秘匿計算を用いた医療統計システムの構成

諸橋 玄武　　五十嵐 大　　千田 浩司

日本電信電話株式会社　NTT セキュアプラットフォーム研究所
〒 180-8585 東京都武蔵野市緑町 3-9-11
{morohashi.gembu, ikarashi.dai, chida.koji}@lab.ntt.co.jp

あらまし 医学研究などに利用される医療データは患者のプライバシーを守るため高度なセキュリティが要求される．本稿では医療データを秘匿計算を用いて統計処理するシステムにおいて，サーバの構成方法や役割に応じて，送受信されるデータを保護するタイミングについて考察し，安全なシステムの構成を示す．

Statistical Analysis System for Clinical Study Using Secure Multi-party Computation

Gembu Morohashi　　Dai Ikarashi　　Koji Chida

NTT Secure Platform Laboratories, NTT Corporation
3-9-11 Midori-cho, Musashino-shi, Tokyo 180-8585, JAPAN

Abstract In clinical study, it is needed advanced security to preserve patients' privacy. In this paper, we propose a system architecture of the secure multi-party computation for clinical studies.

1　はじめに

昨今クラウドをはじめとする ASP，SaaS といったサービスが普及し，業務効率化やコストダウンの観点からこれらのサービスを利用する機会が増してきている．また，東日本大震災の教訓等から BCP（Business Continuity Plan: 事業継続計画）の策定が求められており，基幹システムの分散配置や冗長化，重要なデータの遠隔バックアップ等にクラウドの利用が検討されている．一方で，これらのサービスを利用する場合はサーバやストレージ機器等の運用については基本的にサービス事業者に任されており，預けた情報がどのサーバやストレージ機器で取り扱われているか，意図しないような運用がなされていないかといったことを利用者側で制御するのは難しい．特に，企業における機密情報や顧客情報等，取り扱いに注意を要するような情報を預けている場合には，（サービス事業者による）情報の喪失や漏洩のリスクについても注意を払わなければならない．

秘密分散はそのような観点から注目されている技術である．秘密分散によって保存したいデータを複数箇所に分散させ，断片データからは元のデータに関する情報の漏洩を防ぐことが可能であり，さらに閾値秘密分散法を利用すれば，すべての断片をそろえなくても元のデータを復元することが可能なため，可用性の担保といった側面も併せ持っている．

医療分野においても，複数の医療機関での共有やデータの共通化を狙って情報化インフラの整備が進められ，診療報酬明細（レセプト）や診療録（カルテ）の電子化、オンライン化が推進されてきている．これら医療情報を取り扱う

システムについても ASP，SaaS の利用拡大が見込まれており，2010 年に厚生労働省から出されている医療情報システムの安全管理に関するガイドライン [1] において，それまで認められていなかった診療録等の医療機関外での保存に関する記述が盛り込まれた．それを受けて総務省から ASP，SaaS 事業者を対象としたガイドライン [2] が提示され，医療情報を取り扱う場面においてはより高度なセキュリティが求められている．

病気の原因の解明や予防，診断，治療の改善を目的とした医学研究の分野では，より多くの症例を集めるために，複数の医療機関が参加する共同研究グループが設立され，収集された医療情報を分析する取り組みが行われている．臨床現場での様々な医療情報を扱うため，(そもそもの提供元である）患者のプライバシー保護の観点からもデータの取り扱いには細心の注意が必要なことから，データを安全に取り扱いつつ，医療統計分析を実施することが可能な技術が求められている．このようなデータの保護と活用の両立の実現を目指すプライバシー保護データ分析（PPDA: Privacy-Preserving Data Analysis）の研究が近年盛んになってきている．PPDA を実現する技術の一つに秘匿計算（秘匿関数計算，マルチパーティ計算）がある．千田ら [3, 4] は，分散された入力値を復元することなく任意の演算結果を得ることが可能な 3 パーティの協調計算による秘匿計算方式を示し，濱田ら [5] によって医療分野における統計分析が現実的な時間で処理できることが示されている．

