausible Deniability* の定式化

このような攻撃に対して安全な状態を，再識別率と *Plausible Deniability* で定式化する．

語の定義を図 9 に示す．あるパーソナルデータ X に含まれるレコード X_i の QID を q_i，OA を x_i としたとき，それを匿名化したデータ Y に含まれる X_i の加工後のデータを Y_i，その QID を p_i，OA を y_i とする．Y_i の QID は k-匿名化されており，p_i の含まれる同値類を P_I とする．このとき，P_I に含まれる y_i 以外のレコードの OA を $y_{i\text{-}z}$ で示す．

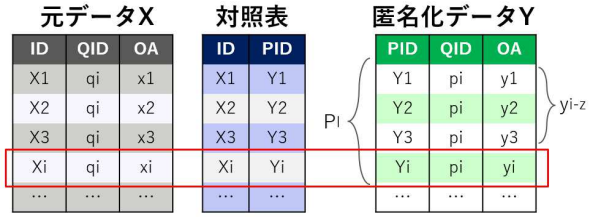

図 9　元データと匿名データの関係表

匿名化アルゴリズムは，元データ X に対して，ある分布 ε に従うノイズ付与や摂動を加えるものとする．

全ての x_i に対してノイズを加えた y_i を作成した際に，ソート攻撃に対して元データとの紐づけが不可能になっており，かつ *Plausible Deniability* においても安全であることを確認する．ここで，存在否定の妥当性を主張できる安全性しきい値を pd と設定する．このしきい値は属性の機微性と分布によって定まるものとする．

$$\forall |\ x_i\ -\ y_i\ | \geqq\ min(|\ x_i - y_{i-z}|) \geqq\ pd \quad (1)$$

この時，P_I に含まれるユーザデータは，他の特徴量を含めたソート攻撃に対して安全であり，かつ，本人の識別（又は濡れぎぬ）が発生した場合でも，その差分が pd 以上存在するため，本人でないと主張する根拠がある．

このノイズ差が pd 以下であるレコード数が，再識別攻撃に成功する数であると設定すれば，再識別率と同時に検証が可能である．即ち $|\ x_i\ -\ y_i\ | = C$，$C=(c_1, c_2, \ldots, c_i)$，$|x_i - y_{i-z}| = C'$ と定義したとき，

$$R = P(\ C\ |\ C \ni c_i > min(C')\ |\ c_i \geqq pd\ ) \quad (2)$$

によって，最大知識攻撃者によるソート攻撃と *Plausible Deniability* 侵害が発生する率を同一に評価できる．本稿では，これを *(R,pd)-privacy* と定義しておく．

この確率が一定値を超えた場合，個人が識別される可能性が高いため，ノイズ付加量を増加させて再処理する，などのアルゴリズムが考えられる．

5. 再識別率を用いた安全性検証方式の提案

本稿では，再識別に関する定義を，元データにおけるソート攻撃の成功率と定義し，その距離が他のレコードと誤認されるレベルまで変更されているかを *Plausible Deniability* 要件と定める安全性指標を提案した．

提案指標 *(R,pd)-privacy* は匿名化を行った Data Publisher が，匿名データを第三者提供する前に安全性を確認する手段として使用することが想定される．

しかし，表 2 で示したとおり，かく乱型の手法が混在した匿名データの場合，安全性を外形的に確認できない課題がある．確認する際に元情報とアルゴリズムを含む最大知識が必要であるため，出力された匿名データが求められる安全性を満たしているかを判断できない．

(R,pd)-privacy は，個人への識別攻撃，または濡れぎぬが発生した際に，否定できる根拠として存在するため，外部から観察できない安全性は個人を安心させるための材料として不十分な場合がある．

また，匿名加工情報と認定し，外部に提供された後は再識別の禁止義務が定まっていることから，再識別に属する手段による安全性検証ができない．

現実的な問題として，OA に含まれるノイズが適正な量であり，かつ，*Plausible Deniability* を確保する必要がある場合は，以下のようなケースを考える．

1) 匿名データと称して，機微性の高い SA を生データのまま外部に提供したいと考える場合
2) 適切な匿名データの中に，特定の VIP 等のデータだけを生データとし，外部に公開する場合

このようなデータは，データを利活用する側にとってもデータ利用後に差し止め請求が来るなどの潜在的なリスクがあり，危険である．

そこで，Data Recipient が匿名データを受領した後に，直接的な再識別処理を行わずに，*(R,pd)-privacy* を満たしたかについて確認する手法を提案する．

