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「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」

研究終了報告書

研究期間　2014年　10月～2020年　3月

研究代表者：宮地　充子
（大阪大学大学院工学研究科、教授）

§1 研究実施の概要

(1)実施概要

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，安全なデータ収集・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，データ所有者及びデータ解析結果のプライバシの実現と，サイバー攻撃など各種攻撃に対する強化も考慮した，データ所有者，解析機関，利用機関を信頼の環で連結するビッグデータ利活用促進のためのセキュリティ技術の体系化と体系化技術の予防安全，医療分野でのテストベッドによる実証を目的とする．グループ全体の成果統合図を図 1-1 に示す．セキュリティ技術の体系化は「セキュリティコア技術グループ」「セキュアデータ流通管理グループ」で行い，テストベッドを「予防安全テストベッド実証グループ」「医療テストベッド実証グループ」で行った．

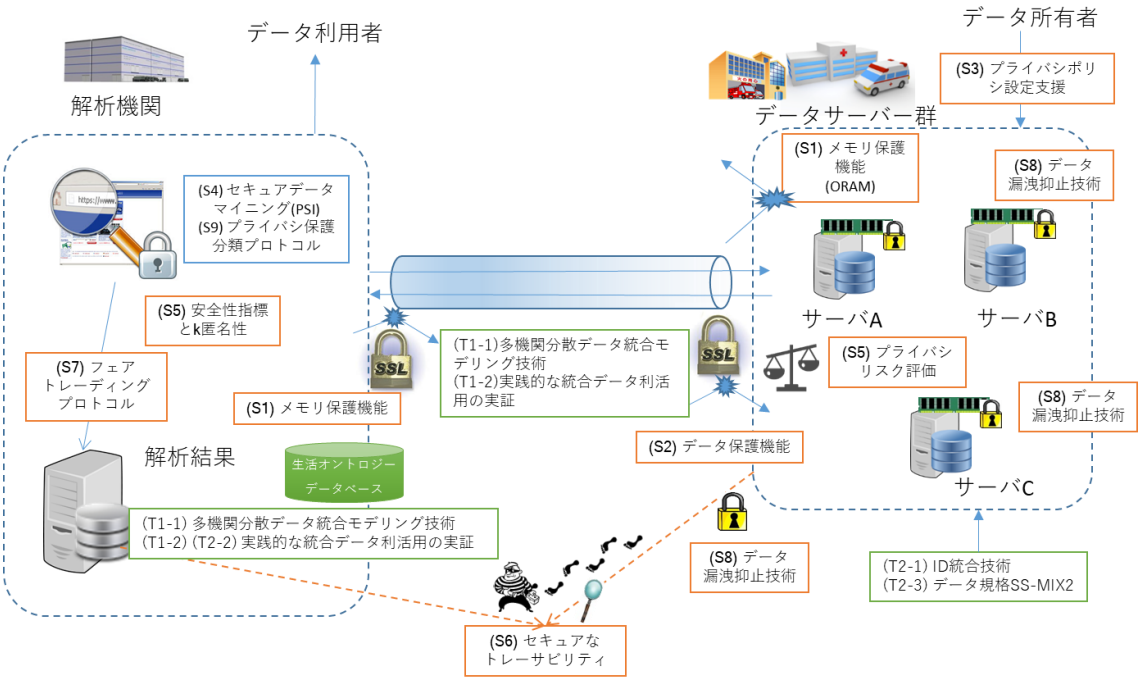

図 1-1 成果統合図

「セキュリティコア技術グループ」においては，ビッグデータ利活用のためのセキュリティ基盤技術として，(S4) プライバシを保護したデータ演算において，安全なデータ集合の突合・統合方式を提案し，プロトタイプシステムを構築，予防安全及び医療分野でのテストベッドを実現した．さらにテストベッド実証結果を反映させて，各種ビッグデータ利活用の汎用セキュリティ技術として，ユーザビリティを向上させたシステムを構築した．また，ビッグデータ利活用促進のためのセキュリティ技術の体系化という観点で，(S2)耐サイバー攻撃システムでは，耐量子暗号の一つである LWE ベースの暗号に対する安全性解析を実現した．(S9)プライバシ保護分類プロトコルにおいては，耐量子安全性を実現する収集したデータに対する秘匿分類プロトコルを構築した．

