electronics

IMPACT FACTOR 2.690

CITESCORE 3.7

Editorial

High Accuracy Detection of Mobile Malware Using Machine Learning

Suleiman Y. Yerima

Special Issue
High Accuracy Detection of Mobile Malware Using Machine Learning

Edited by
Dr. Suleiman Yerima

MDPI

https://doi.org/10.3390/electronics12061408 *electronics*

MDPI

Editorial

High Accuracy Detection of Mobile Malware Using Machine Learning

Suleiman Y. Yerima

Cyber Technology Institute, School of Computer Science and Informatics, De Montfort University,
The Gateway, Leicester LE1 9BH, UK; syerima@dmu.ac.uk

Introduction

As smartphones and other mobile and IoT devices have become pervasive in everyday life, malicious software (malware) authors are increasingly targeting the operating systems that are at the core of these mobile systems. Malware targeting mobile platforms has witnessed an explosive growth in the last decade. As a result of this rapid increase in mobile malware, the limits of traditional signature-based antivirus scanning have been stretched. This has led to the emergence of machine learning-based detection as a complementary solution to traditional antivirus scanning. Although machine learning-based malware detection has continued to attract great research interest, many challenges remain as emerging malware families continue to evolve with more sophisticated capabilities and stealthy evasive techniques.

This Special Issue in Electronics presents some of the most recent research results and innovative machine learning-based approaches to detecting malicious software and attacks that can compromise mobile platforms.

The authors of [1] proposed a novel Android botnet detection system based on image features and manifest file features. The method aims to overcome the limitations of hand-crafted features for machine learning-based botnet detection. Their proposed approach employs Histogram of Oriented Gradients, together with byte histograms obtained from images representing the app executables, and these are subsequently combined with the features derived from manifest files. The proposed system was evaluated using the ISCX botnet dataset, and the experimental results demonstrate its effectiveness with F1 scores ranging from 0.923 to 0.96 using popular machine learning algorithms.

In [2], the authors present a study on generating malware adversarial samples using deep learning models. Gradient-based methods are usually employed in generating adversarial samples; however, they generate the samples on a case-by-case basis, which is very time-consuming for large scale sample generation. To address this issue, a novel method was proposed, which extracts feature byte sequences from benign samples using deep learning. Feature byte sequences represent the characteristics of benign samples and can affect classification decisions. The feature byte sequences are directly injected into malware samples to generate adversarial samples. The proposed method is compared with random injection and gradient-based methods, and the experimental results show that the new method is suitable for generating a large number of adversarial samples.

The authors of [3] propose an ensemble classification-based approach for malware detection. The first-stage classification is performed by a stacked ensemble of dense (fully connected) and convolutional neural networks (CNN), while the final stage classification is performed by a meta-learner. For the meta-learner, 14 classifiers are explored and compared. For baseline comparison, 13 machine learning methods are used: K-Nearest Neighbors, Linear Support Vector Machine (SVM), Radial basis function (RBF) SVM, Random Forest, AdaBoost, Decision Tree, ExtraTrees, Linear Discriminant Analysis, Logistic, Neural Net, Passive Classifier, Ridge Classifier and Stochastic Gradient Descent classifier. The results of experiments performed on the Classification of Malware with PE headers (ClaMP) dataset

check for updates

Citation: Yerima, S.Y. High Accuracy Detection of Mobile Malware Using Machine Learning. *Electronics* **2023**, *12*, 1408. https://doi.org/10.3390/electronics12061408

Received: 12 March 2023
Accepted: 14 March 2023
Published: 15 March 2023

are presented. The best performance is achieved by an ensemble of five dense and CNN neural networks, and the ExtraTrees classifier as a meta-learner.

In [4], the authors presented a comparative study of deep learning techniques with the aim of investigating their efficacy for Android botnet detection, based on static features. To create the deep learning-based botnet detection system, a bespoke tool for automated reverse engineering of Android Package Files (APKs) was developed and used to extract 342 features, which are then used to represent the application as a vector of binary vectors. These vectors were used to train several deep learning models including: Convolutional Neural Networks (CNN), Dense Neural Networks (DNN), Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), as well as more complex networks like CNN-LSTM and CNN-GRU. Evaluation experiments were conducted using 6802 Android applications out of which 1920 were botnet samples from the ISCX botnet dataset. The results showed that the deep-learning models outperformed classical machine learning classifiers and achieved very high accuracy, as well as high precision, recall and F1 scores.

The authors of [5] investigated the relevance of the features of unpacked malicious and benign executables such as mnemonics, instruction opcodes, and API calls to identify a feature that classifies the executable. Prominent features were extracted using Minimum Redundancy and Maximum Relevance (mRMR) and Analysis of Variance (ANOVA). Experiments were conducted on four datasets using machine learning approaches such as Support Vector Machine (SVM), Naïve Bayes, J48, Random Forest (RF), and XGBoost. In addition, they evaluated the performance of deep neural networks such as Deep Dense Network (DDN), One-Dimensional Convolutional Neural Network (1D-CNN), and CNN-LSTM in classifying unknown samples, and observed promising results using APIs and system calls. On combining APIs/system calls with static features, a marginal performance improvement was attained compared to models trained only on dynamic features. Moreover, to improve accuracy, the solution was implemented using distinct deep learning methods and demonstrated a fine-tuned deep neural network that resulted in an F1-score of 99.1% and 98.48% on Dataset-2 and Dataset-3, respectively.