本稿では，五十嵐らの方式 [4] を実用に即したシステムとして実現するため課題とそれを解決するシステムの構成方法について考察する．

2 関連研究

入力された値に関する情報を秘匿したまま任意の演算を処理できる秘匿計算は，Yao [6] による提案から理論的な研究が進められてきた．単純な集計を求める電子投票 [7] は秘匿計算の典型例として知られているが，加算のみならず任意の演算を扱えるような秘匿計算については，処理速度に課題があり，実用に耐えられる方式が出始めたのは近年になってからである．実用化を目指した秘匿計算の開発については，FairplayMP [8, 9] や Sharemind [10, 11] といったプロジェクトや辻井らによる研究 [12] があり，近年盛んになりつつあるが，実装例はまだ少ない．

3 秘匿関数計算

本システムで扱う 3 パーティ秘匿計算 [4] は，秘密分散を用いて入力値を 3 つに分散し，それらを復元することなく算術/論理演算の結果が得られるプロトコルで，semi-honest な 2 out of 3 プロトコルである．すなわち，参加者が正しくプロトコルに従っていれば，分散された 3 つの値のうち任意の 2 つから元の値に復元できるが，断片 1 つのみでは，元の値について何もわからない．また，参加者 1 人のみでは不正に情報を得ることはできない．

3.1 秘密分散

入力: $a \pmod p$ (p は素数)

出力: 参加者 P_1, P_2, P_3 への分散値 $(a_0, a_1), (a_1, a_2), (a_2, a_0)$

1. 乱数 a_0, a_1 を生成する．
2. $a \equiv a_0 + a_1 + a_2 \pmod p$ なる a_2 を求める．
3. P_1 に (a_0, a_1)，P_2 に (a_1, a_2)，P_3 に (a_2, a_0) をそれぞれ出力として渡す．

復元については参加者のうち任意の 2 人の断片が得られれば a_0, a_1, a_2 が得られ，$a \equiv a_0 + a_1 + a_2 \pmod p$ が求められる．

3.2 秘匿関数計算

五十嵐ら [4] の秘匿計算では，入力値を分散した値に対して，加減算，定数和，定数倍，乗算，論理演算（論理積，論理和，否定）の結果を分散された状態で得ることができる．(計算結

果を最終的に得るには，分散された結果を復元する.）

また，比較演算，最大値，最小値，中央値，集合関数演算（集計）は五十嵐ら [4] の方式に対応した濱田ら [5] の方式を踏襲する．

4　システム構成のための要件

本節では秘匿計算システムの構成についての考察を示す．はじめにプロトコルの参加者（システム利用者，サーバ運用者などの登場人物）に対応したハードウェア構成とセキュリティモデルについて考察する．次にそれぞれ持つべき機能のコンポーネント（ノード別に持つべき機能）について，さらにノード間の通信方法，インタフェースについて考察する．

4.1　システムの構成要素とセキュリティモデル

3 パーティ秘匿計算では秘匿計算（を実現するために行う協調計算）の参加者は 3 人であり，そこへデータを入力/入力されたデータを用いた分析を行うユーザが少なくとも 1 人いる．必要最小限の構成要素としては以上の 4 人だが，複数の医療機関からデータを収集することを想定すると，ネットワークを介して様々な場所から複数のユーザがデータの入力，分析を行うことになるため，システムのポータルとなるサーバを用意することを考える．秘匿計算を行ういずれかの参加者（サーバ）と兼ねても構わないが，サーバへの負荷やネットワーク設計上，独立したサーバとしておく．すると，システムの（ハードウェア単位の）構成と，それぞれの概要は以下のようになる.（図 1）

クライアント PC　分析対象となる情報を登録・分散する入力用端末と，蓄えられた情報を秘匿計算によって分析するための分析用端末の 2 種類がある．2 つの機能をまとめ，端末を 1 台で構成しても構わない.

Web サーバ　入力用のフォームや分析演算の指定，演算結果の表示のためのユーザインタフェースと，秘密分散および復元をクライアント PC において実行するためのスクリプトを提供する．また，クライアント PC と秘匿計算サーバのデータ送受信の中継を担う.

秘匿計算サーバ（3 台）分散されたデータの保存と，それらを用いた秘匿計算の実行し，計算結果の断片をクライアント PC へ送付する.

