事業者間パーソナルデータ連携のための k-匿名化クラウドサービスの一検討

米澤　龍樹[1]　佐藤　敬[2]

概要：近年，複数の事業者が保有するパーソナルデータを利活用することにより，新たなサービスの創出が期待されている．一方で，データの利活用にともなうプライバシー侵害の問題が懸念されている．そこで，データの利活用とプライバシーの保護を両立させるために，パーソナルデータをクラウド上で匿名化する k-匿名化クラウドサービスが提案されている．このサービスを複数事業者に展開する場合，事業者が特定のユーザのデータを保有しているか否かを他の事業者に知られる結合匿名テーブル問題が生じる恐れがある．本稿では，複数の事業者が保有するパーソナルデータの結合から k-匿名化までの一連の処理をクラウド上で実現する方式を提案し，提案方式の安全性について議論する．

キーワード：k-匿名化，プライバシー保護

A note on a *k*-anonymity Cloud Service for Personal Data Cooperation among Business Operators

Abstract: Nowadays it is desired that new services will be created based on further utilizing personal data among multiple business operators. Whereas, there is a growing concern over the leakage of personal information. A k-anonymity cloud service that makes personal data anonymous on the cloud was proposed in order to achieve both data utilization and privacy protection. When deploying this service among multiple business operators, there is a problem that the existence of users may be disclosed to operators comparing anonymously processed information.
In this note, we propose a method to realize a k-anonymity cloud service for multiple business operators. The proposed method allows us to combine personal data owned by each business operator and to generate anonymously processed information satisfying k-anonymity. Furthermore, we discuss the security of the proposed method.

Keywords: k-anonymity, privacy protection

1. はじめに

インターネットやスマートフォンの普及によって，事業者は膨大なパーソナルデータを収集し，自身のビジネスに広く活用している．例えば，ショッピングサイトにおいてユーザの嗜好に合った商品を紹介するレコメンドシステムでは，ユーザの購入履歴や閲覧履歴といったパーソナルデータが利用されている．近年，複数の事業者が保有するパーソナルデータを利活用することにより，新たなサービスの創出が期待されている．図 1 は，複数事業者間のパーソナルデータ連携による利活用の一例を示したものである．銀行とショッピングサイト運営会社が連携し，銀行が保有

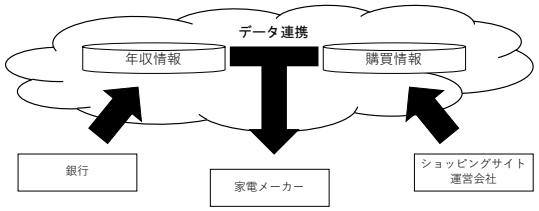


図 1　パーソナルデータ連携の例

[1] 北九州市立大学大学院国際環境工学研究科
Graduate School of Environmental Engineering, The University of Kitakyushu, Kitakyushu, Fukuoka 808-0135, Japan
[2] 北九州市立大学国際環境工学部
Faculty of Environmental Engineering, The University of Kitakyushu, Kitakyushu, Fukuoka 808-0135, Japan

しているユーザの年収情報とショッピングサイト運営会社が保有しているユーザの購買情報を家電メーカーに提供することで，家電メーカーは「年収が高い人が好む商品」や「年収が低い人が好む商品」など従来得ることができなかった年収と購買情報の相関関係を正確に知ることが可能になる．昨年施行された改正個人情報保護法において，パーソナルデータの保護と利活用を目的に匿名加工情報に関する規定が整備されたことで，より一層パーソナルデータの利活用が活発になると考えられる．

一方で，パーソナルデータの利活用にともなうプライバシー侵害が懸念されている．文献 [1] では，2002 年に米国マサチューセッツ州において，氏名を削除して公開された医療データと既に公開されてた投票者の名簿を照合することによって，マサチューセッツ州知事の医療データの特定が可能であることが示されている．我が国でも 2013 年に個人を特定できないように加工された SUICA の利用履歴データを販売しようとして社会問題となったことがある [2]．

