(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5771158号
(P5771158)
(45)発行日　平成27年8月26日(2015.8.26)　(24)登録日　平成27年7月3日(2015.7.3)

(51)Int.Cl.　　F I
G06F 21/62 (2013.01)　　G06F 21/62 354
H04L 9/14 (2006.01)　　H04L 9/00 641

請求項の数 5　(全 24 頁)

(21)出願番号　特願2012-10211(P2012-10211)
(22)出願日　平成24年1月20日(2012.1.20)
(65)公開番号　特開2012-165374(P2012-165374A)
(43)公開日　平成24年8月30日(2012.8.30)
審査請求日　平成27年1月19日(2015.1.19)
(31)優先権主張番号　13/021,538
(32)優先日　平成23年2月4日(2011.2.4)
(33)優先権主張国　米国(US)

早期審査対象出願

(73)特許権者　502096543
パロ・アルト・リサーチ・センター・インコーポレーテッド
Ｐａｌｏ　Ａｌｔｏ　Ｒｅｓｅａｒｃｈ　Ｃｅｎｔｅｒ　Ｉｎｃｏｒｐｏｒａｔｅｄ
アメリカ合衆国、カリフォルニア州　９４３０４、パロ・アルト、コヨーテ・ヒル・ロード　３３３３
(74)代理人　100079049
弁理士　中島　淳
(74)代理人　100084995
弁理士　加藤　和詳

最終頁に続く

(54)【発明の名称】時系列データのプライバシー保護アグリゲーション

(57)【特許請求の範囲】
【請求項１】
データアグリゲータと機密データを共有する方法において、
ユーザの組及び前記データアグリゲータと関連付けられた秘密鍵の合計がゼロに等しい、前記ユーザの組におけるローカルユーザのための前記秘密鍵を受信することと、
クライアントコンピュータにおいて、前記ローカルユーザと関連付けられたデータ値の組を選択することと、
暗号化データ値の組を生成するように、一部において前記秘密鍵に基づいて前記データ値の組における個々のデータ値を暗号化し、それにより、前記データアグリゲータが、前記ユーザの組と関連付けられた個々のデータ値を復号することなく且つアグリゲート値を復号しながら前記ユーザの組と情報授受することなく、前記ユーザの組にわたって前記アグリゲート値を復号するのを許容することと、
前記データアグリゲータに対して前記暗号化データ値の組を送信することと、
を備え、
前記データ値の組は、時系列を含み、
前記時系列において前記個々のデータ値を暗号化することに先立って、前記秘密鍵は受信され、
前記アグリゲート値は、前記ユーザの組と関連付けられた前記個々のデータ値の合計を含み、
ユーザｉ及び期間ｔのための個々のデータ値$x_{i,t}$を暗号化することは、式

【数１】

$$c_{i,t}=g^{xi,t.}H(t)^{sk_i}$$

を計算することを含み、
ｃ$_{i,t}$は、前記ユーザｉ及び前記期間ｔと関連付けられた前記暗号化データ値であり、ｇは、生成元であり、ｓｋ$_i$は、前記ユーザｉと関連付けられた前記秘密鍵であり、Ｈ（ｔ）は、ハッシュ関数である
方法。
【請求項２】
前記ローカルユーザのための前記秘密鍵を受信することが、信頼できるソースから前記秘密鍵を受信することを備える、請求項１に記載の方法。
【請求項３】
前記ローカルユーザのための前記秘密鍵を受信することが、安全なマルチパーティプロトコルを使用することを備える、請求項１に記載の方法。
【請求項４】
前記個々のデータ値を暗号化することが、ランダムノイズによって変更されたデータ値の組を生成するように、少なくとも前記データ値の部分集合に対してランダム値を加算することを含む、請求項１に記載の方法。
【請求項５】
前記ランダム値は、前記ユーザの組によって前記アグリゲート値に導入された総ノイズを、
【数２】

$$o\left(\frac{\Delta}{\varepsilon}\sqrt{n}\right)$$

に最小にするために予め判定された分布から選択され、
Δは、前記アグリゲート値の機密度であり、ｎは、前記ユーザの数である
請求項４に記載の方法。

【発明の詳細な説明】
【発明の概要】
【０００１】
システムは、ユーザのプライバシーを漏洩することなく、データアグリゲータに対してユーザのデータを提供する。システムは、ユーザの組におけるローカルユーザのための秘密鍵を判定する。ここで、ユーザの組とデータアグリゲータとに関連付けられた秘密鍵の合計はゼロに等しい。システムはまた、ローカルユーザと関連付けられたデータ値の組を

選択する。そして、システムは、暗号化データ値の組を生成するように、一部において秘密鍵に基づいて組における個々のデータ値を暗号化し、それにより、ユーザの組と関連付けられた個々のデータ値を復号することなく且つアグリゲート値を復号しながらユーザの組と情報授受することなく、データアグリゲータがユーザの組にわたってアグリゲート値を復号するのを許容する。システムはまた、データアグリゲータに対して暗号化データ値の組を送信する。

【０００２】

いくつかの実施形態において、システムは、信頼できるソースから秘密鍵を受信することによってローカルユーザのための秘密鍵を判定する。

【０００３】

いくつかの実施形態において、システムは、安全なマルチパーティプロトコルを使用することによってローカルユーザのための秘密鍵を判定する。

【０００４】

いくつかの実施形態において、データ値の組は、データ値の時系列を含む。さらに、システムは、時系列における個々のデータ値の暗号化に先立って秘密鍵を判定してもよい。

【０００５】

いくつかの実施形態において、個々のデータ値を暗号化することは、ランダムノイズによって変更されたデータ値の組を生成するように、少なくともデータ値の部分集合に対してランダム値を加算することを含む。ここで、Δはアグリゲート値の機密度であり、ｎはユーザ数である。

【０００６】

いくつかの実施形態において、個々のデータ値を暗号化することは、個々のデータ値の高次モーメントも暗号化し、それにより、データアグリゲータがユーザの組にわたるデータ値の分布を判定するのを許容することを含む。

【０００７】

他の実施形態において、システムは、ユーザの組と関連付けられたデータ値の組についてのアグリゲート値を計算する。システムは、データアグリゲータのための秘密鍵を判定する。ここで、ユーザの組とデータアグリゲータとに関連付けられた秘密鍵の合計はゼロに等しい。システムはまた、対応するユーザの組から暗号化データ値の組を受信する。そして、システムは、ユーザの組と関連付けられた個々のデータ値を復号することなく、ユーザの組にわたってアグリゲート値を判定するために秘密鍵を使用し、データアグリゲータに対してアグリゲート値を提供する。

【０００８】

いくつかの実施形態において、アグリゲート値は、ユーザの組と関連付けられた個々のデータ値の積を含む。

【図面の簡単な説明】

【０００９】

【図１】図１は、実施形態にかかるプライベートストリームアグリゲーションについての典型的な計算環境を示している。

【図２】図２は、実施形態にかかる信頼できないデータアグリゲータと安全にデータを共有するためのプロセスを示すフローチャートを表している。

【図３】図３は、実施形態にかかる参加者の組によって提供される暗号化データからアグリゲート値を判定するためのデータアグリゲータによって実行されるプロセスを示すフローチャートを表している。

【図４】図４は、実施形態にかかる参加者の組とデータアグリゲータとの間における情報の典型的な流れを示す概略図を表している。

【図５】図５は、実施形態にかかるプライバシー保護データアグリゲーションを容易とするコンピューターシステムを示している。

【図６】図６は、実施形態にかかるプライバシー保護データアグリゲーションを容易とする装置を示している。　表１は、実施形態にかかる分散型の差分プライバシーデータラン

ダム化を実行するアルゴリズムを表している。　図において、同様の参照符号は、同一の図面要素を参照している。

【発明を実施するための形態】

【００１０】

本発明の実施形態は、ユーザが信頼できないアグリゲータに対して暗号化データのストリームをアップロードするのを許容し且つアグリゲータが各時間間隔のアグリゲート統計値を復号するために秘密鍵を使用するのを許容するプライベートストリームアグリゲーション（ＰＳＡ）システムを提供する。ＰＳＡ技術は、アグリゲータが所望の統計値及びその予備知識から推定することができるもの以外の任意の意図されたものでない情報を学習することができないことを意味する記憶しないアグリゲータである。ＰＳＡ技術は、アグリゲータに公開される統計値が特定の個人が参加するか否かにあまり影響を受けない個々の参加者についての分散型の差分プライバシーを保証する。

【００１１】

信頼できないデータアグリゲータは、個々のデータ点に対してアクセスすることなく、時系列データについてのアグリゲート統計値（例えば参加者のデータの合計又は積）を計算することができる。例えば、データアグリゲータは、毎週、ｎ個の企業の総売上高の把握をすることを求めてもよい。本発明の実施形態によって提供されるプライバシーモデルは、個々の企業それぞれがデータアグリゲータに対して毎週それらの収入のノイズのある暗号化をアップロードするのを許容する。さらに、データアグリゲータは、参加企業の収入のノイズのある合計を復号することができるが、個々の企業と関連付けられた毎週の収入データを取得することはできない。それゆえに、データアグリゲータは、個々の企業に特有の付加情報を推測することができない。

