

明 細 書

発明の名称：

データ秘匿型統計処理システム、統計処理結果提供サーバ装置及びデータ入力装置、並びに、これらのためのプログラム及び方法

関連する出願

[0001] 本出願では、２０１３年１０月２３日に日本国に出願された特許出願番号２０１３－２２０６７３、及び２０１４年８月２９日に日本国に出願された特許出願番号２０１４－１７６５９０の利益を主張し、当該出願の内容は引用することによりここに組み込まれているものとする。

技術分野

[0002] 本発明は、個人のプライバシー等に関する秘匿すべきデータに対して、データの秘匿性を維持しながら統計処理を行い、その結果を提供するための技術に関する。

背景技術

[0003] 近年、個人情報や行動記録などの「ライフログ」を解析し、さまざまなビジネスシーンに活用する事例が増えている。例えば、ＰＯＳデータなどの購買履歴、電子マネーの利用履歴、交通網の乗車履歴、車のＧＰＳ情報、携帯電話やスマートフォンなどの通話履歴や利用履歴、血圧や体重等のヘルスケアに関する測定履歴、さらに病歴等、あらゆる場面でデータを解析することが求められている。

[0004] 「ライフログ」から得られる情報は有益なものが多く、行動パターンの推定やリコメンド、ターゲットマーケティング、新製品や新手法の研究開発等、多くの応用が考えられる。一方で、データを解析する際のプライバシー情報の取り扱いに関する懸念も大きい。

[0005] また、クラウド・コンピューティングの技術を利用し、個人や企業等のユーザが、自身のデータを、手元の機器に保存せずに、ネットワークを介してデータセンタ等へ送信して保存できるようにするサービスも、普及している

。この場合も、クラウド上に保存されるデータにプライバシー情報が含まれると、情報漏洩への懸念が大きくなる。

- [0006] プライバシー情報を保護しつつデータを解析して有用な知識を見出すための技術としては、プライバシー保護データマイニング（PPDM）と呼ばれる技術が開発されており（非特許文献1参照）、保存されているデータ自体が第三者に流出しても秘密の情報は漏洩しないようにするための技術としては、秘密分散と呼ばれる技術が提案されている（特許文献1～3参照）。

先行技術文献

特許文献

- [0007] 特許文献1：特開2013-20314号公報
特許文献2：特表2012-530391号公報
特許文献3：特開2005-250866号公報

非特許文献

- [0008] 非特許文献1：佐久間 淳、小林重信「プライバシー保護データマイニング」、人工知能学会誌 Vol. 24 No. 2 (2009)

発明の概要

発明が解決しようとする課題

- [0009] PPDMには、信頼できる第三者機関の存在を想定し、秘匿すべきオリジナルデータを当該第三者機関へ渡す方式があるが、そのように信頼できる第三者機関は、実際には実現困難である上に、秘密の情報が集約される当該第三者機関から情報が漏洩すると、その被害が甚大となることから、現実的な解決策にはならない。
- [0010] PPDMのうち、信頼できる第三者機関を利用しない方式は、ある組織が有するオリジナルデータを外部から秘匿しながら、そのオリジナルデータの集合に対する解析の結果を当該組織の外部において求めるものであり、解析処理を行う外部の者は、オリジナルデータを渡されることなく、何らかの秘匿化処理が行われたデータを渡されて、解析処理を行う。その際、外部の者

が、渡されたデータから、組織内に秘匿されているオリジナルデータを求めることができないようにするために、種々の手法が開発されている。

[0011] しかしながら、信頼できる第三者機関を利用しない方式も、組織の内部では、秘匿すべきオリジナルデータを保持していることを前提としている。したがって、当該組織の保持しているオリジナルデータが第三者に流出し、プライバシー情報が漏洩するリスクについては、P P D M自体は、無防備であることになる。

[0012] そうすると、従来の技術では、P P D Mとオリジナルデータの保持を暗号化された状態で行う技術とを組み合わせることにより、秘匿すべきデータの安全性を保つことになるが、いくら暗号化された状態であっても、オリジナルデータが存在する以上は、暗号化強度に応じて膨大となるものの計算能力と時間さえかければ、復号化してオリジナルデータを求めることができってしまう。よって、情報漏洩のリスクは、排除できないものとして残ることになる。

[0013] これに対して、秘密分散技術は、秘密情報を幾つか（ N 個とする）のデータに分割して、分散させて保持することにより、 N 個のうちの K 個（ $K < N$ ）のデータが第三者に流出して、それらが集められても、秘密情報を復元することはできないようにして、情報漏洩を防ぐものである。

[0014] このように秘密情報を分散させるということは、オリジナルデータを保持しないということであり、 N 及び K の値を大きくすることにより、確実に情報漏洩のリスクを低減することができる。つまり、 K 箇所で、保持しているデータの流出が起きても、秘密情報は漏洩しないことが担保されるため、 K の値を十分に大きくし、それぞれのデータ保持場所でのセキュリティを高くすることにより、 K 個の全ての場所からデータが流出する可能性を、極めて小さくすることが可能である。

[0015] しかしながら、秘密分散技術により安全に保持されている秘密情報を解析したい場合、分割された状態のままで解析をすることができないため、一旦、データ全部を一箇所に集めてオリジナルの秘密情報を復元した上で、解析

処理を行わなければならない。そうすると、たとえ通常の保存時に秘密分散技術を利用していても、解析をする際にオリジナルデータを保持することになるため、データの流出が即、情報漏洩につながるというリスクは、依然として残ることになる。

[0016] 本発明は、上述した状況に鑑み、オリジナルデータの受け渡しも保存も行わないことにより、オリジナルデータを保持しないようにして、秘密すべき情報が漏洩するリスクを低減しつつ、オリジナルデータの集合について統計処理の結果を得ることを可能にすることを目的とする。

課題を解決するための手段

[0017] 本発明の原理に従う一つの例に係るデータ秘匿型統計処理システムは、それぞれが秘密すべきオリジナルデータを取得する手段を備える複数のデータ入力装置と、それぞれが複数の入力データに基づいて所定の演算を行う手段を備える複数の演算装置と、前記複数の演算装置のそれぞれが前記オリジナルデータの部分データを前記入力データとして演算を行った結果を利用することにより、前記複数のデータ入力装置により取得された複数のオリジナルデータに基づく統計処理の結果を、該オリジナルデータを取得することなく求める手段を備えるデータ処理装置を備える。

発明の効果

[0018] 本発明によれば、オリジナルデータを保持しないようにして秘密すべき情報が漏洩するリスクを低減しつつ、オリジナルデータの集合について統計処理の結果を得ることが可能になる。

図面の簡単な説明

[0019] [図1]本発明の実施の形態に係るデータ秘匿型統計処理システム（以下、「本システム」という）において総和を求める例を説明する図
 [図2]本システムにおいて総和を求める別の例を説明する図
 [図3]本システムにおいて2乗和を求める例を説明する図
 [図4]本システムにおいて2乗和を求める別の例を説明する図
 [図5]本システムにおいて内積を求める例を説明する図

[図6]本システムの構成例を示す図

[図7]本システムにおける統計処理結果提供サーバの構成例を示す図

[図8]本システムにおける処理手順の例（１）～（３）を説明する図

[図9]本システムにおける処理手順の例（４）～（６）を説明する図

[図10]本システムにおける処理手順の例（７）～（９）を説明する図

[図11]本システムにおける処理手順の例（１０）～（１２）を説明する図

[図12]本システムにおける処理手順の例（１３）～（１５）を説明する図

[図13]本システムにおける処理手順の例（１６）～（１８）を説明する図

[図14]本システムにおける処理手順の例（１９）～（２１）を説明する図

[図15]本システムにおける処理手順の例（２２）～（２４）を説明する図

[図16]本システムの別の構成例を示す図

[図17]本システムにおける別の処理手順の例（１）～（２）を説明する図

[図18]本システムにおける別の処理手順の例（３）～（５）を説明する図

[図19]本システムにおける別の処理手順の例（６）～（８）を説明する図

[図20]本システムのさらに別の構成例を示す図

[図21]本システムにおけるさらに別の処理手順の例（１）～（２）を説明する図

[図22]本システムにおけるさらに別の処理手順の例（３）～（６）を説明する図

[図23]本システムにおけるさらに別の処理手順の例（７）～（１０）を説明する図

[図24]本システムを教育の分野において応用する例を説明する図

[図25]本システムを医療の分野において応用する例を説明する図

[図26]本システムを流通業（小売業）の分野において応用する例を説明する図

[図27]本システムをテレマティクスの分野において応用する例を説明する図

発明を実施するための形態

[0020] 上述した本発明の原理に従う一例に係るデータ秘匿型統計処理システムの

構成によれば、各データ入力装置が取得したオリジナルデータは、部分データにされて複数の演算装置に分散して渡されるため、いずれの演算装置もオリジナルデータを取得せず、データ処理装置もオリジナルデータを取得しない。よって、オリジナルデータを保持しないことで、秘匿すべき情報が漏洩するリスクを低減することが可能になる。一方で、各演算装置が、部分データについて演算を行い、データ処理装置が、複数の演算装置からの演算結果を利用することで、オリジナルデータの集合についての統計処理の結果を得ることが可能になる。

[0021] 上記の構成において、前記データ入力装置が、全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを秘密の比率に従って分割し、所定の個数の部分データを生成する手段と、前記所定の個数の部分データのそれぞれを、前記複数の演算装置のうち対応する演算装置へ、保護された通信路により送信する手段を備えるようにしてもよい。

[0022] これにより、オリジナルデータをM個に分割して、M個の演算装置に送信した場合、最大(M-1)個までの部分データが第三者に流出しても、オリジナルデータを復元することができないようにすることが可能になる。したがって、M個の演算装置がそれぞれの部分データを保存して、一部の演算装置からデータが第三者に流出したとしても、オリジナルデータの秘匿性を維持可能とすることができる。また、データ入力装置からの通信路の保護により、通信の傍受によって全部の部分データ（即ちオリジナルデータ）が第三者に取得されることを防止することができる。

[0023] なお、秘密の比率は、各々のデータ入力装置によって異なるようにすることが望ましい。また、各々のデータ入力装置が生成する部分データの個数は、一つの統計処理の対象となる集合に属するオリジナルデータの全てについて同一とする方が、運用管理が簡単化されるが、異なることを許容しても構わない。

[0024] 上記の構成において、前記演算装置が、前記複数のデータ入力装置から受信した複数の前記部分データに基づいて所定の演算を行って得た演算結果を

、前記データ処理装置へ送信する手段を備え、前記データ処理装置が、前記複数の演算装置から受信した複数の前記演算結果に基づいて所定の統計処理を行う手段を備えるようにしてもよい。

[0025] これにより、M個の演算装置のそれぞれが、N個のデータ入力装置から部分データを受信し、N個の部分データから演算した結果をデータ処理装置へ送信し、データ処理装置が、M個の演算結果を処理することで、N個のオリジナルデータについて統計処理を行った結果を求めることが可能になる。

[0026] このとき、各演算装置は、N個のオリジナルデータに対応するN個のデータを受け取るが、それは部分データであってオリジナルデータの情報を含んでおらず、データ処理装置は、オリジナルデータを構成するM個の部分データに対応するM個の演算結果を受け取るが、それはオリジナルデータの集合についての情報であって個々のオリジナルデータの情報を含んでいない。よって、各演算装置とデータ処理装置に、いずれのオリジナルデータも取得させることなく、統計処理の結果が得られることになる。

[0027] 上記の構成において、前記所定の個数の部分データは、前記オリジナルデータを分割した各々の部分の値から生成されるものを含み、前記演算装置が行う所定の演算は、前記複数の部分データの総和の演算を含み、前記データ処理装置が行う所定の統計処理は、前記所定の個数の演算結果の総和を計算する処理を含むものとしてもよい。

[0028] これにより、N個のオリジナルデータの総和 $(X_1 + X_2 + \dots + X_N)$ という統計処理の結果を求めることが、オリジナルデータを取得することなく可能になる。例えば、i番目のデータ入力装置 ($i = 1, 2, \dots, N$) が、 $X_i = x_{1i} + x_{2i} + \dots + x_{mi}$ となるように、m個の部分データ x_{ji} を生成し、j番目の演算装置 ($j = 1, 2, \dots, m$) が、N個の部分データの総和 $(x_{j1} + x_{j2} + \dots + x_{jN})$ の値を求め、データ処理装置が、m個の演算装置が求めた値の総和を求めれば、 $(X_1 + X_2 + \dots + X_N)$ の値が求められる。

[0029] 上記の構成において、前記所定の個数の部分データは、前記オリジナルデータを分割した各々の部分の値から生成されるものと、互いに異なる2つの

部分同士を乗算した値に基づいて生成されるものを含み、前記演算装置が行う所定の演算は、前記複数の部分データの総和及び2乗和の少なくとも一方の演算を含み、前記データ処理総理が行う所定の統計処理は、前記所定の個数の演算結果のうち前記各々の部分の値に対応するものについて2乗和を計算する処理と、前記所定の個数の演算結果のうち前記部分同士を乗算した値に対応するものについて総和を計算する処理を含むものとしてもよい。

[0030] これにより、N個のオリジナルデータの2乗和 ($X_1^2 + X_2^2 + \dots + X_N^2$) という統計処理の結果を求めることが、オリジナルデータを取得することなく可能になる。例えば、i番目のデータ入力装置 ($i = 1, 2, \dots, N$) が、 $X_i = x_{1i} + x_{2i} + \dots + x_{mi}$ となるように、m個の部分データ x_{ji} を生成し、さらに、m個の部分データ [$\sum_{j \neq k} (x_{ji} x_{ki})$] (以下、「 x'_{ji} 」と記す) を生成し、j番目の演算装置 ($j = 1, 2, \dots, m$) が、N個の部分データ x_{ji} の2乗和 ($x_{j1}^2 + x_{j2}^2 + \dots + x_{jN}^2$) の値を求め、j番目の演算装置 ($j = m+1, m+2, \dots, 2m$) が、N個の部分データ x'_{ji} の総和 ($x'_{j1} + x'_{j2} + \dots + x'_{jN}$) の値を求め、データ処理装置が、2m個の演算装置が求めた値の総和を求めれば、($X_1^2 + X_2^2 + \dots + X_N^2$) の値が得られる。

[0031] 別の例として、i番目のデータ入力装置 ($i = 1, 2, \dots, N$) が、 $X_i = x_{1i} + x_{2i} + \dots + x_{mi}$ となるように、m個の部分データ x_{ji} を生成し、さらに、 $m+1$ 番目の部分データ [$\sum_j (\sum_{j \neq k} (x_{ji} x_{ki}))$] (以下、「 x''_i 」と記す) を生成し、j番目の演算装置 ($j = 1, 2, \dots, m$) が、N個の部分データ x_{ji} の2乗和 ($x_{j1}^2 + x_{j2}^2 + \dots + x_{jN}^2$) の値を求め、 $m+1$ 番目の演算装置が、N個の部分データ x''_i の総和 ($x''_1 + x''_2 + \dots + x''_N$) の値を求め、データ処理装置が、 $m+1$ 個の演算装置が求めた値の総和を求めるのでも、($X_1^2 + X_2^2 + \dots + X_N^2$) の値が得られる。

[0032] 代替となる構成として、上記の構成において、前記所定の個数の部分データは、前記オリジナルデータを分割した各々の部分を2乗した値から生成されるものと、互いに異なる2つの部分同士を乗算した値に基づいて生成されるものを含み、前記演算装置が行う所定の演算は、前記複数の部分データ

の総和の演算を含み、前記データ処理装置が行う所定の統計処理は、前記所定の個数の演算結果の総和を計算する処理を含むものとしてもよい。

[0033] これによっても、 N 個のオリジナルデータの2乗和 ($X_1^2 + X_2^2 + \dots + X_N^2$) という統計処理の結果を求めることが、オリジナルデータを取得することなく可能になる。例えば、 i 番目のデータ入力装置 ($i = 1, 2, \dots, N$) が、 $X_i = x_{1i} + x_{2i} + \dots + x_{mi}$ となるように x_{ji} を定めて、 m 個の部分データ x_{ji}^2 と、 m 個の部分データ x'_{ji} とを生成し、 j 番目の演算装置 ($j = 1, 2, \dots, m$) が、 N 個の部分データ x_{ji}^2 の総和 ($x_{j1}^2 + x_{j2}^2 + \dots + x_{jN}^2$) の値を求め、 j 番目の演算装置 ($j = m+1, m+2, \dots, 2m$) が、 N 個の部分データ x'_{ji} の総和 ($x'_{j1} + x'_{j2} + \dots + x'_{jN}$) の値を求め、データ処理装置が、 $2m$ 個の演算装置が求めた値の総和を求めれば、($X_1^2 + X_2^2 + \dots + X_N^2$) の値が得られる。