5.1 匿名化アルゴリズム共有による安全性検証方式

本節では，Data Recipient が *(R,pd)-privacy* を満たしたかを確認するために，匿名データ Y と適用された匿名加工アルゴリズムを用いて，再度ノイズ付与処理を施した結果に対して *(R,pd)-privacy* で計測することによってノイズの付与量の正当性を推定する手法を提案する．

Data Publisher は，公開された匿名データ Y の OA のある値 y_i は，元データに含まれる値 x_i に対して，ある分散 ε に従ってノイズ e_i を付与した $x_i + e_i$ である．

$$\varepsilon = V\{x_i + e_i\} \quad (3)$$

匿名化データ Y に対して，再度匿名化アルゴリズムを適用すると，再度同じ ε に従った e'_i を加えた分布が追加される．多くのノイズ付与処理は，処理前と処理後のデータに対して，分散や最大値/最小値，平均などの値が大きく変化しないようにされているのが一般的である．これにより x_i と y_i のスケールに大きな違いが無い限り，同レベルのノイズ処理が施される．

匿名化データY

PID	QID	OA
Y1	pi	y1=x1+e1
Y2	pi	y2=x2+e2
Y3	pi	y3=x3+e3
Yi	pi	yi=xi+ei
…	…	…

OAを再加工

再匿名化データY'

PID	QID	OA
Y1	pi	y1=x1+e1+e'1
Y2	pi	y2=x2+e2+e'2
Y3	pi	y3=x3+e3+e'3
Yi	pi	yi=xi+ei+e'i
…	…	…

図 10　匿名データと再匿名データの定義

そのため，再度ノイズを付与した分散 $\varepsilon' = V\{x_i + e_i + e'_i\}$ と，ε を比較することで *(R,pd)-privacy* を満たしていると推定する．この時，使用している匿名化アルゴリズムが同じであることは別の手段を用いて保証されているものとする．

処理した結果，安全性基準を満たさない値が多数存在する場合，付与するノイズ量が少なく設定されている可能性がある．また，ε'の分布がデータ作成者の主張である ε 分布との誤差が大きい場合，個別処理で分布を変更する処理が加えられている可能性がある．これらのノイズ量の適正量について，Data Recipient は，元情報を知らなくとも確認できる点が優れている．

また，個別の Data Owners についても，pd レベルを設定していることから特定の属性（職業＝政治家など）のデータを不正に持ち出すことが困難となるため，*Plausible Deniability* を守る効果について主張できる．

しかし，ε と ε' は分散であるため，1 度の施行の結果を比較しただけでは，安全性の検証ができない．

そのため，Data Recipient は同アルゴリズムを用いて大量のε' を作成し，シミュレーションによって安全性を検証することが必要となる．匿名データに付与されたノイズ量が適切である確率は，シミュレーションによる試行 N 回における適切なノイズの出現確率とみなすことで，安全性検証結果とする手法が有効である．

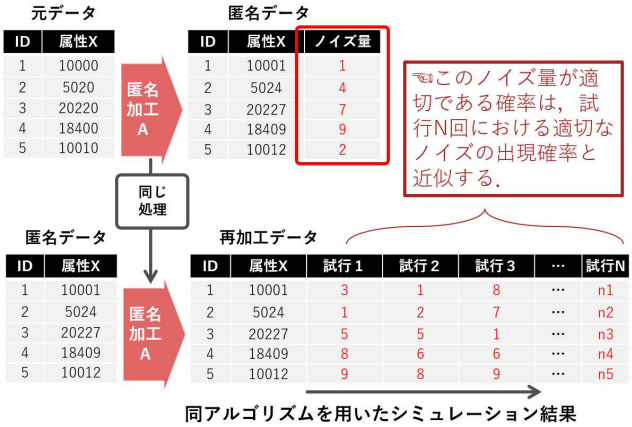

図 11　匿名データと再匿名データの関係性

6. まとめ

本稿では，PWS Cup で使用されてきた「再識別率」の定式化に向け，その安全性について検討した．まず最大知識攻撃者モデルによって PPDP における部分知識からなる攻撃手法との関係性が成立することを示した．しかし，プライバシーフレームワークには *Plausible Deniability* という基本原理があり，それは再識別率だけでは表現できない．それらの課題を過去の PWS Cup で行われた議論を振り返ることで確認し，再識別率と *Plausible Deniability* の定式化によって安全性を定義可能であることを示した．そこで，PWS Cup で一般的に用いられるソート攻撃に対する耐性と *Plausible Deniability* を満たす指標として *(R,pd)-privacy* を定義した．また，その活用ユースケースについて検討し，元となるパーソナルデータを提供せず，アルゴリズムを共有することで *(R,pd)-privacy* を推定する手法を提案した．

