MalDet: How to Detect the Malware?

Samridhi Sharma
Department of CSE,
Seth Jai Parkash Mukand Lal Institute of Engineering and Technology,
Harayana, India.

Shabnam Parveen
Assistant Professor, Department of CSE,
Seth Jai Parkash Mukand Lal Institute of Engineering and Technology,
Harayana, India.

ABSTRACT

Malware is malicious software. This software used to interrupt computer functionality. Protecting the internet is probably a enormous task that the contemporary epoch of computers have seen. Day by day the threat levels large thus making the network susceptible to attacks. Many novel strategies are brought into the field of cyber security to guard websites from attacks. But still malware has remained a grave reason of anxiety to web developers and server administrators. With this war takes place amid the security community and malicious software developers, the security specialists use all possible techniques, methods and strategies to discontinue and eliminate the threats while the malware developers utilize new types of malwares that avoid implemented security features. Easing this dilemma entirely is a rising area of research. This paper is aimed at classification of malware and the various ways of detecting them.

Keywords

Attacks, Obfuscation, Malware Normalizer, reverse engineering.

1. INTRODUCTION

With the increasing development of communication and information systems, a new term and acronym attacked the digital world called as malware. It is a general term, which stands for malicious software and has numerous forms (codes, scripts, active content and others). It has been intended to attain some targets such as, collecting responsive data, accessing computer systems which are private, even sometimes harming the systems. The malware can reach the systems in different ways and through multiple media; the most common way is the downloading process from the internet, once the malware finds its way to the systems, based on the purpose of the malware the drama will start. In a number of cases, the malware will not completely damage the system, instead influence the performance and generate overload process; in case of spying, the malware hides itself in the system, which cannot be detected by the anti-virus software, these hidden malware send important information about the computer to the source. Based on the above challenges, it is primary to carry out an in-depth analysis to recognize the malware for better detection and removal chance of malware. They are designed for the financial gains [1]. They are categorized as to whether they are static or dynamic. In dynamic analysis (also known as behavioural-based analysis), discovery is based on information composed from the operating system at runtime (i.e., during the execution of the program), such as system calls, network access and files, and memory modifications. In static analysis, the detection is based on information extracted openly or unreservedly from the executable binary/source code. The major benefit of static analysis is in providing quick classification. Since antivirus Programs that have the potential to violate the privacy and security of a system. According to the Symantec Internet Threat Report 499,811 new malware samples were received in the second half of 2007 detection. So it becomes necessary to detect the malware.

This paper is organized as follows: Section two has covered the recent state of the malware security. Section three discusses about the malware classification, section four presents the malware dectector. Section five studies mechanism of malware detection, and finally section six explains malware normalization process.

2. RELATED WORK

Technology has turn out to be an building block in recent times where both industry and investigate worlds totally rely on the technology and its functions. Though like the other side of the coin, these developments have also opened the doors for the hacking and attacking community, and in a a small number of years the malware has grow to be a most important security threat, moving computers and networks expansively. Firstly, the process of invading the systems is done for fun purpose only by the hackers and attackers until online trade gained its fame particularly in banking, financial transactions. To identify different types of malware, S.Divya et al. [2] Study the categories of malware, their vulnerabilities and the existing handling mechanisms. Their study concludes two parameters false positive rate and infection ratio in detecting the malware. Due to the mechanism of malware propagation, we can now clearly experience the impact of malware on a variety of computer network infrastructures, technologies and service such as online social networking [3]. Yossi Spiegeletal., [4] discover the choice amid vending new multimedia commercially and bundling it alongside ads and allocating it for free as adware. To sold the software commercially only when its perceived quality is high. It display that adware is extra lucrative after the observed quality of the multimedia is moderately low. Invernizziet.al, [5] present EVILSEED approach to search for the web pages that are malicious and concluded that this approach is efficient than crawler based approaches. Karan B. Maniar [6] has shown that there are many different types of cyber security threats, but at the same time, there are numerous ways to avert those threats. H. B. Kazemian et al., [7] has proposed several machine learning models for text classification to classify the web pages as either malicious or not. There results concluded 89% supervised learning and 87% for unsupervised algorithms.

