匿名化の正当性を検証可能な匿名化署名方式の提案

冨樫 由美子[1,*]　山本 恭平[1]　佐藤 尚宜[1]　吉野 雅之[1]

概要：個人情報保護法の施行・改正やデジタル・ガバメントの実現に向けた各種施策により，官民問わず，匿名加工情報や仮名加工情報のデータ利活用が加速していくと考えられる．データの利活用において，データの正当性は利活用結果の正当性を保証するために重要となる．データの正当性を保証する技術として，デジタル署名があるが，単純に適用した場合，匿名化処理の正当性を検証できない．本論文では，データに対する匿名化処理の正当性検証技術として，墨塗り署名技術を応用した匿名化の正当性検証方式を報告する．

キーワード：匿名化，墨塗り署名

Proposal of Anonymization Signature Method for Verification of Anonymization

Yumiko Togashi[1,*]　Kyohei Yamamoto[1]　Hisayoshi Sato[1]　Masayuki Yoshino[1]

Abstract: It is expected that the utilization of anonymizationously processed information and pseudonym processed information will be accelerated by various measures for the enforcement and revision of the Personal Information Protection Law and the realization of digital government. In the utilization of data, the legitimacy verification of the data is important to guarantee the legitimacy of the utilization result. There is a digital signature as a technology to guarantee the correctness of data, but if it is simply applied, the correctness of the anonymization process cannot be verified. In this paper, we report a method for verifying the validity of anonymization by applying the black-painted signature technology as a technique for verifying the validity of anonymization processing for data.

Keywords: Anonymization, Sanitizable Signature

1. はじめに

個人情報保護法[1]により，特定の個人を識別することができる記述を削除・置換した匿名加工情報の利活用が進んでいる. 個人情報保護法は定期的に見直しが行われており，2020 年 6 月公布，2022 年 4 月全面施行予定の「個人情報の保護に関する法律等の一部を改正する法律（通称，令和 2 年改正個人情報法）[2]」では，内部分析に限定する等を条件に一部対応義務を緩和する「仮名加工情報」が新設されている．また，2021 年 5 月に公布された「デジタル社会の形成を図るための関係法律の整備に関する法律（通称，令和 3 年改正個人情報保護法）[3]」では，これまで行政機関，独立行政法人，民間事業者向けにそれぞれに存在していた 3 本の個人情報保護法を 1 本の法律に統合するとともに，地方公共団体の個人情報保護制度の全国的な共通ルールを規定し所管を一元化することで，個人情報の定義等が国・民間・地方で統一され，行政機関等での匿名加工情報の取扱いに関する規律が明確化されている．

政府は，デジタル・ガバメントの実現に向け，データ流通環境の整備や行政手続きのオンライン化など，官民データの活用に資する施策を推進する法律の整備や, デジタル・ガバメント実行計画の策定を推進しており，行政内あるいは行政間，官民間でのデータ連携が加速していくと考えられる．

以上の点から，今後，行政・民間問わず，匿名加工情報のほか，内部組織での仮名加工情報の活用も進んでいくと考えられる．

データの利活用において，データが不当に改変されていないこと（データの正当性）は、データの利活用結果の正当性を保証するために重要となる．不正なデータを利用した場合，データから得られる知見も不正となり，データに基づく施策やサービスが不適切となる恐れがある．

電子データに対して、データが改変されていないことを検証可能な技術として､デジタル署名技術がある. しかし，単純にデジタル署名を適用した場合，データに匿名化処理を施すと，正当性を検証できない． データホルダが許可した匿名化処理以外の改変が行われていないことを検証可能な技術として，墨塗り署名技術を応用し，削除・置換による匿名化の正当性を検証可能とする方式が提案されている[8]．本報告では，従来手法と比較して，署名者，匿名加工者，検証者間でやり取りするデータを削減可能な匿名化の

1 株式会社　日立製作所　研究開発グループ
Research & Development Group, Hitachi, Ltd.
* yumiko.togashi.ju@hitachi.com

正当性検証方式について報告する．

2. 関連研究

本研究に関連する技術として，署名後の文書に対して部分情報の秘匿を可能にする墨塗り署名技術がある[5][6]．この墨塗り署名に対して，更新できるデータをあらかじめデータベースに用意しておくという制限を加えることで，署名後の文書に対して部分情報を別のデータと入れ替えることを可能とする方式がある[7]．また，藤原らは，あらかじめ一般化を考慮した更新用データを元データに追加し，部分情報の削除，置換が可能な検証可能匿名化方式を提案している[8]．