【００１２】

図１は、実施形態にかかるプライベートストリームアグリゲーションについての典型的な計算環境を示している。具体的には、計算環境１００は、コンピュータネットワーク１０２と、データアグリゲータ１０４と、参加者１０６、１０８及び１１０とを含むことができる。いくつかの実施形態において、計算環境１００はまた、計算環境１００についての信頼できるセットアップを実行する信頼できるエンティティ１１２を含むこともできる。アグリゲータ１０４、参加者１０６～１１０、及び、信頼できるエンティティ１１２は、サーバコンピュータ、デスクトップ、ラップトップ、携帯型の計算装置、又は、無線センサ等のコンピュータネットワーク１０２と連結された計算装置を含むことができる。さらに、コンピュータネットワーク１０２は、有線ネットワーク、無線ネットワーク、又は、それらの組み合わせを含むことができる。

【００１３】

計算システム１００は、セットアップ、ノイズのある暗号化、及び、アグリゲートの復号という少なくとも３つのキー動作を経験する。初期セットアップ動作中において、信頼できるエンティティ１１２は、参加者１０６～１１０のそれぞれに対して秘密鍵を割り当てることができ、アグリゲータ１０４に復号能力を提供するアグリゲータ１０４に対して秘密鍵を割り当てることができる。具体的には、アグリゲータ１０４及び参加者１０６～１１０と関連付けられた秘密鍵についての値の合計は、ゼロに等しい。この特徴は、アグリゲータの秘密鍵のみがアグリゲート値を復号（例えば個々の参加者から受信したデータ値の合計を復号）することができるように、参加者１０６～１１０がそれらのデータを暗号化するのを許容する。

【００１４】

アグリゲータ１０４及び参加者１０６～１１０は、安全なマルチパーティプロトコルを使用して初期セットアップ動作を実行することができ、それにより、信頼できるエンティティ１１２の助けを必要としない。例えば、アグリゲータ１０４及び参加者１０６～１１０は、それらの秘密鍵を公開する必要なく、その合計がゼロに等しい秘密鍵の組を判定するために互いにネゴシエートすることができる。

【００１５】

ノイズのある暗号化動作中において、参加者（例えば参加者１０６）は、その秘密鍵を使用してデータ値を暗号化することができる。いくつかの実施形態において、参加者は、データ値を暗号化する前にノイズをデータ値に加えることができる。例えば、参加者が一定時間にわたって時系列データ値のデータ組を暗号化するとき、暗号化データ値の部分集合は、ランダムノイズを含むことができる一方で、暗号化データ値の残りの部分集合がランダムノイズを含まない。

【００１６】

さらに、アグリゲート暗号化動作中において、アグリゲータ１０４は、一部においてその秘密鍵に基づいて、且つ、アグリゲータ１０４が参加者１０６～１１０から受信した暗号化値の組に基づいて、アグリゲート値を判定する。ここで留意すべきは、アグリゲータ１０４が参加者１０６～１１０のいずれかについての秘密鍵を知らないことから、アグリゲータ１０４は、参加者１０６～１１０のいずれかから取得した個々のデータ値を復号することができないということである。しかしながら、本発明の実施形態によって提供されるプライベートストリームアグリゲーション技術は、アグリゲータ１０４が暗号化データ値の組についてのアグリゲート値を判定するために、その秘密鍵を使用するのを許容する。

【００１７】

以下の段落は、計算環境１００と関連付けられたプライベートストリームアグリゲーションシステムについての様々な典型的な用途を提供する。典型的なそれぞれの用途において、プライベートストリームアグリゲーション技術は、参加者（例えば個々のユーザ又は機関）が、それらの個々のデータ値をデータアグリゲータに対して公開する必要なく、データアグリゲータに対してそれらの個々のデータを提供するのを許容する。

【００１８】

センサネットワークは、建物の安全を監視したり、交通量を測定したり、又は、環境汚染物質を追跡したりする等のために、様々な用途のために一般に配備されている。典型的な用途において、配備されたセンサノードは、それらの読み取り値を、パターンを識別するか又は統計値を判定するためにデータをマイニングする中央ステーションに対して周期的に送信する。多くの状況において、特にセンサが複数の参加機関にわたって配備されている場合には、個々のセンサによって作られる読み取り値は、プライバシーセンシティブであってもよい。本発明の実施形態によって提供されるプライベートストリームアグリゲーション技術は、これらの参加企業に中央ステーション（例えばアグリゲータ１０４）がいかなる特定のセンサノード（例えば参加者１０６～１１０）からも詳細情報を取得しないというプライバシー保証を提供する。それゆえに、このプライバシー保証は、参加企業がその機関にわたってセンサノードを配備することによって重要な研究プロジェクトに貢献するのを促進することができる。

【００１９】

他の例は、電気に関する「スマートグリッド」及び「スマートメータリング」技術の出現である。スマートメータは、月に１回とは対照的に１５分毎に電気、ガス、水道使用量を読み取ること等によって従来の計器よりも非常に細かい精度で電気、ガス、水道使用量を読み取る。この電気、ガス、水道のきめの細かいサンプリングは、公共企業に機密情報を推定するのに十分な詳細情報を提供することができる。例えば、公共企業は、特定の家庭用器具の使用量とともに、家庭における個人数及びそれらの睡眠／作業習慣を推定することができてもよい。本発明の実施形態によって提供されるプライベートストリームアグリゲーション技術は、公共企業（例えば図１におけるアグリゲータ１０４）に、いかなる特定の家庭からも実際のきめの細かい電気、ガス、水道使用量を公開することなく、家庭の組（例えば参加者１０６～１１０）にわたる電気、ガス、水道使用量のきめの細かいサンプリングを取得するための能力を提供することができる。これは、家庭がスマートグリッド技術に対して持つ可能性がある懸念及び心配を軽減することができ、アグリゲート統計値は、スマートグリッドオペレータがそれらの監視労力及び価格最適化を十分達成できるほどであろう。

【００２０】
医学研究は、非常に医学データの利益を享受するが、プライバシーの懸念は、このデータが収集されて広められる範囲を制限する。本発明の実施形態によって提供されるプライベートストリームアグリゲーション技術は、介護人又は遠隔測定装置によって絶えずアップロードされるデータから、個人又は機関の群に及ぶ高レベルの統計値及び高レベルの統計値のみを研究者が取得するのを可能とする。
【００２１】
多くの研究プロジェクト又はソフトウェア技術は、個人にわたるプライバシー懸念に拍車をかける可能性のある、ある程度の母集団調査、感知、及び、監視を実行する。例えば、ある企業のソフトウェアは、参加者の利用可能性を推定し、同僚が任意の時点でその参加者との通信のために最高の手段を識別するのを助けるために、カメラ、ｗｉｆｉ及びコンピュータ処理から取得した参加者のデータを使用することができる。しかしながら、参加者の作業習慣についての詳細情報は、この利用可能性情報から推定されることができる。したがって、何人かのユーザは、それらの利用可能性におけるこの情報が企業管理人によって悪用され得ることを恐れて参加するのを嫌がることがある。本発明の実施形態によって提供されるプライベートストリームアグリゲーション技術は、企業が参加者の群にわたる統計情報を取得するのを可能とする一方で、参加者が選択された個人と詳細な利用可能性情報を共有するのみとするのを許容する。
【００２２】
クラウドコンピューティングが人気を得るのにともない、個人及び機関は、第三者のクラウドサービスにおいてデータを保存している量が増加している。クラウドサービスプロバイダーは、様々な社会的及び経済的な目標を実現するためにこのデータにおける有益な統計値を計算することを望む。しかしながら、参加企業は、クラウドサービスのより多くの使用を行わない理由のトップとして、それらのデータのセキュリティ及びプライバシーについての懸念を挙げている。本発明の実施形態によって提供されるプライベートストリームアグリゲーション技術は、クラウドサービスプロバイダーが、個々の参加者から機密情報を取得することなく、経時的に複数の参加者からあるアグリゲート統計値を追跡するのを可能とする。
【００２３】
図２は、信頼できないデータアグリゲータと安全にデータを共有するためのプロセスを示すフローチャートを表している。
【００２４】
システムは、参加者の組及びデータアグリゲータと関連付けられた秘密鍵の合計がゼロに等しいように、参加者の組におけるローカル参加者のための秘密鍵を判定することによって開始することができる（工程２０２）。例えば、工程２０２は、信頼できるソースから秘密鍵を受信すること、又は、データアグリゲータ及び参加者の組のための秘密鍵を判定するために安全なマルチパーティプロトコルを使用することを含むことができる。
【００２５】
次に、システムは、データアグリゲータと共有するためにデータ値の組を選択することができ（工程２０４）、ランダムノイズによって変更されたデータ値の組を生成するように少なくともデータ値の部分集合に対してランダム値を加えることができる（工程２０６、任意）。ここで留意すべきは、データ値が時系列を含むことができるということである。そして、システムは、組における個々のデータ値を暗号化し（工程２０８）、データアグリゲータに対して暗号化データ値を送信する（工程２１０）。
【００２６】
アグリゲート値は、特定の期間についての参加者の組と関連付けられた個々の値の合計を含むことができる。この場合において、参加者ｉ及び期間ｔについての個々のデータ値$x_{i,t}$を暗号化することは、以下の式を計算することを含む。