パーソナルデータのプライバシーを保護するために，k-匿名化などのデータ加工により個人を特定できないようにする操作が広く行われている．吉野ら [6] は，パーソナルデータをクラウド上で匿名化する k-匿名化クラウドサービスを提案している．k-匿名化クラウドサービスはパーソナルデータの二次利用を目的として，事業者のパーソナルデータをプライバシー保護したままデータ利用者へ送信方式である．しかし，このサービスを複数の事業者が連携して利用する場合には，匿名化されたパーソナルデータと事業者が保有するパーソナルデータを比較することで，事業者が特定のユーザのパーソナルデータを保有しているか否かを他の事業者に知られる結合匿名テーブル問題が生じる恐れがある．

本稿では，既存の k-匿名化クラウドサービスを複数の事業者で利用できるように拡張し，複数の事業者が保有するパーソナルデータの結合から k-匿名性を満たす匿名化データの生成までの一連の処理をクラウド上で実現する事業者間パーソナルデータ連携のための k-匿名化方式を提案する．本稿の構成は以下のとおりである．第 2 章では，匿名化技術と吉野らの k-匿名化クラウドサービスについて概説する．第 3 章と第 4 章では，提案方式とその安全性について説明し，第 5 章でまとめと今後の課題について述べる．

2. 匿名化技術

2.1 k-匿名化

本稿では，パーソナルデータが表形式で表現されているデータベースを対象とする．データは複数の属性を有し，1 件のデータ（レコードという．）はそれらの属性に割り当てられた値（属性値という．）の組として表現される．

表 1 にパーソナルデータの例を示す．この例では，1 つの属性「氏名」の属性値のみで，または「年齢」と「性別」の 2 つの属性値を組み合わせることによって個人を識別できる．前者のように 1 つの属性で個人識別が可能であるとき，その属性を識別子といい，後者のように属性値の組で個人識別が可能であるとき，それらの属性を擬似識別子という．個人が特定されたくない情報の項目をセンシティブ属性という．「購入品」はセンシティブ属性にあたる．

表 1 パーソナルデータ

氏名	年齢	性別	購入品
Alice	28	女性	家電
Bob	30	男性	雑貨
Charlie	20	男性	家電
Dave	29	男性	雑貨
Ellen	25	女性	雑貨
Frank	32	男性	食料品

識別子または擬似識別子から個人を特定できないようにパーソナルデータを加工することを匿名化という．匿名化に関する代表的な指標が k-匿名性である．k-匿名性とは，擬似識別子に着目するとき，擬似識別子の値が同一となるレコードが k 個以上存在することを保証するものである．k-匿名性を満たすようにデータを加工することを k-匿名化という．

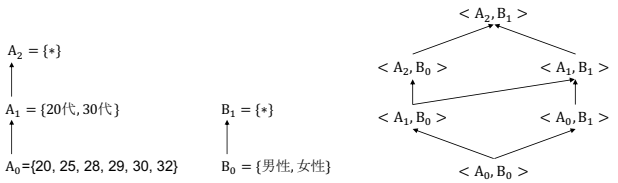
k-匿名化を行う手法として，データの属性を削除したり，データの属性値を置き換える一般化などの操作が知られている．表 2 は表 1 のデータに対して 2-匿名化処理を行った例を示している．

表 2 2-匿名化されたパーソナルデータ

年齢	性別	購入品
20 代	女性	家電
30 代	男性	雑貨
20 代	男性	家電
20 代	男性	雑貨
20 代	女性	雑貨
30 代	男性	食料品

一般化による k-匿名化アルゴリズムとして，Mondrian 法 [3] や Incognito 法 [4] が提案されている．Mondrian 法は，トップダウン型の k-匿名化アルゴリズムであり属性値 X と Y の 2 軸に対して，レコード数が k に近づくまで分割を行う方式である．Incognito 法は，擬似識別子に関する束構造を用いて匿名化を行うアルゴリズムである．まず，図 2(a) のような各属性ごとに階層を作成する．図 2(a) の年齢の例では A_0 を 1 回匿名化すると A_1 になり，A_0 を 2 回匿名化すると A_2 になることを示している．これらの階層を用いて図 2(b) に示すような束を形成する．図 2(b) では，$< A_0, B_0 >$ から順に k-匿名化を満たすか検証を行い，k-匿名化を満たしたところで検索を終了する．表 2 の例では