[0034] 別の例として、 i 番目のデータ入力装置 ($i = 1, 2, \dots, N$) が、 $X_i = x_{1i} + x_{2i} + \dots + x_{mi}$ となるように x_{ji} を定めて、 m 個の部分データ x_{ji}^2 と、1 個の部分データ x''_i とを生成し、 j 番目の演算装置 ($j = 1, 2, \dots, m$) が、 N 個の部分データ x_{ji}^2 の総和 ($x_{j1}^2 + x_{j2}^2 + \dots + x_{jN}^2$) の値を求め、 $m+1$ 番目の演算装置が、 N 個の部分データ x''_i の総和 ($x''_1 + x''_2 + \dots + x''_N$) の値を求め、データ処理装置が、 $m+1$ 個の演算装置が求めた値の総和を求めるのでも、($X_1^2 + X_2^2 + \dots + X_N^2$) の値が得られる。

[0035] 上述した例では、総和を求めるのに m 個の演算装置を用い、2乗和を求めるのに $2m$ 個もしくは $m+1$ 個の演算装置を用いているが、いずれの場合も、同時に $(m-1)$ 箇所データ流出があっても、オリジナルデータの秘匿性は維持可能である。

[0036] また、各々の演算装置は、データ入力装置から受信したデータが何であるかに関わらず、受信したデータに対して総和及び2乗和の演算をして、この2つの演算結果をデータ処理装置に送信するという、統一的な処理を行う構成とし、データ処理装置が、行いたい統計処理に従ってデータ処理装置からの演算結果を選択 (例えば、 $1 \sim m$ 番目の演算装置については2乗和の方の

結果を選択し、 $m+1 \sim 2m$ 番目の演算装置については総和の方の結果を選択する等）して、計算対象とするように構成してもよい。

[0037] 上記のようにオリジナルデータの集合の総和及び2乗和という統計処理の結果を求めることができる構成を利用して、標本平均の算出、標本分散の算出、標本偏差の算出、最尤推定、T分布を用いた区間推定、母比率の信頼区間の推定、母分散の推定、母平均の検定、母集団A及びBに関する母平均の差の検定、母比率の検定、母集団A及びBの母分散の比較検定、分散分析のうちの少なくとも1つの結果を、最終的な統計処理の結果として求める構成とすることも可能である。

[0038] 上記の構成において、前記複数のデータ入力装置は、同数ずつの互いに対応する第1のデータ入力装置と第2のデータ入力装置とを含み、前記第1のデータ入力装置及び前記第2のデータ入力装置が行う送信は、前記所定の個数の部分データのそれぞれを、所定の個数の2乗個の前記演算装置のうち対応する所定の個数の演算装置へ送信するものであり、前記演算装置が行う所定の演算は、前記第1のデータ入力装置からの部分データ列と前記第2のデータ入力装置からの部分データ列との内積を求める演算を含み、前記データ処理装置が行う統計処理は、前記所定の個数の2乗個の演算装置から受信した所定の個数の2乗個の前記演算結果の総和を計算する処理を含むものとしてもよい。

[0039] これにより、第1のオリジナルデータの集合（ N 個のオリジナルデータ X_i ）と第2のオリジナルデータの集合（ N 個のオリジナルデータ Y_i ）の内積（ $X_1Y_1 + X_2Y_2 + \dots + X_NY_N$ ）という統計処理の結果を求めることが、オリジナルデータを取得することなく可能になる。例えば、 i 番目の第1のデータ入力装置（ $i = 1, 2, \dots, N$ ）が、 $X_i = x_{1i} + x_{2i} + \dots + x_{mi}$ となるように、 m 個の部分データ x_{ji} を生成し、 i 番目の第2のデータ入力装置（ $i = 1, 2, \dots, N$ ）が、 $Y_i = y_{1i} + y_{2i} + \dots + y_{mi}$ となるように、 m 個の部分データ y_{ki} を生成し、 j, k 番目の演算装置（ $j, k = 1, 2, \dots, m^2$ ）が、 N 個の部分データ x_{ji} と N 個の部分データ y_{ki} との内積（ $x_{j1}y_{k1} + x_{j2}y_{k2} + \dots + x_{jN}y_{kN}$ ）の

値を求め、データ処理装置が、 m^2 個の演算装置が求めた値の総和を求めれば、 $(X_1Y_1 + X_2Y_2 + \dots + X_NY_N)$ の値が求められる。

[0040] 上記のように2つのオリジナルデータの集合の内積という統計処理の結果を求めることができる構成を利用して、共分散の算出、相関係数の算出、回帰分析のうちの少なくとも1つの結果を、最終的な統計処理の結果として求める構成とすることも可能である。

[0041] 上記のデータ秘匿型統計処理システムにおいて、前記データ入力装置が、前記オリジナルデータを分割する際に発生させた乱数を用いて前記秘密の比率を定め、分割した後に前記秘密の比率の記憶を消去する手段をさらに備えるようにしてもよい。

[0042] これにより、オリジナルデータを構成する複数の部分データのうちの一つしか第三者に流出しておらず、オリジナルデータの秘匿性が維持されるべき場合でも、秘密の比率を知られてしまうと、オリジナルデータが復元できてしまうという情報漏洩のリスクを、低減することが可能になる。その都度ランダムに秘密の比率を定めるようにすれば、比率が推測されてしまう可能性が少なくなるし、秘密の比率の記憶を消去しておけば、情報漏洩の可能性も少なくなる。

[0043] 上記のシステムにおいて、前記演算装置が、前記複数のデータ入力装置から受信した複数の前記部分データのそれぞれを、該部分データを送信したデータ入力装置と関連付けて保存する手段と、前記データ入力装置との関連付けを示す要求に応じて、前記複数の部分データのうち、該データ入力装置と関連付けて保存されている部分データを返送する手段をさらに備えるようにしてもよい。

[0044] これにより、データ入力装置が取得したオリジナルデータを即座に分割し、複数の演算装置に分散して保存させることで、データ入力装置もオリジナルデータを保持しないことになり、秘匿すべき情報が漏洩するリスクを確実に低減することが可能になる。

[0045] 上記の構成において、前記データ入力装置との関連付けを有する装置が、

前記オリジナルデータを分割して生成した部分データの全てを、前記複数の演算装置のうち対応する演算装置から取得して、前記オリジナルデータを復元する手段を備えるようにしてもよい。

[0046] これにより、秘密の比率の記憶が残っていなくても、オリジナルデータの本来の保有者は、分散して保存されている複数の部分データを全て集めることで、オリジナルデータを復元することが可能になる。

[0047] 代替となる構成として、前記データ入力装置との関連付けを有する装置が、前記オリジナルデータを分割した各々の部分のうち一つについての前記比率を記憶する手段と、前記オリジナルデータを分割して生成した部分データのうち、記憶されている前記一つの比率に該当する部分データを、前記複数の演算装置のうち対応する演算装置から取得して、前記オリジナルデータを復元する手段とを備えるようにしてもよい。

[0048] これにより、オリジナルデータの本来の保有者は、分散して保存されている複数の部分データのうちの一つを取得することで、オリジナルデータを復元することが可能になる。

[0049] 上記のシステムにおいて、前記データ処理装置が、前記複数のデータ入力装置のそれぞれへ、前記部分データを前記複数の演算装置のうちいずれへ送信すべきかを指示する手段と、前記複数の演算装置のそれぞれへ、前記複数のデータ入力装置から受信した複数の前記部分データのうちのいずれを対象として所定の演算を行うべきかを指示する手段を備えるようにしてもよい。

[0050] これにより、どのような統計処理の結果を得たいかに応じて、その都度、利用する演算装置を選択したり、演算装置の数を指定したりすることが可能になり、時々の状況に応じた負荷分散や、安全度の細かい設定等が可能になる。また、各演算装置が有する部分データが、所望の統計処理の対象となるオリジナルデータの部分データであるかそうでないかを、演算装置に伝えることが可能になり、統計処理の対象に含めてしまうと結果に誤差等が生じる部分データを、演算から外すことが可能になる。

[0051] 上記のシステムにおいて、前記複数のデータ入力装置のそれぞれが、前記

部分データを前記複数の演算装置のうちいずれへ送信すべきかを決定する手段を備え、前記複数の演算装置のそれぞれが、前記複数のデータ入力装置から受信した複数の前記部分データのうちいずれを対象として所定の演算を行うべきかを決定する手段を備えるようにしてもよい。

[0052] これにより、各データ入力装置自身が送信先の演算装置を選択することと、各演算装置自身が統計処理の対象に含めるべき部分データを選別することとが可能になり、データ処理装置が、各オリジナルデータの内容を取得しただけでなく、各オリジナルデータに関する情報をも取り扱わないようにすることができ、さらなるデータの安全性を実現することができる。

[0053] 上記のいずれの構成においても、前記複数の演算装置の数を、一つのオリジナルデータから得られる部分データの個数である所定の数と同じかそれよりも大きくし、前記所定の個数の部分データが、それぞれ異なる演算装置に送信されるようにするとよい。

[0054] 上述したシステムにおいて、前記複数の演算装置は、それぞれ異なる事業者により提供されるサービスに属するものであり、前記データ処理装置は、前記複数の演算装置とは異なる事業者により運営されるものであってもよい。

[0055] これにより、例えば、統計処理の主体となる事業者が、データ処理装置を運営管理し、既存の複数のクラウドサービス事業者から提供されるデータの保存及び演算サービスを利用して、統計処理結果提供サービスを行うようなことが可能になる。

[0056] 本発明の原理に従う一つの例に係る統計処理結果提供サーバ装置は、秘匿されるべきオリジナルデータを取得することなく、複数の前記オリジナルデータに基づく統計処理の結果を提供するサービスのためのサーバ装置であり、それぞれが複数の入力データに基づいて所定の演算を行う手段を有する複数の演算装置と通信する手段と、前記複数の演算装置のそれぞれに、前記オリジナルデータの部分データを前記入力データとして演算を行わせ、該演算の結果を取得する手段と、前記複数の演算装置からの演算結果に基づいて、

所定の統計処理を行う手段を備える。複数の前記部分データは、全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを秘密の比率に従って分割することにより、生成されたものとする。

[0057] この構成によれば、オリジナルデータが部分データにされて複数の演算装置に分散して渡されるため、いずれの演算装置もオリジナルデータを取得せず、サーバ装置もオリジナルデータを取得しない。よって、オリジナルデータを保持しないことで、秘匿すべき情報が漏洩するリスクを低減することが可能になる。一方で、サーバ装置が、複数の演算装置に部分データを入力とする演算を行わせ、その結果を利用することで、オリジナルデータの集合についての統計処理の結果を得ることが可能になる。また、第三者が一部の部分データを取得してもオリジナルデータは復元されないため、オリジナルデータの秘匿性を維持可能とすることができる。なお、秘密の比率は、オリジナルデータを分割する機器内にのみ、少なくとも分割時に存在するだけで、誰にも知られないか、オリジナルデータの保有者しか知ることのできないものである。

[0058] 上記のサーバ装置において、前記複数の演算装置に前記オリジナルデータに属する全ての部分データが揃ったことを確認する手段と、前記確認のされた部分データのそれぞれを対象として対応する演算装置における前記所定の演算が行われるように、前記複数の演算装置のそれぞれへ指示を行う手段をさらに備えるようにしてもよい。

[0059] これにより、統計処理の対象に含めてしまうと結果に誤差等が生じる部分データを、演算から外すことが可能になる。例えば、同じオリジナルデータに属するある部分データは対応する演算装置に受信され保存されているが、別の部分データは対応する演算装置に受信されていない場合に、各演算装置が、自身に保存されている部分データ全てを対象に演算してしまうと、それらの演算装置からの演算結果を処理した結果は、誤ったものになってしまう。ここで、複数の演算装置を統括的に利用するサーバ装置が、全ての部分データが揃っているものを各演算装置に伝えれば、正しい統計処理結果を得る

ことが可能になる。

[0060] 上記の構成において、サーバ装置が、前記確認のために、前記複数の演算装置のそれぞれから、該演算装置に保存された部分データが属するオリジナルデータの識別番号を受信する手段をさらに備えるようにしてもよい。

[0061] これにより、サーバ装置は、複数の演算装置を見渡して、全ての部分データが揃っている状態になったか否かを、各演算装置から個々の部分データを取得するようなことなく、確認することが可能になる。

[0062] 上記の構成において、サーバ装置が、前記確認がされたオリジナルデータの識別番号の集合をシーケンス番号と対応付けて前記複数の演算装置へ通知する手段と、前回の通知の後に前記確認がされたオリジナルデータの識別番号の集合を次のシーケンス番号と対応付けて前記複数の演算装置へ通知する手段をさらに備え、前記複数の演算装置のそれぞれに、1つのシーケンス番号の指定と共に前記所定の演算の指示を送信することにより、指定されたシーケンス番号以前の複数のシーケンス番号に対応する識別番号の集合を併せて前記所定の演算の対象とする部分データが特定されるようにしてもよい。

[0063] これにより、数多くの部分データが各演算装置に受信され蓄積されていく間に、随時、各演算装置が有する数多くの部分データのうちどれが、全ての部分データが揃っている状態になったかという情報を、サーバ装置が各演算装置に共有させることが可能になる。

[0064] 上記の構成において、サーバ装置は、あるオリジナルデータの集合を対象として、前記複数の演算装置に前記所定の演算を行わせた結果を取得した後に、該あるオリジナルデータの集合に限定された数のオリジナルデータを加えた集合を対象として、前記複数の演算装置に前記所定の演算を行わせた結果を取得することを禁止する手段をさらに備えるようにしてもよい。

[0065] 上述したように、サーバ装置は、N個の部分データから演算した結果を、M個の演算装置のそれぞれから受信して処理することで、N個のオリジナルデータについて統計処理を行った結果を求めるのであるから、ある時点で、 $i = 1, \dots, N$ のオリジナルデータについての統計処理結果を求め、次の時

点で、 $i = 1, \dots, N, N + 1$ のオリジナルデータについての統計処理結果を求めて、両者の差分を取れば、 $i = N + 1$ のオリジナルデータを求めることができる。

[0066] そのような時点で演算結果を取得することを禁止することにより、サーバ装置が、実質的に各演算装置から個々の部分データを取得してオリジナルデータを復元するような、悪意のある動作を行わないことを担保することが可能になる。

[0067] 上記のサーバ装置において、それぞれが前記オリジナルデータを取得して前記部分データを生成する手段を有する複数のデータ入力装置と通信する手段と、利用可能な演算装置の中から、前記所定の統計処理を行うための前記複数の演算装置を選択する手段と、前記複数のデータ入力装置のそれぞれへ、選択された前記複数の演算装置に対して複数の前記部分データを送信できるように、前記複数の演算装置の情報を通知する手段をさらに備えるようにしてもよい。

[0068] これにより、どのような統計処理の結果を得たいかに応じて、その都度、利用する演算装置を選択することが可能になり、また、データ入力装置の数が多くても、サーバ装置からの通知によって一元的に部分データの送信先を設定することができるため、運用管理が簡単化される。

[0069] 本発明の原理に従う一つの例に係るデータ入力装置は、秘匿すべきオリジナルデータを取得する手段と、全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを秘密の比率に従って分割し、所定の個数の部分データを生成する手段と、それぞれが複数の入力データに基づいて所定の演算を行う手段を有する複数の演算装置のうちの対応する演算装置への前記複数の入力データの1つとして、前記所定の個数の部分データのそれぞれを、保護された通信路により送信する手段とを備える。そして、前記複数の演算装置のそれぞれが、複数の前記データ入力装置からの部分データに基づいて前記所定の演算を行った結果を、前記複数の演算装置とは異なるサーバ装置が利用することにより、複数の前記データ入力装置により

取得された複数の前記オリジナルデータに基づく統計処理の結果が、該オリジナルデータが秘匿されたまま求められる。

[0070] この構成によれば、秘匿すべきオリジナルデータが漏洩するリスクを低減しつつ、サーバ装置が、複数の演算装置に部分データを入力とする演算を行わせ、その結果を利用することで、オリジナルデータの集合についての統計処理の結果を得ることが可能になる。