データ入力時はクライアント PC から分散された入力値の断片をそれぞれの秘匿計算サーバへ送付するが，中継する Web サーバが複数の断片を収集したり，複数の秘匿計算サーバと結託されてしまう恐れがある．また，中継サーバによる成りすましについても同様に防がなければならない．そのため，クライアント PC と秘匿計算サーバの間でデータの断片をやりとりする場合は，断片データの秘匿性と完全性を担保しなければならない.

4.2　機能コンポーネント

- クライアント PC

  秘密分散/復元ライブラリ　入力されたデータの分散と分散されたデータの送信，サーバから断片データ（計算結果）の取得と復元を行う.

  分析機能ユーザインタフェース　秘匿計算の演算のリクエストと，演算結果の表示を行う．秘匿計算による統計分析演算は 3.2 節で示した基本演算の組み合わせによって実現するため，どの演算をどの順番で，どのデータを対象に行うかといった情報を組み立ててリクエストを行う.

- Web サーバ

  分散データ送受信部　クライアントにて分散された入力値を 3 台の秘匿計算サーバにそれぞれ送付する．また，秘密計算サーバ側から送付される秘匿計算

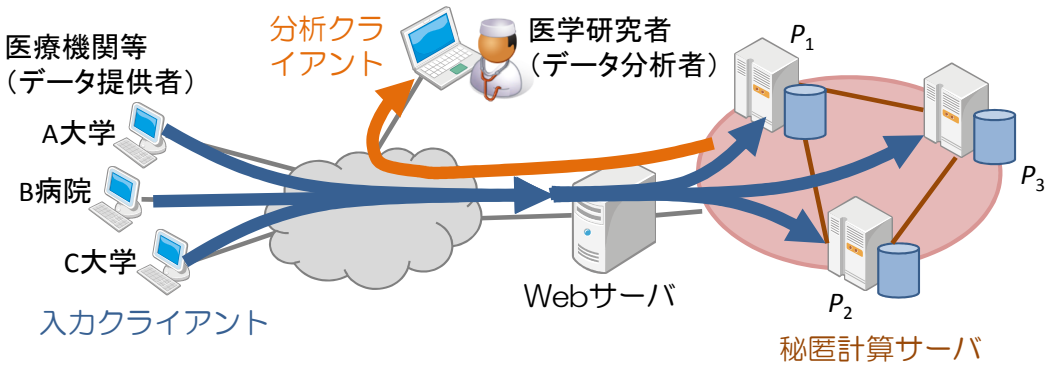

図 1: システムの構成

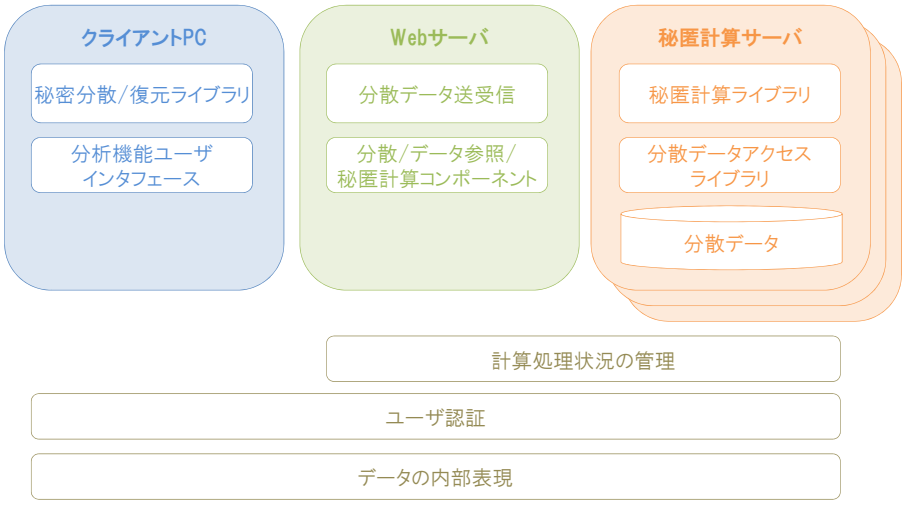

図 2: コンポーネント図

結果の断片をクライアント PC に送付する．

分散/データ参照/秘匿計算コンポーネント 秘匿計算の（クライアントからの）リクエスト，秘匿計算サーバの状態（計算中か，完了か）の参照，および秘匿計算結果の取得を行う．

