

Malware Detection Vectors and Analysis Techniques: A Brief Survey

Dipjyoti Deka¹, Nityananda Sarma², Nithin J. Panicker³

¹⁻³*Department of Computer Science and Engineering, Tezpur University
Tezpur, Assam*

¹dipjyoti20@gmail.com

²nitya@tezu.ernet.in

³getnithinjoseph@gmail.com

Abstract— Over the last decade, a war has been taking place which involves the computer security community and the black hat hackers. The security community uses every possible routine and strategies in order to cease and also to remove the damage caused by the malicious programs while at the same time the black hat community advances their techniques that can easily bypass the installed security measures. In this paper, a study has been carried out on different classes of Malware and their propagation techniques. We highlight on different detection vectors for malware along with different tools to analyze them. Some challenging issues which need further research are also presented.

Keywords— Black hat, Hacker, Malicious, Programs, Detection, Vectors, Malware, Propagation.

I. INTRODUCTION

In today's world of computing, Malware is a very potential threat that makes the computer security more vulnerable. The act of malicious activity continues to grow and getting very sophisticated. It exposes the computer system to the possibility of being attacked or harmed by means of Internet or data communication. The term Malware comes from merging the two words "Malicious" and "Software"[1]. A Malware is a piece of program that is installed on a system silently that is without awareness of the owner of the system with a motive to steal sensitive information, access private data and also sometimes it harms the system by altering the functionality of some legitimate applications as results slow down the system. Malware, is any software used to disrupt computer operations, gathers sensitive information and gains access to private computer systems or displays unwanted advertising[2]. There are various ways by which malware can invade the system and also getting control of the system. Among those the common source of malware infection is data transfer. Now a days, the Internet is the most common way of malware to compromise a system. Once malware enters a system, it scans for vulnerability of the system and perform unintended actions and eventually slow down the system[1]. In a system, malware generally targets the executable files, registry files, memory, system files, MBR files and their

infection does not limit within a system, they can extend their attack level into network traffic which leads to slow down usual network traffic. Generally malware developers or malware programmers use malicious content to get unauthorized access to private or financial data for financial or personal gain. A black-hat hacker is a hacker who "violates computer security for little reason beyond maliciousness or for personal gain"[3]. The main purpose of this paper is to give a study on different forms of malware, their propagation techniques, various detection vectors for malware and tools to analyze those vectors. The rest of the paper is organized as follows. Section II discusses classification of malware and propagation techniques. Section III presents different methods of malware analysis. Section IV gives details about some of detection vectors for malware analysis. Section V presents some of the recent techniques for malware detection. Section VI presents some challenging issues for detection of malware. Finally Section VII presents conclusion and future work.

II. MALWARE CLASSIFICATION

At present where the world is getting advanced in technology, but at the same time the number of malware attacks is also getting increased and still it continues to grow. With time goes on, typical form of malware has been developed to violet computer security and the researcher also in a process of detecting and analyzing them. In recent studies, malware has been categorized into different types[4]. In today's modern world, malware developer or writer always try to improve malware program to get rid of malware detection and analyzing techniques. In general malware can be classified as first generation and second generation malware[5]. Under first generation of malware, the internal content of the malware remain same. But in case of second generation of malware the internal structure of the of malware keep on changing in every sequence of their attack. Under second generation category malware can be further classified as Encrypted, Oligomorphic, Polymorphic and Metamorphic Malwares[5].

A. Encrypted Malware

There are basically two parts of encrypted malware: a main body or encrypted body and a decrypted code[5],[6].

The actual content of this malware is present in the main body which is encrypted and when the malware starts executing on the host computer, the decrypted body converts the main content into machine executable code[6]. The first encrypted malware was reported as CASCADE[7].

B. Oligomorphic Malware

Encrypted malware has only one decrypted function and this limitation leads to implementation of oligomorphic malware. This type of malware has collection of different decryption patterns which are randomly used for new host computer[6]. The oligomorphic Malware is the advanced version of encrypted malware. The first oligomorphic malware was reported as Whale in 1992[8].