3. MALWARE CLASSIFICATION

Newly, the figure of information security threats caused by malware has rapidly increased, which leads to urgently studying the threats and accordingly categorizing them, to simplify the process of discovering and handling them, in order to detect them and find appropriate solutions. In this section we have listed and talk about the main and most common categories as follows:

3.1 Virus:

Virus is a computer program that has the capability to damage and self-replicating in order to contaminate host. Viruses are associated or attached to software effectiveness (e.g. PDF document). Launching the infected PDF document could then activate the virus, and a sequence of events may occur based on the function of the virus. A virus may extend from an infected computer to other through network or corrupted media such as floppy disks, USB drives. Viruses have under attack binary executable file (such as .COM and .EXE files in MSDOS , PE files in Windows etc.), boot records and/or partition table of floppy disks and hard disk, general purpose script files, documents that contains macros, registry entries in Windows, buffer overflow, format string etc. [8]

3.2 Worm:

This is one of the kinds of destructive programs, which is self replicating in nature. This works without user authorization and sends copies of itself to other computers invisibly by using the network. There drawback is that they cause harm to internet that by destroying the bandwidth. The task of worm include encrypt files in as crypto viral extortion attack or send junk email which are dissimilar to virus. Example Sasser, My Doom, Blaster, Melissa etc

3.3 Spyware:

Spyware is a combined name given for software which observes and assemble private information about the host like the content regularly visited, email address, personal information like credit card number. It usually enters a computer when open or trial software is downloaded.

3.4 Adware:

Adware also known as advertising-supported software automatically plays, displays, or downloads advertisements to a computer following malicious software is installed or application is used. This piece of code is usually entrenched into open software. The most frequent adware programs are free games, peer-to-peer clients like KaZaa, BearShare etc.

3.5 Trojan:

Torjans are the kind of malware that steals the confidential information from remote computers. This is one of the dangerous type of malware With this they get the passwords and harm the users of the computer.

3.6 Backdoors:

Backdoors are much the similar as Trojans or worms, apart from that they open a “backdoor” on system, as long as a network connection for hackers or other Malware to enter or for viruses or SPAM to be sent.

3.7 Keyloggers:

Records everything you type on your PC in order to collect your log-in names, passwords, and other sensitive information, and send it on to the basis of the keylogging program. These are mostly used by the the business and industry to obtain system practice information.

3.8 Botnets:

A botnet is remotely controlled software – group of autonomous software robots. These are frequently used to send spam /spyware remotely. The configuration of bots is of two types. Simplest bot configuration is where the bots are connected to single central hub. This configuration is not much effective as maintenance of various connections over single server is hard. Second configuration is hierarchical structure where bot master linked to hundreds of bots which is further linked to a lot of bots. Therefore this configuration would effective much larger level.

4. MALWARE DECTECTOR.

The detection mechanism of Malware detector ‘D’ is stated as a function whose domain and range are the collection of program ‘P’ which are executables and the collection {malicious, benign} [9]. Malware detector can be defined as revealed underneath.

$$D\ (p) = \begin{cases} \text{malicious} & \text{if p contains malicious code} \\ \text{benign} & \text{Otherwise.} \end{cases}$$

The detector scrutinizes the program ‘p’ ε P to check whether a program is benign program or malicious program. The objective of testing is to discover out false positive, false negative, hit ratio. The detection process is carried out by detector which looks for the signatures of malware. Signature is defined as the binary blueprint of the machine code of a specific virus. Antivirus programs evaluate their catalog of virus signatures through the records on the hard disk and removable media (including the boot sectors of the disks) in addition to inside RAM. The antivirus retailer revises the signatures regularly and formulates them accessible to consumers by means of the Web.

a) False positive: A false positive arises when a virus scanner incorrectly identify a 'virus' in a non-infected folder or files. False positives are a consequence when the signature used to detect a specific virus is not original to the virus.

b) False negative: A false negative arises when a virus scanner is unsuccessful to identify a virus in an infected file. The antivirus scanner is ineffective to identify the virus because the virus is novel and no signature so far exists, or it may be ineffective to identify since pattern settings or still out of order signatures.

c) Hit ratio: A hit ratio arises when a malware detector detects the malware. This occurs as the signature of malware equals with the signatures accumulates in the signature catalog.

