International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 11, Issue 3 [May.-June. 2022] PP: 13-15

Malware Detection for Cyber Security using Machine Learning

Mahesh Borkar[1], Akanksha Taware[2], Shraddha Erande[3], Neha Auti[4]
Guide – Prof. Akshada Patil[1]
[1,2]*Department of Computer Engineering, Alard College of Engineering and Management, Pune-57.*

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*
Key Words*: Machine Learning, SVM.*

Date of Submission: 17-05-2022 Date of Acceptance: 31-05-2022

I. INTRODUCTION

Malware, or malicious software, is software created to infect a machine without the user'sKnowledge 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

1.1 PROBLEM DEFINATION

Malware poses a threat to computing systems worldwide, and security experts work tirelessly to detect and classify malware as accurately and quickly as possible.

Previous studies showed that malware behavior can be represented by sequences of executed system calls and that machine learning algorithms can leverage such sequences for the task of malware classification

Accurate malware classification is helpful for malware signature generation and is thus beneficial to antivirus vendors; this capability is also valuable to organizational security experts, enabling them to mitigate malware attacks and respond to security incidents.

In this project, we propose an improved methodology for malware detection, based on support vector machine algorithm

1.2 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.

II. LITERATURE SURVEY

Fanny Lalonde Levesque[1]. In this system, 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

Hye Min Kim [2]. This project presents a system that classifies malware by using common behavioral characteristics along with malware families. We measure the similarity between malware families with carefully chosen features commonly appeared in the same family. With the proposed similarity measure, we can classify malware by malware's attack behavior pattern and tactical characteristics.

Zen van[3]. The system 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.

Sanjeev Das[4]. The system propose hardware-enhanced architecture, GuardOL, to perform online malware detection. GuardOL is a combined approach using processor and FPGA.Our approach aims to capture the malicious behavior (i.e., high-level semantics) of malware.

III. DESIGN METHODOLOGY

Based on past work and assessment, the project is accomplished with understanding of an intelligent assistant capable of taking user command and analyses it and respond the user by using voice media. Python libraries and speech reorganization APLs are used to integrate the personal voice assistant python speech to text model is used.

3.1 SYSTEM ARCHITECTURE

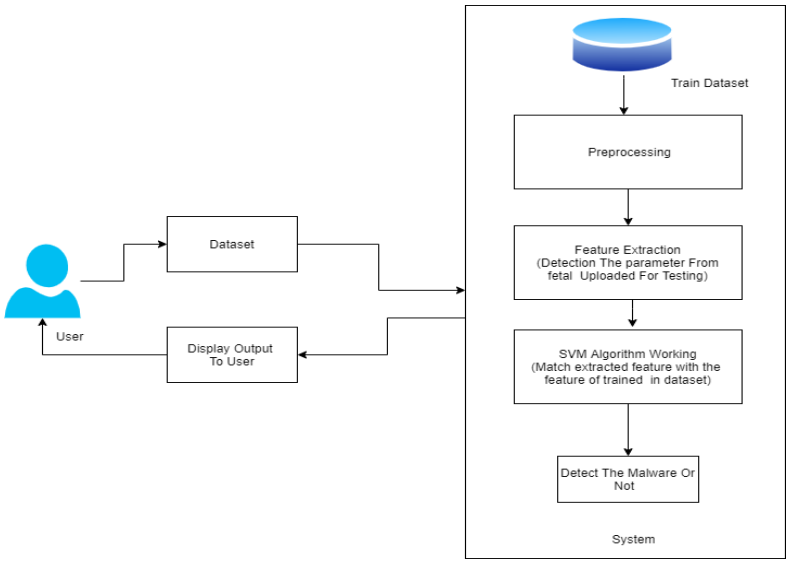

Fig: system Architecture

The proposed system is design such way that it is helpful for user in all aspects related to the day to day task. This system is flexible and robust for user to use. Having a system which can be accessed by mobiles as well as desktops laptops. It is also helpful for disabled peoples.

IV. WORKING

Python -we are using python 3.7 and 3.9 versions not working well with TensorFlow module currently,3.7 is more stable version of python.

TensorFlow-This is an endwise open-source stage for machine learning .it is comprehensive, stretchy environment of tools, lending library and public properties that lets investigate drive the state of the art in machine learning and developers' informal figure and organize machine learning motorized application.

4.1 **Data flow diagram**:

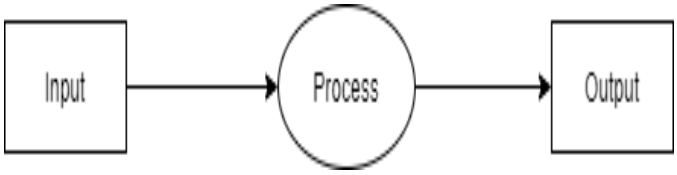

Fig: DFD level 0

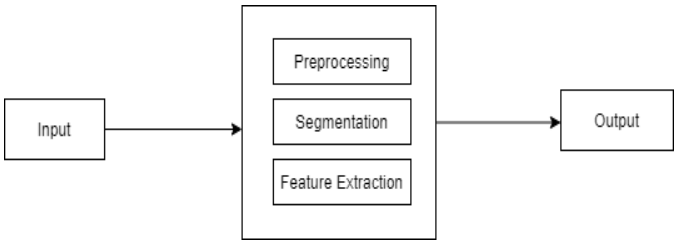

Fig-DFD level 01

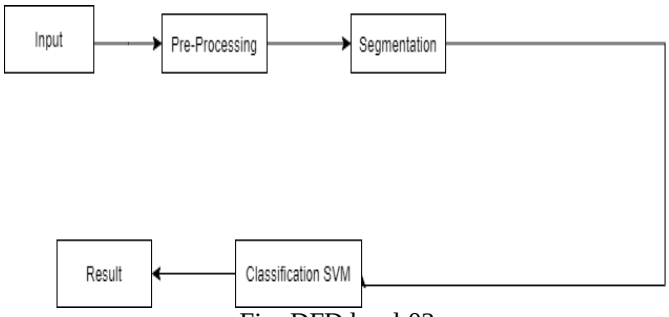

Fig: DFD level 02

V. SCOPE OF SYSTEM

This makes the prevention of malicious attacks an essential part of the battle against cybercrime we are applying machine learning algorithms to predict the malware infection rates of computers based on its features. We are using support vector machine learning algorithms.

5.1 ADVANTAGES

- Malware detection is crucial with malware's prevalence on the Internet because it functions as an early warning system for the computer secure regarding malware and cyber attacks.
- Pay attention to your browser's security warnings. Many browsers come with built-in security tool that scans and warn you before you visit an infected webpage or Download a malicious file.
- It keeps hackers out of the computer and prevents the information from getting CCompromised.

5.2 LIMITATION:

- Requires internet connection.
- It is useless for malware whose code is sufficiently obfuscated.
- Machine learning displays a risk of running inefficient algorithms and making limited prediction when not trained properly.

5.2 APPLICATIONS

- Antivirus
- Android Applications
- We have use this system in offices for cyber attack prevention
- Government level also for security purpose

VI. CONCLUSIONS

In this system we introduced a framework for malware detection based on support vector machine algorithm which result in good accuracy.

This framework was applied to the application-specific malware detection scenario which targets detecting malware infected runs of known applications.

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 convonutional 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.,"Flow-based malware detection using convolutional neural network",2018 International Conference on Information Networking, 2018.