$< A_1, B_0 >$ の時に 2-匿名性を満たしていることが分かる．



(a) 年齢と性別の階層　(b) 束構造

図 2 Incognito

2.2 *k*-匿名化システム

パーソナルデータをクラウド上で *k*-匿名化する *k*-匿名化クラウドサービス [6] が提案されている．図 3 に模式図を示す．

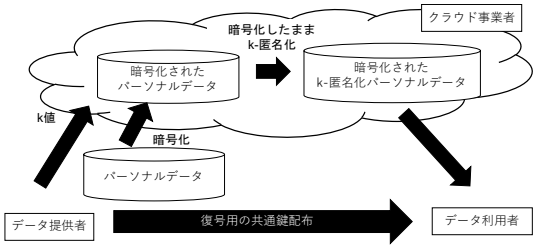


図 3 *k*-匿名化クラウドサービス

k-匿名化クラウドサービスでは，事業者が保持している *k*-匿名化のパラメータである k 値と暗号化したパーソナルデータをクラウドに送る. また，事業者はデータ利用者に *k*-匿名化したデータを受け取るための復号鍵を送る. 事業者から *k* 値と暗号化されたデータを受け取ったクラウド事業者はデータを暗号化したまま *k*-匿名化する. その後，*k*-匿名化したデータをデータ利用者に送る．最後に，データ利用者は事業者から受け取った復号鍵でデータの復号を行い，*k*-匿名化されたデータを得ることができる．

2.3 複数事業者によるデータ連携の形態

複数事業者によるパーソナルデータ連携の形態として，垂直分割と水平分割と呼ばれる形態が存在する．水平分割とは，表 3 に示すような，パーソナルデータをユーザ毎に異なる事業者が保有している分割形態である．垂直分割とは，表 4 に示すような，パーソナルデータを属性毎に異なる事業者が保有している分割形態である．この場合，複数事業者のパーソナルデータを結合し，匿名化することによってできたテーブルを結合匿名テーブルと呼ぶ．

表 3 水平分割データの 2-匿名化の例

(a) 事業者 1

氏名	年齢	性別	購入品
Alice	28	女性	家電
Bob	30	男性	雑貨
Chirle	20	男性	家電

(b) 事業者 2

氏名	年齢	性別	購入品
Dave	29	男性	雑貨
Ellen	25	女性	雑貨
Frank	32	男性	食料品

(c) 結合匿名テーブル

年齢	性別	購入品
20 代	女性	家電
30 代	男性	雑貨
20 代	男性	家電
20 代	男性	雑貨
20 代	女性	雑貨
30 代	男性	食料品

表 4 垂直分割データの 2-匿名化の例

(a) 事業者 1

氏名	年齢
Alice	28
Bob	30
Chirle	20
Dive	29
Ellen	25
Frank	32

(b) 事業者 2

氏名	性別	購入品
Alice	女性	家電
Bob	男性	雑貨
Chirle	男性	家電
Dive	男性	雑貨
Ellen	女性	雑貨
Frank	男性	食料品

(c) 結合匿名テーブル

年齢	性別	購入品
20 代	女性	家電
30 代	男性	雑貨
20 代	男性	家電
20 代	男性	雑貨
20 代	女性	雑貨
30 代	男性	食料品

水平分割データを対象とした複数事業者間のパーソナルデータ連携のための *k*-匿名化技術の 1 つとして Hua ら [5] が提案した方式がある．また，垂直データを対象にした複数事業者間のパーソナルデータ連携のための *k*-匿名化技術の 1 つとして竹之内ら [7] が提案した方式がある．

3. 事業者間データ連携のための k-匿名化クラウドサービス

3.1 モデル

ユーザは事業者 1（銀行）と事業者 2（ショッピングサイト運営会社）のサービスを利用でき，事業者間で共通の識別子（ユーザ ID）が割り当てられているものとする．パー

ソナルデータの連携は，垂直分割モデルで行うものとし，事業者 1 はユーザ ID と年齢の情報を，事業者 2 はユーザ ID と購入品の情報を保有している．そして、各事業者から送信される情報をもとに，クラウド事業者は個人が特定できないように加工された匿名加工情報を作成する．

以降，各事業者が保有するパーソナルデータが表 5 で与えられると仮定して説明する．なお，ユーザ ID を識別可能な識別子とし，属性値は U_i で表す．

表 5 各事業者が保有するパーソナルデータ

(a) 事業者 1 のデータ

ユーザ ID	年齢
U_1	20
U_2	25
U_3	35
U_6	30
U_8	40
U_9	45
U_{11}	40
U_{12}	25
U_{14}	50

