iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53253

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Cloud Based Malware Detection System Using Support Vector Machine

Yash Das[1], Akash Kate[2], Neha Chavan[3], Prashant Wakhare[4]
Department of Information Technology, AISSMS IOIT, Pune, India

Abstract: ***One of the most promising technologies for effectively storing information and offering services online today is cloud computing. There are several benefits to using this quickly evolving technology to defend computer-based systems against cyber-related threats. This quickly evolving technology can offer several benefits in defending Cyber Physical Systems (CPS) and the Internet of Things (IoT) against various cyber-attacks. The use of cloud environments to identify malware can be a promising solution because dangerous software (also known as malware) is growing fast and there is no well-known mechanism for detecting it. As a result, we suggest a framework that makes use of machine learning techniques to extract the best features from the data set we've provided and generate accuracy reports. The Support Vector Algorithm is what we're employing.***
Keywords: Machine Learning, Malware Detection, Support Vector Algorithm (SVM), Malware file, Cloud Computing.

I. INTRODUCTION

Malware is defined as malicious software that attempts to infiltrate or damage a computer system without the owner's consent. Malware is categorized into two types: file infectors and stand-alone malware. There are also different ways to classify malware, depending on their particular action, such as worms, backdoors, trojans, rootkits, spyware, adware, etc. Since all current malware applications tend to use multiple polymorphic layers to avoid detection or to update themselves automatically to a newer version at short intervals in order to avoid detection by antivirus software, it is becoming increasingly difficult to detect malware using standard, signature-based methods. The review paper presents a detailed a review of cloud – based malware detection approach and makes the following contribution:

1) Provide a summary of current academic studies on cloud – based malware detection approach.
2) Presents a vision to understand the benefit of cloud for protection of cyber -physical systems from malware.
3) Explains the trends in creation of malware and hiding techniques.
4) Discuss the current challenges and suggest new techniques for malware detection.
5) Presents a cloud -based malware detection framework, which is based on signature, behavior, deep -learning and heuristic -based approaches.

The economic damage caused by malware attacks has reached its peak in recent years. According to the researcher, cyber-attacks damage the world economy by trillions of dollars. Viruses and worms were used earlier to launch attacks, but over time trojans and ransomwares have largely replaced them. In addition to the methods used to spread attacks, the damages that have been inflicted over the years have also evolved. A variety of social engineering techniques, including phishing scams, malicious emails, and exploiting software vulnerabilities, are used to spread attacks. A majority of the attacks steal information from individuals, like credit card details on banking systems, encrypt computer data to block victims' access to computers, and cause millions of dollars of damage to millions of users worldwide.

II. RELATED WORK

The application of cloud computing technologies has increased recently in numerous fields. Malware detection is one of the most current applications. Numerous scholarly investigations have been conducted in this subject. The methodologies utilized in the cloud and the cloud-based detection strategy have numerous advantages over other detection systems. Prior to going into detail about each cloud-based detection technique, various techniques are briefly discussed in the literature.

Ye et al.'s [9] Valkyrie system, which integrates file content and file relations for malware detection, is based on a semi-parametric classification model. The anti-malware software produced by Comodo includes this mechanism. The accuracy of the suggested approach, according to the authors' findings, beats that of existing well-known antivirus scanners including Kaspersky, McAfee, VirusScan, and Bit fender. File associations and content, however, have different characteristics. As a result, directly integrating these attributes can result in data with lower consistency and correlation.

