Volume 3, Issue 4, April 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey on Techniques in Detection and Analyzing Malware Executables

Kirti Mathur
M.Tech. Scholar,
Department of CSE
Rajasthan Technical University, India.

Saroj Hiranwal
Reader
Department of CSE
Sri Balaji College of Engineering &Technology,India.

Abstract: The computer technology has emerged as a necessity in our day to day life to deal with various aspects like education, banking, communication, entertainment etc. Computer system's security is threatened by weapons named as malware to accomplish malicious intention of its writers. Various solutions are available to detect these threats like AV Scanners, Intrusion Detection System, and Firewalls etc. These solutions of malware detection traditionally use signatures of malware to detect their presence in our system. But these methods are also evaded due to some obfuscation techniques employed by malware authors. This survey paper highlights the existing detection and analysis methodologies used for these obfuscated malicious code.

Keywords: Malware, Polymorphism, Metamorphism, Evasion, Anti-Evasion etc.

I. Introduction

Malware is a malicious code that propagates over the connected systems in network. This scenario is increasing day by day with advanced computing technology and communication network. Malware can be considered as the entity in which new features can be easily added to enhance its dark side effects in the form of various attacks. These malwares can be dangerous with all their side effects on the infected machines like disabling malware detectors or AV Scanners which installed for the security purposes [1].According to statistics, 70-80% of the malware comes from popular sites [2]. The number of malware has grown rapidly. The rate of malware attacks and security solutions is not yet leveling. In fact, according to O'Farrell (2011) and Symantec Global Internet Security Threat Report Trends for 2010 (Symantec, 2010), attacks against Web browsers and malicious code variants installed by means of these attacks have increased [3]. This paper describes different kinds of methodologies, detection and analysis techniques for handling threats in form of malwares for our presently working machines as found in relevant literature. The paper is organized as follows: Section II presents the classification of malwares, whereas in Section III, we discuss the variants of malwares related to the study. Section IV illustrates the detection process of malwares by explaining different detection techniques which followed by Section V that elaborates malware analysis techniques. Section VI presents the evasion scenario while in section VII we deal with anti- evasion approaches of malwares. In Section VIII survey of existing work has been presented. Section IX outlines our proposed work which finally concluded in section X.

II. Classification

According to [4] malware is classified into three generations based on their payload, enabling vulnerability & propagation mechanism. First Generation malwares carry properties of virus which replicates or propagates by some human actions, emails and file sharing whereas malware shares properties of worms in Second Generation. These are hybrid in nature with some features of virus and Trojans which do not need any human intervention for replication. In Third Generation malwares are geographical region or organization specific. These malwares employs multiple attack vectors and usually attacks security technologies and products.

Further we can see various classifications in general for malware [5]:

1. Viruses- computer programs which replicate themselves and infect various system's files.
2. Worms- are also self-replicating computer software that is able to send itself to other computers on a network or the Internet.
3. Trojans- a software program that emulates behavior of an authentic program and hijacks user password to gain control of system remotely.
4. Spyware- software which is installed on a computer system without any concern of the user to collect all his personal and confidential information.
5. Rootkits- are designed to take control of infected machine by gaining administrator access of the system. The name comes from the term root under UNIX.

6. Botnet- A botnet is remotely controlled autonomous software. It is usually a zombie program which is controlled for any network infrastructure.
7. Adware- also called as advertising-supported software whose functionality is to displays or downloads the advertisements to a computer after the installation of malicious software or application.

III. Variants Of Malware

In malware detection, we can have our major focus for the different variants produced or already exist in relevance of the security threats they pose to the system. A detailed discussion about malware variants is given as follows:

I. Polymorphic Malware

If a virus is programmed to look different each times it replicated, but keeping the original code intact. Such a virus is known as polymorphic virus. A polymorphic malware consists of encrypted malicious code along with the decryption module. Polymorphic code is a method now commonly implemented in malware that uses a polymorphic generator to mutate the code while keeping the original algorithm intact.