【数１】

$$c_{i,t} = g^{x_{i,t}} \cdot H(t)^{sk_i}$$

具体的には、ｃ$_{ｉ，ｔ}$は、参加者ｉ及び期間ｔと関連付けられた暗号化値であり、ｇは、生成元であり、ｓｋ$_{ｉ}$は、参加者ｉと関連付けられた秘密鍵であり、Ｈ（ｔ）は、ハッシュ関数である。

【００２７】

アグリゲート値は、参加者の組と関連付けられた個々の値の積を含む。この場合において、参加者ｉ及び期間ｔについての個々のデータ値ｘ$_{ｉ，ｔ}$を暗号化することは、以下の式を計算することを含む。

【数２】

$$c_{i,t} = x_{i,t} \cdot H(t)^{sk_i}$$

【００２８】

図３は、参加者の組によって提供される暗号化データからアグリゲート値を判定するためのデータアグリゲータによって実行されるプロセスを示すフローチャートを表している。

【００２９】

システムは、参加者の組及びデータアグリゲータと関連付けられた秘密鍵の合計がゼロに等しいように、データアグリゲータのための秘密鍵を判定することによって開始することができる（工程３０２）。例えば、工程３０２は、信頼できるソースから秘密鍵を受信することを含むことができるか、又は、データアグリゲータ及び参加者の組のための秘密鍵を判定するために安全なマルチパーティプロトコルを使用することを含むことができる。

【００３０】

次に、システムは、対応する参加者の組から暗号化データ値の組を受信することができる（工程３０４）。システムは、暗号化データ値の組についてのアグリゲート値を判定するために、データアグリゲータと関連付けられた秘密鍵を使用する（工程３０６）。そして、システムは、データアグリゲータに対してアグリゲート値を提供する（工程３０８）。

【００３１】

アグリゲート値を推定することは、以下の式を計算することを含む。

【数３】

$$V = H(t)^{sk_0} \prod_{i=1}^{n} c_{i,t}$$

具体的には、ｃ$_{ｉ，ｔ}$は、参加者ｉ及び期間ｔと関連付けられた暗号化値であり、ｎは、参加者の総数であり、ｓｋ$_{0}$は、データアグリゲータと関連付けられた秘密鍵であり、Ｈ（ｔ）は、ハッシュ関数である。

【００３２】

アグリゲート値は、アグリゲート値を判定することがさらにＶの離散対数を計算することを含むように、参加者の組と関連付けられた個々のデータ値の合計を含む。いくつかの実施形態において、アグリゲート値は、参加者の組と関連付けられた個々のデータ値の積を含む。

【００３３】

汎用の用途において、１つのデータアグリゲータ及びｎ人の参加者がいてもよい。表記

の便宜のために、参加者は、１，・・・，ｎに番号付けられ、データアグリゲータは、０に番号付けられる。例えば、秘密鍵$sk_{i\neq 0}$は、参加者と関連付けられ、秘密鍵sk_0は、データアグリゲータと関連付けられる。さらに、［ｎ］：＝｛１，２，・・・，ｎ｝及びＤを、参加者のデータについての許容値の所定領域を表すものとする。それゆえに、期間ｔ∈Ｎについて、参加者ｉ∈［ｎ］と関連付けられたデータは、値$x_{i,t}\in D$を有する。

【００３４】

説明を簡単にするために、参加者のデータ値についての表記は、必ずしも下付き文字ｔを含まなくてもよく、ｉ∈［ｎ］についてのデータ値x_iの組は、共通の期間と対応すると考えられることができる。それゆえに、$x=(x_1,\cdots,x_n)\in D^n$を、ある一定期間における全ての参加者からの値のベクトルを意味するものとする。さらに、アグリゲータは、範囲Ｏに属している所望の統計値を計算するために関数ｆ（ｘ）を使用する。それゆえに、アグリゲータは、関数$f:D^n\rightarrow O$によって表されるアグリゲート統計値を計算する。

【００３５】

各参加者は、アグリゲータがユーザの入力について任意の補足情報を有し得るときでさえも、強いプライバシー保証を達成するために独立したランダムノイズを発生させることができる。具体的には、各参加者は、標本空間Ωからのノイズr_iが$r:=(r_1,\cdots,r_n)\in\Omega^n$によって表されるように、他の参加者から独立したランダムノイズr_iを発生させることができる。表記目的のために、ハット付きの変数は、（例えばランダム値ｒ及びランダム関数χと関連付けられた）参加者のデータのランダム化されたバージョンを意味し、ハット付きでない変数は、元の参加者データを意味する。それゆえに、$\chi:D\times\Omega\rightarrow D$を、アグリゲータに対して値を暗号化してアップロードする前に、データのノイズのあるバージョン$x_i:=\chi(x_i,r_i)$を計算するために参加者によって使用されることができるランダム化関数を意味するものとする。そして、アグリゲータは、所望の統計値ｆ（ｘ）に導入されるノイズが所定レベルの範囲内であるように、$x^\wedge:=(x_1^\wedge,x_2^\wedge,\cdots,x_n^\wedge)$の暗号化値からノイズのある統計値ｆ（ｘ^）を計算する。

【００３６】

参加者ｉは、その参加者にとっての習慣である所望のデータ分布にしたがってノイズr_iを発生させることができる。さらに、各参加者ｉは、データx_iに対して異なるランダム化関数$\chi_i(x_i,r_i)$を適用することができる。さらにより一般的なシナリオにおいて、各参加者は、暗号化された入力におけるランダム化されたアグリゲート関数$f:D^n\times\Omega^n\rightarrow O$を計算することができるアグリゲータに対してデータ$x_i$及びランダム度$r_i$を送信する前に、別個に$x_i$及び$r_i$を暗号化してもよい。しかしながら、説明を簡単にするために、以下の段落は、$f^\wedge(x,r)=f(x^\wedge)$であるときの特別な場合を考える。さらにまた、以下の段落は、データを暗号化する前に参加者が同じランダム関数χを適用するシナリオをカバーする。

【００３７】

本発明の実施形態によって提供されるプライベートストリームアグリゲーション（ＰＳＡ）機構は、Ｓｅｔｕｐ（）、ＮｏｉｓｙＥｎｃ（）及びＡｇｇｒＤｅｃ（）といういくつかのキー関数から構成される。

【００３８】

Ｓｅｔｕｐ（1^λ）：Ｓｅｔｕｐ（）関数は、セキュリティパラメータλを取り入れ、各期間においてアグリゲート統計値を復号するためにアグリゲータによって使用されるアグリゲータsk_0についての秘密鍵とともに、公開パラメータｐａｒａｍ、各参加者についての秘密鍵sk_iを出力する。各参加者ｉは、秘密鍵sk_iを取得し、データアグリゲータは、能力sk_0を取得する。信頼できるセットアップ段階の後、さらなる相互通信は、データアグリゲータに対して暗号化データ値をアップロードすることを除いて、参加者とデータアグリゲータとの間で必要とされない。

