応専

安全な情報処理を目指す秘密計算技術の研究動向と実用化に向けた取り組み

千田浩司（NTTセキュアプラットフォーム研究所）

入力プライバシー保護

最近では，「パーソナルデータ☆1」を社会や産業の発展のために二次利用する機運が高まってきている．しかしパーソナルデータの扱いは，情報提供者である個人のプライバシーについて十分配慮する必要がある．

パーソナルデータを扱う際のプライバシーの考え方として，入力プライバシーと出力プライバシーがある☆2．パーソナルデータを用いた情報処理を行う際，入力データから生じ得るプライバシーの問題が入力プライバシーであり，その対策は入力プライバシー保護と呼ばれる．同様に，計算結果から生じ得るプライバシーの問題が出力プライバシーであり，その対策は出力プライバシー保護と呼ばれる．

入力プライバシーについて，パーソナルデータの本来の利用者（一次利用者）がたとえばクラウドに情報処理を依頼すれば，クラウドにもパーソナルデータが渡ることになる．また一次利用者とは異なる二次利用者にパーソナルデータが渡る場合も想定される．したがって**図-1**に示すように，パーソナルデータの扱いを巡ってさまざまなステークホルダが入力プライバシーの問題に直面することになる．

出力プライバシー保護については，公的統計の分野で古くから統計的開示抑制（Statistical Disclosure Control）と呼ばれる実務的な研究が進められている．また最近では，プライバシー保護データマイニング（PPDM：Privacy Preserving Data Mining）の研究分野においても目覚ましい研究成果が見られる．詳しくは本特集の「6. データベース問合せにおけるプライバシー保護モデル」を参照されたい．

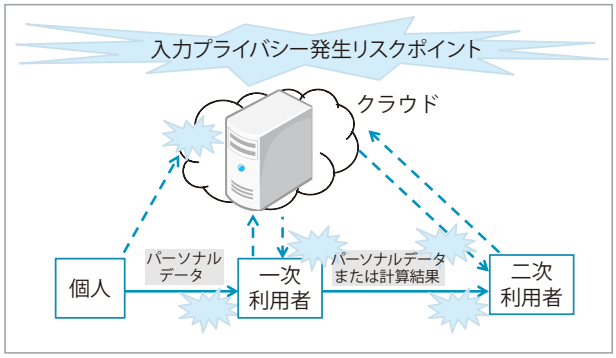

図-1　入力プライバシーの発生リスクポイント

本稿では，有望な入力プライバシー保護技術として期待される「秘密計算」（Secure Computation）☆3技術に焦点を当て，研究動向や仕組み，そして実用化に向けた取り組みについて紹介する．

秘密計算の研究動向

暗号理論では古くから，所定の情報処理（関数）の入力データを秘匿しつつ計算結果を導出可能な技術である秘密計算の研究が進められている．特に最近では，パーソナルデータを扱う際の情報漏洩を根本的に防ぐ新しい情報処理技術として，秘密計算への期待が高まっている．

秘密計算の研究は，汎用的な情報処理が可能な方式と，個別の情報処理に特化した方式に大別され

☆1 個人に関する情報全般を指す用語．総務省「パーソナルデータの利用・流通に関する研究会」（http://www.soumu.go.jp/main_sosiki/kenkyu/parsonaldata/）等，「個人情報の保護に関する法律」（個人情報保護法）で定義される個人情報に限らず，より広い意味で用いられる．

☆2 ここでの「入力」，「出力」は，個人情報保護法における「取得」，「提供」に相当する．

☆3 秘匿関数計算（Secure Function Evaluation）等とも呼ばれる．

る．前者については，2009 年に Craig Gentry が提案した完全準同型暗号[1] に基づく方式が注目を集めている．これは暗号化されたデータを入力として，誰もが暗号化された計算結果を求めることができ，真の計算結果に戻せるのは復号鍵を所持する者のみという特徴を持つ（**図 -2**）．したがってクラウド等の情報処理主体に対して入力データも計算結果も知られずに済む．しかし処理時間が課題であり，大規模なデータ分析や複雑な処理に対して実用レベルの性能を実現するためには，今後さらなる研究のブレークスルーが求められるだろう．