C. Polymorphic Malware

Polymorphic malware is far more advanced and complicated than encrypted and oligomorphic malware. This type of malware generally uses obfuscation technique and has the ability to create unlimited number of descriptors[8]. The common obfuscation techniques are dead-code or junk code insertion, instruction replacement, variable or register substitution etc[9],[10]. The first known polymorphic malware was 1260, appeared in 1990[8].

D. Metamorphic Malware

This type of malware is also known as body-polymorphic malware[11]. Metamorphic malware does not have a encrypted body but instead it changes its body(structure, code ,size) but its functionality remains same[5],[6]. The first metamorphic malware was found as Win95/Regswap[13]. In this section we have discussed about various forms of malware and their propagation techniques as shown in the Table I.

TABLE I
DIFFERENT FORMS OF MALWARE

Type	Features	Propagation Techniques	Targeted Domains
Virus	1. Self - replicate 2. Attach its code into legitimate program 3. Active when host program runs [12],[14]	Inject its malicious content into legitimate application remain inactive and when the infected application such as executable file runs it becomes active and spreads itself from one program to another[12],[15]	.COM, .EXE, PDF, JPEG, Boot records, general purpose script files, registry files etc[16].
Worms	1. Self-replicating program 2. Uses network to spread[16].	Uses network to send its copies of itself to other system without user authorization[16].	Its main effect is on network bandwidth and junk emails[17].

Spyware	1. Keep track of user's activity without their knowledge [15]	Generally spyware is installed on user system by bundling itself with legitimate software[18].	Spy user's information like frequently visited website, email, credit card numbers, password etc[16].
Trojan	1. Hide itself 2. Disguise as a legitimate program 3. Give access to remote hijacker to control user system[15],[17].	Pretend to be benign program and installed on user system without user's consent[16].	Steals user's login credentials. Monitors user system, damage system files etc[14].
Rootkit	1. Ability to get privileged access to a computer while actively hiding its presence 2. Use masking network connection, hiding registry keys [15],[19],[28].	These are designed to be enabled administrator level access to computer or network[18].	Conceals other malware Password stealing Provide attacker an unauthorized access to Control user system[20].
Botnet	1. Control user system remotely 2. Zombie program (trojan,worm) controls centrally over public and private network infrastructure [1]. 3. Each instance of botnet is called bot.	In general botnet is referred as Zombie network. It's a collection of infected system which are under common control over network infrastructure. It spreads spam or spyware remotely to Control user system. Each will wait for command from the central system who controls the botnet[16].	Spy user's sensitive information Steal login credentials etc.

III. MALWARE ANALYSIS TECHNIQUES

Malware Analysis has been a very challenging research topic in the field of Computer Security. The outcome of the Malware analysis is to understand the context of the malware intrusion incident, its threat level. The main motive of the malware analysis is to collect information about the all infected files of a compromised system, the inner working of the malware and its behavior. There are basically two

methods involved for malware analysis: Static and Dynamic and these two processes are further categorized into basic and advanced[22].

A. Basic Static Analysis

This technique is fast and safe which examines malware binaries without running it[15]. This technique does not give us the view of actual instruction of malware executable file[22]. However it retrieves information about some of the basic functionality of malware like its hash signature, strings etc. But basic static analysis does not work at all on sophisticated malware.

B. Basic Dynamic Analysis

This type of technique on the other hand requires running malware binaries and examines its behavior. This technique gathers information about various functionality of malware and also helps to remove infection of compromised system[22]. In order to implement this technique malware analyzer should set up a safe environment that allows to monitor the running malware without damage to their system. Even though this technique involves running malware but it does not effective with all type of malware as it cannot identify some of the typical malware functionalities[22].

C. Advanced Static Analysis

Unlike basic static analysis, this techniques needs to run malware binaries which calls for reverse engineering process of malware internal content[22]. This technique involves loading malware executable into a dissembler and examines the instruction of malware program and also give us a proper understanding about what malware program does[22]. In order to use this technique, it requires deep knowledge about assembly code, programming and operating system concepts[22].