- 秘匿計算サーバ

秘匿計算ライブラリ 3.2 節で示した秘匿計算の基本演算ライブラリ．

秘匿計算 DB 制御 秘匿計算サーバで保存しているデータの断片を制御する．秘密分散/復元操作のみならず，秘匿計算の演算対象となるデータを参照する場合にも必要となる．

4.3 通信内容の暗号化

4.3.1 クライアント側による秘密分散

前述の通り，クライアント PC と秘匿計算サーバ間で断片データを送受信する場合には，断片データの安全性を確保しなければならない．秘密分散機能や暗号化機能を盛り込んだクライアント用プログラムを用意すれば，秘匿計算サーバとの情報のやりとりに際して，データを安全に送付する手段を容易に提供できるが，データ提供元となる医療機関が多岐にわたり，クライアント PC の種類（OS やソフトの対応状況，スペックなど）を統一するのは難しい．個別のソフトウェアを用意すると，システム利用者側にも初期設定等の負担をかけてしまうことから，秘密分散や復号処理の機能を一般的な Web ブラ

ウザで実行可能になるように，JavaScript によるライブラリを用意するのが理想的である．(関連研究として小田ら [13] は共通鍵暗号，公開鍵暗号，電子署名，楕円演算など様々な暗号プロトコルに必要な演算を JavaScript を用いて実現し，ユーザの負担を軽減しつつ，ASP・SaaS 利用時の安全性を高める方式を提案している．)

4.3.2 通信経路での暗号化

クライアント PC と秘匿計算サーバ間でデータのやりとりを行うと，Web サーバが中継するかたちで必ず介入してくる．先に述べたとおり，Web サーバが他の秘匿計算サーバと結託し不正がはたらかないようにするためには，Web サーバにデータを盗み見られないよう暗号化しておく必要がある．

5 システム全体の機能要件

本節では，4 節のシステム構成，機能コンポーネントに共通する，もしくは横断的な機能要件について考察する．

5.1 データの内部表現

効率的な秘匿計算を行うため，演算対象となるすべてのデータはあらかじめ定められた素数 p によって固定ビット長となる．そのため，数値などは変換を行う必要がある．濱田らによるソートアルゴリズム [14] では，固定小数点の数値しか扱えないため，整数以外の数値はある程度制限された値の範囲内でしか利用できない．入力値のうち，どの属性の値が何桁で収まるのか，といった利用データ依存の設計になってしまうことを避けるため，属性ごとに利用される型などについて，スキーマと呼ぶ平文情報をあわせて秘匿計算サーバには格納しておく．スキーマに書かれた情報に基づいて処理を行うことにより，分散された値の取り扱いで不都合が生じないようにしている．

文字列についても，いくつかの固定ビット長整数の単位に変換して取り扱われる．このとき，実際の文字列の長さによらず，文字列の最大長まで文字が入っているものとみなして，分散や秘匿計算を行っている．文字列を可変長にしてしまうと，文字列の長さに関する情報が知られてしまう可能性があるため，文字列については固定長で取り扱う．

5.2 秘匿計算サーバでの処理状況の管理

分散された情報に対する操作を行う場合，3 台のサーバのうちいずれかで処理漏れが生じると，分散データの不整合が起きてしまう．そのため，正しい順番で処理が行われていることを担保する仕組みが必要となる．各秘匿計算サーバに演算処理のリクエストを送信する Web サーバ側で，処理の行われる順に番号を付与し，秘匿計算サーバ側はその番号に基づいて処理の前後関係を確認する．パケットロスなどの要因により処理が行われない，あるいは処理に失敗した場合，秘匿計算サーバから完了通知を送る仕組みを備えることにより Web サーバ側で管理できる．(正常な応答がなかった場合には一定時間リトライをかける．) 秘匿計算サーバ側だけでこのような仕組みを実現しようとすると，お互いの処理状況（対象データがどこでどのような演算を行っているか）が知られ，秘匿計算サーバ同士の結託に結びつかないような方法を採るには検証方法が複雑化してしまう．

