Review Article

iMedPub Journals
www.imedpub.com

American Journal of Computer Science and Information Technology
ISSN 2349-3917

2023
Vol.11 No.7:002

A Review on Autonomous AI-Based Robust Model of Malware Detection and Classification for 2D Images

Tharusha Lakshan[*] and Venoli Shehara Gamage

Department of Computer Science and Information Technology, Informatics Institute of Technology, Ramakrishna Road, Colombo, Sri Lanka
***Corresponding author:** Tharusha Lakshan, Department of Computer Science and Information Technology, Informatics Institute of Technology, Ramakrishna Road, Colombo, Sri Lanka, Tel: 0715923471; E-mail: tharusha.20191003@iit.ac.lk

Received date: June 03, 2023, Manuscript No. IPACSIT-23-16906; **Editor assigned date:** June 06, 2023, PreQC No. IPACSIT-23-16906 (PQ); **Reviewed date:** June 21, 2023, QC No. IPACSIT-23-16906; **Revised date:** August 04, 2023, Manuscript No. IPACSIT-23-16906 (R); **Published date:** August 11, 2023, DOI: 10.36648/2349-3917.11.7.002

Citation: Lakshan T, Gamage VS (2023) A Review on Autonomous AI-Based Robust Model of Malware Detection and Classification for 2D Images. Am J Compt Sci Inform Technol Vol:11 No:7

Abstract

This paper is aimed at developing an online intranet College Management System (CMS) that is of importance to either an educational institution or a college. The system (CMS) is an intranet based application that can be accessed throughout the institution or a specific department. This system may be used for monitoring attendance for the college. Students and staff logging in may also access or can search any of the college information.

Attendance of the staff and students as well as marks of the students will be updated by the staff. This system is being developed for an engineering college to maintain and facilitate easy access to information. For this, the users must be registered with the system after which they can access as well as modify data as per the permissions given to them. CMS is an intranet based application that aims at providing information to all levels of management within an organization. This system can be used as a knowledge/ information management system for the college. A given student/staff (technical/non-technical) can access the system to either upload or download some information from the database.

Keywords: Engineering; Organization; College Management System (CMS); Institution; HTML; Software

Introduction

Malware attacks are becoming more powerful day by day. After the technological revalution, there are are so many malwares, viruses are spread to the applications. This is the most dangerous for application details and private information's. Actually it's a huge damage for the privacy damage [1]. Malwares are created unintentionally or intentionally, they can attack as well as can huge damage (harm) in the applications. This section will provide a simple overview regarding analysis of malware and the end need for image processing techniques based classification and detection model. We'll talk about the best technologies to employ when introducing a generic adversarial protection strategy against image malware attacks for practical applications and real world hostile environments. Model robustification techniques for malware image classification, starting with.

We talk about malware image attacks and identified physical world corruptions. Examined in particular are potential evaluation approaches. Also, this critically assesses the drawbacks and advantages of previous research in the field of autonomous AI image classification models and adversarial machine learning. Malware is still one of the most potent hazards in the cyber environment, despite the tremendous advancements in cyber security systems and their ongoing growth [2]. Malware analysis uses methods from a variety of disciplines, including network analysis and program analysis, to examine dangerous samples in order to gain a greater knowledge of their behavior and the way they change over time. Each advancement in security technology is typically swiftly followed by a comparable evasion in the never-ending arms race between malware creators and analysts. The features that novel defensive measures make use of determine some of their effectiveness. An easy way to avoid a detection rule based on the MD5 hash of a known malware, for instance, is to use obfuscation or other more sophisticated approaches. For illustration, detection by using simple methods like obfuscation or more complex strategies like polymorphism or metaphism, a rule based on the MD5 hash of a known malware can be easily evaded.

Literature Review

Refer to Ye, et al. for a thorough discussion of these methods. These techniques alter the malware's binary and consequently its hash, but they do not alter its behavior. On the other hand, it is considerably harder to get around detection criteria that capture the semantics of a malicious sample since malware developers must make more intricate alterations. A key objective of malware analysis is to gather new data that can be utilized to strengthen security protocols and make evasion as challenging as feasible. A logical choice to help such a knowledge extraction process is machine learning. In fact, this is a common literary tendency, with a wide range of methods, goals, and outcomes [3]. In order to facilitate 2D image malware analysis of Windows executables, or Portable Executables (PEs), this review will examine and organize the available literature that uses machine learning and image processing approaches. Alternatively, all apps and web applications. Any security analysts, such as security conscious reverse engineers or software developers, who may profit from the application of machine learning to automate some malware analysis processes and reduce the burden, make up the intended audience of this survey.

Malware detection and classification

Malware identification and classification is the very deep and advanced process becuse getting malwares, it can be categorized the group of them. In other word malware can be classifying different types of malware and each one has their respective malware families. Malware analysis, which is the study of a specific malware sample such as a virus, worm, trojan horse, rootkit, or backdoor, is used for categorization. Detection and analyzing the big number of malware samples by malware analysis is a huge process and it takes lot of time as well as it's depend on various categories as an example: Machine power [4].