藤原らの方式は，元データに乱数や一般化用の更新データを追加した拡張データを作成することで検証可能な匿名化方式を実現しているが，署名者から匿名加工者，匿名加工者から検証者へやりとりするデータサイズに関しては未検討である．そこで，本研究では，「あるセルの値が加工されるとき，そのセルの値と同じ属性値を持つ同属性の他のセルについても同様の加工がおこなわれる」という制限を加えることで，署名者，匿名加工者，検証者間でやり取りするデータ量を抑制する検証可能匿名化方式を提案する．提案手法では，元データとは別に乱数データ，一般化階層木を管理し，それぞれの情報を参照しながら署名生成，匿名加工，署名検証の処理を行う．

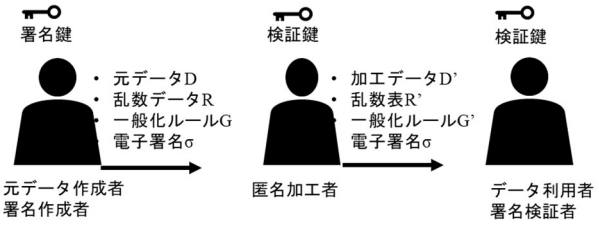


図 1　検証可能匿名化データ提供の流れ

3. 提案手法

3.1 提案手法の概要

署名後の構造化データに対して，削除・仮名化・一般化が可能な匿名化方式を提案する．提案手法は，あるセルの値が加工されるとき，そのセルの値と同じ属性値を持つ同属性の他のセルについても同様の加工がおこなわれることを前提とする．

検証可能な匿名化方式は，署名生成，匿名加工，署名検証の 3 つの処理で構成される．以下，各処理の詳細を示す．

3.2 署名生成

まず，署名生成に必要な前処理について説明する．初めに，元データの各属性に対して，匿名化方針（無加工・仮名化・削除・一般化）を決定する．次に，無加工以外の属性に対して，属性ごとに乱数データを作成する．乱数データは，当該属性に含まれる属性値を抽出し，属性値の種類数だけ乱数を生成し，属性値と乱数のペアで構成する．さらに，一般化対象属性に対しては，属性値を一般化するルール（住所の場合，市区町村→都道府県→地方→国など）を示す一般化階層木を用意する（図 2）．

次に，署名生成処理について説明する．署名生成処理では，元データに対するハッシュ値を 1 つ生成し，そのハッシュ値に対する署名を生成する．元データに対するハッシュ値を計算するために，まず，元データのセルごとのハッシュ値を計算する．セルごとのハッシュ値の計算方法は，加工方法により異なる．

無加工属性

セル値のハッシュ値が当該セルのハッシュ値とする．

仮名化属性

セル値とその値に対応する乱数を結合したデータのハッシュ値を，当該セル値のハッシュ値とする（図 3）．

削除属性

仮名化属性と同様に，セル値とその属性値に対応する乱数を結合したデータのハッシュ値を，当該セルのハッシュ値とする．ただし，削除の場合，属性内に同じ属性値が複数含まれる場合は，当該属性値に対応する乱数から新たな乱数を生成し，ハッシュ値の計算に用いる．例えば，属性値 A に対する乱数が R であり，属性中に属性値 A が 3 回出現する時，各属性値に対してそれぞれ R, R+1,R+2 の乱数を結合する（図 4）．

一般化属性

削除処理と同様，同じ属性値には異なる乱数を用いてハッシュ値を計算する．「属性値と乱数を結合したデータのハッシュ値を一階層上の属性値に対する乱数として使用する」というルールで一般化を繰り返し，最上位階層と対応する乱数の結合データのハッシュ値を当該セルのハッシュ値とする（図 5）．

匿名化方針に従い，セルごとのハッシュ値を計算した後，セルごとのハッシュ値からレコードごとのハッシュ値を計算する（図 6）．さらに，レコードごとのハッシュ値からデータ全体のハッシュ値を計算する（図 7）．最後にデータ全体に対するハッシュ値に対する署名を署名鍵を用いて生成する．