5.3 ユーザ認証

Web サーバによるなりすましを防ぐため，クライアント PC を使う利用者の認証においても，秘匿計算を利用した認証を行うことが望ましい．PKI 認証など仕組みを導入すれば回避できるが，クライアント PC への設定等の負担軽減から，従来のパスワード認証を採用した場合，パスワード情報をどこに持てばよいかが問題になる．Web サーバやいずれかの秘匿計算サーバに置いてしまうと，不正に利用される恐れがある．そのため，システムに登録するパスワード情報を分散し，認証の際は与えられたパスワードが一致するかどうかを協調計算によって求める．

6 まとめ

本稿では千田らによる秘匿計算の実用化を目指したシステムの構成について考察した．現在本稿での考察を元にシステムの実装を進めており，完成したシステムを利用した実証実験等を通じて実用化の検証に取り組んでゆく．

参考文献

[1] 厚生労働省．医療情報システムの安全管理に関するガイドライン．第 4.1 版，2010. http://www.mhlw.go.jp/shingi/2010/02/s0202-4.html.

[2] 総務省．「ASP・SaaS 事業者が医療情報を取り扱う際の安全管理に関するガイドラインに基づく SLA 参考例」及び「ASP・SaaS 事業者が医療情報を取り扱う際の安全管理に関するガイドライン第 1.1 版」の公表．報道資料，2012. http://www.soumu.go.jp/menu_news/s-news/01ryutsu02_01000009.html.

[3] 千田，濱田，五十嵐，高橋.“ 軽量検証可能 3パーティ秘匿関数計算の再考 ”. CSS 2010 予稿集，2010，2B4-1.

[4] 五十嵐，千田，濱田，高橋.“ 計量検証可能 3パーティ秘匿関数計算の効率化及びこれを用いたセキュアなデータベース処理 ”．暗号と情報セキュリティシンポジウム（SCIS 2011）予稿集，2011，2C3-6.

[5] 濱田, 大竹, 五十嵐, 竹之内, 千田, 冨士, 高橋, 村田, 熊田.“ 秘匿関数計算システムによる医療データのプライバシ保護統計分析 ”. 信学技報, vol.111, no.470, LOIS2011-102, 2012，pp.177–181.

[6] Yao, A. C. “Protocols for secure computations.” Proc. of FOCS ’82. Los Alamitos, IEEE Press, 1982, pp.160–164.

[7] Damgård, I.; Jurik, M. “Client/server tradeoffs for online elections.” Public Key Cryptography 2002. Berlin, Springer-Verlag, 2002, pp.125–140, (LNCS 2274).

[8] FairplayMP. http://www.huji.ac.il/project/Fairplay/FairplayMP.html, (accessed 2012-08-24).

[9] Ben-David, A.; Nisan, N.; Pinkas, B. “FairplayMP — A secure multi-party computation system.” Proceedings of the ACM CCS 2008. pp.257–266.

[10] Cybernetica. “Sharemind: Your secure service platform for data collection and analysis.” Cybernetica White Paper. http://sharemind.cyber.ee/files/images/Sharemind%20secure%20service%20platform%202012.pdf, (accessed 2012-08-24).

[11] Bogdanov, D.; Jagomägis, R.; Laur, S. “A universal toolkit for cryptographically secure privacy-preserving data mining.” Intelligence and Security Informatics Pacific Asia Workshop (PAISI 2012). Kuala Lumpur, Malaysia, 2012-05-29. Springer, 2012, pp.112–126, (LNCS 7299).

[12] 辻井，山口，土井，角尾，藤田，庄司.“ Paillier 暗号と RSA 暗号の連携による暗号化状態処理の一方式—第 2 報 ～電子行政・医療介護ネットワークにおける個人情報の保護と利用の両立を目指して～ ”．信学技報，vol.112，no.126，ISEC2012-33，2012，pp.175–179.

[13] 小田，諸橋.“ Web 暗号 ”．信学技報，vol.102，no.272，LOIS2009-51，2009，pp.75–80.

[14] 濱田，五十嵐，千田，高橋.“ 秘匿関数計算上の線形時間ソート ”. 暗号と情報セキュリティシンポジウム（SCIS 2011）予稿集，2011，2C3-5.