ISSN:0975-9646

A Review of Malware Detection Based on Pattern Matching Technique

Manish Kumar Sahu , Manish Ahirwar , A.Hemlata

Dept. of Computer Science & Engineering,
University Institute of Technology,
RGPV, Bhopal, India

Abstract— **Malware detection developer faced a problem for a generation of new signature of malware code. A very famous and recognized technique is pattern based malware code detection technique. This leads to the evasion of signatures that are built based on the code syntax. In this paper, we discuss some well known method of malware detection based on semantic feature extraction technique. In current decade, most of authors focused on malware feature extraction process for generic detection process. The effectiveness of the signature based technique for malware detection invites for moderation and improvement of the current system and method. Some authors used rule mining technique, some other used graph technique and some also focused on feature clustering process of malware detection.**

Keywords— **Malware, Signature based dynamic pattern .**

I. INTRODUCTION

The current decade of information technology and data science suffering from threats of malware, Malware is collection of virus, zombie, worm, Trojan horse and much malicious software. The variant of malware software is self propagated and invited some attack in the target computer. Malware application software damage any executable program, data and operating system files [1,4]. It generates more traffic in the network and produce a denial of service. When the user runs damage file of the program, it store in memory and damage some anther program and file store in memory. The process of operating system openness, malware application program control task management of operating system

A propagation of malware software and reduction of damage control some malware detector is available in market [9]. A malware detector is a program that attempts to categorize malware application program. A virus scanner uses signatures and other heuristics to discover malware, and thus is an example of a malware detector. Given the confusion that can be caused by malware, malware detection is an important task in current scenario. The aim of a malware developer is to modify or morph their malware to shirk detected by a malware detector. A common technique used by malware developer to evade detection is a program obfuscation [11, 12]. Polymorphism and metamorphism are two common obfuscation techniques used by malware developer. Malware detection method used to detect or identify the malware software. Generally, malware detection technique can be classified into pattern-based detection, abnormal-based detection and rule-based detection [15]. Signature-based or sometime called as misuse detection as maintain database of known software technique and detects software by comparing behavior against the database. It shall require less amount of system resource to detect the software. It also claimed that this technique can detect known attack accurately. However the disadvantage of this technique is ineffective against previously unseen attacks and hence it cannot detect new and unknown software methods as no signatures are available for such attacks [14]. Abnormal-based detection analyses user behavior and the statistics of a process in normal condition, and it checks whether the system is being used in a different approach. That attack will result in behavior different from that normally observed in a system and an attack can be detected by comparing the current behavior with pre-established normal behavior. Specification-based detection will rely on program specifications that describe the intended behavior of security-critical programs [24]. The aim of the policy specification language to provide a simple way of specifying the policies of privileged of detection programs. It monitors execution program involve and detecting deviation of their behavior from the specification, rather than detecting the occurrence of specific attack patterns. The current scenario of research focus on hybrid signature based malware detection. The hybrid method is a combination of static and dynamic approach of pattern detection. The process of pattern detection also focusses in feature extraction of the malware file for detection process and apply some graph theory and mining technique for detection technique [25]. The above section discuss the introduction of malware and malware detection. In section II describe related work of malware detection. In section III compression of malware detection technique. In section IV discuss types of malware and finally conclude in section V.

II. RELATED WORK

In this section describe some recent related work in the field of malware detection using the pattern based technique. Pattern based technique follows several approaches such as data mining, soft computing and tree based technique. These techniques used for the process of feature extraction of malware code. Some method related to in this paradigm discuss here.

[1] In this paper author proposed an intelligent instruction sequence based malware classification system based on the weighted subspace clustering method. This system uses a novel feature for malware characterization which is the function based instruction sequence segments. This method integrates the advantages of function and NGrams, and overcome the drawbacks of N-Grams which include noise information and are much more time-consuming. The proposed model an integrated system consisting of three major modules: feature exactor, malware categorizer using weightedsubspace clustering methods and a malware signature generator.

[2] In this paper author proposed Bin Graph, a new mechanism that accurately discovers metamorphic malware. Bin Graph leverages the semantics of the malware, since the mutant malware is able to manage their syntax only. To this end, we initial take out API calls from malware and convert to a hierarchical behavior graph that represents with identical 128 nodes based on the semantics. The sub graph analysis can provide not only metamorphic malware classification, but also behavior analysis of modules used by malware variants. Such information can be useful to AV vendors and make the malware authors harder to develop metamorphic malware. To represent the semantics of a binary, Bin-Graph construct a behavior graph using the API call sequence. Therefore, the extracting accurate call sequence is required.

[3] In this paper author analyzed the process of malware detection in terms of positive as well as a negative security factor in a design framework for combating malware threats, in mission critical environment, usage of the specification based application behavior white listing is more effective. Also with the growing attacks by exploiting software vulnerabilities, the end user wants an assurance that functionality is implemented correctly and software provides only the desired features. The application behavior white listing will also help to provide formal security assurance of IT systems.

[4] In this paper author proposed a novel method to develop a malware signature that is resistant to obfuscation method. The proposed signature is based on kernel object characteristics while avoiding dependency on specific malicious code information that may utilize to evade created signatures. In addition, a method is proposed to recognize kernel object's features that effectively contribute to the development of a robust malware detection signature. Kernel object profile and an invariant detection technique are, also, proposed to support the procedure of evasion-resistant signature development. To support the proposed techniques, a sample tool is developed to generate malware detection signatures based on obtaining profiles.

