「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」
2014年度採択研究代表者

2018 年度
実績報告書

宮地　充子

大阪大学大学院工学研究科
教授

ビッグデータ統合利活用促進のためのセキュリティ基盤技術の体系化

§1. 研究成果の概要

ビッグデータの解析結果は新製品開発など様々な活用が期待され，データ収集・解析・利用の促進・定着は重要である．ビッグデータ流通システムの促進・定着には，データ所有者，解析機関，利用機関の各エンティティが win-win の関係を築けることが必須である．本研究課題はデータ所有者に着目し，データのプライバシ，データ解析結果のプライバシを実現し，データ所有者，解析機関，利用機関を信頼の環で連結する．またサイバー攻撃など各種攻撃に対する強化も考慮し，安全かつプライバシに配慮したビッグデータの流通プラットフォームを実現する．グループ全体の成果統合図を図 1　成果統合図に示す．

主な内容はビッグデータ利活用のためのセキュリティ技術基盤として，安全なデータ集合の突合・統合方式のユーザビリティと機能の向上を実現した．秘匿分類プロトコル方式においては,秘匿線形計算プロトコルおよび秘匿大小計算プロトコルの設計および理論的安全性の証明を行った．耐サイバー・耐量子攻撃システムの研究では，データ秘匿に利用するストリーム暗号の新たな安全性解析手法を実現し[1]，Ring-LWE 問題に基づく準同形性をもつ耐量子暗号の安全性の実験的検証を行った．匿名化技術のリスク評価手法においては，履歴データに対する匿名化手法を提案し，既存の匿名化手法との比較を行い，有用性の評価を実施した[2]．また文書データに対しては，実データ分析に向けてアプリケーションの開発を行った．データ漏洩抑止技術においてはデータ流通プラットフォームへの組み込み方法の検討を完了した．予防安全チームでは安全なデータ集合の突合・統合方式を応用したプロトタイプシステムによるヒアリングを進めており，次年度に現場での本格的な検証を行う．医療チームでは，テストベッドプロトタイプ構築に向けてシステム機能群の開発を進めており，昨年度までに開発した多機関分散型診療情報ストレージを対象としたデータ収集基盤に対し，収集後データのプライバシリスク評価機能，データ収集履歴のトレーサビリティ機能を追加実装した．特に，安全なデータ集合の突合・統合方式のプロトタイプの設計概要と 100 万件・3ノードの実証環境の性能結果について[3]で報告した．2018 年度医療情報学会でビッ

グデータのセキュアな利活用事例の 1 セッションを企画し，研究成果の発表を行った．

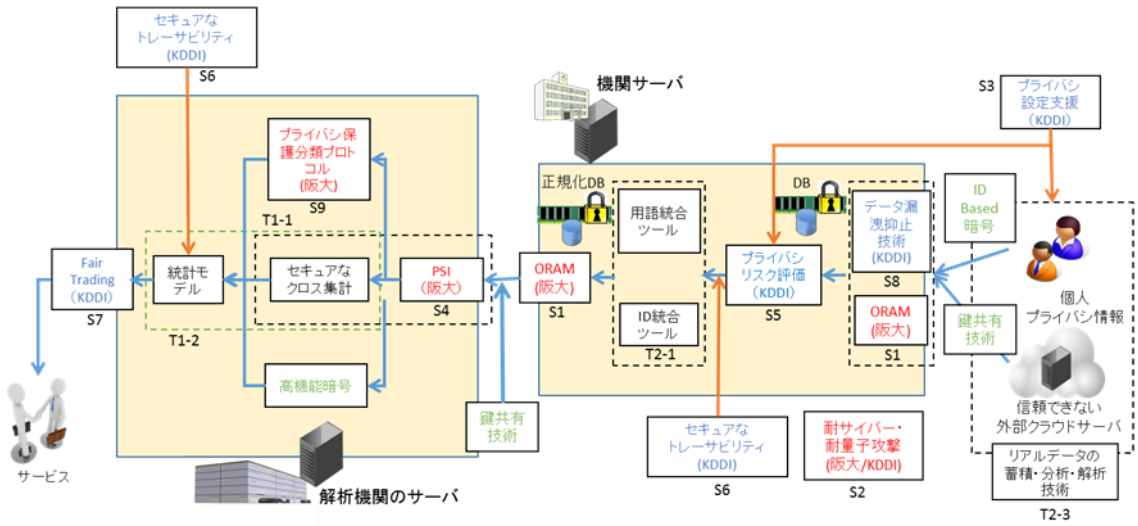

図 1 成果統合図

以下各グループの個別内容について述べる．

「セキュリティコア技術グループ」

プライバシを保護したデータ集合の突合・統合(PSI)アルゴリズム(S4)においては，医療分野と連携し，社会実装を進めた．また，PSI の Web サービスのプロトタイプ版において，利便性向上のために Python をインターフェイスとする PSI コア部を gmpy2(Python の GMP パッケージ)を利用し，Pure Python で再構築をおこなった．これにより，他のシステムとの連携が容易になる．

プライバシ保護分類プロトコル(S9)においては，耐量子公開鍵暗号として有望視されている HQC に基づいた秘匿線形計算プロトコルおよび秘匿大小計算プロトコルを秘匿分類プロトコルの基盤構成要素として設計し，またそれらの理論的安全性の証明を行った．

耐サイバー・耐量子攻撃システム(S2)においては，分解体上の Ring-LWE 問題は効率的な準同型耐量子準同型暗号に対して，格子攻撃及び環構造に基づく攻撃とその改良版について攻撃実験を行い，従来の耐量子準同型暗号と同等以上の安全性を実現できると考えられる結果を得た．