©IJRASET: All Rights are Reserved

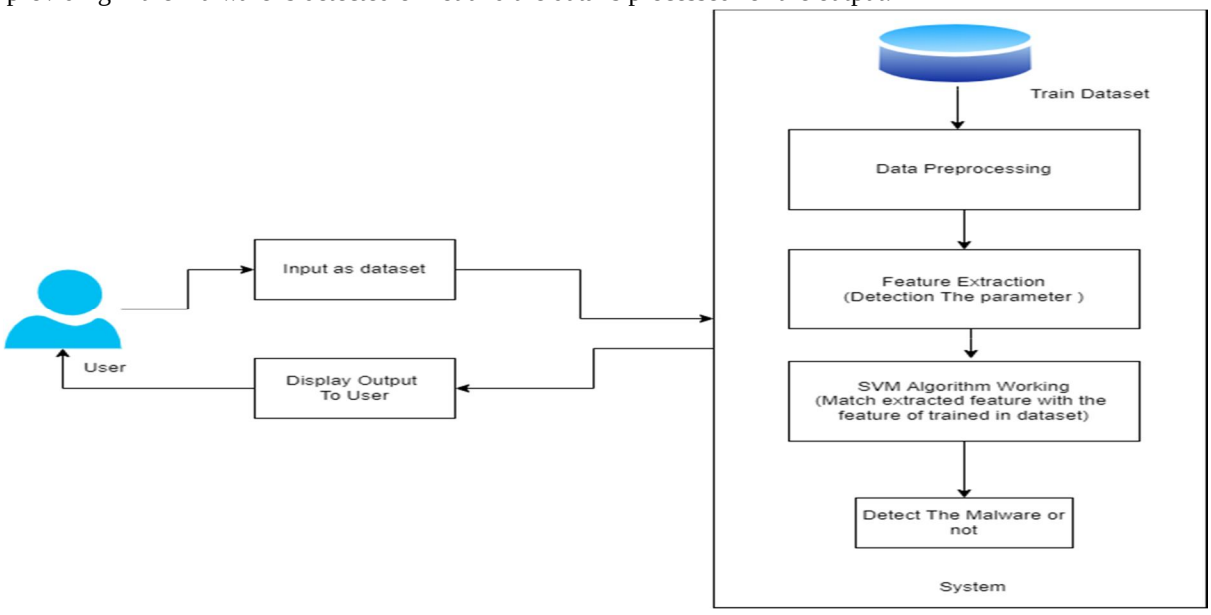
SplitScreen is a brand-new malware detection technique that Cha et al. suggested [10]. A secondary screening stage is used before the signature matching stage in this distributed malware detection system. The two-stage screening step of SplitScreen is divided into client-server operations. The suggested approach was included to ClamAV as an addition, increasing scanning throughput with a more than two-fold increase in signature sets while utilising just half the RAM. when the authors noted, when the number of signatures rises, SplitScreen performs more quickly and uses less memory. The suggested approach is adaptable to a variety of low-end consumer and managed devices. It would be preferable to maximise server efficiency and push some work onto the client side because just one server is utilised on the cloud side.

Win et al. studied cyber-attacks targeting the virtualization infrastructure underlying cloud computing services [11]. The proposed a malware and rootkit detection system that defends guests from several attacks. The system was combined with Support Vector Machine (SVM) based external monitoring on the host, with system call monitoring and system call hashing in the guest kernel. They indicated the efficiency of the proposed approach by appreciating it against well-known user-level malware and kernel-level rootkit attacks. According to the authors, the implemented solution eliminated the demand to us a signature database for malware classification.

Penning et al. [18] summarize mobile malware threats, attacks and cybercriminal motivations behind malware. They discuss in more detail current prevention methods, their limitations, and difficulties encountered in preventing malware on mobile devices. In addition, they propose a cloud framework for mobile malware detection. The proposed framework requires a collaboration between mobile subscribers, app stores and IT security professionals. According to authors, the cloud-based malware detection is a potential approach to mobile security. The work should be developed by examining more studies and how to benefit cloud services and collaborations. Gupta et al. proposed a novel model for malware detection in the cloud [12]. The aim of the studies is to detect the malicious activities with some techniques and warn guests VMs about it. In this paper, DNA sequence detection process, the extracted the DNA sequence from a file formats and used symbols to detect malware files. During the behavioral detection operation, they observed the behavior of the file and determined whether it was a malicious program using the Anubis sandbox. A prototype of the proposed approach (PMDM) is partially implemented on the Eucalyptus. According to authors, PMDM is inexpensive, need less runtime, and ensures well performances for large number of files compared to other known systems. However, this study can be improved further by using a bigger dataset.

