(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6065833号
(P6065833)
(45)発行日　平成29年1月25日(2017.1.25)　(24)登録日　平成29年1月6日(2017.1.6)

(51)Int.Cl.　F I
G06F 21/62 (2013.01)　G06F 21/62 354

請求項の数 15 (全 29 頁)

(21)出願番号　特願2013-518146(P2013-518146)
(86)(22)出願日　平成24年5月24日(2012.5.24)
(86)国際出願番号　PCT/JP2012/064016
(87)国際公開番号　WO2012/165518
(87)国際公開日　平成24年12月6日(2012.12.6)
審査請求日　平成27年4月17日(2015.4.17)
(31)優先権主張番号　特願2011-124398(P2011-124398)
(32)優先日　平成23年6月2日(2011.6.2)
(33)優先権主張国　日本国(JP)

(73)特許権者　000004237
日本電気株式会社
東京都港区芝五丁目７番１号
(74)代理人　100109313
弁理士　机　昌彦
(74)代理人　100124154
弁理士　下坂　直樹
(72)発明者　竹之内　隆夫
東京都港区芝五丁目７番１号
日本電気株式会社内

審査官　児玉　崇晶

最終頁に続く

(54)【発明の名称】分散匿名化システム、分散匿名化装置及び分散匿名化方法

(57)【特許請求の範囲】
【請求項１】
ユーザ識別子とユーザに関する情報とを関連付けて記憶する記憶手段と、
入力される複数の識別子のうち、前記記憶されているユーザ識別子と、当該ユーザ識別子とは異なる識別子であるダミー識別子とを複数のグループに分割し、かつ、前記ダミー識別子が前記複数のグループに分散されるように分割する分割手段と、
前記グループ毎に含まれる前記識別子に関する情報を他装置に送信する送信手段と、
前記送信手段から送信された情報に基づいて前記他装置で分割されたグループ毎に含まれる識別子に関する情報を、前記他装置から受信する受信手段と、
自装置と、前記他装置とのいずれにも存在する識別子の割合が予め定めた匿名指標を満たすか否かを前記グループ毎に判定する判定手段と、
を含む分散匿名化装置。
【請求項２】
前記入力される複数の識別子のうち、前記記憶されているユーザ識別子とは異なる識別子を前記ダミー識別子として設定する設定手段を含み、
前記グループ毎に含まれる前記識別子に関する情報は、前記グループ毎における前記識別子の内容を示す情報である
請求項１記載の分散匿名化装置。
【請求項３】
前記判定手段は、さらに前記複数のグループが含む全ての識別子に対する前記ユーザ識

別子の数の割合である存在指標を満たすか否かを前記グループ毎に判定する、
請求項１又は２に記載の分散匿名化装置。
【請求項４】
前記判定手段が匿名指標又は存在指標を満たさないと判定した場合に、最後に行った分割をキャンセルして、結合匿名化テーブルを生成する生成手段と、
をさらに含む請求項１乃至３のいずれか１項に記載の分散匿名化装置。
【請求項５】
前記ダミー識別子に、前記ユーザに関する情報として値を関連付ける操作手段と、
をさらに含む請求項１乃至４のいずれか１項に記載の分散匿名化装置。
【請求項６】
前記操作手段は、前記ダミー識別子に、前記ユーザに関する情報として幅を持った値を関連付ける、
請求項５に記載の分散匿名化装置。
【請求項７】
前記操作手段は、前記ユーザ識別子のユーザに関する情報である値の分散に基づいて、前記ダミー識別子の値を関連付ける、
請求項５又は６に記載の分散匿名化装置。
【請求項８】
前記生成手段は、一以上のダミー識別子のデータを残した結合匿名化テーブルを生成する、
請求項４に記載の分散匿名化装置。
【請求項９】
前記生成手段は、一以上のユーザ識別子のデータを削除した結合匿名化テーブルを生成する、
請求項４に記載の分散匿名化装置。
【請求項１０】
コンピュータが、
ユーザ識別子とユーザに関する情報とを関連付けて記憶し、
入力される複数の識別子のうち、前記記憶されているユーザ識別子と、当該ユーザ識別子とは異なる識別子であるダミー識別子とを複数のグループに分割し、かつ、前記ダミー識別子が前記複数のグループに分散されるように分割し、
前記グループ毎に含まれる前記識別子に関する情報を他装置に送信し、
前記送信された情報に基づいて前記他装置で分割されたグループ毎に含まれる識別子に関する情報を、前記他装置から受信し、
自装置と、前記他装置とのいずれにも存在する識別子の割合が予め定めた匿名指標を満たすか否かを前記グループ毎に判定する、
分散匿名化方法。
【請求項１１】
前記コンピュータが、
前記入力される複数の識別子のうち、前記記憶されているユーザ識別子とは異なる識別子を前記ダミー識別子として設定し、
前記グループ毎に含まれる前記識別子に関する情報は、前記グループ毎における前記識別子の内容を示す情報である
請求項１０記載の分散匿名化方法。
【請求項１２】
前記判定の際、さらに前記複数のグループが含む全ての識別子に対する前記ユーザ識別子の数の割合である存在指標を満たすか否かを前記グループ毎に判定する、
請求項１０又は１１に記載の分散匿名化方法。
【請求項１３】
コンピュータに、

ユーザ識別子とユーザに関する情報とを関連付けて記憶し、
入力される複数の識別子のうち、前記記憶されているユーザ識別子と、当該ユーザ識別子とは異なる識別子であるダミー識別子とを複数のグループに分割し、かつ、前記ダミー識別子が前記複数のグループに分散されるように分割し、
前記グループ毎に含まれる前記識別子に関する情報を他装置に送信し、
前記送信された情報に基づいて前記他装置で分割されたグループ毎に含まれる識別子に関する情報を、前記他装置から受信し、
自装置と、前記他装置とのいずれにも存在する識別子の割合が予め定めた匿名指標を満たすか否かを前記グループ毎に判定する、
処理を実行させるプログラム。
【請求項１４】
コンピュータに、
前記入力される複数の識別子のうち、前記記憶されているユーザ識別子とは異なる識別子を前記ダミー識別子として設定し、
前記グループ毎に含まれる前記識別子に関する情報は、前記グループ毎における前記識別子の内容を示す情報である
処理を実行させる請求項１３記載のプログラム。
【請求項１５】
コンピュータに、
前記判定の際、さらに前記複数のグループが含む全ての識別子に対する前記ユーザ識別子の数の割合である存在指標を満たすか否かを前記グループ毎に判定する、
処理を実行させる請求項１３又は１４に記載のプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、分散して保持されている情報を結合する際の匿名化技術に関する。
【背景技術】
【０００２】
分散して保持されている情報を結合する際に、個人の特定や属性の推定を防ぐための匿名化（分散匿名化）の技術が知られている。
例えば、非特許文献１の技術は、２つの事業者の間でデータを結合する際に、まず、２つの事業者がそれぞれ保持する個人情報を抽象化して初期匿名テーブルを生成する。そして、非特許文献１の技術は、抽象化された個人情報を、匿名性を満たすことを確認しながら徐々に具体化していく。
個人情報の具体化のために、一方の事業者は、個人情報の分割点の候補を決定し、分割点で分割したユーザ識別子のリストを、もう一方の事業者に通知する。センシティブ情報を保持している事業者は、通知された分割点でデータを分割した際に、ｋ匿名性とｌ多様性という二つの指標が満たされるか否かを、確認する。ここでセンシティブ情報とは、結合後のデータの情報処理に用いるため、変更したくない情報のことを言う。それらの二つの指標を満たしているデータからは、個人を特定することができない。
それらの二つの指標を満たしているデータのみを利用者に提供することで、提供するデータからの、個人の特定を防ぐことができる。言い換えると、非特許文献１の技術により、個人のセンシティブ情報の特定を防ぐことができる。
【先行技術文献】
【非特許文献】
【０００３】
【非特許文献１】"Ｐｒｉｖａｃｙ－Ｐｒｅｓｅｒｖｉｎｇ Ｄａｔａ Ｍａｓｈｕｐ", Ｎｏｍａｎ Ｍｏｈａｍｍｅｄ, Ｂｅｎｊａｍｉｎ Ｃ. Ｍ. Ｆｕｎｇ, Ｋｅ Ｗａｎｇ, Ｐａｔｒｉｃｋ Ｃ. Ｋ. Ｈｕｎｇ, Ｉｎ ＥＤＢＴ’０９ Ｐｒｏｃｅｅｄｉｎｇｓ ｏｆ ｔｈｅ １２ｔｈ Ｉｎｔｅｒｎａｔｉｏｎａｌ Ｃｏｎｆｅｒｅｎｃｅ ｏ

n Extending Database Technology：Advances in Database Technology, 2009.
【非特許文献２】“OpenID Authentication 2.0－Final”, OpenID Foundation, 20 07, http：／／openid.net／specs／openid－authentication－2_0.html, http：／／openid－foundation－japan.github.com／openid－authentication.html

【発明の概要】

【発明が解決しようとする課題】

【０００４】

非特許文献１に記載の技術の課題は、分散匿名化処理の過程で、他の事業者にユーザのデータの存在が漏洩してしまうことにある。例えば、事業者Ａと事業者Ｂとがそれぞれ保持しているデータを結合する際の匿名化処理について説明する。匿名化処理のプロトコルの途中で、事業者Ａが事業者Ｂに対して個人情報（例えばユーザＩＤ等）を通知したとする。この場合、事業者Ａからの通知によって、事業者Ｂに、「通知されたユーザＩＤのユーザのデータが、少なくとも事業者Ａが保持するデータに存在すること」が、漏洩してしまう。

本発明の目的の一つは、分散匿名化処理の過程でユーザに関するデータの存在が漏洩することのない分散匿名化システム、分散匿名化装置、分散匿名化方法及びプログラムを提供することにある。

【課題を解決するための手段】

【０００５】

上記目的を達成するため、本発明における分散匿名化システムは、全ユーザの識別子を管理する識別子管理装置と、第一の分散匿名化装置と、第二の分散匿名化装置と、を含む分散匿名化システムであって、前記識別子管理装置は、前記管理している識別子を前記第一の分散匿名化装置及び第二の分散匿名化装置に通知を行い、前記第一の分散匿名化装置は、ユーザ識別子と個人情報とを関連付けて記憶する第一の記憶手段と、前記識別子通知手段から通知された全識別子のうち、前記第一の記憶手段が記憶するユーザ識別子に該当しない識別子をダミー識別子として設定する第一の設定手段と、前記設定せれたダミー識別子を含む全識別子をグループに分割する第一の分割手段と、前記分割された各グループにおける識別子の内容を示す第一の分割情報を前記第二の分散匿名化装置に送信する第一の送信手段と、前記第二の分散匿名化装置から送信された第二の分割情報を受信し、該分割情報に基づいてさらに前記全識別子をグループに分割する第一の受信手段と、を含み、前記第二の分散匿名化装置は、ユーザ識別子と個人情報とを関連付けて記憶する第二の記憶手段と、前記識別子通知手段から通知された全識別子のうち、前記第二の記憶手段が記憶するユーザ識別子に該当しない識別子をダミー識別子として設定する第二の設定手段と、前記設定されたダミー識別子を含む全識別子をグループに分割する第二の分割手段と、前記分割された各グループにおける識別子の内容を示す第二の分割情報を前記第一の分散匿名化装置に送信する第二の送信手段と、前記第一の分散匿名化装置から送信された第一の分割情報を受信し、該分割情報に基づいてさらに前記全識別子をグループに分割する第二の受信手段と、を含み、前記第一の分散匿名化装置と前記第二の分散匿名化装置との少なくともいずれか一方は、前記第一の分散匿名化装置と前記第二の分散匿名化装置とのいずれにも存在する識別子の割合が予め定めた匿名指標を満たすか否かを前記分割後のグループ毎に判定する判定手段と、前記判定手段が匿名指標を満たさないと判定した場合に、最後に行った分割をキャンセルして、結合匿名化テーブルを生成する生成手段と、をさらに含む。