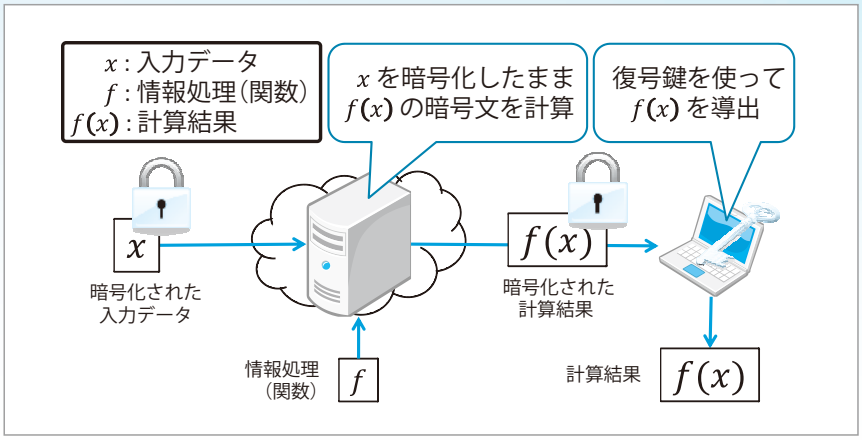

図 -2　完全準同型暗号に基づく秘密計算の処理イメージ

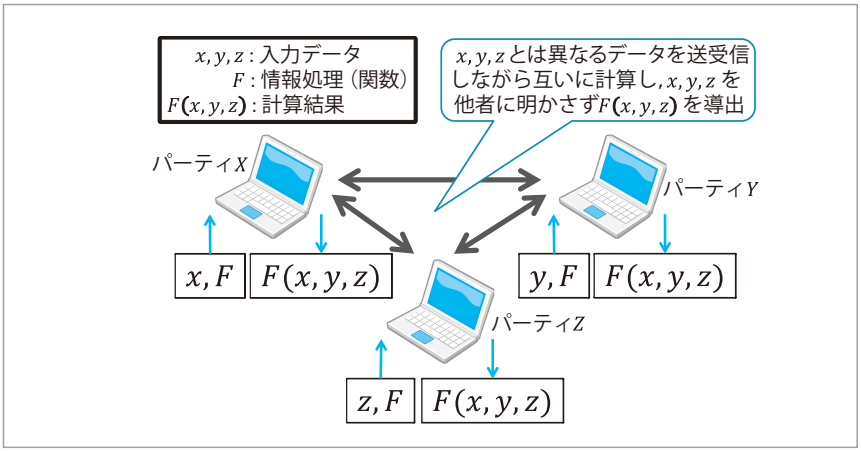

図 -3　セキュアマルチパーティ計算の処理イメージ

また 1986 年に Andrew Yao が提案して以来，現在も発展を続けているセキュアマルチパーティ計算[2] に基づく汎用的な秘密計算も代表的なアプローチとして知られる．セキュアマルチパーティ計算は，複数人がそれぞれ秘密のデータを持ち，互いに秘密のデータを明かすことなく所定の情報処理の計算結果を導出する手法である（**図 -3**）．セキュアマルチパーティ計算も処理時間が課題とされてきたが，アルゴリズム改良の研究が進み，ある程度汎用的な情報処理が実用レベルの性能で実現できるようになってきた．

一方，後者の個別の情報処理に特化した秘密計算は，主に汎用的な秘密計算よりも処理時間を抑える目的で設計され，PPDM の研究分野等でさまざまな方式が提案されている．たとえば PPDM の先駆的な研究成果として，ID3 アルゴリズムによる決定木を計算するための，セキュアマルチパーティ計算に基づく秘密計算が提案されている[3]．また MONOMI[4] 等，データベースを暗号化しつつ各種要求を処理できるシステムも実用段階に近づいている．ただし処理できる演算は限られており，多様な情報処理を実現するためには，汎用的な秘密計算の援用等が求められるだろう．

入力プライバシー保護に有効な秘密計算

入力プライバシー保護の満たすべき要件は，各ステークホルダが受け取るパーソナルデータを利用目的に沿って必要最小限とし，かつ各ステークホルダからのパーソナルデータ漏洩のリスクを極力抑えることであろう．すると図 -1 の例において，クラウドが秘密計算によって情報処理を行えば，クラウドにパーソナルデータを一切開示せずに済む．さらに個人が自身のパーソナルデータを秘匿処理したうえでクラウドに直接提供すれば，一次利用者や二次利用者でさえも，パーソナルデータについて所定の計算結果以外の情報を知ることができない．したがっ

