「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

H26 年度
実績報告書

宮地　充子

北陸先端科学技術大学院大学　情報科学研究科
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 （北陸先端科学技術大学院大学　情報科学研究科　教授）

② 研究項目

- セキュアデータ管理
- 耐サイバー攻撃
- セキュアデータマイニング
- セキュアデータ運用

(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:三宅　優(KDDI研究所　グループリーダ)

② 研究項目

- 耐サイバー攻撃
- セキュアデータ運用
- セキュアデータマイニング

(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

- セキュアデータマイニング
- 予防安全テストベッド

(3)医療テストベッド実証グループ

① 主たる共同研究者：山本　隆一（東京大学大学院医学系研究科　特任准教授）
② 研究項目
- セキュアデータマイニング
- 医療テストベッド

## §2. 研究実施の概要

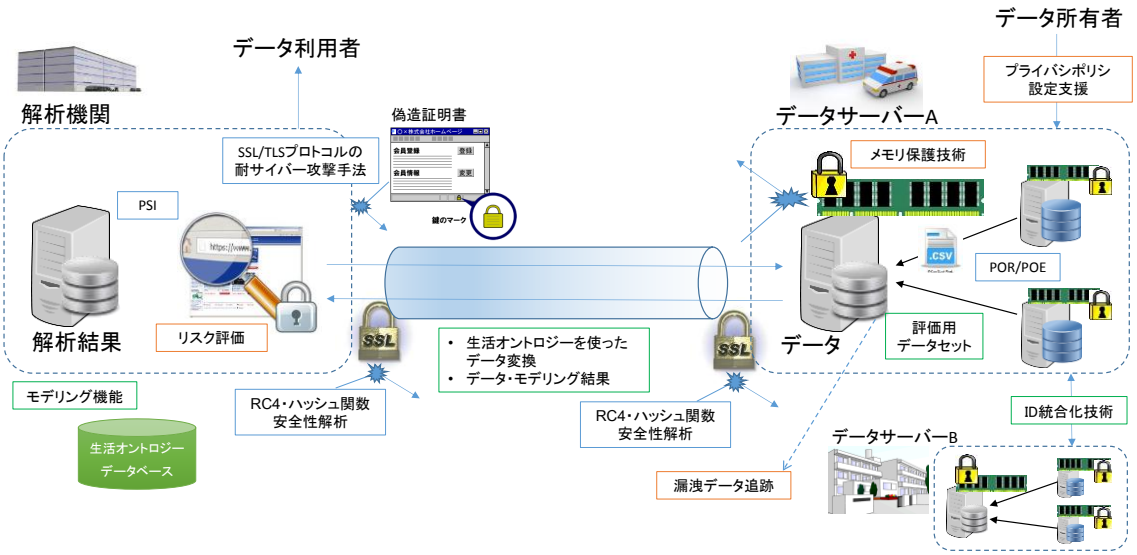


図 1:成果統合図

チーム全体の成果統合図を図 1 に示す. 各チームの研究実施概要は下記のとおりである.

**「セキュリティコア技術グループ」**

耐サイバー攻撃:偽造証明書を排除する証明書検証システムを考案してプロトタイププログラムを作成・検証し, さらに無線 LAN(WPA)における RC4 暗号の安全性を解析し, 理論的脆弱性を発見した[J-1]. セキュアデータ管理:データ所有者がクラウドストレージ上の符号語データを一旦復号することなく符号語データのままでサーバの復旧ができるデータ管理検証手法(POR)を提案し, 演算処理の実装評価により効率性を検証した. セキュアデータマイニング:n 機関がそれぞれ持つデータの共通集合を共通集合以外を秘匿しながら計算する手法(PSI)を各機関のデータ集合の大きさに依存しない通信量で実現する手法を検討した. 一般に PSI においては, 各機関のデータサーバの計算量より通信量がボトルネックになることが多く, データサイズに依存しない通信量での実現は意義が大きい.

**「セキュアデータ流通管理グループ」**

実装上の脆弱性をつく SSL/TLS 等へのセキュリティ機能への攻撃を防ぐメモリ保護技術を研究開発し、OpenSSLに実装して効果を検証した。また、セキュアデータ流通プラットフォーム実現に向けて、匿名化されたデータが流出した場合にその流出元を特定する漏洩データ追跡手法、匿名化されたデータのリスク評価、複雑なプライバシ設定を半自動化するプライバシポリシ設定支援手法、さらにはデータの価値と匿名性のバランスを取る手法の考察[K-1]の研究開発を実施した。

**「予防安全テストベッド実証グループ」**

平成26年度は、平成27年度以降のアルゴリズム開発やテストベッドを用いた実証実験に不可欠なデータベースとして、医療機関データ、救急搬送データ、災害給付データからなるデータベース

を作成した。これらの複数のデータベースから統計モデルを作成する際に、ターミノロジーの標準化を行う技術として、発達行動に関する統計データである Denver II、製品を扱う JICFS/IFDB を利用したオントロジーデータベース技術を作成した。誤飲事故データを例題に、統計モデル作成を行い、統計モデルに基づく傷害シミュレーション[A-3]へ応用することで、モデリング機能の有効性を確認した。予防安全分野へのアプトカムインパクトの検討を目的に掲げており、その準備として、統計モデルや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（医師、工業デザイナー、研究者などから構成）を4回、開催した。

**「医療テストベッド実証グループ」**

テストベッドの作成のステップとして、東大病院に蓄積された SS-MIX2 形式で保存された診療情報を、母集団とは異なる特性にはなるものの、男女比や年齢階層等に矛盾を生じないことを目的として、男女比、5歳区分の年齢構成を維持し、また、検査項目をグルーピングした上で、 統計法における国勢調査ミクロデータの匿名化手法等として一般的に用いられているランダムスワッピングを行い、再特定性はないものの、原データには戻り得ない評価用データセットを作成した。また ID 統合化技術については IHE PIX-PDQ に関して調査を行い適切な Master Patient Index を置くことで、理論的に実現可能であることが判明したが、多量の情報を扱うビッグデータで現実的に適応可能か、すなわちスケーラビィリティについては今後検討したい。

**「参考文献」**

[J-1] 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), Lecture Notes in Computer Science, Springer-Verlag, to appear.

[K-1] S. Kiyomoto, Y. Miyake, “Data Value Estimation for Privacy-Preserving Big/Personal Data Businesses,” FMfI2014 Post-proceedings, 2015, accepted.

[A-3] Y. Nishida, D. Nakazato, K. Kitamura, H. Mizoguchi, T. Yamanaka, “Childhood Home-Injury-Situation Simulation for Individual Environments Based on Child Physical Model and Injury Semantic Structure Model,” the 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) , 2015 (in press)