【００３９】
ＮｏｉｓｙＥｎｃ（ｐａｒａｍ，ｓｋ$_{i}$，ｔ，ｘ，ｒ）：時間ステップｔの間、各参加者は、ノイズｒでそのデータｘを符号化するためにＮｏｉｓｙＥｎｃ関数を呼び出す。結果は、ノイズｒによってランダム化されたデータ点ｘのノイズのある暗号化である。ＮｏｉｓｙＥｎｃ関数は、時々ＮｏｉｓｙＥｎｃ（ｐａｒａｍ，ｓｋ$_{i}$，ｔ，ｘ＾）として書かれる。ここで、ｘ＾：＝χ（ｘ，ｒ）は、参加者のデータのノイズのあるバージョンであり、χは、基礎となるランダム化関数である。
【００４０】
ＡｇｇｒＤｅｃ（ｐａｒａｍ，ｓｋ$_{0}$，ｔ，ｃ$_{1}$，ｃ$_{2}$，・・・，ｃ$_{n}$）：復号アルゴリズムは、同じ期間ｔについての公開パラメータｐａｒａｍ、能力ｓｋ$_{0}$、及び、暗号文ｃ$_{1}$，ｃ$_{2}$，・・・，ｃ$_{n}$を取り入れる。各ｉ∈［ｎ］について、ｃ$_{i}$＝ＮｏｉｓｙＥｎｃ（ｓｋ$_{i}$，ｔ，ｘ$_{i}$＾）とする。ここで、それぞれ、ｘ$_{i}$＾：＝χ（ｘ$_{i}$，ｒ$_{i}$）である。ｘ：＝（ｘ$_{1}$，・・・，ｘ$_{n}$）及びｘ＾：＝（ｘ$_{1}$＾，・・・，ｘ$_{n}$＾）とする。復号アルゴリズムは、ターゲットとされた統計値ｆ（ｘ）のノイズのあるバージョンであるｆ（ｘ＾）を出力する。
【００４１】
いくつかの実施形態において、アグリゲート関数ｆ（ｘ＾）は、参加者の組についてのデータ値のノイズのある合計を生成する。この文脈において、参加者のデータｘ$_{i}$は、素数ｐについてのＺ$_{p}$に属し、アグリゲート関数は、以下として定義される。
【数４】

$$sum(\hat{\mathbf{x}}) := \sum_{i=1}^{n} \hat{x}_i$$

さらに、各参加者は、整数の組からノイズｒ$_{i}$を発生させることができ、ランダム化関数χ（ｘ$_{i}$，ｒ$_{i}$）：＝ｘ$_{i}$＋ｒ$_{i}$ｍｏｄｐを適用することができる（すなわち、参加者はデータを暗号化する前に付加的ノイズを組み入れる）。
【００４２】
同様に、他のいくつかの実施形態において、アグリゲート関数ｆ（ｘ＾）は、参加者の組についてのデータ値のノイズのある積を生成する。この文脈において、アグリゲート関数は、以下として定義される。
【数５】

$$product(\hat{\mathbf{x}}) := \prod_{i=1}^{n} \hat{x}_i$$

【００４３】
図４は、本発明の実施形態にかかる参加者の組とデータアグリゲータとの間における情報の典型的な流れを示す概略図を表している。具体的には、参加者４０２～４０６は、参加者が対応する暗号化データ値ｃ$_{i}$を生成するために秘密鍵ｓｋ$_{i}$を使用する前にノイズｒ$_{i}$をデータ値ｘ$_{i}$に加えることができるように、対応する暗号化データ値を生成することができる。さらに、アグリゲータ４０８は、参加者４０２～４０６から暗号化データ値を受信し、アグリゲート値４１０を生成するためにその秘密鍵ｓｋ$_{0}$を使用する。
【００４４】
アグリゲータ不記憶要件は、アグリゲータが、公開された統計値ｆ（ｘ＾）から推論されることができること及び既に知っているいかなる補足データ以外の何も学習しないという保証を提供する。さらに、この要件を達成することは、適切なアグリゲータ能力（例えばアグリゲータの秘密鍵ｓｋ$_{0}$）がない関係者が何も学習しないということを保証する。
【００４５】

直観的に、アグリゲータ不記憶要件は、以下のセキュリティ概念を満たす。
・アグリゲータは、各期間についてのノイズのある合計のみを学習することができ、それ以上は何も学習することができない。例えば、アグリゲータは、全ての参加者の暗号文の真部分集合からいかなる部分的な情報も学習することができない。
・アグリゲータの秘密鍵を知ることなく、数人の参加者が残りのユーザに対して連携を形成する場合であっても、敵対者は、暗号化データについて何も学習しない。
・アグリゲータが参加者の部分集合と共謀した場合、又は、暗号化データの部分集合が漏洩された場合、アグリゲータは、必然的に残りの参加者の合計を学習することができる。しかしながら、アグリゲータは、残りの参加者についての個々のデータ値を学習することができない。

【００４６】

アグリゲータ不記憶セキュリティの概念は、合計以外の一般的な統計値まで広げられてもよい。しかしながら、敵対者が成功裏の攻撃から取得することができる情報を制限するのに格別の注意が払われなければならない。例えば、敵対者が参加者の組Ｋ⊆［ｎ］を漏洩した場合、敵対者は、その後、これらの参加者に代わって何でも暗号化することが可能である。したがって、敵対者は、漏洩された参加者Ｋの組についてのいかなる所望の平文ベクトルx_K＝｛x_i｜ｉ∈Ｋ｝も入力することができ、平文ベクトルを暗号化することができ、そして、x_Kに基づいてアグリゲート統計値を復号するためにＡｇｇｒＤｅｃ関数を呼び出すことができる。それゆえに、セキュリティ定義は、これが敵対者についての最良且つ唯一の戦略であるという事実を反映しなければならない（すなわち、敵対者は、この攻撃から収集される情報以外の他の情報を学習することができない）。この要件は、以下の段落によって定義される。

【００４７】

定義１（アグリゲータ不記憶セキュリティ）：挑戦者によってどの平文ベクトル（すなわち、ｘ又はｘ’）が暗号化されているかについて区別する際の無視できる程度の利点以上を有する確率的多項式時間敵対者がない場合、ＰＳＡは、記憶しないアグリゲータである。２つの平文ベクトルがそれらのアグリゲート値に対して同等であるとき、この要件は、通常満たされる。合計統計値について、この要件は、以下として定義される。

【数６】

$$\sum_{i\in U} x_i = \sum_{i\in U} \hat{x}'_i$$

この状態は、他の統計値と関連付けられた一般的なクエリーを満たすのがより困難であり得る。

【００４８】

定義２（１回暗号化セキュリティ）：それぞれの正当な参加者は、期間につき１回のみ暗号化すると期待されている。ＰＳＡは、
１）上述したセキュリティゲームにおいて無視できる程度の利点以上を有する確率的多項式時間敵対者がない。
２）以下の制約が∀ｉ∈Ｕ、∀（ｘ，ｒ）∈Ｄ×Ωを保持する場合、「１回暗号化」モデルにおいて記憶しないアグリゲータであるといわれる。ここで、タプル（ｉ，ｔ＊，ｘ，ｒ）は、いかなる暗号化クエリーにおいても現れない。

【００４９】

以下の段落は、アグリゲータ不記憶セキュリティを達成するための暗号化構築を記述する。Ｇを、ディフィー・ヘルマン判定法が困難である素数の位数ｐの巡回群を意味するものとする。さらに、Ｈ：Ｚ→Ｇを、整数を数学的群Ｇに写像する（ランダムオラクルとしてモデル化された）ハッシュ関数を意味するものとする。

【００５０】

Ｓｅｔｕｐ（1^{λ}）：信頼できるセットアップ段階の間、信頼できるディーラーは、以

下であるようにランダム発生器ｇ∈Ｇ及びｎ＋１個のランダム秘密ｓｋ$_{i}$∈Ｚ$_{p}$を選択する。

【数７】

$$\sum_{i=0}^{n} sk_i = 0$$

ここで留意すべきは、公開パラメータは、ｐａｒａｍ：＝ｇとして初期化されるということである。さらに、データアグリゲータは、能力ｓｋ$_{0}$を取得し、参加者ｉは、秘密鍵ｓｋ$_{i}$を取得する。

【００５１】

それゆえに、期間ｔの間、各参加者は、ｉ∈［ｎ］についてＲ$_{i,t}$＝Ｈ（ｔ）^{sk_i}を計算し、アグリゲータは、Ｒ$_{0,t}$＝Ｈ（ｔ）^{sk_0}を計算する。ここで留意すべきは、ｓｋ$_{i}$についての合計がゼロに等しいことから、それは、以下にしたがうということである。

【数８】

$$\prod_{i=0}^{n} R_{i,t} = 1$$

この特性は、参加者が信頼できるセットアップ段階の後に互いに通信するのを必要とすることなく、ＮｏｉｓｙＥｎｃ（）及びＡｇｇｒＤｅｃ（）演算が独立して機能するのを許容する。さらに、ディフィー・ヘルマン判定法が巡回群Ｇについて困難であると仮定すると、それは、数Ｒ$_{i,t}$が外見的にはランダムオラクルモデルの下でランダムであるということになる。

【００５２】

ＮｏｉｓｙＥｎｃ（ｐａｒａｍ，ｓｋ$_{i}$，ｔ，ｘ＾）：参加者ｉが時間ステップｔについての値ｘ＾∈Ｚ$_{p}$を暗号化するため、参加者は、以下の暗号文を計算する。

【数９】

$$c \leftarrow g^{\hat{\mathbf{x}}} \cdot H(t)^{sk_i}$$

各参加者がデータ値を暗号化する前にデータ値に対してノイズを加えるとみなされると仮定すると、ランダム化された平文値は、用語ｘ＾＝ｘ＋ｒｍｏｄｐによって表される。

【００５３】

ＡｇｇｒＤｅｃ（ｐａｒａｍ，ｓｋ$_{0}$，ｔ，ｃ$_{1}$，ｃ$_{2}$，・・・，ｃ$_{n}$）：アグリゲータは、以下のようにアグリゲート値Ｖを計算する。

