戦略的創造研究推進事業　CREST
研究領域「イノベーション創発に資する人工知能基盤技術の創出と統合化」
研究課題「プライバシー保護データ解析技術の社会実装」

研究終了報告書

研究期間　2019年4月～2022年3月

研究代表者：花岡　悟一郎
（（国研）産業技術総合研究所　サイバーフィジカルセキュリティ研究センター、首席研究員）

§1 研究実施の概要

(1)実施概要

本研究においては、入出力情報を秘密に保ったままデータ処理を実行可能なプライバシー保護データ解析技術について、広範な適用範囲に対して誰でも利用可能な汎用的技術と金融データ解析を特に念頭においた専用的技術の双方に関して研究開発を行うことで、社会実装を推進していくことを目的としている。また、具体的な達成指標として、①外部企業・機関からの要望に応えて作成した実アプリケーションの数、および、②具体的にデータの提供を受けた金融機関の数と得られた精度を設定している。本研究期間においては、①に関し、外部企業・機関から依頼を受けて、ZenmuTech 社が本研究課題による成果に基づき開発を行った秘匿計算システムが現在までに 4 件あり、②に関しては、現在 5 つの金融機関と契約を結び、実取引データの解析委託を受けて不正送金検知のデータ解析を進め、2 銀行でのプライバシー保護協調学習により、1 日あたり 900 件の検知まで許容した場合、それぞれの銀行で85%以上、90%以上の精度が得られている。したがって、本研究期間における上記の目的は十分に達成されたものと考えられる。目的の達成に関し、各グループにおいては、特に以下の活動を行った。

花岡グループでは、高速な汎用的秘匿計算アルゴリズムの理論設計および安全性評価を進め、これらの成果をもとに汎用的秘匿計算技術「QueryAhead」の開発を行った。さらに、QueryAhead を用いることで外部企業や組織からの要望に応え、実運用を想定した秘匿計算システムの開発を 4 件行った。具体的には、A 社より秘匿不動産マッチングシステム等、B 技術組合より秘匿線形回帰システム等についてそれぞれ複数の依頼を受け、QueryAhead を用いることでわずかな工数のみで開発を完了している。これらのシステムはいずれも A 社や B 技術組合によって実利用が検討されている。また、理論設計および安全性評価を行った秘匿計算アルゴリズムについては、ACM CCS, ACM AsiaCCS, CRYPTO, EUROCRYPT, ASIACRYPT, TCC, PKC などの情報セキュリティ分野のトップ国際会議に合計 37 件もの成果が採録されており、特に EUROCRYPT 2020 においては最優秀論文賞を授賞している。

盛合グループでは、小澤グループ、菅原グループとともに、金融機関 5 行(千葉銀行、三菱 UFJ 銀行、中国銀行、三井住友信託銀行及び伊予銀行)と連携し、プライバシー保護連合学習を活用した不正送金検知の実証実験を行った。実証実験では、プライバシー保護連合学習技術「DeepProtect」を活用し、目標としていた不正送金の検知精度 80%以上を達成するとともに、一銀行では検知できなかった不正送金の被害に遭った取引の検知や、不正送金に悪用された口座の早期検知に成功した。また、本技術の民間企業への技術移転も開始した。

浅井グループでは、幅広いユーザが、多様な秘匿化計算を実環境で効率的に実行するためのフレームワークを構築することを目指した技術開発を行なった。準同型暗号に関しては、効率的なゼロ知識証明、任意 1 入力関数プロトコル、悪意のある参加者に対しても安全で効率的なプロトコルを開発するとともに、実装の効率化を進め、電子投票システムなどの応用に提供した。秘密分散技術に関しては、群論的手法に基づいた通信回数が少なく効率的なシャフル、循環シフト、行列演算のプロトコルを開発し、QueryAhead(上述)に提供した。また、機械学習に必須な max/min/argmax/argmin およびテーブル参照の秘匿計算を効率的に行う手法を開発した。これらの技術を応用し、ゲノム配列を含む多様な文字列データの解析に役立つ秘匿全文検索アプリケーションを開発・実装した。さらに、CPU のセキュリティ技術の一つである Intel SGX を用いて、実用的な速度で大規模な個人ゲノムデータを解析することのできるシステムを開発した。

小澤グループでは、盛合グループとともに、秘匿計算手法を使ったニューラルネットや決定木アンサンブルなど、実用性を重視したプライバシー保護機械学習モデルの開発を中心に行った。複数組織がもつデータを互いに明かさず協調学習し、説明可能性に優れた分類が行えるプライバシー保護決定木アンサンブルモデルを提案した。具体的には、勾配ブースティング法の一種である XGBoost を個々の組織がもち、組織固有のデータで求めたモデルを共通

鍵で暗号化し、Semi-honest な外部サーバを介して各組織のモデルを統合する協調学習スキームを提案した。金融機関 5 行と連携した不正送金検知の実証実験において、提案した FL-XGBoost は有効に機能し、KPI である不正検知率 80%を超える性能を達成した。さらに、組織間協調学習が可能なプライバシー保護異常検知モデルや組織間でデータを共有しなくても高次元データを可視化できるプライバシー保護データ可視化モデルを開発した。

