「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」
平成 26 年度採択研究代表者

H29 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

§2. 研究実施の概要

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, 匿名化技術のリスク評価手法を提案・プロトタイプを作成し, 分類方法を秘匿する方式及びデータ漏洩抑止技術においてはビッグデータ利活用時における既存方式の課題を明確化した. また, ビッグデータ利活用のためのセキュリティプラットフォーム技術の体系化を目指し, IoT 機器向けハッシュ関数の改良も行った. さらに安全なデータ集合の突合・統合方式は予防安全チーム及び医療分野において実証実験の準備段階にあり, 匿名化技術のリスク評価手法は医療分野において実証実験の準備段階にある. また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, 発表を行った[3].

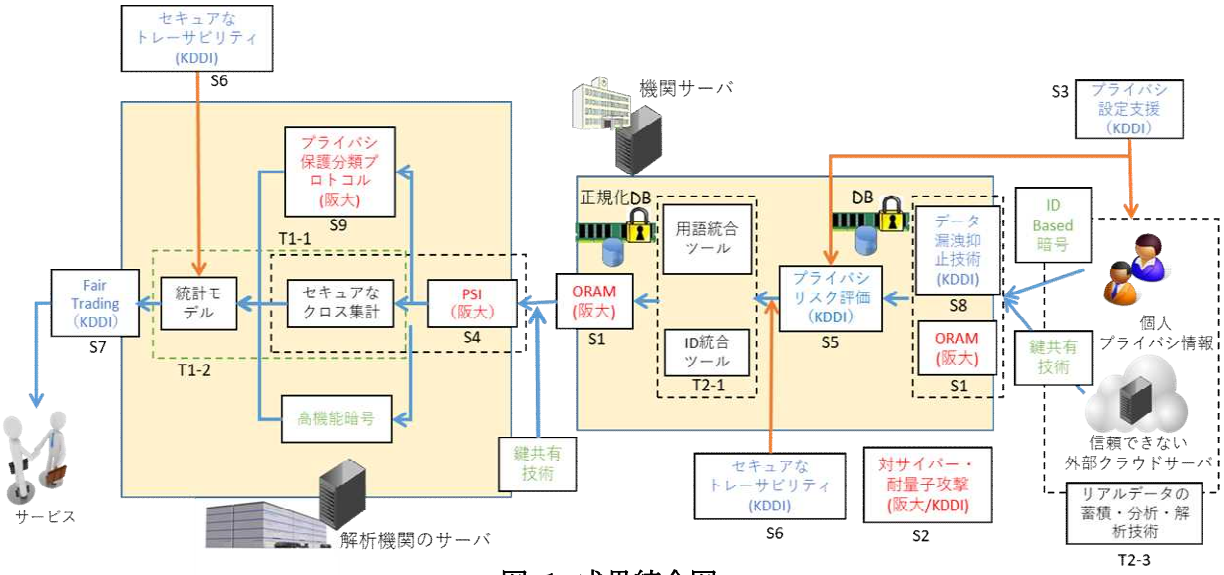

図 1 成果統合図

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演算(S4)においてはデータ集合積演算の研究をデータ集合和演算に拡張した. 特に, 集合演算の効率的な処理に必要となるブルームフィルターの更新機能を付加し, 偽陽性が起こるケースを削減した. また提案済のデータ集合積演算(PSI)機能の Web サービスのプロトタイプ版において, ユーザビリティと耐久性の観点で, ソフトの改良を行った. プライバシ保護分類プロトコル(S9)においては, 既存のプライバシ保護分類プロトコル, 特に二者間秘匿計算プロトコルの中で分類に応用できる技術の調査・検討を行った. 基盤となる構成技術として数論系, 格子系, 符号系の技術の中から, 耐量子攻撃性と高速実装の可能性からノイズ付き符号化を選んだ. 格子ベース暗号で用いられる計算効率化技術との組み合わせを考察し, さらに線形関数を利用したプライバシ保護分類プロトコルの基本設計を行った. 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(S10)においては軽量なハッシュ関数の改良を提案し[2], 最優秀論文賞を受賞, また, 軽量な楕円曲線暗号に関して招待講演を行った[1].

「セキュアデータ流通管理グループ」

(S3) プライバシポリシ設定支援については，推測モデルをユーザ属性及び回答傾向にて分割したデータから生成した場合についての精度向上について評価した結果をまとめ，DPM2017 で発表した．また，プライバシポリシに自動的にプライバシリスクに関するラベルを付与することで，プライバシポリシの要約を行う技術の検証に向け，プライバシポリシの収集及びポリシへのラベル付け作業を実施した．さらに，(S5)の成果と連携し，プライバシ機能を有するクラウドベースのデータ流通基盤の実装について検討した．(S5)匿名化技術のリスク評価手法では，現実的な攻撃者モデルを想定して，サンプリング，ノイズ付加，k 匿名化を組み合わせた際の安全性を統一的に評価するリスク評価法を提案し[4]，機能として持つ匿名化・リスク評価ツールをライブラリ化し，「医療テストベッド実証グループ」に提供した．更に開発した匿名化・リスク評価ツールの汎用性を高めるため，履歴データへの対応が可能なツールのプロトタイプを開発した．また(S6)セキュアなトレーサビリティ，(S8)データ漏洩抑止技術については，Fair Trading プロトコルとデータ漏洩抑止技術を統合したシステムを設計し，統合可能であることを確認した．また，左記の研究成果について，欧州5か国の研究機関 8 機関と合同で提案書を作成し，H2020「ICT-13-2018-2019: Supporting the emergence of data markets and the data economy」に応募した．(S7)セキュアかつフェアなデータ対価決定プロトコルについては，Fair Trading プロトコルを設計し解析を行った結果を IEEE ICIM2018 に投稿し，採録となった．