【数１０】

$$V = H(t)^{sk_0} \prod_{i=1}^{n} c_i .$$

その際、ここで留意すべきは、ｉ∈［ｎ］についてｃ$_{i}$＝ＮｏｉｓｙＥｎｃ（ｐａｒａｍ，ｓｋ$_{0}$，ｔ，ｘ$_{i}$＾）であるということである。さらに、アグリゲータ及び参加者の組についてのｓｋ$_{i}$がゼロになることから、以下である。

【数１１】

$$\prod_{i=0}^{n} H(t)^{sk_i} = 1$$

したがって、それは、Ｖが以下の形式からなることになる。

【数１２】

$$V = g^{\sum_{i=1}^{n} \hat{x}_i}$$

【００５４】

以下の合計を復号するために、アグリゲータは、ｇを底とするＶの離散対数を計算する。

【数１３】

$$\sum_{i=1}^{n} \hat{x}_i$$

平文空間が小さいとき、復号は、総当たり調査によって達成されることができる。より良好なアプローチは、おおよそ平文空間における平方根の復号時間を必要とするポラードのラムダ法を使用することである。例えば、各参加者の入力が範囲｛０，１，・・・，Δ｝にあると仮定する。そして、参加者の合計は、範囲｛０，１，・・・，ｎΔ｝に含まれる。この場合において、復号は、ポラードの方法を使用して√（ｎΔ）時間を必要とする。換言すれば、ｎΔが多項式時間において成功裏の復号を確実にするためにセキュリティパラメータλにおける多項式であることが必要とされる。ここで留意すべきは、付加準同形の暗号化スキーム（例えば、エルガマル暗号化及びＢＧＮ　準同形暗号化）として使用されるとき、小さい平文空間の制限は、ディフィー・ヘルマンベースの暗号化スキームに特有であるということである。

【００５５】

定理１：ディフィー・ヘルマン判定法の問題が群Ｇにおいて困難であり、ハッシュ関数Ｈがランダムオラクルであると仮定すると、上述した段落において提供された構築は、「１回暗号化」モデルにおいてアグリゲータ不記憶セキュリティを満たす。

【００５６】

提案された暗号構築において、暗号化は、１つのディフィー・ヘルマン群における少なくとも１つのハッシュ演算（例えばＳＨＡ－２５６）、２つのモジュラー指数、及び、乗算を含む。実行時間は、ハッシュ関数及び群乗算を計算するための時間が指数についての時間と比較して非常に小さいことから、２つのモジュラー指数によって支配される。ｅＢＡＣＳプロジェクトによって報告されたベンチマーク数によれば、最新の６４ビットデスクトップコンピュータにおいて、古典的なディフィー・ヘルマン群モジュラー１０２４ビット素数を使用してモジュラー指数を計算するためにおおよそ３ｍｓかかる。「ｃｕｒｖｅ２５５１９」等の高速楕円曲線を使用して、モジュラー指数を計算するために０．３ｍｓのみかかる。したがって、暗号化は、最新のコンピュータにおいておおよそ０．６ｍｓで行われることができる。アグリゲート統計値の復号は、離散対数をとることを必要とし、総当たり法を使用する場合、それぞれの可能な平文を試みるのに０．３ｍｓ必要とする１つのモジュラー指数をとる。したがって、本発明の実施形態によって提供されるＰＳＡ技術は、平文空間が小さい状況において実用的である。例えば、各参加者の平文が通信の

ための利用可能性を示す１ビットを含むとき、且つ、おおよそ１０００人の参加者がいるとき、復号は、総当たりアプローチを使用して約０．３ｓで行われることができる。復号のためにポラードのラムダ法を適応することは、さらなる速度向上を提供することができ、それゆえに、約√（ｎΔ）まで実行時間を短縮する。ここで、ｎは参加者数であり、各参加者の値が組｛０，１，・・・，Δ｝にあると仮定する。

【００５７】

本発明の実施形態は、アグリゲータが各個人の値ではなくノイズのある統計値のみを学習するということを保証する。それゆえに、個々の参加者についてのプライバシーは、最後の統計値ｆ（ｘ＾）が十分なランダム度を累積する限り保証される。これは、各個人がより少ないランダムノイズを個々のデータ値に対して加えるのを許容する。さらに、参加者のある一部が漏洩されてデータアグリゲータと共謀することを決定したとき、それらは、それらのデータ又はランダム度をアグリゲータに対して公開することがある。この場合において、残りの漏洩されていない参加者のランダム度がそれらのプライバシーを保護するのに十分であることが望ましい。

【００５８】

このプライバシーの概念は、公開された統計値におけるノイズが参加者から収集されるという事実を反映する分散型の差分プライバシー（ＤＤプライバシー）と称される。以下の段落は、分散型の差分プライバシーの概念を定義する。

【００５９】

アグリゲータが以下のようにして発生させられるｎ人の参加者のランダム化されたデータｘ＾∈Ｄ^{n}において関数ｆ：Ｄ^{n}→Ｏを評価するということを思い出してみよう。各参加者（ｉ）は、ある一定の分布にしたがって独立したランダム度ｒ$_{i}$∈Ωを発生させ、ｘ$_{i}$：＝χ（ｘ$_{i}$，ｒ$_{i}$）を生成するようにデータｘ$_{i}$においてランダム化関数χ：Ｄ×Ω→Ｄを適用する。ｘ∈Ｄ^{n}及びｒ∈Ω^{n}を考えると、表記ｘ＾＝ｘ＾（ｒ）：＝（χ（ｘ$_{1}$，ｒ$_{1}$），χ（ｘ$_{2}$，ｒ$_{2}$），・・・，χ（ｘ$_{n}$，ｒ$_{n}$））は、いかにｒにおけるｘ＾の従属が間接的であるかを表す。さらに、参加者の部分集合Ｋを考えると、以下の式（数１４）をＫの補集合であるものとする（すなわち、以下の式（数１５）である）。

【数１４】

$$\mathbf{r}_K := \{r_i : i \in K\} \text{及び} \overline{K}$$

【数１５】

$$\overline{K} = \{1,2,\ldots,n\} \setminus K$$

【００６０】

以下の分散型の差分プライバシー要件は、各期間ｔ∈Ｎに適用される。

【００６１】

定義３（（ε，δ）－ＤＤプライバシー）：ε＞０、０≦δ＜１、及び、０＜γ≦１と仮定する。データランダム化手順（すなわち、結合分布ｒ：＝（ｒ$_{1}$，・・・，ｒ$_{n}$）及びランダム化関数χによって与えられる）は、以下の状態が保持する場合には、関数ｆに関し且つ割合γの漏洩されていない一部の参加者の下で（ε，δ）－分散型の差分プライバシー（ＤＤプライバシー）を達成する。任意の隣接ベクトルｘ，ｙ≦Ｄ^{n}について、任意の部分集合Ｓ⊆Ｏについて、及び、少なくとも以下の式（数１６）のγｎ人の大きさの漏洩されていない参加者の任意の以下の式（数１７）の部分集合について。

【数１６】

$$\Pr[f(\hat{\mathbf{x}}) \in S \mid \mathbf{r}_K] \leq \exp(\varepsilon) \cdot \Pr[f(\hat{\mathbf{y}}) \in S \mid \mathbf{r}_K] + \delta.$$

【数１７】

$$\overline{K}$$

【００６２】

定義３において、２つのベクトルｘ，ｙ∈Ｄ^{n}は、それらが厳密に１つの座標において異なる場合、近傍又は隣接ベクトルであるといわれる。これは、厳密に１人のユーザがデータ値を変えるときにシナリオに対応する。

【００６３】

Ｋが漏洩されたノードの組であるとき、上述した定義は、残りの正当な参加者のランダム度が差分プライバシーを確実にするのに十分であることを必要とする。したがって、確率は、漏洩された参加者からのランダム度ｒ$_{K}$が条件とされている。換言すれば、確率は、正当な参加者からの以下のランダム度が引き継がれる。

【数１８】

$$\mathbf{r}_{\overline{K}}$$

ＤＤプライバシーの定義は、以下の式（数１９）である限り、漏洩されていない参加者の任意の以下の式（数２０）の組について、定義３の式が保持することを必要とする。

【数１９】

$$|\overline{K}| \geq \gamma n$$

【数２０】

$$\overline{K}$$

【００６４】

以下の段落は、（ε，δ）－差分プライバシー保証を達成する暗号構築を構成する方法を示している。この暗号技術は、それを暗号化する前に、データ値に対してノイズを加えることによってそれら自身のデータの差分プライバシーを確実にすることに参加者を関与させる。分散型の差分プライバシー保証を達成するために、以下の２つの課題は、対処される必要がある。