(b) 事業者 2 のデータ

ユーザ ID	購入品
U_1	家電
U_4	家電
U_5	家電
U_6	雑貨
U_7	雑貨
U_9	食料
U_{10}	洋服
U_{14}	洋服
U_{15}	本

3.2 結合匿名テーブル問題

竹之内ら [7] は，複数事業者間のパーソナルデータ連携において各事業者の保有しているユーザの集合が一致しない場合に，匿名化されたパーソナルデータと事業者が保有するパーソナルデータを比較することで，事業者が特定のユーザのパーソナルデータを保有しているか否かを他の事業者に知られる問題が存在することを指摘している．この問題を結合匿名テーブル問題という．

表 5 のパーソナルデータに関する結合匿名テーブルは表 6 で与えられる．

表 6 結合匿名テーブル問題

年齢	購入品
20 or 30	家電 or 雑貨
20 or 30	家電 or 雑貨
45 or 50	食品 or 洋服
45 or 50	食品 or 洋服

この場合，事業者 1，事業者 2 ともに結合匿名テーブルと自身が保有しているデータを比較することで，ユーザ U_1, U_2, U_6, U_8 が他の事業者のサービスを利用していることを特定可能であることがわかる．

3.3 結合匿名テーブル問題の回避

本節では，結合匿名テーブルの属性に対して順序情報を利用した一般化を行うことで匿名結合テーブル問題を回避できることを示す．表 7 に順序を利用した一般化の例を示す. 表 7 では年齢の属性値が 2 つの区間 [20, 35] , [40, 50] で表現されている．このような一般化を行うことで，区間 [20, 35] ではユーザ U_2, U_3. U_{12} も該当するため，事業者 1 は，$U_1, U_2, U_3, U_6, U_{12}$ のうち誰が事業者 2 のサービスを利用しているか特定できなくなる．

表 7 順序を利用した一般化による匿名加工情報の生成

年齢	購入品
[20, 35]	家電 or 雑貨
[20, 35]	家電 or 雑貨
[40, 50]	雑貨 or 食料 or 洋服 or 本
[40, 50]	雑貨 or 食料 or 洋服 or 本

3.4 提案方式

3.4.1 構成要素

複数事業者間のパーソナルデータ連携のための k-匿名化クラウドサービスのフローを図 4 に示す．

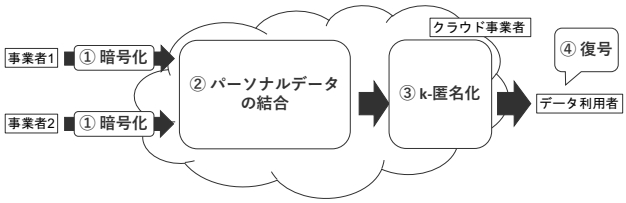


図 4 提案方式のフロー

提案方式は，「暗号化」，「パーソナルデータの結合」，「k-匿名化」，「復号」の 4 つのステップから構成される．以下にそれぞれのステップの概要を述べる．

3.4.2 暗号化

各事業者が保有しているパーソナルデータを暗号化してクラウド事業者へ送るステップである．提案方式では，一方向性ハッシュ関数，公開鍵暗号を利用して，データの匿名化処理を行う．

前提

- 事業者 k は一方向性ハッシュ関数 f_k を所持している．
- 事業者とクラウド事業者は，データ利用者を受信者とする公開鍵暗号 $E(\cdot)$ を使ってデータを暗号化してデータ利用者に送信できるものとする．

一方向性ハッシュ関数 f_k は，任意の $i < j$ において，$f_k(i) < f_k(j)$ が成り立つような順序保存性を有するものとする．こうした関数の例としては，Agrwal ら [8] が提案した共通鍵ベースの順序保存暗号 (Order preserving encryption) があげられる．これらの関数は，データの大小関係を保ったまま暗号化するために用いる一方向性関数であるが，順序の比較だけが関心事であるため復号の必要はないことに注意されたい．

