IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 4, April 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004008 |

# Malware Detection For Cybersecurity

**Neha Bhapkar, Yogini Kamthe , Rutuja Deshmukh, Shantanu Girme, Prof. Sushma Akhade**

UG Student, Dept. of Computer Engineering, KJCOEMR, Pune, Maharashtra, India

UG Student, Dept. of Computer Engineering, KJCOEMR, Pune, Maharashtra, India

UG Student, Dept. of Computer Engineering, KJCOEMR, Pune, Maharashtra, India

UG Student, Dept. of Computer Engineering, KJCOEMR, Pune, Maharashtra, India

Assistant Professor, Dept.of Computer engineering, KJCOEMR, Pune, Maharashtra, India

**ABSTRACT:** In this modern, technological age, the internet has been adopted by the masses. And with it, the danger of malicious attacks by cybercriminals have increased. These attacks are done via Malware, and have resulted in billions of dollars of financial damage. This makes the prevention of malicious attacks an essential part of the battle against cybercrime. In this paper, we are applying machine learning algorithms to predict the malware infection rates of computers based on its features. We are using supervised machine learning algorithms and gradient boosting algorithms. We have collected a publicly available dataset, which was divided into two parts, one being the training set, and the other will be the testing set. After conducting four different experiments using the aforementioned algorithms, it has been discovered that LightGBM is the best model with an AUC Score of 0.73926.

**KEYWORDS**: Preprocessing, Feature Extraction., Segmentation

## I. INTRODUCTION

Malware, or malicious software, is software created to infect a machine without the user's knowledge or consent. It is actually a generic definition for all sorts of threats that can affect a computer. A simple classification of malware consists of file infectors and stand-alone malware. The objectives of a malware could include accessing private networks, stealing sensitive data, taking over computer systems to make use of its resources, or disrupting computing or communication operations. The sheer size of IoT networks being deployed today presents an "attack surface" and poses significant security risks at a scale never before encountered. In other words, a single device/node in a network that becomes infected with malware has the potential to spread malware across the network, eventually ceasing the network functionality. Simply detecting and quarantining the malware in IoT networks does not guarantee to prevent malware propagation. On the other hand, use of traditional control theory for malware confinement is not effective, as most of the existing works do not consider real-time malware control strategies that can be implemented using uncertain infection information of the nodes in the network or have the containment problem decoupled from network performance. In this work, we propose a two-pronged approach, where a runtime malware detector that employs Hardware Performance Counter values to detect the malware and benign applications is devised. This information is fed during runtime to a stochastic model predictive controller to confine the malware propagation without hampering the network performance. With the proposed solution, a runtime malware detection accuracy of 92.21% with a runtime of 10ns is achieved, which is an order of magnitude faster than existing malware detection solutions. Synthesizing this output with the model predictive containment strategy lead to achieving an average network throughput of nearly 200% of that of IoT networks without any embedded defense.

## II. MOTIVATION

The purpose of malware analysis is to obtain and provide the information needed to rectify a network or system intrusion. Our goals will be to find out exactly what happened, and to make sure that all infected machines and files are located.

### I. Objective

To analysis the malware prediction by using SVM(Support Vector Machine)Algorithm. To save time. To work predicts a computer driven system's chances of getting attacked by various malwares in the base level in the time of manufacturing of the System based on different specifications of the Operating System and the device.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004008 |

## III. LITERATURE SURVEY

Fanny Lalonde Levesque.," Risk prediction of malware victimization based on user behavior"[1], Understanding what types of users and usage are more conducive to malware infections is crucial if we want to establish adequate strategies for dealing and mitigating the effects of computer crime in its various forms. Real-usage data is therefore essential to make better evidence-based decisions that will improve users' security. To this end, we performed a 4-month field study with 50 subjects and collected real-usage data by monitoring possible infections and gathering data on user behavior. In this paper, we present a first attempt at predicting risk of malware victimization based on user behavior. Using neural networks we developed a predictive model that has an accuracy of up to 80

Zhen Wan, "Multilevel Permission Extraction in Android Applications for Malware Detection"[2], With the widespread use of Android applications in security-sensitive scenarios, more and more Android malware has been discovered. Existing work on malware detection fail to automatically learn effective feature interactions, which are, however, the key to the success of many prediction models. In order to detect malware efficiently and accurately, in this paper, we propose Multilevel Permission Extraction, an approach to automatically identify permission interactions that are effective in distinguishing between malicious and benign applications. We then utilize the extracted information to classify malicious and benign applications by machine learning based classification algorithms. We evaluate our approach in a large data set consisting of 4,868 benign applications and 4,868 malicious applications. The experimental results show that our malware detection approach can achieve over 95.8achieve a better malware detection rate of 97.88

Matu´s Uchn ˇ ar´," Behavioral malware analysis algorithm comparison"[3], Malware analysis and detection based on it is very important factor in the computer security. Despite of the enormous effort of companies making antimalware solutions, it is usually not possible to respond to new malware in time and some computers will get infected. This shortcoming could be partially mitigated through using behavioral malware analysis. This work is aimed towards machine learning algorithms comparison for the behavioral malware analysis purposes