A typical implementation of a polymorphic code is to encrypt malware and include the encryptor/decryptor within the code. Polymorphic malwares have specially designed mutation engines. These Polymorphic viruses have a constant payload which is being encrypted with a different decryptor at each instance. Just by switching the order of instructions, a new decryption routine generated by the Mutation Engine. Polymorphic viruses typically encrypt the body of the virus and front-end it with a variable decryption routine. Thus, the body cannot be scanned because it is encrypted, and mutation engine is capable of generating too many different decryption routines [6] [7].

II. Metamorphic Malware

These are a kind of body-polymorphic, where body of virus itself changes from one instance to another. Metamorphic malwares use different types of obfuscation techniques to reprogram themselves into a new transformed code, i.e. similar to original code. The metamorphic nature of the malware enables malicious code to mutate while spreading across the network and making signature based detection completely ineffective [1]. First metamorphic malware recovers its base version and then it incorporates a different metamorphic payload within the base version to evade detection. Various Techniques employed in metamorphism:

- Disassembly-Depermutation / Shrinking
- Expansion
- Permutation
- Assembling
- Other transformations

IV. Detection: How To Identify A Malicious Executable?

It is necessary to understand the behavioral aspects of a malware which only possible by executing malware binary. Execution of malware binary can have normal layouts or traces or may show signs of some abnormalities with abrupt termination in execution. Detection is all about identifying whether code is genuinely benign or malicious. Robust malware detection depends on the capability of handling obfuscated malware efficiently. Code obfuscation changes malware syntax but not its intended behavior, which has to be preserved. Both reverse engineering and deobfuscating techniques generally begin with some sort of static program analysis, which can be depicted as an abstraction of program semantics [8].

For this purpose we have a component called as Malware detector, which can be defined as a system that attempts to identify malware using signatures and other heuristics parameters [1]. Two common obfuscation techniques are Polymorphism and Metamorphism. The malware detector can be concluded as the safeguard interface of our system whose functionality depends upon the behavior of executables and not of its own residence on the system. Two input components for malware detector [9]-

- ✓ Signature or behavioral parameters of given code.
- ✓ Executable code under inspection

So, when malware detector has been provided with above two input, it can employ its detection mechanism to conclude about the code as malware or benign.

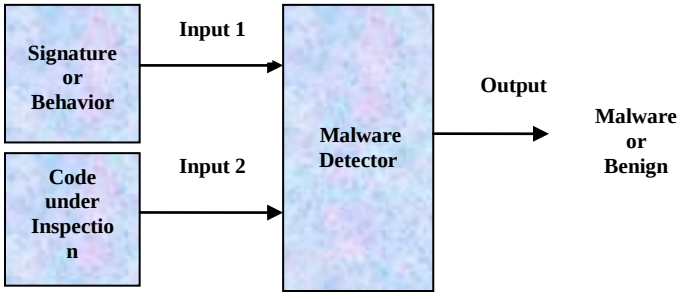

Fig 1. Malware Detector

I. Methodology of Malware Detection

Two techniques/methods employed for malware detection [9]:

- **Signature-Based Techniques-** Most of the antivirus tools are based on detection with signature based techniques. These signatures are created by examining the disassembled code of malware binary. Various disassemblers and debuggers are available which help in disassembling the portable executables. Disassembled code is analysed and features are extracted. These features are used in constructing the signature of particular malware family.
- **Behavior-Based Techniques-** Here mainly the goal is to analyze the behavior of known or unknown malwares. Behavioral parameters include various factors such as source/ destination address of malwares, types of attachments and other countable statistical features.

Consequently each of the above technique can be further applied using static analysis, dynamic analysis or hybrid analysis.