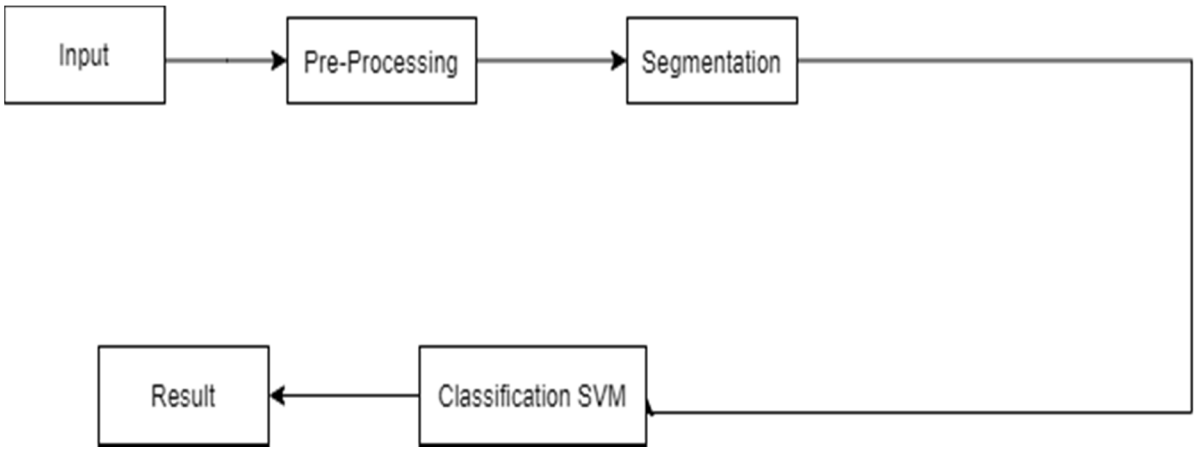
III. SYSTEM ARCHITECTURE

The user needs to create his profile on the log in page providing their required details. Later the user needs to input or download the required dataset. Further the trained dataset is preprocessed and then it is used for feature extraction. In this processes, detection of parameter is done. After the feature extraction process the SVM algorithm is applied. In the working of SVM algorithm, the extracted feature is matched with the trained dataset. This process helps in the detection of malware. Further the algorithm gives the results by providing if the malware is detected or not and the data is processed for the output.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

A. Data flow diagram



IV. METHODOLOGY

Support vector machines are used to detect malware. The SVM algorithm creates a line or decision boundary that segregates n-dimensional space into classes so that we can easily place new data points into the correct category in the future.

A. Gathering Data

The first step of machine learning is to gather data once you know exactly what you want and have the necessary equipment. It is crucial to gather high quality and sufficient data for this step to ensure that the predictive model is as accurate as possible. Data collected is tabulated and called Training Data.

B. Data Preparation

After the training data is gathered, you move on to the next step of machine learning: Data preparation, where the data is loaded into a suitable place and then prepared for use in machine learning training. Here, the data is first put all together and then the order is randomized as the order of data should not affect what is learned.

C. Choosing a Model

In the workflow, the next step is to select a model from among the many that researchers and data scientists have developed over time. Make the right choice that should get the job done. We are using SVM algorithms for these processes.

D. Training

Once the before steps have been completed, you move on to what is often referred to as the bulk of machine learning - training. This is where the data is used to incrementally improve the model's ability to predict. As part of the training process, we initialize some random values for A and B in our model, predict the output using those values, compare it with the model's prediction, and then adjust the values to match the previous predictions. Each cycle of updating is referred to as one training step.

E. Evaluation

Once training is complete, this step checks to see if it is good enough. That dataset you set aside earlier comes in handy here. In an evaluation, the model is tested against data that has never been seen or used for training and is intended to represent how it might behave in real life.

F. Parameter Tuning

Any further improvement in your training can be achieved by tuning the parameters after the evaluation. The training implicitly assumed a few parameters. The learning rate determines how far the line is shifted based on the information from the previous training step during each step.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

All of these variables affect how accurate the training model is, as well as how long the training takes. For more complex models, initial conditions play a significant role in training outcome. It is possible to see differences depending on whether a model starts off training with zeroes or with some distribution of values, which then leads to the question of which distribution is to be used. There are many factors to consider at this stage of training, so defining what makes a good model is crucial. Parameters such as these are called Hyper parameters. It depends on the dataset, model, and training process how these parameters are adjusted or tuned. If you are satisfied with these parameters, you can move on to the next step.

G. Prediction

A machine learning algorithm uses data to answer questions. The final step is to answer a few questions. At this point, machine learning becomes valuable. Now you can use your model to predict the outcome of what you want. From where you create a model to where you predict its output, the above-mentioned steps act as a learning path.