5. MALWARE DETECTION TECHNIQUES

The detection of malware is broadly classified into 4 types which are discussed here.

5.1 Signature-Based Malware Detection

Techniques saleable antivirus scanners look for signatures which are typically a sequence of bytes within the malware code to declare that the program scanned is malicious in nature. Essentially there are three kinds of malwares: basic, polymorphic, metamorphic malwares. In basic malware the program entry point is changed such that the control is transferred to malicious payload. Detection is relatively if the signature can be found for the viral code. Figure 1 show basic malware.

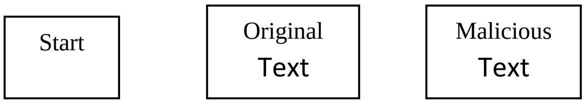

Fig 1. Basic form of virus

Polymorphic viruses change as observance the original code intact. A polymorphic malware consists of encrypted malicious text and the decryption module. To allow the polymorphic virus the virus has got polymorphic engine in the virus body. The polymorphic engine generates new variations every time it is implemented. Signature based detection for such a virus is difficult because each variation new signature is generated which makes signatures based detection difficult. Strong static analysis based on API sequencing is used for polymorphic virus detection. Figure 2 shows polymorphic malware's.

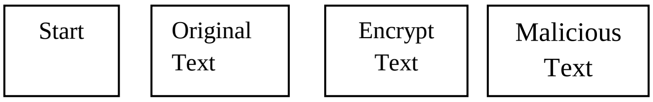

Fig.2 Polymorphic virus

Metamorphic malware have the ability to change the program itself by definite obfuscation techniques so that the new malware generated never look like the parent malware. Such malwares escape the detections from the malware detector since each new modification produced will have different signature, therefore it is unfeasible to accumulate the signatures of multiple modification of same malware illustration. To thwart detection a metamorphic engine has to be executed with some sort of disassembler in order to parse the input code. Following disassembly, the engine will convert the program text and will create new text that will preserve its functionality and would still seem different from the unique text. These viruses are growing in number Figure 3 shows metamorphic malware and multiple signatures for multiple modifications.

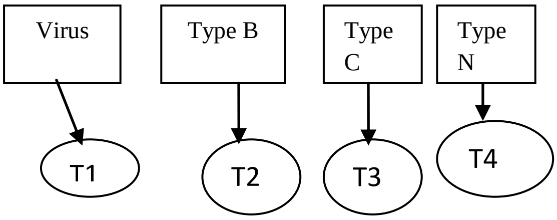

Fig 3. Metamorphic Virus

Let 'T' be the set of malware signatures, Ti Є T are signatures of metamorphic variant belonging to single metamorphic sample 'M'.

The major harms with the signature based detection method is as follows:

- Extraction of signatures and distribution is a complicated assignment.
- The signature generation includes manual interference and needs severe code analysis.
- The signatures can be easily evaded as and when new signatures are created.
- The number of signature storehouse keeps on rising at an alarming rate.

5.2 Specification-based Detection

Specification-based malware detection is a method where a detection algorithm straights the absence of pattern-matching. This detection originates from anomaly based detection. Specification-based detection is the derivate of anomaly based detection. In this approximation of the requirements of application or system is done instead of the implementation of a system or application. This detection method presents a training phase which attempts to recognize the all legitimate behaviour of a program or system which needs to examined. The key drawback of specification based system is that it if very complicated to accurately state the actions the system or program. Panorama is a tool in which system wide information flow of the program is captured under inspection over a system, and ensures the behaviour against a legitimate set of rule to detect malicious action. [10, 11]

5.3 Behaviour based Detection

Behaviour based detection [12] recognizes the action carry out by the malware. In this method dissimilar syntax of the different program is collected having same behaviour, thus this single behaviour signature can recognize a variety of model of malware. The behaviour detector basically consists of following components which are as follows:

- *Data Collection*: This module accumulates the dynamic or static information's are captured.
- *Interpretation:* This module translates the unprocessed information collected by data collection module into intermediary illustration.
- *Matching Algorithm:* It is used to compare the representation with the behaviour signature.

