案 実践的なサイバーセキュリティの研究分野　研究構想の具体例

資料5

信頼ある分散型データの活用を実現するセキュリティ基盤技術

目標　プライバシー等を保護しつつ分散型データを活用するためのセキュリティに関する基盤技術の確立を目指す。

概要　分散型データの活用に際して重要となるプライバシーや非開示情報の保護について、分散させたままこれらを保護し、保護したまま制御、活用するための基盤技術の確立を目指す。加えて、そのような基盤技術を確立しつつ、様々なサイバー攻撃観測データを統合的に共有しアカデミックに分析し、より高度なセキュリティ対策を見出すための研究も併せて推進する。これにより、データの保護と活用を両立した信頼ある自由なデータ流通（DFFT）に資する。

1. 研究内容と背景

- データエコノミーが進展した社会においては、データの活用は、マーケティングなどの商業的な目的だけでなく、感染症対策などの公益的な目的にもつながり得るが、その際、プライバシーや非開示情報の保護を両立させることは極めて重要である。
- 特に、我が国の状況に鑑みると（メガプラットフォーマーがいない一方、分散的に多くのデータが存在すること等）、分散型データの活用が重要と考えられ、そのためには、セキュリティ技術により分散的に存在するデータに保護をかけること、保護したまま分散型データを制御、活用できることが必要であり、非構造化データを構造化することを含め、そのための基盤技術の確立を目指す。
- また、そのような基盤技術を確立しつつ、サイバー攻撃観測データに活用することで、多角的かつ多地点に分散した観測データを、匿名性やプライバシー・非開示情報を保ったまま統合的に共有し、アカデミックな分析を行い、より高度な対策を見出すことが可能となるため、これも併せて推進する。

2. 具体的な研究例

セキュリティ分野の研究者と、機械学習やデータサイエンス分野等の研究者のコラボレーションによる研究実施が期待される。

A　プライバシー等を保護した社会的・経済的データの共有・分析基盤

個人情報や行動履歴等を活用したマーケティングやリコメンデーション、感染症対策などといった社会的・経済的価値創出に際して、Cの共通技術と連携しつつ、データ提供者が自己の情報をコントロールできる権利を維持し、必要な情報のみを共有・分析するための技術の研究を実施する。

例　・キーワードを秘匿したまま検索できることを狙いとした研究（検索可能暗号等）
・データ所有者が自己の情報をコントロールできることを狙いとした研究（分散・自己主権ID、Biometric Information Protection等）
・データを活用する多様な組織間で、データの相互運用を可能とし、円滑に活用できる環境の実現を狙いとした研究

B　プライバシー等を保護したサイバー攻撃観測データの共有・分析基盤

サイバー攻撃観測データについて、Cの共通技術と連携しつつ、情報交換の際にデータがどのように処理されるべきかなどを検討し、データの構造化、もしくは非構造化データのまま活用するための研究を実施する。また、多様な組織からサイバー攻撃観測データの共有が促進されるべく、インセンティブ設計技術の研究を実施する。

例　・サイバー攻撃に付随する多様なログ等を統合的に扱うためのデータの構造化を狙いとした研究
・データ提供者が、自身の提供するデータの量・価値に応じて適切なインセンティブ（分析できるデータの量や期間等）を設定できるようにすることを狙いとした研究（ダイナミックプライシング、オークション理論等）

C　共通技術の深化・高度化及び新たな革新技術の創出

様々な分散型データに含まれる、一部もしくは全部のプライバシー情報や非開示情報等を秘匿したまま分析を可能にする技術など、データの保護・活用に関して、A、Bと連携しながら技術の深化・高度化を図るとともに、新たな革新技術の創出に挑む。その際、既存のプライバシー強化技術（PETs技術）の活用も念頭に置きつつ、必要に応じて新たなPETs技術の創出を目指すことも含む。

＜保護・活用のための基盤技術＞
例　準同型暗号、秘密計算、（局所）差分プライバシー　（主に保護）
例　プライバシー保護データマイニング、Federated Learning (FL)　（主に活用）