In [6], the authors presented an approach called eRBCM to detect malware. The eRBCM system was designed using the reinforcement learning approach, which utilizes the strength of Monte–Carlo simulations and builds a strong machine learning model to detect complex malware patterns. It combines the most beneficial elements of MOCART's reinforcement learning and RF's exploration capabilities. A large number of experiments were conducted using different malware benchmarks, including ARP attack, ICMP attack, and Microsoft Malware. eRBCM was consistently better than its competitors in terms of learning the new malware patterns and detecting unknown malware. This was mainly explained by eRBCM's self-adaptability to exploration and intelligent tuning of the balance for the trade-off between exploration and exploitation.

The authors of [7] present a study on detecting drive-by exploits in images using deep learning. With steganographic techniques being combined with polyglot attacks to deliver exploits in web browsers, machine learning approaches have been proposed for detecting steganography in images. However, exploit code hiding has not been systematically addressed; hence the paper proposes the use of deep learning methods for such detection, accounting for the specifics of the situation in which the images and the malicious content are delivered using Spatial and Frequency Domain Steganography algorithms. The methods were evaluated by using benchmark image databases with collections of JavaScript exploits, for different density levels and steganographic techniques in images. A convolutional neural network was built to classify the infected images with a validation accuracy around 98.61% and a validation AUC score of 99.75%.

In [8], the authors propose a Salp Swarm Algorithm (SSA) as a trainer for Multilayer perceptron (MLP) in the context of digital forensics. SSA is an effective meta-heuristic algorithm that belongs to the swarm-based family. It has a single parameter that decreases in an adaptive manner relative to increasing iteration. It also performs an extensive exploration in the initial iterations and then adaptively switches to exploit the most promising areas of the search space. Furthermore, SSA also preserves the best-found solution so that it never loses the optimal solution. Lastly, follower salps change their locations adaptively following other members of the population, so it has the power to alleviate the local minima problem. In this paper, seven metaheuristic algorithms are compared to the proposed approach: Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Genetics Algorithm (GA), Differential Evolution algorithm (DE), and BackPropagation. In the majority of cases, the SSA-based MLP outperformed the other approaches when evaluated on the digital forensics dataset created from audit logs, registry, and file system.

The authors of [9] provide a systematic review of machine learning-based Android malware detection techniques. This paper aims to enable researchers to acquire in-depth knowledge in the field and to identify potential future research and development directions. The paper critically evaluates 106 carefully selected articles and highlights their strengths and weaknesses as well as potential improvements. Finally, the machine learning-based methods for detecting source code vulnerabilities are discussed, because it might be more difficult to add security after the app is deployed.

In [10], the authors present a systematic literature review and examination of the state of the art of Business Email Compromise (BEC) phishing detection techniques with the aim of providing a detailed understanding of the topic to allow researchers to identify the main principles of BEC phishing detection. Based on a selected search strategy, 38 articles (of 950 articles) were chosen for closer examination. The selected articles were discussed and summarized to highlight their contributions as well as their limitations. In addition, the features of BEC phishing used for detection were provided, and the ML algorithms and datasets that were used in BEC phishing detection models were discussed. In the end, open issues and future research directions of ML-based BEC phishing detection were also discussed.

Acknowledgments: I would like to thank all the authors for the papers they submitted to this Special Issue. I would also like to acknowledge all the reviewers for their careful and timely reviews to help improve the quality of this Special Issue.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Yerima, S.; Bashar, A. A Novel Android Botnet Detection System Using Image-Based and Manifest File Features. *Electronics* **2022**, *11*, 486. [CrossRef]
2. Ding, Y.; Shao, M.; Nie, C.; Fu, K. An Efficient Method for Generating Adversarial Malware Samples. *Electronics* **2022**, *11*, 154. [CrossRef]
3. Damaševičius, R.; Venčkauskas, A.; Toldinas, J.; Grigaliūnas, Š. Ensemble-Based Classification Using Neural Networks and Machine Learning Models for Windows PE Malware Detection. *Electronics* **2021**, *10*, 485. [CrossRef]
4. Yerima, S.; Alzaylaee, M.; Shajan, A.; Vinod, P. Deep Learning Techniques for Android Botnet Detection. *Electronics* **2021**, *10*, 519. [CrossRef]
5. Ashik, M.; Jyothish, A.; Anandaram, S.; Vinod, P.; Mercaldo, F.; Martinelli, F.; Santone, A. Detection of Malicious Software by Analyzing Distinct Artifacts Using Machine Learning and Deep Learning Algorithms. *Electronics* **2021**, *10*, 1694. [CrossRef]
6. Alrammal, M.; Naveed, M.; Tsaramirsis, G. A Novel Monte-Carlo Simulation-Based Model for Malware Detection (eRBCM). *Electronics* **2021**, *10*, 2881. [CrossRef]
7. Iglesias, P.; Sicilia, M.; García-Barriocanal, E. Detecting Browser Drive-By Exploits in Images Using Deep Learning. *Electronics* **2023**, *12*, 473. [CrossRef]
8. Alazab, M.; Khurma, R.A.; Awajan, A.; Wedyan, M. Digital Forensics Classification Based on a Hybrid Neural Network and the Salp Swarm Algorithm. *Electronics* **2022**, *11*, 1903. [CrossRef]
9. Senanayake, J.; Kalutarage, H.; Al-Kadri, M. Android Mobile Malware Detection Using Machine Learning: A Systematic Review. *Electronics* **2021**, *10*, 1606. [CrossRef]
10. Atlam, H.; Oluwatimilehin, O. Business Email Compromise Phishing Detection Based on Machine Learning: A Systematic Literature Review. *Electronics* **2023**, *12*, 42. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.