様式3

論 文 内 容 の 要 旨

氏 名 （ 芝原 俊樹 ）	
論文題名	Robust Malicious Communication Detection for Multi-layered Defense by Overcoming Anti-analysis Techniques （解析妨害機能への対策による多層防御のための堅牢な悪性通信検知）

論文内容の要旨

Cybersecurity is actively studied because cyberattacks can damage enterprise reputations and cause enormous financial losses. Multi-layered defenses, which are combinations of various detection systems to improve detection capabilities, are generally deployed to mitigate diverse and dynamically changing attacks. Detection systems can be roughly categorized according to two perspectives: detection targets or input data. The systems can be divided according to detection targets into those for pre-infection and post-infection detection. The systems can be further divided according to input data into those for network-based and host-based detection.

In this thesis, we focus on network-based detection systems, because they are effective for both pre-infection and post-infection detection. Effective network-based detection requires exhaustive beforehand collection of malicious communications, which has become difficult because attackers are employing anti-analysis techniques. Specifically, attackers use cloaking to conceal malicious websites and use environment-aware malware to conceal communications between malware samples and attackers. In this thesis, we propose systems and a method for overcoming such anti-analysis techniques.

First, we propose a system for detecting malicious websites without collecting all malicious data. Compromised websites have similar traits because attackers use search engines to automatically discover vulnerable websites. We therefore build a classifier by leveraging both malicious and compromised websites. The proposed system detects 143 more malicious websites employing anti-analysis techniques than does a conventional system. This system enhances exhaustiveness of collected malicious websites and improves detection capabilities for network-based pre-infection detection.

Next, we propose a system for detecting communications to malicious websites from simple logs such as proxy logs. We focus on sequences of destination URLs, because some artifacts of malicious redirections can be extracted from simple logs by considering several nearby URLs. We compare three approaches for classifying URL sequences: an individual-based approach, a convolutional neural network (CNN), and a novel event de-noising CNN (EDCNN). Evaluation results show that only our EDCNN achieves practical classification performance, a true positive rate (TPR) of 99.1% and a false positive rate of 3.4%. Using detected malicious communications, we can improve capabilities for network-based pre-infection detection.

Then, we propose a system for efficiently collecting HTTP requests with dynamic malware analysis. Specifically, our system analyzes a malware sample over a short period, then determines whether analysis should be continued or suspended. In the proposed system, we apply a recursive neural network, which has recently exhibited high classification performance in the field of natural language processing. In an evaluation with 42,856 malware samples, our proposed system collects 94% of novel HTTP requests and reduces analysis time by 82% in comparison with a system that continues all analyses. We can improve network-based post-infection detection by using the collected HTTP requests.

We also propose a method for detecting dynamically changing attacks even when some malicious communications cannot be collected by anti-analysis techniques. Specifically, we investigate how to improve existing deep neural network-based systems in terms of detecting unknown families, namely new types of malicious websites or malware samples. We focus on the tendency that some features are inherent across different families because malicious data include similar code or produce similar behaviors to exploit vulnerabilities or to cost-effectively achieve a successful attack. Therefore, we build a classifier that prioritizes family-invariant features. Our evaluation results show that our method outperforms conventional optimization methods, which optimizes a classifier only so that it accurately classifies malicious and benign data, by at most 19%, 19%, and 7% in terms of TPR for malicious websites, Android applications, and PE files, respectively. This method can improve network-based detection and compensate for degradation caused by anti-analysis techniques.

様式 7

論文審査の結果の要旨及び担当者

氏　名（　芝原　俊樹　）		
	（職）	氏　名
論文審査担当者		
主 査	教授	村田 正幸
副 査	教授	渡辺 尚
副 査	教授	長谷川 亨
副 査	教授	東野 輝夫
副 査	教授	松岡 茂登

論文審査の結果の要旨

多様かつ動的に変化するサイバー攻撃による被害を低減するために、多くの企業では多層防御と呼ばれる複数の検知システムを組み合わせたサイバー攻撃対策システムが用いられている。多層防御で用いられている検知システムは、検知対象の違いから、マルウェア感染の兆候を検知する事前対策と、感染後のマルウェアの挙動を検知する事後対策に分類される。さらに、検知システムに入力されるデータの違いから、ネットワーク上の通信を用いるネットワークベース検知と、ホスト上のAPIコールやレジストリ操作を用いるホストベース検知に分類される。本論文では、これらのシステムの中で、事前対策と事後対策の両方に有効なネットワークベース検知に着目している。

ネットワークベース検知では、既知の悪性通信と類似する通信を検知するため、高精度な検知システムを構築するためには、事前に悪性通信を網羅的に収集する必要がある。事前対策のためには、ハニークライアントと呼ばれる囮のウェブブラウザを用いて大量のウェブサイトの中から悪性サイトを特定し、悪性サイトにアクセスした際の通信を悪性通信として収集している。事後対策のためには、サンドボックスと呼ばれる安全な環境で実際にマルウェアを動作させ、その際のマルウェアから送信された通信を悪性通信として収集している。これらに加えて、多層防御を導入しているネットワークの通信ログを記録し、その中から見逃された悪性通信を特定することで、悪性通信を収集する取り組みも行われている。しかし、攻撃者が開発した解析妨害機能によって、悪性通信の収集は困難になっている。具体的には、攻撃者は、悪性サイトをクローキングによって隠蔽し、マルウェアと攻撃者との通信を環境依存マルウェアによって隠蔽している。

本論文の研究成果は、解析妨害機能への対策によって堅牢なネットワークベース検知の実現を可能としたことである。まず、クローキングによってハニークライアントによるアクセスで悪性サイトの一部のコンテンツが収集できなくても悪性サイトを高精度に特定するために、攻撃者に改ざんされて攻撃に悪用されるWebサイトは常に観測できることに着目したシステムを提案した。また、プロキシログに含まれている見逃された悪性サイトへの通信を特定するために、悪性なリダイレクト関係のURLの特徴を効果的に学習できる新たな畳み込みニューラルネットワークに基づくシステムを提案した。次に、環境依存マルウェアから効率的に悪性通信を収集するために、マルウェアごとに新規の悪性通信を送信する可能性を予測し、可能性が高いマルウェアをサンドボックスで長時間解析するシステムを提案した。これらの提案システムでは、従来システムより網羅的に悪性通信を収集することができるため、ネットワークベース検知の精度を向上させることができる。さらに、収集された悪性通信が網羅的でない場合も考慮し、未知のマルウェアファミリに対する深層学習の検知精度を向上させる最適化手法も提案している。この提案手法によって、悪性通信が網羅的に収集できていなくても、従来のネットワークベース検知より高い検知精度を達成することができる。

以上のように本論文は、解析妨害機能の影響の最小化による悪性通信収集の網羅性向上と、収集された通信が網羅的でない条件でのネットワークベース検知の検知精度向上を達成することで、堅牢なネットワークベース検知を実現するために有用な研究成果を上げている。よって、博士（情報科学）の学位論文として価値のあるものと認める。