て，秘密計算は前記の要件を非常に高いレベルで満たすことが期待される．

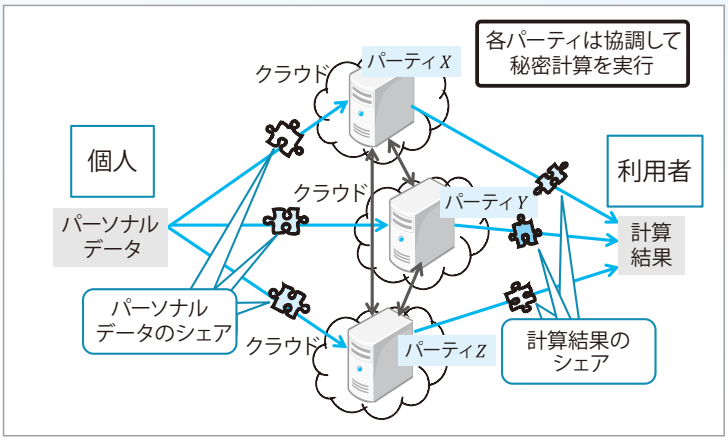

図 -4　セキュアマルチパーティ計算に基づく秘密計算の全体像の例

◆ 汎用的な情報処理を実現する秘密計算の仕組み

汎用的な情報処理を実現する秘密計算の代表的な構成要素は，入力データを秘匿しつつ加減算や乗算を実現する仕組みである．たとえば a, b を秘密の値，$E()$ を暗号化関数としたとき，a, b の暗号文 $E(a), E(b)$ から，a や b を元に戻さず加減算結果の暗号文 $E(a \pm b)$ や乗算結果の暗号文 $E(ab)$ が得られるようにする．特に a, b が 1 ビットの値であれば，$E(ab)$ は a,b の論理積 $a \wedge b$ の暗号文と見なすことができる．また a, b の論理和 $a \vee b$ は $a+b-ab$ と加減乗算で表現できるため，a, b の論理和の暗号文 $E(a+b-ab)$ も計算可能となる．同様に a, b の排他的論理和の暗号文 $E(a+b-2ab)$ や，a の否定の暗号文 $E(1-a)$ 等も計算可能となる．

以上から，入力データをビット列と見なすことで，入力データを秘匿しつつ汎用性の高い組合せ論理回路（出力が現在の入力だけで決まる論理回路）の演算が可能となる．また最近では，ソーティング等，組合せ論理回路では非効率な演算について，組合せ論理回路を直接用いない実現方式の研究も進められている．

◆ 具体例

文献 5）を例に，セキュアマルチパーティ計算に基づく秘密計算の加減算および乗算の仕組みを簡単に紹介する．全体像を**図 -4** に示す．なお加減算や乗算は前述したように組合せ論理回路の秘密計算の構成要素であり，後述するように各種統計や医療分析等の要素演算としても用いられる．

まず入力データの秘匿処理として秘密分散法（Secret Sharing Scheme）を用いる．秘密分散法の基本原理は，$N \geqq K \geqq 2$ を満たす整数 N, K について，入力データを符号化して N 個のデータ（各データをシェアと呼ぶ）を求め，K 個未満のシェアからは入力データは一切分からないが，任意の K 個のシェアが集まれば入力データを復元できるというものであり，N=3, K=2 としたとき，以下に示すように単純に実現可能である．

秘密分散の例

入力：2 以上の整数 m ,0 以上 m 未満の整数 a

出力：a のシェア

処理手順：

i.　0 以上 m 未満の整数 a_0, a_1 をランダムに生成

ii.　a_2:=a-a_0-a_1 (mod m) を計算

iii.　$(a_0, a_1), (a_1, a_2), (a_2, a_0)$ を a のシェアとして出力

a-a_0-a_1 (mod m) は，a-a_0-a_1 を m で割った余りを表す．ここで 3 つの整数の組からなる a のシェア (a_0, a_1)，$(a_1, a_2), (a_2, a_0)$ について，1 つのシェアでは a を復元できないが，2 つのシェアが揃えば a_0, a_1, a_2 が揃うため，a=a_0+a_1+a_2 (mod m) を復元できる．そのため不正者が容易に 2 組揃えられないよう，各整数の組は別々の主体（パーティ）が管理し，各パーティがそれぞれアクセス制御を行うことが運用上望ましい．

次に秘密計算の加減算の例について説明する．a のシェア同様，0 以上 m 未満の整数 b のシェアを $(b_0, b_1), (b_1, b_2),(b_2, b_0)$ とすれば，c_0:= $a_0 \pm b_0$ (mod m), c_1:= $a_1 \pm b_1$ (mod m), c_2:= $a_2 \pm b_2$(mod m) として，$(c_0, c_1), (c_1, c_2), (c_2, c_0)$ は c:=$a \pm b$ (mod m) のシェアとなる．すなわち，シェア同士を要素ごとに加減算すればよい．

