様式３

AI技術を活用した不正プログラム解析手法の高度化

官民研究開発投資拡大プログラム（PRISM）

「AI技術領域」

令和４年度成果

令和５年３月

警察庁

資料1 「AI技術を活用した不正プログラム解析手法の高度化」の概要

アドオン額：74,250千円（警察庁）
元施策・有/PRISM事業・新規

課題と目標

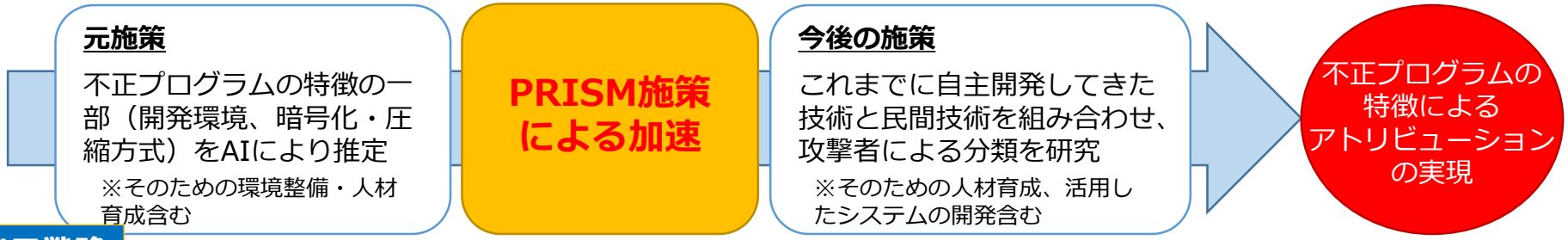
■（課題）不正プログラムの難読化の進行により、詳細解析に要する期間が長期化。
■（目標）AI技術の活用等により、我が国の技術を標的とするサイバー攻撃で利用される不正プログラムの分類や機能特定の自動化をはじめとする各種不正プログラム検体の動作解明プロセスを効率化・高度化する機能の実装を促す。

「AI技術を活用した不正プログラム解析手法の高度化」の概要

■元施策：
機械学習を活用した解析手法に関する調査を行い、不正プログラム検体の開発環境や暗号化・圧縮方式を推定する手法を自主開発し、一部運用を開始。

■PRISMで実施する理由：
警察への先端技術の継続的な導入及び不正プログラム解析の高度化・効率化により、民間企業（セキュリティ企業等）における技術開発が促進されるなどの投資誘発効果のほか、警察における不正プログラム解析に要する財政支出の効率化が見込まれるため、PRISMで実施する。

■テーマの全体像 ：



出口戦略

○ 不正プログラム解析技術を警察において迅速に導入することで、サイバー事案の予防、サイバー空間の安全・安心の確保を実現
○ 攻撃者の公表・非難を行う「パブリックアトリビューション」に寄与し、サイバー攻撃抑止に貢献
○ サイバー事案の手口に係る情報についての警察からの注意喚起に活用
○ 官民連携の枠組みにより不正プログラム解析の継続的な高度化を実現

民間研究開発投資誘発効果等

○ 投資誘発効果：AI技術研究開発分野、サイバーセキュリティ研究開発分野等の研究機関や民間企業における技術開発を促進
→ ユーザー（警察庁）との十分な意見交換を通じ、研究開発・技術実装を推進
○ 財政支出の効率化：警察における不正プログラム解析の省力化・効率化により年間9,000万円相当の効果
○ 民間貢献：民間から不正プログラムやIoC情報等データセット、最新技術に関する知見の提供（5,000万円程度）

資料2　「AI技術を活用した不正プログラム解析手法の高度化」の概要

アドオン（警察庁）：74,250千円
元施策名：サイバー空間の脅威への対処に必要な解析用資機材の整備（R3補正：約130,000千円、R4：約280,000千円）、情報技術解析のための外部訓練の委託（R4：約4,000千円）、属性情報推定装置の整備（R2：約25,000千円）等

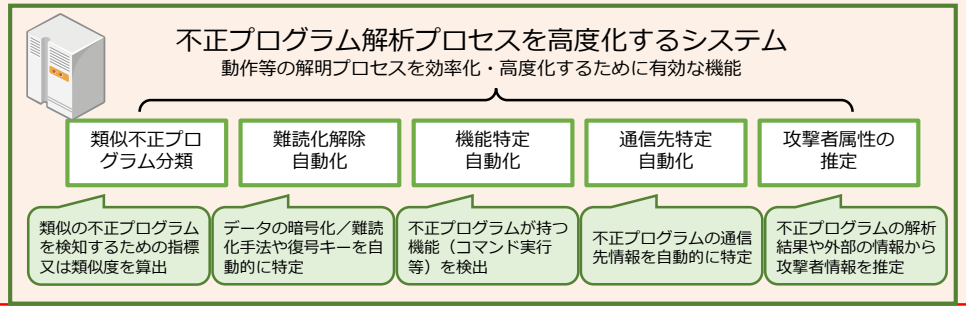
警察における不正プログラムの解析結果の活用

犯罪捜査等への活用
　不正プログラムの詳細解析結果を犯罪捜査等に活用。
○　犯行に用いられたプログラムが犯罪の構成要件を満たすものであるか確認
○　犯罪者を追跡するために必要な不正プログラムの接続先等を確認　等