Machine learning malware classification methods

Making manual detection rules can no longer keep up with the influx of malware attacks and other new threats [5]. Therefore, it is becoming more and more important to automate malware analysis using machine learning. The Assembler Source Code (.asm) file and the Binary Text Asset Data (.bytes) file are used as input data in various categorization algorithms to accomplish this. These files are converted after being taken from the malware software. These input files can be automatically examined and categorized using machine learning, which can significantly increase the efficiency and precision of virus detection. Therefore, in the current environment where new malware attacks are continually appearing, machine learning-based malware classification approaches are becoming more and more prevalent.

Existing work

Deep malware image classification system using CNN: In that research the authors focus primarily on the difficulty of recognizing malware that spreads through photos, a strategy frequently exploited by attackers to get around traditional signature based detection techniques. Deep malware image classification is a research paper by Mohammad Sabouri and Mehdi Kargar that proposes a deep learning-based approach to classify malware images [6]. The suggested method classifies photos as malicious or benign (android malware images) by using features extracted from the images by a Convolutional Neural Network (CNN). To train and evaluate their model, they employed a dataset of 10,000 malware images and 10,000 benign photos.

The researchers also compared their approach to other cutting edge techniques and found that it performed better than theirs in terms of accuracy rates. The researchers also compared their approach to other cutting edge techniques and found that it performed better than theirs in terms of accuracy rates [7]. Upon reviewing that study, the fact that this research suggests a method for accurately classifying malware photos using deep learning algorithms is one of its benefits. Given the expanding use of graphics as a vehicle for malware infection, this is especially crucial. Using a sizable dataset and a comparison with other cutting edge methods, the authors also offer a thorough assessment of their methodology.

However, get overall in this project, the researchers didn't provide in depth analysis and contribution of the features extracted by CNN. As well as insights into the types of assaults that use photos as a vector for infection could be gained by understanding the unique characteristics that contribute to the classification of malware images [8].

Signature based recommendation systems: The researchers begin by outlining the idea of malware and the necessity of efficient malware detection techniques. One of the earliest and most popular ways for identifying malware at the time was signature based, which they discussed.

The way that signature based approaches operate, according to the researchers, is by comparing a suspicious file's digital signature to a database of known malware signatures. The file is recognized as malicious if its digital signature matches any signature in the database [9].

That study offers a thorough overview of the many kinds of signature based strategies, such as hash based, string based and pattern based ones. The authors also go over the advantages and disadvantages of each strategy and give examples of programs for detecting malware using signatures [10]. Thus, their study's conclusion includes a discussion of the shortcomings of signature based strategies, such as their inability to identify malware that was previously undiscovered and their vulnerability to evasion tactics. Additionally, they recommend combining signature based techniques with other detection strategies to increase the overall effectiveness of malware detection.

Heuristic based system for malware detection: When examining that project, one can see that it offers a thorough analysis of heuristic based malware detection methods. The researchers start off by defining malware and the effects it has on computer systems. As an alternative to signature-based strategies for malware detection, they then present heuristic based techniques. The international journal of computer applications published that study [11].

They go on to describe how heuristic based techniques operate by examining a program's behavior to determine whether it is harmful. Heuristic based methods can identify previously unknown malware by spotting suspicious behavior, in contrast to signature based methods, which depend on a database of known malware signatures.

As a result, they provide a thorough overview of the various heuristic-based methodologies, including static and dynamic analysis techniques. Researchers have outlined and evaluated the advantages and disadvantages of each strategy and offered illustrations of heuristic based malware detection programs [12].

They come to the conclusion that heuristic-based methods can be used to supplement signature-based methods and are a useful addition to malware detection strategies. They contend that combining the two methods can offer a malware detection system that is completer and more effective.

Distance learning based for identification of malware structural information: Malware samples were suggested by Kong as part of their proposed distance learning-based methodology for classifying malware. Identified a fragment using the two unbiased classification models, file and type. According to G. Conti the fragments are treated as if they were a grayscale image, and the classification work done on the pieces offers a draft answer for automatic categorization. Data mining is used by Kong D, et al. to identify and categorize malware [13]. The conventional method of identifying and categorizing malware relies on scanning technology based on the code, which presents a difficult problem to PE. Introduce the field of antivirus to address flaws in data mining and machine learning approaches. Malware can be identified and categorized using a variety of techniques, such as graph based malware detection instruction sequence based classification and API–based classification.

Discussion

Proposed architecture

Problem statement: Today, maintaining computer system privacy is of utmost importance. Currently, computer systems have improved algorithms to detect and classify (more securely, guarantee secrecy, availability and integrity). But in cryptography, the encryption algorithms offer privacy [14]. However, the user must establish secrecy rather than privacy. Malware image categorization is used to determine whether or not an image is compromised by malware. The user then tries to use malware classification to determine to which malware family it belongs. Malware can incorporate malicious data into any image file type (in this case, 2D images are being considered). In this article, we talk about how to recognize 2D malware graphics.

Background and motivation: Malware images identify and classify play a huge role to maintain and improve the safe keeping and transferring of secret information through images. The need to handle the malware affected 2D im and give the sensitive information to only authorized personnel in a secure method is still a key topic in the domain. Example: Digital signatures It will always play a significant role in industries like defense services.

