A Research on Different Types of Malware and Detection Techniques

Chandini S B, Rajendra A B,Nitin Srivatsa G

Abstract—Malware has become a serious threat. Malware analysis is one of the challenging domain. The increase in the malware exploitation has made the detailed study of the malware, understand the different types of malware and its behavior model and analyze the existing detection system with their short comes to identify the research gaps [8] to solve the specific problem. So in this paper, we have presented the different malware taxonomy and different malware detection techniques with its features and also presented the malware model and the research gaps in the malware analysis domain.

Keywords: Polymorphic virus, Malware genesis, Self-replicating cellular automata, Zero-day threat, obfuscation technique, and Anomaly-based detection.

I. INTRODUCTION

Malware is a malicious code which comes with the intention to destruct the system [1]. Some of the symptoms of the infected machine are slow in performance, lack of storage space, crashes, error message, unexpected pop-up windows, missing files, high network activity, website redirected to an unusual website and disabling computer protection and many more [2].

Different types of malware are as follows:

Virus

This is malicious which enters to the computer and causes harm to files by altering files or data [2]. It requires human action to get activated. It can enter the system as an attachment of files, images, greetings and also as downloads from the internet. The different types of virus are:

Boot Sector virus

This type of virus infects the boot sector of a computer. This malicious code resides on a disk, it could be either floppy disk, compact disk or hard disk. It infects computers by altering the contents of the boot sector program. It replaces the legitimate contents with its own infectious. Due to the recent advancement in threat detection this virus is mitigated now.

File Virus

This virus infects executable files. It is also known as a file injector. It stays in memory and tries to infect all programs that load on to memory by inserting infected code into an executable file.

Resident Virus

This type of virus hides within the computer memory and gets activated whenever the operating system starts or execute a specific action.

Non-resident Virus

This type of virus does not reside in memory. It infects the target and transfers the control to the infected application program. It consists of finder module and replicating modules finder will find the new targets to infect the new file and the replicates will infect the file.

Macro Virus

This virus is written in a macro language. It is transmitted through phishing emails containing a malicious attachment and spreads through sharing the infected document.

Polymorphic Virus

This type of virus changes the behavior every time it infects the new file in order to evade from malware scanner. It is difficult to detect this type of virus because of its transformation nature or obfuscation technique [6].

Metamorphic virus

This type of virus changes its signature and code with each infection. It is very difficult to detect and build this type of virus.

Stealth virus

This type of virus hide in memory, files, and boot sector and it uses various method to avoid detection by the scanner. It infects the boot sector in the storage and it uses various tricks to hide the changes it made to the files or the boot sectors. An antivirus should be able to identify the stealth virus by monitoring the evidence in the memory and as well in the area viruses usually attacks

Trojan

This is a destructive program written to steal sensitive information from the victim's machine. It masquerades itself as a non-malicious and they do not replicate and infect other files instead it survives by unseen by antivirus. Some of the action Trojan takes is creating back doors, spying, sending messages, remote machine accessing and creating zombie network which is used in DDoS.

Revised Version Manuscript Received on August 19, 2019.

Chandini S B,Department of Information Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, Karnataka, India.(email:chandini@vvce.ac.in)

Rajendra A B,Department of Information Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, Karnataka, India.

NitinSrivatsa G, Department of Information Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, Karnataka, India.

Spyware

This is malicious software that is installed on the victim's machine without his knowledge which is used to monitor and collect information about a particular user. Anti-spyware tools can be used to prevent spyware.

Adware

This is software comes in the form of advertisements to the user and collect the data regarding user marketing type which can be used to analyze user internet behavior and render the customized ads for the user it collects the data with the user permission and it gets into the machine through freeware, shareware, and infected websites.

Worms

It is a standalone self-replicating malicious software that spreads to the other computers without any intervention its main motive is to harm the network by consuming bandwidth by increasing the payload some of the main categories of worms are:

Email worms: This worms spread through the email.

Internet scanning worms: This worms spread through the internet.

Mobile worms: This worms which spread through Bluetooth features or any mobile communication applications.

