RES MILITARIS Social Science Journal

ENHANCING MALWARE DETECTION WITH DEEP LEARNING FOR ROBUST INTELLIGENT SECURITY SOLUTIONS

[1]RANADHEER REDDY VALLEM,[2]NAKKA RAVI,[3]NOOKA KUMAR SUNDEEP

[123]Assistant Professor

Department of CSE

Vaagdevi Engineering College, Bollikunta, Khila Warangal, Warangal, Telangana, India

ABSTRACT: "Malware or malware remains a major concern in present digital life as computer users, companies and governments see the rise of malware attacks. Current malware detection solutions are use static and dynamic analysis of malware signatures and behaviours; this is time consuming and ineffective in identifying unknown malware. The latest malware uses polymorphism, shapeshifting, and other avoidance methods to quickly change the behavior of the malware and create more malware. Recently, machine learning algorithms (MLA) have been used to effectively identify malware because new malware is often different from existing malware." This requires a lot of engineering skills, technical training and artistic expression. The engineering process can be avoided altogether by using advanced MLA techniques such as deep learning. Although some recent studies have progressed in this direction, the efficiency of algorithms is highlighted by the data.

"In order for newly developed methods to be effective in zero-day malware detection, it is necessary to reduce the bias and self-testing of these methods. To fill this gap in the literature, this study evaluates the role of classical MLA and deep learning in malware detection, classification and classification using public and private databases. Training and testing distinguish between public and private data used in clinical trials and collected at different times. We also provide a new image processing system with state-of-the-art MLA and in-depth courses." A qualitative analysis of the method shows that deep learning outperforms traditional MLA. Overall, this project uses deep learning techniques to classify malware in real time and provides powerful insights. Visualization and deep learning, based on a combination of static, dynamic and image processing in a big data environment, is a new method for zero-day malware discovery.

1. INTRODUCTION

"With the rise of malware threats, it is more necessary than ever to protect our computers and mobile phones with antivirus software. Machine learning is an important technique for malware detection. It's been trained with millions of samples, so even when new types of malwares are discovered, it can learn something to recognize them at scale. It is artificial intelligence that can be used to detect malware. It works by extracting information from files and comparing them to known malware names.

It also includes scanning all or part of the system, removing malware signatures, comparing signatures to known behaviours, and detecting malware. The battle between malware developers and censors is fierce. Both the research and hacker community do the same; one creates malware detection systems, while the other creates malware that will attack computers and network resources. The Malware Checker detects malware and attempts to detect it before the user's computer is infected. This disease has many effects, including:"

- Illegal information.
- The file size has changed.
- Erases all content on the DVD.
- Partition table corrupted, data on disk becomes unreadable.

• Various issues but terrible video/sound effects

2. LITERATURE REVIEW

Robust Intelligent Malware Detetion Using Deep Learning

VINAYAKUMAR R1 , MAMOUN ALAZAB2 , (Senior Member, IEEE) , SOMAN KP1 , PRABAHARAN

Malware or malware remains a major concern in this digital life as computer users, companies and governments see the rise of malware attacks. Current malware detection solutions use static and dynamic analysis of malware signatures and behavior; this is time consuming and ineffective in identifying unknown malware. Recent malware uses evasion techniques such as polymorphism and metamorphism to rapidly change malware behavior and create more malware. Recently, machine learning algorithms (MLA) have been used to effectively identify malware because new malware is often different from existing malware. This requires a lot of engineering skills, technical training and artistic expression. The level of feature engineering can be avoided entirely by using advanced MLA methods such as deep learning. Although some recent studies have moved in this direction, the effectiveness of the algorithm is geared towards the data presented. To achieve new development methods for effective zero-day malware detection, there is a need to reduce the bias and self-testing of these methods. To fill this gap in the literature, this study evaluates classical MLA and deep learning for malware detection, classification and classification using public and private data. The training and separate testing of public and private data used in clinical trials are separated and collected at different times.