菅原グループでは、実社会において機密性および可用性が高く、また運用面での利便性の高いシステムの開発・改良を行った。導入先企業における実証実験のなかであがった要望を取り込むことで、実用性の高い機能を実装できた。また、統合学習の役割を司る中央サーバは大規模災害が発生した際でもサービス提供を継続できるような構成を実現している。ハードウェア面だけでなく、ソフトウェア面においても、複数のモデルのバージョンを管理可能にし、また暗号化したデータを通信経路上で送受信することで、より機密性高く、実社会での運用に耐えうるシステムが開発できた。

次いで、研究ビジョンとその達成状況について述べる。研究ビジョンは、「個人情報や企業の機密情報等のあらゆる機微情報を、安全性を保ったまま任意のデータ処理に適用可能とするプライバシー保護データ解析技術を 2028 年までに完成させる。これにより、すべての機密データの総合的な活用がなされ、様々な高度情報サービスが実現した社会を創出する。」である。達成状況は、A 社や B 技術組合より依頼を受け、秘匿計算技術に関し、実利用を見据えた実アプリケーションの開発を4件行い、さらに、金融機関 5 行と連携し、プライバシー保護連合学習を活用した不正送金検知の実証実験を行い、検知精度 80%以上を達成する等、個人情報や企業の機密情報の利活用に向けたプライバシー保護データ解析技術の社会展開の進展が有意に確認できる結果を得られた。

(2)顕著な成果

＜優れた基礎研究としての成果＞

1. 理論的に最適な通信回数を達成する秘匿シャッフルプロトコルの設計

概要: 2 者の秘匿計算における秘匿シャッフル(ランダム置換)プロトコルについて、実行時の通信回数が最適となる方式の実現方法を示した。提案方式は、特殊な乱数組を用いてランダム置換を始めとする群作用を効率化するための新しいフレームワークの構築によって得られたものであり、従来の 2 者秘匿シャッフルと比較して 100 倍以上の高速化を達成している。この応用として、DB 処理において特に重要なソート、JOIN などの処理の効率化にも成功した。本提案方式は、QueryAhead の中核的な機能を担っている。本成果は、情報セキュリティにおけるトップ国際会議 ACM CCS 2021 に採録されている。

2. 理論的に最適な放送暗号の設計

概要: 秘密計算の要素技術の一つでもある放送型暗号について、30 年近くにわたる重要な未解決問題であった暗号文長と鍵長がいずれも定数サイズになる方式の実現方法を明らかにした。提案方式は、楕円曲線理論および格子理論の異なる代数的構造を同時に活用することにより得られたものとなっており、今後の新たな暗号方式の設計において重要な知見を与えるものと考えられる。本成果は、暗号理論におけるトップ国際会議 EUROCRYPT 2020 において最優秀論文賞を受賞した。

3. プライバシー保護協調学習の高度化及び新方式の提案

概要: プライバシー保護協調学習において、学習参加者の数やニューラルネットワークの深層による通信量が、従来の技術と比べ半分以上削減できるアルゴリズムの提案や、ネットワークフォールトにロバストなアルゴリズムの設計・評価を行い、ともに IEEE Access に採録された。また、深層学習ベースの DeepProtect と異なり、判定結果の分析が行いやすい FL-XGBoost も提案し、国際会議 ICONIP2020 で発表、金融機関との実証実験で高い精度が出ることが検証された。

＜科学技術イノベーションに大きく寄与する成果＞

1．汎用的秘匿計算技術 QueryAhead の開発

概要：広範な適用範囲に対して誰でも利用可能な汎用的秘匿計算技術「QueryAhead」の開発を行った。QueryaAheadは、基本演算（加算／減算／乗算）、SELECT、WHERE、FROMアルゴリズムに加え、除算、JOIN、ORDER BY、基本統計（平均／最大／最小／合計）、GROUP BY について秘匿計算を容易に実行可能とする技術であり、特に一般的なデータベース処理の秘匿化を簡便に実装できるだけでなく、それにとどまらない広範なデータ処理の秘匿化が可能である。

2． QueryAhead に基づく、外部企業・機関からの要望に応えた秘匿計算システムの開発

概要：前項で記載した汎用秘匿計算技術 QueryAhead を用いて、外部企業・組織と連携しながら、秘匿計算技術の実アプリケーションへの適用を推し進めている。具体的には、A 社より秘匿不動産マッチングシステム等、B 技術組合より秘匿線形回帰システム等についてそれぞれ複数依頼を受け、開発を行っている。。これらのシステムはいずれも A 社や B 技術組合によって実利用が検討されている。いずれも実装は秘匿計算の専門家ではない一般のエンジニアによって短期間に行われており、これらの開発事例は誰でも簡単に秘匿計算を実装できることを目指した QueryAhaed の利便性、および QueryAhead が秘匿計算の社会展開加速に寄与する技術であることを示している。

