Study on Malware Analysis Techinques

SruthiKrishna G

Abstract—Nowadays the wide spread of malwares has increased manifold and has emerged as major challenge in the computer world. Malware is the malicious code that will run with or without user involvement. Malware can be in the form of a script, executables etc. Malware can do several activities in the system, like it can log all the details, encrypt the data, it can entirely damage the system. So the presence of the malware need to be detected. Malware analysis can be divided as static, dynamic and hybrid analysis. In this paper its deals with a comparative study on different malware analysis techniques.

Index Terms—, dynamic, malware, machine learning, static

1 Introduction

Cyber Crimes are computer related crimes. It is a crime that involves the combination of network and computer. The computer related crime rate is increasing day by day, as the technologies is developing. Among different crimes malware related crime rate is increasing exponentially. Malware is a malicious program or code. It is used by the attackers in different levels of exploitation. Examples of malware are virus, worms, Trojan spyware etc. Malware can spread easily either through scripts, executables, active contents etc. Malwares can do different activities include logging all activities done on the system, encrypting the content of the data stored etc. Importance of malware analysis is that it can even damage the entire system.McAfee Report of 2016 shown in Fig 1.1 indicates the importance of malware analysis.

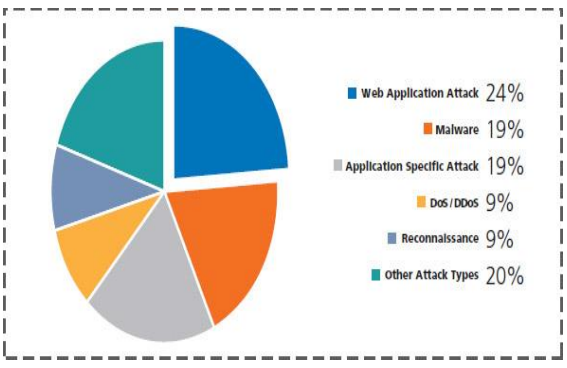

Fig 1.1 Malware Analysis attck Graph

The graph shown in the figure indicates that the 24% is web application based attack is standing out.Then after that 19% is the malware based attacks .so the rate of malware crime is increasing day by day.So these malware presence must be detected.

2 Malware Types

Malwares are malicious softwares, with can replicate easily with the help of a host or not.

Virus – It is a malicious software **.**Which can be wriiten to any normal softwares .Virus can replicate very easily with out the permission of the user.It will reproduce itself or it will modify other softwres.

Worms – Similar to virus family.Worms can spread over the network very easily.It will replicate to different machine in a fast rate.

Trojan – Trojan horse is shorted as Trojan **.**It is a computer program, which will make the user believe that it was a legitimate software. It will pretend to be a real program. It may give permission for someone to remote access our system.

Adware – It is type of malware which will be showing the advertisments on our system. Since it is showing advertisement it is called as Adware's.

Spyware - As the name says that it is actually doing the spy activity .It will track all the activites in the system.Notonly tracking the activities it will log and it will send to the third party who had created this.

Rootkit – It detection is highly difficult, because its existence is hided. It is used to provide unauthorized access to the attacker. Then the attacker can modify the system and collect all the information in the system.

- *Sruthi Krsihna G is with the Cyber forensics and information security, ER and DCI institute of technology, E-mail: sruthikrishna9413@ gmail.com.*

Backdoor –It will not directly harm the system, they are used before doing another type of malware attack. It mainly provides a background for malware attack.

Key Logger – It will log all the key pressed on the system.it wll log the contents typed like passwords, card-numbers etc.

Ransomware – It will encrypt all the data on the system and the victim will ask to pay money for the key to decrypt the data.

3. MALWARE ANALYSIS TECHNIQUES

3.1 Static Analysis

Static analysis is performed statically .It means it is done without the execution of the file [1]. Malware analysis is based on the analysis of their design information.First step is the extraction of the design information.

- **File metadata –** By analyzing the structure we will find Metadata informations.If we are analyzing any executables, its header will contain the metadata information's like modified time, last accessed time etc. By analyzing we can understand that somebody other than as has accessed our system or made changes in the file or exe.Some miss activity can be predicted so we can go for further analysis on that particular exe or file before running that.

- **String Analysis –** It is based on the Strings in the exe.By extracting the string of the data, malware analysis can be done.