「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 28 年度までに開発してきた分散した傷害データを，PSI を用いて統合して分析するシステムについて，機能拡張を行うとともに，スケーラビリティ検証を行った．機能拡張に関しては，これまで開発してきた重傷に至る要因分析機能に加え，社会的ニーズが高い，傷害の経年変化を分析するトレンド分析機能を実装した．また，新たにセキュリティコア技術グループが開発した「WEB サービス化された PSI 機能」と統合し，実際の学校現場での実証を行うための実用システムを開発した[5]．スケーラビリティ検証に関しては，日本スポーツ振興センターと連携し，1 年に学校環境(保育園・幼稚園・小中高学校)で発生する傷害データサイズ(悉皆データサイズ)に相当する 100 万件規模の傷害ビッグデータや，野球や柔道などの全国の過去 8 年のスポーツ外傷データ(悉皆データサイズ)に適用することで開発機能の検証を行った．

実践的な統合データ利活用の実証(T1-2)では，開発した統計モデルや収集した事故データの予防安全技術開発への活用法を検討する研究会を開催し，現場で発生する事故事例，研究所で実施した統計解析の結果の共有，それらの統計データを活用したプロダクトデザイン(データ駆動型デザイン)を行った．平成 29 年度は，平成 28 年度に試作・ユーザビリティ評価をしてきた耳かきによる鼓膜損傷や耳道損傷を防ぐための耳かきについて，傷害データとともにデザインアイディアについて企業にプレゼンし，企業と具体的な製品化について検討を始めることができた．傷害を予防する観点と使いやすさや価格などに関して議論を進め，製品化に向けた準備プロセスを進めることができた．また，傷害データの分析から明らかとなったリスクについて，具体的な予防策を考案するための取り組みとして，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ミクロな調査を行い，リスク要因を詳細に明らかにし，教育コンテンツの作成や予防グッズ

のユーザビリティ調査など具体的な活動につなげることができた.

「医療テストベッド実証グループ」

ID 統合技術(T2-1) においては, IHE-ITI における PIX/PDQ/XCA の実装状況について昨年度に引き続き調査を行った. 多コミュニティ間の ID 検索に関して総当たりの検索になる以外, 大きな課題は確認できていない. 医療等分野における識別子(ID)の政府検討方針が示されたが, 情報の二次利用を行うための識別子として使用できる可能性がなくなっており, 具体的な実装検討は中止とした. 平成 29 年度をもって, 本テーマは完了とする.

実践的な統合データ利活用の実証(T2-2)においては,「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」により開発された, (S1)(S4)(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. (S1)については, 医療現場で頻用される診療情報提供書や放射線検査画像のオンライン運用に必要な機能開発を終え, 大容量データを交換するためのセキュアなクラウドサービスの開発を行い, 既設の商用ソフトウェアとの連携機能を実装した. 東京大学医学部附属病院周辺の医療機関間で放射線検査画像データのオンライン運用への試験的導入を行うべく, 関係機関との協議を開始した. (S4)については, (T2-3)で開発中の SS-MIX2 標準化ストレージにおける大規模データを模擬的に複数作成し, (S4)の成果を実装した評価を行った. 結果, 互いに 10%の重複を有する3つの 100 万件規模の分散データを同ストレージ上で 1 分程度の処理時間で突合, 抽出可能なことが確認できた. (S6)については, 収集した診療データの二次利用に際した患者同意情報のトレーサビリティ基盤の開発に際して, 平成 29 年度はデータ連携方式の検討を行い, 患者同意書様式の電子化実装の検討を行った. 記述規格候補として, HL7 CDAR R2: Privacy Consent Directives, Release 1 の調査, 実装検討を行い, 実運用に適用可能であることを確認した. また, 同規格による同意文書サンプルを試作した.

リアルデータの蓄積・分析・解析技術(T2-3) においては, DPC, SS-MIX2, レセプトについて, 横断解析が可能とするためのデータ蓄積基盤の構築を終えた. とくに, 多機関に分散蓄積された診療データの横断的探索を行う二次利用において, (S4)(S6)の成果を展開するために, データの収集と同時に解析, 検索可能なテストベッド機能の開発が完了し, 性能評価を終えた. 平成 29 年 12 月に東京大学医学部附属病院の実データを利用する検証について倫理申請が承認されたため, 同病院の 8 年分規模の蓄積データを利用して検証を進めている. テストベッドとしては, SS-MIX2 標準化ストレージに格納される内視鏡検査情報を除く全てのイベントデータを(S4)ライブラリと連携可能な基盤が構築できた. なお, センサデータは現在, 経産省の事業で, かなり精度にばらつきがあり, 直接収集しても有意な分析に支障があることが明らかになっており, ばらつき補正が困難であるため, 本事業としてはスコープ外との結論とした.

代表的な原著論文

[1] Miyaji, (Keynote Speak) “Elliptic Curve Cryptosystems for IoT devices”, The 19th International Conference on Information and Communications Security (ICICS 2017).
[2] Nomaguchi, Miyaji and Su, “Evaluation and Improvement of Pseudo-Random Number Generator for EPC Gen2”, The 16th IEEE TrustCom'17, 2017.(最優秀論文賞)
[3] 公募企画シンポジウム: 安心・安全なビッグデータの流通プラットホームとセキュリティ基盤技

術, 37 回医療情報学連合大会(第 18 回日本医療情報学会学術大会), 2017 年.
[4] Mimoto, Kiyomoto, Tanaka and Miyaji, “(p, N)-identifiability: Anonymity Under Practical Adversaries”, The 16th IEEE TrustCom’17.
[5] Kitamura, Nishida, “Living safety technology based on integration of multi-organizational distributed data”, International Conference for Leading and Young Computer Scientists, 2018