3．金融機関 5 行と連携したプライバシー保護協調学習による不正送金検知の実証実験

概要：金融機関 5 行と連携し、プライバシー保護協調学習を活用した不正送金検知の実証実験を行った。同種の不正取引を検知する銀行グループに分けて実験を行い、本研究課題での達成目標としていた複数組織による協調学習で単独組織での学習より高い精度が達成される事例を示すことができ、銀行へのシステム導入につなげた。また、プライバシー保護機械学習エンジンのプラットフォーム化を進め、今後の組織を超えたプライバシー保護データ利活用に繋げた。

＜代表的な論文＞

1．Nuttapong Attrapadung, Goichiro Hanaoaka, Takahiro Matsuda, Hiraku Morita, Kazuma Ohara, Jacob Schuldt, Tadanori Teruya, Kazunari Tozawa. “Oblivious Linear Group Actions and Applications, ” Proc. of ACM CCS 2021, to appear.

概要：2 者の秘匿計算における秘匿シャッフル（ランダム置換）プロトコルについて、実行時の通信回数が最適となる方式の実現方法を示した。提案方式は、特殊な乱数組を用いてランダム置換を始めとする群作用を効率化するための新しいフレームワークの構築によって得られたものであり、従来の 2 者秘匿シャッフルと比較して 100 倍以上の高速化を達成している。この応用として、DB 処理において特に重要なソート、JOIN などの処理の効率化にも成功した。本提案方式は、QueryAhead の中核的な機能を担っている。本成果は、情報セキュリティにおけるトップ国際会議 ACM CCS2021 に採録されている。（上記＜優れた基礎研究としての成果＞ 1．に関する論文成果。）

2．Shweta Agrawal and Shota Yamada, “Optimal Broadcast Encryption from Pairings and LWE,” Proc. of EUROCRYPT 2020, pp.13-43, 2020.

概要：秘密計算の要素技術の一つでもある放送型暗号について、30 年近くにわたる重要な未解決問題であった暗号文長と鍵長がいずれも定数サイズになる方式の実現方法を明らかにした。提案方式は、楕円曲線理論および格子理論の異なる代数的構造を同時に活用することにより得られたものとなっており、今後の新たな暗号方式の設計において重要な知見を与えるものと考えられる。本成果は、暗号理論におけるトップ国際会議 EUROCRYPT 2020 において最優秀論文賞を受賞した。（上記＜優れた基礎研究としての成果＞ 2．に関する論文成果。）

3．Fuki Yamamoto, Lihua Wang, Seiichi Ozawa, “New Approaches to Federated XGBoost Learning for Privacy-Preserving Data Analysis,” Proc. of ICONIP2020, pp.558-569, 2020.

概要：高性能かつ説明性に優れた決定木アンサンブルである XGBoost に協調学習スキーム

を導入したFL-XGBoostを提案した。提案手法では、各組織のローカル情報で個別に学習したあと、モデルを暗号化して、勾配情報などの評価値とともに中央サーバに送られる。中央サーバでは、評価値に基づいて一つのモデルを選択し、これを各組織で共有する。本提案手法は、学習の高速性と有効性を兼ね備えており、現在、特許申請中である。（上記＜優れた基礎研究としての成果＞ 3. に関する論文成果。）

§2 研究実施体制

(1)研究チームの体制について

① 花岡グループ
研究代表者:花岡 悟一郎 (産業技術総合研究所 サイバーフィジカルセキュリティ研究センター 首席研究員／研究チーム長)
研究項目
・プライバシー保護データ解析技術の社会実装

② 盛合グループ
主たる共同研究者:盛合 志帆 (情報通信研究機構 サイバーセキュリティ研究所 研究所長)
研究項目
・プライバシー保護データ解析技術の高度化と社会実装

③ 浅井グループ
主たる共同研究者:浅井 潔 (東京大学 新領域創成科学研究科 教授)
研究項目
・プライバシー保護情報処理の高度化と汎用化

④ 小澤グループ
主たる共同研究者:小澤 誠一 (神戸大学 数理・データサイエンスセンター 教授)
研究項目
・プライバシー保護機械学習の開発と社会実装

⑤ 菅原グループ
主たる共同研究者:菅原 貴弘 (株式会社エルテス 代表取締役)
研究項目
・プライバシー保護データマイニング手法の事業化及び社会実装

(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

(1)の研究チームに加えて、金融機関との実証実験を進める上で、(株)NTT データ経営研究所と連携、取引データ解析では秘密計算スタートアップの EAGLYS(株)と連携している。また、(株)NTT データ経営研究所を通じてマネーロダリングシステム対応国内トップシェアの Oculus®シリーズを持つ NTT データジェトロニクス(株)と成果導入方法に関する議論も行っている。このほか、研究成果の技術移転先の一つとして(株)インフォーズとの連携、本技術に興味を持つ KPMG ジャパンとの連携を進めてきた(2019 年 11 月の「秘密計算ハッカソン」の開催など)。さらに、金融庁や全国地方銀行協会と定期的な情報交換を行っている。