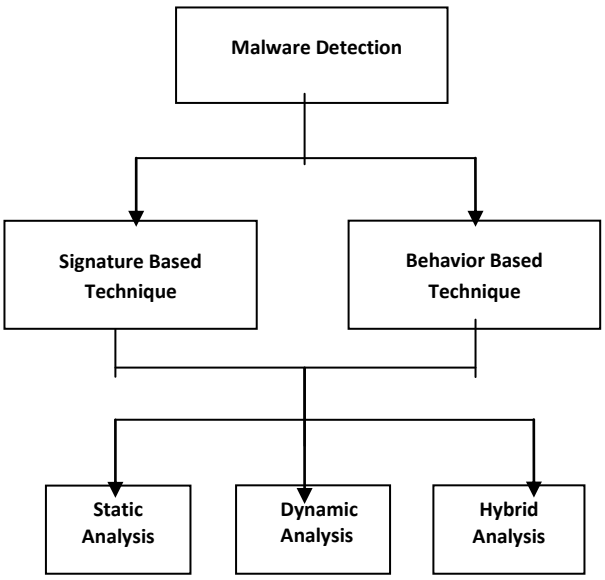

Fig 2. Malware Detection and Analysis

V. Malware Analysis Techniques

The detection process accompanied by signature based and behavior based techniques which further accomplished the whole process with static, dynamic and hybrid analysis.

I. Static Analysis

It is the process of analyzing executable code without actually executing the file. In static analysis we extract the low level information from codes which gathered by decompiling or disassembling the codes with the use of any disassembler tools. Static analysis has the advantage to reveal the code structure of the program under consideration. And this is only due to the ease of examining various parts of a program which normally do not executed. Static analysis may fails in analyzing unknown malware that uses code obfuscation techniques. Researchers for performing static analysis of malwares must possess a good knowledge of assembly language and the working operating system. Program analyzers, disassemblers and debuggers can be treated as working tools in Static analysis. The major advantage of static analysis over its dynamic counterpart is that its free from the overhead of execution time [10].

II. Dynamic Analysis

It is also called as behavioral analysis, involves executing the malware and monitoring its behavior, system interaction, and the effects on the host machine. In dynamic analysis, infected files are analyzed in simulated environment like a virtual machine, simulator, emulator, sandbox etc. As the malware writers make use of anti-emulation and anti-virtual machine tools for hiding the malware functions, so make the working environment invisible to the malware [1]. Dynamic analysis may fails to detect an activity which shows the behavioral changes in codes by different trigger conditions during the course of its execution.

Table I: Static Analysis vs. Dynamic Analysis [3]

Static Analysis	Dynamic Analysis
Fast and safe	Easy detection of unknown malware.
Difficult analysis of unknown malware	Difficult to analyze multipath malware
Easy analysis of multipath malwares	Neither fast nor safe

III. Hybrid Analysis

Hybrid analysis includes the combinatorial approach of both dynamic and static analysis. It firstly analyses about the signature specification of any malware code then combine it with the other behavioral parameters for enhancement of complete analysis. Due to this approach hybrid analysis overcomes the limitations of both static and dynamic analysis.

VI. Evasion Of Malware

The term 'evasion' literally explains how to protect or escape the one from some unusual or unfavorable scenarios [7]. In malwares there are different evasion approaches to evade the malcodes from external antivirus/antimalware scanners.

- **Code obfuscation[1][4]:** In code obfuscation the main goal is to hide the underlying logic of the program so as to prevent the others from having any related knowledge of the code. The malicious code remains incomprehensible and all its harmful functionality whenever activated. When we apply some obfuscation transformations to a code, then it results in a kind of self-decrypting encryption.
- **Packing:** Packing refers to encrypt or compress the executable file. In Packing, original code remains hidden till the runtime or the unpacking process of executable codes which results in the immunity of code for static analysis. Packed malware codes can be treated as subset of obfuscated codes which are compressed and cannot be analyzed so, consequently unpacking phase is necessary to reveal the overall semantic of the code [11].

VII. Anti- Evasion Approaches Of Malware

Two approaches which are used as just reverse of evasion techniques are:

• Deobfuscation

• Unpacking

In deobfuscation firstly there is focus to identify about the status of obfuscation and also the helping packer used. Next is to defeat the protection and restoring the de-obfuscated form of the code. For these above mentioned processes there are some of the available packers/unpackers to be used as open source. For example -UPX Packer. One of the other method used for deobfuscation by analysts known as dumping of PE files traced from memory [12].