上記目的を達成するため、本発明における分散匿名化装置は、データとして存在するユーザの識別子であるユーザ識別子と個人情報とを関連付けて記憶する記憶手段と、外部から通知された複数の識別子である全識別子のうち、前記ユーザ識別子に該当しない識別子をダミー識別子として設定する設定手段と、前記設定されたダミー識別子を含む全識別子

をグループに分割する分割手段と、前記分割された各グループにおける識別子の内容を示す分割情報を他装置に送信する送信手段と、自装置と、前記他装置とのいずれにも存在する識別子の割合が予め定めた匿名指標を満たすか否かを前記分割後のグループ毎に判定する判定手段と、を含む。

上記目的を達成するため、本発明における分散匿名化方法は、コンピュータが、データとして存在するユーザの識別子であるユーザ識別子と個人情報とを関連付けて記憶し、外部から通知された複数の識別子である全識別子のうち、前記ユーザ識別子に該当しない識別子をダミー識別子として設定し、前記設定されたダミー識別子を含む全識別子をグループに分割し、前記分割された各グループにおける識別子の内容を示す分割情報を他装置に送信し、自装置と、前記他装置とのいずれにも存在する識別子の割合が予め定めた匿名指標を満たすか否かを前記分割後のグループ毎に判定する。

上記目的を達成するため、本発明における不揮発性媒体に記録されたプログラムは、コンピュータに、データとして存在するユーザの識別子であるユーザ識別子と個人情報とを関連付けて記憶し、外部から通知された複数の識別子である全識別子のうち、前記ユーザ識別子に該当しない識別子をダミー識別子として設定し、前記設定されたダミー識別子を含む全識別子をグループに分割し、前記分割された各グループにおける識別子の内容を示す分割情報を他装置に送信し、自装置と、前記他装置とのいずれにも存在する識別子の割合が予め定めた匿名指標を満たすか否かを前記分割後のグループ毎に判定する、処理を実行させる。

【発明の効果】

【０００６】

本発明の効果の一例は、他の事業者にユーザのデータの存在が漏洩する危険性なしで、分散匿名化処理を実行可能なことである。

【図面の簡単な説明】

【０００７】

【図１】図１は、第１実施形態に係る分散匿名化システム１０００の構成を示すブロック図である。

【図２】図２は、第一の分散匿名化装置１００の構成を示すブロック図である。

【図３】図３は、第二の分散匿名化装置２００の構成を示すブロック図である。

【図４】図４は、本発明の第１実施形態に係る分散匿名化システム１０００の動作を示すフローチャート図である。

【図５】図５は、第一の設定部１３０が再構成したダミー識別子を含むテーブルの例を示す図である。

【図６】図６は、ダミー識別子に適当な個人情報の値を割り当てたテーブルの例を示す図である。

【図７】図７は、本発明の第１実施形態における、事業者Ａが保持する初期匿名テーブルの例を示す図である。

【図８】図８は、本発明の第１実施形態における、事業者Ｂが保持する初期匿名テーブルの例を示す図である。

【図９】図９は、図７のテーブルを身長１７０で分割したデータを表す図である。

【図１０】図１０は、図８のテーブルを受信した分割情報に基づいて分割したデータを表す図である。

【図１１】図１１は、図１０のテーブルを年齢３０で分割したデータを表す図である。

【図１２】図１２は、図９のテーブルを受信した分割情報に基づいて分割したデータを表す図である。

【図１３】図１３は、図１１のテーブルを年齢４０で分割したデータを表す図である。

【図１４】図１４は、図１２のテーブルを、受信した分割情報に基づいて分割したデータを表す図である。

【図１５】図１５は、双方に存在する人数を計算したテーブルを示す図である。

【図１６】図１６は、第１実施形態に係る本発明により生成された最終的な結合された匿

名化テーブルを示す図である。
【図１７】図１７は、第２実施形態に係る第一の分散匿名化装置５００の構成を示すブロック図である。
【図１８】図１８は、第１実施形態に係る第一の操作部１４０が、ダミー識別子に適当な個人情報の値を関連付けたテーブルを示す図である。
【図１９】図１９は、第２実施形態に係る第一の操作部１４５が、ダミー識別子に個人情報の値として幅を持った値を関連付けたテーブルを示す図である。
【図２０】図２０は、第１実施形態に係る第一の設定部１３０によって、ダミー識別子が設定されたテーブルを示す図である。
【図２１】図２１は、第３実施形態に係る第一の操作部１４５が、ユーザ識別子の値の分散に基づいて、ダミー識別子の値を関連付けたテーブルを示す図である。
【図２２】図２２は、第４実施形態に係る第一の分散匿名化装置６００の構成を示すブロック図である。
【図２３】図２３は、第一の生成部１９５が、全てのダミー識別子を残して生成した結合匿名化テーブルの例を示す図である。
【図２４】図２４は、第一の生成部１９５が、ユーザ識別子を１つ削除して生成した結合匿名化テーブルの例を示す図である。
【図２５】図２５は、第５実施形態に係る分散匿名化装置７００の構成を示すブロック図である。
【図２６】図２６は、第５実施形態に係る分散匿名化装置７００の動作のフローチャート図である。
【図２７】図２７は、第１実施形態に係る第一の分散匿名化装置１００のハードウェア構成の一例を示すブロック図である。
【図２８】図２８は、事業者Ａの装置が保持する個人情報のテーブルの例を示す図である。
【図２９】図２９は、事業者Ｂの装置が保持する個人情報のテーブルの例を示す図である。
【図３０】図３０は、事業者Ａの装置が保持する個人情報の初期匿名テーブルの例を示す図である。
【図３１】図３１は、事業者Ｂの装置が保持する個人情報の初期匿名テーブルの例を示す図である。
【図３２】図３２は、図３０のテーブルを身長１７０で分割したテーブルを表す図である。
【図３３】図３３は、図３１のテーブルを受信した分割情報に基づいて分割したデータを表す図である。
【図３４】図３４は、図３３のテーブルを年齢３０で分割したデータを表す図である。
【図３５】図３５は、図３２のテーブルを受信した分割情報に基づいて分割したデータを表す図である。
【図３６】図３６は、図３４のテーブルを年齢４０で分割したデータを表す図である。
【図３７】図３７は、図３５のテーブルを、受信した分割情報に基づいて分割したデータを表す図である。
【図３８】図３８は、最終的な結合された匿名化テーブルを示す図である。
【図３９】図３９は、本発明のプログラムを記録する、記録媒体の例を示す図である。
【発明を実施するための形態】
【０００８】
＜第１実施形態＞
まず、本発明の実施形態の理解を容易にするために、本発明の背景を説明する。
事業者Ａと事業者Ｂという異なる事業者がそれぞれ保持する個人情報を、匿名性及び多様性を保ちつつ結合する場合を説明する。
ここでは例として、事業者Ａは病院であり、事業者Ａは、身長と病気に関する個人情報

を保持しているとする。また、事業部Ｂはスポーツセンターであり、事業者Ｂは、年齢に関する個人情報を保持しているとする。また、各事業者が保持する個人情報は、識別子管理事業者が管理する共通の識別子に対応している。

本例においては、病気に関する個人情報がセンシティブ情報であるとする。センシティブ情報以外の個人情報は準識別子と呼ばれる。なお、外見からはわからず、他人に知られたくない情報（病気に関する情報）をセンシティブ情報とし、それ以外の外見からある程度推測可能な情報（身長、年齢等）を準識別子として区別しても良い。

分散匿名化の技術には、非特許文献１の技術を用いるものとする。匿名性及び多様性が保たれているか否かは、予め定めたｋ匿名性及びｌ多様性の指標を満たすか否かによって判定される。ｋ匿名性とは、準識別子の組み合わせが同じユーザをｋ人以上にすることを要求する指標である。ｌ多様性とは、準識別子の組み合わせが同じユーザのセンシティブ情報をｌ通り以上にすることを要求する指標である。以降の本例の説明では、個人情報のテーブルが２匿名性及び２多様性を満たすことを要求するものとする。

まず、識別子管理事業者は、結合の対象となるユーザの識別子を各事業者に対して通知する。例えば、ｕｓｅｒ１～ｕｓｅｒ１２の識別子が各事業者に通知されたものとする。

事業者Ａの装置は、通知された識別子のユーザに関して、図２８に示す個人情報のテーブルを保持しているとする。図２８に示すように、事業部Ａの装置は、ｕｓｅｒ１、ｕｓｅｒ３、ｕｓｅｒ５、ｕｓｅｒ７、ｕｓｅｒ８、ｕｓｅｒ１０、ｕｓｅｒ１１、ｕｓｅｒ１２の計８つの識別子のユーザに関する個人情報を保持している。

事業者Ｂの装置は、通知された識別子のユーザに関して、図２９に示す個人情報のテーブルを保持しているとする。図２９に示すように、事業部Ｂの装置は、通知された識別子の全ユーザ（ｕｓｅｒ１～ｕｓｅｒ１２の識別子のユーザ）に関する個人情報を保持している。

非特許文献１の技術は、各個人情報を抽象化した初期匿名テーブルを生成する。例えば、非特許文献１の技術は、事業者Ａの装置が保持する図２８のテーブルから図３０に示す初期匿名テーブルを生成する。また、非特許文献１の技術は、事業者Ｂの装置が保持する図２９のテーブルから図３１に示す初期匿名テーブルを生成する。

非特許文献１の技術は、図３０や図３１のように抽象化されたテーブルから、匿名性及び多様性を満たすことを確認しながら徐々にテーブルの個人情報を具体化していく。

事業者Ａの装置は、個人情報が特定されない、安全な個人情報の分割点を決定する。ここでは、事業者Ａの装置は、準識別子である身長の平均値を分割点に決定する。具体的には事業者Ａの装置は、身長１７０を分割点に決定する。

図３２は、図３０のテーブルを身長１７０で分割したデータを表す図である。図３２に示すように、身長１７０を分割点とすると、ユーザ（識別子）は、｛ｕｓｅｒ１、ｕｓｅｒ３、ｕｓｅｒ５、ｕｓｅｒ７｝及び｛ｕｓｅｒ８、ｕｓｅｒ１０、ｕｓｅｒ１１、ｕｓｅｒ１２｝に分割される。事業者Ａの装置は、ユーザ（識別子）の分割情報（｛ｕｓｅｒ１，３，５，７｝及び｛ｕｓｅｒ８，１０，１１，１２｝という２つのグループにユーザ（識別子）を分割することを示す情報）を事業者Ｂに送信する。分割情報は、例えば分割点で分割したユーザ識別子のリストでも良い。

この時、事業者Ｂは、送信されたユーザ（識別子）の分割情報から、事業者Ａが保持するデータに、どのユーザのデータが存在するかがわかってしまう。具体的には、事業者Ｂは、事業者Ａが保持するデータに、ｕｓｅｒ１、ｕｓｅｒ３、ｕｓｅｒ５、ｕｓｅｒ７、ｕｓｅｒ８、ｕｓｅｒ１０、ｕｓｅｒ１１及びｕｓｅｒ１２の識別子に相当する、８人のユーザのデータが存在することがわかってしまう。

この問題が、上述した「分散匿名化処理の過程で、他の事業者にユーザのデータの存在が漏洩してしまう」という問題（問題１）である。本実施形態に係る分散匿名化システムは、問題１に加えて、後述する結合データから、ユーザのデータの存在が漏洩するという問題も解決する。