＜制御のための基盤技術＞
例　関数型暗号、ゼロ知識証明

3. 想定する研究の進め方（PI人数規模のイメージなど）

A PI 5名、B PI 2-3名、C PI 7名　程度あるいはそれ以上

実践的なサイバーセキュリティの研究分野　研究構想の具体例

人工知能セキュリティ

目標　人工知能（機械学習）が浸透する社会において機械学習とセキュリティに関する基盤技術の確立を目指す。

概要　機械学習の重要性の高まりを受け、機械学習に立脚したシステムに対する情報セキュリティの重要3要素（機密性、完全性、可用性（CIA））の確立を根源的な狙いの1つとする。また、機械学習技術を高度に活用したセキュリティ技術の開拓を狙いとする。いずれも理論から応用に至る包括的な研究により基盤技術の確立を目指す。

1. 研究内容と背景

- 将来の機械学習技術を高度に適用する社会においては、機械学習を活用したシステム全般に、より高度なレベルのセキュリティと信頼性が求められる。機械学習に対するセキュリティ強化を狙いとした研究が意欲的に研究され始めているが、包括的なフレームワークの確立には至っていない。
- 機械学習を応用したシステムを対象とし、機械学習に対する情報セキュリティの重要三要素---機密性 (Confidentiality)、完全性 (Integrity)、可用性 (availability) を確立することを狙いとし、理論から応用に至る包括的な基盤技術に関する研究を行う。
- 機械学習を従来のセキュリティ対策技術に対して高度に応用した飛躍的な技術の開発研究、およびいわゆるオフェンシブセキュリティ（攻撃者視点の研究）のアプローチにより、攻撃者による機械学習の悪用がもたらす脅威と対策技術に関する基盤的研究を行う。

2. 具体的な研究例

機械学習とセキュリティの研究者のコラボレーションによる研究実施が期待される。機械学習のCIA 確立は基礎研究を中心とし、機械学習のセキュリティ技術への応用は、応用研究を中心とする。

A　機械学習のCIA確立

＜機密性＞

機械学習が扱うデータ、および訓練済みのモデルを保護することを狙いとした研究

例・ モデル抽出攻撃 (model extraction)と対策 （理論、応用）

・ データ再構築攻撃 (model inversion)と対策 （理論、応用）

・ プライバシー保護データマイニング技術（理論、応用） 基礎研究として差分プライバシーを含む秘密計算の機械学習への適用（理論、応用） 基礎研究として秘密計算関連研究全般を含む

＜完全性＞

機械学習を応用したシステムに対する悪意がある入力に対する保護を狙いとした研究

例・ 敵対的機械学習 (adversarial machine learning) に関する研究 （理論、応用）

・ 敵対的サンプル (adversarial example) の生成および検出方法に関する研究（理論、応用）

＜可用性＞

MLaaS（machine learning as a service）のような機械学習アルゴリズムの出力を提供するクリティカルなシステムにおいて、不正なクエリの発行やデータ汚染が行われた際もシステムが影響を受けずに利用可能な状態を維持する技術の研究

例・ 機械学習アルゴリズムに対する不正な入力、パターンの発見、検出技術（応用）

B　機械学習のセキュリティ技術への応用

＜防御技術＞

本テーマは機械学習を高度に応用し、一般的なセキュリティ対策技術の飛躍的な性能向上を図る狙いとした研究である。スパムフィルタやマルウェア検出においてMLの適用が進むが、ML技術の進展に応じ、更なる発展の見込みがある。また、攻撃技術として機械学習が使われた場合、いかに機械学習で対抗できるかに関する基礎検討を進める。

例・機械学習を用いたマルウェアの検出・分類技術

・機械学習を用いた悪性ウェブサイトの検出・分類技術

・機械学習を用いた侵入検知技術

＜攻撃技術＞

攻撃者が高度に機械学習を利用することで生じる脅威に関する研究

例・ 機械学習を用い、実際には存在しない画像、動画、音、テキストを巧妙に生成する技術と対策方法の研究

・ 本来秘匿されるべきデータを機械学習を適用することにより、高精度で推定するサイドチャネル攻撃の評価と対策方法の研究

3. 想定する研究の進め方（PI人数規模のイメージなど）

AはCIA確立の元となる基礎研究を含め理論を中心に進める。B は応用を中心に進める。

A PI 5名 （理論4、応用1） B PI 5名（理論1、応用4） 程度あるいはそれ以上