In Table 1 consists of different types of malware were in a parameter like Nature, Harm Memory, Self-Replicating, Behavior, infected through and Examples are summarized. Table consist of different types of malware was in Trojan will not self-replicate remaining all virus self-replicate and spread to other files and harm the memory. Polymorphic, Metamorphic and stealth are dynamic in nature and it is difficult to identify this virus. Most of the virus spreads through malicious email attachment, Internet downloaded malicious executable files, USB drive, and Instant message, etc.

Table 1 Malware with properties

Malware Type	Nature	Harm Memory	Self-Replicating	Behavior	Infected through	Examples
Boot Sector Virus	Static	Yes	Yes	Targets specifically a boot sector	USB, Pendrive, Hard Disk	Form, Michelangelo,stoned[10]
File virus	Static	Yes	Yes	Infects the executable files	Internet downloaded executable files	Jerusalem, cascade[10]
Resident virus	Static	Yes	Yes	Loads to RAM to infect other files	File attachment	CMJ, Meve, and MrKlunky[11]
Non-resident Virus	Static	Yes	Yes	Infects the target host and transfer control to other application	Executable files	Executable virus
Macro Virus	Static	Yes	Yes	Written in macro language to infect software application	Phishing email	DMV, Nuclear, Word Concept[12]
Polymorphic Virus	Dynamic	Yes	Yes	It encrypts (morph)itself to avoid detection	Downloading infected email attachment	Involuntary, Stimulate, Cascade, Phoenix, Evil, Proud, Virus 101[13]
Metamorphic virus	Dynamic	Yes	Yes	It changes its signature pattern after each iteration uses obfuscation technique	Downloading infected email attachment	Zmist, Virlock [14]

Stealth virus	Dynamic	Yes	Yes	Memory resident virus	Malicious email attachment	Brain, Frodo, Joshi, Whale
Trojan	Static	Yes	No	Open backdoor, exploit.	MP3 files, image, games,and movie	Rootkit, Vundo
Worms	Dynamic	Yes	Yes	It spreads to another machine without human intervention	Email, Instant messaging, relay Chats, File sharing	Iloveyou, Codered, Blaster, Sasser, Conficker

Malware genesis and some of the attacks and their impact

During the 1980s, the first virus appeared in the machine and was called "ELK cloner"[3]. The first virus was written on the Apple - II, circa. It was created by a ninth standard graduate, Richard skrenta.

Table 2 Major Malware attacks and their impact

Year	Event	Description
1999	Melissa	Melissa malware attack might now seem quaint given the malware detection and prevention techniques but it showed how destructive a major cyberattack can be. Pretend as a word file which contains passwords to big adult sites it gained the attention of the victims and when opened it executed a macro to resend the virus to first 50 contacts' in the users. The surge in email traffic hit the US government and corporations including the likes of big companies like Microsoft and Intel. Melissa was responsible for a total of 1.2 billion dollars in damages. It was coded and released by David L. Smith.[15]
2000	ILOVEYOU	Referred to as Love Bug or Love letter, the worm spread across millions of computers. The email contained a subject with ILOVEYOU prompting victims to foolishly open it to contain an attachment stating love letter with a VBS (visual basic script) extension which was ignored in the windows at that time. Similar to the Melissa virus, this spread to all the contacts in the address book of the user. Total damage was estimated to be around 20 billion dollars.[15]
2003	SQL Slammer	SQL Slammer spreads very fast. Exploiting vulnerability in Microsoft's SQL Server and database products, the Distributed denial of service attack crashed the internet. The famous "Warhol Worm" phrase today is thanks to this attack which resulted in a billion dollars' worth of damages. Bank of America's ATM's was made useless and Continental Airlines had to cancel many reservations because of the attack.[15]
2004	Mydoom	The fastest spreading malware in the history of cyber attacks first emerged in the year 2004. Using phrases like "Mail Delivery System" and "Error" to make users to access and open emails, it paced over the web with full effect. It reportedly affected 25% of total emails. Infected users saw the random launch of programs and network opening created to allow access to others to the machines. ThisDDoS attack affected companies like Google, Microsoft,and Lycos. The estimated damages were about 38.5 billion dollars the highest till date.[15]
2007	Zeus	This Trojan horse was first reported in 2007 and it was used to wipe data of US transportation department. Around 74k FTP accounts were compromised including accounts from banks and corporations such as Cisco and Amazon. This Trojan used Zeus botnet and was deployed to steal credentials of social networks, banks,and email accounts. Total estimated damages were around 70 million dollars.[15]
2010	Stuxnet	The first reported look at future cyber warfare. Stuxnet was deployed using a USB flash drive. It infected software controlling in a nuclear plant in Iran. It caused chaos around the globe as the nuclear codes were stolen. This futuristic was described as the "first digital weapon". This terrifying affair was made into a documentary titled "Zero days".[15]
2014	Sony Pictures Hack	Three years prior to this attack, 77 million user's data were stolen and the service was taken offline for 10 days. Coming back to this date, the famous hacker group Guardians of Peace or GOP hacked Sony and around 100 terabytes of data were stolen including emails, movie script and phone numbers of 100 celebrities. The hack took place and the malware-infected Sony companies computers and made it inoperable.[15]
2017	WannaCry	Dubbed as the biggest malware attack by cybersecurity experts, WannaCry infected computers around 150 countries. They exploited the security vulnerability in older versions of the Windows OS. Thisransomware encrypted data and demanded a ransom to unlock them.[15]