事業者Ｂの装置は、事業者Ａの装置から分割情報を受信する。続けて、事業者Ｂの装置は、図３１に示す初期匿名テーブルを分割情報に基づいて分割する。図３３は、図３１の

テーブルを、受信した分割情報に基づいて分割したデータを表す図である。

次に、事業者Ａの装置は、図３２のテーブルの匿名性及び多様性が保たれているかを確認する。図３２の身長が１７０以下のグループ（一行目のグループ）は、４匿名及び２多様で匿名性及び多様性が保たれている。

具体的には、匿名性は、準識別子（身長及び年齢）の組み合わせが同じユーザが４人いるので４匿名である。また、多様性は、準識別子の組み合わせが同じユーザのセンシティブ情報（病気に関する個人情報）がガンと心臓病の２通りなので２多様である。

また、図３２の身長が１７０以上のグループ（二行目のグループ）も、４匿名及び２多様で匿名性及び多様性が保たれている。

なお、本例においては、センシティブ情報を保持しているのは事業者Ａのみなので、匿名性及び多様性の確認は事業者Ａの装置のみが行えばよい。

事業者Ａの装置が「事業者Ａが保持するテーブルの匿名性及び多様性が保たれていること」を確認すると、次に事業者Ｂの装置は、次の分割点を決定する。ここでは、事業者Ｂの装置は、準識別子である年齢の平均値を分割点に決定する。具体的には事業者Ｂの装置は、年齢３０を分割点に決定する。

図３４は、図３３のテーブルを年齢３０で分割したデータを表す図である。図３４に示すように、年齢３０を分割点とすると、ユーザ（識別子）は、｛ｕｓｅｒ１、ｕｓｅｒ３｝、｛ｕｓｅｒ５、ｕｓｅｒ７｝及び｛ｕｓｅｒ８、ｕｓｅｒ１０、ｕｓｅｒ１１、ｕｓｅｒ１２｝に分割される。事業者Ｂの装置は、ユーザ（識別子）の分割情報（｛ｕｓｅｒ１，３｝｛ｕｓｅｒ５，７｝及び｛ｕｓｅｒ８，１０，１１，１２｝という３つのグループにユーザ（識別子）を分割することを示す情報）を事業者Ａに送信する。

事業者Ａの装置は、事業者Ｂの装置から分割情報を受信して、図３２のテーブルを分割情報に基づいて分割する。図３５は、図３２のテーブルを受信した分割情報に基づいて分割したデータを表す図である。

次に、事業者Ａの装置は、図３５のテーブルの匿名性及び多様性が保たれているかを確認する。匿名性は、上から２匿名、２匿名及び４匿名であり、２匿名性の指標を満たす。また、多様性はいずれも２多様であり多様性の指標を満たす。

次に例えば事業者Ａの装置は、適当な分割点がないと判断したとする。その場合、事業者Ａの装置は、分割点がない旨を事業者Ｂに送信する。事業者Ｂの装置は、事業者Ａから分割点がない旨を受信すると、適当な分割点を決定する。事業者Ｂの装置は、例えば年齢４０を分割点に決定する。

図３６は、図３４のテーブルを年齢４０で分割したデータを表す図である。図３６に示すように、年齢４０を分割点とすると、ユーザ（識別子）は、｛ｕｓｅｒ１、ｕｓｅｒ３｝、｛ｕｓｅｒ５、ｕｓｅｒ７｝、｛ｕｓｅｒ８、ｕｓｅｒ１０｝、及び｛ｕｓｅｒ１１、ｕｓｅｒ１２｝に分割される。事業者Ｂの装置は、ユーザ（識別子）の分割情報（｛ｕｓｅｒ１，３｝｛ｕｓｅｒ５，７｝、｛ｕｓｅｒ８，１０｝及び｛ｕｓｅｒ１１，１２｝という４つのグループにユーザ（識別子）を分割することを示す情報）を事業者Ａに送信する。

事業者Ａの装置は、事業者Ｂから分割情報を受信して、図３５のテーブルを分割情報に基づいて分割する。図３７は、図３５のテーブルを、受信した分割情報に基づいて分割したデータを表す図である。

次に、事業者Ａの装置は、図３７のテーブルの匿名性及び多様性が保たれているかを確認する。いずれの行も２匿名及び２多様であり、匿名性及び多様性の指標が満たされている。

これ以上分割を行った場合、匿名性及び多様性を満たさないことは明らかなので、事業部Ａ及び事業部Ｂはデータの分割を終了し、それぞれの分割データを出力し合い、データを結合する。

図３８は、最終的な結合された匿名化テーブルを示す図である。図３８に示すように、準識別子の組み合わせのグループ毎に２匿名性及び２多様性の指標が保たれている。そのため、図３８からは個人のセンシティブ情報を特性することはできない。具体的には、事

業者Ｂは、図３８を見ても、どのユーザがなんの病気かを知ることはできない。
しかし、事業者Ｂは、自身が保持するデータから「４０歳以上のユーザは、識別子がｕｓｅｒ１１及びｕｓｅｒ１２である、２名のユーザしかいないことがわかる。そのため、事業者Ｂは、図３８を見ることで、少なくとも識別子がｕｓｅｒ１１及びｕｓｅｒ１２である、２名のユーザのデータが事業者Ａが保持するデータに存在することがわかる。
すなわち、上述した問題１とは別に、「最終的な結合された匿名化テーブルから、他の事業者にユーザのデータの存在が漏洩してしまう」という問題（問題２）がある。
以上の問題１及び問題２は、例えば「具体的な病気の特定まではできないが、あるユーザがガン又は心臓病のため病院に通っている」ことが漏洩することを意味している。
以下に説明される本発明の第１実施形態によれば、これまでに説明した問題１及び問題２が解決される。
まず、図１～図３を参照して、本発明の第１実施形態に係る分散匿名化システム１０００の機能構成を説明する。
図１は、第１実施形態に係る分散匿名化システム１０００の構成を示すブロック図である。図１に示すように分散匿名化システム１０００は、第一の分散匿名化装置１００と、第二の分散匿名化装置２００と、識別子管理装置３００と、情報提供装置４００とを含む。本実施形態においては、分散匿名化装置は２台として説明するが、２台に限定されず、２台以上の複数の装置を含むシステムでも良い。
第一の分散匿名化装置１００は、例えば上述した事業部Ａが分散匿名化処理のために備える装置である。
第二の分散匿名化装置２００は、例えば上述した事業部Ｂが分散匿名化処理のために備える装置である。第二の分散匿名化装置２００は、第一の分散匿名化装置１００と協同して個人情報のテーブルの分割を繰り返す。
識別子管理装置３００は、第一の分散匿名化装置１００及び第二の分散匿名化装置２００が共通で使用する識別子を管理する。識別子管理装置３００は、第一の分散匿名化装置１００又は第二の分散匿名化装置２００の少なくともいずれか一方に存在する、全ユーザの識別子を管理している。
識別子管理装置３００は、管理している全識別子を第一の匿名化装置１００及び第二の匿名化装置２００に通知する。又は識別子管理装置３００は、管理している全識別子ではなく、分散匿名化システム１０００の処理の対象となる識別子を特定し、特定された全識別子を通知しても良い。
識別子管理装置３００が管理する識別子は、例えば国民ＩＤでも良い。又は識別子管理装置３００が管理する識別子は、非特許文献２に記載されているＯｐｅｎＩＤでも良く、これらに限定されない。
また、予め第一の分散匿名化装置１００及び第二の分散匿名化装置２００が、全識別子のデータを保持していても良い。例えば、第一の分散匿名化装置１００及び第二の分散匿名化装置２００は、第一の分散匿名化装置１００及び第二の分散匿名化装置２００に登録することを許可されているユーザの国民ＩＤを保持していても良い。
図２は、第一の分散匿名化装置１００の構成を示すブロック図である。図２に示すように、第一の分散匿名化装置１００は、第一の取得部１１０と、第一の記憶部１２０と、第一の設定部１３０と、第一の操作部１４０と、第一の分割部１５０と、第一の送信部１６０と、第一の受信部１７０と、第一の判定部１８０と、第一の生成部１９０とを含む。
第一の取得部１１０は、識別子管理装置３００からの通知を受け、母集団となる全識別子を取得する。第一の取得部１１０は、取得した全識別子のデータを第一の設定部１３０に出力する。
第一の記憶部１２０は、ユーザ識別子と個人情報とを関連付けて記憶する。ここで「ユーザ識別子」は、ある装置に注目した時に、実際にその装置が記憶しているユーザの識別子を意味する。例えば、「第一の記憶部１２０が記憶するユーザ識別子」は、第一の記憶部１２０が記憶している識別子を意味する。即ち、「第一の記憶部１２０が記憶するユーザ識別子」は、後述する第二の記憶部２２０が記憶しているが、第一の記憶部１２０が記

憶していないユーザの識別子は含まない。
第一の設定部１３０は、第一の取得部１１０から通知された複数の識別子である全識別子のうち、第一の記憶部１２０が記憶するユーザ識別子に該当しない識別子をダミー識別子として設定する。第一の設定部１３０は、ダミー識別子であると設定した識別子にダミーフラグをたてても良い。第一の設定部１３０は、ダミー識別子が設定されたデータを第一の操作部１４０に出力する。なお、第二の分散匿名化装置２００は、全識別子のうち、どの識別子が第一の設定部１３０が設定したダミー識別子であるか否かを、特定できない。
第一の操作部１４０は、第一の設定部１３０から出力されたデータから、分割の開始の状態のテーブル（以下「初期匿名テーブル」という）を生成する。なお、第一の操作部１４０は、初期匿名テーブルの生成の前に、ダミー識別子に適当な個人情報の値（準識別子及びセンシティブ情報の値）を関連付けても良い。第一の操作部１４０は、初期匿名テーブルのデータを第一の分割部１５０に出力する。
第一の分割部１５０は、第一の操作部１４０から出力された初期匿名テーブルに含まれる全識別子のデータをグループに分割する。分割方法は特に限定されない。第一の分割部１５０は、所定の準識別子の値の平均値を分割点として、データを２つのグループに分割しても良い。又は、第一の分割部１５０は、周知のヒューリスティック関数を用いて分割点を決定しても良い。
また、第一の分割部１５０は、周知のヒューリスティック関数に加えて、ダミー識別子の情報エントロピー量を考慮して分割点を決定しても良い。ダミー識別子の情報エントロピー量を考慮することで、第一の分割部１５０は、ダミー識別子が分割後のデータに適度に分散されて入るような分割点を決定する。
例えば、ダミー識別子の情報エントロピー量は、以下のような式で計算される。
ｐ＝「分割後のグループ内でのダミー識別子の数」／「分割後のグループ内での識別子の数（ユーザ識別子の数とダミー識別子の数の合計）」
ダミー識別子の情報エントロピー量＝－１×ｐ×ｌｏｇ（ｐ）
例えば、第一の分割部１５０は、上記のダミー識別子の情報エントロピー量を、分割後に作成される２つのグループ内（分割点以上と未満の２つのグループ）について計算する。ここで、２つのグループの情報エントロピー量を足した値をＳとする。分割後の２つのグループに同じくらいの配分でダミー識別子が含まれるように分割された場合、Ｓの値が最大になる。
Ｓの値を周知のヒューリスティック関数に加えて分割点を決定することで、ダミー識別子が分割後のグループのデータに適度に分散されて入るような分割点が選ばれることになる。Ｓの値を考慮して分割点を決定することで、第一の分割部１５０は、分割の回数を増やすことができる。
なお上述したように、第一の分散匿名化装置１００と第二の分散匿名化装置２００とは、互いのダミーデータがわからない。具体的には第一の分散匿名化装置１００は、自装置が保持するデータのうち、どの識別子がダミーであるかはわかるが、第二の分散匿名化装置２００がどの識別子をダミーとしているかは、知り得ない。
そこで、第一の分割部１５０は、ＭＰＣ（Ｍｕｌｔｉ Ｐａｒｔｙ Ｃｏｍｐｕｔａｔｉｏｎ）又はＳＭＰＣ（Ｓｅｃｕｒｅ Ｍｕｌｔｉ Ｐａｒｔｙ Ｃｏｍｐｕｔａｔｉｏｎ）を用いて、第二の分散匿名化装置２００が保持するダミー識別子の情報も考慮して、分割点となる値を計算しても良い。第一の分割部１５０は、ＭＰＣ等を用いることで、第一の分散匿名化装置１００及び第二の分散匿名化装置２００が互いの個人情報を出すことを一切必要とせずに、分割点となる値を計算することができる。
また、第一の分散匿名化装置１００と第二の分散匿名化装置２００とが保持するそれぞれのデータを考慮して分割点を決定したい場合も、第一の分割部１５０は、ＭＰＣ又はＳＭＰＣを用いて分割点となる値を計算しても良い。第一の分割部１５０がＭＰＣ又はＳＭＰＣを用いる場合は、具体的には、第一の分散匿名化装置１００が保持する身長の値と、第二の分散匿名化装置２００が保持する年齢の値とを考慮して、最適な分割点を決定した