D. Advanced Dynamic Analysis

This technique generally requires debugger to analyze running malware binaries[22]. Debugger is a software used to give details of internal state of the program. Advanced dynamic analysis gives the clear picture of internal state of the malware program and also using this technique one can examines or monitors each and every steps of the malware program. Advanced dynamic analysis is successful in retrieving information details about most of the typical malware.

IV. MALWARE DETECTION VECTORS

The most essential input to detect a malware is the deep knowledge about analyzing malware behavior[14]. There are various vectors or parameters that helps to differentiate between malicious behaviors and benign behaviors. In this section some of the important detection vectors for malware and also the respective tools to analyze are summarized as shown in the Table II.

TABLE II
DIFFERENT DETECTION VECTORS FOR MALWARE ANALYSIS

Name of vector	Scope of Detection	Details of vector	Existing Tools for Analysis
Hash Signature	Static Analysis	Known as fingerprint of Malware[22]. MD5 and SHA1 two algorithms are used to generate hash signature of a executable file (16 bytes and 20 bytes)[22]	Any Antivirus Software
Strings	Static Analysis	String is a sequence of characters of a program which contains the list of host contacted by the program and also some dll or system functions that are invoked by the program[22].	Process Explorer [23]
Image or icons	Static Analysis	In general legitimate executable contains verified image or icon. But malware executable does not have its image or icons.	Process Explorer [23]
Registry Activities	Dynamic Analysis	Registry activities includes addition, deletion or modification of registry files by some processes.	Process Monitor [24]
Inbound and Outbound network Traffic	Dynamic Analysis	Generally network traffic give details about source and destination and also internal details of the data packet.	Wireshark, ApateDNS, Inetsim [22],[24] [25].
Autostart Location	Dynamic Analysis	Some executable install itself in variety of locations like registry, startup folder etc.	Autoruns [22].
System API calls	Dynamic Analysis	In general application and services invokes application programmable interface (API) for execution. API calling gives information about the Thread ID, name of the DLL made by API calls and also give details about name of parameter passing, return values.	API Monitor V2, Process Monitor [24,26].
Memory Image	Dynamic Analysis	Memory address gives details about instruction address, process address etc from main memory that enables to detect malicious behavior.	Volatile.py ,Memoryze [22],[27]

V. RECENT TECHNIQUES IN MALWARE DETECTION

Heuristic based Techniques are generally employed machine learning and data mining methods in order to find the behavior of the running malicious program[35]. In recent works [36],[37] one of the efficient method of machine learning technique is the Hidden Markov Model(HMM) to detect malware. Data mining techniques has also been a great deal of research technique to investigate malicious behavior[38]. In general, Heuristic based Techniques take place in two phases: Training(learning) phase and Detection(monitors) phase[39]. In the course of training phase, the system is examined in the absence of attack[39] and the outline of the normal behavior is created using machine learning approach[39]. During the phase of monitoring this outline is compared with the current behavior and any difference in the behavior leads to a alert of being malicious attacked.

VI. ISSUES FOR DETECTION OF MALWARE

From time to time malware developers have been code in order to invade computer and network security by defeating every possible detection methods that the researcher have developed for many years. Many of these detection methods have their own limitation. In this section, some challenging problems for detecting malware have been addressed with respect to the various detection methods.

A. Issues in static based detection

In this technique, the main input is malicious binaries or executable file and a malware can be recognized from its known hash signature. But this technique does not work on new or unknown malware which signature is not available[29]. Static based detection is useless for advanced and sophisticated malware as they are capable of changing its signature pattern[30]. A proper knowledge of assembly language and internal function of operating system is necessary to conduct static based detection[15].

B. Issues in dynamic based detection

The main disadvantage of this technique is that it is time consuming and requires a safe environment[15]. Generally this technique is conducted under virtual environment like sandboxes etc to get rid of attack from malware[31-33]. But now a day's detecting malware in virtual environment is becoming very difficult as malware developer uses anti-virtual machine technique to hide malicious features. This technique is focused on finding abnormal behavior of an application which indicate the presence of malicious activity[17]. But determining a normal program or system behavior is very difficult task which leads to high false alarm rate[34]. Presently, malware writer uses obfuscation technique to conceal malicious functions that makes the extraction of malicious behavior difficult[1]. Dynamic based detection method uses reverse engineering of malware executables in order to learn the malicious content. However, for malwares, equipped with obfuscation technique are very difficult to carry out the reverse engineering process and as a results its malicious content cannot be acquired[1].