最後に秘密計算の乗算の例について説明する．乗算の場合は，単純にシェア同士を要素ごとに乗算しても成り立たない．そのため，各パーティは自身が所持するシェアをほかのパーティに明かさず，シェアそのものとは異なる値をほかのパーティと送受信しながら乗算のシェアを求める．

秘密計算の乗算の例

共通入力：2 以上の整数 m

パーティ X の入力：a, b のシェア $(a_0, a_1), (b_0, b_1)$

パーティ Y の入力：a, b のシェア $(a_1, a_2), (b_1, b_2)$

パーティ Z の入力：a, b のシェア $(a_2, a_0), (b_2, b_0)$

パーティ X, Y, Z の出力：c:=$ab \pmod m$ の各シェア

処理手順：

i. パーティ X は以下を行う

（ア）0 以上 m 未満の整数 c_0, r_1, r_2 をランダムに生成

（イ）c_1:=$(a_0+a_1)(b_0+b_1)-r_1-r_2-c_0 \pmod m$ を計算

（ウ）パーティ Y, Z にそれぞれ $(c_1, r_1), (c_0, r_2)$ を送信

（エ）(c_0, c_1) を c のシェアの 1 つとして出力

ii. パーティ Y, Z はそれぞれ y:=$a_1b_2+a_2b_1+r_1 \pmod m$, z:=$a_2b_0+a_0b_2+r_2 \pmod m$ を計算してパーティ Z, Y に送信

iii. パーティ Y, Z は c_2:=$y+z+a_2b_2 \pmod m$ を計算し，それぞれ $(c_1, c_2), (c_2, c_0)$ を c のシェアの 1 つとして出力

シェアが正しいことは，c=$c_0+c_1+c_2 \pmod m$ が成り立っていることから確認できる．また，各パーティがほかのパーティより受信したデータから，ほかのパーティのシェアを推定することは理論的に不可能であることが保証されている．

実用化に向けた取り組み

最近ではセキュアマルチパーティ計算に基づく秘密計算の開発プロジェクトがいくつか立ち上がっており，具体的なパーソナルデータの活用を見据えた動きも見られる．

エストニアの開発プロジェクト Sharemind[☆4] は，秘密計算のソフトウェア開発キットを公開しており，利用者が秘密計算による情報処理の実行コードを作成できるようにしている．2013 年 2 月には，秘密計算による大規模なゲノム解析のプロトタイプを開発したことを報告している．3 億個のデータの情報処理に相当する，1,000 人分のゲノムおよそ 300,000 カ所の解析が可能であるという．個人の遺伝子型は最もセンシティブな情報の 1 つであるとし，ゲノムバンクやパーソナルゲノムサービス提供者に対するプライバシー保護の必要性について述べている．また，人工衛星の所有者たちが自身の人工衛星の軌道情報を互いに明かさず衝突確率を求める実証も行っているようである．

スイスの開発プロジェクト SEPIA[☆5] は，秘密計算によって各種演算を実行するための Java クラスライブラリを公開している．特にネットワーク管理を行うための演算が充実している．各ドメインのアラート情報，異常検知，ネットワークパフォーマンス統計等を集約して分析すれば，ネットワーク管理に有用であろう．しかし各ドメインの管理主体が異なれば，ネットワークデータの提供によって外部から攻撃されるリスクが増し，またネットワークデータは一般にパーソナルデータを含むため，安易な集約は難しい．そこで秘密計算によって各ドメインのネットワークデータを秘匿しつつネットワーク管理に必要な各種分析を行うツールが提供されている．SEPIA では，加算，乗算，等号判定，大小比較といった要素演算の秘密計算に加え，イベント相関，エントロピー（Tsallis entropy），集合演算（Set Intersection, Set Intersection Cardinality, Set Union）等の秘密計算もサポートされている．

筆者らは，疫学研究等における診療情報の安全な二次利用の実現に向け，秘密計算による医療分析が実行可能なプロトタイプの開発や評価実験を行っている．特に診療情報データベースを想定したレセプト電算処理形式の疑似データ約 50,000 人分につい