[0071] 上記のデータ入力装置において、送信した前記所定の個数の部分データを、それぞれ対応する前記演算装置に、許可された者からのみアクセスできるように保存させる手段と、取得した前記オリジナルデータの記憶を消去する手段とをさらに備えるようにし、前記オリジナルデータは、前記許可された者によりそれぞれ対応する前記演算装置に取得された前記所定の個数の部分データに基づいて復元されるものとしてもよい。

[0072] これにより、本来の保有者が後でオリジナルデータを取得したい場合への備えを、データ入力装置においてオリジナルデータを保存するのではなく、複数の演算装置に分散して保存されている部分データを取得してオリジナルデータを復元可能とすることにより、実現することができるため、秘匿すべき情報が漏洩するリスクを確実に低減することが可能になる。

[0073] 上記のデータ入力装置において、前記サーバ装置にアクセスするための情報を記憶する手段と、前記対応する演算装置を特定するための情報を前記サーバ装置から受信する手段とをさらに備えるようにしてもよい。

[0074] これにより、データ入力装置は、サーバ装置にアクセスするための情報さえ記憶していれば、オリジナルデータを幾つに分割してどのように部分データを生成し、どの複数の演算装置へ渡すか等を、サーバ装置からの指定に従って行うことが可能になる。

[0075] 上記のデータ入力装置において、前記部分データに、システムにおいて一意となる識別情報を付与する手段と、前記識別情報に基づいて求められる値が、前記複数の演算装置のそれぞれが担当する範囲のうちのいずれの範囲に属するかに従って、前記対応する演算装置を特定する手段とをさらに備える

ようにしてもよい。

[0076] これにより、データ入力装置は、自身で、各部分データの送信先となる演算装置を決定することが可能になり、サーバ装置が、各オリジナルデータに関係する情報を取り扱わないようにすることができるとともに、一つのオリジナルデータから得られる各部分データが、それぞれ異なる演算装置に送信されるようにすることも可能になり、さらなるデータの安全性を実現することができる。

[0077] 上記のデータ入力装置において、一つのオリジナルデータから得られる部分データの全てがいずれかの演算装置に受信されたことを確認した後に、確認済みであることを示す情報をいずれかの演算装置へ送信して登録する手段をさらに備えるようにしてもよい。

[0078] この構成と、以下に例示する各演算装置の構成とにより、各演算装置が有する部分データのうち、統計処理の対象に含めてしまうと結果に誤差等が生じる部分データを、演算から外すことが可能になる。

[0079] 本発明の原理に従う一つの例に係る演算装置は、秘匿されるべきオリジナルデータを取得することなく、複数の前記オリジナルデータに基づく統計処理の結果を提供するサービスのためのサーバ装置と通信する手段と、それぞれが前記オリジナルデータを自装置の内部に秘匿する手段を有する複数のデータ入力装置から、複数の前記オリジナルデータのそれぞれに属する部分データを受信する手段と、複数の入力データに基づいて所定の演算を行う手段とを備える。そして、前記サーバ装置は、複数の前記演算装置からの演算結果に基づいて、所定の統計処理を行うものであり、前記演算装置は、前記複数のデータ入力装置から受信した複数の前記部分データのうち、該部分データに対応して、一つのオリジナルデータから得られる部分データの全てがいずれかの演算装置に受信されたことが確認済みであることを示す情報が登録されているものを、前記入力データとして選択する手段と、選択された前記入力データについての前記所定の演算の結果を、前記サーバ装置へ送信する手段とをさらに備える。

[0080] 上述したデータ秘匿型統計処理システム、統計処理結果提供サーバ装置及びデータ入力装置の発明は、それぞれ互いに他の要素の発明としても、本システム全体もしくは個々の装置において行なわれる方法の発明としても、汎用のコンピュータシステムを本システムとして動作させるためのプログラム（又はそのプログラムを記録した記録媒体）の発明としても、汎用のコンピュータを個々の装置として動作させるためのプログラム（又はそのプログラムを記録した記録媒体）の発明としても、勿論成立するものである。以下に、そのうちの幾つかを例示する。

[0081] 本発明の原理に従う一つの例に係るプログラムは、他のコンピュータと通信する機能を有するコンピュータを、データ秘匿型統計処理システムにおけるデータ処理装置として動作させるためのプログラムである。前記他のコンピュータとして、それぞれが複数の入力データに基づいて所定の演算を行う手段を有する複数の演算装置があり、前記データ処理装置は、秘匿されるべきオリジナルデータを取得することなく、複数の前記オリジナルデータに基づく統計処理の結果を提供するものである。そして、前記プログラムは、前記コンピュータに、前記複数の演算装置のそれぞれに、前記オリジナルデータの部分データを前記入力データとして演算を行わせ、該演算の結果を取得する手段と、前記複数の演算装置からの演算結果に基づいて、所定の統計処理を行う手段を備えさせるものであり、複数の前記部分データは、全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを秘密の比率に従って分割することにより、生成されたものである。

[0082] 本発明の原理に従う別の例に係るプログラムは、秘匿すべきオリジナルデータを取得する機能と他のコンピュータと通信する機能とを有するコンピュータを、データ秘匿型統計処理システムにおけるデータ入力装置として動作させるためのプログラムである。前記他のコンピュータとして、それぞれが複数の入力データに基づいて所定の演算を行う手段を有する複数の演算装置がある。そして、前記プログラムは、前記コンピュータに、全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを

秘密の比率に従って分割し、所定の個数の部分データを生成する手段と、前記所定の個数の部分データのそれぞれを、前記複数の演算装置のうちの対応する演算装置への前記複数の入力データの1つとして、保護された通信路により送信する手段とを備えさせるものであり、前記複数の演算装置のそれぞれが、複数の前記データ入力装置からの部分データに基づいて前記所定の演算を行った結果を、前記複数の演算装置とは異なるサーバ装置が利用することにより、複数の前記データ入力装置により取得された複数の前記オリジナルデータに基づく統計処理の結果が、該オリジナルデータが秘匿されたまま求められる。

[0083] 本発明の原理に従うさらに別の例に係るプログラムは、他のコンピュータと通信する機能を有するコンピュータを、データ秘匿型統計処理システムにおける複数の演算装置のうちのひとつとして動作させるためのプログラムである。前記他のコンピュータとして、秘匿されるべきオリジナルデータを取得することなく、複数の前記オリジナルデータに基づく統計処理の結果を提供するサービスのためのサーバ装置と、それぞれが前記オリジナルデータを自装置の内部に秘匿する手段を有する複数のデータ入力装置とがある。そして、前記プログラムは、前記コンピュータに、前記複数のデータ入力装置から、複数の前記オリジナルデータのそれぞれに属する部分データを受信する手段と、複数の入力データに基づいて所定の演算を行う手段と、前記複数のデータ入力装置から受信した複数の前記部分データのうち、該部分データに対応して、一つのオリジナルデータから得られる部分データの全てがいずれかの演算装置に受信されたことが確認済みであることを示す情報が登録されているものを、前記入力データとして選択する手段と、選択された前記入力データについての前記所定の演算の結果を、前記サーバ装置へ送信する手段とを備えさせるものであり、前記サーバ装置は、前記複数の演算装置からの演算結果に基づいて、所定の統計処理を行うものである。

[0084] 本発明の原理に従う一つの例に係る統計処理結果提供サービス方法は、秘匿すべきオリジナルデータを取得する手段を備える複数のデータ入力装置の

それぞれが、全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを秘密の比率に従って分割して得られた所定の個数の部分データを出し、複数の入力データに基づいて所定の演算を行う手段を備える複数の演算装置のそれぞれが、複数の前記データ入力装置のそれぞれから出力された前記部分データを前記入力データとして前記演算を行った結果を出し、データ処理装置が、前記複数の演算装置のそれぞれから出力された前記演算の結果を利用することにより、前記複数のデータ入力装置により取得された複数のオリジナルデータに基づく統計処理の結果を、該オリジナルデータを取得することなく求める。

[0085] 以下、本発明の実施の形態について、例示のために、図面を用いて説明する。本システムは、プライバシー保護を考慮したクラウド型のデータ処理を行うためのシステムである。

[0086] 現在、多数のセンサやＩＣカードが普及しており、例えば、車は数億台、スマートフォンは十億台超、センサは数十億～数兆個という、膨大な数のデータ生成源（本システムにおけるデータ入力デバイスとなり得るもの）が存在する。これらを対象とするＭ２Ｍ（マシンｔｏマシン）型のサービスも、種々考案されている。

[0087] これらのサービスの多くが、データの蓄積と解析処理を、データの本来の所有者ではない第三者がリソースを提供するクラウドを利用して、行うことが想定される。そうすると、クラウドにおいて扱われるデータには、プライバシー情報が大量に含まれることになり、データがクラウドの外部へ流出したときの情報漏洩のリスクが、より深刻になる。したがって、クラウドを利用する際には、情報漏洩のリスクを低減するために、クラウドにおけるデータの蓄積から解析処理まで通して、クラウド上のデータを秘匿化した状態にすることが、強く望まれる。

[0088] そこで、本システムでは、データ生成源からオリジナルデータを集める際に、オリジナルデータを秘匿化できるような分割（以下、「秘匿分割」ということがある）を行う。そして、オリジナルデータはどこへも渡さず、分割

されたデータを複数のクラウドへ渡して、蓄積及び解析処理をさせる。このようにすれば、単一のクラウドからデータが流出しても、そこからオリジナルデータを復元することはできなくなる。

[0089] そして、本システムでは、各クラウドで個別に、統計解析処理を行い、クラウドとは別の解析事業者（「統計処理結果提供サービス事業者」ともいう）が、各クラウドの処理結果を集めて、本来の統計処理の結果を得る。ここで、各クラウドサービスを提供する事業者は、複数のクラウドから一度にデータが流出する可能性を低くするためにも、複数のクラウド上のデータを合わせてオリジナルデータを導き出そうとする試みをさせないためにも、別々の事業者とすることが好ましい。どのクラウドサービスを利用するかは、解析事業者もしくはデータ生成源の所有者が決めればよい。

[0090] なお、クラウドサービスでは一時的な計算資源の利用も可能であるため、本システムを、データの永続保存が不要（オリジナルデータの復元が不要）な場合に適用する際には、クラウドサービスを用いて必要なときに必要なだけの計算資源を確保し、演算処理を行ったのち、不要になった計算資源を開放する（その演算処理のために保存した部分データを全て消去する）ようにしてもよい。これにより、より情報漏洩に対する安全性を高めることができることに加えて、物理的に冗長な計算資源を維持せずに済むようにすることが可能である。

[0091] 解析事業者は、データ生成源の所有者とは異なる者でもよいし、例えば、一つの企業が自身の所有する多数のデータ生成源からのデータについて第三者のクラウドサービスを利用して蓄積及び解析処理を行う場合には、解析事業者がデータ生成源を所有する企業自身であってもよい。各データ生成源の所有者がそれぞれ異なる個人であって、解析事業者とも、解析事業者による統計処理結果を提供される利用者企業とも、異なるという応用例もあり得る。

[0092] 本システムでは、このようにオリジナルデータを秘匿分割して複数のクラウドに分散させた状態を保ちながら処理を行って、多数のオリジナルデータ

の総和、2乗和、内積等を求めることができる。例えば、総和と2乗和が求められるだけでも、統計処理として、平均値や分散値を求めたり、基本的な推定や検定を行ったりすることができるため、多様な応用が可能である。しかも、オリジナルデータをどこにも存在させず、秘匿分割された状態のまま、且つ、一つのオリジナルデータから秘匿分割により生成された複数のデータが一箇所に集まることなく分散された状態のまま、統計処理の結果を求められるため、安全性を十分に高くすることが可能である。

[0093] 図1は、各オリジナルデータを2つに分割して、N個のオリジナルデータの総和を求める本システムの例を示している。図中、説明のため、各々のデータ入力デバイス10-1～10-Nが、各々のオリジナルデータ $x_1 \sim x_N$ を分割して、クラウドサービス設備30-1及び30-2へアップロードするように描かれているが、本システムでは、1つのデータ入力デバイスが複数のオリジナルデータについて、取得、秘匿分割、アップロードを行うことも、勿論可能である。なお、Nは、2以上の整数であり、億や兆の単位の数とすることも可能である。

[0094] 各々のデータ入力デバイス10-iは、オリジナルデータ x_i を取得すると、 $x_i = x_{1i} + x_{2i}$ を満たすように x_i を分割する。どのような比率で分割するかは、その都度、デバイス内で乱数を発生させる等により、ランダムに定め、その比率は、秘密とする（この処理を、「ランダムシェアによる秘匿分割」という）。

[0095] これにより、個々の x_{1i} 、 x_{2i} は、 x_i について完全秘匿性を持つことになる（このことを、「 $H(x_i | x_{1i}) = H(x_i) \ \& \ H(x_i | x_{2i}) = H(x_i)$ 」と表す）。これにより、単一のクラウドにおけるデータ流出ではオリジナルデータを復元できないことが、担保される。

[0096] そして、各々のデータ入力デバイス10-iは、第1のクラウドサービス設備30-1へ部分データ x_{1i} をアップロードし、第2のクラウドサービス設備30-2へ部分データ x_{2i} をアップロードする。

[0097] 各々のクラウドサービス設備30-jは、アップロードされてきたデータ

を保存する。各データ入力デバイスからのアップロードのタイミングはそれぞれ任意の時点でよく、ある時点で、第1のクラウドサービス設備30-1にはN個の部分データ $\{x_{11}, x_{12}, \dots, x_{1N}\}$ が保存されており、第2のクラウドサービス設備30-2にはN個の部分データ $\{x_{21}, x_{22}, \dots, x_{2N}\}$ が保存されている状態になる。

[0098] この時点で、第1のクラウドサービス設備30-1は、N個の部分データ x_{1i} の総和を計算した結果 $f(X_1)$ を統計処理結果提供サーバ50へ送信し、第2のクラウドサービス設備30-2は、N個の部分データ x_{2i} の総和を計算した結果 $f(X_2)$ を統計処理結果提供サーバ50へ送信する。Nが膨大な数である場合、クラウド上の計算機リソースを使って処理ができることも、重要な利点となる。

[0099] 統計処理結果提供サーバ50は、送信されてきた結果について、総和を求める処理を行う。「 $f(X_1) + f(X_2)$ 」の値は、 $(x_{1i} + x_{2i})$ の i を1からNまで合計した値と等しくなるから、オリジナルデータ x_i の総和を求めたことになる。本システムにより提供されるサービスの利用者は、統計解析の結果のみを参照する。

[0100] なお、統計処理結果提供サーバ50は、各クラウドからN個の部分データについて計算処理を行った結果である $f(X_i)$ のみを取得し、個々の部分データには関知しないため、統計処理結果提供サーバ50を運営する解析事業者に対しても、オリジナルデータの秘匿性を高く保つことが可能である。

[0101] 図1は、各オリジナルデータを2個に分割する例であるが、図2は、各オリジナルデータを m 個（2より多い数）に分割して、N個のオリジナルデータの総和を求める本システムの例を示している。図2の場合、 m 箇所の独立した異なるクラウド上で、分散して処理を行うことになる。

[0102] 各々のデータ入力デバイス10- i は、オリジナルデータ x_i を取得すると、 $x_i = x_{1i} + x_{2i} + \dots + x_{mi}$ を満たすように x_i を分割する。どのような比率で分割するかは、その都度、デバイス内で乱数を発生させる等により、ランダムに定め、その比率は、秘密とする。