謝辞　過去の PWS Cup のルール策定，運営を通じて，実行委員の皆さま，及び，参加者の皆さまと，データの安全性に関する多くの議論を行ってきました．心より感謝致します．

参考文献

[1] 菊池 浩明, 山口 高康, 濱田 浩気, 山岡 裕司, 小栗 秀暢, 佐久間 淳, "匿名加工・再識別コンテストIce & Fireの設計", コンピュータセキュリティシンポジウム2015論文集, 2015(3), pp.363-370, (2015).

[2] 菊池 浩明, 小栗 秀暢, 野島 良, 濱田 浩気, 村上 隆夫, 山岡 裕司, 山口 高康, 渡辺 知恵美, "PWSCUP: 履歴データを安全に匿名加工せよ", コンピュータセキュリティシンポジウム2016論文集, 2016(2), pp.271-278, (2016).

[3] 菊池 浩明, 小栗 秀暢, 中川 裕志, 野島 良, 波多野 卓磨, 濱田 浩気, 村上 隆夫, 門田 将徳, 山岡 裕司, 山田 明, 渡辺 知恵美, "PWSCUP2017:長期間の履歴データの再識別リスクを競う", コンピュータセキュリティシンポジウム2017論文集, (2017).

[4] Domingo-Ferrer, J., Ricci, S. and Soria-Comas, J., "Disclosure risk assessment via record linkage by a maximum-knowledge attacker", Privacy, Security and Trust (PST), 2015 13th Annual Conference on, pp.28-35, (2015).

[5] Fung, B., Wang, K., Chen, R. and Yu, P.S., "Privacy-preserving data publishing: A survey of recent developments", ACM Computing Surveys (CSUR), 42(4), pp.14, (2010).

[6] L.Sweeney, "k-anonymity: a model for protecting privacy", International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, pp.557-570, (2002).

[7] Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M., "l-diversity: Privacy beyond k-anonymity", ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1), pp.3, (2007).

[8] Li, N., Li, T. and Venkatasubramanian, S., "t-closeness: Privacy beyond k-anonymity and l-diversity", Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on, pp.106--115, (2007).

[9] Danezis, G., "Introduction to Privacy Technology", Katholieke University Leuven, COSIC: Leuven, Belgium, (2007).

[10] ElShekeil, S.A. and Laoyookhong, S., "GDPR Privacy by Design", Stockholm University Master's degree project, (2017).

[11] Deng, M., Wuyts, K., Scandariato, R., Preneel, B. and Joosen, W., "A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements", Requirements Engineering, 16(1), pp.3-32, (2011).

[12] Kalloniatis, C., Kavakli, E. and Gritzalis, S., "Addressing privacy requirements in system design: the PriS method", Requirements Engineering, 13(3), pp.241--255, (2008).

[13] Schaar, P., "Privacy by design", Identity in the Information Society, 3(2), pp.267--274, (2010).

[14] Rost, M. and Bock, K., "Privacy by design and the new protection goals", DuD, January, (2011).

[15] Oetzel, M.C. and Spiekermann, S., "A systematic methodology for privacy impact assessments: a design science approach", European Journal of Information Systems, 23(2), pp.126--150, (2014).

[16] Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J., Metayer, D.L., Tirtea, R. and Schiffner, S., "Privacy and Data Protection by Design-from policy to engineering", arXiv preprint arXiv:1501.03726, (2015).

[17] Wuyts, K., Scandariato, R. and Joosen, W., "LIND (D) UN privacy threat tree catalog", (2014).

[18] Nojima, R., Oguri, H., Kikuchi, H., Nakagawa, H., Hamada, K., Murakami, T., Yamaoka, Y. and Watanabe, C., "How to Handle Excessively Anonymized Datasets", Journal of Information Processing, 26 pp.477--485, (2018).

[19] 山岡裕司, 伊藤孝一, others, "k-匿名性による特定可能性分析に基づいたデータプライバシーのリスク分析", 研究報告コンピュータセキュリティ (CSEC), 2016(31), pp.1--8, (2016).

[20] Chawla, S., Dwork, C., McSherry, F., Smith, A. and Wee, H., "Toward privacy in public databases", Theory of Cryptography Conference, pp.363--385, (2005).