Guozhu Meng, Matthew Patrick§ , Yinxing Xue, Yang Liu‡ , Jie Zhang," :Securing Android App Markets via Modelling and Predicting Malware Spread between Market"[4], The Android ecosystem has recently dominated mobile devices. Android app markets, including official Google Play and other third party markets, are becoming hotbeds where malware originates and spreads. Android malware has been observed to both propagate within markets and spread between markets. If the spread of Android malware between markets can be predicted, market administrators can take appropriate measures to prevent the outbreak of malware and minimize the damages caused by malware. In this paper, we make the first attempt to protect the Android ecosystem by modelling and predicting the spread of Android malware between markets. To this end, we study the social behaviors that affect the spread of malware, model these spread behaviors with multiple epidemic models, and predict the infection time and order among markets for well-known malware families. To achieve an accurate prediction of malware spread, we model spread behaviors in the following fashion: 1) for a single market, we model the within-market malware growth by considering both the creation and removal of malware, 2) for multiple markets, we determine market relevance by calculating the mutual information among them, 3) based on the previous two steps, we simulate a Susceptible Infected (SI) model stochastically for spread among markets. The model inference is performed using a publicly-available well-labeled dataset ANDRADAR. To conduct extensive experi ments to evaluate our approach, we collected a large number (334,782) of malware samples from 25 Android markets around the world. The experimental results show our approach can depict and simulate the growth of Android malware on a large scale, and predict the infection time and order among markets with 0.89 and 0.66 precision, respectively.

Josh McGiff, William G. Hatcher, James Nguyen, Wei Yu, Erik Blasch, Chao LuAutomatic Capability Annotation for Android Malware"[5], Android malware poses serious security and privacy threats to the mo bile users. Traditional malware detection and family classification technologies are becoming less effective due to the rapid evolution of the malware landscape, with the emerging of so-called zero-day-family malware families. To address this is sue, our paper presents a novel research problem on automatically identifying the security/privacyrelated capabilities of any detected malware, which we refer to as Malware Capability Annotation (MCA). Motivated by the observation that known and zero-day-family malware families share the security/privacyrelated capabilities, MCA opens a new alternative way to effectively analyze zero-day-family malware (the malware that do not belong to any existing families) through exploring the re lated information and knowledge from known malware families. To address the MCA problem, we design a new MCA hunger solution, Automatic Capability An notation for Android Malware

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004008 |

(A3CM). A3CM works in the following four steps: 1) A3CM automatically extracts a set of semantic features such as permissions, API calls, network addresses from raw binary APKs to characterize malware samples; 2) A3CM applies a statistical embedding method to map the features into a joint feature space, so that malware samples can be represented as numerical vectors; 3) A3CM infers the malicious capabilities by using the multi-label classification model; 4) The trained multi-label model is used to annotate the malicious capabilities of the candi date malware samples. To facilitate the new research of MCA