- [0103] このランダムシェアによる秘匿分割により、個々の $x_{1i}, x_{2i}, \dots, x_{mi}$ は、 x_i について完全秘匿性を持ち、また、例えば、 $x_{1i} \sim x_{(m-1)i}$ の値が分かっても、 x_{mi} の値が分からなければ、 x_i を復元できないことから、同時に $(m-1)$ 箇所のデータ流出があっても、秘匿性は維持されることになる。
- [0104] そして、各々のデータ入力デバイス $10-i$ は、 m 個のクラウドサービス設備 $30-j$ のそれぞれへ対応する部分データ x_{ji} をアップロードする。アップロードのタイミングはデータ入力デバイス毎に独立としてよいが、ある時点で、いずれのクラウドサービス設備 $30-j$ にも N 個の部分データ $\{x_{j1}, x_{j2}, \dots, x_{jN}\}$ が保存されている状態になる。
- [0105] この時点で、各々のクラウドサービス設備 $30-j$ は、 N 個の部分データ x_{ji} の総和を計算した結果 $f(X_j)$ を統計処理結果提供サーバ 50 へ送信する。統計処理結果提供サーバ 50 は、送信されてきた結果について、総和を求める処理を行う。「 $f(X_1) + f(X_2) + \dots + f(X_m)$ 」の値は、 $(x_{1i} + x_{2i} + \dots + x_{mi})$ の i を 1 から N まで合計した値と等しくなるから、オリジナルデータ x_i の総和を求めたことになる。
- [0106] 図3は、各オリジナルデータを2つに分割して、 N 個のオリジナルデータの2乗和を求める本システムの例を示している。図1では、 i が 1 から N までの x_i の総和を求める処理を $f(X_i)$ と記述しているが、図3及び図4では、同じ総和を求める処理を $f_{\Sigma}(X_i)$ と表し、 i が 1 から N までの x_i の2乗和を求める処理を $f_s(X_i)$ と記述する。
- [0107] また、図3では、統計処理結果提供サーバ 50 が、第1のクラウドサービス設備 $30-1$ からの2乗和 $f_s(X_1)$ と、第2のクラウドサービス設備 $30-2$ からの2乗和 $f_s(X_2)$ と、第3のクラウドサービス設備 $30-3$ からの総和 $f_{\Sigma}(X_{12})$ とを用いて、 N 個のオリジナルデータの2乗和 $f_s(X)$ を求める点を説明しているが、同時に、第1のクラウドサービス設備 $30-1$ からの総和 $f_{\Sigma}(X_1)$ と、第2のクラウドサービス設備 $30-2$ からの総和 $f_{\Sigma}(X_2)$ とを用いて、 N 個のオリジナルデータの総和 $f_{\Sigma}(X)$ を求めることも可能である。

- [0108] 各々のデータ入力デバイス $10-i$ は、オリジナルデータ x_i を取得すると、ランダムシェアによる秘匿分割を行い、 x_i は $x_i = x_{1i} + x_{2i}$ を満たすように分割される。統計処理の結果として2乗和を求めたい場合、各データ入力デバイス $10-i$ はさらに、 x_{1i} と x_{2i} を乗算した値を求めて、 x_{1i} 、 x_{2i} 、 $x_{1i} \times x_{2i}$ の3つを、 x_i の部分データとして生成する。図3のように $x_{1i} \times x_{2i}$ をも生成してアップロードすべきか、図1のように x_{1i} と x_{2i} だけでよいかを、統計処理結果提供サーバ50が各データ入力デバイス $10-i$ に指示するようにしてもよい。
- [0109] そして、各々のデータ入力デバイス $10-i$ は、第1のクラウドサービス設備30-1へ部分データ x_{1i} をアップロードし、第2のクラウドサービス設備30-2へ部分データ x_{2i} をアップロードし、第3のクラウドサービス設備30-3へ部分データ $x_{1i} \times x_{2i}$ をアップロードする。この場合、3つのクラウドのうちの1つにおいてデータが流出しても、オリジナルデータは復元されない。
- [0110] 各々のクラウドサービス設備30-jは、アップロードされてきたデータを保存する。各データ入力デバイスからのアップロードのタイミングはそれぞれ任意の時点でよく、ある時点で、第1のクラウドサービス設備30-1にはN個の部分データ $\{x_{11}, x_{12}, \dots, x_{1N}\}$ が保存されており、第2のクラウドサービス設備30-2にはN個の部分データ $\{x_{21}, x_{22}, \dots, x_{2N}\}$ が保存されており、第3のクラウドサービス設備30-3にはN個の部分データ $\{x_{11} \times x_{21}, x_{12} \times x_{22}, \dots, x_{1N} \times x_{2N}\}$ が保存されている状態になる。
- [0111] この時点で、第1のクラウドサービス設備30-1は、N個の部分データ x_{1i} の総和と2乗和をそれぞれ計算した結果 $f_{\Sigma}(X_1)$ と $f_S(X_1)$ を統計処理結果提供サーバ50へ送信し、第2のクラウドサービス設備30-2は、N個の部分データ x_{2i} の総和と2乗和をそれぞれ計算した結果 $f_{\Sigma}(X_2)$ と $f_S(X_2)$ を統計処理結果提供サーバ50へ送信し、第3のクラウドサービス設備30-3は、N個の部分データ $x_{1i} \times x_{2i}$ の総和と2乗和をそれぞれ計算した結果 $f_{\Sigma}(X_{12})$ と $f_S(X_{12})$ を統計処理結果提供サーバ50へ送信する。

- [0112] 統計処理結果提供サーバ50は、送信されてきた結果の中から、 $f_s(X_1)$ と $f_s(X_2)$ と $f_{\Sigma}(X_{12})$ とを選択して、 $f_{\Sigma}(X_{12})$ については2倍した上で、これら全てを合計する処理を行う。「 $f_s(X_1) + 2f_{\Sigma}(X_{12}) + f_s(X_2)$ 」の値は、 $(x_{1i} + x_{2i})^2$ の i を1から N まで合計した値と等しくなるから、オリジナルデータ x_i^2 の総和（即ち x_i の2乗和）を求めたことになる。
- [0113] 図3の構成において、統計処理結果提供サーバ50が、送信されてきた結果の中から、 $f_{\Sigma}(X_1)$ と $f_{\Sigma}(X_2)$ とを選択して、総和を求める処理を行えば、オリジナルデータ x_i の総和が求められる。いずれの場合も、第3のクラウドからの $f_s(X_{12})$ の結果は用いられないし、2乗和のみを求める場合、第1及び第2クラウドからの $f_{\Sigma}(X_j)$ の結果は用いられない。また、図3の構成で総和のみを求める場合、第1及び第2クラウドからの $f_s(X_j)$ の結果は用いられず、第3のクラウドからのいずれの結果も用いられないことになる。
- [0114] 結果が用いられない計算処理を行うことはリソースの浪費とも捉えられるが、クラウド上には豊富な計算機リソースがあることに加え、統計処理結果提供サーバ50で行いたい統計処理の内容に関わらず、各クラウドにおける計算処理の内容を同一にすることには、以下の利点がある。
- [0115] 図3の構成では、各クラウドサービス設備30-jは、アップロードされてきたデータが、 x_i を分割した部分 x_{ji} であるのか、2つの部分を乗算した $x_{ji} \times x_{ki}$ であるのか、さらにいえばオリジナルデータなのか部分データなのかさえ、関知することなく、単に、入力されたデータに対して、 i が1から N までの総和及び2乗和を計算するという処理を、画一的に行う。よって、各クラウドにおいて行われる計算処理の内容から、統計処理結果提供サーバ50で行われる統計処理の内容や、各クラウドに保存されているデータの意味等を推測されることがなく、安全性をより高めることが可能である。
- [0116] 図3は、各オリジナルデータを2個に分割する例であるが、図4は、各オリジナルデータを m 個（2より多い数）に分割して、 N 個のオリジナルデータの2乗和を求める本システムの例を示している。図4の場合、 $2m$ 箇所の

独立した異なるクラウド上で分散して処理を行うことになる。この場合、 $2m$ 個のクラウドのうちの $(m-1)$ 個においてデータが流出しても、オリジナルデータは復元されない。

- [0117] 各々のデータ入力デバイス $10-i$ は、オリジナルデータ x_i を取得すると、ランダムシェアによる秘匿分割を行い、 $x_i = x_{1i} + x_{2i} + \dots + x_{mi}$ を満たすように x_i を分割する。そして、まず、 m 個の部分データ x_{ji} ($j = 1, 2, \dots, m$) を生成する。
- [0118] 各データ入力デバイス $10-i$ はさらに、 x_{ji} の値と x_{ji} 以外の x_{ki} を合計した値とを乗算した値を x'_{ji} として、 m 個の部分データ x'_{ji} ($j = 1, 2, \dots, m$) を生成する。例えば、 $m=4$ であれば、 $x'_{1i} = x_{1i}x_{2i} + x_{1i}x_{3i} + x_{1i}x_{4i}$ と、 $x'_{2i} = x_{2i}x_{1i} + x_{2i}x_{3i} + x_{2i}x_{4i}$ と、 $x'_{3i} = x_{3i}x_{1i} + x_{3i}x_{2i} + x_{3i}x_{4i}$ と、 $x'_{4i} = x_{4i}x_{1i} + x_{4i}x_{2i} + x_{4i}x_{3i}$ とを生成する。
- [0119] そして、各々のデータ入力デバイス $10-i$ は、 m 個のクラウドサービス設備 $30-j$ ($j = 1, 2, \dots, m$) のそれぞれへ対応する部分データ x_{ji} をアップロードし、さらに m 個のクラウドサービス設備 $30-j$ ($j = m+1, m+2, \dots, m+m$) のそれぞれへ対応する部分データ x'_{ji} をアップロードする。アップロードのタイミングはデータ入力デバイス毎に独立としてよいが、ある時点で、いずれのクラウドサービス設備 $30-j$ にも i が 1 から N までの N 個の部分データが保存されている状態になる。
- [0120] この時点で、各々のクラウドサービス設備 $30-j$ は、 N 個の部分データ ($j = 1 \sim m$ については x_{ji} 、 $j = m+1 \sim 2m$ については x'_{ji} になるが、各クラウドはその相違を関知しない) の総和及び2乗和をそれぞれ計算した結果 ($j = 1 \sim m$ については $f_{\Sigma}(X_i)$ と $f_s(X_i)$ 、 $j = m+1 \sim 2m$ については $f_{\Sigma}(X'_{ji})$ と $f_s(X'_{ji})$ になるが、各クラウドはその相違を関知しない) を統計処理結果提供サーバ 50 へ送信する。
- [0121] 統計処理結果提供サーバ 50 は、送信されてきた結果の中から、 $j = 1 \sim m$ のクラウドからの結果については $f_s(X_i)$ を、 $j = m+1 \sim 2m$ のクラウドからの結果については $f_{\Sigma}(X'_{ji})$ を選択して、これら全てを合計する処理

を行う。「 $f_s(X_1) + f_s(X_2) + \dots + f_s(X_m) + f_\Sigma(X'_1) + f_\Sigma(X'_2) + \dots + f_\Sigma(X'_m)$ 」の値は、 $(x_{1i} + x_{2i} + \dots + x_{mi})^2$ の i を1から N まで合計した値と等しくなるから、オリジナルデータ x_i^2 の総和（即ち x_i の2乗和）を求めたことになる。

[0122] 図4の構成でも、図3と同様に、オリジナルデータ x_i の総和と2乗和の両方を求めることができ、各クラウドから出力される結果のうち、 $j = 1 \sim m$ のクラウドからの $f_\Sigma(X_i)$ が総和に利用され、 $j = 1 \sim m$ のクラウドからの $f_s(X_i)$ と $j = m + 1 \sim 2m$ のクラウドからの $f_\Sigma(X'_j)$ とが2乗和に利用されることになる。

[0123] 上述したように総和と2乗和が得られると、以下に例示するように、基本的な統計解析手法への幅広い応用が可能になる。

[0124] 標本平均 m は、 $m = \sigma / N = f_\Sigma(X) / N$ により求めることができ、最尤推定は、母集団が正規分布に従うとき、母集団に最尤平均値 $= m$ として行うことができる。

[0125] 標本分散 s^2 は、 $s^2 = (S - \sigma^2) / N = (f_s(X) - \{f_\Sigma(X)\}^2) / N$ により求めることができ、標準偏差 s は、標本分散 s^2 の正の平方根として求めることができる。

[0126] T分布を用いた区間推定は、 $T = (m - \mu) / (s / N^{1/2})$ が自由度 $(N - 1)$ のt分布に従うことから、例えば、母平均 μ の信頼度95%の信頼区間を、

$$m - 1.96 \times s / N^{1/2} \leq \mu \leq m + 1.96 \times s / N^{1/2}$$

と推定することができる。以上により、母集団の平均を推定することが可能になる。

[0127] 母比率の信頼区間の推定は、標本比率 r （例えば、 N 人のうち r 人がYESと答えた）が、 $r = f_\Sigma(X)$ により求められると、母比率 R の95%信頼区間を、

$$r - 1.96 \times (r(1-r) / N)^{1/2} \leq R \leq r + 1.96 \times (r(1-r) / N)^{1/2}$$

と推定することができる。これは、YES／NOや選択式（又は機械の on／off）の統計データに応用することができる。

- [0128] 母分散の推定は、母集団は分散 σ^2 の正規分布で、N個の標本の不偏分散を s^2 とすると、 $Z = (N-1) \times s^2 / \sigma^2$ が自由度 $(N-1)$ の χ^2 分布に従うことから、母分散 σ^2 につき、同分布の下側 95% 点 k_1 、上側 95% 点 k_2 との関係が、

$$(N-1) \times s^2 / k_2 \leq \sigma^2 \leq (N-1) \times s^2 / k_1$$

となることが推定することができる。これにより、母集団のばらつきを推定することが可能になる。

- [0129] 母平均の検定（t 検定）は、 $T = (m - \mu) / (s / N^{1/2})$ が自由度 $(N-1)$ の t 分布に従うことを応用して、行うことができる。また、母集団 A，B に関する母平均の差の検定は、 $T = (m_A - m_B) / (Z_1^{1/2} \times Z_2^{1/2})$ が自由度 $(N_A + N_B - 2)$ の t 分布に従うことを応用して、行うことができる。但し、

$$Z_1 = 1 / N_A + 1 / N_B$$

$$Z_2 = ((N_A - 1) \times s_A^2 + (N_B - 1) \times s_B^2) / (N_A + N_B - 2)$$

である。これにより、母集団の平均を検定することが可能になる。

- [0130] 母比率の検定（ χ^2 検定）は、 $\chi^2 = (N-1) \times s^2 / \sigma^2$ は自由度 $(N-1)$ の χ^2 分布に従うことを応用して、行うことができる。また、母集団 A，B の母分散の比較検定（F 検定）は、 $F = (s_A^2 / \sigma_A^2) / (s_B^2 / \sigma_B^2)$ が自由度 k_A, k_B の F 分布に従うことから、母分散が同じであると仮定すると、 $F = s_A^2 / s_B^2$ が自由度 $N_A - 1, N_B - 1$ の F 分布に従うことを応用して、行うことができる。これにより、母集団のばらつきを検定することが可能になる。

- [0131] 一次元配置の分散分析は、例えば、施策 1，2，…，k に効果の差があるかどうかを検討するために行うことができ、全体平均 $m = \sum_i \sum_j x_{ij} / N$ （但し、 $N = \sum_i N_i$ ）、グループ平均 $m_i = \sum_j x_{ij} / N_i$ 、グループ間変動 $Q_1 = \sum_i (m_i - m)^2$ 、グループ内変動 $Q_2 = \sum_i \sum_j (x_{ij} - m_i)^2$ であるとき、 $F = Q_1 / Q_2$ は自由度 $(k-1), k \times (N-1)$ の F 分布に従うことを応用して、行

うことができる。これは、例えば、施策、投薬、改修、改善、キャンペーン、広告等の取り組みの効果を確認する際に有効である。

[0132] 二次元配置の分散分析は、上述した一次元配置の分散分析の簡単な拡張から、繰り返し有りと繰り返し無しの両方のケースについて、行うことができる。これは、複数の取り組みの組合せによる効果を確認する際に有効である。

[0133] 以上、一つの要素に対する統計解析について説明したが、本システムは、複数の要素に対する統計解析にも、適用可能である。例えば、2つの要素への応用として、内積、共分散、相関係数、さらには、回帰方程式、決定係数等を求めることも可能である。

[0134] 図5は、2つの要素の各オリジナルデータ x_i , y_i をそれぞれ2つに分割して、N対のオリジナルデータの内積を求める本システムの例を示している。図5は、各オリジナルデータを2個に分割する例であるが、各オリジナルデータを m 個（2より多い数）に分割して、 m^2 個の独立した異なるクラウド上で分散して処理を行うことにより、N対のオリジナルデータの内積を求めることも、勿論可能である。