表 8 に示すように，事業者 1 は保有するパーソナルデータの各レコード（ユーザ ID と年齢の属性値の組 (U_i, A_i)）

に対して匿名化処理を施すために，$(U_i, f_1(A_i), E(A_i))$ を計算して，k の値と一緒にクラウド事業者に送信する．また，事業者 2 は保有するパーソナルデータの各レコード（ユーザ ID と購入品の属性値の組 (U_i, I_i)）に対して匿名化処理を施すために，$(U_i, f_2(I_i), E(I_i))$ を計算して，k の値と一緒にクラウド事業者に送信する．

表 8 各事業者からクラウド事業者への送信データ

(a) 事業者 1 の送信データ

ユーザ ID	E_{op1}(年齢)	PKE(年齢)
U_1	$f_1(20)$	$E(20)$
U_2	$f_1(25)$	$E(25)$
U_3	$f_1(35)$	$E(35)$
U_6	$f_1(30)$	$E(30)$
U_8	$f_1(40)$	$E(40)$
U_9	$f_1(45)$	$E(45)$
U_{11}	$f_1(40)$	$E(40)$
U_{12}	$f_1(25)$	$E(25)$
U_{14}	$f_1(50)$	$E(50)$

(b) 事業者 2 の送信データ

ユーザ ID	f_2(購入品)	PKE(購入品)
U_1	f_2(家電)	E(家電)
U_4	f_2(家電)	E(家電)
U_5	f_2(家電)	E(家電)
U_6	f_2(食品)	E(食品)
U_7	f_2(食品)	E(食品)
U_9	f_2(雑貨)	E(雑貨)
U_{10}	f_2(洋服)	E(洋服)
U_{14}	f_2(洋服)	E(洋服)
U_{15}	f_2(本)	E(本)

3.4.3 パーソナルデータの結合

クラウド事業者は各事業者から送信されたデータを結合し，結合後はユーザ ID の削除を行う．得られた結合テーブルを表 9 に示す．

表 9 結合テーブル

年齢	年齢	購入品	購入品
$f_1(20)$	$E(20)$	f_2(家電)	E(家電)
$f_1(30)$	$E(30)$	f_2(雑貨)	E(雑貨)
$f_1(45)$	$E(45)$	f_2(食品)	E(食品)
$f_1(50)$	$E(50)$	f_2(洋服)	E(洋服)

3.4.4 k-匿名化

k-匿名化の手順は次の 3 つのステップからなる．

(1) 各属性ごとに階層木を作成する．
(2) k 値を決定する．
(3) 作成した階層木を用いて k-匿名化処理を行う．

階層木の作成

属性毎に階層木を作成する．ここで，階層木は k-匿名化を行う際に使用する段階的に匿名化を行うための情報を保有する補助情報である．提案 k-匿名化クラウドサービスでは木を使用して，結合匿名テーブル問題の回避を行う．提案 k-匿名化クラウドサービスが使用する木の条件は以下のとおりである．

- 各接点ノードは，属性値の多重集合とする．
- 各接点ノードには，結合テーブルに含まれる値と結合テーブルに含まれない値が最低 1 つずつ含まれる．
- 各属性値が一定回数の匿名化が可能になるようにバランス木とする．

今回使用する木は 2 分木を想定し，バランス木とするために，「全てのノードにおいて右部分木と左部分木の深さの差が 1 以下になる」という性質を持っている AVL 木 [9] を参考に木の作成を行う．図 5 のような木を作成するためのアルゴリズムを Algorithm 1 に示す．提案方式において使用する年齢に関する階層木の例を図 5 に示す．

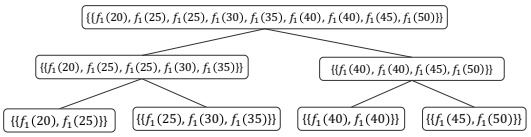


図 5 階層木の例 (年齢)

パラメータ k の決定

パラメータ k は各事業者から送信された k の値のうち最大のものとする．k 値は小さいほどユーザが特定される確率が大きくなるため，各事業者の中でプライバシーの保護レベルを高めたい事業者の要望に答えるためである．

k-匿名化

作成した階層木の情報を用いて，2.1 節で示した incognito を用いて k-匿名化を行う．k-匿名化後はデータ利用者に結合匿名テーブルを送信する．各事業者の一方向性ハッシュ関数 $f_k(\cdot)$ によって暗号化されたデータは利用者が復号できないため，利用者の公開鍵を用いて暗号化された部分のみを送信する．