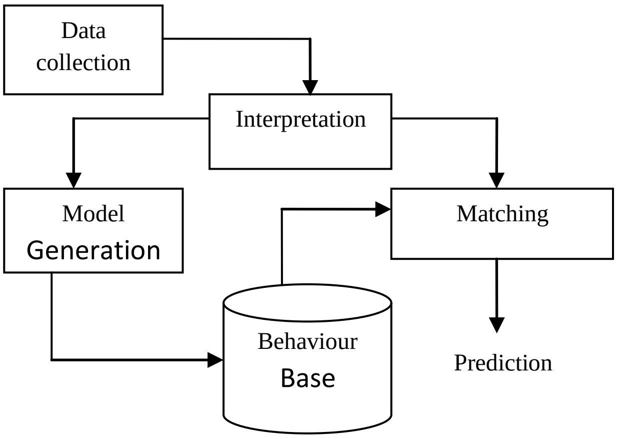

Fig 4. Behaviour Detector

5.4 Obfuscation

Obfuscation is a technique which hides the intended meaning and makes it confusing so that others cannot find the true meaning. Software dealers utilize the obfuscation so that the software would be difficult to reverse engineer. Malware writers obtain it as benefit and obfuscate the malicious program using various obfuscation conversions so that the Malware is hard to reverse engineer and therefore its malicious objective cannot be erudited.

5.4.1 Obfuscation Theory

Given a program P and a conversion function C generates program P' such that the following properties holds true:

- P' is hard to reverse engineer.
- P' grasps the functionality of P.
- P' performs similar to P.

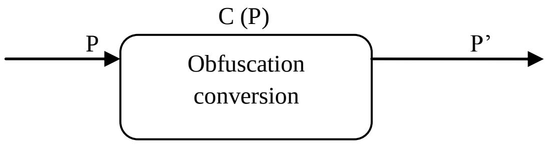

Fig 5. Obfuscation

Example of obfuscation techniques include metamorphic and polymorphic so that they can defeat the signature based detection .With this technique signatures of Malware can be easily distorted. Let us first look at some example of obfuscation technique modifying the signature of the code given below.

Original Code

Hex Opcodes	**Assembly**
71	push ecx
70	push eax
7B	pop ebx
8D 4B 38	lea ecx,[ebx+38h]
70	push eax
E8 00000000	call 0h
7B	pop ebx
83 C3 1C	addebx,1Ch

Signature
7170 7B8D 4B38 70E8 0000 0000 7B83 C31C

Now suppose the original code is obfuscated by inserting a bunch of junk instruction like nops. Then the obfuscated code and the new signature is as follows:

Original Code

Hex Opcodes	**Assembly**
71	push ecx
90	**nop**
70	push eax
7B	pop ebx
8D 4B 38	lea ecx,[ebx+38h]
70	push eax
90	**nop**
E8 00000000	call 0h
7B	pop ebx
83 C3 1C	add ebx, 1Ch

Signature
7190 707B 8D4B 3870 90E8 0000 0000 7B83 C31C

Therefore the modified signature is not detected by Malware scanner due to this the false negative rate will raise extremely. General obfuscation techniques fall into following main types: Dead-code-insertion, Code transportation, Register Renaming, Instruction Substitution

5.4.1.1 Dead-code-insertion: This is can be complete by either putting group of nops (that does not accomplish anything), or putting some number of push x followed by pop x, where x refers to register.

5.4.1.2 Code Transportation: Code transportation is done by inserting jump instruction or unconditional branch instructions in order that the original control flow of the program is preserved.

5.4.1.3 Register Renaming: With this an unused instruction is used rather than using the register in an instruction

5.4.1.4 Instruction Substitution: A series of instructions is linked to a set of another series of instructions which are semantically alike to the novel one. Each series of novel instructions can be replaced by some random instructions.

6. MALWARE NORMALIZATION

Malware normalizer take input the obfuscated edition of Malware and eradicates the obfuscation approved on the program and generates the normalized executables. With this detection rate of detecting the malware increases.

Malware normalization can be recognized as a procedure and method to detect the obfuscated duplicates of malware and escalating the pace of catching the malware by the detector, the output of the normalization will be the unique signature of the malware which has been obfuscated and as a result the signature will be evaluated to the signatures to confirm it, then it will be saved in the catalog of recognized signatures in order to reduce the time of scrutinizing and detecting after that times.