[0135] 第1の要素に属するオリジナルデータ x_i を取得する各々のデータ入力デバイス $10-i$ は、取得したオリジナルデータ x_i に対してランダムシェアによる秘匿分割を行い、 x_i は $x_i = x_{1i} + x_{2i}$ を満たすように分割される。第2の要素に属するオリジナルデータ y_i を取得する各々のデータ入力デバイス $20-i$ は、取得したオリジナルデータ y_i に対してランダムシェアによる秘匿分割を行い、 y_i は $y_i = y_{1i} + y_{2i}$ を満たすように分割される。

[0136] そして、各々のデータ入力デバイス $10-i$ は、第1及び第2のクラウドサービス設備 $30-1$, 2 へ部分データ x_{1i} をアップロードし、第3及び第4のクラウドサービス設備 $30-3$, 4 へ部分データ x_{2i} をアップロードし、各々のデータ入力デバイス $20-i$ は、第1及び第3のクラウドサービス設備 $30-1$, 3 へ部分データ y_{1i} をアップロードし、第2及び第4のクラウドサービス設備 $30-2$, 4 へ部分データ y_{2i} をアップロードする。

[0137] 各々のクラウドサービス設備 30-j は、アップロードされてきたデータを保存する。各データ入力デバイスからのアップロードのタイミングはそれぞれ任意の時点で行く、ある時点で、第 1 のクラウドサービス設備 30-1 には第 1 の要素の N 個の部分データ $\{x_{11}, x_{12}, \dots, x_{1N}\}$ 及び第 2 の要素の N 個の部分データ $\{y_{11}, y_{12}, \dots, y_{1N}\}$ が保存されており、第 2 のクラウドサービス設備 30-2 には第 1 の要素の N 個の部分データ $\{x_{11}, x_{12}, \dots, x_{1N}\}$ 及び第 2 の要素の N 個の部分データ $\{y_{21}, y_{22}, \dots, y_{2N}\}$ が保存されており、第 3 のクラウドサービス設備 30-3 には第 1 の要素の N 個の部分データ $\{x_{21}, x_{22}, \dots, x_{2N}\}$ 及び第 2 の要素の N 個の部分データ $\{y_{11}, y_{12}, \dots, y_{1N}\}$ が保存されており、第 4 のクラウドサービス設備 30-3 には第 1 の要素の N 個の部分データ $\{x_{21}, x_{22}, \dots, x_{2N}\}$ 及び第 2 の要素の N 個の部分データ $\{y_{21}, y_{22}, \dots, y_{2N}\}$ が保存されている状態になる。

[0138] この時点で、第 1 のクラウドサービス設備 30-1 は、N 対の部分データ x_{1i} と y_{1i} の内積を計算した結果 $f_p(X_1, Y_1)$ を統計処理結果提供サーバ 50 へ送信し、第 2 のクラウドサービス設備 30-2 は、N 対の部分データ x_{1i} と y_{2i} の内積を計算した結果 $f_p(X_1, Y_2)$ を統計処理結果提供サーバ 50 へ送信し、第 3 のクラウドサービス設備 30-3 は、N 対の部分データ x_{2i} と y_{1i} の内積を計算した結果 $f_p(X_2, Y_1)$ を統計処理結果提供サーバ 50 へ送信し、第 4 のクラウドサービス設備 30-4 は、N 対の部分データ x_{2i} と y_{2i} の内積を計算した結果 $f_p(X_2, Y_2)$ を統計処理結果提供サーバ 50 へ送信する。

[0139] 統計処理結果提供サーバ 50 は、送信されてきた結果の全てを合計する処理を行う。「 $f_p(X_1, Y_1) + f_p(X_1, Y_2) + f_p(X_2, Y_1) + f_p(X_2, Y_2)$ 」の値は、 $(x_{1i} + x_{2i})$ と $(y_{1i} + y_{2i})$ を乗算した値を i が 1 のものから N のものまで合計した値と等しくなるから、オリジナルデータ x_i と y_i の内積を求めたことになる。

[0140] 上述したように内積、さらに必要に応じて総和及び 2 乗和が得られると、以下に例示するように、様々な統計解析手法への幅広い応用が可能になる。

[0141] 共分散 Cov_{XY} は、 m_X 、 m_Y をそれぞれ X 、 Y の標本平均として、

$$Cov_{XY} = 1/N \times \sum (x_i - m_X)(y_i - m_Y)$$

であり、 $m_X = f_{\Sigma}(X)/N$ 、 $m_Y = f_{\Sigma}(Y)/N$ であるから、

$$Cov_{XY} = (f_P(X, Y) - f_{\Sigma}(X)f_{\Sigma}(Y))/N$$

として求められる。

[0142] 相関係数 CC_{XY} は、 s_X 、 s_Y をそれぞれ X 、 Y の標本偏差として、

$$CC_{XY} = Cov_{XY} / s_X s_Y$$

として求められる。ここで、 $s_X = [(f_S(X) - \{f_{\Sigma}(X)\}^2)/N]^{1/2}$

、 $s_Y = [(f_S(Y) - \{f_{\Sigma}(Y)\}^2)/N]^{1/2}$ である。

[0143] そして、上述したように、平均 m_X 、 m_Y 、分散 s_X^2 、 s_Y^2 、共分散 Cov_{XY} が求められれば、回帰分析における 1 次式の係数を求める公式への応用が可能であり、変動、残差平方和、決定係数も計算可能である。

[0144] 図 6 は、図 1 ～ 5 を参照して説明した本システムが採り得る構成の一例を示している。データ入力デバイス 10-1 ～ 10-N（図示しないが、内積を求める場合の 20-1 ～ 20-N も、同様の構成を有する）と、クラウドサービス設備 30-1 ～ 30-M と、統計処理結果提供サーバ 50 とが、ネットワーク 40（例えば、インターネット）を介して接続されている。

[0145] なお、各データ入力デバイス 10 と各クラウドサービス設備 30 との間、各クラウドサービス設備 30 と統計処理結果提供サーバ 50 との間、統計処理結果提供サーバ 50 と各データ入力デバイス 10 との間に、それぞれ別個の通信網（例えば、無線網と有線網等）がある構成としてもよい。

[0146] それぞれの間の通信の安全性に関しては、現存する十分な安全性を持った通信の暗号化を行う。特に、各データ入力デバイス 10 と各クラウドサービス設備 30 との間は、個々の通信は、分割されたデータしか含んでいないが、あるデータ入力デバイスから m 個のクラウドサービス設備への通信の全てを傍受すれば、オリジナルデータが復元できてしまうことから、例えば、オンラインショッピングや電子決済、商取引、ネットバンキング等などで利用されているのと同程度の安全性を持った暗号化技術を用いることが好ましい

。

[0147] 図6に示すとおり、各データ入力デバイス10は、データ取得部110と、取得したオリジナルデータを秘匿分割する秘匿分割部120と、秘匿分割により得られた部分データを暗号化された通信路で各クラウドサービス設備30へアップロードするアップロード部130を備える。データ取得部110は、機器が自動的にオリジナルデータを生成するものでもよいし、人間がオリジナルデータを入力するものでもよいし、別のデータベース等からオリジナルデータを抽出するものでもよい。

[0148] 各データ入力デバイス10の備える制御部140は、統計処理結果提供サーバ50中の管理部（管理サーバ）500からの指示に従って、秘匿分割部120におけるデータの分割数や生成すべき部分データの種別を制御する。制御部140はまた、管理サーバ500からの指示に従って、アップロード部130における各部分データのアップロード先を制御する。

[0149] 但し、アップロード先となるクラウドサービス設備が予め決まっている場合は、統計処理結果提供サーバ50と通信することなく、制御部140の内部に埋め込まれている制御情報に従って、これらの制御を行えばよい。

[0150] 各クラウドサービス設備30は、各データ入力デバイス10からアップロードされたデータを保存するデータ保存部310と、保存された多数の部分データについて総和（322）、2乗和（324）、内積（326）等の演算処理を行う計算部320を備える。いずれの演算処理も、データ入力デバイスの数Nに対して計算量 $O(N)$ で演算が可能であり、億や兆のような単位の大きな数のNに対しても、実用的なレベルでシステムをスケール（拡張）することができる。

[0151] なお、計算部320は、本システムの用途に応じて必要な演算処理だけを備えれば十分であり、例えば、内積を求める用途には使わないことが予め決まっている場合は、内積演算部は備えなくてよい。あるいは、用途の拡大に備えて、計算部320に、種々の演算部を構成可能にしておき、管理サーバ500からの指示に従って、統計処理毎に用いる演算部を選択するようにし

てもよい。

- [0152] 各クラウドサービス設備 30 の備える制御部 330 は、統計処理結果提供サーバ 50 中の管理部（管理サーバ）500 からの指示に従って、計算部 320 が所定の演算処理を行うタイミングと、その演算処理の対象としてデータ保存部 310 から読み出すべきデータとを特定する。
- [0153] 各データ入力デバイス 10 は、例えば、コンピューティング機能を有する機器に本方式のためのプログラムをインストールすることにより、構成される。その機器は、汎用のコンピュータでも、プログラムが予め組み込まれて製造される専用の機器でもよい。秘匿分割前にオリジナルデータを一時記憶している部分や、秘匿分割のための秘密の比率を使用する部分等については、特に、ハードウェア又はソフトウェア上のセキュリティを高めたモジュール内に設けるようにしてもよい。
- [0154] 各データ入力デバイス 10 が、専用の機器で、記憶容量が少ない場合等には、機器内に埋め込んでおく初期情報を最小限にするため、統計処理を司るマネージャ（管理サーバ 500）のアドレス（URL もしくは IP アドレス等）と、マネージャとの通信を暗号化するための鍵（公開鍵方式又は共通鍵方式）とを初期情報とし、各クラウド 30 のアドレス等は、マネージャを利用して取得するようにしてもよい。
- [0155] 各クラウドサービス設備 30 は、一般に提供されているクラウドサービスの設備を利用して、実現することができる。
- [0156] 統計処理結果提供サーバ 50 は、例えば、汎用のサーバに本方式のためのプログラムをインストールすることにより、構成可能であり、統計処理結果提供サービス自体を、クラウド上の計算サービスとして実現してもよい。
- [0157] 図 7 は、統計処理結果提供サーバ 50 の内部構成の一例を示している。統計処理結果提供サーバ 50 は、統計処理部 570 とともに、各データ入力デバイス 10 や各クラウドサービス設備 30 を制御する機能を備える管理部（管理サーバ）500 と、その統計処理の結果を利用者へ提供するための結果提供インタフェース 590 を備える。

- [0158] 統計処理結果提供サーバ50が、独立の複数の利用者に対して結果を提供するために、独立の複数の統計処理を行うことを可能とする場合、各々の統計処理について、管理サーバ500の機能が設けられることになり、その各々をマネージャと呼ぶ。例えば、各マネージャに異なるURLを割り当てる等により、マネージャの識別が可能である。
- [0159] 図6及び後述する図7における各部の機能は、ハードウェア又はソフトウェア、もしくはハードウェアとソフトウェアの組み合わせにより、実現することができる。複数の統計処理が並存する場合、着目する統計処理1を司るマネージャ50-1が、管理サーバ500として機能する。
- [0160] 図8～15は、本システムにおける処理手順の一例を説明するためのものである。本例の手順を実現する管理サーバ500が、例えば、図7に示す各部を備えるものとなる。
- [0161] 本例の手順を開始する前に、統計処理結果提供サービス事業者は、その統計処理のために用いるクラウドの個数と、各クラウドに必要な計算リソース（台数、CPU、メモリ等）を見積もり、本システムの設計を行う。そして、独立なクラウドサービス事業者を必要な個数選択して、クラウドリソースを契約する。その後、以下の手順を行い、必要な統計処理結果を得たら、情報漏洩のリスクを確実になくすため、データを初期化（完全削除）して、クラウドリソースを解約する。
- [0162] 図8は、マネージャの通知部510と各データ入力デバイス10との間で行われる準備段階の手順を示す。各データ入力デバイスは、予め決められたマネージャに問い合わせ[1]、マネージャは、M個の利用可能なクラウド群の中から、図1の例では2つのクラウドを選択して[2]、その情報を各データ入力デバイスへ通知する[3]。マネージャは、図3～5の例ではどのクラウドにどの種類のデータをアップロードするかを示す情報も、各データ入力デバイスへ通知する[3]。また、マネージャは、処理対象データ利用クラウド登録部520において、各々のオリジナルデータのID（1デバイス1データの場合、データ入力デバイスのIDでもよい）に対応させて、

データ入力デバイスへ通知した内容を記憶する〔２〕。

[0163] 図９は、各データ入力デバイス１０が、マネージャから通知された内容に従って、秘匿分割〔４〕により得られた各部分データを各クラウドサービス設備へアップロードする〔５〕〔６〕手順を示す。各データ入力デバイス１０は、部分データに加えて、マネージャのアドレス等の識別情報と、データのＩＤも、アップロードする。なお、〔５〕と〔６〕は、同時に行っても、時間差を設けて行ってもよく、また、各データ入力デバイス１０における〔４〕～〔６〕の実行タイミングは、独立でよい。つまり、データ入力デバイス間で同期を取る必要はなく、各々のデータ入力デバイス１０でオリジナルデータ取得されたタイミングで、〔４〕～〔６〕が実行される。

[0164] 図１０は、各クラウドサービス設備３０が、それぞれのタイミングで、マネージャのアップロード状態把握部５３０へ、アップロードされたデータのＩＤを通知する〔８〕〔９〕手順を示す。これらの通知を受けたマネージャは、処理対象データ利用クラウド登録部５２０において各データＩＤに対応させて登録されている複数のクラウドのうち、通知されたクラウドにアップロード済みのマークを付す等により、状態一時記憶部５３０に、登録された複数のクラウドの一部から通知を受けた状態になったデータＩＤについて、その状態を記憶する〔９〕。これにより、マネージャは、部分データ自体を受信することなく、どのデータの部分データがどのクラウドに保存されたかを管理することが可能になる。

[0165] 図１１は、マネージャの計算対象データ特定部５５０が、各クラウドサービス設備３０との間で、全てのクラウドに部分データが受信されているデータＩＤを共有するための手順を示す。マネージャは、状態一時記憶部５３０に記憶されているデータＩＤが、登録されている全てのクラウドから通知を受けている状態になると、そうなったデータＩＤ又はデータＩＤ群に対応するシーケンス番号を払い出し、シーケンス情報登録部５６０に、払い出したシーケンス番号とそのＩＤ又はＩＤ群とを登録する〔１０〕。そして、登録されたＩＤ又はＩＤ群の記憶を状態一時記憶部５３０から消去する〔１０〕

。

[0166] マネージャの計算対象データ特定部550は、その後、所定のタイミングで、シーケンス番号と対応するID又はID群とを、各クラウドサービス設備30に通知する[11]。この通知は、シーケンス番号が払い出される毎に行ってもよいし、幾つかのシーケンス番号の情報をまとめて通知するようにしてもよい。各クラウドサービス設備30は、自身が保存しているアップロードされた部分データのIDと、通知されたシーケンス番号との対応を記憶する[12]。

[0167] 図9に示すように、例えば、ID=3の部分データが、クラウドBには届いているが、クラウドAには届いていない場合、図10に示す管理により、図11で、クラウドA及びBの全てに部分データが届いているID=1, 2だけが、シーケンス番号=1に対応することが通知される。

[0168] 図12は、図9の続きであり、ID=4の各部分データと、ID=5の各部分データが、各データ入力デバイス10での秘匿分割により生成され[13]、各クラウドサービス設備へアップロードされている[14][15]。

[0169] 図13は、図12のアップロードを受けた各クラウドが、図10で説明したように、マネージャへの通知を行い[16][17]、マネージャが状態を記憶する[18]様子を示している。

[0170] 図14は、図13の通知を受けたマネージャが、図11で説明したシーケンス番号を払い出した後に、登録されている全てのクラウドから通知を受けている状態になったデータID又はデータID群に対応して、新たなシーケンス番号を払い出し[19]、各クラウドへ通知して[20]、対応を記憶させる[21]様子を示している。

[0171] 例えば、ID=3の部分データがクラウドBに届かないまま、クラウドA及びBの全てにID=4, 5の部分データが届いたとすると、マネージャは、ID=4, 5を、新たなシーケンス番号=2に対応させて、登録する。