V. DATASETS

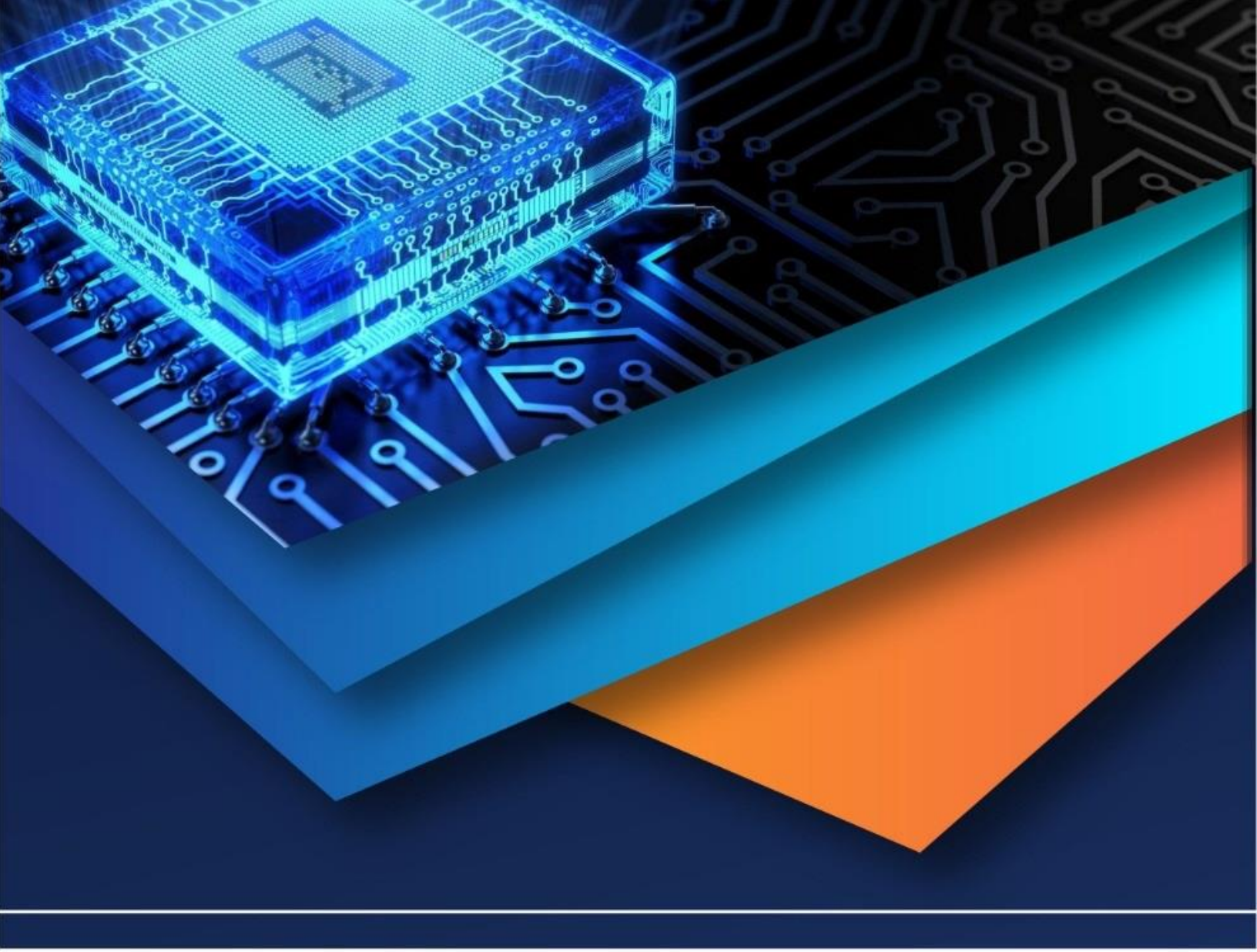
The dataset used for the processes is the malware detection dataset. The dataset consist of trained dataset provided with the parameter.

VI. CONCLUSION

In conclusion, we have suggested a malware detection method that makes use of machine learning. The SVM algorithm is employed by the proposed system. The system we suggest intercepts all running binaries and malware and submits them to a signature database for a thorough check against unknown exploits and malware. This integrates malware detection systems with cloud computing environments. As more users choose to cloud computing platforms for their computing needs, cloud computing is gaining importance. The best method for malware detection in this case is SVM, which requires care to build richer feature representation for greater generalisation and can accurately detect malicious activity.

REFERENCES

[1] Gavrilut D., Cimpoesu M., Anton D., & Ciortuz L., “Malware Prediction Using Machine Learning”, International Multiconference on Computer Science and Information Technology, 2020.
[2] Baset, M, “Machine Learning For Malware Detection”, 2021.
[3] V. Ambalavanan, "Cyber threats detection and mitigation using machine learning" in Handbook of Research on Machine and Deep Learning Applications for Cyber Security, Hershey, PA, USA:IGI Global, pp. 132-149, 2020.
[4] The Comprehensive National Cybersecurity Initiative, Jun. 2020, [online] Available: https://obamawhitehouse.archives.gov/issues/foreign-policy/cybersecurity/national-initiative.
[5] 10 Years After the Landmark Attack on Estonia Is the World Better Prepared for Cyber Threats?, Jun. 2020, [online] Available: https://foreignpolicy.com/2017/04/27/10-years-after-the-landmark-attack-on-estonia-is-the-world-better-prepared-for-cyber-threats/.
[6] Deepti Gupta, Smriti Bhatt, Maanak Gupta, Olumide Kayode, and Ali Saman Tosun, “Access control model for google cloud iot. In 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity),” IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS). IEEE, 198–208, 2020.
[7] Ren, Z., Wu, H., Ning, Q., Hussain, I., & Chen, B. “End-to-end malware detection for android IoT devices using deep learning,” Ad Hoc Networks, 101, 102098, 2020.
[8] Virussign. Malware Downloading Website. Accessed: Jan. 15, 2021. [Online]. Available: https://virussign.com/
[9] Malwarebenchmark. Malware Downloading Website. Accessed: Jan. 15, 2021. [Online]. Available: http://malwarebenchmark.org/
[10] Yazı A. F. Elezaj O. Ahmed J Catak, F. O, “Deep learning based Sequential model for malware analysis using Windows exe API Calls,” PeerJ Computer Science 6, e285, 2020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)