IPA

2019 年度 未踏 IT 人材発掘・育成事業 採択案件評価書

1. 担当 PM

藤井 彰人
（KDDI 株式会社 執行役員 ソリューション事業本部 サービス企画開発本部長）

2. クリエータ氏名

櫻井 碧（早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻）

3. 委託金支払額

2,304,000 円

4. テーマ名

生命情報解析向けインタプリタを搭載した秘密計算用クラウド

5. 関連 Web サイト

BI-SGX の GitHub リポジトリ：https://github.com/hello31337/BI-SGX

6. テーマ概要

ゲノムデータのようなプライバシー性の非常に高い情報に対して、セキュアに統計の計算等を行うためには、ソースコードを含めて厳重なセキュリティの確保が必要であり、かつ現実的な処理時間になるよう計算コストが低く抑え、ユーザにとって利用しやすいシステムが必要となる。

本プロジェクトは、秘密計算を実行可能な生命情報解析に特化したクラウドプラットフォームを開発した。生命情報はプライバシー性が高く、厳格なセキュリティとその実行管理が求められるが、Intel SGX を活用して秘密計算プラットフォームを実現し、保護領域上に利用が容易なインタプリタを実装した点が特徴である。

7. 採択理由

本提案は、秘密計算を実行可能な生命情報解析に特化したクラウドプラットフォームの開発を目指すプロジェクトである。生命情報はプライバシー性が高く、厳格なセキュリティとその実行管理が求められるが、Intel SGX を活用して

これを実現しするだけでなく、保護領域上にインタプリタの実装も予定している点が特徴である。性能や利用しやすさなどにも言及しており、ドメインを特化した提案ではあるが、他の領域にも発展可能な内容である。クラウド市場の拡大は説明するまでもないが、IaaS でのベアメタルの利用拡大に合わせて、秘密計算クラウドの活用の現実味が増しており、グローバルに提案可能なフレームへの拡大も期待して採択した。

8. 開発目標

本プロジェクトの目的は、Intel SGX の実用的なパフォーマンスを活用し、研究者の負担を最小限に留めながら秘密計算を実行可能なクラウドプラットフォームを実現することである。また、生命情報を保有するデータ所有者が安心して利用することのできる、セキュアクラウドストレージとしての機能も提供し、データを安全に用いて秘密計算をする機能の実現を目標とした。

その他にも、機密情報を抜き出せてしまうような実行定義を阻止し、出力情報が匿名化されていることを担保する「アウトプットプライバシの保護」や、SGX が不得手とするサイドチャネル攻撃への対抗手段も実装することを目標とした。

9. 進捗概要

本プロジェクトでは、Intel SGX が RAM 上に生成する保護領域「Enclave」上でインタプリタを駆動させることで、SGX 公式 SDK の極めて不可解かつ煩雑な仕様に悩まされることなく解析内容を記述し解析できる秘密計算用クラウドシステム「BI-SGX」（Bioinformatic Interpreter on SGX-based Secure Computing Cloud）を開発した。BI-SGX のシステム全体の概要図を図 1 に示す。