Additionally, we are introducing a new image processing system with views best suited for both MLA and deep learning. A qualitative analysis of this method shows that deep learning outperforms traditional MLA. Overall, this work presents a powerful visualization of malware in real time using scalable hybrid deep learning techniques. Visualization and deep learning as a combination of static, dynamic and image processing in the big data environment is a new development method for zero-day malware detection. A robust intelligent zero-day cyber-attack detection technique

Vikash Kumar; Ditipriya Sinha

With the introduction of the mainstream internet, such as e-commerce, online business, healthcare and other everyday products, exposure to various risks has increased exponentially. Zero-day attacks on unknown vulnerabilities in software or systems drive further research in the field of cyber-attacks. Current methods use machine learning/DNN or weak algorithms to prevent these attacks. With this strategy, the detection of zero-day attacks misses many parameters such as the frequency of the byte stream in network traffic and their relationship. Low-traffic attacks are difficult to cover by neural network models because they require more traffic to make accurate predictions. This article presents a new robust and intelligent network attack detection model to detect the above issues, using the context and mapping techniques of heavy hitters to detect zero-day attacks. The working strategy consists of two phases (a) signature generation and (b) evaluation phase. The model uses signatures created during training to evaluate performance. Analysis of the results of the proposed zero-day stop search shows that the binary distribution has an accuracy of 91.33% and an accuracy of 90%.35% for multiclass classification of real attack data. The performance of the CICIDS18 benchmark data shows an efficiency of 91.62% for the model for binary class classification. So, the plan shows that it supports the results in the zero-day attack analysis.

A Dynamic Robust DL-Based Model for Android Malware Detection
Ikram UlHaq;Tamim Ahmed Khan;AdnanAkhunzada
"The rise in Android-based smart devices has led to technological advances aimed at improving overall quality of life, making it a billion-dollar business. Despite the hype in the Android market, the prevalence and potential of malicious mobile malware has emerged as a threat to the popular Android platform and an ideal target for various cyber-attacks. In contrast, multi-vector malware is very difficult to detect in an efficient and timely manner because it is often hidden behind legitimate third-party software and can be easily spawned by any file extension. The authors propose a hybrid deep learning (DL)-based intelligent multi-vector malware detection mechanism to alleviate this most worrying issue. The proposed method uses non-linear communication and Bidirectional Short-Term Memory (BiLSTM) to detect persistent malware. The proposed system has been extensively evaluated using publicly available data, performance benchmarks, and state-of-the-art hybrid DL-driven architectures and benchmark DL algorithms. In addition, the proposed framework was cross-validated and performed well in terms of both time efficiency and detection accuracy."

A Robust Malware Detection System Using Deep Learning on API Calls
Liu, Yingying; Wang, Yiwei
"With the development of technology, malware has become a big problem for computer security nowadays. In our work, we use a malware detection engine that uses deep learning of API calls. From the Cuckoo sandbox we extract the so-called API as part of the malicious program. We filtered and analysed unfulfilled API calls to extract valid API sequences. We evaluate BLSTM on a large dataset of 21,378 samples compared to GRU, BGRU, LSTM and Simple RNN. Experimental results show that BLSTM has the best performance for malware detection with an accuracy of 97.85%."

Robust Malware Detection using Residual Attention Network
Shamika Ganesan; VinayakumarRavi; MoezKrichen;SowmyaV; RoobaeaAlroobaea; Soman KP
"Recent advances in cybersecurity combined with the power of artificial intelligence and human intelligence in intrusion detection. The massive increase in the amount of new malware created every day and the continued risk of zero-day attacks require research into the malware detection system. There has been a lot of research on machine learning and the use of Convolutional neural networks (CNNs). There is a shift from machine learning and deep learning based on using malware byte data to using image-based detection to better evaluate malicious data. While CNNs are effective at capturing local features, monitoring techniques play an important role in identifying patterns of variation in malware. In this article, we explore the use of residual attention for malware detection and compare it to existing CNN-based methods and traditional machine learning algorithms that use GIST features. The plan can provide clear "tracking" for parts of malware that are important in distinguishing good data, thereby mitigating the downside, which is the main point to get into a cybersecurity perspective. This technique is resistant to modification patterns in malware and achieves 99.25% accuracy, outperforming traditional malware detection techniques."