The goal of this study is to help services discover and classify malware picture data in order to more effectively detect malware as either an offensive or defensive strategy.

Proposed solution in detail: In this part, a model called VGG-16 with low dimension normalized input photos is proposed to be used for image based malware identification and classification using machine learning. The important component of the suggested model is to scale down and normalize the virus images before training in order to enhance classification performance and shorten training time. This indicates that the proposed model's input photos have been normalized to the same size. Here, the user can enter 2D photos of any size; the model will resize all images to 224 × 224 size.

In practice, dimensions is chosen so that the original image's width is divisible by dimension, for example, dimensions=2, 4 and 8 for 32, 64 and 128-pixel wide images. Primarily resize all imported photos to 224 pixels in that model. The malware image has been shrunk, however the texture has not been lost despite the reduction in width [15]. The suggested model can be trained more quickly than others while maintaining a high level of classification accuracy.

It is acknowledged that the suggested model treats grayscale pixel values as the malware image's features when they are used to train the model. As a result, the model's training duration is significantly impacted by the size of the input malware image. The training process takes longer the larger the input image size is.

Conclusion

In this research, we used own dataset getting from various images of popular malware data sense making data set and etc. as well as after analyzing those researchers shows that malware detection and classification approaches are recently supremely formulated research area considering and comparing all the different approaches used to detect and classifying malware images shows that the machine learning and deep learning techniques based systems are more accurate other than models. This research is mainly focused, when user given image to the model identified it is malware affected or not. If it is malware affected the model show the malware family it belongs and accuracy of the classification of that process. Thus this research is well designed to implement the final outcome. As well as develop and evaluate the get high efficiency model eventually. Especially use CNN and VGG-16 model for that proposed model. The research mainly achieved and reached its goals and significantly added to the corpus of knowledge in the fields of transfer learning, deep learning, image processing and machine learning. The study acknowledged its shortcomings and noted opportunities for further improvement. In conclusion, this research effort has the ability to assist disaster relief teams in quickly identifying photographs that have been compromised by malware in actual product design applications, enabling quicker and more accurate decision-making regarding the supplied image data sample.

Acknowledgements

I thank my supervisor Ms. Venoli Shehara Gamage who provided insight and expertise in the cybersecurity domain that greatly assisted in the research.

References

1. Vinayakumar R, Alazab M, Soman KP, Poornachandran P, Venkatraman S (2019) Robust intelligent malware detection using deep learning. IEEE access 7:46717-46738
2. Sriram S, Vinayakumar R, Sowmya V, Alazab M, Soman KP (2020) Multi-scale learning based malware variant detection using spatial pyramid pooling network. In IEEE INFOCOM 2020-IEEE on computer communications workshops (INFOCOM WKSHPS). IEEE, Canada. 740-745
3. Son TT, Lee C, Le-Minh H, Aslam N, Dat VC (2022) An enhancement for image-based malware classification using machine learning with low dimension normalized input images. J Inf Secur Appl 69:103308
4. Zou B, Cao C, Tao F, Wang L (2022) IMCLNet: A lightweight deep neural network for image-based malware classification. J Inf Secur Appl 70:103313
5. Gibert D, Planes J, Mateu C, Le Q (2022) Fusing feature engineering and deep learning: A case study for malware classification. Expert Syst Appl 207:117957.
6. Liu L, Wang BS, Yu B, Zhong QX (2017) Automatic malware classification and new malware detection using machine learning. Front Inf Technol Electron Eng 18:1336-1347
7. Makandar A, Patrot A (2017) Malware class recognition using image processing techniques. In 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI). IEEE, India. 76-80
8. Raff E, Nicholas C (2006) A survey of machine learning methods and challenges for windows malware classification. arXiv preprint arXiv 2:09271
9. Kalash M, Rochan M, Mohammed N, Bruce ND, Wang Y, et al. (2018) Malware classification with deep convolutional neural networks. In2018 9th IFIP international conference on New Technologies, Mobility and security (NTMS). IEEE, France. 1-5
10. Al Nafea R, Almaiah MA (2021) Cyber security threats in cloud: Literature review. In 2021 international conference on information technology (ICIT). IEEE, Jordan. 779-786
11. Taylor PJ, Dargahi T, Dehghantanha A, Parizi RM, Choo KK (2020) A systematic literature review of blockchain cyber security. Digit Commun Netw 6:147-156
12. Nataraj L, Yegneswaran V, Porras P, Zhang J (2011) A comparative assessment of malware classification using binary texture analysis and dynamic analysis. In Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence 21-30
13. Kong D, Yan G (2013) Discriminant malware distance learning on structural information for automated malware classification. InProceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 11:1357-1365
14. Conti G, Bratus S, Shubina A, Lichtenberg A, Ragsdale R, et al. (2010) A visual study of primitive binary fragment types. White Paper, Black Hat USA
15. Kruczkowski M, Szynkiewicz EN (2014) Support vector machine for malware analysis and classification. In 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). IEEE, Poland. 2:415-420