い場合等である。
なお、以降は説明の便宜のため、第一の分割部１５０は、準識別子の値の平均値を分割点としてデータを分割するものとする。
第一の分割部１５０は、グループに分割したデータを第一の送信部１６０に出力する。
第一の送信部１６０は、第一の分割部１５０が全識別子のデータを分割した各グループにおける識別子の内容を示す、分割情報を第二の分散匿名化装置２００に送信する。分割情報は、例えば分割点で分割したユーザ識別子のリストでも良い。
第一の受信部１７０は、第二の送信部２６０から送信される分割情報を受信する。また、第一の受信部１７０は、受信した分割情報に基づいて全識別子のデータを分割する。また、第一の受信部１７０は、分割後のデータを第一の判定部１８０に出力する。
第一の判定部１８０は、第一の分散匿名化装置１００と第二の分散匿名化装置２００とのいずれにも存在する識別子の割合が予め定めた匿名指標を満たすか否かを、分割後のグループ毎に判定する。ここで匿名指標とは、上述したｋ匿名性及びｌ多様性でも良い。
また第一の判定部１８０は、さらに分割後のグループが含む全ての識別子に対する、ユーザ識別子の数の割合である存在指標を満たすか否かを、分割後のグループ毎に判定する。
ここで、存在指標とは、「最終的な結合された匿名化テーブルから、他の事業者にユーザのデータの存在が漏洩してしまう」という問題を解決するための指標である。存在指標は、ダミーを含めた全体の人数のうち、実際のユーザはどれだけ存在するかという、ユーザの存在率を表す。具体的には、存在指標は、分割されたグループ内に実際に存在するユーザ識別子の数をａ、ダミー識別子の数をｂとした場合における、ａ／（ａ＋ｂ）で表される。
分割後のあるグループ内にダミーが一つも入っていない場合（ｂ＝０）、存在指標は１となる。存在指標が１であるということは、分割後のデータと、第一の分散匿名化装置１００が保持するデータとを照らし合わせることで、あるユーザのデータが第二の分散匿名化装置２００が保持するデータの中に存在することを理解（認識）できることを意味する。又は、存在指標が１であるということは、分割後のデータと、第二の分散匿名化装置２００が保持するデータとを照らし合わせることで、あるユーザのデータが第一の分散匿名化装置１００が保持するデータの中に存在することを理解（認識）できることを意味する。
存在指標の値が１未満であれば、上述の「最終的な結合された匿名化テーブルから、他の事業者にユーザのデータの存在が漏洩してしまう」という問題は起こらない。存在指標の閾値は、任意に適当な値が設定されれば良い。以降では説明の便宜のため存在指標の値は１未満であれば良く、具体的には存在指標の閾値を０．９９９９・・・とする。
なお、分割後のデータに互いに異なるダミー識別子が混じっている場合、第一の分散匿名化装置１００及び第二の分散匿名化装置２００は、真に存在するユーザの識別子がわからない。そのため、第一の判定部１８０は、正確な存在指標を計算することができない。この場合、第一の判定部１８０は、上述したＭＰＣ等を用いて、存在指標を満たしているか否かを計算しても良い。一方の装置のデータにのみダミー識別子が含まれている場合は、ＭＰＣによる計算は必ずしも必要ではない。
第一の判定部１８０は、匿名指標又は存在指標のいずれかが満たされていないと判定すると、データを第一の生成部１９０に出力する。
第一の生成部１９０は、第一の判定部１８０からデータが出力されると、該データに最後に行われた分割をキャンセルし、匿名指標及び存在指標が満たされている状態に戻す。この時、第一の送信部１６０は、最後に行った分割をキャンセルする旨の通知を第二の分散匿名化装置２００に送信する。第二の受信部２７０がキャンセルする旨の通知を受信すると、第二の分散匿名化装置２００においても、最後に行った分割がキャンセルされる。
第一の生成部１９０は、第一の分散匿名化装置１００及び第二の分散匿名化装置２００の両装置において、最後に行った分割をキャンセルした２つのテーブルから、最終的な結合された匿名化テーブル（以下「結合匿名化テーブル」という。）を生成する。

第一の生成部１９０は、生成した結合匿名化テーブルを情報提供装置４００に出力する。

情報提供装置４００は、第一の生成部１９０から出力された結合匿名化テーブルを情報利用者へ提供する。

図３は、第二の分散匿名化装置２００の構成を示すブロック図である。図３に示すように、第二の分散匿名化装置２００の構成は、第一の分散匿名化装置１００と同様でも良い。

次に図４を参照して、本発明の第１実施形態に係る分散匿名化システム１０００の動作について説明する。

図４は、本発明の第１実施形態に係る分散匿名化システム１０００の動作を示すフローチャート図である。図４に示すように、識別子管理装置３００は、管理している全識別子を第一の分散匿名化装置１００及び第二の分散匿名化装置２００に通知する（ステップＳ１）。

第一の分散匿名化装置１００において、第一の取得部１１０が全識別子の通知を受けると、第一の設定部１３０は、第一の記憶部１２０が実際に記憶している識別子をユーザ識別子とし、それ以外の識別子をダミー識別子として設定する。第二の分散匿名化装置２００においても同様に、第二の取得部２１０が全識別子の通知を受け、第二の設定部２３０がダミー識別子を設定する。当然のことながら、第一の記憶部１２０と第二の記憶部１２０とが記憶している識別子は異なるので、互いのユーザ識別子及びダミー識別子は異なる。

なお、識別子管理装置３００は、管理している全識別子ではなく、分散匿名化処理の対象となる識別子を特定し、対象となる全識別子を通知するようにしても良い。

次に、第一の操作部１４０は、全識別子のデータから初期匿名テーブルを生成する。第一の操作部１４０は、初期匿名テーブルの生成の前に、ダミー識別子に適当な個人情報の値（準識別子及びセンシティブ情報の値）を関連付けても良い。

次に、第一の分割部１５０は、全識別子のデータを分割するのに分割点の候補が存在するか否かを判定する（ステップＳ２）。分割点の候補が存在すると判定すると、第一の分割部１５０は、候補となった該分割点で全識別子のデータを分割する。続けて、第一の分割部１５０は、分割したグループのデータを第一の送信部１６０に出力する。そして、処理はステップＳ３に進む。分割点の候補が存在しないと判定した場合は、処理はステップＳ６に進む。

ステップＳ３において、第一の送信部１６０は、分割した各グループにおける識別子の内容を示す分割情報を第二の分散匿名化装置２００に送信する。

次に、第二の分散匿名化装置２００において、第二の受信部２７０は、第一の送信部１６０から送信される分割情報を受信する。第二の受信部２７０は、受信した分割情報に基づいて全識別子のデータを分割する（ステップＳ４）。

次に、第一の判定部１８０と、第二の判定部２８０とは、共に分割後のデータが匿名指標、多様指標及び存在指標を満たしているか否かを判定する（ステップＳ５Ａ、Ｓ５Ｂ）。仮にセンシティブ情報が一方の装置しか保持していないのであれば、センシティブ情報を保持している装置のみが匿名指標、多様指標及び存在指標を満たしているか否かを判定しても良い。

第一の判定部１８０と第二の判定部２８０が共に（又はセンシティブ情報を保持している一方の装置が）指標を満たしていると判定すると、第二の分割部２５０は、全識別子のデータをさらに分割するのに適当な分割点の候補が存在するか否かを判定する（ステップＳ６）。

分割点の候補が存在すると判定した場合は、処理は、ステップＳ３～５Ａ、５Ｂと同様のステップＳ７～９Ａ、９Ｂに進む。分割点の候補が存在しないと判定した場合は、処理はステップＳ２に進む。ステップＳ２及びステップＳ６の両方で分割点の候補が存在しないと判定した場合は、処理はステップＳ１０に進む。

ステップＳ５Ａ、５Ｂ、９Ａ又は９Ｂにおいて、指標を見たしていないと判定すると、

第一の生成部１９０と第二の生成部２９０は、最後に行った分割をキャンセルし、互いのデータを最後の指標を満たしている状態に戻す。続けて、第一の生成部１９０又は第二の生成部２９０は、指標を満たしている状態の２つのテーブルから結合匿名化テーブルを生成する。次に、第一の生成部１９０又は第二の生成部２９０は、生成した結合匿名化テーブルを情報提供装置４００に出力する。

情報提供装置４００は、結合匿名化テーブルを情報利用者へ提供する（ステップＳ１０）。

次に、図５～図１６を参照して、図４の各ステップを、具体的に例を用いて説明する。前提として、事業者Ａが、第一の分散匿名化装置１００を有するものとする。また、事業者Ｂが、第二の分散匿名化装置２００を有するものとする。また、識別子管理事業者が、識別子管理装置３００を有するものとする。

また、以降の例は、上述した例と同様の状況を前提とする。具体的には、事業者Ａは病院であり、事業者Ａは、身長と病気に関する個人情報（図２８に示すテーブル）を保持しているとする。また、事業部Ｂはスポーツセンターであり、事業者Ｂは、年齢に関する個人情報（図２９に示すテーブル）を保持しているとする。各事業者が保持する個人情報は、識別子管理事業者が管理する共通の識別子で対応している。また、病気に関する個人情報をセンシティブ情報とし、個人情報のテーブルが２匿名性及び２多様性を満たすことを要求する。

なお、以降の例では、２匿名性及び２多様性に加えて、個人情報のテーブルが１未満の存在率（存在指標）を満たすことを要求する。

図４のステップＳ１において、識別子管理事業者は、管理している全識別子を事業者Ａ及び事業者Ｂに通知する。ここでは、識別子管理事業者は、ｕｓｅｒ１～ｕｓｅｒ１２の識別子を各事業者に通知する。

事業者Ａにおける第一の取得部１１０が全識別子（ｕｓｅｒ１～ｕｓｅｒ１２）の通知を受けると、第一の設定部１３０は、図２８に示す情報と照らし合わせる。照らし合わせた結果、第一の設定部１３０は、実際に記憶しているｕｓｅｒ１、ｕｓｅｒ３、ｕｓｅｒ５、ｕｓｅｒ７、ｕｓｅｒ８、ｕｓｅｒ１０、ｕｓｅｒ１１及びｕｓｅｒ１２の計８つの識別子をユーザ識別子とする。第一の設定部１３０は、それ以外の識別子であるｕｓｅｒ２、ｕｓｅｒ４、ｕｓｅｒ６及びｕｓｅｒ９をダミー識別子として設定する。

