Study of Malware Based On Pattern Matching Techniques

[1]Jaspreet Kaur, [2]Sugandha Sharma
[1]Research Scholar, Deptt of CSE Chandigarh University Punjab (India)
[2]Assistant Professor Chandigarh University Punjab (India)

***Abstract* -Malware is security threat that can break computer operation without knowing user's information and it is difficult to identify its behavior. We can use signature based matching technique, encryption and decryption engines, metamorphism based method and KNN (K- nearest neighbor) algorithm to identify the behavior of malware. Among all these techniques a pattern based technique is well famous for the detection of malware. For the moderation and improvement of the current system the signature based technique is preferred.**

***Keywords* - Malware, its types and detection based techniques**

I. INTRODUCTION

Malware :is a one type of software which can harm the computer's operating system and may also can steal the personal information from the computer, Malware can be made by using any programming language by the programmer. It is very difficult to define a malware with a single term or a single name. A malware can be consider as a malicious software or malcode or as a malicious code .There are different types of malware like adware, bots,Trojan horses, viruses, bugs, rootkits, spyware and worms. The explanations of these are as given below:

Types of Malware:

Adware: It is also known as advertising-supported software. As its name suggest that it is software that shows ads when you are using internet. The examples of this are that some time when you open a site and you see the pop up ads. The main motive of these ads is to steal information of user. When user clicks on these type of ads it fetch the personal information of the user.

Bots: The purpose of the bots is that the programmer can control the infected computer remotely. The collection of bots is controlled by a single Mater which is known as botnet.

Trojan Horses: This type of software pretends to be useful software but in actual it is not. Once it is downloaded then the malicious part in it changes the settings of the computer. And some time it also infects the other system files also. The main purpose behind the creation of Trojan is to discover the financial information and also taking over computer's system resources. Trojans are also regarded as to create back doors to give malicious users access to the One's computer system.

Viruses: virus is a type of malware which adds itself to the other programs. And it spread to the other computers by attaching it to with other executable programs. Virus does not run independently. It needs certain type of action trigger on it to execute. Viruses can also spread through script files, documents, and pen-drives.

Bugs: Bug is unwanted output by compiling in a source code .This is mainly due to human mistake. If there is a minor mistake by the programmer then it can lead to the totally different than what we want.

Rootkits: Rootkit is mainly designed to access remotely computers. It is most dangerous form of malware because it hides its presence that's why it is very difficult to detect of find the ant viruses or any other techniques. To remove this many experts give advice to completely wipe the data of hard disk and install the all software's from the starting.

Spyware: Spyware is a type of malware which spy on all user sensitive actives without the user knowledge. These activities include computer's passwords, accounts related password and keystroke etc. spyware collects the sensitive data of user and transfer to the malicious person. Spyware also has some extra powers like it can also change the settings of the software.

Worms: worms are also like viruses but it runs independently and does not need any human interactions to propagate.

Malware symptoms: There are the following symptoms of the malware:

- We can see unwanted ads which are not related to your site.
- Settings of the computer automatically changed and hard to retrieve the previous one.
- Web browser contains additional components like toolbars which are not required by the user.
- It slows down the speed of the computer and also of the web browser.
- Sometime the files of the computer modified and deleted.
- Strange behavior of your PC.
- Sometime your emails send automatically without user's knowledge.

II.MALWARE DETECTION TECHNIQUES

- Signature based technique
- Anomaly based technique
- Emulation based technique

- Obfuscation based technique
- Heuristics based technique
- CBR based technique
- Bioinformatics based technique
- KNN algorithm technique

Signature Based Pattern Matching Technique: This technique is most popular in Malware detection technique. Antivirus scanners can find the signatures of malware which is a sequence of bytes with the malware code. The malware are 3 types: basic, polymorphic and metamorphic. In basic malware in which the entry point of program is changed. Fig1 shows the basic malware.

Fig 2:Basic Malware

Metamorphic malware is malicious software that is capable of changing its code. Metamorphic malware have virus mutation to the next level and each new variant create different signature which is difficult to recognize. Metamorphic malware is more difficult for anti-virus software to recognize but not impossible. Polymorphic malware changes encrypted code by adding an additional component. [3]

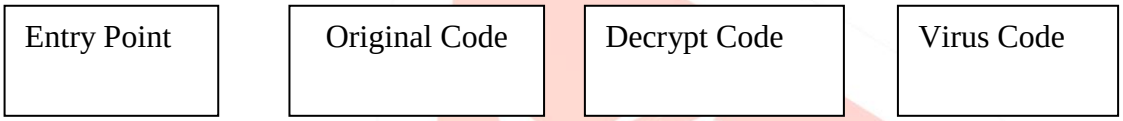

Fig 3: Polymorphic Malware

Anomaly based detection Technique: In this technique any abnormal activity is occur in the program it can be arise. By using this technique any abnormal activity is arise the system can give alert message .It is more reliable technique as compared to another technique because it can be detect new virus. In which the detector can learn the behavior of host and it can detect the zero day attackers. Zero day attackers are the unknown of previous malware detector.

Emulation based detection Technique: In this technique it can be detect the beviour of malware and the sequence of malware. It technique is used to decrease the time of detection. It is used to detect the polymorphic malware as well as metamorphic malware.[4].

Heuristics based detection: In this technique it can be used 2 approaches: static approach and dynamic approach .In static approach to find out the known pattern for matching if it can be present in program. In dynamic approach it can be open other executable files which can be modify its contents. It can be promise to detect the unknown malware and which it can make system more vulnerable by taking the real malware as another.[5].