「セキュアデータ流通管理グループ」では(S2)耐サイバー攻撃システムでは IoT デバイス等を含む環境に適用可能なメモリ保護技術を提案し，既に実用レベルにある．(S3)プライバシポリシ設定支援では，機械学習を用いたポリシ設定推定技術のシステムを開発し，被験者を用意してその有効性の検証を行った．(S5)プライバシリスク評価については，医療テストベッド実証グループにて開発したシステムでの実データを用いた評価実験を実施，既に実用レベルにある．(S8) データ漏洩抑止技術と(S6) セキュアなトレーサビリティについては，ブロックチェーンを活用したシステムを

設計し，プロトタイプを実装して処理可能なトランザクション数およびログ検索時間を評価した．

「予防安全社会実装グループ」「予防安全テストベッド実証グループ」においては，多機関に分散した傷害データを，MPSI を用いて統合して分析可能なシステムを開発してきた．分析機能として，類似状況におけるクリフ分析機能，トレンド分析機能，特徴的な条件を見つけ出す機能，全体像の可視化機能を提案し，開発を行ってきた．開発したシステムを実際に学校で発生した事故データに適用することで有用性の実証を行った．

開発したシステムで分析した結果は，小児科医，デザイナー，消防庁などから構成される研究会で情報共有を行い，予防のための新たな製品のデザインについて，トマトカッターと耳かきについて試作を行うとともに，ユーザビリティテストも行った．また，分析によって明らかとなったリスクについて，深掘り調査を行って，予防策の考案まで実施する取り組みを行っており，これまでに野球での眼部の傷害，サッカーゴールの転倒事故，跳び箱での傷害について実施し，シンポジウムを開催し，情報発信を行った．

「医療テストベッド実証グループ」においては，セキュアコア技術グループ，セキュアデータ流通管理グループが開発した各要素技術を医療向けテストベッドに適用しつつ，実際のデータを適用した実装・評価を実施した．共通データモデルとしては，厚生労働省標準でもある SS-MIX2 標準化ストレージを採用し，多施設に分散した標準化ストレージを横断検索可能とする情報基盤を構築した．要素技術としては，同標準化ストレージに対して，1)プライバシを保護したデータ演算基盤，2)データ二次利用における患者同意情報による抽出制御，3)抽出されたデータセットのプライバシリスク評価機能，4)患者自身による二次利用履歴の検証機能，を実装した基盤の開発を終えた．開発したテストベッドは，パブリッククラウドへ適用した．

(2)顕著な成果

＜優れた基礎研究としての成果＞

1．Atsuko Miyaji, Kazuhisa Nakasho, Shohei Nishida, “Privacy-Preserving Integration of Medical Data: A Practical Multiparty Private Set Intersection”, Journal of Medical Systems, vol. 41, no. 3, Plenum Press, DOI: 10.1007/s10916-016-0657-4., 1-10, 2017.
概要：
有用な医療データは異なる組織において管理されていることが多く，疫学研究等における詳細分析では，患者のプライバシや組織データベースの機密性を侵害することなく，これらのデータセットを統合する必要性がしばしば生じる．MPSI（Multiparty Private Set Intersection）プロトコルは，各機関は，それぞれが所有するデータ以外の情報を入手することなく，共通データを抽出するプロトコルであり，複数機関が保有するデータの秘匿計算における最重要課題の一つである．本論文では，アウトソーシング計算を利用し，通信量と計算量が MPSI への参加機関数に依存しないプロトコル方式を提案するとともに，本方式を実装することにより実用性を検証した．

2．Koji Kitamura, Kenta Imai, Yoshifumi Nishida, Hiroshi Takemura, Tatsuhiro Yamanaka, ``Potential Risk Assessment System by Integrating Injury Data at Multiple Schools″, The 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015), pp. 1991-1998, July 2015.
概要：
個人情報保護などの観点から学校における事故データは共有されることがないため，事故が起きる前にハイリスクな事故が起きる可能性を把握することが難しく，事故予防が進みにくい．これに対し，MPSI を活用して多機関の事故データを秘匿したまま共有し，日々の軽傷事故と類似した状況で起きている重傷事故を把握可能とするデータ共有・分析システムを開発した．開発したシステムを 69 校の小学校で起きた 5817 件の事故データに適用し，事故が起きた小学校名の情報は秘匿したまま，重症度が高い事故を検出したり，その事故情報を共有可能であることを示した．

