戦略的創造研究推進事業　CREST
研究領域「ビッグデータ統合利活用のための
次世代基盤技術の創出・体系化」
研究課題「自己情報コントロール機構を持つ
プライバシ保護データ収集・解析基盤の構築と
個別化医療・ゲノム疫学への展開」

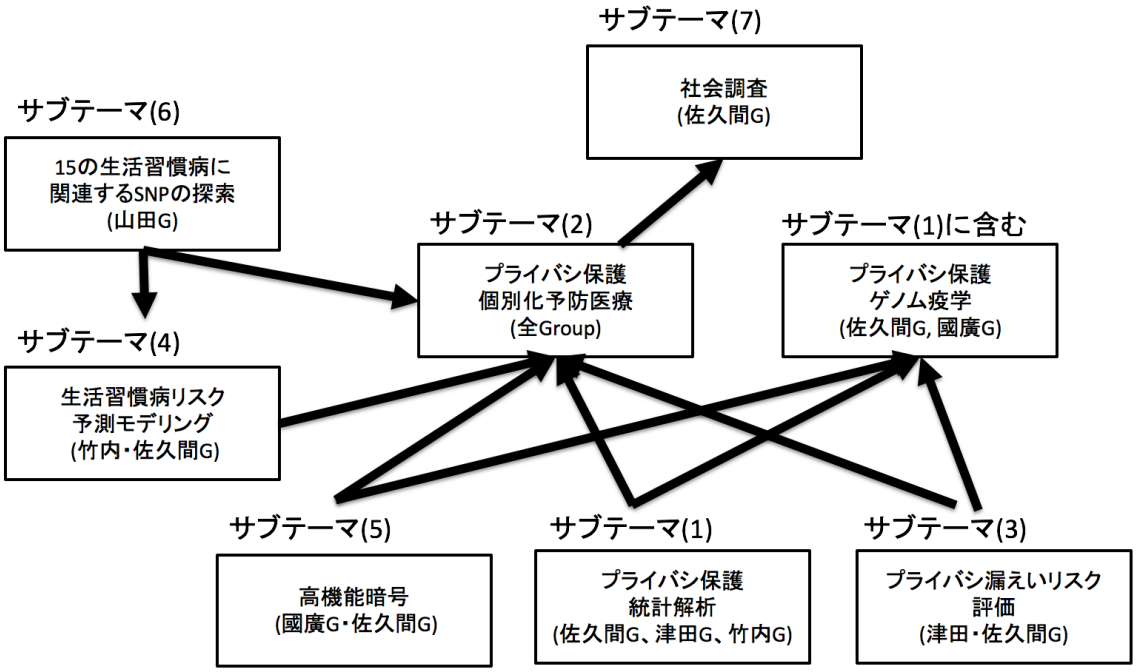
研究終了報告書

研究期間　2013年　10月～2019年　　3月

研究代表者：佐久間　淳
（筑波大学システム情報系　教授）

§1 研究実施の概要

(1)実施概要



プライバシ保護統計解析(サブテーマ 1): 情報を暗号化したまま加算や乗算が可能である完全準同型暗号を用いて、様々な統計解析や統計的検定を実現するための秘密計算手法を設計し、ライブラリを構築した。サブテーマ(2)プライバシ保護個別化予防医療のリスク予測部分にこの技術を応用したほか、ゲノム疫学におけるプライバシ保護に応用した。

プライバシ保護個別化予防医療(サブテーマ 2): サブテーマ(1)で開発した準同型暗号の利用技術を用いて，個人ゲノム情報と医療情報をプライバシ保護しつつ結合し，生活習慣病の罹患リスクをレポーティングするシステムを構築した．ゲノム検査機関および二つの医療機関を共同し，構築システムを用いて，参加者に遺伝子検査に基づく生活習慣病の予防に関するアドバイスを行った。利用するSNPはサブテーマ(6)の知見を、リスクモデリングにはサブテーマ(4)の知見を利用した。サブテーマ(3)の知見から、検査結果の公開時に伴うプライバシ漏えいリスクが評価できる。

プライバシ漏えいリスク評価(サブテーマ 3): 個別化予防医療における疾患リスク値やゲノム疫学における検定統計量などから、プライベートな情報が漏えいリスクを評価する技術を構築した。サブテーマ(2)の実応用時に必要な技術である。

生活習慣病予測モデリング(サブテーマ 4): サブテーマ(6)で発見した生活習慣病関連 SNP を利用し、生活習慣病のリスクを予測するモデルを構築するために必要となる様々な機械学習技術を研究した。構築されたモデルはサブテーマ(2)のプライバシ保護個別化予防医療において利用される。