事業者Ｂにおける第二の記憶部２２０は、全識別子を保持するため（図２９参照）、第二の設定部２３０は、全識別子をユーザ識別子とし、ダミー識別子設定は行わない。

第一の設定部１３０は、ダミー識別子に相当する実際には存在しないユーザもあたかも存在するかのようにしてテーブルを再構成する。

図５は、第一の設定部１３０が再構成したダミー識別子を含むテーブルの例を示す図である。図５においては、識別子の若い順に、身長が低い方から並んでいる例を示している。保持しているデータの値がバラバラ（身長の並びがバラバラ）の場合は、第一の設定部１３０は、データを順番に並びかえて適当な位置にダミーを挿入しても良い。

次に、第一の操作部１４０が、ダミー識別子に適当な個人情報の値（準識別子及びセンシティブ情報の値）を関連付けても良い。図６は、ダミー識別子に適当な個人情報の値を割り当てたテーブルの例を示す図である。第一の分散匿名化装置１００は、第一の操作部１４０を含まず、図５の状態で以降の処理を進めても良い。

第一の操作部１４０及び第二の操作部２４０は、各個人情報を抽象化した初期匿名テーブルを生成する。例えば、第一の操作部１４０は、図５のテーブルから図７に示す初期匿名テーブルを生成する。また、第二の操作部２４０は、事業者Ｂが保持する図２９のテーブルから図８に示す初期匿名テーブルを生成する。

図７及び図８に示すように、初期匿名テーブルは、識別子（ＩＤ）と、準識別子（年齢及び身長に関する情報）、センシティブ情報（病気に関する情報）及びダミー数を含む。

図４のステップＳ２において、第一の分割部１５０は、個人情報が特定されない安全な個人情報の分割点が存在するか否かを判定する。ここでは、第一の分割部１５０は、準識別子である身長の平均値が分割点として適当であると判定したとする。第一の分割部１５

０は、身長１７０を分割点に決定する。
図９は、図７のテーブルを身長１７０で分割したデータを表す図である。図９に示すように、身長１７０を分割点とすると、ユーザは、｛ｕｓｅｒ１～ｕｓｅｒ７｝及び｛ｕｓｅｒ８～ｕｓｅｒ１２｝に分割される。
図４のステップＳ３において第一の送信部１６０は、ユーザ（識別子）の分割情報（｛ｕｓｅｒ１～ｕｓｅｒ７｝、｛ｕｓｅｒ８～ｕｓｅｒ１２｝という２つのグループにユーザ（識別子）を分割することを示す情報）を事業者Ｂに送信する。
この時、事業者Ｂは、送信されたユーザ（識別子）の分割情報を見ても、事業者Ａが保持するデータに、どのユーザのデータが存在するかはわからない。事業者Ａは、ダミーを含めて全識別子の情報を送信しているからである。ダミーを含めることで、上述した「分散匿名化処理の過程で、他の事業者にユーザのデータの存在が漏洩してしまう」という問題（問題１）が解決される。
図４のステップＳ４において、事業者Ｂの第二の受信部２７０は、事業者Ａから分割情報を受信して、図８のテーブルを分割情報に基づいて分割する。図１０は、図８のテーブルを受信した分割情報に基づいて分割したデータを表す図である。
図４のステップＳ５Ａにおいて、第一の判定部１８０は、図９のテーブルの匿名性及び多様性が保たれているかを確認する。図９の身長が１７０以下のグループ（一行目のグループ）は、ユーザ７名のうち３名がダミーなので４匿名である。また、図５のテーブルより２多様である。従って匿名性及び多様性が保たれている。また、図９の身長が１７０以上のグループ（二行目のグループ）も４匿名及び２多様で匿名性及び多様性が保たれている。
なお、本例においては、センシティブ情報を保持しているのは事業者Ａのみなので、匿名性及び多様性の確認は事業者Ａのみが行えばよい。この場合、図４のステップＳ５Ｂの処理は行われない。
また、本例においては、ダミー識別子は事業者Ａのデータのみに含まれているので、指標を満たしていることの確認は難しくない。仮に事業者Ｂのデータにもダミー識別子が含まれている場合には、第二の判定部２８０が、ＭＰＣを用いて事業者Ａのデータ及び事業者Ｂのデータが共に指標を満たしていることを確認しても良い。
事業者Ａが保持するテーブルの匿名性及び多様性が保たれていることを確認すると、図４のステップＳ６において、事業者Ｂの第二の分割部２５０は次の適当な分割点が存在するか否かを判定する。ここでは、第二の分割部２５０は、準識別子である年齢の平均値が分割点として適当であると判定したとする。第二の分割部２５０は、年齢３０を分割点に決定する。
図１１は、図１０のテーブルを年齢３０で分割したデータを表す図である。図１１に示すように、年齢３０を分割点とすると、ユーザは、｛ｕｓｅｒ１～ｕｓｅｒ３｝、｛ｕｓｅｒ４～ｕｓｅｒ７｝及び｛ｕｓｅｒ８～ｕｓｅｒ１２｝に分割される。
図４のステップＳ７において、第二の送信部２６０は、ユーザの分割情報（｛ｕｓｅｒ１～ｕｓｅｒ３｝、｛ｕｓｅｒ４～ｕｓｅｒ７｝及び｛ｕｓｅｒ８～ｕｓｅｒ１２｝という３つのグループにユーザを分割することを示す情報）を事業者Ａに送信する。
図４のステップＳ８において、事業者Ａの第一の受信部１７０は、事業者Ｂから分割情報を受信して、図９のテーブルを分割情報に基づいて分割する。図１２は、図９のテーブルを受信した分割情報に基づいて分割したデータを表す図である。
図４のステップＳ９Ａにおいて、第一の判定部１８０は、図１２のテーブルの匿名性及び多様性が保たれているかを確認する。匿名性は、識別子の数からダミー識別子の数を引いて、上から２匿名、２匿名及び４匿名であり、２匿名性の指標を満たす。また、多様性は、図５のテーブルから、いずれも２多様であり多様性の指標を満たす。
次に図４のステップＳ２に戻り、第一の分割部１５０は、適当な分割点がないと判断したとする。その場合、第一の送信部１６０は、分割点がない旨を事業者Ｂに送信する。
第二の受信部２７０が、事業者Ａから分割点がない旨を受信すると、図４のステップＳ６において、第二の分割部２５０は適当な分割点を決定する。第二の分割部２５０は、例

えば年齢４０を分割点に決定する。

図１３は、図１１のテーブルを年齢４０で分割したデータを表す図である。図１３に示すように、年齢４０を分割点とすると、ユーザは、｛ｕｓｅｒ１～ｕｓｅｒ３｝、｛ｕｓｅｒ４～ｕｓｅｒ７｝、｛ｕｓｅｒ８～ｕｓｅｒ１０｝及び｛ｕｓｅｒ１１～ｕｓｅｒ１２｝に分割される。

図４のステップＳ７において、第二の送信部２６０は、ユーザの分割情報（｛ｕｓｅｒ１～ｕｓｅｒ３｝、｛ｕｓｅｒ４～ｕｓｅｒ７｝、｛ｕｓｅｒ８～ｕｓｅｒ１０｝及び｛ｕｓｅｒ１１～ｕｓｅｒ１２｝という４つのグループにユーザを分割することを示す情報）を事業者Ａに送信する。

図４のステップＳ８において、第一の受信部１７０は、事業者Ｂから分割情報を受信して、図１２のテーブルを分割情報に基づいて分割する。図１４は、図１２のテーブルを、受信した分割情報に基づいて分割した状態を表す図である。

図４のステップＳ９Ａにおいて、第一の判定部１８０は、図１４のテーブルの匿名指標、多様指標及び存在指標が保たれているかを確認する。第一の判定部１８０は、図１４の４行目のグループのダミー数が０（ｂ＝０）であり、ａ／（ａ＋ｂ）＝２／（２＋０）＝１となり存在指標を満たさないと判定する。

第一の判定部１８０が指標を満たさないと判定すると、第一の生成部１９０は、最後に行った図１２から図１４への分割をキャンセルする。また、第一の送信部１６０は、事業者Ｂにキャンセルの通知を送信する。キャンセルの通知を受信すると、第二の生成部２９０は、最後に行った図１１から図１３への分割をキャンセルする。

第一の生成部１９０又は第二の生成部２９０は、キャンセルしたそれぞれのテーブルについて、ＭＰＣを用いて双方に存在する人数を計算する。

図１５は、双方に存在する人数を計算したテーブルを示す図である。

第一の生成部１９０又は第二の生成部２９０は、キャンセルした２つのテーブルから結合匿名化テーブルを生成する。

図１６は、第１実施形態に係る本発明により生成された最終的な結合された匿名化テーブル（結合匿名化テーブル）を示す図である。

なお、図１６に示す結合匿名化テーブルは、第一の生成部１９０又は第二の生成部２９０ではなく、両装置からテーブルが出力された情報提供装置４００によって生成されても良い。

図４のステップＳ１０において、情報提供装置４００は、結合匿名化テーブル（図１６に示すテーブル）を情報利用者へ提供する。

ここで、最終的に出力される図１６に示すテーブルを参照しても、事業者Ｂはどのユーザのデータが確実に事業者Ａのデータに存在するかはわからない。具体的には、事業者Ｂは、図１６を参照することで「－３０」であるｕｓｅｒ１～ｕｓｅｒ３の３名のうち２名のデータが事業者Ａのデータに存在することがわかるが、その２名を特定することはできない。また、事業者Ｂは、図１６を参照することで「３０－」であるｕｓｅｒ４～ｕｓｅｒ１２の９名のうち６名のデータが事業者Ａのデータに存在することがわかるが、その６名を特定することはできない。

第１実施形態に係る本発明は、匿名指標及び多様指標に加え存在指標を満たすことを確認する。存在指標を満たさない場合、第１実施形態に係る本発明の分散匿名化装置は、ユーザの存在が特定されうる分割をキャンセルすることで、問題２を解決する。ここで、問題２は、「最終的な結合された匿名化テーブルから、他の事業者にユーザのデータの存在が漏洩してしまう」という問題である。

以上説明したように、第１実施形態に係る分散匿名化システム１０００によれば、他の事業者にユーザのデータの存在が漏洩する危険性なしで、分散匿名化処理を実行することができる。その理由は、第１実施形態に係る分散匿名化システム１０００は、他の事業者に送信するデータの中に、実際には存在しないダミーのデータを含めて送信するからである。

また、第１実施形態に係る分散匿名化システム１０００によれば、他の事業者にユーザ

のデータの存在が漏洩する危険性のない、結合匿名化テーブルを生成することができる。その理由は、第１実施形態に係る分散匿名化システム１０００は、存在指標という新たな指標を導入し、存在指標を満たさない場合、ユーザの存在が特定されうる分割をキャンセルして最終的なテーブルを生成するからである。

＜第２実施形態＞

次に図１７～図１９を参照して、本発明の第２実施形態に係る第一の分散匿名化装置５００の機能構成を説明する。

図１７は、第２実施形態に係る第一の分散匿名化装置５００の構成を示すブロック図である。図１７に示すように、第一の分散匿名化装置５００は、第１実施形態における第一の分散匿名化装置１００と比較して、第一の操作部１４０に代えて、第一の操作部１４５を含む点で異なる。第一の操作部１４５以外の構成部については第１実施形態と同様の構成であるため、同様の番号を付し、説明を省略する。

第一の操作部１４５は、第１実施形態における機能に加え、ダミー識別子に個人情報の値として幅を持った値を関連付ける。

図１８及び図１９は第一の操作部１４５の機能を説明するための図である。

図１８は、第１実施形態に係る第一の操作部１４０が、ダミー識別子に適当な個人情報の値を関連付けたテーブルを示す図である。図１８に示すように、ユーザ識別子であるｕｓｅｒ１の身長の値は１５５である。また、ユーザ識別子であるｕｓｅｒ３の身長の値は１６２である。例えば、第一の操作部１４０は、２つのユーザ識別子の間のダミー識別子であるｕｓｅｒ２に身長の値として１５８の値を関連付ける。