Intelligent Vision-Based Malware Detection and Classification Using Deep Random Forest Paradigm
Roseline, S. Abijah; Geetha, S.; Kadry, Seifedine; Nam, Yunyoung
The malware threat to modern computing is growing rapidly. Malware authors continue to use various techniques such as code obfuscation to create malware variants and evade detection of existing malware.

Classifying unknown malware variants with similar properties according to their families is a significant challenge, even when the classifier is trained with known variants of the same family. Identifying and extracting different features of each malware is another challenge in detecting malware systems. Features that enable classes' ability are difficult to develop by modifying them in every malware.
"Traditional malware detection methods use static signatures and behavior-based methods that are ineffective in identifying and identifying advanced and zero-day malware. To address these issues, this work uses a visual approach where malware is represented as 2D images and offers a powerful machine learning-based anti-malware system. The proposed method is based on a layered approach that follows the main features of deep learning but is more efficient. The proposed method does not require hyper parameter tuning or iteration and can reduce model complexity. The proposed model outperforms other state-of-the-art methods with an identification rate of 98%. Malimg, BIG 2015, and MaleVis malware datasets are 65%, 97.2%, and 97.43%, respectively. The results show that the solution is effective in detecting new and advanced malware due to its different features."

LAB to SOC: Robust Features for Dynamic Malware Detection
Rhode, Matilda; Tuson, Lewis; Burnap, Pete; Jones, Kevin
"Modern machine learning models achieve over 95% accuracy in case studies of dynamic malware detection issues, but the models that feed these models are rarely publicly available. Not only does this pose challenges for academics and business professionals, it doesn't specifically point out the machine learning model of data from a single source. This article simulates "lab" experiments with different types of data, machine learning algorithms, and features tested using data from two sources to explore the power of these models on different models. The first source is the same as the training data, and the second source is the commercial malware dataset provided by the organization's advanced malware detection system. These preliminary results show that commonly used API calls for Windows executables are less powerful than behavioral metrics such as CPU usage, RAM usage, and receiving and sending packets, leading to more predictable results in many variables."

Robust Malware Detection for Internet of (Battlefield) Things Devices Using Deep Eigenspace Learning
Azmoodeh, Amin; Dehghantanha, Ali; Choo, Kim-Kwang Raymond
"The Internet of Things (IoT) in the military usually consists of various Internet connections and nodes such as medical equipment and combat uniforms. IoT devices and nodes are prime targets for cybercriminals, especially government sponsors or national authorities. One attack vector is the use of malware. In this article, we propose a deep learning-based method to detect Internet Of Battlefield Things (IoBT) malware in a device's Operational Code (OpCode) string. We transform the OpCodes into a vector space and use a deep space learning method to classify bad and bad practices. We also demonstrate the strength of our malware detection system and its robustness against unauthorized attacks. Finally, we have made our malware samples available on Github in the hope that they will be useful for future research (eg., to facilitate evaluation of future malware detection methods)."

A review of artificial intelligence-based malware detection using deep learning
Mustafa Majid, A.-A., Alshaibi, A. J., Kostyuchenko, E., &Shelupanov, A.
"The hostile malware spread has seen many problems around the world. It is common to see malware distributed in many countries to make money. With the increase in malware distribution activities, it is now possible to have malware models to train

machine learning models. Therefore, machine learning has become essential for malware detection. The performance of traditional machine learning models is limited due to in-depth training. The advent of deep learning models paved the way for more advanced training and improved detection accuracy with minimal errors. This article reviews the literature on deep learning for malware detection. Deep learning methods for malware detection include CNNs, RNNs, LSTMs, and autoencoders. LSTM has been shown to have better performance in mobile memory. Autoencoders have been shown to have better unattended encoding and decoding for detecting malware. There are many programs for Android malware detection that use machine learning and deep learning. The information contained in this document can contribute to further research in deep learning, which is important to the advancement of government."