【００６５】

参加者個人についての差分プライバシーを確実にするために、公開された統計値は、適切な大きさのランダムノイズｒを含まなければならない。現実の設定において、参加者は、互いに信頼しないことがあり、参加者の部分集合が漏洩されて、データアグリゲータと共謀することがあり得る。最悪の場合において、あらゆる参加者が他のｎ－１人の参加者が漏洩されてアグリゲータと共謀すると思っている場合には、参加者は、それ自身のデータのプライバシーを確実にするために十分なノイズを加えることを必要とするであろう。しかしながら、これにより、アグリゲータが所望値を超えてエラーを累積した統計値を計算する結果をもたらすであろう。

【００６６】

少なくとも割合γの一部の参加者が正当であって漏洩されていない場合、ノイズ発生タスクは、これらの参加者間で分散されることができる。各参加者は、より少ないノイズを加えてもよく、最後の統計値におけるノイズが十分に大きい限り、個々のプライバシーは保護される。それゆえに、参加者がγについて下界の事前推定を有すると考えられる。しかしながら、それらは、どの参加者が漏洩されるかを厳密に知っている必要はない。各参加者は、γに依存する分布からノイズを発生させることになっている。正当な参加者は、このプロトコルにしたがうが、漏洩された参加者は、それらのノイズをデータアグリゲータに対して公開してもよく、又は、ノイズを加えないことを選択してもよい。この構築は、高確率で、公開された統計値が正当な参加者からの十分なノイズを累積することを保証

する一方で、許容レベルの範囲内で最後の統計値のエラーを保持する。

【００６７】

他の挑戦は、暗号構築と関連付けられた代数制約の範囲内での作業を含む。典型的な暗号化スキームは、平文値が離散値の群から抽出されることを必要とする。したがって、暗号構築は、離散群の範囲内でデータ及びノイズ値を符号化することが可能でなければならない。さらに、記憶しないアグリゲータを達成する暗号構築技術は、小さい平文空間において機能しなければならない。したがって、より共通に使用されるラプラス分布を使用する代わりに、対称幾何分布を使用することにより、そのような離散群によって作業することが可能である。

【００６８】

本発明の実施形態は、離散データ値の群によって作業するために対称幾何分布を使用する。ここで留意すべきは、対称幾何分布は制限されないことから、それは、群の大きさ又は平文空間の大きさからオーバーフローすることがあるということである。それゆえに、本発明の実施形態は、そのようなオーバーフローの確率が、アグリゲータが高い成功確率でノイズのある統計値を成功裏に復号するのを許容するのに十分小さいということを確実にする。

【００６９】

定義４（幾何分布）：α＞１とし、Ｇｅｏｍ（α）を、ｋにおける確率質量関数が以下であるように整数値をとる対称幾何分布を意味するものとする。

【数２１】

$$\frac{\alpha-1}{\alpha+1}\cdot\alpha^{-|k|}$$

さらに、Ｇｅｏｍ$^{+}$（α）を、ｋにおける確率質量関数が以下であるように正の整数値をとる片側幾何分布を意味するものとする。

【数２２】

$$(\alpha-1)\alpha^{-k}$$

【００７０】

対称幾何分布Ｇｅｏｍ（α）は、その確率密度関数が以下の式（数２３）であるラプラス分布Ｌａｐ（ｂ）の離散バージョンとしてみられることができる（αは以下の式（数２４）である）。

【数２３】

$$x\mapsto \frac{1}{2b}\exp\left(-\frac{|x|}{b}\right)$$

【数２４】

$$\alpha\approx\exp\left(\frac{1}{b}\right)$$

Ｇｅｏｍ分布の以下の特性は、整数値を出力する差分プライベート機構を設計するのに役に立つ。

【００７１】

特性１：ε＞０とし、ｕ及びｖが｜ｕ－ｖ｜≦Δであるような２つの整数であると仮定する。さらに、ｒを以下の分布を有する確率変数であるとする。

【数２５】

$$Geom(\exp(\frac{\varepsilon}{\Delta}))$$

そして、任意の整数ｋについて、Ｐｒ［ｕ＋ｒ＝ｋ］≦ｅｘｐ（ε）・Ｐｒ［ｖ＋ｒ＝ｋ］とする。

【００７２】

特性１は、ターゲットとされた統計値ｆ（ｘ）が機密度Δを有する場合、Δに比例する大きさを有する幾何ノイズを加えることは、差分プライバシーを達成するのに十分であることを示唆する。上述したように、参加者は、アグリゲータ又は互いを信頼しない。それゆえに、アグリゲータは、アグリゲータに対して真の統計値を公開することは差分プライバシーを明らかに違反することから、ノイズ発生のタスクによって信頼されてはならない。さらに、個々の参加者は、さもなければこの指定された参加者は真の統計値も学習することが可能であることから、このタスクによっても信頼されてはならない。

【００７３】

合計についてのＤＤプライバシーを達成するために、ｘ＝（x_1，・・・，x_n）∈D^n及びｒ＝（r_1，・・・，r_n）∈Ω^nを、所定期間における全ての参加者からのデータ及びノイズ値をそれぞれ表すものとする。それゆえに、Ｄ＝Ｏ＝Z_p（すなわち、加法モジュロｐを備える巡回群）であり、Ω＝Ｚである。また、以下を有するアグリゲート関数ｓｕｍ：$D^n \rightarrow O$を考える。

【数２６】

$$sum(\mathbf{x}) = \sum_{i=1}^{n} x_i p$$

さらに、各参加者は、以下の同じランダム化関数を使用する。

【数２７】

$$\chi(x_i, r_i) \coloneqq x_i + r_i p$$

【００７４】

任意の２つの要素ｕ，ｖ∈Z_pについて、｜ｕ－ｖ｜を、ｕ＝ｖ＋ｓｐ又はｖ＝ｕ＋ｓｐであるように最小の非負整数ｓであるとする。さらに、Z_pにおける要素に対して整数を加えるとき、加算は、モジュロｐを使用して実行されると考えられることができる。

【００７５】

また、各参加者の元データが領域｛０，１，・・・，Δ｝に含まれると仮定する。それゆえに、合計の機密度は、１人の参加者の変更に関してΔである。換言すれば、単一の参加者がデータを変える場合、合計は、最大でΔだけ変わる。以下のノイズが出力に組み込まれる場合、ε－差分プライバシーが達成されるということを特性１から思い出してみよう。

【数２８】

$$Geom(\exp(\frac{\varepsilon}{\Delta}))$$

この場合において、参加者は、最後の出力統計値に対して共同してノイズを与える。目標は、少なくともγｎ人の参加者が正当で且つ漏洩されていない場合、同様の大きさのノイズが累積されることを確実にすることである。このようにして、差分プライバシーが保証されるのみならず、累積されたノイズは、エラーが小さいように最後の出力において制限されることを確実にされる。

【００７６】
それゆえに、本発明の実施形態によって提供される暗号化機構は、（ε，δ）－ＤＤプライバシーを保証し、その一方で、おおよそ以下の大きさの小さいエラーを確実にする。
【数２９】

$$O\left(\frac{\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}}\right)$$

割合γの一定の一部の参加者が正当である限り、誤差項は、参加者数ｎから独立している。以下の大きさの累積されたノイズが差分プライバシーを確実するのに必要であることを考えれば、結果は最適に近い。
【数３０】

$$\Theta\left(\frac{\Delta}{\varepsilon}\right)$$

さらにまた、以下の式（数３１）であるときの極端な場合（すなわち、他の全ての参加者が漏洩されることがあると各参加者が考える、又は、一定数のそれらのみが正当である）を考えると、累積されたノイズは、以下の式（数３２）であろう。
【数３１】

$$\gamma = O\left(\frac{1}{n}\right)$$

【数３２】

$$O\left(\frac{\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}}\right) = O\left(\frac{\Delta}{\varepsilon}\sqrt{n}\right)$$

各参加者がこの場合においてプライバシーを確実にするために以下の大きさの対称ノイズを加えなければならないことから、これはまた直観的にうなずける。
【数３３】

$$\Theta\left(\frac{\Delta}{\varepsilon}\right)$$

したがって、以下の式（数３４）の大きさのｎ個の独立した対称ノイズの合計は、高確率で以下の式（数３５）の大きさの最後のノイズをもたらすことにつながる。
【数３４】

$$\Theta\left(\frac{\Delta}{\varepsilon}\right)$$

【数３５】

$$O\left(\frac{\Delta}{\varepsilon}\sqrt{n}\right)$$

【００７７】

定理２（低いエラーを有するＤＤプライベート手順）：ε＞０及び０＜δ＜１とする。各参加者のデータがZ_pにおける幅Δの間隔内の整数から到来すると仮定する。ここで、Δは以下である。