VII. CONCLUSION AND FUTURE WORK

At present Malware Developer or the Black Hat community always one step ahead of the computer security professional by implementing new strategies and techniques to conceal

malicious code. As a result the existing techniques and tools for detection and analyzing of malware are not enough or effective and make the anti-malware solutions challenging issue in the field of computer and network security. This survey suggests that to lessen malware attacking rate the deployment of improved and advanced autonomous countermeasures are necessary. The recent trends in malware analysis solutions, machine learning and data mining techniques offers a promising role in the field of developing autonomous approach in order to minimize malware problems to some extent. In future we will apply some of the machine learning and data mining techniques for anti-malware solutions.

ACKNOWLEDGMENT

This work has been supported under the project entitled "Implementation of Distributed Honeynet System at North East to Collect, Distribute and Analyze live Cyber-attack data." funded by CDAC-Mohali and Deity.

REFERENCES

- [1] Vinod, P., et al., "Survey on malware detection methods", *Proceedings of the 3rd Hackers' Workshop on computer and internet security (IITKHACK'09)*. 2009.
- [2] Malware Definition. [Online]. Available: <http://techterms.com/definition/malware>
- [3] Moore, Robert (2005), *Cybercrime: Investigating High Technology Computer Crime*, Matthew Bender & Company. p. 258. ISBN 1-59345-303-5.
- [4] M. F. Zolkipli and A. Jantan, "An approach for malware behavior identification and classification", in *Computer Research and Development (ICCRD) 2011, 3rd International Conference on*, 2011, vol. 1, pp. 191 – 194.
- [5] Sharma, Ashu, and Sanjay Kumar Sahay, "Evolution and detection of polymorphic and metamorphic malwares: a survey.", *arXiv preprint arXiv:1406.7061* (2014).
- [6] Rad, Babak Bashari, Maslin Masrom, and Suhaimi Ibrahim, "Camouflage in malware: from encryption to metamorphism", *International Journal of Computer Science and Network Security* 12.8 (2012): 74-83.
- [7] Beaucamps, P., "Advanced polymorphic techniques", *International Journal of Computer Science*, 2001, 25:400411.
- [8] P. Szor, *The Art of Computer Virus Research and Defense*: Addison-Wesley Professional, (2005).
- [9] L. Xufang, P. K. K. Loh, and F. Tan, "Mechanisms of Polymorphic and Metamorphic Viruses", 2011 *European Intelligence and Security Informatics Conference (EISIC)*, pp. 149-154, (2011).
- [10] D. Bruschi, L. Martignoni, and M. Monga, "Code normalization for self-mutating malware", *IEEE Security & Privacy*, vol. 5, no. 2, pp. 46-54, (2007).
- [11] Rad, Babak Bashari, Maslin Masrom, and Suhaimi Ibrahim, "Camouflage in malware: from encryption to metamorphism", *International Journal of Computer Science and Network Security* 12.8 (2012): 74-83.
- [12] An introduction to malware - CERT UK. [Online]. Available: <https://www.cert.gov.uk/wp-content/uploads/2014/.../An-introduction-to-malware.pdf>.
- [13] Ferrie, P., Corporation, S. and Monica, S., "Hunting for metamorphic", *Proceedings of the Virus Bulletin Conference 2001*, Czech Republic, Prague, 2001, Sep 27-28, 123144.
- [14] What is the Difference: Viruses, Worms, Trojans and Bots? [Online] Available: <http://www.cisco.com/web/about/security/intelligence/virus-worm-diffs.html>.
- [15] Dolly Uppal, Vishakha Mehra and Vinod Verma, "Basic survey on Malware Analysis, Tools and Techniques ", *International Journal on Computational Sciences & Applications (IJCSA)* Vol.4, No.1, February 2014.
- [16] Pranit Gaikwad, Dilip Motwani, Vinayak Shinde, "Survey on Malware Detection Techniques", *International Journal of Modern Trends in Engineering and Research (IJMTER)* Volume 02, Issue 01, [January - 2015] e-ISSN: 2349-9745, p-ISSN: 2393-8161.
- [17] Damshenas, Mohsen, Ali Dehghantanha, and Ramlan Mahmoud, "A survey on malware propagation, analysis, and detection", *International*