Robust PDF Malware Detection with Image Visualization and Processing Techniques
Corum, Andrew; Jenkins, Donovan; Zheng, Jun
"One of the most popular files, PDF is often used by attackers as a vector to spread malware due to its flexible file format and ability to embed different types of content. In this article, we propose a new learning-based method to detect PDF malware using image processing and manipulation techniques. PDF files are first converted to grayscale images using image visualization. Various image features represent different features of PDF malware and quality of extracted PDF files. Finally, a learning algorithm is used to create a classification model that classifies the new PDF file as bad or bad. The effectiveness of the proposed method was evaluated using the Contagio PDF malware dataset. The results show that the proposed method is effective for PDF malware detection. It also shows that this scheme is more resistant to repeated spoofing attacks than state-of-the-art work."

SEdroid: A Robust Android Malware Detector using Selective Ensemble Learning
Wang, Ji; Jing, Qi; Gao, Jianbo; Qiu, Xuanwei
"In response to the rapid increase in Android malware count and the poor performance of manual detection, deep learning started to become the Android malware detection service last year. However, these models depend on the quality of the dataset and perform poorly when the training data is not good enough. In the real world, the quality of data cannot be guaranteed without human review, and even Google Play can have bad practices that cause the education model to fail. To solve this challenge, we propose an effective Android malware detection method based on cluster learning, which seeks to provide optimal solutions not limited by the quality of the data. The proposed model uses a genetic algorithm to help find the best combination of subjects and improve the quality of the model." Our results show that the proposed method is better than other methods in the same field.

Adversarial Deep Learning for Robust Detection of Binary Encoded Malware
Al-Dujaili, Abdullah; Huang, Alex; Hemberg, Erik; OReilly, Una-May
"The malware evolves to avoid detection. Model-based malware detectors such as SVMs and neural networks are vulnerable to so-called competing samples, which are minor modifications to detect malware that allow emerging malware to evade detection. A continuous value method useful for nonlinear graph examples has been developed using the saddle point optimization formulation. We are inspired by them to create similar separation patterns, e.g.A binary domain that characterizes malware. Another unique challenge of malware is that malicious samples must be created in a way that

preserves their functionality. We present a method that can be used to enable relational malware samples to run on binary domains. Using the saddle point formulation, we put competing examples of training models that are robust against them. We evaluate the performance of these methods and others when it comes to Portable Execution (PE) files. The comparison encourages us to apply online metrics calculated during training to assess expected strength."

Robust Malware Detection Models: Leaning from Adversarial Attacks and Defenses

Hemant Rathore a, *, Adithya Samavedhi a, Sanjay K. Sahay a, Mohit Sewak

"The last decade has witnessed the growth of smartphones and their users, which have become the main source of malware. Currently, malware detection engines cannot cope with the volume, distribution, and diversity of incoming malware. That's why the anti-malware community is exploring the use of machine learning and deep learning to develop malware detection models. But research in other fields has shown that machine learning/deep learning is easy. Therefore, in this study, we propose a framework for establishing strong malware detection patterns against competitive attacks.

First, we created twelve different malware detection models using various classification algorithms. Then, relative to the enemy, we propose a gradient-based enemy attack network to obtain the argument of the above model. The attack is to convert the maximum number of malware samples into different samples with small differences per sample. Average detection rate of 98.68% was achieved for 12 permission-based malware detection models and 90 recommended attacks.71% voted against twelve target-based malware searches. We've also set up a simple mandate/purpose list that candidates can use to mislead the discovery process. Next, we introduce our intrusion prevention strategy to prevent search engine attacks. The hybrid distillation-based preservation strategy increases the average accuracy of 12 permission-based analysis models and 59 by 54.21%."

Static Malware Detection & Subterfuge: Quantifying the Robustness of Machine Learning and Current Anti-Virus

Fleshman, William; Raff, Edward; Zak, Richard; McLean, Mark; Nicholas, Charles

"As machine learning (ML)-based malware detection becomes more common, it should be able to identify them better than most antivirus (AV) programs today. It is impractical to set up a consensus test setup for standard malware detection systems for pure distribution. Instead, we address this issue by designing a new experiment in which we measure the performance variation between well-known and not-well-known information in the presence of a negative variation. The change in efficiency is coupled with the avoidance process, and then the system's stability against this method is valuable. Through these experiments, we were able to demonstrate how to measure the value of machine learning-based systems more powerful than AV products in detecting malware that is trying to evade adaptation but will be slow to adapt to new attacks."