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.
概要:
電子カルテデータの収集および二次利用を伴う臨床研究を安全に実施するために必要となるインフラとして，電子カルテデータを外部ストレージへ保存する際の国内標準規格である SS-MIX2 標準化ストレージに対し，本研究プロジェクトで開発したプライバシ保護演算機能を実装した多機関分散型のセキュアなデータ収集基盤を開発した．本稿は，実装のデザインを示すとともに，100 万件規模の実データによる検証において，実用に支障のない性能が得られたことを報告した．

＜科学技術イノベーションに大きく寄与する成果＞

1．匿名化・リスク評価ツールの商用展開
概要:
パーソナルデータを含むデータセットに対し，データセットの加工およびリスク評価機能を備えた匿名化・リスク評価ツールを開発し，東大病院との実証実験を実施した．さらに商用化に向けた改修を行った．改修したツールはすでに CREST プロジェクト内外でトライアルを実施しており，さらには認定個人情報保護団体である日本データ通信協会主催のハンズオンの教材として採用された．現在本ツールについては本格的な展開に向けて活動を行っている．

2．MPSI を用いた多機関データの統合による学校環境下での事故リスク分析
概要:
MPSIを活用して多機関に分散した事故データを秘匿したまま統合して，共有や分析可能なシステムを構築した．また，統合したデータを対象に，給付金額や医療費をKPIとして，類似状況で発生した事故のうち特に重症度が高い事故を見つけ出す機能も開発した．このシステムを用いて学校での事故データについて分析した結果は，朝日新聞で学校の事故の特集として採用され，13回の連載記事となり，社会に学校の事故の問題を訴求した．

3．MPSI に基づくデータ統合システム
概要:
情報漏洩を懸念することなく，複数機関が所有するデータ統合を実現するシステムを構築した．現在，各種機関への展開中である．本システムにより，下記の機能が実現できる．
・各機関の希少データを統合することで，プライバシを保護した精度の高い解析．
・各機関の持つ異なる属性を統合することで，プライバシを保護した多角的な解析．
さらに，該当システムには下記の特徴がある．
高機密性:各機関が許可したデータのみが許可された機関でのみ閲覧可能．
高汎用性:データ数・機関数に非依存．対象機関・突合項目・出力項目を自由に設定可能．
導入容易性:データ預託機関は不要であり，各機関はクライアントサーバモデルで実行可能．
高速性:アウトソーシング計算機によりデータ秘匿性を保持しながら，機関数に非依存な処理効率を実現．

＜代表的な論文＞
Atsuko Miyaji, Kazuhisa Nakasho, Shohei Nishida, “Privacy-Preserving Integration of Medical Data: A Practical Multiparty Private Set Intersection”, Journal of Medical Systems, vol. 41, no. 3, Plenum Press, 1-10, 2017.

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

Ryoma Ito and Atsuko Miyaji, “New Linear Correlations related to State Information of RC4 PRGA using IV in WPA”, The 22nd International Workshop on Fast Software Encryption (FSE 2015), LNCS, Springer-Verlag, vol 9054, 557-576, 2015.

§2 研究実施体制

(1)研究チームの体制について

① 「セキュリティコア技術グループ」
研究代表者:宮地　充子(大阪大学大学院工学研究科　教授)
研究項目
(S1) 検証機能付きデータ管理/削除
データ管理サーバに対する安全性強化技術の構築
(S2) 耐サイバー・耐量子攻撃システム
耐量子暗号による情報セキュリティ技術の再構築
(S4) プライバシを保護したデータ演算
多機関分散管理のデータベースをプライバシを保護しながら突合を実現し，多機関に保存されるデータの関連付けを実現するシステムの構築
(S6) セキュアなトレーサビリティ
プライバシポリシを満たすサービス提供者にデータ配布・追跡を実現する手法の構築
(S9) プライバシ保護分類プロトコル
データ分類における参加者データのプライバシ保護を実現する手法の構築