VIII. Survey Of Existing Work

In existing scenario we have malware detection and analysis as the foundation of this paper work. So we surveyed about malware detection and analysis approaches which has been summarized in the table given below:

S. No.	Title	Author Name	Year of Publication	Techniques	Feature	Conclusion
1.	Automatic Malware Detection Using Common Segment Analysis and Meta-Features	Tahan et al.[3]	2012	Machine Learning techniques	Suggestion to use meta features instead of plain features like n-grams for malware detection	The goal of this paper is to develop and evaluate a novel methodology and supporting algorithms for detecting malware files by using common segment analysis
2.	Survey on malware evasion techniques: state of the art and challenges	Marpaung et al. [7]	2012	Evasion and Mitigation techniques	Various evasion and mitigation approaches. Robust technique as CFI	Evasion and mitigation techniques used are surveyed. Also comparison of evasion techniques with the parameters related to attacks and difficulties for detection.
3.	Malware Detection Based on Hybrid	Elhadi et al. [9]	2012	Static and dynamic Analysis technique	Combination of static and	Proposed Framework

	Signature Behaviour Application Programming Interface Call Graph				dynamic analysis with signature and behavior based techniques	combines signature-based and Behavior-based analysis using API Graph System
4.	An Introduction to Malware	Robin Sharp et al. [5]	2012	Thesis	Types of malwares discussed with basic principles of operation and dissemination	Introduction about malware
5.	Malware Threats & Mitigation Strategies : A Survey	Rehmani et al. [2]	2011	Malware designs and latest attack models	Latest mitigation strategies studied	The different attack models of the malware have been discussed so as to find out best mitigation strategy depending on the type of attack.
6.	Malware Obfuscation Techniques: A Brief Survey.	You et al. [6]	2010	Obfuscation Techniques	Obfuscation techniques reviewed by studying encrypted, oligomorphic, polymorphic and metamorphic malwares	Study of Various obfuscation techniques used by malwares to evade detection.
7.	Comparision and Benchmarking of Automatic Malware Unpacking Techniques	Getu et al.[11]	2010	Thesis	It basically starts by assessing research works related to malware analysis and detection, and focus on packed malware analysis techniques.	This thesis work, compares and benchmark currently existing automatic malware Unpacking techniques, and explores new approaches to design automated malware unpackers.
8.	Malware Analysis	Saffaf et al. [12]	2009	Thesis	forensics techniques for fighting current and future malware.	The purpose of this thesis is to identify the methods and tools used by anti-virus firms to protect Internet's users from the threats of malware.

9.	Survey on Malware Detection Methods	p. vinod et al. [1]	2009	Series of Malware detection techniques	Problems related to traditional signature based detection method is highlighted	Challenge lies in the development of good disassembler, similarity analysis algorithm so that the variants of malwares can be detected in shorter time there by reducing the computation overhead.
10.	Malware Detection via Classifying With Compression	Gong et al. [13]	2009	Malware detection with compression	Adaptive Data Compression for Malware detection	Method proposed for malware detection using compression which results with high accuracy and low false positives rate.
11.	Data Mining Methods for Malware Detection	Siddiqui et al. [10]	2008	Thesis	Investigation of data mining methods for malware detection	Proposed a framework as an alternative to traditional signature based detection methods.
12.	Code Obfuscation and Malware Detection by Abstract Interpretation	Preda et al. [8]	2007	PhD Thesis	Malware detection based on program semantics and abstract interpretation	Formal approach to code obfuscation and malware detection
13.	Static Analyzer of Vicious Executables (SAVE)	Sung et al.[4]	2004	Signature based malware detection	The hypothesis is That all versions of the same malware share a common core signature that is a combination of several features of the code.	Emphasis is on detecting metamorphic or polymorphic malware

IX. Our Proposed Approach

Here as we carry our work to check maliciousness of binary using "Bi-features". The proposed approach will be implemented using three classification algorithms viz. Random Forest, Instance based classifier using k nearest neighbours and naive bayes algorithm. In this approach, Bi-features will be used to enhance the accuracy of existing Static Mono-features analysis. Bi-feature is a feature composed using two features to make the feature analysis to be more absolute for any code under inspection. It will reduce the false positives occurred with Mono-feature analysis. To check the accuracy of the results parameters like true positive, false positive, true negative, false negative will be used. Generally when a virus scanner detects a 'virus' in non-infected file erroneously, it is called as false positive occur at this instant. While when a virus scanners fails

to detect a virus in an infected file, it is the occurrence of false negative. So with the combination of Static Bi-feature analysis approach, fast, accurate and safe results implemented during malware detection and analysis.