【数３６】

$$\Delta \geq \frac{\varepsilon}{3}$$

また、以下であるように少なくともｎ人の参加者の割合γの一部が漏洩されていないと仮定する。

【数３７】

$$\gamma \geq \frac{1}{n}\log\frac{1}{\delta}$$

そして、合計に関して（ε，δ）－ＤＤプライベートである$r=(r_1, \cdots, r_n)$を発生させるようにランダム化された手順が存在する。さらに、全ての$x \in (Z_p)^n$について、以下の式（数３８）であるように全ての０＜η＜１について、ｒのランダム選択にわたる少なくとも１－ηの確率で、以下の式（数３９）である。

【数３８】

$$\log\frac{2}{\eta} \leq \frac{1}{\gamma}\log\frac{1}{\delta}$$

【数３９】

$$|\, sum(\mathbf{x}) - sum(\hat{\mathbf{x}}) \,| \leq \frac{4\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}\log\frac{1}{\delta}\log\frac{2}{\eta}}$$

ここで、$x^{\wedge} := x^{\wedge}(r) := (x_1 + r_1, x_2 + r_2, \cdots, x_n + r_n)\ p$である。

【００７８】

表１は、定理２における保証を達成する手順を記載している。さらなる分析は、補助定理１及び定理３の下で以下に提供される。具体的には、定理３は、どの程度ノイズのある統計値ｓｕｍ（ｘ＾）が真の出力ｓｕｍ（ｘ）から偏差を有するかを分析する。ｘ＾以後、それは以下の大きさを制限するのに十分である。

【数４０】

$$Z := \sum_{i=1}^{n} r_i$$

【表１】

DDプライベートデータランダム化手順
$\alpha \coloneqq \exp(\frac{\varepsilon}{\Delta})$とし、$\beta \coloneqq \frac{1}{\gamma n}\log\frac{1}{\delta} \leq 1$とする。 ｘ＝（ｘ１，・・・，ｘｎ）を、所定期間における全ての参加者のデータを意味するものとする。 各参加者について、ｉ∈［ｎ］は、 以下の分布にしたがってノイズｒ$_i$をサンプリングする。 $$r_i \leftarrow \begin{cases} \mathrm{Geom}(\alpha) & \text{確率}\ \beta \\ 0 & \text{確率}\ 1-\beta \end{cases}$$ は、 $\hat{x}_i \leftarrow x_i + r_i \bmod p$を計算することによってデータをランダム化する。

【００７９】
補助定理１：ε＞０及び０＜δ＜１とする。少なくとも参加者の割合γの一部が漏洩されていないと仮定する。そして、上述したランダム化手順は、以下について、合計に関して（ε，δ）－ＤＤプライバシーを達成する。
【数４１】

$$\beta = \min\{\frac{1}{\gamma n}\log\frac{1}{\delta}, 1\}$$

【００８０】
定理３（制限エラー）：ε＞０及び０＜δ＜１とする。各参加者のデータが幅Δの間隔内の整数から到来すると仮定する。ここで、Δは以下である。
【数４２】

$$\Delta \geq \frac{\varepsilon}{3}$$

少なくともｎ人の参加者の以下の一部が漏洩されていないと仮定する。
【数４３】

$$\gamma \geq \frac{1}{n}\log\frac{1}{\delta}$$

表１におけるランダム化された手順が、以下によってｒ：＝（ｒ$_1$，・・・，ｒ$_n$）を生成するように実行されると仮定する。

【数４４】

$$\alpha \coloneqq \exp(\frac{\varepsilon}{\Delta}) \text{及び} \beta \coloneqq \frac{1}{\gamma m}\log\frac{1}{\delta} \leq 1$$

そして、以下の式（数４５）であるように全ての０＜η＜１について、少なくとも１－η の確率で、以下の式（数４６）であると仮定する。

【数４５】

$$\log\frac{2}{\eta} \leq \frac{1}{\gamma}\log\frac{1}{\delta}$$

【数４６】

$$|\sum_{i=1}^{n} r_i| \leq 4\sqrt{\frac{1}{\gamma}\log\frac{1}{\delta}\log\frac{2}{\eta}} \cdot \frac{\sqrt{\alpha}}{\alpha-1} \leq \frac{4\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}\log\frac{1}{\delta}\log\frac{2}{\eta}}$$

【００８１】

定理３によれば、累積エラーは、高確率で以下によって制限される。

【数４７】

$$O(\frac{\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}})$$

各参加者の値が領域Ｄ＝｛０，・・・，Δ｝から抽出されると仮定する。そして、アグリゲータは、単に、以下の範囲内で合計を復号しようとしなければならない。ここで、ｐは、使用中の数学的群の大きさである。

【数４８】

$$[-O(\frac{\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}}), n\Delta + O(\frac{\Delta}{\varepsilon}\sqrt{\frac{1}{\gamma}})]p$$

復号は、高確率で成功するはずである。

【００８２】

定理３は、各r_iの積率母関数を分析することによって証明される測定濃度結果である。漏洩されていない参加者の一定の割合γの一部がある限り、エラー制限はｎから独立していることに注目する。Ｇｅｏｍ（α）分布の分散が以下であることから、高確率で、エラーは、最後の回答に対して、多くてもＧｅｏｍ（α）の１つのコピーを加えるよりも悪い固定要因であり、それは、ε－差分プライバシーを確実にするのに必要とされるノイズの最小量である。

【数４９】

$$\frac{2\alpha}{(\alpha-1)^2}$$

【００８３】

分析者は、大抵は、母集団にわたる分布を調査したいであろう。いくつかの実施形態において、ＰＳＡシステムは、アグリゲータがｎ人の参加者のデータの近似分布を周期的に

評価するのを許容するために拡張されることができる。例えば、分布がガウシアンであると知られていると仮定すると、各参加者がその２乗と元の値を暗号化するのに十分である。これらの暗号化値を使用して、アグリゲータは、平均及び分散（又は二次モーメント）を介して分布を回収することができる。他の分布について、参加者は、より高次モーメントを暗号化することを必要としてもよい。一般に、各参加者がより多くのモーメントを暗号化するほど、アグリゲータは、より良好に分布を評価することができる。

【００８４】

いくつかの実施形態において、ＰＳＡシステムは、個々の値に対してではなく、アグリゲート値（例えば個々の値の合計）に対して公開アクセスするのを可能とする。例えば、アグリゲータ能力は、$sk_0=0$に設定されることができ、それゆえに、アグリゲータの能力を公開とすることができる。また、ｎ人の参加者は、ゼロになる値sk_1，・・・，sk_nを受信する。ｎ人の参加者及び任意の公開アグリゲータは、通常通りアグリゲート統計値の暗号化及び復号を実行する。アグリゲート合計を取得するために、離散対数が計算されなければならず、そのため、平文空間は小さくなければならない。

【００８５】

いくつかの実施形態において、ＰＳＡシステムについてのプロトコルは、アクセス制御階層を提供するためにネストさせられることができ、より高レベルのエンティティは、それらの下で全ての葉ノードを介してプールされる統計値に対してアクセスする。セットアップ段階において、レベル$j>1$におけるエンティティは、レベル未満におけるエンティティの秘密の合計が与えられる。また、$j=1$（すなわち、１つのネストレベルのみ）である場合、葉ノードより上の各エンティティは、それ未満の参加者の秘密の合計の負が与えられる。

【００８６】

いくつかの実施形態において、ＰＳＡ機構は、合計の代わりに積の記憶しない計算をサポートする。これを達成するために、参加者は、以下としてデータ値χを暗号化する。

【数５０】

$$c \leftarrow \chi \cdot H(t)^{sk_i}$$

もはや平文が指数ではないことから、積についてのこのスキームは、平文空間が小さいことを必要としない。

【００８７】

図５は、本発明の実施形態にかかるプライバシー保護データアグリゲーションを容易とするコンピューターシステムを示している。コンピューターシステム５０２は、プロセッサ５０４と、メモリ５０６と、記憶装置５０８とを含む。さらにまた、コンピューターシステム５０２は、表示装置５１０、キーボード５１２、及び、ポインティングデバイス５１３と接続されることができる。記憶装置５０８は、オペレーティングシステム５１４、アプリケーション５１６、及び、データ５２６を記憶することができる。

【００８８】

アプリケーション５１６は、コンピューターシステム５０２によって実行されるとき、コンピューターシステム５０２にこの開示に記載されている方法及び／又は処理を実行させる命令を含むことができる。具体的には、アプリケーション５１６は、秘密鍵を判定するための命令（セットアップモジュール５２０）と、平文値を暗号化するための命令（暗号化モジュール５２２）と、暗号化データ値の組からアグリゲート値を復号するための命令（アグリゲート復号モジュール５２４）とを含むことができる。

【００８９】

データ５２６は、この開示において記載された方法及び／又は処理によって入力として必要とされる又は出力として生成される任意のデータを含むことができる。具体的には、データ５２６は、少なくとも、秘密鍵、平文データ値、ランダムノイズ値、暗号化データ値、及び、アグリゲート値を記憶することができる。