3. PROBLEM STATEMENT

In the last few years, great progress has been made in the development of malware detection systems that not only detect malware, but also manage their development. However, malware detection has become more difficult. One reason is that malware developers have become experts in using and developing advanced obfuscation prevention techniques (for example, obfuscation methods) that hide the malicious behaviours of malware. Also, existing and new computers are becoming more distributed, diverse and powerful, creating more opportunities for malware that can take different forms depending on the

machine. While machine learning-based MDSs are effective at detecting new threats, they still fail to detect malware in a changing environment. Malware-related patterns weeded and cut by existing machine learning are limited to specific environments and infrastructures and need not be retained after using advanced obfuscation techniques for malware and/or running the malware on multiple systems. These vulnerabilities will reduce the chances of malware detection. To solve the above problems, we offer a powerful MD framework that incorporates deep learning techniques to improve detection accuracy in dynamic environments.

The proposed model learns the "good" representation of strategies (malware) that are strong for scanning prevention and redirection. The learned notation can then be used to train a classifier in malware detection. More specifically, our framework is based on deep neural networks, where a:MalConv is a design concept for malware detection that has 3 variants, namely (1) preprocessing (2) convolution and (3) fully connected building blocks. to train neural networks. MalConv allows MD to learn how to rebuild after the original malware is infected.

4. SYSTEM ARCHITECTURE

Architecture is a graphical representation of data from information systems that models its processes. It is used as a preliminary step in the development of the process and does not require further explanation. The architecture specifies how the data is accessed and output from the system, how the data is processed by the system, and where the data is stored. Unlike standard scheduling, which focuses on flow control, it does not show information about the timing of the process or how well the process is performing or stabilizing. Logical data flowcharts can be drawn using four simple notations. for example, it represents process and data storage. We use these symbols as Gain and Sarson symbols. Boxes indicate external locations, curved boxes indicate processes, rectangular boxes indicate data storage, and arrows indicate data flow.

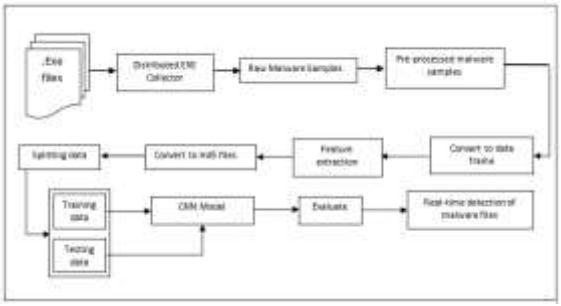

Fig.1 Proposed System architecture

Algorithm

Input: Exe file no I, Ember dataset as P, Deep learning model as M
Output: Malware detection results as R

1. Start
2. Input the Ember dataset
3. Read the json files
4. m ←Read Meta data
5. P ←Pre-processing (f, m)
6. H ←Convert into H5 files (P)
7. Initialize a model M
8. Add convolutional layers
9. Add max pooling layers
10. Add cropping layers
11. Add full connected layers
12. Configure dropout
13. M ←TrainModel(H)
14. For each Epoch e in n
15. For each batch b in m
16. Update M
17. End For End For
18. R←Predict (M, I)
19. Return R

5. CONCLUSION

In this article, we propose a new framework for detecting malware in a changing environment. The proposed method is based on deep neural networks that allow us to extract powerful and useful features to improve identification accuracy in the changing environment. Deep learning is easy in a hostile environment. Prolific competing network methods that can be used to build models during testing or deployment will be easily manipulated by deep learning architectures. The robustness of deep learning architectures is not discussed in the proposed study.

This is one of the key guidelines for future work, as malware flaws are an essential part of implementation in security-critical environments. An incorrect classification can cause enormous damage to the organization. In particular, Malconv is used as a building block during training of deep neural networks. The proposed model was used to learn how to reproduce malware after noise is applied. This is useful for extracting features that are resistant to analytical processes and unstable

environments. Our model is based on real world data. The results show the performance of the proposed method compared to the most advanced malware detection methods.