- **API and URL'S** –Application program interface is a protocols or tool for building application software.Malicous API's and URL's can be identified .The information of these can be analyzed from the ASCII code of EXE itself. A dataset of already existing API can be used for the comparison and can arrives at a conclusion that it is a malicious URL or API.Analysing the header of the executable can give different information of any suspicious behavior[2]

- **Disassembly code –** Most advanced method. That involves the reverse engineering of executables [3]**.** Reverse engineering involves the extraction of designs information.Analysing the design information by using the dataset of malware. Different algorithms can be used for the malware classification and similarity calculation. Machine learning approach can be used for the similarity analysis between the dataset and the input EXE or file.N – gram based algorithm [4] or Hidden Markov model [5] can be used to identify the similarity and to detect the malicious code.
  N –gram is a fixed sliding window of byes. The size of the window is malware samples set is classified and the input EXEs n gram profile need to be extracted. Then the similarity calculation is done. It is based on the threshold value.
  Hidden Markov model is a statical model based on probability calcuations.It is a machine learning technique. Observations are done with the trained HMM data set and the current analyzing set.Probality calculation is done. If the probability is high, the observation sequence is similar to the training sequences.

3.2 Dynamic Analysis

In this the behavior of the file is monitored while it is executing. If a file contain any malware, while executing the file at that time itself the system will went to some changes. So this dynamic analysis is done while running the system in virtual environment. It is faster than static analysis. Machine learning methods includes k-Nearest Neighbors, Decision Trees, Random Forests, Support Vector Machines and Naive Bayes[6].The model is trained in a way that files similar to the training set are assigned with high probabilities and other files are given low probabilities.

2.3 Hybrid Analysis

Hybrid analysis is the combination of static and dynamic analysis. It involves the entropy analysis of given file either packed or not packed by hidden malicious code. It mainly based on the calculation of opcode frequency and execution trace of executables. Accuracy rate is higher, because it is a combination of both.

4. TOOLS FOR STATIC AND DYNAMIC ANALYSIS

Static analysis is done after obtaining the design information only. Design information includes its assembly code, binary and ASCII code. IDA pro is a disassembler that can generate the assembly code. By analyzing the output, we can find malicious API's, URL etc. IDA Pro can help in malware dynamic analysis also. It is used to identify or to collect the instruction traces. Another tools that promotes the static analysis are OllyDbg and PE explorer.

In Dynamic Analysis, the behavior is traced and analyzed. Ether is a tool which is open source used for dynamic analysis. This tool completely resides outside the target OS.It will analyze the behavior of malware. Another tool is the BSA, Buster Sandbox Analyzer which check the presence of potentially malicious behavior content. It may be a File, Executable etc.it will also notify the changes in the port, Registry. Among the dynamic tools more user friendly will be BSA.

5. COMPARISON BETWEEN STATIC AND DYNAMIC ANALYSIS

From the study, we can arrive at a comparison between the malware analysis techniques. Hybrid analysis is the combination of static and dynamic. The table 5.1 shows the comparison features.

Table 5.1 Comparison between Static and dynamic analysis

Static Analysis	Dynamic Analysis
Fast And Safe	Time Consuming
Good In Analyzing Multi-path Malware	Not Possible
Based on byes, system calls, signatures etc.	Behavioral based
Cannot detected new malwares easily	Can detect
Accuracy is high	Low accuracy

6. CONCLUSION

Malware related crimes are increasing in huge rate. It presence must be detected. This paper deals with the study on static, dynamic and hybrid malware analysis. Important methods used in these three analysis technique are explained. As conclusion from the study we can say that it is better is use hybrid analysis, but it's difficultly rate in implementation is higer.Static analysis is safer and better results are obtained than dynamic analysis.

Acknowledgment

I wish to thank all who supported me in fulfilment of this paper.

REFERENCES

[1] Nirmal Singh, Dr Sawtantar Singh Khurmi,"Malware Analysis Clustering and classifications: A Literature Review" IJCST vol 6, Issue 1 SPL-1 Jan March 2015.

[2] Tzu Yen Wang .ChinHsiungWu, "Detecting unknown malicious Executables using portable Executable headers," in Fifth International Joint Conference.

[3] Taher Ahamed Ghaleb "The role of open software in program analysis for reverse engineering" IEEE 22nd international conference

[4] Abdurrahman Pektas Mehmet Eris, "Proposal of n-gram based algorithm for Malware classification," SECURWARE 2011: The Fith International Joint conference on INC, IMS and IDC.

[5] Saja Alqurashi, Omar Batarfi,"A comparison of malware detection techniques based on hidden markov model" Journel of information security 2016, 7,215-223.

[6] Kateryna Chumachenko ,"Machine Learning methods for malware detection and classification" Bachelor's Thesis Information technology.

[7] Gandeva B. Satrya ,Niken D.W.Cahyani,Ritchie F Andreta"The detection of 8 types malware botnet using hybrid malware analysis in executable file windows operating system" August 2015 conference.

Sruthi Krishna G completed Btech degree from Lourdes Matha College of science and technology in 2016.Currently doing Mtech in Cyber Forensics and Information Security in ER and DCI Institute of technology.