

「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

H28 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
  - ・(S1) 検証機能付きデータ管理/削除
  - ・(S4) プライバシを保護したデータ演算
  - ・(S6) セキュアなトレーサビリティ
  - ・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機能分析

### (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 グループリーダー)
- ② 研究項目
  - ・(S3) プライバシポリシー設定支援
  - ・(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
  - ・(S8) データ漏洩抑止技術
  - ・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機能分析

### (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 に示す。各チームの概要は下記のとおりである。

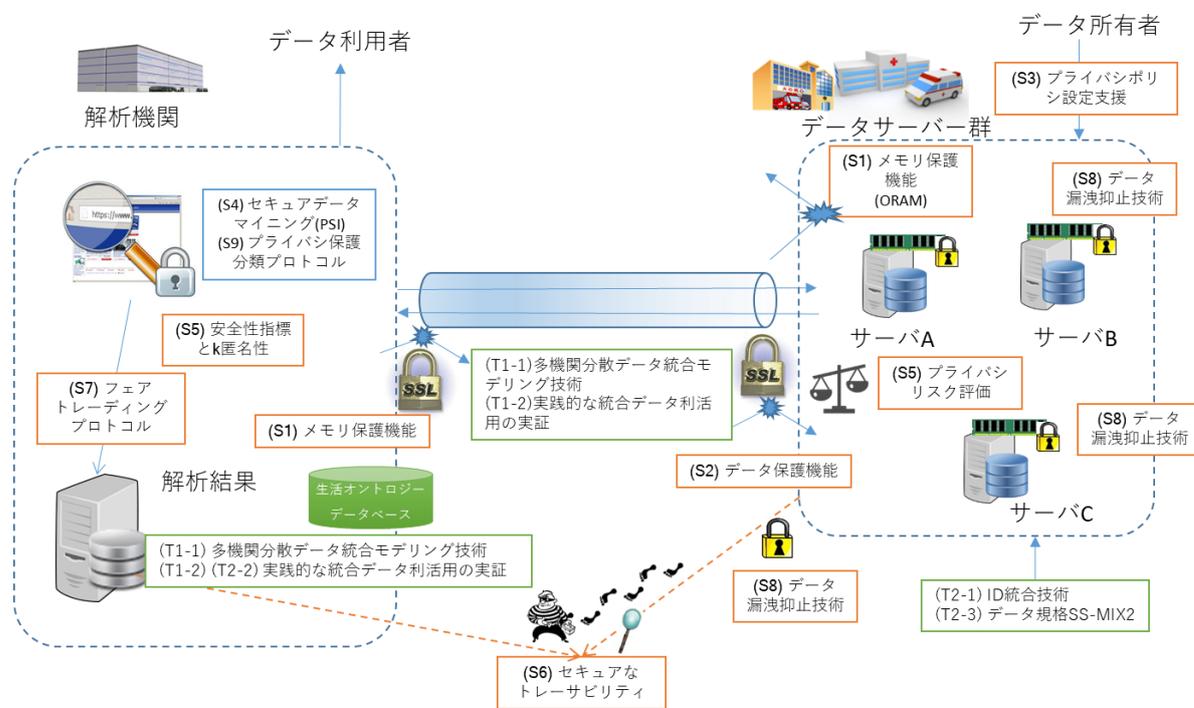


図 1 成果統合図

### 「セキュリティコア技術グループ」

(S1) 検証機能付きデータ管理/削除:クライアントのデータアクセスパターン秘匿として、再帰的行列 ORAM(RM-ORAM)と呼ばれる、記憶装置に制約がある場合に適した新たな ORAM の構成を提案し、その理論的解析を与えた。この RM-ORAM は再帰呼び出しによって計算コストと通信コストが若干増加する代わりに、クライアント側で必要となる記憶容量を大幅に削減することが可能となる。パラメータを適切に選ぶことで、RM-ORAM の漸近的な全体の性能は既存の ORAM 方式より優れていることが検証された。

(S4) プライバシを保護したデータ演算:プライバシを保護しながら、多機関が独立に収集したデータの統計処理を実現できる PSI(Private Set Intersection)方式についての研究を継続した[10]。本年度は 27 年度に提案したアウトソーシング型多機関 PSI プロトコルを実用化するために ICT 技術に不慣れな利用者にとっても容易に実行可能な Web アプリケーションの構築などの取り組みを行なった。これにより、次年度以降のテストベッド実施にむけての準備が完了した状況にある。

(S6) セキュアなトレーサビリティ:クラウドサービスにおけるデータ流通と安全な委託計算のための相互トレーサビリティのフレームワークを提案した。本構成ではデータ所有者とデータ利用者の両方が流通するデータを追跡可能とした。特に、この構成においてデータ所有者はセキュリティとプライバシポリシーに合致する特定のクラウドサービス提供者のグループに対して効率良く自分のデータを配布および追跡が可能となり、また追跡可能性を維持しているデータは共有された後でも更新が可能である。