署名者は，検証鍵を公開する．また，匿名加工者に，元データ，乱数データ，一般化階層木，署名値を提供する．

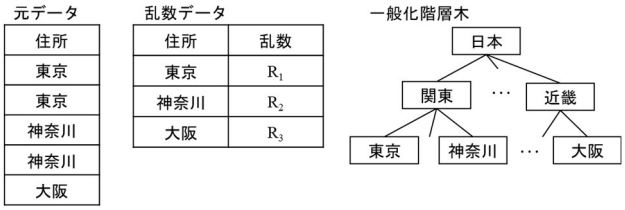


図 2　データ例

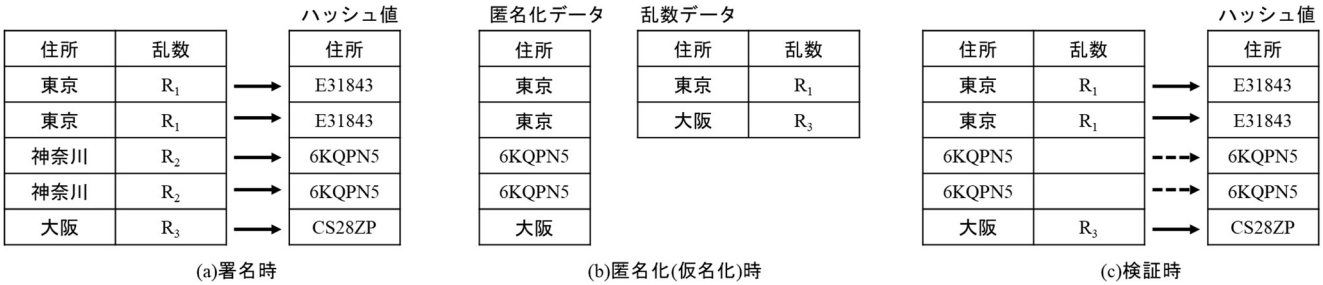


図 3　仮名化対象属性における処理

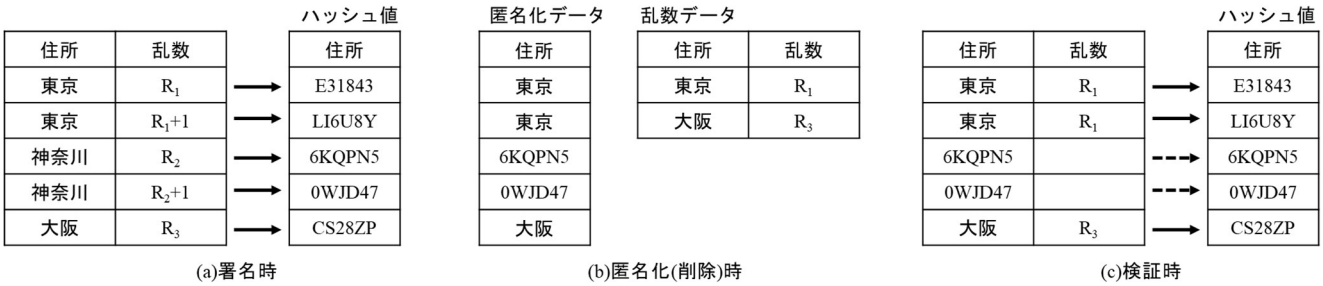


図 4　削除対象属性における処理

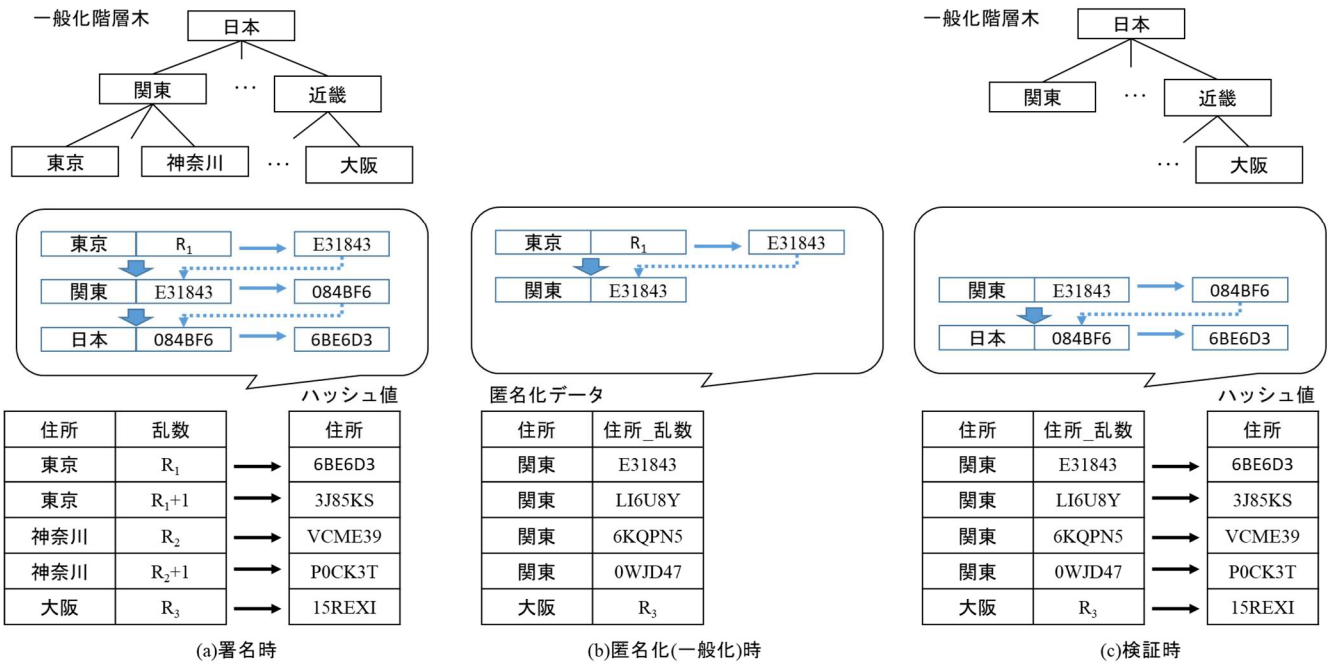


図 5　一般化対象属性における処理

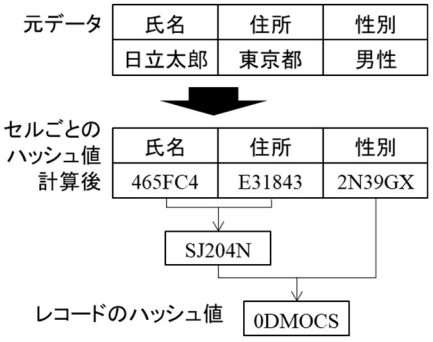


図 6　レコードに対するハッシュ値生成

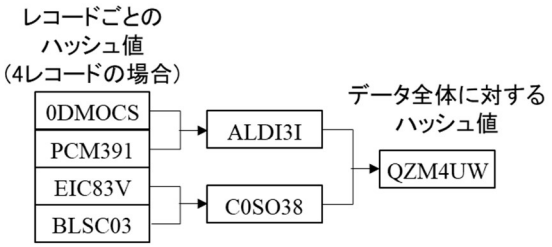


図 7　データに対するハッシュ値生成

3.3 匿名化処理

匿名加工者が行う，削除，仮名化，一般化のそれぞれの匿名化処理について説明する．匿名加工者は署名生成者が設定した匿名加工方針の範囲で，データ利用者の要望に応じた匿名加工を実施する．匿名化処理が行われない属性については，元データの値が匿名化データとして検証者に提供される．

仮名化処理

仮名化対象に対して，署名生成時と同様に，属性値とその属性値に対応する乱数を結合したデータのハッシュ値を生成し，当該属性値に置換する．仮名化の場合，同じ属性値は同じハッシュ値で置換される．図 3 では，神奈川県のみが仮名化される例を示している．検証者には，匿名化データと，仮名化対象である神奈川県とその乱数が削除された乱数データが提供される．

削除処理

削除対象に対して，署名生成時と同様に，削除対象の属性値と乱数を結合したデータのハッシュ値を，当該属性の属性値に置換する．属性内に同じ属性値が複数含まれる場合は，当該属性値に対応する乱数から新たな乱数を生成し，ハッシュ値の計算に用いる．検証者には，匿名化データと，削除対象である神奈川県とその乱数が削除された乱数データが提供される．