## IV. SYSTEM ARCHITECTURE

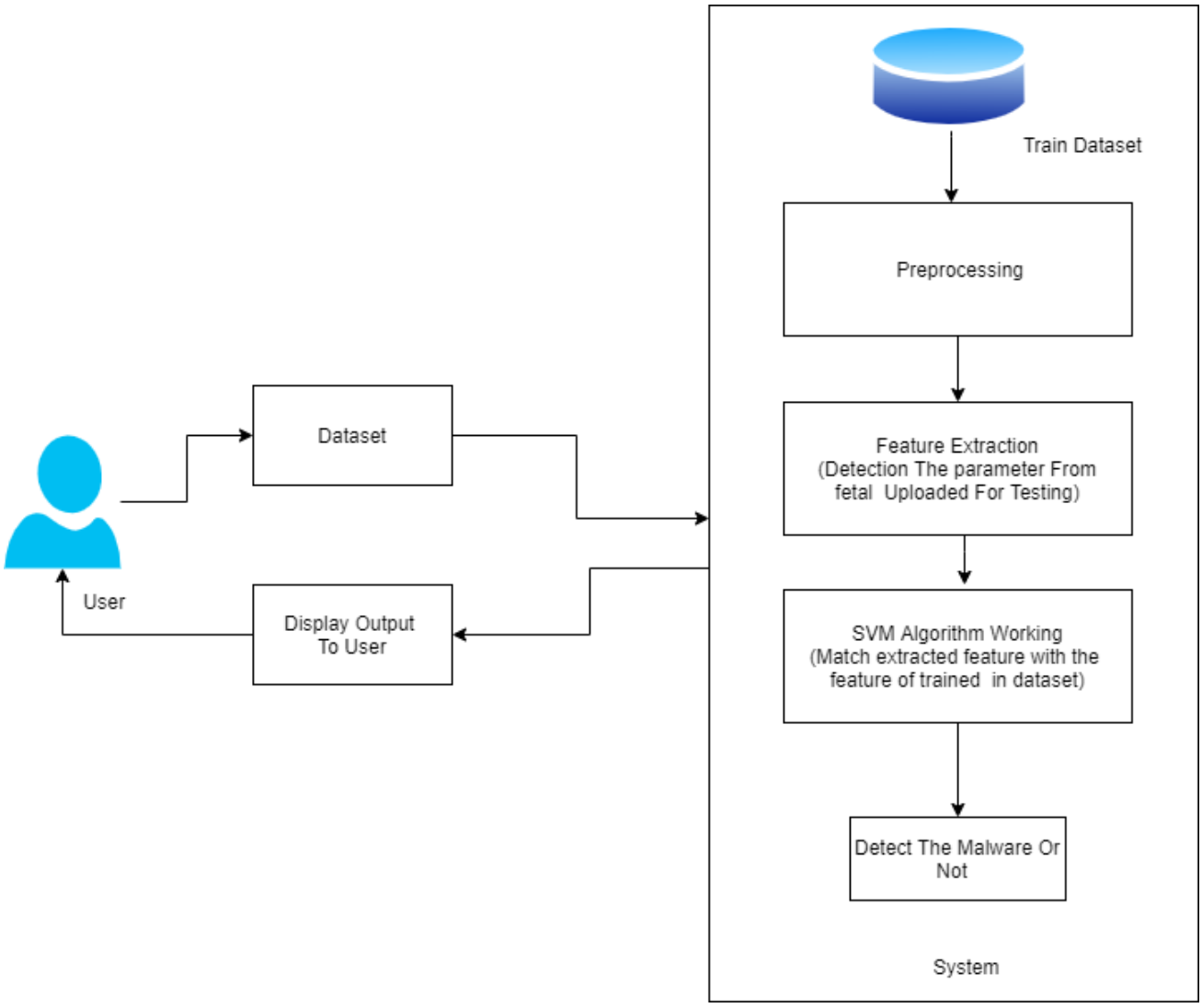


Fig. system architecture

## V. ALGORITHM

**SVM ALGORITHM** :- In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. A separator between the categories is found, then the data are transformed in such a way that the separator could be drawn as a hyperplane. We use SVM for identifying the classification of genes, patients on the basis of genes and other biological problems. Protein fold and remote homology detection – Apply SVM algorithms for protein remote homology detection. Handwriting recognition – We use SVMs to

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004008 |

recognize handwritten characters used widely. SVMs are used in applications like handwriting recognition, intrusion detection, face detection, email classification, gene classification, and in web pages. This is one of the reasons we use SVMs in machine learning. It can handle both classification and regression on linear and non-linear data.

## VI. CONCLUSION

SVM is better classification technique which can be used for detection of malware. Needs attention to construct better feature representation for better generalization. A portable self-defense device with a camera, gps module, an alarm system and a compressed gas can was developed and interfaced together with a single switch control, culprit image was captured sent to the current geographic location with the image to http web server, simultaneously spraying the gas and generating alarm. This setup was obtained for the security of the women.

## REFERENCES

1 Gavrilut D., Cimpoesu M., Anton D., Ciortuz L., "Malware Prediction Using Machine Learning", International Multiconference on Computer Science and Information Technology, 2009.
2 Rhode, M., Burnap, P., Jones, K., "Early-stage malware prediction using recurrent neural networks", computers security, 2018.
3 Baset, M, "Machine Learning For Malware Detection", 2016.
4 Yeo, M., Koo, Y., Yoon, Y., Hwang, T., Ryu, J., Song, J., Park, C.,"Flowbased malware detection using convolutional neural network",2018 International Conference on Information Networking, 2018.
[5] Prof. Basavaraj Chougula1, Archana Naik2, Monika Monu3, Priya Patil4 and Priyanka Das5, SMART GIRLS SECURITY SYSTEM, 1,2,3,4&5KLE's College of Engineering and Technology, Dept. of Electronics & Communication, Belgaum, IJAIEM, ISSN 2319 – 4847,Volume 3, Issue 4, April 2014
[6] G. P. HELMERS, ALARM HAND BAG, APPLICATION FILED mu. 7. 1914, Patented Aug. 3, 1915, GEBHARD P. HELMEBS, 0F BALTIMORE, MARYLAND, "ALARM HAND-BAG", Patented Aug. 3, 1%15, Application filed January 7, 1914. Serial No. 810,766.
[7] U.S. Pat. No. 3,683,114 issued to Egan et al. discloses an automatic dialing and reporting system which is responsive to an alarm condition. The Egan et al. device seizes a telephone line and initiates automatic transmission over the telephone line.
[8] U.S. Pat. No. 4,044,712, Aug. 30, 1977 to Goodman and Jaremus requires active, overt action to trigger the Pressurized Fluid Powered Horn after the attacker makes his intentions known. Nor has the device provided for any deterent value through broadcast, by bright warning orange coloring, that it is being utilized.
[9] U.S. Pat. No. 4,759,309 discloses a passive air, gas aerosol or pressurized fluid activated personal self-protection screech alarm device that is armed prior to the person utilizing it entering into a potentially dangerous area or situation

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details