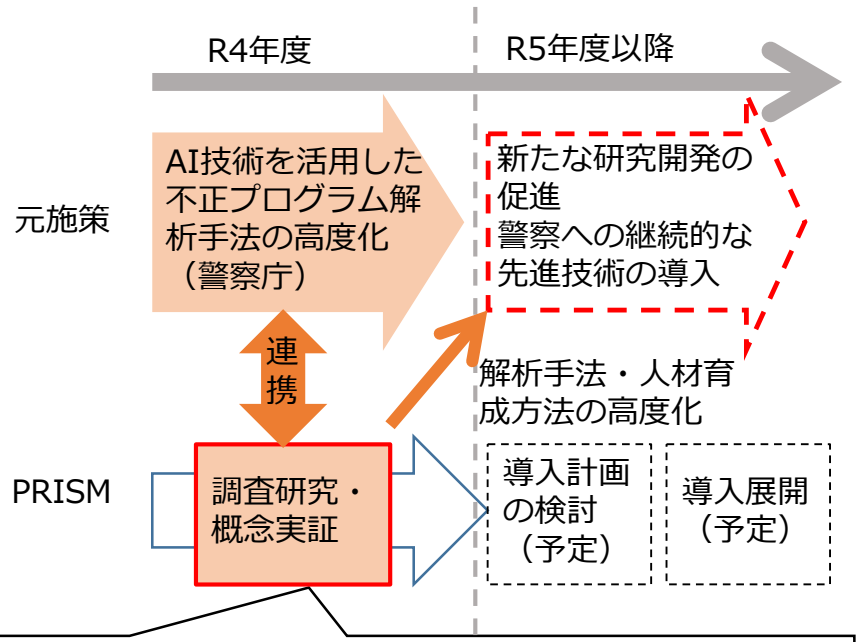
防犯対策等への活用
　犯行手口や犯罪者像の分析及び、分析結果の海外法執行機関等との情報交換により、未知の犯罪者固有の属性情報を収集。
　判明したサイバー事案の手口に係る情報を、重要インフラ事業者等に対する注意喚起や助言に活用。

課題：不正プログラムの難読化の進行
　　詳細解析に要する期間の長期化

【PRISM】
AI技術の活用等により、我が国の技術を標的とするサイバー攻撃で利用される不正プログラムの分類や機能特定の自動化をはじめとする各種不正プログラム検体の動作解明プロセスを効率化・高度化する機能の実装化を促す。



【開発のイメージ】



①　既存技術・ツールの文献調査
- 不正プログラム解析に関する文献等をAI及び不正プログラムの知見に基づき体系的に整理

②　現状把握及び課題の整理
- 様々な技術・ツールの実用可能性及び警察における解析に導入した場合の効果を検証

③　解析手法の検討及びロードマップの作成
- 短期的な警察における解析の効率化、及び目標達成に向けた研究開発を含む中長期的なロードマップの作成

資料3　「AI技術を活用した不正プログラム解析手法の高度化」の目標達成状況

○施策全体の目標　AI技術の活用等により、我が国の技術を標的とするサイバー攻撃で利用される不正プログラムの分類や機能特定の自動化をはじめとする各種不正プログラム検体の動作解明プロセスを効率化・高度化する機能の実装化を促す。

事業名等（※個別に目標を設定している場合）	当年度目標	目標の達成状況
既存技術・ツールの文献調査	• 不正プログラム検体の動作解明作業に関する既存研究等を調査し、類似不正プログラムの分類、難読化解除・機能特定・通信先特定の自動化、攻撃者属性の推定を実現するに当たって有用な技術を体系的に整理 • 省力化及び高度化に有効な機能や手法、ツール等の不正プログラムの解析に有用である技術の調査を実施	• 有識者に対するヒアリングを実施。 • 解析を省力化する先端技術や手法の調査を実施。 ➢ 製品：製品の特長、運用コスト等 ➢ 論文：手法の効果や実装コスト、運用コスト等 • 既存技術に関する調査結果の取りまとめを実施。
現状把握及び課題の整理	• 警察で現在実施している不正プログラム解析業務の手順や手法について課題を整理 • 警察が継続的に先端技術を導入する際の課題（官民連携の在り方、警察が備えるべき人材等）を整理	• 警察庁及び民間企業における解析実施状況についてヒアリングを実施。 • 官民連携の在り方、人材育成方針等を整理。 • 民間と連携した技術導入のアプローチを検討。人材育成方針及びスキルマップを策定。
解析手法の検討及びロードマップの作成	• 解析業務を高度化する手法や機能について、その実現可能性や警察に導入した際の効果を検証 • 解析業務の高度化・効率化に向け、直近で導入すべき技術に係る方針及び継続的な先端技術の導入に係るロードマップを策定	• 評価方法を策定。 • 実検体を使用した実証実験及び効果の検証を実施。 • 短期的な警察における解析の効率化、及び目標達成に向けた研究開発を含む中長期的なロードマップを策定。