[0172] ここで、過去に遡って統計処理をする用途がないものであれば、マネージ

ャは、シーケンス番号＝１に対応して登録されているＩＤ＝１，２を、シーケンス番号＝２に対応するものとして追加し、シーケンス番号＝１についての登録を削除してしまっても構わない。また、各クラウドは、マネージャから通知されたとおり、ＩＤ＝１，２はシーケンス番号＝１に対応し、ＩＤ＝４，５はシーケンス番号＝２に対応するものとして記憶して、後にシーケンス番号＝２が指定された場合に、指定されたシーケンス番号及びそれより小さいシーケンス番号に対応するＩＤ群のデータが指定されたと解釈してもよいし、その解釈を示すようにシーケンス番号を書き換えて記憶してもよい。

[0173] 図１５は、マネージャが統計処理の結果を求める段階の手順を示す。マネージャの統計処理部５７０における計算依頼部５７５が、部分データを保存する全てのクラウドに対して、現在の（過去に遡って統計処理をする場合には指定した時点の）シーケンス番号を引数として、計算処理をするよう依頼する〔２２〕。このとき、マネージャから各クラウドへ渡される情報は、シーケンス番号のみとすることができる。図３又は図４の例では、各クラウドで行われる処理は、総和及び２乗和の計算である。

[0174] 依頼を受けた各クラウドサービス設備３０では、既に、指定されたシーケンス番号にどのＩＤ群が対応するかが記憶されているため、これらのＩＤ群の部分データにつき計算処理を行って、その結果の値をマネージャへ返送する〔２３〕。

[0175] マネージャの統計処理部５７０における集計部５７７は、依頼した全てのクラウドから結果が返送されると、それらの値を合計等して、求める統計値を算出する〔２４〕。図３のように、一部のクラウドからの値については２倍する等、どのクラウドからの結果かによってマネージャにおいて行う処理が異なる場合は、処理対象データ利用クラウド登録部５２０に記憶されているクラウドとアップロードされるデータの種類の対応を示す情報を参照する。

[0176] 上述したように、マネージャが管理するシーケンス番号を利用することで、全てのクラウドに部分データが揃ったデータ（上記の例では、ＩＤ＝１，

2, 4, 5) を対象にして、統計処理の結果を求めることができ、データの整合性が保証される。

[0177] なお、シーケンス番号を利用して、マネージャが、各クラウドに対して、計算処理の対象としてよいデータIDの情報を頻繁に共有することにより、通信の負荷を分散させるとともに、統計処理のために計算を依頼する際のレスポンスを速くすることが可能になる。

[0178] すなわち、データIDの情報共有をせず（計算対象データ特定部550を備えず）に、マネージャが、各クラウドに計算処理を依頼する際に、対象とすべき（全てのクラウドに部分データが揃っている）データIDの全てを通知する（上記の例では、シーケンス番号＝2の代わりに、ID＝1, 2, 4, 5という情報を通知する）構成でも、本システムを実現することができるが、膨大な数のデータに対して統計処理を行う場合、シーケンス番号を利用した情報共有をしておく方が望ましい。

[0179] 本システムにおいては、マネージャと、他の装置との間のAPI（インタフェース）は、オリジナルデータは勿論、オリジナルデータを構成する個々の部分データも、一切、受け渡しされることがないように構成される。また、オリジナルデータを取り扱う各データ入力デバイスと、他の装置との間のAPIは、データ入力デバイスからのアクセスのみが行われ（図8の[1]、図9の[5][6]等）、データ入力デバイスが外部からアクセスされることがないように構成される。そして、オリジナルデータは存在せず秘匿化されているものの、部分データは保持する各クラウドと、他の装置との間のAPIは、クラウドから部分データを取り出すことがないように構成される。これらのAPIによっても、秘匿すべきデータの安全性が保たれる。

[0180] 上記のAPIに加えて、マネージャの統計処理部570を、あるシーケンス番号に対応するデータ群についての処理をした後、一定量（例えば、1万等、個々のデータについての推測を事実上不可能にするくらいの大量）以上のデータIDが処理対象として追加されるまでは、次の計算依頼を各クラウドへ送信しない構成とすると、さらに安全である。これは、マネージャが、

例えば、シーケンス番号＝2（ID＝1，2，4，5）の総和を求めた後に、シーケンス番号＝3（ID＝1，2，4，5，7）の総和を求めると、引き算することにより、追加された個別要素であるID＝7のオリジナルデータが求められてしまうためである。

[0181] 図6～15で説明した本システムの構成例においては、各データ入力デバイスの生成した各部分データがそれぞれのクラウドサービス設備に保存されるかという情報を、統計処理結果提供サーバ（マネージャ）が管理しているため、悪意を持った攻撃者に同サーバがクラックされると、各データの所有者や保存場所等のヒントを攻撃者に与える可能性がある。

[0182] そのような可能性までも低減するためには、各データ入力デバイスと統計処理結果提供サーバが通信することがなく、各データ入力デバイス自身が各部分データを保存するクラウドサービス設備（アップロード先）を決定できるようにして、統計処理結果提供サーバは各データ入力デバイスを特定する情報を取り扱わないようにするとよい。

[0183] 具体例としては、各データ入力デバイスが、コンシステントハッシュ（例えば、D.Karger et al. "Consistent Hashing and Random Trees: Distributed Caching Protocols for Relieving Hot Spots on the World Wide Web," Proceedings of the 29th Annual ACM Symposium of Theory of Computing, pp.654-663 (1997)、I.Stoica et al. "Chord: A scalable peer-to-peer lookup service for internet applications," ACM SIGCOMM Computer Communication Review 31(4), p.149 (2001)等を参照）の仕組みを利用して、データの保存先となるクラウドサービス設備を決定することができる。

[0184] 図16は、そのように構成した本システムの例であり、図6及び図7の例と同一の符号を付したブロックは、図6及び図7に関して説明したのと同様の機能を有する。

[0185] 図16では、データ入力デバイス15-1～15-Nと、クラウドサービス設備35-1～35-Mと、統計処理結果提供サーバ55が、ネットワーク40を介して接続されているが、各データ入力デバイス15と統計処理結

果提供サーバ55との間では、通信が行われない。

- [0186] 各データ入力デバイス15は、データ取得部110と、秘匿分割部120と、秘匿分割により得られた部分データを暗号化された通信路で各クラウドサービス設備35へアップロードするアップロード部130に加えて、コンシステントハッシュによりアップロード先を決定するためのキー生成部160及びハッシュ計算部170を備える。
- [0187] 各データ入力デバイス15の備える制御部150は、秘匿分割部120におけるデータの分割数や生成すべき部分データの種別を制御することに加えて、キー生成部160に、秘匿分割されたデータ毎に一意となるキー（例えば、UUID（ユニバーサリユニーク識別子）、IPv6（インターネットプロトコルバージョン6）アドレス等）を生成させ、ハッシュ計算部170に、生成されたキーと、時刻と、シーケンス番号との合計を求めさせ、その合計値から、ハッシュ値を算出させることを行う。
- [0188] 例えば、各クラウドサービス設備35に、予め所定の範囲を持った値群（range）を割り当てておくことにより、算出されたハッシュ値がそのrangeに含まれるクラウドサービス設備を、データのアップロード先として特定することができる。この仕組みにより、制御部150が、部分データ毎に算出されたハッシュ値に従って、アップロード部130における各部分データのアップロード先を指定することで、各データ入力デバイスは、統計処理結果提供サーバ（マネージャ）に対して、アップロード先となるクラウドを問い合わせる必要がなくなる。
- [0189] 各クラウドサービス設備35の備える制御部335は、統計処理結果提供サーバ55中の管理部（管理サーバ）505からの指示に従って、計算部320が所定の演算処理を行うタイミングを特定する。その演算処理の対象としてデータ保存部310から読み出すべきデータは、制御部335自身が特定する。
- [0190] 統計処理結果提供サーバ55は、管理サーバ505と、結果提供インタフェース590とを備える。管理サーバ505は、統計処理部572を備え、

各クラウドサービス設備 3 5 へ計算処理を依頼し（計算依頼部 5 7 6）、各依頼に対して返却される計算結果を集計して（集計部 5 7 8）、統計処理の結果を求める。

[0191] 図 1 6 における統計処理結果提供サーバ 5 5（管理サーバ 5 0 5）は、図 7 の統計処理結果提供サーバ 5 0（管理サーバ 5 0 0）と異なり、アップロード先のクラウドを各データ入力デバイスに通知するための機能や、アップロード状態を把握したり計算対象となるデータを特定したりするための機能を備えない。よって、統計処理結果提供サーバ 5 5（マネージャ）は、個々のデータに関連するヒントを一切持たない。

[0192] また、マネージャは、自身の行う統計処理について、どのクラウドが利用され得るか（上述した range が割り当てられていると各データ入力デバイスが認識しているクラウドがどれか）は把握しており、統計処理を行う際には、利用され得る全てのクラウドに対して総和および 2 乗和の計算を依頼するが、各クラウドにおける計算がどのデータ入力デバイスからのデータを対象として行われたものかは把握できないため、マネージャに対しても、データのセキュリティを担保することが可能となる。

[0193] さらに、コンシステントハッシュを利用することで、クラウドの数が増えてもスケーラビリティを担保することが可能であり、分散処理に強いシステムを実現することができるという利点もある。

[0194] 図 1 7 ～ 1 9 は、図 1 6 の構成例において、各データ入力デバイス X_i が、取得したデータ A_i を、2 つの部分データ a_i 及び b_i に秘匿分割して、複数（本例では 4 つだが、多数とすることが可能）のクラウドから任意に選択した 2 つのクラウドにアップロードし、統計処理を行うための処理手順の一例を示している。

[0195] 図 1 7 は、各データ入力デバイス 1 5 の内部で行われる準備段階の手順を示す。各データ入力デバイスは、2 つの部分データのアップロード先のクラウドを決定するために、UUID を利用して、2 つのキー（ k_1 及び k_2 ）を生成する [1]。そして、それぞれのキー（ k_1 及び k_2 ）に対して、時刻（ t_i

me) と、シーケンス番号 n (1 及び 2) とを足し、それぞれの合計値のハッシュ値 (h_1 及び h_2) を算出する。

[0196] ここで、各クラウドには、0000～ffffまでの値が割り振られており、リング(環)が形成されている。例えば、クラウドが4つの場合、クラウドAには0000～3fffまでの範囲の値群を、クラウドBには4000～7fffまでの範囲の値群を、クラウドCには8000～bfffまでの範囲の値群を、クラウドDにはc000～ffffまでの範囲の値群を割り当てることができる。本例では、割り当てる範囲を等分割しているが、あるクラウドに割り当てる値群の範囲を、別のクラウドに割り当てる値群の範囲よりも大きくしてもよい。そして、算出されたハッシュ値 (h_1 及び h_2) が、割り当てられた値群の中に含まれるクラウドを、それぞれ対応する部分データ (a_i 及び b_i) のアップロード先として決定する [2]。

[0197] 図18は、各データ入力デバイス15が、秘匿分割 [3] により得られた各部分データ (a_i 及び b_i) を、各クラウドサービス設備35へアップロードする [4] [5] 手順を示す。各データ入力デバイス15は、部分データのみをアップロードするのでもよいし、部分データに加えてマネージャのアドレス等(統計処理の識別情報となるもの)をアップロードしてもよい。

[0198] また、[4] と [5] は、同時に行っても、時間差を設けて行ってもよいところ、一つのデータから秘匿分割で得られた全ての部分データが各クラウドに保存されるまでのタイムラグの間に、当該データを対象とする統計処理が行われてしまうと、結果に誤りが生じる。これを防ぐために、所定時間以上前の時刻が付されたデータに計算対象を限定する機能を、各クラウドが有している場合等には、部分データに加えて、時刻をアップロードしてもよい。但し、図16の構成例では、データのIDは、アップロードされない。

[0199] [4] と [5] の具体的な処理は、次のようになる。各データ入力デバイス X_i は、それぞれのタイミングで、[2] の $n=1$ で生成したハッシュ値 h_1 に相当するクラウドに対して、[3] の部分データ a_i (及び必要に応じて時刻)を送信する。図18の例では、データ入力デバイス X_1 はクラウドBに対

して、データ入力デバイス X_2 はクラウドAに対して、データ入力デバイス X_3 はクラウドAに対して、部分データ a_i を送信している。

[0200] 上述した部分データ a_i のアップロード先での保存が、キーバリューストアによって行われる場合、部分データ a_i は、対応するハッシュ値 h_1 と共に送信される。そうすると、各クラウドは、ハッシュ値 h_1 をキーとし、部分データ a_i （及び必要に応じて時刻）をバリューとして、データ保存部310への保存を行い、データ入力デバイス X_i へ受領確認通知を行う[4]。

[0201] 同様に、各データ入力デバイス X_i は、それぞれのタイミングで、[2]の $n=2$ で生成したハッシュ値 h_2 に相当するクラウドに対して、[3]の部分データ b_i （及び必要に応じて時刻）を送信する。図18の例では、データ入力デバイス X_1 はクラウドCに対して、データ入力デバイス X_2 はクラウドCに対して、データ入力デバイス X_3 はクラウドDに対して、部分データ b_i を送信している。

[0202] 部分データ b_i は、対応するハッシュ値 h_2 と共に送信され、各クラウドにおいて、ハッシュ値 h_2 をキーとし、部分データ b_i （及び必要に応じて時刻）をバリューとして、データ保存部310への保存が行われる。そして、データ入力デバイス X_i へ受領確認通知が返される[5]。

[0203] 図19は、統計処理結果提供サーバ（マネージャ）55が、複数のクラウドを利用して、統計処理の結果を求める段階の手順を示す。マネージャは、本統計処理のために利用する全てのクラウドに対して、各クラウドに対象となるデータが実際にアップロードされているか否かに関わらず（各データ入力デバイスが任意にアップロード先を選択するため、一部のクラウドがどのデータ入力デバイスからも選択されていないという状態も生じ得るが、マネージャではその状態を把握せずに）、計算処理（例えば、総和及び2乗和の計算）をするよう依頼する[6]。

[0204] 依頼を受けた各クラウドサービス設備35では、データ保存部310に保存されている部分データを対象として計算処理を行い、その結果の値をマネージャへ返送する[7]。このとき、上述したタイムラグを考慮して、デー

タ保存部310に保存されているデータのうち、現在時刻から所定時間以上前となる時刻が付されたものだけを、計算処理の対象とするようにしてもよい。また、一度統計処理の対象となった部分データを再び対象としないよう、計算処理をした部分データをデータ保存部310から削除するか、計算処理の対象を未処理の部分データに限定するようにしてもよい。

[0205] マネージャは、依頼した全てのクラウドから結果が返送される（対象となるデータが実際にアップロードされていないクラウドからは、ゼロという値が返送される）と、それらの値を合計等して、求める統計値を算出する〔8〕。

[0206] 以上の構成により、少なくとも図1及び図2の例の総和を求めることが可能である。図3及び図4の例の2乗和を求めるためには、図17に例示したクラウドのリングを少なくとも2つ設けておき、 m 個の部分データ x_{ji} については、第1のリングに属する複数のクラウドの中から各部分データに対して決定されたクラウドへアップロードし、 m 個の部分データ x'_{ji} については、第2のリングに属する複数のクラウドの中から各部分データに対して決定されたクラウドへアップロードする。

[0207] そして、マネージャ55は、各クラウドが第1のリングと第2のリングのいずれに属しているのかは把握しており、第1のリングに属するクラウドからの結果については $f_s(X_i)$ 即ち総和を選択し、第2のリングに属するクラウドからの結果については $f_z(X'_i)$ 即ち2乗和を選択して、これらを合計する処理を行う。これにより、オリジナルデータ x_i の2乗和を求めることができる。また、第1のリングに属するクラウドからの結果のうちの $f_s(X_i)$ を選択して、これらを合計する処理を行えば、オリジナルデータ x_i の総和が求められる。

[0208] 図16～19で説明した構成例において、一つのデータを秘匿分割して得られた複数の部分データのうち、一部はクラウドに保存されているが、残りが保存されていない状態である場合に、そのような状態のデータを確実に除いて、統計処理の結果を求めることができるように、マーカーという仕組み

を導入してもよい。

[0209] 具体的には、各データ入力デバイスが、秘匿分割により得られた各部分データ用のハッシュ値に加えて、マーカー用のハッシュ値を算出し、一つのデータを構成する全ての部分データがクラウド上に保存されたことを確認した後で、クラウド上にマーカーを設定する。このマーカーを指し示す情報を、各データ入力デバイスが各部分データをクラウドに保存する際に、部分データと一緒に保存しておく。