高機能暗号(サブテーマ 5): 準同型暗号を含む様々な高機能暗号に関する基礎研究を行った。開発技術の一部はサブテーマ(1)に組み込まれ、プライバシ保護統計解析の効率化、高機能化に役立てた。

15 の生活習慣病に関する SNP の探索(サブテーマ 6): 15,896 例のエクソームアレイによる SNPs

解析を行い，約 39 億個の SNPs 情報および 13 種類の生活習慣病に関する臨床情報・ライフスタイル情報を包括する大規模なデータベースを構築した.構築したデータベースを用いて 15 種類の生活習慣病の発症に強く関連する遺伝子群および SNPs を特定した.特定された SNP はサブテーマ(4)のリスクモデリングの特徴ベクトルとして利用され、サブテーマ(2)の個別化予防医療でリスク評価に用いる。

社会調査(サブテーマ 7): 個人ゲノムを用いた医療応用に関するアンケート調査を中心とした社会調査を行う。個人ゲノムやサブテーマ(2)の検査結果の医療機関同士の共有や家族、友人間の共有などについて調査する。

(2)顕著な成果

＜優れた基礎研究としての成果＞

1. Wen-jie Lu, Yoshiji Yamada, and Jun Sakuma, Privacy-preserving Genome-wide Association Studies on Cloud Environment using Fully Homomorphic Encryption, BMC Medical Informatics and Decision Making, 2015, 15 (Suppl 5), S1, 2015.

概要:

ゲノムワイド関連解析(GWAS)による疾患に関連する SNP の同定は個別化医療の発展に貢献するが，複数機関から情報を収集する場合，プライバシの保護が解析の障害となる．この研究では，準同型暗号を用いて，GWAS に用いられるカイ二乗検定量や相関係数などの統計量を，遺伝情報を暗号化したままクラウドに委託計算する手法を開発した．標準的な GWAS の規模である 10000 サンプル，20 万 SNPs を用いたカイ二乗検定において，既存研究では 2000 日以上の計算と 5600TB の記憶容量を必要とするところ，提案法では 63 時間 (8 core 並列で 8 時間)の計算時間で計算を完了することができることを示した．

2. Wen-jie Lu, Shohei Kawasaki, Jun Sakuma, Using Fully Homomorphic Encryption for Statistical Analysis of Categorical, Ordinal and Numerical Data. The Network and Distributed System Security Symposium 2017 (NDSS 2017), 16 pages, online proceedings.

概要:

情報を暗号化したまま加算や乗算が可能である完全準同型暗号を用いて、数値属性データ、順序属性データ、離散属性データなどを暗号化したまま統計解析を実現するための秘密計算手法を構築した。この論文では、多くの統計計算が行列演算と大小比較演算で記述されることに着目し、完全準同型暗号を用いた効率の良い行列演算と大小比較演算のためのアルゴリズムを開発した。あわせて、統計解析に必要な高精度の数値演算を暗号文上で実現するための平文拡張演算を開発した。これらの新たな手法を組み合わせることで、数値属性、順序属性、離散属性を含む数万レコードの暗号化データを対象として、標準的な記述統計(最大・最小、平均、分散、共分散、カウント、ヒストグラム、分割表、中央値、最頻値、k パーセンタイル)や予測統計(線形回帰、主成分分析)、統計的検定(カイ二乗検定、正確ロジスティック回帰)などの評価を数秒から 10 分程度で実現することに成功した。例えば線形回帰では、従来の手法に比べて 200 倍以上の効率化となることを示すことができた。

3. Takashi Yamakawa, Shota Yamada, Goichiro Hanaoka, Noboru Kunihiro, “Self-bilinear Map on Unknown Order Groups from Indistinguishability Obfuscation and Its Applications,” in Proc. of CRYPTO2014 (2) LNCS 8617, pp. 90-107, 2014. (DOI: 10.1007/978-3-662-44381-1_6)

概要:

國廣らは，補助情報付きの暗号学的自己双線形写像を導入し，識別不可性難読化を用いた上で，方式の提案を行った．ついで，素因数分解仮定の元で，提案方式が安全であることを示した．提案した自己双線形写像を用いることにより，望ましい性質を持つ多重線形写像の構成，複数人鍵共有，任意の回路に対する属性ベース暗号の構成などが可能となる．さらに，提案