Obfuscation Technique: This technique is used as a dead code. It can be inserting a large number of codes actually these codes are nothing. In this technique the malware Can be transformed the functionality of this code is same but it is very difficult to understand. [6]

Bioinformatics techniques: The bioinformatics techniques effectively used for Protein and DNA matching method we can identify malware variants with minimum false alarms and misses. Metamorphic malware mutates its code on each replication preserving the functionality of the program. The code is morphed by a small mutation engine also known as metamorphic engine. These malware use different obfuscation techniques to evade the conventional signature based scanner employing exact string matching methods. The size of metamorphic engine is designed to be small in order to bypass detection Structural transformation in the malicious code is introduced with limited set of instruction replacements as total change in the malicious code is impossible (as it would alter the functionality of the variant). Detection of metamorphic malware is still possible as they are less diverse compared to benign samples. The primary reason is in order to preserve maliciousness some generic code is embedded which cannot be transformed to large extent.[3]

Case Based Reasoning (CBR) Technique: In this the user can solve the problem by using the knowledge reuse. In this paper study the different methods have been discussed based on shared nearest neighbor, effectively combining them BM25 and calculation method the Neighborhood Information of samples for the classification purpose [9].

KNN algorithm Classification Technique: KNN (k – Nearest Neighbor) algorithm is used for classification in which K is positive integer. KNN algorithm aims the object is assigned to class which is closest neighbor. if k=1 then it can be assigned a single value for a object. In this paper the KNN algorithm is used in text categorization system. However the KNN classification process in which first find out the k nearest document in the training set. Normally speaking in NNN classification process the class distribution in training set is uneven. The researcher discussed the KNN algorithm which is used different numbers of nearest neighbors for different categories as compared to fixed number across all categories. [10].In this paper the researcher

consider each of the characteristics in our training set as a different dimension in some space and take the value an observation has for this characteristic to be its coordinate in that dimension so getting a set of points in space. We can then consider the similarity of two points to be the distance between them in this space under some appropriate metric [13].

III. CONCLUSION

Malware is a security threat that can break computer operation. Signature based pattern matching technique is most popular for malware detection. But there is one drawback of this technique it cannot detect new virus. In this review paper we also represent the various pattern matching techniques to detect the malware in program. There are various pattern matching techniques are explained in this paper such as anomaly based, emulation based signature based and KNN algorithm for classification etc.

IV. REFERENCES

[1] NwokediIdika and Aditya P. Mathur "A Survey of Malware Detection Techniques" pp.1-47, February 2, 2007.
[2] Dennis Distler "Malware Anylasis an introduction" pp.1-64, December 14, 2007.
[3] Vinod.P, Laxmi.V.,Chauhan.G "MetamOrphicMalware Exploration Techniques Using MSA signatures"In: International Conference on Innovations in Information Technology (IIT),pp 232-237(2012).
[4] Kumar,N.D., Mishra.L., Charan.M.S., Kumar.B.D " The New Age Of Computer Virus and Their Detection"In: International Journal of Network Security & Its Applications (IJNSA), Vol.4, pp 79-96 ,(2012).
[5] Rafique,M.Z.,Chen,P.,Huygens,C., Joosen,W "Evolutionary Algorithms for Classification of Malware Families through Different Network Behaviors" Pp, 1-8, (2014).
[6] Imtithal A. Saeed,Ali Selamatand Ali M. A. Abuagoub "A Survey on Malware and Malware Detection Systems" In:International Journal of Computer Applications,Vol.67,pp 25-31,(2013).
[7] Agrawal,H., Bahler,.L, Micallef,J., Snyder,S., Virodov,A.: Detection of Global, Metamorphic Malware Variants Using Control and Data Flow Analysis. Pp 1-6 ,IEEE(2013).
[8] Emad,S.A., Hashemi,.S. " A General Paradigm for Normalizing Metamorphic Malwares".In:10th International Conference on Frontiers of Information Technology. pp 349-353(2012).
[9] Cai,Y.L.,Ji,.D., Cai,.D.F. "A KNN Research Paper Classification Method Based on Shared Nearest Neighbo". Pp 336-340, (2010).
[10] Baoli,.L, Shiwen,.Y., Qin"An Improved k-Nearest Neighbor Algorithm for Text Categorization".In:20th International Conference on Computer Processing of Oriental Languages, pp 1-6 ,(2003).
[11] Zolkipli, M. F., Jantan ,A. "An Approach for Malware Behavior Identification and Classification" pp 191-194 ,IEEE (2011).
[12] Li. X, Loh,P.K.K, Tan.F "Mechanisms of Polymorphic and Metamorphic Viruses" In: European Intelligence and Security Informatics Conference, pp 150-154 (2011).
[13] Sutton,.O " Introduction to K Nearest Neighbor Classification and Condensed Nearest Neighbor Data Reduction"pp 1-10,(2012).
[14] Rad,B.B,Masrom,M.,Ibrahim,.S "Camouflage in Malwarefrom Encryption to Metamorphis".In: International Journal of Computer Science and Network 74 Security(IJCSNS), VOL.12, pp 74-83(2012).
[15] Egele ,M "A Survey on Automated Dynamic Malware-Analysis Techniques and Tools" Vol. 44, pp 1-41(2012).
[16] Berkat,A. "Metamorphic Computer Virus Detection by Case-Based Reasoning (CBR) Methods" In: International Journal of Software Engineering & Applications (IJSEA) Vol.2, pp 1-10 (2011).