[0210] これにより、統計処理結果提供サーバが計算処理をクラウドに依頼した際に、クラウドは、保存された部分データに紐づくマーカーが、設定されている場合、すなわち、そのデータを構成する全ての部分データが、それぞれいずれかのクラウドに保存済みである場合にのみ、データを計算の対象に含めることができ、データ入力デバイスからクラウドへのアップロードが途中であるデータが計算されるのを、確実に防ぐことができる。

[0211] 上記の仕組みは、3相コミットメント（例えば、Dale Skeen, "A Formal Model of Crash Recovery in a Distributed System," IEEE Transactions on Software Engineering 9(3), pp.219-228 (May 1983)等を参照）の技術を応用して実現することも可能である。上記のマーカーが、3相コミットメントの調整者に対応し、各データ入力デバイスが、3相コミットメントの参加者に対応するが、その際、各データ入力デバイスは、一意なキーにUUID等を利用するため、毎回アドレスが変わることで自らを隠蔽することになる。

[0212] 図20は、そのように構成した本システムの例であり、図16の例と同一の符号を付したブロックは、図16に関して説明したのと同じ機能を有する。

[0213] 図20では、データ入力デバイス17-1～17-Nと、クラウドサービス設備37-1～37-Mと、統計処理結果提供サーバ55が、ネットワーク40を介して接続されているが、各データ入力デバイス17と統計処理結果提供サーバ55との間では、通信が行われない。

[0214] 各データ入力デバイス17は、データ取得部110と、秘匿分割部120

と、キー生成部 160 及びハッシュ計算部 170 と、アップロード部 190 とを備え、アップロード部 190 は、秘匿分割により得られた部分データを各クラウドサービス設備 37 へアップロードする機能に加えて、マーカーを設定する情報（以下、「マーカー情報」という）をいずれかのクラウドサービス設備 37 へアップロードする機能を有する。

[0215] 各データ入力デバイス 17 の備える制御部 180 は、図 16 の制御部 150 が有する機能に加えて、マーカー用に、キー生成部 160 における一意なキー（UUI D 等）の生成と、ハッシュ計算部 170 における生成されたキーと、時刻と、シーケンス番号との合計値からのハッシュ値の算出とを、行わせる機能を有する。また、制御部 180 は、アップロード部 190 と連携して、秘匿分割により得られた全ての部分データがクラウド上に保存されたことを確認した後に、マーカー情報をアップロードする。

[0216] 各クラウドサービス設備 37 の備えるデータ保存部 317 は、アップロードされた各部分データと共に、マーカー情報の保存先を指し示す情報を保存する機能を有し、各クラウドサービス設備 37 は、データ保存部 317 に加えて、アップロードされたマーカー情報を保存するためのマーカー保存部 350 と、自身のもしくは他のクラウドサービス設備 37 のマーカー保存部 350 におけるマーカー情報の保存状況を問い合わせるためのマーカー問い合わせ部 340 とを備える。

[0217] 各クラウドサービス設備 37 の備える制御部 337 は、統計処理結果提供サーバ 55 中の管理部（管理サーバ）505 からの指示に従って、計算部 320 が所定の演算処理を行うタイミングを特定する。制御部 337 は、その演算処理をデータ保存部 317 に保存された部分データのうちのどれを対象として行うべきかを、マーカー問い合わせ部 340 と連携して、特定する。

[0218] 図 21～23 は、図 20 の構成例において、各データ入力デバイス X_i が、取得したデータ A_i を、2つの部分データ a_i 及び b_i に秘匿分割して、複数（本例では4つだが、多数とすることが可能）のクラウドから任意に選択した2つのクラウドにアップロードし、マーカー m_i を用いて整合性を担保しながら

ら、統計処理を行うための処理手順の一例を示している。

[0219] 図21は、各データ入力デバイス17の内部で行われる準備段階の手順を示す。各データ入力デバイスは、2つの部分データ及びマーカ情報のアップロード先のクラウドを決定するために、UUI Dを利用して、3つのキー(k_0 、 k_1 、 k_2)を生成する[1]。

[0220] そして、それぞれのキー(k_0 、 k_1 、 k_2)に対して、時刻(time)と、シーケンス番号n(0、1、2)とを足し、それぞれの合計値のハッシュ値(h_0 、 h_1 、 h_2)を算出する。そして、算出されたハッシュ値(h_0 、 h_1 、 h_2)が、割り当てられた値群の中に含まれるクラウドを、それぞれ対応するマーカ及び部分データ(m_i 、 a_i 、 b_i)のアップロード先として決定する[2]。

[0221] 図22は、各データ入力デバイス17が、秘匿分割[3]により得られた各部分データ(a_i 及び b_i)を、各クラウドサービス設備37へアップロードし[4][5]、その受領確認が得られた後に、それら部分データに対応するマーカ(m_i)を、クラウドサービス設備37へアップロードする[6]手順を示す。

[0222] 各データ入力デバイス17は、それぞれの部分データと一緒に、マーカ情報の保存先を指し示す情報(m_i に対応するハッシュ値 h_0)をアップロードする。これらに加えて、図16の構成例と同様に、マネージャのアドレス等(統計処理の識別情報となるもの)をアップロードしてもよい。なお、図20の構成例でも、データのIDは、アップロードされない。

[0223] また、一つのデータから秘匿分割で得られた複数の部分データのうちの一部分についてのアップロードのトランザクションがエラーとなった場合に、残りの部分データについてのトランザクションを解除する(保存されたデータを削除する等)ために、トランザクションにかかる時間の上限を超えたこと(タイムアウト)を検出する機能を、各クラウドが有している場合等には、部分データに加えて、時刻をアップロードしてもよい。

[0224] [4]～[6]の具体的な処理は、次のようになる。各データ入力デバイ

ス X_i は、それぞれのタイミングで、[2]の $n=1$ で生成したハッシュ値 h_1 に相当するクラウドに対して、[3]の部分データ a_i 及びハッシュ値 h_0 （及び必要に応じて時刻）を送信する。図22の例では、データ入力デバイス X_1 はクラウドBに対して、データ入力デバイス X_2 はクラウドAに対して、データ入力デバイス X_3 はクラウドAに対して、部分データ a_i 及びハッシュ値 h_0 を送信している。

[0225] 上述した部分データ a_i 及びハッシュ値 h_0 のアップロード先での保存が、キーバリューストアによって行われる場合、部分データ a_i 及びハッシュ値 h_0 は、対応するハッシュ値 h_1 と共に送信される。そうすると、各クラウドは、ハッシュ値 h_1 をキーとし、部分データ a_i 及びハッシュ値 h_0 （及び必要に応じて時刻）をバリューとして、データ保存部317への保存を行い、データ入力デバイス X_i へ受領確認通知を行う[4]。

[0226] 同様に、各データ入力デバイス X_i は、それぞれのタイミングで、[2]の $n=2$ で生成したハッシュ値 h_2 に相当するクラウドに対して、[3]の部分データ b_i 及びハッシュ値 h_0 （及び必要に応じて時刻）を送信する。図22の例では、データ入力デバイス X_1 はクラウドCに対して、データ入力デバイス X_2 はクラウドCに対して、データ入力デバイス X_3 はクラウドDに対して、部分データ b_i 及びハッシュ値 h_0 を送信している。

[0227] 部分データ b_i 及びハッシュ値 h_0 は、対応するハッシュ値 h_2 と共に送信され、各クラウドにおいて、ハッシュ値 h_2 をキーとし、部分データ b_i 及びハッシュ値 h_0 （及び必要に応じて時刻）をバリューとして、データ保存部317への保存が行われる。そして、データ入力デバイス X_i へ受領確認通知が返される[5]。

[0228] 各データ入力デバイス X_i は、[4]及び[5]の受領確認通知を受信する（データのクラウド上での保存に成功する）と、[2]の $n=0$ で生成したハッシュ値 h_0 に相当するクラウドに対して、マーカー（ m_i ）を設定する値（例えば、1）を送信する。図22の例では、データ入力デバイス X_1 はクラウドAに対して、データ入力デバイス X_2 はクラウドBに対して、データ入力デ

バイス X_3 はクラウド D に対して、マーカ (m_i) を設定する値を送信している。

[0229] 上述したマーカ (m_i) のクラウド上での設定が、キーバリューストアによって行われる場合、マーカを設定する値 (例えば、1) は、対応するハッシュ値 h_0 と共に送信される。そうすると、各クラウドは、ハッシュ値 h_0 をキーとし、値 1 をバリューとして、マーカ保存部 350 への保存を行い、データ入力デバイス X_i へ受領確認通知を行う [6]。

[0230] 図 23 は、統計処理結果提供サーバ (マネージャ) 55 が、複数のクラウドを利用して、統計処理の結果を求める段階の手順を示す。マネージャは、本統計処理のために利用する全てのクラウドに対して、各クラウドに対象となるデータが実際にアップロードされているか否かに関わらず、計算処理 (例えば、総和及び 2 乗和の計算) をするよう依頼する [7]。

[0231] 依頼を受けた各クラウドサービス設備 37 では、データ保存部 317 に部分データと一緒に保存されているハッシュ値 h_0 (マーカ情報の保存先を指し示す情報) を読み出し、そのハッシュ値 h_0 に相当するクラウドに対して、マーカが設定されているか、すなわち、ハッシュ値 h_0 をキーとしてマーカを設定する値 (1) がマーカ保存部 350 に保存されているかの確認を行う [8]。

[0232] 図 23 の例では、クラウド A は、自身が保存している部分データ a_2 、 a_3 についてのマーカ問い合わせ [8] を、それぞれクラウド B、D に対して行い、クラウド B は、自身が保存している部分データ a_1 についてのマーカ問い合わせ [8] をクラウド A に対して行い、クラウド C は、自身が保存している部分データ b_1 、 b_2 についてのマーカ問い合わせ [8] を、それぞれクラウド A、B に対して行い、クラウド D は、自身が保存している部分データ b_3 についてのマーカ問い合わせ [8] を自身の内部で行っている。

[0233] 問い合わせを受けたクラウドは、問い合わせに係るキー (ハッシュ値 h_0) とバリューの組を自身が保存していればそのバリュー (1) を、マーカ (m_i) の値として、問い合わせ元のクラウドへ返送する。保存していなければ

、エラーを示す値（１以外の値）を、マーカーの値として返送する。

[0234] 問い合わせ元のクラウドでは、[８]で返送されたマーカー（ m_i ）の値が１であれば、そのハッシュ値 h_0 と一緒に保存されていた部分データを対象として計算処理を行い、その結果の値をマネージャへ返送する[９]。マーカーの値が１以外である部分データは、計算対象に含めないことにより、一つのデータを構成する全ての部分データがクラウド上に揃っているデータのみに基づいて、正確な統計処理を行うことが可能である。

[0235] また、問い合わせ元のクラウドは、問い合わせ先のクラウドから１という値が返送されなかったマーカーのハッシュ値 h_0 と一緒に保存されていた時刻をチェックして、その時刻が現在時刻から所定時間（例えば１０分）以上前であれば、トランザクションが正常に完了されなかったとみなして、一緒に保存されていた部分データを削除してもよい。所定時間以内であれば、まだトランザクションの途中である可能性があるともみなして、その部分データを計算対象には含めずに、そのまま残せばよい。

[0236] マネージャは、依頼した全てのクラウドから結果が返送される（対象となるデータが実際にアップロードされていないクラウドからは、ゼロという値が返送される）と、それらの値を合計等して、求める統計値を算出する[１０]。

[0237] なお、図６～１５で説明した例と、図１６～１９で説明した例と、図２０～２３で説明した例とは、それぞれ適宜組み合わせて実施することも可能である。

[0238] 例えば、図５の例の内積を求める構成として、１つのデータ毎に４つのクラウドを各データ入力デバイス自身が（マネージャから指示を受けることなく）特定できるようにしつつ、各データ入力デバイスから各クラウドへ、部分データと共にデータのＩＤ（ i ）をアップロードする（クラウドからマネージャへは報告しない）ことで、統計処理結果提供サーバ（マネージャ）が管理する情報を少なくすることが可能になる。このときに、４つのクラウドのいずれか又は他のクラウドにマーカーを登録し、各クラウドが内積を計算

する対象をマーカーが登録されている部分データに限ることで、マネージャが管理を行わずに正確な統計処理結果を求めることも可能になる。

[0239] また、例えば、図 16～19 について説明したのと同様に、図 20～23 においても、2 乗和を求めるために、クラウドのリングを少なくとも 2 つ設けることができる。その場合に、マーカーを登録するクラウドとして、第 1 のリングに属するクラウドを選択しても、第 2 のリングに属するクラウドを選択しても、いずれのリングにも属さないクラウドを選択しても構わない。

[0240] 以上、統計処理について述べてきたが、本システムは、オリジナルデータの所有者が、統計処理のために部分データをアップロードする各クラウドを利用して、オリジナルデータを秘匿化して分散保存しておき、他者はこれにアクセスできないようにしつつ、本人は参照したいときにいつでもオリジナルデータを復元できるように構成することも可能である。

[0241] このためには、各クラウドサービス設備 30 のデータ保存部 310 に、鍵によるアクセス権限の検証機能を追加し、例えば、データ入力デバイス 10 から各クラウドサービス設備 30 へ部分データをアップロードする際に、その鍵の情報も付加してアップロードする。そして、各クラウドサービス設備 30 のデータ保存部 310 は、部分データとともに鍵によるアクセス情報を保存しておき、部分データに対するアクセスを受けると、アクセスしてきた者が対応する鍵を所有していることが検証された場合のみ、部分データの取得を許可する。

[0242] 別の例として、各クラウドサービス設備 30 のデータ保存部 310 に、予めデータの所有者の鍵の情報を保存しておき、部分データがアップロードされた際に、その部分データに対応する鍵の情報を付加して（例えば、部分データをその鍵で暗号化して）保存するようにしてもよい。いずれの例でも、オリジナルデータの所有者は、部分データを保存する全てのクラウドにアクセスして、それぞれ鍵を用いて部分データを取得し、全ての部分データを集めることにより、オリジナルデータを復元することができる。

[0243] 図 24～27 は、本システムのあり得る応用例について、そのごく一部を

説明するものである。図24は、教育分野への応用であり、例えば、オンライン試験や模擬試験等を対象とする統計処理に適用可能である。図25は、医療（ヘルスケア）分野への応用であり、例えば、血圧、体重、体脂肪率等を対象とする統計処理に適用可能である。図26は、流通業分野への応用であるが、それに限らず、例えば、生活実態調査等の匿名アンケート調査における統計処理にも適用可能である。図27は、テレマティクス（自動車）分野への応用であり、例えば、速度、加速度等の走行情報を対象とする統計処理に適用可能であるが、それ以外の分野におけるリスク管理への応用等も可能である。

[0244] 以上、本発明の実施形態について例示的に説明したが、本明細書における説明により本発明が限定されるものではなく、上述の実施形態を本発明の範囲内で当業者が適宜、種々に変形、応用して実施できることは勿論である。

請求の範囲

- [請求項1] それぞれが秘匿すべきオリジナルデータを取得する手段を備える複数のデータ入力装置と、
- それぞれが複数の入力データに基づいて所定の演算を行う手段を備える複数の演算装置と、
- 前記複数の演算装置のそれぞれが前記オリジナルデータの部分データを前記入力データとして演算を行った結果を利用することにより、前記複数のデータ入力装置により取得された複数のオリジナルデータに基づく統計処理の結果を、該オリジナルデータを取得することなく求める手段を備えるデータ処理装置と、
- を備えることを特徴とするデータ秘匿型統計処理システム。
- [請求項2] 前記データ入力装置は、
- 全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを秘密の比率に従って分割し、所定の個数の部分データを生成する手段と、
- 前記所定の個数の部分データのそれぞれを、前記複数の演算装置のうち対応する演算装置へ、保護された通信路により送信する手段とを備えることを特徴とする請求項1に記載のデータ秘匿型統計処理システム。
- [請求項3] 前記演算装置は、
- 前記複数のデータ入力装置から受信した複数の前記部分データに基づいて所定の演算を行って得た演算結果を、前記データ処理装置へ送信する手段を備え、
- 前記データ処理装置は、
- 前記複数の演算装置から受信した複数の前記演算結果に基づいて所定の統計処理を行う手段を備えることを特徴とする請求項2に記載のデータ秘匿型統計処理システム。
- [請求項4] 前記所定の個数の部分データは、前記オリジナルデータを分割した