図１９は、第２実施形態に係る第一の操作部１４５が、ダミー識別子に個人情報の値として幅を持った値を関連付けたテーブルを示す図である。図１９に示すように、例えば第一の操作部１４５は、ダミー識別子であるｕｓｅｒ２に身長の値として１５６～１６１という幅を持った値を関連付ける。

図１８のようにダミー識別子の値を１つに決めた場合を説明する。このとき、第一の分割部１５０は、分割点を身長１６０とする場合、ｕｓｅｒ２を必ず「－１６０」のグループに含ませる。

一方、図１９のようにダミー識別子の値に幅を持たせた場合を説明する。このとき、第一の分割部１５０は、ｕｓｅｒ２を「－１６０」のグループに含ませるか「１６０－」のグループに含ませるかをプロトコルの途中で判断できる。従って、第一の分割部１５０は、ダミーの偏りがないような、より適切なグループの分割を行うことが可能となる。

以上説明したように、第２実施形態に係る第一の分散匿名化装置５００によれば、ダミーの偏りがないような、より適切なグループの分割を行うことが可能となる。その理由は、第一の操作部１４５が、ダミー識別子に個人情報の値として幅を持った値を関連付けるからである。

＜第３実施形態＞

次に図１７、図２０及び図２１を参照して、本発明の第３実施形態に係る第一の分散匿名化装置５００の機能構成を説明する。

第３実施形態に係る第一の分散匿名化装置５００の構成は、第２実施形態に係る第一の分散匿名化装置５００と同様の構成で良く、図１７で示される。

第３実施形態に係る第一の分散匿名化装置５００は、第一の操作部１４５がダミー識別子に個人情報の値を関連づける方法が、第２実施形態に係る第一の分散匿名化装置５００と異なる。

第３実施形態に係る第一の操作部１４５は、ユーザ識別子の個人情報である値の分散に基づいて、ダミー識別子の値を関連付けても良い。

図２０及び図２１は第３実施形態に係る第一の操作部１４５の機能を説明するための図である。

図２０は、第１実施形態に係る第一の設定部１３０によって、ダミー識別子が設定されたテーブルを示す図である。図２０に示すように、第一の操作部１４０がダミー識別子に何の値も関連付けなかった場合、ダミー識別子の挿入位置が偏ることがある。図２０にお

いて、身長が１７０台のグループ（ｕｓｅｒ１～ｕｓｅｒ４のグループ）にはダミー識別子が２つ存在しており、ユーザの存在率は０．５である。一方、身長が１８０台のグループ（ｕｓｅｒ５～ｕｓｅｒ１２のグループ）にはダミー識別子が２つ存在しており、ユーザの存在率は、０．２５である。

図２１は、第３実施形態に係る第一の操作部１４５が、ユーザ識別子の値の分散に基づいて、ダミー識別子の値を関連付けたテーブルを示す図である。図２１に示すように、例えば第一の操作部１４５は、身長が１７０台のグループのユーザ存在率と、身長が１８０台のグループのユーザ存在率とが同じになるように、ダミー識別子に値を関連付ける。図２１において、身長が１７０台のグループ（ｕｓｅｒ１，ｕｓｅｒ２，ｕｓｅｒ４のグループ）にはダミー識別子が１つ存在しており、ユーザの存在率は０．３３・・・である。一方、身長が１８０台のグループ（ｕｓｅｒ５～ｕｓｅｒ１２及びｕｓｅｒ３のグループ）にはダミー識別子が３つ存在しており、ユーザの存在率は、０．３３・・・である。

図２０のようにダミー識別子の位置が偏っている場合を説明する。このとき、第一の分割部１５０は、例えば身長１８５を分割点とすることができない。身長１８５を分割点とした場合、「１８５－」のグループの存在指標が１になってしまうからである。

一方、図２１のようにダミー識別子の値を分散に基づいて決めた場合を説明する。このとき、第一の分割部１５０は、例えば身長１８５を分割点とすることができる。身長１８５を分割点とした場合、「１８５－」のグループの存在指標は、０．６６・・・となり満たされるからである。

以上説明したように、第３実施形態に係る第一の分散匿名化装置５００によれば、ダミーの偏りがないような、より適切なグループの分割を行うことが可能となる。その理由は、第一の操作部１４５が、ユーザ識別子の個人情報である値の分散に基づいて、ダミー識別子の値を関連付けるからである。

＜第４実施形態＞

次に図２２～図２４を参照して、本発明の第４実施形態に係る第一の分散匿名化装置６００の機能構成を説明する。

図２２は、第４実施形態に係る第一の分散匿名化装置６００の構成を示すブロック図である。図２２に示すように、第一の分散匿名化装置６００は、第１実施形態における第一の分散匿名化装置１００と比較して、第一の生成部１９０に代えて、第一の生成部１９５を含む点で異なる。第一の生成部１９５以外の構成部については第１実施形態と同様の構成であるため、同様の番号を付し、説明を省略する。

第一の生成部１９５は、最終的な結合をした、匿名化テーブルのデータの一部を変更することで、２つの装置が保持するユーザ識別子が包含関係になっている場合にも、対応する。具体的には、例えば事業者Ａが保持するユーザ識別子を事業者Ｂが全て包含している場合にも、事業者Ａは結合匿名化テーブルを参照することで、事業者Ａが保持する、全てのユーザのデータが事業者Ｂのデータに存在することがわかってしまう。このような場合に、第一の生成部１９５が結合匿名化テーブルの一部を変更することで、事業者Ａに、事業者Ｂが保持する、ユーザのデータの存在をわからなくする。

例えば、第一の生成部１９５は、一以上のダミー識別子のデータを残した結合匿名化テーブルを生成しても良い。

図２３は、第一の生成部１９５が、全てのダミー識別子を残して生成した結合匿名化テーブルの例を示す図である。図２３に示すように、第一の生成部１９５は、ダミーのデータも最終的なデータとして残している。図１６と異なりどのデータもダミーである可能性があるため、事業者Ａは、図２３の結合匿名化テーブルを参照してもどのユーザのデータが事業者Ｂのデータに存在するかがわからない。

他の例として、第一の生成部１９５は、一以上のユーザ識別子のデータを削除した結合匿名化テーブルを生成しても良い。

図２４は、第一の生成部１９５が、ユーザ識別子を１つ削除して生成した結合匿名化テーブルの例を示す図である。図２４に示すように、第一の生成部１９５は、「年齢３０、身長１７０－」のグループに属するユーザ識別子を１つ削除している。そのため、図１６

と異なり、事業者Ａは、図２４の結合匿名化テーブルを参照しても、全てのユーザのデータが事業者Ｂのデータに存在することはわからない。具体的には、事業者Ａは、ｕｓｅｒ８とｕｓｅｒ１１のどちらが事業者Ｂに存在し、どちらが存在しないのかがわからない。

以上説明したように、第４実施形態に係る第一の分散匿名化装置６００によれば、２つの装置が保持するユーザ識別子が包含関係になっている場合にも、ユーザのデータの存在を不明にして対応することができる。その理由は、第一の生成部１９５が、最終的な結合をした匿名化テーブルの、データの一部を変更するからである。

＜第５実施形態＞

次に図２５及び図２６を参照して、本発明の第５実施形態に係る分散匿名化装置７００の機能構成を説明する。

図２５は、第５実施形態に係る分散匿名化装置７００の構成を示すブロック図である。図２５に示すように、分散匿名化装置７００は、記憶部７２０と、設定部７３０と、分割部７５０と、送信部７６０と、判定部７８０とを含む。なお、これらは上述した第一の記憶部１２０、第一の設定部１３０、第一の分割部１５０、第一の送信部１６０及び第一の判定部１８０と同様の構成である。

記憶部７２０は、データとして存在するユーザの識別子である、ユーザ識別子と個人情報とを関連付けて記憶する。

設定部７３０は、外部から通知された複数の識別子である、全識別子のうち、ユーザ識別子に該当しない識別子をダミー識別子として設定する。

分割部７５０は、設定部７３０により設定されたダミー識別子を含む全識別子をグループに分割する。

送信部７６０は、分割した各グループにおける識別子の内容を示す分割情報を他装置に送信する。

判定部１８０は、自装置と、前記他装置とのいずれにも存在する識別子の割合が予め定めた匿名指標を満たすか否かを前記分割後のグループ毎に判定する。

図２６は、第５実施形態に係る分散匿名化装置７００の動作のフローチャート図である。なお、図２６においては、動作の説明のため分散匿名化装置７００は図示しない受信部及び生成部を含むものとする。

図２６に示すように、分散匿名化装置７００の設定部１３０は、外部から通知された複数の識別子である全識別子のうち、ユーザ識別子に該当しない識別子をダミー識別子として設定する（ステップＳ１１）。

次に、分散匿名化装置７００は、自装置が分割を行う装置か否かを確認する（ステップＳ１２）。

自装置が分割を行う装置ではないと判断すると、分散匿名化装置７００は他装置からの分割情報の送信を待つ。分散匿名化装置７００の受信部（図示しない）は、他装置から分割情報を受信すると、受信した分割情報に基づいて保持するデータを分割する（ステップＳ１６）。その後分散匿名化装置７００の処理は、ステップＳ１７に進む。

ステップＳ１２において、自装置が分割を行う装置であると判断すると、分散匿名化装置７００の分割部１５０は、全識別子のデータを分割するのに分割点の候補が存在するか否かを判定する（ステップＳ１３）。

分割点の候補が存在すると判断すると、分割部１５０は、該分割点で全識別子のデータを分割し、処理はステップＳ１５に進む。分割点の候補が存在しないと判定した場合は、処理はステップＳ１８に進む。ステップＳ１８において、他装置に分割点の候補が存在する可能性があれば、分散匿名化装置７００は、他装置に分割情報の送信を依頼する旨の通知を出力し、処理はステップＳ１２に進む。他装置にも分割点の候補が存在しないことがわかっている場合には、処理はステップＳ１９に進む。

ステップＳ１５において、送信部１６０は、分割した各グループにおける識別子の内容を示す分割情報を他装置に送信する。

次に、判定部１８０は、分割後のデータが匿名指標及び多様指標を満たしているか否かを判定する（ステップＳ１７）。分散匿名化装置７００がセンシティブ情報を保持してい

ないのであれば、判定部１８０は、匿名指標及び多様指標を満たしているか否かを判定しなくても良い。

指標を満たしていると判定すると、処理はステップＳ１２に進む。指標を満たしていないと判定すると、処理はステップＳ１９に進む。

ステップＳ１９において、分散匿名化装置７００の生成部（図示しない）は最後に行った分割をキャンセルし、互いのデータを最後の指標を満たしている状態に戻す。生成部は他装置と各グループの共通ユーザ数を共有する。生成部は、他装置と共通ユーザ数を計算する際、ＭＰＣ又はＳＭＰＣを用いて計算しても良い。生成部は、共通ユーザ数を共有すると、結合匿名化テーブルを生成する。

以上説明したように、第５実施形態に係る分散匿名化装置７００によれば、他の事業者にユーザのデータの存在が漏洩する危険性なく、分散匿名化処理を実行することができる。

以上、各実施形態を参照して本発明を説明したが、本発明は以上の実施形態に限定されるものではない。本発明の構成や詳細には、本発明のスコープ内で同業者が理解し得る様々な変更をすることができる。

図２７は、第１実施形態に係る第一の分散匿名化装置１００のハードウェア構成の一例を示すブロック図である。

図２７に示すように、第一の分散匿名化装置１００を構成する各部は、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）１と、ネットワーク接続用の通信ＩＦ（Ｉｎｔｅｒｆａｃｅ）２と、メモリ３と、プログラムを格納する記憶装置４とを含む、コンピュータ装置によって実現される。ただし、第一の分散匿名化装置１００の構成は、図２７に示すコンピュータ装置に限定されない。