一般化処理

一般化処理対象に対して，署名生成時と同様に属性値と乱数を結合したハッシュ値を一般化階層木の一階層上の属性値に対する乱数として使用する，というルールで一般化を繰り返し，一般化後の属性値とその属性値に対する乱数を生成する．一般化対象属性の匿名化データに対象属性の乱数を格納する列を追加し，対象の属性値に一般化後の属性値を，乱数列に一般化後の属性値に対する乱数を格納する（図 5 は，東京と神奈川を関東に一般化した例である）．検証者には，匿名化データと，一般化後の一般化階層木データが提供される．

3.4 署名検証処理

署名検証処理は，署名検証者が，匿名化データ，乱数データ，一般化階層木データ，公開鍵を用いて署名検証処理を行う．

署名検証処理では，まず，匿名化データ，乱数データ，一般化階層木データから，匿名化データ全体に対するハッシュ値を生成する．匿名化データのハッシュ値は，署名生成時と同様，まずセルのハッシュ値を生成し，レコードのハッシュ値，データ全体に対するハッシュ値の流れで生成する．元データのハッシュ値を検証鍵で復号した値と，匿名化データのハッシュ値が一致するか否かで匿名化データの正当性を検証する．

匿名化データのセルのハッシュ値の生成方法について説明する．無加工属性に対するハッシュ値生成方法は署名生成時と同様である．削除属性並びに仮名化属性については，当該属性値に対する乱数データが存在する場合（加工がおこなわれていない場合）は，署名生成時と同様にハッシュ値計算を行い，当該属性値が乱数データに存在しない場合（つまり，加工がおこなわれている場合）は匿名化データの値をそのままセルのハッシュ値とする．一般化対象属性の場合，乱数データが存在する属性値に対しては署名生成処理と同様の処理を行う．匿名化データに乱数列が追加されている場合は，当該属性の属性値と対応する乱数と一般化階層木を用いて，最上位階層と乱数のハッシュ値計算までを行う．

4. 考察

4.1 データサイズについての考察

藤原らの手法とのデータサイズの違いについて考察する．レコード数をN個とし，ある属性に含まれる属性値の種類がM種類であったとする．なお，藤原らの手法では仮名化が考慮されていないため，拡張データ生成時に，提案手法と同様に同じ属性値に対して同じ乱数を付与するという処理を行う前提で考える．

4.1.1 乱数データの増加数

まず，１つの属性に対する乱数データの増加数をセルサ

イズの観点で比較する．

署名者，署名者→匿名加工者

従来手法では，元データに乱数列を追加し，すべての属性値に対して異なる乱数を付与するため，元データに対してN個の乱数が追加される．一方，提案手法では，元データに乱数列は追加せず，別途，属性値と乱数の対応表を生成する．提案手法では，属性値の種類数分の乱数のみを生成し，対応表に格納するため，対応表のデータサイズは$M \times 2$となる．以上から，$M \times 2 < N$のとき，提案手法はデータ削減の効果がある．

匿名加工者→署名検証者

M'種類，N'個のセル値を削除あるいは仮名化した場合，従来手法では，$N - N'$個の乱数データが存在する（従来手法では，乱数列に「Delete」を格納しているが，乱数の個数で比較しているため，カウント外とする）．提案手法では，加工後の乱数データサイズは，$(M - M') \times 2$となる．以上から，$(M - M') \times 2 < N - N'$のとき，提案手法はデータ削減の効果がある．

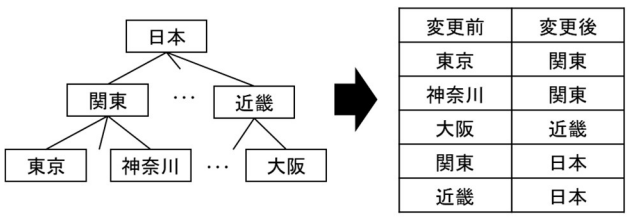
一般化の場合は，従来手法，提案手法ともに匿名加工後の乱数データサイズはN個となる．