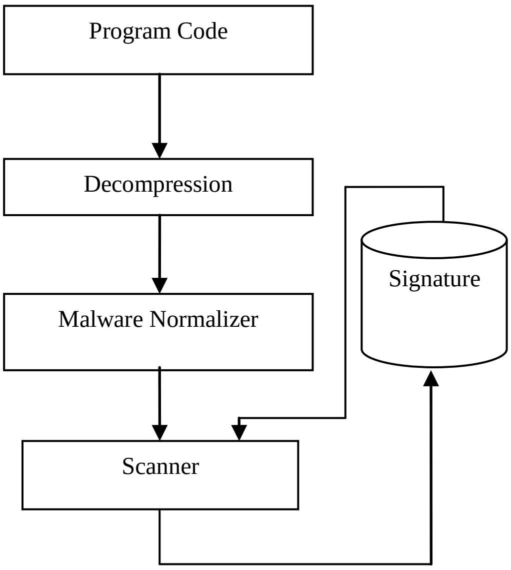

Fig.6. Malware Normalization and signature comparison

7. CONCLUSION

In this survey a sequence of malware detection techniques have been obtainable. Destruction to computers is increasing rapidly with the spreading of malware. Many issues have been evolved due to the malware changing their signature when detection is done. The challenge is to have a growth of high-quality disassembler , so the variations of malware's can be detected in less time thus dropping the calculation overhead. Although the developing procedure of malware along with their detection systems are quickly rising, the lessons can be taken as a major suggestion for the developers in the field.

8. REFERENCES

[1] R. Ford and W. H. Allen, “Malware Shall Greatly Increase...,” *Secur. Priv. IEEE*, vol. 7, no. 6, pp. 69–71, 2009.

[2] S.divya “A Survey on Various Security attacks vulnerabilities and detection Techniques ”International Journal of Enginerring and Technology”

[3] H. Gao, J. Hu, T. Huang, J. Wang, and Y. Chen, “Security issues in online social networks,” *Internet Comput. IEEE*, vol. 15, no. 4, pp. 56–63, 2011

[4] Yossi Spiegel, "Commercial software, adware, and consumer *privacy."* International Journal of Industrial Organization 31, no. 6 (2013): 702-713

[5] Luca Invernizzi et.al “*EVILSEED: A Guided Approach to Finding Malicious Web* Pages”,2012 IEEE 2012 IEEE Symposium on Security and Privacy.

[6] Karan B. Maniar ”Overview of Cyber Security” International Journal of Engineering Trends and Technology (IJETT)Volume 15 Number 3– Sep 2014

[7] H. B. Kazemian and S. Ahmed. "*Comparisons of machine learning techniques for detecting malicious webpages."* Expert Systems with Applications 42, no. 3 (2015): 1166-117.

[8] P. Vinod, R. Jaipur, V. Laxmi, and M. Gaur, “Survey on malware detection methods,” in *Proceedings of the 3rd Hackers' Workshop on Computer and* International Journal of Cyber-Security and Digital Forensics (IJCSDF) 2(4): 10-29 The Society of Digital Information and Wireless Communications, 2013 (ISSN: 2305-0012) *Internet Security (IITKHACK'09)*, 2009, pp. 74–79.

[9] Mihai Christodorescu and Somesh Jha ,” Testing Malware Detectors”,in Proc. ISSTA’04, July 11 - 14, 2004.pages 33-44, Boston, MAUSA, ACM Press.Heng Yin ,Dawn Song ,Manuel

[10] Egele,Christopher Krugel , and EnginKirda , “Panorama: Capturing System – wide Information Flow forMalware Detection and Analysis”, in Proc CCS’07, October 29 – November 2, 2007, Alexandria, Virginia, USA,ACM Press.

[11] Andreas Moser , Christopher Krugel , and Engin Kirda, “ExploringMultiple Execution Paths for Malware Analysis”, Secure Systems Lab, Technical University Vienna.

[12] Greoigre Jacob,Herve Debar,Eric Fillol,”Behavioral detection of malware:from a survey towards an established taxonomy”,Springer-Verlag France 2008.