「セキュアデータ流通管理グループ」

(S5)匿名化技術のリスク評価手法では，履歴データを想定した匿名化手法を提案した．具体的には履歴データを行列として扱い，行列分解を匿名化手法の一つとして提案した．さらに行列分解と既存の匿名化手法(k-匿名化およびノイズ付加)とを組み合わせることで，行列分解を用いない場合と比較して特徴量を維持した匿名化データが作成可能であることを示した[2]．また文書データに対応したリスク評価手法についても検討し，既存研究の調査および実データでの実験に向けて試作ツールを開発した．また(S6)セキュアなトレーサビリティ，(S7)セキュアかつフェアなデータ対価決定プロトコルを含む，これまでセキュアデータ流通管理グループで検討ししてきた技術の実用化に向けて，データ流通プラットフォームへの実装の要件定義を明確にした．

「予防安全テストベッド実証グループ」

多機関分散データ統合モデリング技術(T1-1)においては，平成 29 年度までに開発してきた分散した傷害データを PSI を用いて統合して，類似状況の分析やトレンド分析を行うシステムについて，さらに機能を拡張するために，事故発生時の状況をより適切に分析可能となる手法として，事故状況情報のグラフ構造化による分析について検討を行った．また，現場や事故予防関係者などから，事故の全体像から着目すべき事故や条件を把握したい要望があり，それに対応するために，特徴的な条件を見つけ出す手法を開発し，統合した事故データに適用し，機能検証を行った．

実践的な統合データ利活用の実証(T1-2)では，現状のプロトタイプシステムをベースに機能や活用方法について現場に紹介し，フィードバックやさらなる要望を得た．また，学校と同様に事故データが各施設に分散しており，予防に活かすことが難しいという課題を抱える介護施設での事故についても，過去に収集した事故データを対象に，プロトタイプシステムを適用し，評価した．また，平成 30 年度も継続して，データを活用した創造的傷害予防研究会を開催し，データの活用について検討を行った．

「医療テストベッド実証グループ」

(T2-2) において，(S5)(S6)の成果を中心に，テストベッドへの適用と評価，実運用への適用を考慮した機能開発を進めた．(S5)については，(S4)を適用した多機関横断検索が可能なテストベッドに対し，プライバシリスク評価ライブラリを WebAPI 化したサービスを実装し，個人特定度を算出，提示するユーザインターフェイスを開発し，(T2-3)で構築中の検証用データベースと連携させた．
(S6)については，収集した診療データの二次利用の履歴を患者自らが検索可能なサービス基盤の試作を行った．Block Chain Network 技術（以下，BCN）を採用した場合，BCN ノードとしてのトランザクション性能が秒間約 100 件程度であることを確認できたため，性能向上のため，二次利用ログを患者単位に集約する方式を考案し，評価を進めている．(T2-3)においては，東京大学医学部附属病院の電子カルテデータから，SS-MIX・レセプトデータ・DPC データの約 7 年分の登録が完了した．本プロトタイプの設計概要と，100 万件・3ノードの実証環境での(S4)の性能測定結果について，[3]で報告した．

【代表的な原著論文】

発表論文

[1] Ryoma Ito and Atsuko Miyaji, ``New Iterated RC4 Key Correlations″, The 23rd Australasian Conference on Information Security and Privacy(ACISP 2018), Lecture Notes in Computer Science, 10946(2018), Springer-Verlag, 154-171.

[2]Tomoaki Mimoto, Shinsaku Kiyomoto, Seira Hidano, Anirban Basu, Atsuko Miyaji, ``The Possibility of Matrix Decomposition as Anonymization and Evaluation for Time-sequence Data″, The 16th Annual Conference on Privacy, Security and Trust(PST2018), IEEE, 1-7, 2018.

[3] K. Tanaka, R. Yamamoto, K. Nakasho, and A. Miyaji, ″Development of a Secure Cross-Institutional Data Collection System Based on Distributed Standardized EMR Storage″, Stud Health Technol Inform, vol. 255, pp. 35-39, 2018.

§2. 研究実施体制

(1)セキュリティコア技術グループ

① 研究代表者:宮地　充子 （大阪大学大学院工学研究科　教授）

② 研究項目

・(S2) 耐サイバー・耐量子攻撃システム

・(S4) プライバシを保護したデータ演算

・(S9) プライバシ保護分類プロトコル

・(S10)ビッグデータ利活用のためのセキュリティ技術の体系化

(2)セキュアデータ流通管理グループ

① 主たる共同研究者:清本　晋作(KDDI 研究所　グループリーダ)

② 研究項目

・(S3) プライバシポリシ設定支援

・(S5) 匿名化技術のリスク評価手法

・(S6) セキュアなトレーサビリティ

・(S7) セキュアかつフェアなデータ対価決定プロトコル

・(S8) データ漏洩抑止技術

・(S10)ビッグデータ利活用のためのセキュリティ技術の体系化

(3)予防安全テストベッド実証グループ

① 主たる共同研究者:西田　佳史(産業技術総合研究所　首席研究員)

② 研究項目

・(T1-1) 多機関分散データ統合モデリング技術

・(T1-2) 実践的な統合データ利活用の実証

(4)医療テストベッド実証グループ

① 主たる共同研究者: 田中　勝弥 （東京大学大学院医学系研究科　講師)

② 研究項目

・(T2-1) ID 統合技術

・(T2-2) 実践的な統合データ利活用の実証

・(T2-3) リアルデータ(DPC, SS-MIX2, レセプト, センサデータ)の蓄積・分析・解析技術