II. MALWARE MODEL

Self-Replicating Model

In 1984 Cohen presents a generic mathematical model for the computer virus [3]. It was identical to the Neumann's self–replicating cellular automata model. There are three major components in the system according to von Neumann's as shown in Figure 1.

- Machine
- Constructor
- Information on the tape

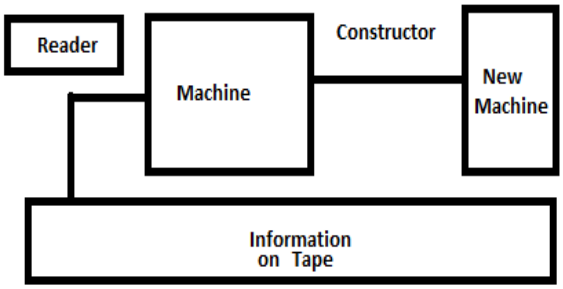

Figure 1.Self-replicating Machine

The machine consists of component like tape which provides instruction for the machine, reader which reads the instruction and construction which takes the instruction to build a new machine. The new machine is a replica of the old and it constructs a new machine.

Malware has a different phase which is mentioned below:

Dormant phase in which the virus will be idle and will be activated by some events.

Propagationphase in which the virus places its copy to another program.

Triggering phasein which it performs action what it was intended to do.

Execution phaseit reflects its behavior to the system by opening pop windows or slow downing the system etc.

III. MALWARE DETECTION TECHNIQUES

Malware is affecting throughout the world. Studies show that the attacks are getting worse day by day. So, to protect against malware better defenses has to install. The quality of the defense is based on the technique it uses. To build a better defense we need to first understand the technique and understand the strengths and limitations of using that technique.

Malware detection can be broadly classified into two types:

1. Anomaly-based detection
2. Signature-based detection

Anomaly-Based Detection

Anomaly-based detection technique has two phases. A learning phase and a monitoring phase [5]. In the learning phase, it attempts to learn the normal behavior of the system. The detector fully learns about the host in the learning phase. The anomaly detection technique has the ability to stop the devastating zero-day attacks.

Anomaly-based detection is not efficient it has more false alarm rate and the design is complex involving what features to be taught in the training phas

Different types of anomalies based detection are as follows:

Dynamic Anomaly-based detection

In dynamic Anomaly-based detection, information gathered during the execution of the program is used to detect the malicious code [5]. The program under inspection is monitored during execution, in the monitoring phase and is checked for abnormal behavior which was learned during the training phase.

Different approaches in Dynamic Anomaly-Based Detection:

1. PAYL a tool which calculates expected payload
2. Data mining approach for Malware detection techniques
3. Using computer Forensic methods for privacy-invasive software
4. A short sequence for system calls
5. Finite State Automata(FSA) for detecting anomalous programs
6. Process profiling for system calls
7. NATE (Network Analysis of Anomalous Traffic Events)

Static Anomaly-Based Detection

The main difference between Dynamic and static based Anomaly-based detection is that in static based anomaly detection the characteristics of the file structure are used to detect the malicious code [5]. The main advantage of this detection technique is that the program carrying the malware cannot be executed and it will be dealt with using static based detection technique.