### 「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、属性間の相関に応じて一般化階層木を構築するアルゴリズムを提案した。また、現実的な攻撃者を想定したリスク評価手法を提案した。さらに、予防安全テストベッド実証グループと連携して医療データセットの匿名化手法について検討し、UI を用いた一般化階層木の作成機能等を含めた匿名化・リスク評価ツールの試作開発を行った。(S3) プライバシポリシー設定支援については、サポートベクトルマシンを用いた推定を行い、提供に対する抵抗意識に関するアンケート 80 項目の回答結果に対して約 86%の精度で推測できることを明らかにした[13]。また、実証実験用システムの開発を行い、約 300 人の被験者を対象に推測手法の有効性を検証した。(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ではこれらを実現するために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・評価した。また、(S2)の担当分については、平成 26 年度に研究が完了している。

### 「予防安全テストベッド実証グループ」

(T1-1) 多機関分散データ統合モデリング技術:傷害データに関してテキストマイニング技術と医療費に着目したクリフ分析を統合した分析システムを開発し、類似状況で特にリスクが高い事故を把握することが可能となった。69 校の小学校の傷害データを対象に分析を行い、分析手法の有用性を検証した[15]。

(T1-2) 実践的な統合データ利活用の実証:(T1-1)で開発した分析システムとセキュリティコア技術グループが開発した PSI とを統合したシステムの開発を行った。これにより、複数の小学校で発生した事故情報を、どこの小学校で起きた事故かは知られることなく、統合して分析することが可能となった。これにより、小学校同士が互いの傷害データを活用し、自環境で起きた事故に類似した状況で起き得るリスクが高い事故を把握して対策をとるといった、新たな傷害予防の連携が可能な状態となった。

### 「医療テストベッド実証グループ」

(T2-1) ID 統合技術: PIX/PDQ の広域拡張である XCA について実証状況および課題を検討した。XCA は患者 ID が既知の条件下で運用可能な規格であり、運用事例も地域、製品が限定的であった。他事例からは、保険証記号・番号を検索キーに含めることで施設間の照合率が向上することが報告されており、保険証の資格確認を包含する「医療等 ID」は、XCA 上での MPI として効率的であると考えられた。一方で、二次利用における患者同意情報の広域な共有は実現されておらず、今後の検討課題とした。

(T2-2) 実践的な統合データ利活用の実証:28 年度は、「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)(S5)の成果について、テストベッドへの適用と評価を行った。(S1)については、SS-MIX2 標準化ストレージのバックアップを対象とする検証を行い、性能が確認できたため、本成果を利用した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プロトタイプ開発を行った。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本クラウドサービスを介した多機関間での情報交換に必要な運用フローの検討に着手した。(S4)(S5)については、1 万件規模の検体検査結果を模したテストデータを作成し、基本的な動作検証を行った。(S4)については、1 万件のデータセットを用いた抽出処理の場合、4 コア・8 スレッド (Intel Xeon E5-2620 v4) の環境上で要する時間は約 200 秒程度の性能であった。(S5)について

は、同じデータセットを用いて「セキュアデータ流通管理グループ」が開発した Web 型の匿名加工ツールがレスポンスよく操作できることが確認できた。双方とも基本的な動作検証が実施できたため、より大規模かつ複雑なデータセットを想定したテストベッド構築に着手した。

(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: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、この事業と連携し、ばらつき補正が可能かどうかを見極めて取り組むこととした。(H29 年度中に結論を得る。)なお、医療・健康管理に用いる IoT のセキュリティ要件に関しても、現時点ではコンセンサスは得られていない。セキュリティ要件についても今後の検討課題とした。

SS-MIX2 については、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、現状の SS-MIX2 標準化ストレージでは患者横断的なデータ処理がファイルシステムの全検索をとれない、実質困難である。このため、(S4)(S6)の成果を展開する目的で、SS-MIX2 データの格納と同時に解析、検索可能なテストベッド機能の試作を行った。FUSE ベースで実装されている Oracle Database File System を採用し、HL7v2 形式で記述、格納されるファイルをリアルタイムに追跡し、機械的にデータベース内の別テーブルへ格納可能なモジュールを PL/SQL で実現した。試作したテストベッドを使用して、処方オーダーメッセージについて動作検証を完了した。

#### 発表論文

[10] Miyaji, Atsuko, Kazuhisa Nakasho, and Shohei Nishida, “Privacy-Preserving Integration of Medical Data”. *Journal of medical systems* 41.3 (2017): 37.

[13] Nakamura T., Kiyomoto S., Tesfay W.B., Serna J, “Easing the Burden of Setting Privacy Preferences: A Machine Learning Approach”, *CCIS, Vol.691, Springer, 2017.*

[15] Kenta Imai, Koji Kitamura, Yoshifumi Nishida, Hiroshi Takemura, Tatsuhiro Yamanaka, “Smart Share of Serious Injury among Schools in the Era of Cloud Computing,” *Injury Prevention, Vol. 22, No. 2, pp. A236, 2016*