② 「セキュアデータ流通管理グループ」
主たる共同研究者:清本　晋作((株)KDDI 研究所　グループリーダ)
研究項目
(S2) 耐サイバー・耐量子攻撃システム
実装上の脆弱性をつく SSL 等へのセキュリティ機能への攻撃を防ぐメモリ保護手法の構築
(S3) プライバシポリシ設定支援
複雑な環境における個人のプライバシポリシの初期設定を支援する手法の構築
(S5) 匿名化技術のリスク評価手法
データセットに含まれるプライバシ漏洩リスクの定量化およびその対策手法の構築
(S6) セキュアなトレーサビリティ
ブロックチェーンを用いてデータの流通をトレースする手法の構築
(S7) セキュアかつフェアなデータ対価決定プロトコル
公平なデータ取引を実現するセキュアかつフェアなデータ対価決定手法の構築
(S8) データ漏洩抑止技術
匿名化されたデータが漏えいした場合に流出元を特定するための手法の構築

③ 「予防安全社会実装グループ」「予防安全テストベッド実証グループ」
主たる共同研究者:西田　佳史(東京工業大学工学院機械系　教授)，北村　光司(産業技術総合研究所人工知能研究センター　主任研究員)
研究項目
(T1-1) 多機関分散データ統合モデリング技術
多機関に分散したデータをセキュアに統合した活用手法の構築
(T1-2) 実践的な統合データ利活用の実証
多機関に分散したデータをセキュアに統合した活用手法の実データへの適用による有用性の検証及び課題解決のための調査・検討

④ 「医療テストベッド実証グループ」
主たる共同研究者:　田中　勝弥(国立がん研究センター情報統括センター　情報システム企画課長)
研究項目

(T2-1) ID 統合技術
医療用共通番号を採用した場合のシステム要件調査
(T2-2) 実践的な統合データ利活用の実証
医療テストベッドを構築するための要件検討および実装・評価
(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
各種リアルワールドデータを二次利用するためのデータ形式, 要素技術の開発・検証

(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
「セキュリティコア技術グループ」
「癌サバイバーと脳卒中などの循環器疾患発症に関する研究」への適用に向けた検討を行っている. さらに, 山本景一准教授(和歌山県立医科大学情報基盤センター副センター長)との共同による癌検診データを用いた健康増進プロジェクトへの適用などのネットワーク形成を構築しつつある. 今後, これらのシステム適用を目指し, さらなる改良を行う予定である. また, Yan Wan 准教授(University of Texas at Arlington), Shengli Fu 教授 (University of North Texas)との連携により, ドローンを用いて収集するビッグデータの利活用に向けて, 本プロジェクトで行った各種研究を適用し, ドローンの特徴に応じたセキュリティ技術の開発に向けた共同研究を開始している.

「セキュアデータ流通管理グループ」
プライバシ設定支援に関しては, ドイツのゲーテフランクフルト大学と連携を構築し, 日独の比較を行うなど共同で研究を行った. また, この連携を生かして欧州の研究ファンドであるHorizon2020へのプロジェクト提案活動なども行った. また, 耐サイバー耐量子攻撃システムの研究に関しては, オーストラリアのウーロンゴン大学と連携を構築し共同で研究を進めている. さらに, 匿名化技術のリスク評価手法については, 一般財団法人 日本データ通信協会と連携してハンズオンセミナーを開催し, 研究成果をアピールするとともに, 商用展開として複数の医療機関と議論を開始している.

「予防安全社会実装グループ」「予防安全テストベッド実証グループ」
これまでに, 創造的傷害予防研究会では, 小児科医, デザイナー, 消防庁などとの検討を進めて, 子どもの傷害予防のための連携を深めている. 具体的に, 予防のための試作などを行い, 製造メーカーへ製品化の打診なども進めている. また, これまでの深掘り調査では, 弁護士から構成される日本スポーツ法支援・研究センター, 日本中学校体育連盟, 国士舘大学, 日本体育大学, 理学療法士など連携した調査を実施しており, 新たなネットワーク構築へと展開している.

「医療テストベッド実証グループ」
開発した医療テストベッドシステムについて, 研究分担者が所属する国立がん研究センター内の他臨床 AI プロジェクト(PRISM 等)との連携協議を進めている. また, 開発成果としてのテストベッドのパブリッククラウド上でのサービス化について, 民間企業1社との共同研究, 製品化の打診を進めている.