A file print analysis is one type of static anomaly detection technique.

During the training phase, the model has characterized the various file types which are present on the system based on the structure which means byte composition. The models learn the behavior of the structural file and it intends to handle the malware. Any file under inspection which varies greatly from the given model will be marked as suspicious.

The fileprint analysis helped deal with 95% of pdf files which contained malware. More work has to be done to find malware embedded in the middle of the pdf.

Hybrid Anomaly-based detection

Ghostware is a malware that hides its existence from the operating systems querying utilities [5]. This is usually done by intercepting the querying results and modifying them so no traces can be found. The detection of file hiding ghostware was tedious and the hybrid technique is used to detect this type of virus.

Signature-based detection

Signature-based detection uses classification of what is known to be malicious to decide the program is malicious or not [5]. Programs or the application have attributes that can be used to create a unique signature. Various algorithms can be used to determine its digital signature. When a malware scanner identifies an object as malicious, its signature is

added to a database of known malware. These repositories may contain hundreds of millions of signatures that identify the malicious program or the application. However, these techniques have drawbacks like susceptible to the evasion and zero-day attack. This can be overcome in anomaly-based detection.

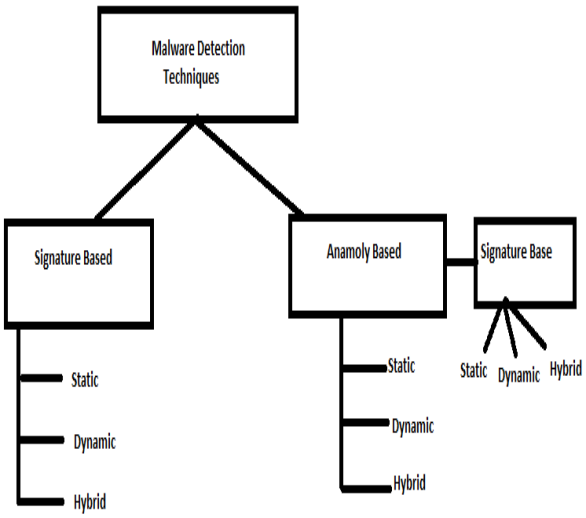

Figure 2 Malware Detection Techniques

Specification-based detection

Specification-based detection is a type of anomaly-based detection which tries to address the high false alarm rates of the anomaly based detection techniques. So, in other words,specification-based detection technique can be called as the derivative of anomaly based. Specification-based detection tries to approximate the requirements of the system. The main limitation of using this technique is it is too difficult to specify completely and accurately the entire valid behavior a system which exhibit.

Similar to anomaly-based detection technique, Specification is also of three types:

Dynamic based technique
Static based technique
Hybrid based technique

Table 3 consist of the Static, Dynamic and Hybrid technique comparison with the parameter like time, resource, detection of a new instance of the virus, effectiveness, efficiency, technique, an example [9]. In which it says that static is more effective and hybrid technique is more efficient.

Table 3 Comparison table of the malware detection technique

Type / Parameters	Static	Dynamic	Hybrid
Time Required	Low	High	High
Resource Consumption	Low	High	High
Detect new instance of the virus	No	Yes	Yes
Effectiveness	High	Medium	Low
Efficient	Low	Medium	Yes
Technique	Signature-based	Behavior-based	Heuristic technique
Example	Virustotal	Ethereal, Process Explorer	Bridesmaid, Samadroid

IV. RESEARCH GAPS IN MALWARE ANALYSIS& RESULTS

Malware is affecting throughout the world. Studiesshow that the attacks are getting worse day by day. Malware analyzer or the researcher needs to identify the different vulnerability and the malicious behavior which exploits the system. The attacker develops new techniques to evade detection. Here we present some of the root causes of the major sources of the gaps.