FUTURE SCOPE

In the future, we hope to learn about future attacks to detect and analyze the next generation of malware. Despite the benefits gained through the abstraction process, existing solutions are highly vulnerable to zero- day attacks. In fact, the malware developer has an advantage over the malware analysts in knowing the current malware protection and can create new models accordingly. A new trend in malware analysis is to explore the possibility of these differences by predicting what future malware will look like to allow analysts to take action against malware and stay one step ahead of this arms race.

REFERENCES

[1] VINAYAKUMAR R1, MAMOUN ALAZAB2 , (Senior Member, IEEE), SOMAN KP1 , PRABAHARAN. (2019). Robust Intelligent Malware Detection Using Deep Learning. IEEE, pp.1-24.

[2] Vikash Kumar;Ditipriya Sinha; (2021). A robust intelligent zero-day cyber-attack detection technique . Complex & Intelligent Systems, p1-24.

[3] Ikram UlHaq;Tamim Ahmed Khan;AdnanAkhunzada; (2021). A Dynamic Robust DL-Based Model for Android Malware Detection. IEEE Access, p1-13.

[4] Liu, Yingying; Wang, Yiwei (2019). [IEEE 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) - Chengdu, China (2019.3.15-2019.3.17)] A Robust Malware Detection System Using Deep Learning on API Calls. , p1456–1460.

[5] Shamika Ganesan;VinayakumarRavi;MoezKrichen;SowmyaV;RoobaeaAlroobaea;Soman KP; (2021). Robust Malware Detection using Residual Attention Network. 2021 IEEE International Conference on Consumer Electronics (ICCE),

[6] Roseline, S. Abijah; Geetha, S.; Kadry, Seifedine; Nam, Yunyoung (2020). Intelligent Vision-Based Malware Detection and Classification Using Deep Random Forest Paradigm. IEEE Access, 8, p206303–206324.

[7] Rhode, Matilda; Tuson, Lewis; Burnap, Pete; Jones, Kevin (2019). [IEEE 2019 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks – Industry Track - Portland, OR, USA (2019.6.24-2019.6.27)] Industry Track - LAB to SOC: Robust Features for Dynamic Malware Detection. , p13–16.

[8]Azmoodeh, Amin; Dehghantanha, Ali; Choo, Kim-Kwang Raymond (2018). Robust Malware Detection for Internet Of (Battlefield) Things Devices Using Deep Eigenspace Learning. IEEE Transactions on Sustainable Computing, p1–9.

[9] Mustafa Majid, A.-A., Alshaibi, A. J., Kostyuchenko, E., &Shelupanov, A. (2021). A review of artificial intelligence-based malware detection using deep learning. Materials Today: Proceedings.

[10]Corum, Andrew; Jenkins, Donovan; Zheng, Jun (2019). [IEEE 2019 2nd International Conference on Data Intelligence and Security (ICDIS) - South Padre Island, TX, USA (2019.6.28-2019.6.30)] Robust PDF Malware Detection with Image Visualization and Processing Techniques. , p108–114.

[11] Wang, Ji; Jing, Qi; Gao, Jianbo; Qiu, Xuanwei (2020). [IEEE 2020 IEEE Wireless Communications and Networking Conference (WCNC) - Seoul, Korea (South) (2020.5.25-2020.5.28)] SEdroid: A Robust Android Malware Detector using Selective Ensemble Learning. , p1–5.

[12] Al-Dujaili, Abdullah; Huang, Alex; Hemberg, Erik; OReilly, Una-May (2018). [IEEE 2018 IEEE Security and Privacy Workshops (SPW) - San Francisco, CA, USA (2018.5.24-2018.5.24)] Adversarial Deep Learning for Robust Detection of Binary Encoded Malware. P1-7.

[13] Rathore, H., Samavedhi, A., Sahay, S. K., & Sewak, M. (2021). Robust Malware Detection Models: Learning from Adversarial Attacks and Defenses. Forensic Science International: Digital Investigation, 37, 301183. P1-10.