例えば、第一の取得部１１０、第一の送信部１６０及び第一の受信部１７０は、通信ＩＦ２によって実現されても良い。

ＣＰＵ１は、オペレーティングシステムを動作させて第一の分散匿名化装置１００の全体を制御する。また、ＣＰＵは１、例えばドライブ装置などに装着された記録媒体からメモリ３にプログラムやデータを読み出し、これにしたがって各種の処理を実行する。

例えば第一の設定部１３０、第一の操作部１４０、第一の分割部１５０、第一の判定部１８０及び第一の生成部１９０は、ＣＰＵ１及びプログラムによって実現されても良い。

記録装置４は、例えば光ディスク、フレキシブルディスク、磁気光ディスク、外付けハードディスク、半導体メモリ等であって、コンピュータプログラムをコンピュータ読み取り可能に記録する。記憶装置４は、例えば、インターフェースを変換するための変換ルールを格納していても良い。また、コンピュータプログラムは、通信網に接続されている図示しない外部コンピュータからダウンロードされても良い。

例えば、第一の記憶部１２０は記録装置４によって実現されても良い。

なお、これまでに説明した各実施形態において利用するブロック図は、ハードウェア単位の構成ではなく、機能単位のブロックを示している。これらの機能ブロックはハードウェア及びソフトウェアの任意の組み合わせによって実現される。また、第一の分散匿名化装置１００の構成部の実現手段は特に限定されない。すなわち、第一の分散匿名化装置１００は、物理的に結合した一つの装置により実現されても良いし、物理的に分離した二つ以上の装置を有線又は無線で接続し、これら複数の装置により実現されても良い。

本発明のプログラムは、上記の各実施形態で説明した各動作を、コンピュータに実行させるプログラムであれば良い。

図３９は、上述のプログラムを記録（記憶）する、記録媒体（記憶媒体）７の例を示す図である。記録媒体７は、情報を非一時的に記憶する不揮発性記録媒体である。なお、記録媒体７は、情報を一時的に記憶する記録媒体であってもよい。記録媒体７は、図２６に示す動作をコンピュータ装置（ＣＰＵ１）に実行させるプログラム（ソフトウェア）を記録する。なお、記録媒体７は、さらに、任意のプログラムやデータを記録してよい。

上述のプログラム（ソフトウェア）のコードを記録した記録媒体７が、コンピュータ装置に供給され、ＣＰＵ１は、記録媒体７に格納されたプログラムのコードを読み出して実

行するようにしてもよい。あるいは、ＣＰＵ１は、記録媒体７に格納されたプログラムのコードを、メモリ３に格納するようにしてもよい。すなわち、本実施形態は、コンピュータ装置（ＣＰＵ１）が実行するプログラムを、一時的に又は非一時的に、記憶する記録媒体７の実施形態を含む。

以上、実施形態を参照して本願発明を説明したが、本願発明は上記実施形態に限定されるものではない。本願発明の構成や詳細には、本願発明のスコープ内で当業者が理解し得る様々な変更をすることができる。

この出願は、２０１１年６月２日に出願された日本出願特願２０１１－１２４３９８を基礎とする優先権を主張し、その開示の全てをここに取り込む。

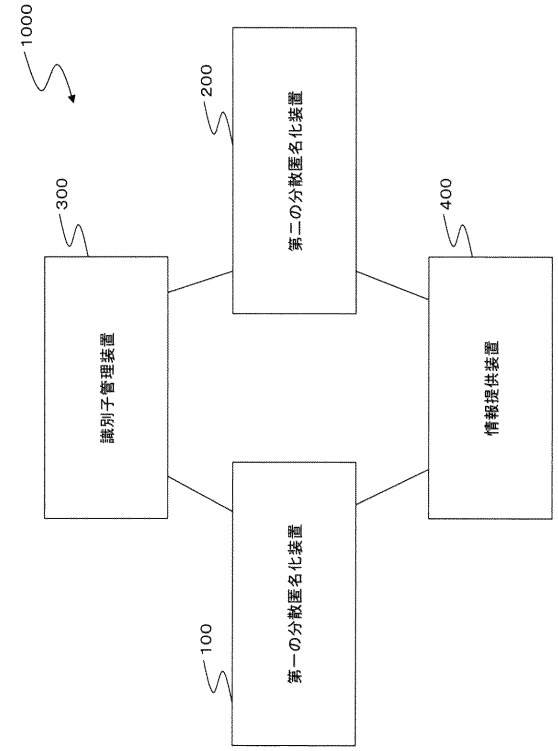
【符号の説明】

【０００９】

１　ＣＰＵ
２　通信ＩＦ
３　メモリ
４　記憶装置
７　記録媒体
１００、５００、６００　第一の分散匿名化装置
１１０　第一の取得部
１２０　第一の記憶部
１３０　第一の設定部
１４０、１４５　第一の操作部
１５０　第一の分割部
１６０　第一の送信部
１７０　第一の受信部
１８０　第一の判定部
１９０、１９５　第一の生成部
２００　第二の分散匿名化装置
２１０　第二の取得部
２２０　第二の記憶部
２３０　第二の設定部
２４０　第二の操作部
２５０　第二の分割部
２６０　第二の送信部
２７０　第二の受信部
２８０　第二の判定部
２９０　第二の生成部
３００　識別子管理装置
４００　情報提供装置
７００　分散匿名化装置
７２０　記憶部
７３０　設定部
７５０　分割部
７６０　送信部
７８０　判定部
１０００　分散匿名化システム