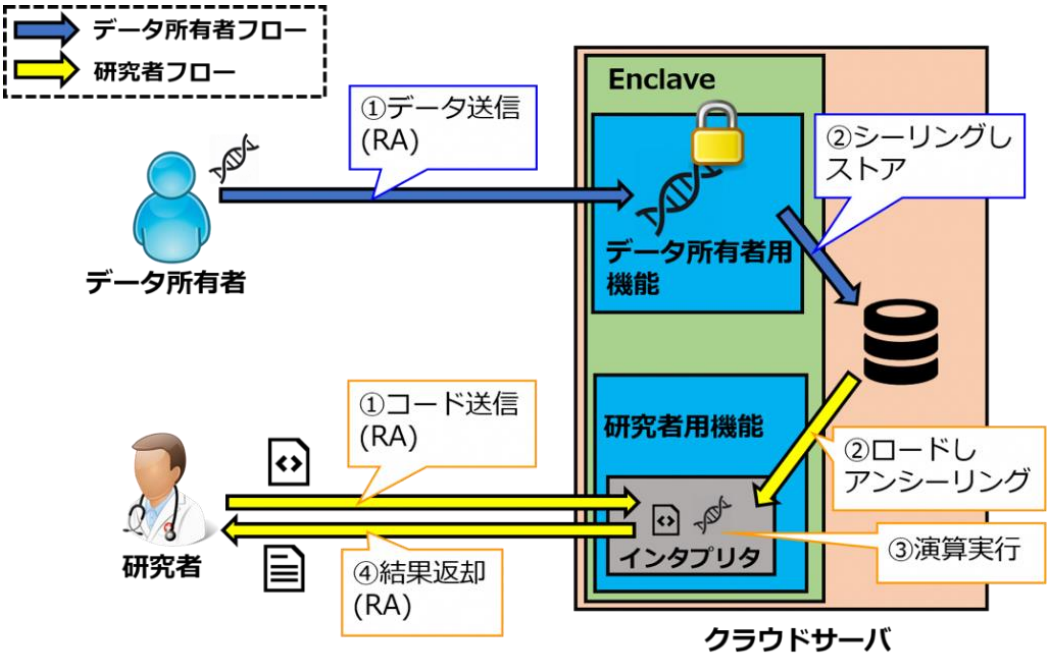

図 1. BI-SGX のシステム全体の概要図

BI-SGX は、ゲノムデータ等の機密情報を持つデータ所有者向けのストレージ機能と、そのストレージ上のデータを用いて様々な解析を行うことのできる秘密計算機能を提供している。データ所有者のアップロードしたデータは、Enclave 内でのみ復号され、かつ SGX のシーリングと呼ばれる 128bit AES/GCM ベースの強固な暗号化によって保護された状態でストアされる。

一方で、研究者は BI-SGX が提供する独自の言語「Qliphoth」を用いることで、容易にクラウド上で実行したい処理を記述することができる。Qliphoth によるアリル頻度分析の秘密計算を要求するコードを図 2 に示す。

```
1:      func main()
2:      var a
3:              a = alleleFreq("21", 10417440, "JPT", "none")
4:      end
```

図 2. Qliphoth の組み込み関数を利用したプログラム例

Qliphoth を使用することにより、研究者は SGX SDK を用いた場合に比べて劇的に小さい負担で実行したい処理を記述可能になる。SGX SDK で開発を行った BI-SGX は、本体のコード量が合計約 40,000 行という膨大なコード量にまで膨れ上がっている。しかしながら、Qliphoth を利用すれば、このように莫大な負担を強いられる SGX SDK を使わずに済み、たった 4 行で後述の GWAS のような比較的大規模な処理についての秘密計算を実行させることができる。

Qliphoth はプログラミング言語として基本的な機能を一通り備えているだけでなく、言語仕様レベルでアウトプットプライバシを侵害するような処理を根本的に排除している。アウトプットプライバシを侵害する処理とは、例えば平均を計算すると謳いながら、単一のデータに対する平均値、即ちデータそのものを取り出す、というような「プロトコル上は正しいが結果的に得られる値がプライバシーを侵害している」処理を指す。Qliphoth では、SQL 等とは異なり、データの集合であるデータセットよりも細かい粒度で条件指定を行うことが根本的に不可能である為、アウトプットプライバシを侵害するような処理を排除することに成功している。

更に、BI-SGX は SGX アプリケーションが概して苦手とするサイドチャネル攻撃に対する防護も実現している。サイドチャネル攻撃の中でも極めて SGX に対して有効な攻撃に、制御チャネル攻撃（Controlled-Channel Attack）が存在する。これは、ページフォルトを悪用して条件分岐先の変数や関数へのアクセスを観測し、その結果から分岐条件となった変数の値を推定するような攻撃である。しかし、BI-SGX では Qliphoth のスクリプトコードは Enclave により完全に保護されており、かつインタプリタ本体は字句解析器がスクリプトコードから取得する「トークン」という、言わばその要素の種別に相当する粒度でしか条

件分岐を行わない。よって、制御チャネル攻撃をもってしても秘密情報を BI-SGX から抽出することが不可能となっている。

10. プロジェクト評価

Intel SGX を活用した、生命情報解析向けの秘密計算クラウドプラットフォーム「BI-SGX」をプロジェクト期間中に開発した。単に Intel SGX 活用の秘密計算処理の可能性、実用性を証明しただけではなく、生命情報解析における実利用を想定した Enclave へのデータ転送、インタプリタ、ゲノムワイド関連解析（GWAS）なども実装されていることは、まさに生命情報解析向けの秘密計算プラットフォームと言え、高く評価すべきである。

Microsoft Azure などのクラウド環境でも Intel SGX を搭載したベアメタルサーバが活用できる時代になっている。BI-SGX が国内だけにとどまらず、生命情報解析の分野においてグローバルに活用されることを期待している。

11. 今後の課題

BI-SGX を実際の生命情報解析の現場でサービスとして活用してもらうためには、ID 管理やリソース管理などビジネスやオペレーションをサポートするための機能が必要になる。生命情報解析の現場でのテスト利用を実施し、これら機能の拡充を期待したい。

加えて、Microsoft Azure などの Public Cloud 環境において、BI-SGX を動作させどこでも利用可能であることを証明することも大切であろう。

本プロジェクトはクリエーター人での開発体制であったが、今後の発展を考えれば、プロジェクトの内容を広く公開して協力者を募ることや、Intel や Microsoft との連携も必要である。