☆4 http://sharemind.cyber.ee/
☆5 http://sepia-project.eu/

て，平均値や中央値等の基本統計演算や，単純集計，クロス集計，また等号判定や大小比較によるフィルタリングを組み合わせた演算（例：ある薬に対する 60 歳以上の男性の投与量平均値）の秘密計算を試行評価し，数秒から数分で処理できることを確認した[6]．50,000 件のデータは，1 カ月分の希少症例データを想定したものである．診療情報は，個人が特定されるリスクが高い希少症例データが特にセンシティブと考えられており，患者のプライバシー保護に十分配慮する必要がある．

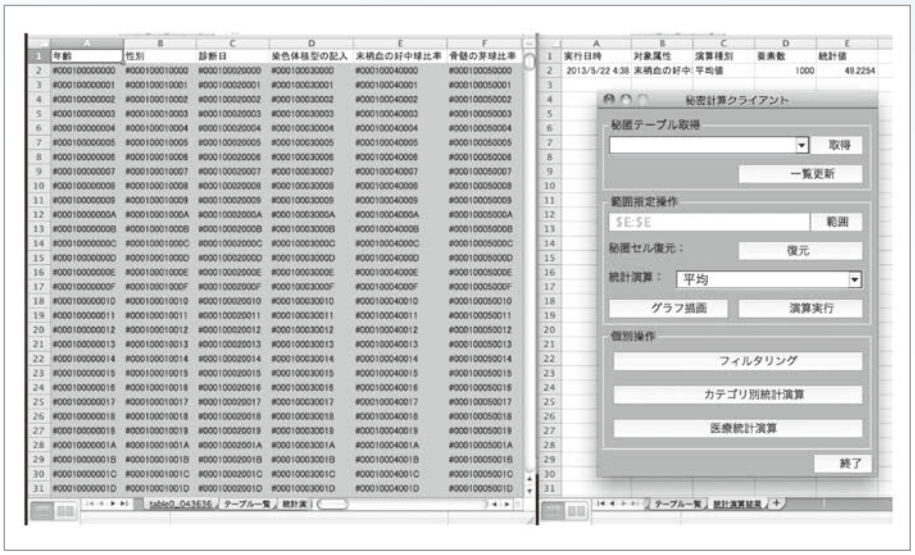

図 -5　Microsoft Excel から実行可能な秘密計算の画面

筆者らが開発したプロトタイプでは，Microsoft Excel や統計処理フリーソフトウェア R を用いて各種分析を行うことができる．**図 -5** は Microsoft Excel を用いた画面例である．左側 1 行目は属性名を表しているが，2 行目以降の各属性の値は秘匿されている（代わりにセルの番号を表示）．分析の操作は右側のユーザフォームから行い，ボタン操作と範囲選択により分析が可能である．右側上部には指定した分析の処理結果が表示されている．一方 R はオープンソースであり，利用者が新たに分析用演算の秘密計算のコードを作成することも可能である．

以上より，技術的には秘密計算を実行する準備が整ってきたといえる．今後パーソナルデータの安全な利用に向け秘密計算を実用化していくためには，法制度・ガイドラインとの関係の整理や国民の正しい理解が不可欠であり，学際的な取り組みが強く望まれる．また，入力プライバシー保護に資するその他の技術的対策や出力プライバシー保護も考慮した複合的なソリューションを確立することも重要な課題である．

参考文献

1) Gentry, C. : Fully Homomorphic Encryption Using Ideal Lattices, STOC2009, pp.169-178 (2009).
2) Yao, A. : How to Generate and Exchange Secrets (Extended Abstract), FOCS'86, pp.162-167 (1986).
3) Lindell, Y. and Pinkas, B. : Privacy Preserving Data Mining, Crypto2000, LNCS 1880, pp.20-24 (2000).
4) Tu, S., Kaashoek, M., Madden, S. and Zeldovich, N. : Processing Analytical Queries Over Encrypted Data, PVLDB2013, pp.289-300 (2013).
5) 千田，五十嵐，濱田，高橋：エラー検出可能な軽量 3 パーティ秘匿関数計算の提案と実装評価，情報処理学会論文誌 Vol.52, No.9, pp.2674-2685 (2011).
6) 諸橋，千田，冨士，間形，藤村，山本：秘匿計算の大規模医療情報データベースへの応用に関する研究，第 17 回日本医療情報学会春季学術大会 プログラム・抄録集，pp.70-71 (2013).

(2013 年 5 月 22 日受付)

千田浩司（正会員）｜chida.koji@lab.ntt.co.jp

2000 年早稲田大学大学院理工学研究科数理科学専攻修士課程修了．同年 NTT 入社．博士（工学）．プライバシー保護技術の研究開発に従事．本会コンピュータセキュリティ研究会（CSEC）幹事．