各々の部分の値から生成されるものを含み、

前記演算装置が行う所定の演算は、前記複数の部分データの総和の演算を含み、

前記データ処理装置が行う所定の統計処理は、前記所定の個数の演算結果の総和を計算する処理を含むことを特徴とする請求項3に記載のデータ秘匿型統計処理システム。

[請求項5]

前記所定の個数の部分データは、前記オリジナルデータを分割した各々の部分の値から生成されるものと、互いに異なる2つの部分同士を乗算した値に基づいて生成されるものを含み、

前記演算装置が行う所定の演算は、前記複数の部分データの総和及び2乗和の少なくとも一方の演算を含み、

前記データ処理装置が行う所定の統計処理は、前記所定の個数の演算結果のうち前記各々の部分の値に対応するものについて2乗和を計算する処理と、前記所定の個数の演算結果のうち前記部分同士を乗算した値に対応するものについて総和を計算する処理とを含むことを特徴とする請求項3又は4に記載のデータ秘匿型統計処理システム。

[請求項6]

前記所定の個数の部分データは、前記オリジナルデータを分割した各々の部分を2乗した値から生成されるものと、互いに異なる2つの部分同士を乗算した値に基づいて生成されるものを含み、

前記演算装置が行う所定の演算は、前記複数の部分データの総和の演算を含み、

前記データ処理装置が行う所定の統計処理は、前記所定の個数の演算結果の総和を計算する処理を含むことを特徴とする請求項3又は4に記載のデータ秘匿型統計処理システム。

[請求項7]

前記データ処理装置により求められる統計処理の結果は、標本平均の算出、標本分散の算出、標本偏差の算出、最尤推定、T分布を用いた区間推定、母比率の信頼区間の推定、母分散の推定、母平均の検定、母集団A及びBに関する母平均の差の検定、母比率の検定、母集団

A 及び B の母分散の比較検定、分散分析のうちの少なくとも 1 つの結果であることを特徴とする請求項 1 ～ 6 のいずれか 1 項に記載のデータ秘匿型統計処理システム。

[請求項8] 前記複数のデータ入力装置は、同数ずつの互いに対応する第 1 のデータ入力装置と第 2 のデータ入力装置とを含み、

前記第 1 のデータ入力装置及び前記第 2 のデータ入力装置が行う送信は、前記所定の個数の部分データのそれぞれを、所定の個数の 2 乗個の前記演算装置のうち対応する所定の個数の演算装置へ送信するものであり、

前記演算装置が行う所定の演算は、前記第 1 のデータ入力装置からの部分データ列と前記第 2 のデータ入力装置からの部分データ列との内積を求める演算を含み、

前記データ処理装置が行う統計処理は、前記所定の個数の 2 乗個の演算装置から受信した所定の個数の 2 乗個の前記演算結果の総和を計算する処理を含むことを特徴とする請求項 3 ～ 6 のいずれか 1 項に記載のデータ秘匿型統計処理システム。

[請求項9] 前記データ処理装置により求められる統計処理の結果は、共分散の算出、相関係数の算出、回帰分析のうちの少なくとも 1 つの結果であることを特徴とする請求項 1 ～ 8 のいずれか 1 項に記載のデータ秘匿型統計処理システム。

[請求項10] 前記データ入力装置が、

前記オリジナルデータを分割する際に発生させた乱数を用いて前記秘密の比率を定め、分割した後に前記秘密の比率の記憶を消去する手段をさらに備えることを特徴とする請求項 2 ～ 9 のいずれか 1 項に記載のデータ秘匿型統計処理システム。

[請求項11] 前記演算装置が、

前記複数のデータ入力装置から受信した複数の前記部分データのそれぞれを、該部分データを送信したデータ入力装置と関連付けて保存

する手段と、

前記データ入力装置との関連付けを示す要求に応じて、前記複数の部分データのうち、該データ入力装置と関連付けて保存されている部分データを返送する手段とをさらに備えることを特徴とする請求項 1 ～ 10 のいずれか 1 項に記載のデータ秘匿型統計処理システム。

[請求項12]

前記データ入力装置との関連付けを有する装置が、

前記オリジナルデータを分割して生成した部分データの全てを、前記複数の演算装置のうち対応する演算装置から取得して、前記オリジナルデータを復元する手段を備えることを特徴とする請求項 11 に記載のデータ秘匿型統計処理システム。

[請求項13]

前記データ入力装置との関連付けを有する装置が、

前記オリジナルデータを分割した各々の部分のうち一つについての前記比率を記憶する手段と、

前記オリジナルデータを分割して生成した部分データのうち、記憶されている前記一つの比率に該当する部分データを、前記複数の演算装置のうち対応する演算装置から取得して、前記オリジナルデータを復元する手段とを備えることを特徴とする請求項 11 に記載のデータ秘匿型統計処理システム。

[請求項14]

前記データ処理装置は、

前記複数のデータ入力装置のそれぞれへ、前記部分データを前記複数の演算装置のうちいずれへ送信すべきかを指示する手段と、

前記複数の演算装置のそれぞれへ、前記複数のデータ入力装置から受信した複数の前記部分データのうちいずれを対象として所定の演算を行うべきかを指示する手段とを備えることを特徴とする請求項 1 ～ 13 のいずれか 1 項に記載のデータ秘匿型統計処理システム。

[請求項15]

前記複数のデータ入力装置のそれぞれは、

前記部分データを前記複数の演算装置のうちいずれへ送信すべきかを決定する手段を備え、

前記複数の演算装置のそれぞれは、

前記複数のデータ入力装置から受信した複数の前記部分データのうちいずれを対象として所定の演算を行うべきかを決定する手段を備えることを特徴とする請求項 1 ～ 13 のいずれか 1 項に記載のデータ秘匿型統計処理システム。

[請求項16] 前記複数の演算装置の数は、一つのオリジナルデータから得られる部分データの個数である所定の数と同じかそれよりも大きく、

前記所定の個数の部分データが、それぞれ異なる演算装置に送信されることを特徴とする請求項 14 又は 15 に記載のデータ秘匿型統計処理システム。

[請求項17] 前記複数の演算装置は、それぞれ異なる事業者により提供されるサービスに属するものであり、

前記データ処理装置は、前記複数の演算装置とは異なる事業者により運営されるものであることを特徴とする請求項 1 ～ 16 のいずれか 1 項に記載のデータ秘匿型統計処理システム。

[請求項18] 秘匿されるべきオリジナルデータを取得することなく、複数の前記オリジナルデータに基づく統計処理の結果を提供するサービスのためのサーバ装置であって、

それぞれが複数の入力データに基づいて所定の演算を行う手段を有する複数の演算装置と通信する手段と、

前記複数の演算装置のそれぞれに、前記オリジナルデータの部分データを前記入力データとして演算を行わせ、該演算の結果を取得する手段と、

前記複数の演算装置からの演算結果に基づいて、所定の統計処理を行う手段とを備え、

複数の前記部分データは、全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを秘密の比率に従って分割することにより、生成されたものであることを特徴とする統計

処理結果提供サーバ装置。

[請求項19] 前記複数の演算装置に前記オリジナルデータに属する全ての部分データが揃ったことを確認する手段と、

前記確認のされた部分データのそれぞれを対象として対応する演算装置における前記所定の演算が行われるように、前記複数の演算装置のそれぞれへ指示を行う手段をさらに備えることを特徴とする請求項18に記載の統計処理結果提供サーバ装置。

[請求項20] 前記確認のために、前記複数の演算装置のそれぞれから、該演算装置に保存された部分データが属するオリジナルデータの識別番号を受信する手段をさらに備えることを特徴とする請求項19に記載の統計処理結果提供サーバ装置。

[請求項21] 前記確認がされたオリジナルデータの識別番号の集合をシーケンス番号と対応付けて前記複数の演算装置へ通知する手段と、

前回の通知の後に前記確認がされたオリジナルデータの識別番号の集合を次のシーケンス番号と対応付けて前記複数の演算装置へ通知する手段とをさらに備え、

前記複数の演算装置のそれぞれに、1つのシーケンス番号の指定と共に前記所定の演算の指示を送信することにより、指定されたシーケンス番号以前の複数のシーケンス番号に対応する識別番号の集合を併せて前記所定の演算の対象とする部分データが特定されることを特徴とする請求項19又は20に記載の統計処理結果提供サーバ装置。

[請求項22] あるオリジナルデータの集合を対象として、前記複数の演算装置に前記所定の演算を行わせた結果を取得した後に、該あるオリジナルデータの集合に限定された数のオリジナルデータを加えた集合を対象として、前記複数の演算装置に前記所定の演算を行わせた結果を取得することを禁止する手段をさらに備えることを特徴とする請求項18～21のいずれか1項に記載の統計処理結果提供サーバ装置。

[請求項23] それぞれが前記オリジナルデータを取得して前記部分データを生成

する手段を有する複数のデータ入力装置と通信する手段と、

利用可能な演算装置の中から、前記所定の統計処理を行うための前記複数の演算装置を選択する手段と、

前記複数のデータ入力装置のそれぞれへ、選択された前記複数の演算装置に対して複数の前記部分データを送信できるように、前記複数の演算装置の情報を通知する手段とをさらに備えることを特徴とする請求項 18～22 のいずれか 1 項に記載の統計処理結果提供サーバ装置。

[請求項24]

秘匿すべきオリジナルデータを取得する手段と、

全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを秘密の比率に従って分割し、所定の個数の部分データを生成する手段と、

それぞれが複数の入力データに基づいて所定の演算を行う手段を有する複数の演算装置のうちの対応する演算装置への前記複数の入力データの 1 つとして、前記所定の個数の部分データのそれぞれを、保護された通信路により送信する手段とを備えるデータ入力装置であって、

前記複数の演算装置のそれぞれが、複数の前記データ入力装置からの部分データに基づいて前記所定の演算を行った結果を、前記複数の演算装置とは異なるサーバ装置が利用することにより、複数の前記データ入力装置により取得された複数の前記オリジナルデータに基づく統計処理の結果が、該オリジナルデータが秘匿されたまま求められることを特徴とするデータ入力装置。

[請求項25]

送信した前記所定の個数の部分データを、それぞれ対応する前記演算装置に、許可された者からのみアクセスできるように保存させる手段と、

取得した前記オリジナルデータの記憶を消去する手段とをさらに備え、

前記オリジナルデータは、前記許可された者によりそれぞれ対応する前記演算装置から取得された前記所定の個数の部分データに基づいて復元されるものとすることを特徴とする請求項24に記載のデータ入力装置。

[請求項26] 前記サーバ装置にアクセスするための情報を記憶する手段と、
前記対応する演算装置を特定するための情報を前記サーバ装置から受信する手段とをさらに備えることを特徴とする請求項24又は25に記載のデータ入力装置。

[請求項27] 前記部分データに、システムにおいて一意となる識別情報を付与する手段と、
前記識別情報に基づいて求められる値が、前記複数の演算装置のそれぞれが担当する範囲のうちのいずれの範囲に属するかに従って、前記対応する演算装置を特定する手段とをさらに備えることを特徴とする請求項24又は25に記載のデータ入力装置。

[請求項28] 一つのオリジナルデータから得られる部分データの全てがいずれかの演算装置に受信されたことを確認した後に、確認済みであることを示す情報をいずれかの演算装置へ送信して登録する手段をさらに備えることを特徴とする請求項24～27のいずれか1項に記載のデータ入力装置。

[請求項29] 秘匿されるべきオリジナルデータを取得することなく、複数の前記オリジナルデータに基づく統計処理の結果を提供するサービスのためのサーバ装置と通信する手段と、

それぞれが前記オリジナルデータを自装置の内部に秘匿する手段を有する複数のデータ入力装置から、複数の前記オリジナルデータのそれぞれに属する部分データを受信する手段と、

複数の入力データに基づいて所定の演算を行う手段とを備える演算装置であって、

前記サーバ装置は、複数の前記演算装置からの演算結果に基づいて

、所定の統計処理を行うものであり、

前記演算装置は、

前記複数のデータ入力装置から受信した複数の前記部分データのうち、該部分データに対応して、一つのオリジナルデータから得られる部分データの全てがいずれかの演算装置に受信されたことが確認済みであることを示す情報が登録されているものを、前記入力データとして選択する手段と、

選択された前記入力データについての前記所定の演算の結果を、前記サーバ装置へ送信する手段とをさらに備えるものであることを特徴とする演算装置。

[請求項30]

他のコンピュータと通信する機能を有するコンピュータを、データ秘匿型統計処理システムにおけるデータ処理装置として動作させるためのプログラムであって、

前記他のコンピュータとして、それぞれが複数の入力データに基づいて所定の演算を行う手段を有する複数の演算装置があり、

前記データ処理装置は、秘匿されるべきオリジナルデータを取得することなく、複数の前記オリジナルデータに基づく統計処理の結果を提供するものであって、

前記プログラムは、前記コンピュータに、

前記複数の演算装置のそれぞれに、前記オリジナルデータの部分データを前記入力データとして演算を行わせ、該演算の結果を取得する手段と、

前記複数の演算装置からの演算結果に基づいて、所定の統計処理を行う手段とを備えさせるものであり、

複数の前記部分データは、全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを秘密の比率に従って分割することにより、生成されたものであることを特徴とするプログラム。

[請求項31] 秘匿すべきオリジナルデータを取得する機能と他のコンピュータと通信する機能とを有するコンピュータを、データ秘匿型統計処理システムにおけるデータ入力装置として動作させるためのプログラムであって、

前記他のコンピュータとして、それぞれが複数の入力データに基づいて所定の演算を行う手段を有する複数の演算装置があり、

前記プログラムは、前記コンピュータに、

全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを秘密の比率に従って分割し、所定の個数の部分データを生成する手段と、

前記所定の個数の部分データのそれぞれを、前記複数の演算装置のうちの対応する演算装置への前記複数の入力データの1つとして、保護された通信路により送信する手段とを備えさせるものであり、

前記複数の演算装置のそれぞれが、複数の前記データ入力装置からの部分データに基づいて前記所定の演算を行った結果を、前記複数の演算装置とは異なるサーバ装置が利用することにより、複数の前記データ入力装置により取得された複数の前記オリジナルデータに基づく統計処理の結果が、該オリジナルデータが秘匿されたまま求められることを特徴とするプログラム。

[請求項32] 他のコンピュータと通信する機能を有するコンピュータを、データ秘匿型統計処理システムにおける複数の演算装置のうちの一つとして動作させるためのプログラムであって、

前記他のコンピュータとして、秘匿されるべきオリジナルデータを取得することなく、複数の前記オリジナルデータに基づく統計処理の結果を提供するサービスのためのサーバ装置と、それぞれが前記オリジナルデータを自装置の内部に秘匿する手段を有する複数のデータ入力装置とがあり、

前記プログラムは、前記コンピュータに、

前記複数のデータ入力装置から、複数の前記オリジナルデータのそれぞれに属する部分データを受信する手段と、

複数の入力データに基づいて所定の演算を行う手段と、

前記複数のデータ入力装置から受信した複数の前記部分データのうち、該部分データに対応して、一つのオリジナルデータから得られる部分データの全てがいずれかの演算装置に受信されたことが確認済みであることを示す情報が登録されているものを、前記入力データとして選択する手段と、

選択された前記入力データについての前記所定の演算の結果を、前記サーバ装置へ送信する手段とを備えさせるものであり、

前記サーバ装置は、前記複数の演算装置からの演算結果に基づいて、所定の統計処理を行うものであることを特徴とするプログラム。

[請求項33]

秘匿すべきオリジナルデータを取得する手段を備える複数のデータ入力装置のそれぞれが、全ての部分を合わせると前記オリジナルデータが復元されるように前記オリジナルデータを秘密の比率に従って分割して得られた所定の個数の部分データを出力し、

複数の入力データに基づいて所定の演算を行う手段を備える複数の演算装置のそれぞれが、複数の前記データ入力装置のそれぞれから出力された前記部分データを前記入力データとして前記演算を行った結果を出力し、

データ処理装置が、前記複数の演算装置のそれぞれから出力された前記演算の結果を利用することにより、前記複数のデータ入力装置により取得された複数のオリジナルデータに基づく統計処理の結果を、該オリジナルデータを取得することなく求めることを特徴とする統計処理結果提供サービス方法。