4.1.2 一般化対象属性のデータ増加数

次に，一般化対象属性について，データ増加数をセルサイズの観点で比較する．

署名者，署名者→匿名加工者

従来手法では，一般化階層木の階層数をLとすると，元データに対して，$L - 1$個の列が追加される．つまりセルの増加数は$N \times (L - 1)$となる．提案手法では，元データに変換用データを追加せず，一般化ルールを参照しながら署名生成，匿名加工，署名検証を行う．一般化階層木を変更前と変更後の列からなる一般化階層木データで構成した場合（図 8），一般化階層木データのサイズは，一般化階層木のノード数をKとすると，$(K - 1) \times 2$となる．$N < (K - 1) \times 2 < N \times 2$となるため，$L > 3$のとき，提案手法はデータ削減の効果がある．



変更前	変更後
東京	関東
神奈川	関東
大阪	近畿
関東	日本
近畿	日本

図 8　一般化階層木データ

匿名加工者→署名検証者

一般化対象属性に対して，すべての属性値を同じ一般化階層数に一般化する場合を考える．一般化後の階層数をL’，一般化後のノード数をK’とすると，従来手法では，$L' - 1$列に一般化用データが存在し，$L - L'$列は空白のセルとなる．提案手法では，一般化階層木データのサイズが$(K' - 1) \times 2$となる．以上から，空白のセルをカウント外とすると，$(K' - 1) \times 2 - N \times (L' - 1)$個のデータ削減効果がある．

4.2 安全性についての考察

提案手法の安全性について考察する．提案手法が満たすべき安全性として、(1)署名の偽造ができないこと（不正な改変を行った匿名化データを生成できないこと）、(2)元データを秘匿できること、の 2 点があげられる。(1)については、墨塗り署名の安全性に(おおよそ)依存できる。(2)について、提案手法では，ハッシュ計算に「乱数」を付加することで元データの推測を困難にする。提案手法では，ある属性値が加工されるとき同じ属性値のすべてのセルを同様に加工することを前提とし、属性値の種類数だけ乱数を生成し，削除と一般化の対象属性で属性値が複数存在する場合には､その属性値に対する乱数から新たな乱数を生成し、ハッシュ計算時に付加する．これらにより，ハッシュ値から元の値の復元が困難となる適切なハッシュ関数を使用した場合、匿名化データから元データの復元が困難であること，削除と一般化においては，データ検証者が匿名化データから元データが同一であるかどうかを識別できないことを実現する．

5. おわりに

本論文では，データに対する匿名化処理の正当性検証技術として，墨塗り署名技術を応用した，仮名化，削除，一般化の匿名加工が可能な匿名化署名方式を提案した．提案手法では，乱数データを属性値の種類数分だけ用意し，必要に応じて属性値に割り当てられた乱数から新たなセルの乱数を生成することにより，署名者，匿名加工者，検証者間でやり取りするデータを抑制しつつ，匿名化の正当性検証が可能である．

提案手法の実用化に向けて，各処理の実行性能評価，実際のデータを想定したデータサイズの評価が今後の課題である．

参考文献

[1] “個人情報の保護に関する法律”. https://www.ppc.go.jp/files/pdf/201212_personal_law.pdf, (参照 2021-07-26).
[2] “個人情報の保護に関する法律等の一部を改正する法律 “. https://www.ppc.go.jp/files/pdf/200612_houritsu.pdf, (参照 2021-07-26).
[3] “デジタル社会の形成を図るための関係法律の整備に関する

法律”. https://www.ppc.go.jp/files/pdf/seibihou.pdf, (参照 2021-07-26).
[4] “デジタル・ガバメント実行計画”. https://cio.go.jp/sites/default/files/uploads/documents/2020_dg_all.pdf, (参照 2021-07-26).
[5] 宮崎邦彦 他. 電子文書墨塗り問題, 信学技法 ISEC2003-20, pp.61-67, 2003
[6] 増渕孝延 他. より効率的な墨塗りシステムの開発と評価. 信学技法 ISEC2004-25, pp.7-13, 2004
[7] 齊藤旭 他. 編集可能コンテンツに対する墨塗り署名を用いた電子署名システムの提案. IPSJ SIG Technical Report CSEC2007-39, pp.49-54,2007
[8] 藤原啓成, 佐藤尚宜. 匿名化の正当性を検証可能な墨塗り匿名化方式の提案, SCIS2019 暗号と情報セキュリティシンポジウム, 2019