- [18] Spyware [Online]. Available: <https://en.wikipedia.org/wiki/Spyware>
- [19] What is a Rootkit?[Online]. Available: <http://www.varcode.com/security/rootkit>.
- [20] Rootkit. [Online]. Available: <https://en.wikipedia.org/wiki/Rootkit>
- [21] Sharma, Ashu, and Sanjay Kumar Sahay. "Evolution and detection of polymorphic and metamorphic malwares: a survey." arXiv preprint arXiv:1406.7061 (2014).
- [22] Sikorski, Michael, and Andrew Honig, *Practical Malware Analysis* Netw. Secur 2012.12 (2012): 4.
- [23] Process Explorer. [Online]. Available: <http://technet.microsoft.com/en-us/sysinternals/bb896653.aspx>.
- [24] Process Monitor. [Online]. Available: <http://technet.microsoft.com/en-us/sysinternals/bb896645.aspx>.
- [25] Wireshark. [Online]. Available: <http://www.wireshark.org/>
- [26] API Monitor. [Online]. Available: <http://www.rohitab.com/apimonitor>.
- [27] Volatile.py. [Online]. Available: <http://www.volatilityfoundation.org/>.
- [28] Rudd, Ethan, et al., "A Survey of Stealth Malware: Attacks, Mitigation Measures, and Steps Toward Autonomous Open World Solutions." arXiv preprint arXiv:1603.06028 (2016).
- [29] Landage, Jyoti, and M. P. Wankhade, "Malware and Malware Detection Techniques: A Survey", *International Journal of Engineering Research and Technology*. Vol. 2. No. 12 (December-2013). ESRSA Publications, 2013.
- [30] OKane, Philip, Sakir Sezer, and Kieran McLaughlin, "Obfuscation: the hidden malware", *IEEE Security & Privacy* 9.5 (2011): 41-47.
- [31] Zabidi, Muhammad Najmi Ahmad, Mohd Aizaini Maarof, and Anazida Zainal", Challenges in high accuracy of malware detection", *Control and System Graduate Research Colloquium (ICSGRC)*, 2012 IEEE. IEEE, 2012.
- [32] Willems, Carsten, Thorsten Holz, and Felix Freiling, "Toward automated dynamic malware analysis using cwsandbox", *IEEE Security and Privacy* 5.2 (2007): 32-39.
- [33] Yoshioka, Katsunari, Takahiro Kasama, and Tsutomu Matsumoto, "Sandbox analysis with controlled internet connection for observing temporal changes of malware behavior", 2009 Joint Workshop on *Information Security (JWIS 2009)*, 2009.
- [34] Idika, Nwokedi, and Aditya P. Mathur", A survey of malware detection techniques", Purdue University 48 (2007).
- [35] Alqurashi, S. and Batarfi, O.(2016)" A Comparison of Malware Detection Techniques Based on Hidden Markov Model", *Journal of Information Security*, 7, 215-223. <http://dx.doi.org/10.4236/jis.2016.73017>
- [36] Annachhatre, C., Austin, T.H. and Stamp, M. (2015), "Hidden Markov Models for Malware Classification", *Journal in Computer Virology and Hacking Techniques*, 11, 59-73. <http://dx.doi.org/10.1007/s11416-014-0215-x>
- [37] Wong, W. and Stamp, M., (2006) "Hunting for Metamorphic Engines", *Journal in Computer Virology*, 2, 211-229. <http://dx.doi.org/10.1007/s11416-006-0028-7>
- [38] Siddiqui, Muazzam Ahmed, "Data mining methods for malware detection" ProQuest, 2008.
- [39] Kuber, Ms Shital Balkrishna, "A Survey on Data Mining Methods for Malware Detection."