3.4.5 匿名加工情報の入手

最後にデータ利用者は自身の秘密鍵を用いてクラウド事業者から受け取った結合匿名テーブルを取得する．データ利用者が取得可能な結合匿名テーブルを表 10(b) に示す．

Algorithm 1 木の作成アルゴリズム

Input T_c : 結合テーブル , T : 事業者のパーソナルデータ (1 属性ずつ)
Output tree: 作成した木

1. T_c および T を昇順ソートする.
2. T_s に T の各属性値が T_c に含まれているか否かの情報を格納する.
3. T および T_s を用いて木の作成を行う.
   3.1 T および T_s か属性値を 1 つ取り出す.
   3.2 tree の根に取り出した属性値を挿入する.
   3.3 tree のノードに子があると右の子ノードへ移動し，取り出した属性値を挿入する.
   3.4 3.3 を葉ノードに到達するまで繰り返す.
   3.5 属性値を挿入した葉のノードが分割可能か判定する.
   分割可能条件 : 分割後の各集合に置いて結合テーブルに含まれる値と結合テーブルに含まれない値がそれぞれ最低 1 つ存在する.
   分割できる点が複数存在する場合 : 分割後の集合において結合テーブルに含まれる値と結合テーブルに含まれない値の比率の差が最小になる場所を分割点とする.
   3.6 分割可能 → 分割後の各集合を分割前の集合の子ノードとする.
   分割不可能 → 3.1 に戻る.
   3.7 分割後，木のバランスが崩れていたらバランスの調整を行う.
3.1〜3.7 を全ての T の属性値で繰り返す.

表 10 のデータを事業者 1 が取得した場合でも，区間 $[20, 35]$ に該当するユーザが U_1, U_2, U_3, U_6, U_{12} の 5 人存在するため誰が事業者 2 のサービスを利用しているか特定することができない．同様に，区間 $[40, 50]$ に該当するユーザも 2 人以上存在していることがわかる．

表 10 2-匿名化による匿名加工情報の生成

年齢	購入品
[20, 35]	家電 or 雑貨
[20, 35]	家電 or 雑貨
[40, 50]	雑貨 or 食料 or 洋服 or 本
[40, 50]	雑貨 or 食料 or 洋服 or 本

4. 安全性

本章では，各事業者が取得した情報から他事業者の情報を取得することができないということの証明を行う．

4.1 モデル

提案 k-匿名化クラウドサービスでは，各事業者，データ利用者, クラウド事業者が存在する場合を想定している. また各事業者は semi-honest で振舞うこととする．semi-honest とは，プロトコルによって得られた情報を解析して他の事業者の情報を得ようとするが，プロトコルを逸脱した攻撃は行わないというモデルである．

4.2 クラウド事業者が得られる情報

まず，クラウド事業者単独でユーザのパーソナルデータが取得可能か検討する．

クラウド事業者が取得可能な情報は，表 8 に示しているとおり，ユーザ ID，f_1, $f_2()$ それぞれで暗号化された順序データ，利用者の公開鍵で暗号化された $E(A_i)$ と $E(I_i)$ である．データの 3 種類がある．これらからユーザのパーソナルデータが取得可能か検討する．

ユーザ ID の漏洩

本提案 k-匿名化クラウドサービスではユーザを直接識別可能な識別子のハッシュ値をユーザ ID とした．ハッシュ関数は一方向関数であり，ハッシュ値から元のデータへ復元することはできないという性質を持っているためハッシュ値からユーザの情報を取得することはできない．また，クラウド事業者は事業者がどのようなハッシュ関数を

用いて，どのような識別子のハッシュ値をユーザ ID としたかの情報を知ることができない．そのため元のデータを推測することも困難である．

順序情報が保存された暗号文の復号

f_1, $f_2()$ それぞれで暗号化された順序データはクラウド事業者は鍵を保有していないため、復号不可能である、しかし、これらのデータからは属性値の順序情報を取得することが可能である．この属性値の順序情報から，U_1 は U_2 より年齢が上であるという情報を知られる危険がある．しかし，クラウドが取得する情報からは属性値がどのような属性に対応するかということを得ることができないため，順序情報からユーザの情報を取得するということも不可能である．

利用者が復号可能な暗号文の復号

利用者の公開鍵で暗号化されたデータはクラウド事業者は鍵を保有していないため復号不可能であり，ユーザの情報を取得することはできない．以上のことから，クラウド事業者が取得した情報からユーザの情報を取得することは不可能である．

