大規模言語モデルによるメモリフォレンジクスに関する研究
Leveraging Large Language Models for Automated Memory Forensics

曾浩然・システム分科会・情報セキュリティ大学院大学

研究背景

ファイルレスマルウェアは、従来のディスク解析を回避し、主にメモリ上で動作する特性を有する。そのため、揮発性データから証拠を抽出するメモリフォレンジックの重要性が高まっている。近年では、LLMを活用し解析を自動化する試みも進められているが、現状ではツール出力の受動的解釈にとどまっている。また、高い誤検知率が依然として課題であり、フォレンジック結果の信頼性を低下させる要因となっている。

研究目的

本研究の目的は、Volatilityフレームワークと統合した自律型LLMエージェントを構築し、メモリフォレンジックの一連のプロセスを自動化することである。LLMが能動的にフォレンジックツールを制御・判断できる仕組みを実現することで、分析効率の向上を図るとともに、最大の課題である誤検知率の削減を目指す。さらに、調査過程における高い説明可能性を維持することも重要な目標とする。

提案手法

本研究では、強化学習およびQLoRAを用いてLLMをファインチューニングし、ツール利用能力の向上を図る。具体的には、出力フォーマットの正確性やツール呼び出しの適切性に基づく詳細な報酬設計を行い、段階的にヒントを削減するカリキュラム学習を導入する。また、領域知識の補強および説明可能性の向上を目的として、RAGとの統合も検討する。さらに、学習データが少数である点やVolatilityの実行時間が長いという実用上の制約に対しては、データ拡張の導入および同一指示に対する実行結果のキャッシュ化を行い、学習効率の最適化を図る。

今後の予定

今後は、まずContext Engineering手法の詳細な調査を実施する。その上で、提案手法への適用可能性を検証し、自律型エージェント構築に伴う具体的な技術的課題について体系的に整理・検討を進める予定である。