MALWARE DETECTION AND PREVENTION USING MACHINE LEARNING

Project Reference No.: *45S_BE_1851*

College ***:*** *G.M. Institute of Technology, Davanagere*
Branch ***:*** *Department of Information Science and Engineering*
Guide(s) ***:*** *Mr. Amith Shekar C*
Student(S) ***:*** *Mr. Nagachandan P*
Ms. Sanjana V G
Ms. Vaishnavi R Surve
Mr. Vinay K S

Keywords:
Malware, Polymorphic, cyber, PE files, Machine learning

Introduction:

Idealistic hackers attacked computers in the early days because they were eager to prove themselves. Hacking machines, is a trend in today's world. Despite recent improvements in software and computer hardware security, both in frequency and sophistication, attacks on computer systems have increased. Regrettably, there are major drawbacks to current methods for detecting and analysing unknown code samples. Numerous reports indicate that malware's effect is worsening at an alarming pace. Although malware diversity is growing, anti- virus scanners are unable to fulfil security needs, resulting in attacks on millions of hosts.

Objectives:

1. To secure the centralized organizational system by detecting the malwares.
2. Implement the ML algorithms such as decision tree and random forest to train the model
3. Check the PE headers in the external file
4. Predict the malware with trained model by feeding the input.

Methodology:

1. In the first step we choose the dataset
2. Chosen dataset contains malware files and the non-malware file in .exe or .dll or .bat format
3. The .exe files contains an PE headers
4. The trained datasets will train the model with the predefined attributes
5. By applying the any external files which are in .exe format model extracts the PE headers and the Prominent characteristics or the attributes and matches with the trained attributes.
6. If a malware is detected, then the alert message/email is sent to an user.

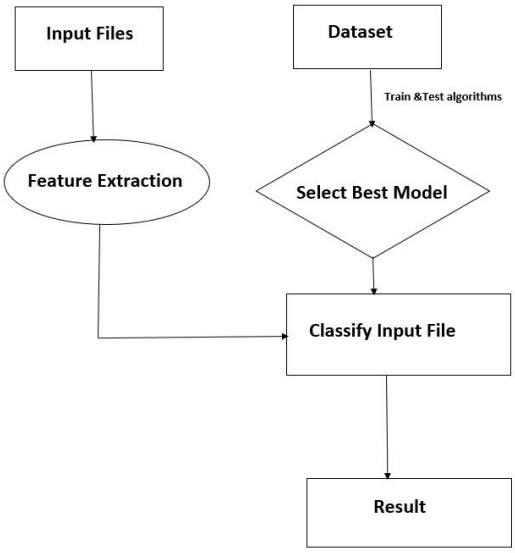

Fig 1: Dataflow diagram

PE header:

The portable executable file header contains the metadata about the executable file itself. The PE format is a file format for the file executable, object code, DLLs and others used in 32-bit and 64-bit versions of Windows operating systems. The PE format is a data structure that encapsulates the information necessary for the windows OS loader to manage the wrapped executable code.

Results and Conclusions:

1. The project will display the output as malware if the file is malicious.
2. The project will detect and display the output as legitimate if the file is not malicious.
3. project shows the error message if the dos header is not found.

```
C:\Users\user\Desktop\running>python checkpe.py MSB.exe
The file MSB.exe is malicious
```

```
C:\Users\user\Desktop\running>python checkpe.py powershell.exe
The file powershell.exe is legitimate
```

```
pefile.PEFormatError: 'Unable to read the DOS Header, possibly a truncated file.'
```

Conclusions:

We have proposed malware detection module based on the machine learning that can be implemented in an organization. This will not only easily detect known viruses but acts as knowledge that will detect the harmful files. our project is cost efficient and protect the invaluable privacy data from the security threats and prevent the immense financial damage.

Scope for future study:

1. The project can be extended to the .bat and .dll formats also.
2. The project shall focus on the mobile devices.
3. User interface can be provisioned in the project.