方式と同様のアイデアを用いることにより，準同型暗号の構成も行った．難読化は，ソフトウェア保護などの分野において古くから研究がなされてきたが，近年になり，暗号理論のアプローチからの理論整備が，国内外で急速に進んでいる．本研究は，この中でも，素因数分解という暗号理論では標準的に用いられている安全性の仮定から構成していることに特徴がある．暗号分野においては，難しさがよくわからない仮定よりも，研究が進んで困難さがよく解明されている仮定を用いることが望ましい．この点でも，本研究は，他にないインパクトを持っている．

＜科学技術イノベーションに大きく寄与する成果＞

1． プライバシ保護個別化予防医療システム実証実験：
概要： 準同型暗号を用いて，個人ゲノム情報と医療情報をプライバシ保護しつつ結合し，生活習慣病の罹患リスクをレポーティングするシステムを構築した．ゲノム検査機関および二つの医療機関を共同し，40名の実験参加者を対象に，構築システムを用いて，遺伝子検査に基づく生活習慣病の予防に関するアドバイスを行った。

２． 生活習慣病感受性遺伝子 SNP 特許
概要：本研究の成果として 5 件の特許を出願した（制御装置、解析装置、復号装置および送信装置（特願 2014-262353）、循環器疾患の遺伝的リスク検出法（特願 2017-042125）、腎関連疾患の遺伝的リスク検出法（特願 2017-049143）、代謝疾患の遺伝的リスク検出法（特願 2017-056544）、脳血管障害の遺伝的リスク検出法（特願 2017-056911）。初めの一つは，高機能暗号を用いた情報管理方法に関する特許である．残りの 4 つは，15 の生活習慣病に強く関連する遺伝子に関する特許である．これらの特許についてはすべて G&G サイエンス株式会社と実施許諾契約を締結した。

3． プライバシ保護統計解析ライブラリ
概要： 情報を暗号化したまま加算や乗算が可能である完全準同型暗号を用いて、数値属性データ、順序属性データ、離散属性データなどを暗号化したまま統計解析を実現するための秘密計算手法に関するライブラリを構築し，github にて公開した．

＜代表的な論文＞

1．Wen-jie Lu, Yoshiji Yamada, and Jun Sakuma, Privacy-preserving Genome-wide Association Studies on Cloud Environment using Fully Homomorphic Encryption, BMC Medical Informatics and Decision Making, 2015, 15 (Suppl 5), S1, 2015.

2．Wen-jie Lu, Shohei Kawasaki, Jun Sakuma, Using Fully Homomorphic Encryption for Statistical Analysis of Categorical, Ordinal and Numerical Data. The Network and Distributed System Security Symposium 2017 (NDSS 2017), 16 pages, online proceedings.

3．Takashi Yamakawa, Shota Yamada, Goichiro Hanaoka, Noboru Kunihiro, “Self-bilinear Map on Unknown Order Groups from Indistinguishability Obfuscation and Its Applications,” in Proc. of CRYPTO2014 (2) LNCS 8617, pp. 90-107, 2014. (DOI: 10.1007/978-3-662-44381-1_6)

§２ 研究実施体制

(1)研究チームの体制について

① 佐久間グループ

・ 研究代表者:佐久間 淳 (筑波大学大学院システム情報工学研究科 教授)

研究項目

・ 自己情報コントロール機能を持つプライバシ保護統計解析方式の提案 (國廣グループと共同)
・ プライバシ保護ゲノム疫学(津田グループ、竹内グループと共同)
・ プライバシ保護個別化医療システムの開発と実証(竹内グループと共同)
・ 個別化医療におけるプライバシ漏洩リスク評価 (新規設定課題・津田・竹内グループと共同)
・ プライバシ保護データ収集・解析の社会科学的分析と制度設計

② 津田グループ

・ 主たる共同研究者:津田 宏治(東京大学大学院新領域創成科学研究科 教授)

研究項目

・ 個別化医療における漏洩リスク評価 (新規設定・佐久間・竹内グループと共同)
・ プライバシ保護ゲノム疫学(佐久間グループ、竹内グループと共同)
・ プライバシ保護ゲノム疫学の実証(佐久間グループ、竹内グループと共同)

③ 竹内グループ

・ 主たる共同研究者:竹内 一郎(名古屋工業大学工学研究科 教授)

研究項目

・プライバシ保護個別化医療システムの開発と実証(佐久間グループと共同)
・プライバシ保護ゲノム疫学(津田グループ、竹内グループと共同)
・ 心筋梗塞の個別化予防システム(山田グループと共同)
・ 個別化医療におけるプライバシ漏洩リスク評価 (新規設定課題・津田・佐久間グループと共同)

④ 國廣グループ

・主たる共同研究者:國廣 昇(東京大学大学院情報理工学系研究科 准教授)

研究項目:

・ 自己情報コントロールを実現する高機能暗号研究
・ 自己情報コントロール機能を持つプライバシ保護統計解析基盤自己情報コントロール機能を持つプライバシ保護統計解析基盤(佐久間グループと共同)

⑤ 山田グループ

・ 主たる共同研究者:山田 芳司(三重大学先端科学研究支援センター 教授)

研究項目

・ 心筋梗塞感受性遺伝子 SNPs 同定
・ 縦断ゲノム疫学研究による SNPs の検証と機能解析

(2)国内外の研究者や産業界等との連携によるネットワーク形成の状況について

・ 佐久間 G・竹内 G・山田 G:共同で、次世代遺伝子検査システムを開発し、遺伝子検査会社およびソフトウェア開発会社と共同で次世代遺伝子検査システムの複数病院における実証実験を実施。特許出願済み。
・ 佐久間 G・津田 G:共同で、北海道大学情報理工学研究科有村博紀教授と秘密計算に基づく非決定性オートマトンの評価法を研究し、ウイルスゲノム(ノロウイルス)検出を用いた実証的実験を実施。論文発表済。特許出願済。

- 佐久間 G・津田 G：共同で、東京大学情報基盤研究センターの荒井ひろみ助教と遺伝子検査結果からの遺伝情報・臨床情報の推測可能性について共同研究、論文発表済
- 佐久間 G：UCSD iDASH の主催する個人ゲノム秘密計算コンペティションに参加。コンペティションにおける議論を元に Shuang Wu (assistant prof.)共同研究を行い、ゲノム文字列の秘密計算について、論文発表済。
- 佐久間 G: 民間企業と機械学習およびプライバシについて、共同研究
- 佐久間 G：産総研の照屋唯紀研究員とアンドロイドを用いた秘密計算実装に関する共同研究、論文発表済
- 佐久間 G、山田 G: 誤操作検出可能な準同型暗号とそのゲノム関連解析への応用について、 NICT と共同研究、論文発表済、プレスリリース
- 竹内 G：京都大学豊浦和明准教授，瀬古敦人准教授と材料科学への機械学習応用に関する共同研究を進めており，論文発表済
- 竹内 G：名古屋大学医学研究科　門松健治教授と医療科学への機械学習応用に関する共同研究を進めている.
- 竹内 G：フランス Telecom Paris Tech の Joseph Salmon 准教授らと機械学習に関する共同研究を進めている.
- 竹内 G：豊田紡織，神戸製鋼，パナソニックと機械学習に関する共同研究を実施している.
- 津田 G：癌研究会の八尾研究員と、癌科学への機械学習応用に関する研究を実施している。
- 津田 G: 東京大学塩見淳一郎教授と材料科学への機械学習応用に関する研究を実施している。論文発表済。
- 津田 G: 東北大学梅津光央教授と、タンパク質科学への機械学習応用に関する研究を実施している。
- 國廣 G：産業技術総合研究所の山田翔太研究員と高機能暗号に関する研究を進めている. 論文発表済.
- 國廣 G，佐久間 G：情報通信研究機構と準同型暗号に関する研究に関して，共同で研究を進めている. 論文発表済.
- 國廣 G：日立製作所と暗号データベースの安全性に関して，共同で研究を進めている. 論文発表済.
- 國廣 G：オックスフォード大学・Ali El Kaafarani 博士と高機能暗号に関して共同で研究を進めている. 論文発表済.
- 山田 G：延世大学医学部（韓国）Yangsoo Jang 教授、Jong Ho Lee 教授、Dong-Jik Shin 准教授「ゲノム研究のグローバリゼーションとアジア間ネットワークの形成」（2004 年～）
- 山田 G：サンパウロ大学医学部（ブラジル）Regina Mingroni-Netto 教授、Tiago Pereira 研究員「高血圧・肥満のゲノム疫学研究」（2006 年～）
- 山田 G：スタンフォード大学医学部（米国）Euan Ashley 教授、Ziad Ali 研究員「血管リモデリングに関連する分子遺伝学研究」（2009 年～）
- 山田 G：オックスフォード大学医学部（英国）Robert Clarke 教授、Derrick Bennett 教授、Sarah Parish 教授「冠動脈疾患・脳血管障害のゲノム疫学研究」（2009 年～）
- 山田 G：ケンブリッジ大学医学部（英国）John Gregson 教授、Daniel Freitag 教授「冠動脈疾患のゲノム疫学研究」（2013 年～）
- 山田 G：モスクワ州立大学医学部（ロシア）Sergey Suchkov 教授「Predictive, preventive, and personalized medicine の国際ネットワークの形成」（2014 年～）