【００９０】
図６は、本発明の実施形態にかかるプライバシー保護データアグリゲーションを容易とする装置を示している。装置６００は、有線又は無線通信チャネルを介して互いに通信してもよい複数の機構を備えることができる。装置６００は、１つ以上の集積回路を使用して実現されてもよく、また、装置６００は、図６に示されるものよりも少しの又は多くの機構を含んでもよい。さらに、装置６００は、コンピューターシステムにおいて一体化されてもよく、又は、他のコンピューターシステム及び／又は装置と通信することができる別個の装置として実現されてもよい。具体的には、装置６００は、通信機構６０２と、セットアップ機構６０４と、暗号化機構６０６と、アグリゲート復号機構６０８とを備えることができる。

【００９１】
いくつかの実施形態において、通信機構６０２は、データアグリゲータに対して暗号化データ値を送信し、及び／又は、参加者の組から暗号化データ値を受信するように構成されてもよい。さらに、セットアップ機構６０４は、秘密鍵を判定するように構成されてもよく、暗号化機構６０６は、平文値を暗号化するように構成されてもよく、アグリゲート復号機構６０８は、暗号化データ値の組からアグリゲート値を復号するように構成されてもよい。

【図１】

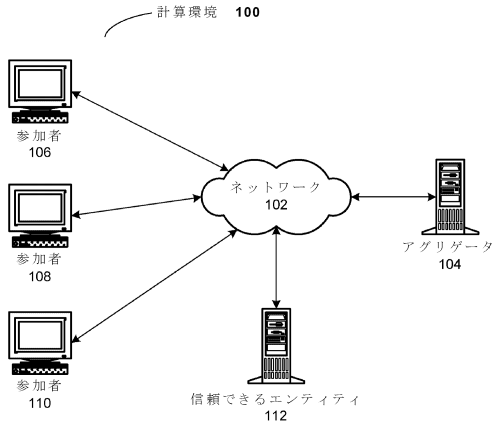

図１

【図２】

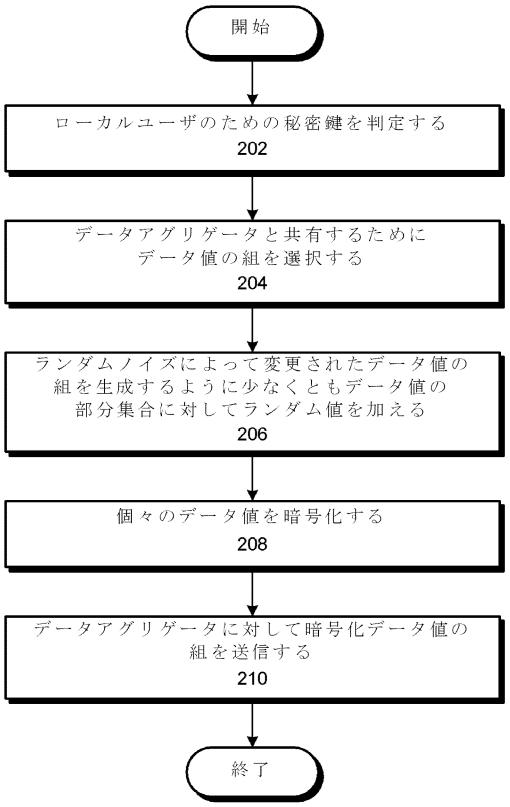

図２

【図３】

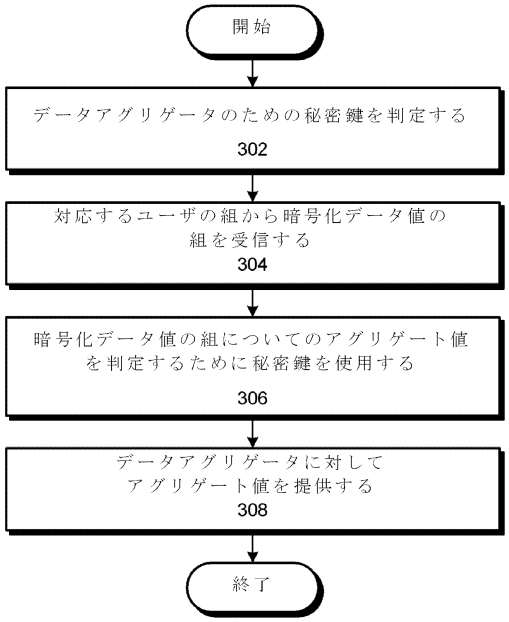

図３

【図４】

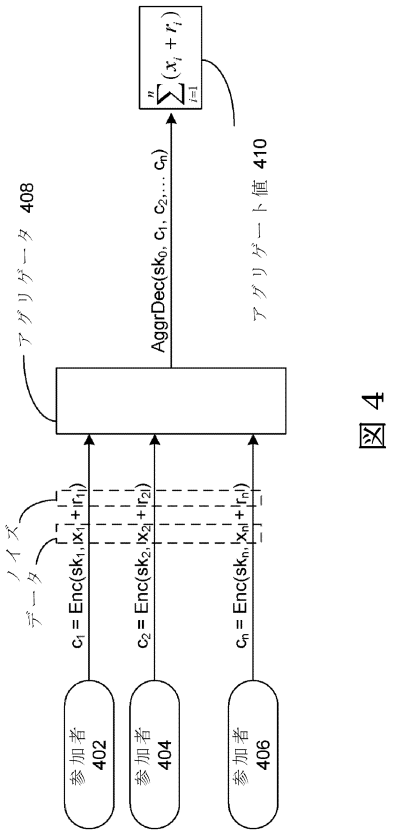

図４

【図５】

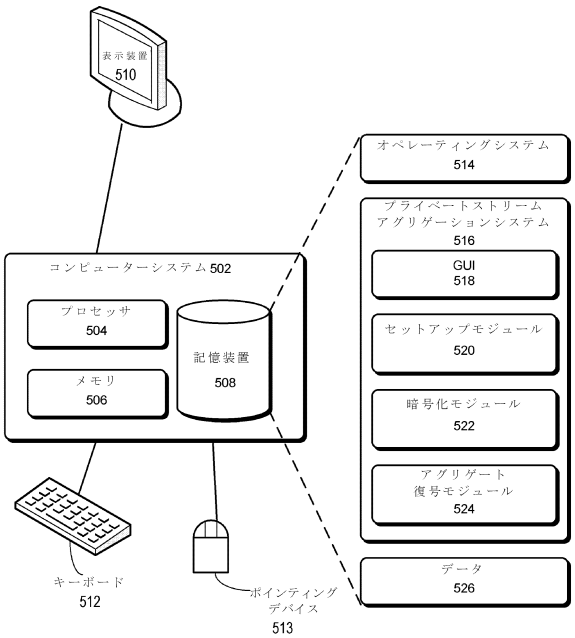

図５

【図６】

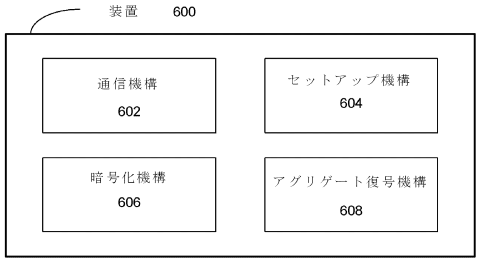

図６

フロントページの続き

(72)発明者 ルンティング・シー
アメリカ合衆国 カリフォルニア州 ９５１３０ サン・ノゼ ベイン・プレイス ４３４０
(72)発明者 リチャード・チョウ
アメリカ合衆国 カリフォルニア州 ９４０８７ サニーベール グロスビーク・アベニュー １６７４
(72)発明者 ツ・ホン・ヒューバート・チャン
中華人民共和国香港特別行政区 ポクフラム・ロード ユニバーシティー・オブ・ホンコン内 チョウ・エイ・チン・ビル コンピュータ・サイエンス学科

審査官 中里 裕正

(56)参考文献 特開２００６－０１８０５３（ＪＰ，Ａ）
特開２００８－１０９３０８（ＪＰ，Ａ）
特開２００８－２３４６０５（ＪＰ，Ａ）
特表２０１０－５２４４１３（ＪＰ，Ａ）
特開２０１２－０５８３４５（ＪＰ，Ａ）
特開２０１２－０８３５３６（ＪＰ，Ａ）
米国特許出願公開第２０１１／０２８３０９９（ＵＳ，Ａ１）
Rastogi, V. and Nath, S., Differentially Private Aggregation of Distributed Time-Series with Transformation and Encryption, Proceedings of the 2010 ACM SIGMOD International Conference on Management of data, ２０１０年, p.735-746

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ０６Ｆ ２１／６２
Ｈ０４Ｌ ９／１４
ＪＳＴＰｌｕｓ／ＪＭＥＤＰｌｕｓ／ＪＳＴ７５８０（ＪＤｒｅａｍＩＩＩ）