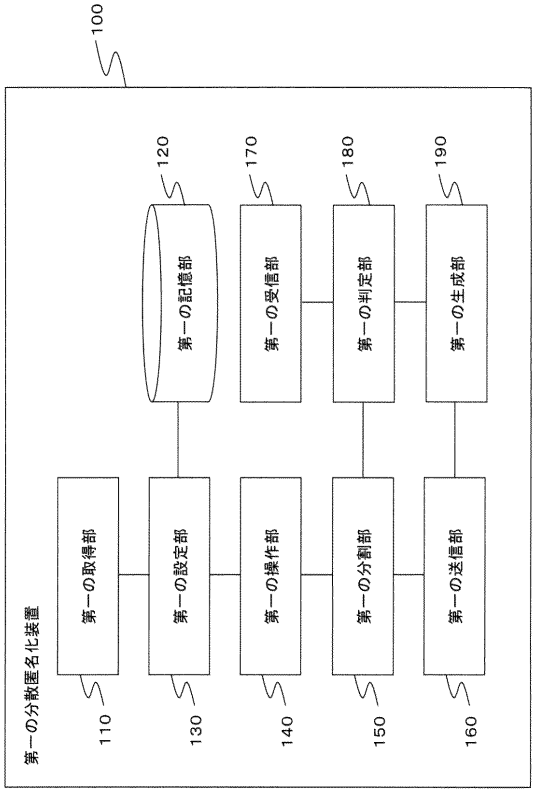
【図１】

図1



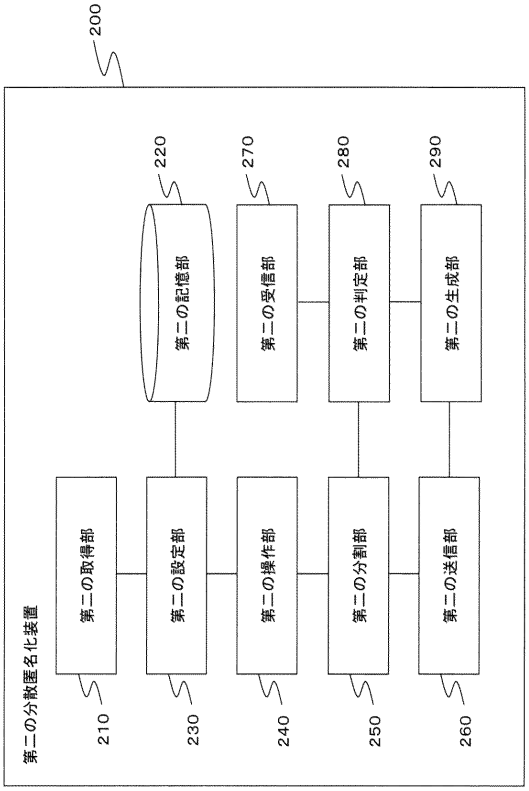
【図２】

図2



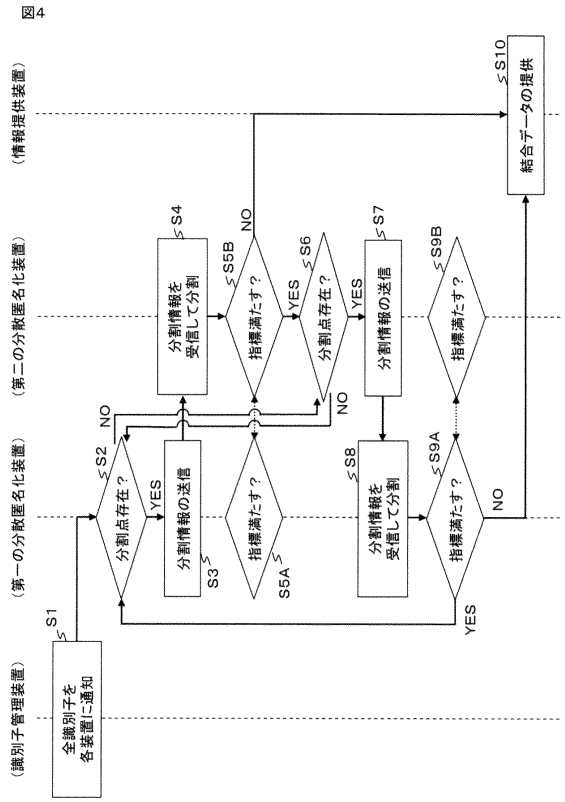
【図３】

図3



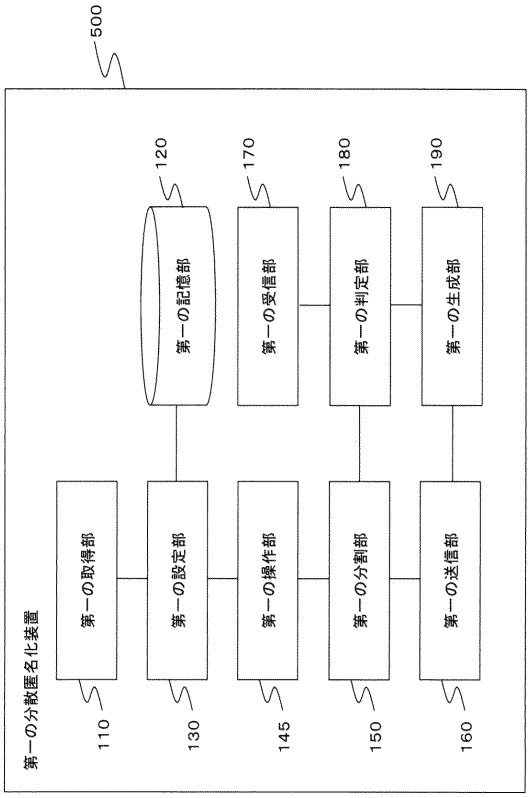
【図４】

図4



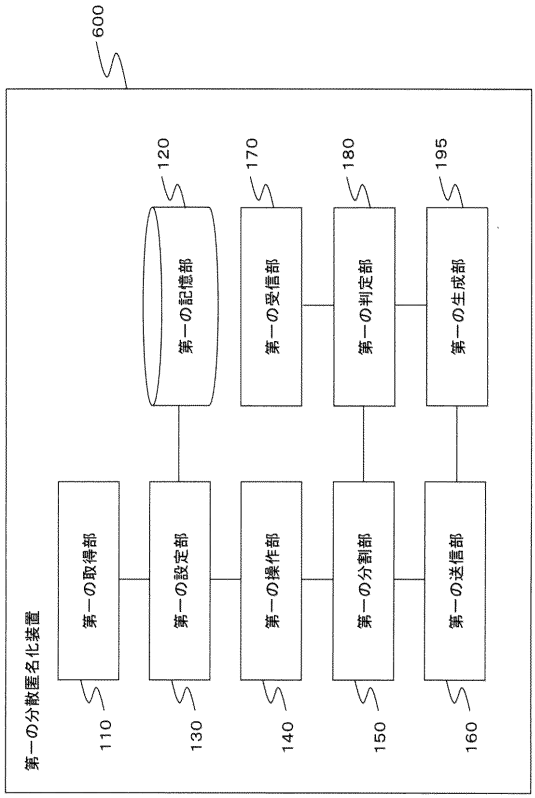
【図１７】

図17



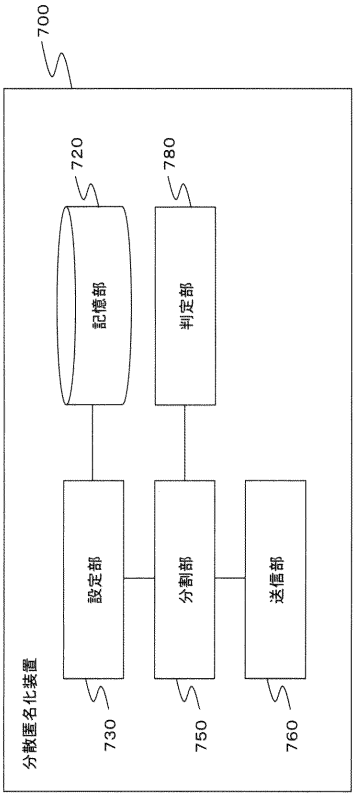
【図２２】

図22



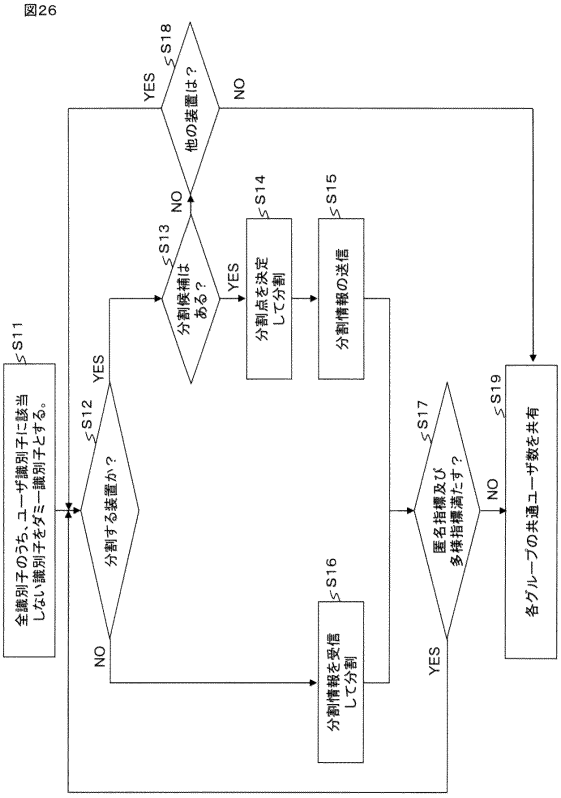
【図２５】

図25



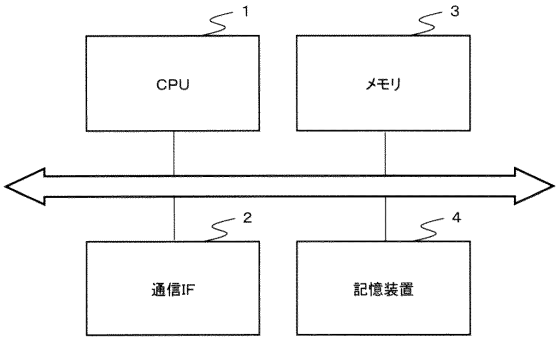
【図２６】

図26



【図２７】

図27



【図３９】

図39

【図５】

図5

ID	身長	病気
user1	160	ガン
user2(ダミー)	-	-
user3	162	心臓病
user4(ダミー)	-	-
user5	165	ガン
user6(ダミー)	-	-
user7	168	心臓病
user8	172	ガン
user9(ダミー)	-	-
user10	175	心臓病
user11	178	ガン
user12	179	心臓病

【図６】

図6

ID	身長	病気
user1	160	ガン
user2（ダミー）	161	ガン
user3	162	心臓病
user4（ダミー）	163	心臓病
user5	165	ガン
user6（ダミー）	166	ガン
user7	168	心臓病
user8	172	ガン
user9（ダミー）	171	心臓病
user10	175	心臓病
user11	178	ガン
user12	179	心臓病

【図７】

図7

年齢	身長	病気	ID	ダミー数
?	any	ガン, 心臓病	user1-12	4

【図８】

図8

年齢	身長	病気	ID	ダミー数
any	?	?	user1～12	0

【図９】

図9

年齢	身長	病気	ID	ダミー数
?	-170	ガン, 心臓病	user1-7	3
?	170-	ガン, 心臓病	user8-12	1

【図１０】

図10

年齢	身長	病気	ID	ダミー数
any	?	?	user1-7	0
any	?	?	user8-12	0

【図１１】

図11

年齢	身長	病気	ID	ダミー数
-30	?	?	user1-3	0
30-	?	?	user4-7	0
30-	?	?	user8-12	0

【図１２】

図12

年齢	身長	病気	ID	ダミー数
?	-170	ガン, 心臓病	user1-3	1
?	-170	ガン, 心臓病	user4-7	2
?	170-	ガン, 心臓病	user8-12	1

【図１３】

図13

年齢	身長	病気	ID	ダミー数
-30	?	?	user1-3	0
30-40	?	?	user4-7	0
30-40	?	?	user8-10	0
40-	?	?	user11-12	0

【図１４】

図14

年齢	身長	病気	ID	ダミー数
?	-170	ガン, 心臓病	user1-3	1
?	-170	ガン, 心臓病	user4-7	2
?	170-	ガン, 心臓病	user8-10	1
?	170-	ガン, 心臓病	user11-12	0

【図１５】

図15

年齢	身長	存在人数	内訳
-30	-170	2名	心臓病1名、ガン1名
30-	-170	2名	心臓病1名、ガン1名
30-	170-	4名	心臓病2名、ガン2名

【図１６】

図16

年齢	身長	病気
-30	-170	心臓病
-30	-170	ガン
30-	-170	心臓病
30-	-170	ガン
30-	170-	心臓病
30-	170-	心臓病
30-	170-	ガン
30-	170-	ガン

【図１８】

図18

ID	身長	病気
user1	155	ガン
user2(ダミー)	158	-
user3	162	心臓病
user4(ダミー)	166	-
user5	175	ガン
user6(ダミー)	176	-
user7	181	心臓病
user8	183	ガン
user9(ダミー)	183	-
user10	184	心臓病
user11(ダミー)	185	-
user12	188	心臓病

【図１９】

図19

ID	身長	病気
user1	155	ガン
user2(ダミー)	156～161	-
user3	162	心臓病
user4(ダミー)	162～174	-
user5	175	ガン
user6(ダミー)	176～180	-
user7	181	心臓病
user8	183	ガン
user9(ダミー)	183	-
user10	184	心臓病
user11(ダミー)	184～187	-
user12	188	心臓病

【図２０】

図20

ID	身長	病気
user1	172	ガン
user2 (ダミー)	-	-
user3 (ダミー)	-	-
user4	175	心臓病
user5	180	ガン
user6 (ダミー)	-	-
user7	181	心臓病
user8	183	ガン
user9 (ダミー)	-	-
user10	184	心臓病
user11	186	ガン
user12	188	心臓病

【図２１】

図21

ID	身長	病気
user1	172	ガン
user2 (ダミー)	173	-
user4	175	心臓病
user5	180	ガン
user6 (ダミー)	180	-
user7	181	心臓病
user8	183	ガン
user9 (ダミー)	183	-
user10	184	心臓病
user11	186	ガン
user3 (ダミー)	187	-
user12	188	心臓病

【図２３】

図23

年齢	身長	病気
-30	-170	ガン
-30	-170	ガン
-30	-170	心臓病
30-	-170	心臓病
30-	-170	ガン
30-	-170	ガン
30-	-170	心臓病
30-	170-	ガン
30-	170-	心臓病
30-	170-	心臓病
30-	170-	ガン
30-	170-	心臓病

【図２４】

図24

年齢	身長	病気
-30	-170	心臓病
-30	-170	ガン
30-	-170	心臓病
30-	-170	ガン
30-	170-	心臓病
30-	170-	心臓病
30-	170-	ガン

【図２８】

図28

ID	身長	病気
user1	160	ガン
user3	162	心臓病
user5	165	ガン
user7	168	心臓病
user8	172	ガン
user10	175	心臓病
user11	178	ガン
user12	179	心臓病

【図２９】

図29

ID	年齢
user1	16
user2	22
user3	19
user4	31
user5	30
user6	33
user7	32
user8	31
user9	35
user10	34
user11	41
user12	40

【図３０】

図30

年齢	身長	病気	ID
?	any	ガン:4, 心臓病:4	user1, 3,5,7,8,10,11,12

【図３１】

図31

年齢	身長	病気	ID
any	?	?	user1-12

【図３２】

図32

年齢	身長	病気	ID
?	-170	ガン:2, 心臓病:2	user1, 3,5,7,
?	170-	ガン:2, 心臓病:2	user8,10,11,12

【図３３】

図33

年齢	身長	病気	ID
any	?	?	user1, 3,5,7,
any	?	?	user8,10,11,12

【図３４】

図34

年齢	身長	病気	ID
-30	?	?	user1, 3
30-	?	?	user5,7
30-	?	?	user8,10,11,12

【図３５】

図35

年齢	身長	病気	ID
?	-170	ガン:1, 心臓病1	user1, 3
?	-170	ガン:1, 心臓病:1	user5,7
?	170-	ガン:2, 心臓病:2	user8,10,11,12

【図３６】

図36

年齢	身長	病気	ID
-30	?	?	user1, 3
30-40	?	?	user5,7
30-40	?	?	user8,10
40-	?	?	user11,12

【図３７】

図37

年齢	身長	病気	ID
?	-170	ガン:1, 心臓病1	user1, 3
?	-170	ガン:1, 心臓病:1	user5,7
?	170-	ガン:1, 心臓病:1	user8,10
?	170-	ガン:1, 心臓病:1	user11,12

【図３８】

図38

年齢	身長	病気
-30	-170	心臓病
-30	-170	ガン
30-40	-170	心臓病
30-40	-170	ガン
30-40	170-	心臓病
30-40	170-	ガン
40-	170-	心臓病
40-	170-	ガン

フロントページの続き

(56)参考文献 特開２００６－０５３７１１（ＪＰ，Ａ）
特開２００７－２６４７３０（ＪＰ，Ａ）
米国特許出願公開第２０１０／０１１４９２０（ＵＳ，Ａ１）
村本 俊祐ほか，背景知識を用いた推測を困難にしデータ歪曲度を極小化するプライバシー保護手法，電子情報通信学会 第19回データ工学ワークショップ（DEWS 2008）論文集，２００８年 4月 7日，pp. 1-8，DEWS2008 C1-4
千田 浩司ほか，集合匿名化クラウドの課題と対策，電子情報通信学会技術研究報告，２０１１年 5月 5日，Vol. 111，No. 30，pp. 117-122，Vol.2011-CSEC-53 No.21/Vol.2011-IOT-13 No.21

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ ２１／６２