X. Conclusion

Malware is posing a threat to user's computer systems in terms of stealing personal and private information, corrupting or disabling our security systems. This paper highlights some existing methodologies incorporated by security researchers to tackle these threats. Our survey paper explains about static, dynamic and hybrid malware analysis techniques. Static analysis extract feature without actual execution so safer than dynamic analysis in which code is executed actually. These techniques handle various kind of structurally different malware like polymorphic, metamorphic, packed etc, which are produced by employing obfuscation techniques. Some shortcomings of dynamic analysis like single execution path, significant performance overhead etc make static analysis more preferable than dynamic. Mainly the focus of this paper is on implementing such an analysis technique for malwares which overcome the limitations associated with the existing counterpart techniques. So, static Bi-feature analysis cope up with many of the limitations associated with static (Mono-feature) analysis or dynamic analysis.

References

[1] Vinod P. V.Laxmi,M.S.Gaur: Survey on Malware Detection Methods, 3rd Hackers' Workshop on Computer and Internet Security, Department of Computer Science and Engineering, Prabhu Goel Research Centre for Computer & Internet Security,IIT, Kanpur, pp-74-79, March,2009.

[2] Rizwan Rehmani , G.C. Hazarika and Gunadeep Chetia : Malware Threats and Mitigation Strategies: A Survey, Journalof Theoretical and Applied Information Technology, Vol. 29 No.2, 2011. Retrieved on: March, 15, April , 2013 http://www.jatit.org/volumes/researchpapers/Vo l29No2/3Vol29No2.pdf

[3] Gil Tahan, Lior Rokach and Yuval Shahar: Mal-ID: Automatic Malware Detection Using Common Segment Analysis and Meta-Features, Journal of Machine Learning Research pp- 949-979, 2012.

[4] A. H. Sung, J. Xu, P. Chavez and S. Mukkamala: Static Analyzer of Vicious Executables (SAVE), Proceedings of the 20th Annual Computer Security Applications Conference (ACSAC'04), IEEE.

[5] Robin Sharp, An Introduction to Malware, Spring 2012. Retrieved on April, 10, 2013 http://orbit.dtu.dk/fedora/objects/orbit:82364/datastreams/file_4918204/content

[6] Ilsun You and Kangbin Yim: Malware Obfuscation Techniques: A Brief Survey, International Conference on Broadband, Wireless Computing, Communication and Applications, 2010.

[7] Jonathan A.P. Marpaung, Mangal Sain and Hoon-Jae Lee: Survey on malware evasion techniques: state of the art and challenges, International Conference of Advanced Communication Technology, pp 19-22, 2012.

[8] Mila Dalla Preda: Code Obfuscation and Malware Detection by Abstract Interpretation Universit`a degli Studi di Verona, Dipartimento di Informatica, TD-02-07, 2007.

[9] Ammar Ahmed E. Elhadi, Mohd Aizaini Maarof and Ahmed Hamza Osman :Malware Detection Based on Hybrid Signature Behaviour Application Programming Interface Call Graph, American Journal of Applied Sciences 9 (3): 283-288, 2012.

[10] Muazzam Ahmed Siddiqui:Data Mining Methods for Malware Detection: University of Central Florida, 2008.

[11] Tewfik Adem Getu: Comparision and Benchmarking of Automatic Malware Unpacking Techniques, Politecnico di Milano, 2009-2010.

[12] Mohammad Nour Saffaf: Malware Analysis Bachelor's Thesis., Helsinki Metropolia University of Applied Sciences, May 27, 2009.

[13] Tao Gong, Xiaobin Tan and Ming Zhu: Malware Detection via Classifying With Compression, The 1st International Conference on Information Science and Engineering, (ICISE), 2009