資料4　「AI技術を活用した不正プログラム解析手法の高度化」の成果

○　既存技術・ツールの文献調査
各項目ごとに調査対象とする製品・ツール及び論文をリストアップし、情報収集、試用を通じて特長や運用コスト等を調査。

製品・ツール及び論文の例	
類似不正プログラムの分類	・局所性鋭敏型ハッシュを算出するツール ・マルウェアデータセット及びそれを用いて学習されたLightGBMによる判定モデル ・特定の不正プログラムに特有のバイト列を再利用している不正プログラムを検出する手法
難読化解除の自動化	・メモリ書き込み可能と実行可能のフラグを基にメモリをダンプしてアンパックするツール ・動的解析に使用するサンドボックスで、メモリ内の実行可能モジュールやシェルコードの抽出を可能とする手法
機能特定の自動化	・対象ファイルの機能(C2サーバーとの通信、ファイルの読み書き等)を自動的に判別するツール ・バイナリ解析用のシンボリック実行プラットフォームを用いて、耐解析を回避する条件を抽出する手法
通信先特定の自動化	・サンドボックスにより動的解析し、メモリイメージから既知の不正プログラムの設定情報を自動で取得するプラグイン ・インターネットアクセスを模擬的に実行するツール ・暗号化されたマルウェアのトラフィックを機械学習で検出する手法
攻撃者属性の推定	・サイバー脅威情報を集約し、様々な情報源を横断的に検索できるWebベースのツール ・攻撃手法をMITRE ATT&CKへマッピングするための抽出を自動化するWebベースのツール
解析の省力化	・不正プログラムが解析環境を検知する要素を削除するツール ・不正プログラムによるサンドボックス検知を回避できるツール ・耐解析技術が実装されたマルウェアを解析するためのオープンソースのフレームワーク

➡ 既存のツール・手法を組み合わせて活用することで一定の高度化、効率化が見込まれるが、
実用化までは至っていない論文も数多く存在。

○　現状把握及び課題の整理
ヒアリング等の結果を踏まえ現状の課題及び在り方を整理

(1) 官民連携の在り方
AI 技術をはじめとする先端技術を民間と協力して導入するためのアプローチとして、次の２案を提示。
・警察が提示した先端技術で解決すべき課題を民間で技術開発
・民間が持つ先端技術を警察で活用可能とする仕組みづくり

(2) 人材育成の在り方
・警察が抱える課題を整理し、人材育成のアウトラインを作成
・スキルマップの策定
＞SecBookを基にしたスキルの定義

○　解析手法の検討及びロードマップの作成

1　実証実験、効果の検証
(1) 評価設計、マルウェア選定／環境構築、効果測定を実施
(2) 検証対象の検体を使用して実機による検証を実施
→正確性、合目的性、網羅性、資源効率性、時間効率性、習得性、操作性、安定性、導入性の観点で評価
(3) 従来手順と効率化済み手順の比較・評価を実施

2　ロードマップの策定
短期と中長期に分けたロードマップを策定

資料5　「AI技術を活用した不正プログラム解析手法の高度化」の民間からの貢献及び出口の実績

○民間からの貢献額：５千万円相当
（内訳）機械学習の精度を向上させるために必要な不正プログラムやIoC情報等データセットの提供（５千万円／年相当）

当年度当初見込み	当年度実績
・データセットの提供：年間当たり５千万円相当	・民間企業が独自に収集した検体(約700万件所有)や動的解析ログ、表層解析情報を提供 ・事業実施期間において民間の分析システム・サービスの使用権及び構築済み環境を提供

○出口戦略
・不正プログラム解析技術を警察において迅速に導入することで、サイバー事案の予防、サイバー空間の安全・安心の確保を実現
・攻撃者の公表・非難を行う「パブリックアトリビューション」に寄与し、サイバー攻撃抑止に貢献
・サイバー事案の手口に係る情報についての警察からの注意喚起に活用
・官民連携の枠組みにより不正プログラム解析の継続的な高度化を実現

当年度当初見込み	当年度実績
（民間技術の継続的な導入） ・警察が継続的に先端技術を導入していくに当たり、官民連携の在り方や警察が備えるべき人材の能力等の課題を整理し、実現に向けたロードマップを作成する。	（民間技術の継続的な導入） 民間企業と警察それぞれの強み等を比較し、警察が継続的に導入していくに当たり、官民連携の在り方や警察が備えるべき人材の能力等の課題を整理するとともに、検証結果を踏まえ、効果が見込まれる手法・ツールの導入方針、長期的な解析システムの構築や警察において導入し続けるための仕組みづくり等を含めたロードマップを作成。
（アトリビューションの実現） ・不正プログラム解析に関する既存技術を調査・整理し、不正プログラムの分類や各種プロセスの自動化の実現可能性や警察へ導入した際の効果等を検討する。	（アトリビューションの実現） 不正プログラム解析に関する既存技術を調査・整理し、机上検証や実機検証を通じて、不正プログラムの分類や各種プロセスの自動化の実現可能性や警察へ導入した際の効果等について、正確性や費用対効果、今後研究を進めていくべき分野等を検討。