4.3 事業者とデータ利用者が結託した場合

次に，クラウド事業者とデータ利用者が結託することでどのようなユーザのパーソナルデータが取得可能か検討する．

クラウド事業者が取得可能な情報を表 11 に示す．

表 11 事業者とデータ利用者が結託して取得可能なデータ例

(a) 事業者のデータ例

ユーザ ID	年齢
U_1	20
U_2	25
U_3	35
U_6	30
U_8	40
U_9	45
U_{11}	40
U_{12}	25
U_{14}	50

(b) データ利用者が利用可能なデータ例

年齢	購入品
[20, 35]	家電 or 雑貨
[20, 35]	家電 or 雑貨
[40, 50]	雑貨 or 食料 or 洋服 or 本
[40, 50]	雑貨 or 食料 or 洋服 or 本

事業者とデータ利用者が結託した場合に取得可能な情報は以下の 2 つである．

- 事業者の保有しているパーソナルデータ
- データ利用者が取得した結合匿名テーブル

この 2 つの情報からユーザのパーソナルデータが取得可能か検討を行う．まず，結合匿名テーブルでは k-匿名性を満たしているためユーザのデータを一意に特定することは不可能である．事業者の保有しているパーソナルデータと結合匿名テーブルの 2 つを比較し，他事業者のユーザのパーソナルデータを取得しようとした場合を考慮する．その場合であっても，本提案 k-匿名化クラウドサービスでは，結合匿名テーブル問題の回避を行なっているため，ユーザの情報を取得することができない．

5. まとめと今後の課題

本稿では，複数の事業者が保有しているパーソナルデータの結合から k-匿名性を満たす匿名加工情報の生成までの一連の処理をクラウド上で実現する方式を提案した．提案方式は，複数事業者のデータ連携の際の課題であった匿名結合テーブル問題を，データの属性値を順序情報に基づいて区間として表現し，結合したデータに含まれない属性値を包含することによって回避した．そして，提案 k-匿名化クラウドサービスの安全性について評価を行なった今後は，提案 k-匿名化クラウドサービスにおける匿名化データの有用性について評価する必要がある．また，今回の提案 k-匿名化クラウドサービスではプライバシー保護指標として k-匿名性のみを対象とした．今後は k-匿名性のみではなく，k-匿名性を拡張した指標である l-多様性などの指標も考慮することでよりプライバシー保護することが可能になると考える．

参考文献

[1] Latanya Sweeny, “k-anonymity: A model for protecting privacy”. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 10, No. 5: pp. 557–570, 2002.

[2] “Suica 利用履歴販売，JR 東は「個人情報に当たらない」との見解”, ITmedia ビジネス online, http://bizmakoto.jp/makoto/articles/1307/19/news141.html (参照：2019 年 1 月)

[3] Kristen LeFevre et al.,“Mondrian multidimensional k-anonymity.” In: Data Engineering, 2006. ICDE’06. Proceedings of the 22nd International Conference on. IEEE, p. 25, 2006.

[4] LeFevre, K, et al. “Incognito: Efficient full-domain k-anonymity.” In: Proceedings of the 2005 ACM SIGMOD international conference on Management of data. ACM, pp. 49–60, 2005.

[5] Jingyu Hua et al., “Practical m-k-Anonymization for Collaborative Data Publishing without Trusted Third Party.” Security and Communication Networks, Vol. 2017, Article ID 9532163, 10 pages, 2017.

[6] 吉野雅之ら. “暗号化データにおける k-匿名化技術: 安全な匿名化クラウドサービスの実現”. 電子情報通信学会技術研究報告 Vol. 115, No. 502, pp. 211–217, 2016.

[7] 竹之内隆夫ら. クラウド上での事業者間データ連携のための分散型パーソナル情報保護エージェント.” 情報処理学会論文誌，Vol. 53, No. 11, pp. 2432–2444, 2012.

[8] Rakesh Agrawal et al., “Order preserving encryption for numeric data.” In: Proceedings of the 2004 ACM SIGMOD international conference on Management of data, pp. 563–574, 2004.

[9] Donald E knuth, The art of computer programming Volume3 Sorting and Serching Second Edition ,Addison-Wesley, pp 458–466, 1998