博士請求論文審査要旨

情報セキュリティ大学院大学
情報セキュリティ研究科

論文題目 ： Semantic Code Learning for Binary Analysis: Applications to Malware Classification and Code Understanding

申請者 ： 染谷 実奈美

審査委員会 ： 主査　教授　大塚 玲
副査　教授　有田 正剛
副査　教授　吉岡 克成（横浜国立大学）
副査　教授　稲葉 緑

Ⅰ. 論文内容の要旨

本論文は“Semantic Code Learning for Binary Analysis: Applications to Malware Classification and Code Understanding”と題し、7 章からなっている。

第 1 章では、研究の概要として、攻撃者と防御者の非対称性を指摘し、機械語バイナリを解析するサイバーセキュリティの課題と、それに対する本論文の機械学習的解決策の概観が記述されている。

第 2 章では、バイナリ解析、機械学習手法、およびマルウェア分類とコード理解に関する関連研究を概観している。まず、マルウェアのバイナリ解析に関する研究がバイト列の直接的な解析からオペコード列、制御フローグラフ、関数呼び出しグラフへと発展するにつれ高精度化してきた技術発展の経緯を解説している。また、LLM (大規模言語モデル)のコード理解能力に関する Shang らの驚くべき結果をはじめとする近年の LLM を用いたバイナリ解析の動向を述べた上で、実務に利用するためには解釈可能性の課題、LLM 能力と情報漏洩の課題、マルウェア学習データ不足の課題の 3 つの解決が必要であると述べている。

第 3 章では、グラフニューラルネットワークと大規模言語モデルを含む基本概念を定義している。まず、DNN(深層ニューラルネットワーク)、CNN(畳み込みネットワーク）、RNN(再帰ネットワーク）、Transformer を概観し、残差接続や正則化で勾配消失や過学習を抑えることができることを解説した上で、バイナリ解析の要となるグラフ表現を取り上げ、GNN(グラフニューラルネットワーク）が埋め込み空間でグラフ構造の分類に適したグラフの埋め込み表現を獲得し、続く注意機構により各グラフ要素が判定に貢献した度合いを可視化できること、さらに、Transformer のマルチヘッド注意機構が長距離依存を効率的に扱い、プログラムコードにおける変数スコープや制御フローを捉える仕組みを述べている。加えて、LLM が巨大コーパスで学習する過程で推論・要約などの能力を創発し、スケーリング則がモデル規模とデータ量の関係を定量化することを紹介し、クラウド依存の LLM の一部の能力を小規模言語モデルに転移させる知識蒸留の概念を解説している。

第 4 章では 、コアアイディアの一つとして、FCGAT (Function Call Graph with Attention)のアーキテクチャ、学習手法、および実際のマルウェアファミリーに対するケーススタディを通じた解釈可能性の解析結果を述べている。FCGAT は、関数呼び出しグラフを入力として、GNN と注意機構により、マルウェア分類の判定結果を出力し、さらに、判定の際に重視された上位の関数を自動で抽出できるため、アナリストのリバースエンジニアリングに有効なツールと成り得る。MalwareBazaar、 Microsoft BIG-2015 の公開データを用いた評価で、それぞれ 98.15 %、 98.18 %の F1 スコアを達成し、GNN を含む全ての既存手法を上回る高性能が得られたと述べている。

第 5 章では、本研究における 2 つ目のコアアイディアとして、バイナリ解析で失われた関数名を自動復元するための言語モデル「RevLlama」を提案し、その設計思想と性能評価の結果を述べている。RevLlama では、段階的蒸留法(Distillation Step by Step)を応用し、主にクラウドで利用される LLM のバイナリ解析能力をローカルで実行可能な言語モデルに追加学習(fine-tuning)で転移させることで実現している。実験結果は、関数名推定タスク等において RevLlama が大規模言語モデルを上回る性能を有することを示している。

第 6 章では、2 つめのコアアイディアにおけるマルウェア学習データの入手困難性の課題に対応して、擬似マルウェアコードの合成による、学習データの拡張手法について述べている。本章では、MITRE ATT&CK 等に記載された攻撃技法を入力として、LLM に擬似マルウェアコードを生成させた後、構造リファクタリング、API／識別子多様化、難読化、コード変換、解析回避の注入などでコードを多様化することで、関数データセットの構築法を提案している。この学習データを小規模な言語モデルに適用することで、関数名推定タスク等において最新の LLM の性能に迫る実験結果が得られることが示されている。

第 7 章では、結論として、強力な AI 技術とバイナリ解析の実務的要請との間に存在するギャップを埋める重要な一歩を示したとしている。解釈可能性、効率性、ドメイン適合性を重視することで、本論文は高精度であるだけでなく、信頼性が高く、実際のサイバーセキュリティ運用に展開可能な次世代リバースエンジニアリングツールの礎を築いたと結論づけている。

Ⅱ. 論文審査結果の要旨

本論文は、サイバーセキュリティにおける重要課題の一つであるマルウェアのバイナリ解析に関し、発展が著しい機械学習技術を効果的に活用するための、具体的かつ基盤的な方法を示したものであって、情報学ならびに情報社会の発展に貢献するところが大きい。よって、本論文は情報学における博士論文として十分価値のあるものと認める。

Ⅲ. 審査経過

本審査委員会は、2025 年 8 月 6 日に論文内容とこれに関連する事項について口述試問を行い、その後、2025 年 8 月 27 日に関連する事項の最終試験審査を行った。審査に当たっては、博士学位のディプロマ・ポリシーに基づいて総合的に評価し、申請者が学位取得にふさわしい知見を持つものと判断した。