- Lack of context during the detection or the analysis phase [4].
- The complication in attack analysis because of the flexibility in the system design.
- Environment constraints is a challenge to replicate the same issue.
- Economic constraints to upgrade the system.
- Attacks are dynamic.
- Lack of efficient detection system.

V. FUTURE SCOPE OF WORK

In this paper different malware analysis technique is discussed with its pros and cons which helps us to understand the existing technique for malware analysis and the existing gaps. Future malware analysis using machine learning, artificial intelligence, quantum computing and deep learning which are advanced algorithm can be used to detect the different malware in an efficient manner and the data can be given to the forensic team to have better cybersecurity.

VI. CONCLUSION

In this survey paper, one can understand the various types of malware and its taxonomy [7]. The malware model is illustrated with its different component and malware phases. And also the major attacks are discussed with its impact. The different existing malware detection techniques are discussed with its features and we have identified some research gaps. Static and anomaly-based detection alone is not sufficient to detect all type of virus.

The hybrid model would be required with more efficiency to detect virus-like a polymorphic virus whichis difficult to analyze. More efficient detection technique needs to be developed which can identify the dynamic malware.

VII. ACKNOWLEDGEMENTS

The manuscript is prepared by taking assistance from Accendere Knowledge Management Services Pvt. Ltd, we are thankful to them. We also express our gratitude to our teachers and mentors for guiding us throughout the work.

REFERENCES

1. R. Zhou, J. Pan, X. Tan, and H. Xi, "Application of CLIPS Expert System to Malware Detection System," in 2008 International Conference on Computational Intelligence and Security, 2008, pp. 309–314.
2. O. Yavanoglu and M. Aydos, "A review on cyber security datasets for machine learning algorithms," in 2017 IEEE International Conference on Big Data (Big Data), 2017, pp. 2186–2193.
3. Vikram, "Viruses : Different types and Examples," Snakebytez, 2010. [Online]. Available: https://www.snakebytez.com/viruses-different-types-and-examples. [Accessed: 10-Dec-2018].
4. P. Szor and Peter, The art of computer virus research and defense. Addison-Wesley, 2005.
5. N. Idika and A. P. Mathur, "A Survey of Malware Detection Techniques," West Lafayette, 2007.
6. H. Seifi and S. Parsa, "Mining malicious behavioural patterns," IET Inf. Secur., vol. 12, no. 1, pp. 60–70, Jan. 2018.
7. SebastianZ, "Security 1:1 - Part 1 - Viruses and Worms | Symantec Connect Community," Symantec, 2013. [Online]. Available: https://www.symantec.com/connect/articles/security-11-part-1-viruses-and-worms. [Accessed: 11-Jan-2019].
8. V. Rao and K. Hande, "A comparative study of static, dynamic and hybrid analysis techniques for android malware detection," Int. J. Eng. Dev. Res., vol. 5, no. 2, pp. 1433–1436, 2017.
9. H. Okhravi, C. R. Meiners, W. W. Streilein, and T. Hobson, "A Study of Gaps in Attack Analysis," 2016.
10. D. A. Mundie and D. M. Mcintire, "An Ontology for Malware Analysis," in 2013 International Conference on Availability, Reliability and Security, 2013, pp. 556–558.
11. "The 10 best (or should that be worst?) malware attacks | IT PRO." [Online]. Available: https://www.itpro.co.uk/security/innovation-at-work/29577/the-10-best-or-should-that-be-worst-malware-attacks. [Accessed: 02-Feb-2019].
12. "Information Security information, news and tips - SearchSecurity." [Online]. Available: https://searchsecurity.techtarget.com/. [Accessed: 10-Apr-2019].
13. "F-Secure | Cyber Security Solutions for your Home and Business." [Online]. Available: https://www.f-secure.com/en/welcome. [Accessed: 12-Feb-2019].
14. "Phoenix Web Design." [Online]. Available: https://www.phoenixwebsitedesign.com/. [Accessed: 05-Mar-2019].
15. "F-Secure Virus Descriptions: Stoned." [Online]. Available: http://www.f-secure.com/v-descs/stoned.shtml. [Accessed: 30-Mar-2019].
16. D. Flower. "Immunoinformatics: "Predicting immunogenicity in silico". Quantum distributor, 1st edition, 2007.