[6] In this paper author proposed a method for metamorphic malware detection using MSA methods. Signature for a malware family is extracted and tested using the unseen samples. It finds that the unseen samples were detected using signatures with low false positives. Also, the detection rate of implementation method is comparable with that of antivirus like Avast, Avira, AVG. Some of the undetected malware executables from all commercial antiviruses were detected by signatures generated using the proposed method.

[10] In this paper, the authors evaluate the effects of the post processing techniques (e.g., rule pruning, rule position, and rule choice) of associative classification in malware detection and propose an efficient way, i.e., CIDCPF, to detect the malware since the "gray list." To the best of our knowledge, this is the first author used post processing techniques of associative classification in malware detection. In addition, to IMDS system, which adopts the CIDCPF method for building classifiers can greatly reduce the number of generated rules and make it easy for our virus analysts to identify the useful ones.

[11] in this paper authors discuss Soft computing techniques for malware detection. These techniques have the ability of learning from the past incidences and can categorize normal and abnormal behavior. A review of the application of these soft-computing techniques in malware detection has also been given in this paper. Despite of so much study, techniques with good accuracy and low false alarm rate are still needing attention.

[15] In this survey a series of malware detection techniques have been given. The problems related to traditional signature based detection method is also highlighted. The rate of new malware's causing distractions to systems worldwide is increasing at an alarming rate. Detection of malware's changing their signatures, frequently has posed many open research issues. The challenge lies in the development of good disassemble, similarity analysis algorithm so that the variants of malware's can be detected in shorter time thereby reducing the computation overhead.

III. Comparative Study

The development of malware detector is challenging job in the current scenario of information technology. Because the developer of malware code is one step ahead of malware detected, now a researcher of malware detector is going on and focuses on dynamic signature generation for detection of malware pattern. In the generation of dynamic pattern various authors used signature based technique for enhancement purpose. Here describe some comparative study of malware detection technique based on signature based.

Table-I gives the comparative study of malware code detection technique in current scenario.

TABLE I
COMPARATIVE STUDY OF MALWARE CODE DETECTION TECHNIQUE

Method	Advantage	Disadvantage	Problem
ISMCS	Instruction sequence based technique, very effective and fast detection.	Feature extraction of malware code is not generic for analysis	The extracted feature grouped on the basis of 2-dimensional relation attribute.
Bin-Graph	The detection rate of malware family of virus is very high	Apply semantic signature process for feature extraction.	Semanticsignatures presenting common behavior acrossvarious kinds of malware families
Profiling Kernel Data Structure Objects	Signature is more resistant toobfuscation methods and resilient in detecting malicious codevariants.	Signature based detection can be bypassed using malicious code obfuscation, used signature are vulnerable to manipulate and tampering by malicious code	Profiling of data structure is changed the sequence of code generation process.
Momentum	Method usingbioinformatics techniques effectively used for Protein andDNA matching. Instead of using exact signature matchingmethods, more sophisticated signature(s) areextracted usingmultiple sequence alignment (MSA).	Compromised with new generated pattern for malware developer	The processing of data in form of train and test take more time for computational task execution.
CIMDS	New method of post processing of malware detection technique.	The rule based classification technique only focus on categorized rule of malware feature.	CIMDS currently only provides binary predictions, i.e., whether a PE file is malicious or not

TABLE III
CATEGORIES OF MALWARE SOFTWARE

Malware categories	Propagation	Infection	Self-defense	Capabilities
Key logger	Infected websites and/or USB or other media	Vulnerable browsers or unpatched OS or application	Replace IO device drivers or APIs	Collect user keystrokes including credentials
Rootkit[20]	Infected websites and/or installs on servers by hackers or insiders	Exploited trusted admin access, vulnerable browsers, or unpatched OS or application	Replacing OS kernel-level API routines	Collect data and impersonate user activity for entire machine and its interfaces
Flaw Exploits [19]	Execution of unexpected commands to flawed software by remote hackers	Vulnerable software-to-database and command execution interfaces	Impersonation of authorized user	Download or upload data from data repositories between target and malware operator site
Bot [15]	Bots are generally delivered via infected websites, or links to malicious websites embedded in phishing email.	User may voluntarily install individual bots based on deceptive messages in email or web instruction, or via browser/OS vulnerabilities.	Bot updates security patches and anti-virus on machine to ensure stable operation and keep other bots out. Lays dormant until activated.	When activated by botnet operator, the operator may direct bot to execute a variety of standard or custom functions.
Denial of Service (DOS) [25]	IP packet delivery	Internet protocols that automate packet processing	Simultaneously attack from multiple sources	Consume computing resources on targets

IV. CATEGORIES OF MALWARE

Malware forms a roll of the verity of software application. It activated may be hardware and software applications such as file system desktop and programmable device. Sophisticated attacks have confirmed data can be stolen through well written malware residing only in system memory without leaving any footprint in the form of important data. Malware has been known to disable information security protection mechanisms such as desktop firewalls and anti-virus programs. Some even have the ability to subvert certification, validation and inspection functions. It has configured initialization files to maintain persistence even after an infected system is rebooted. Upon execution, sophisticated malware may self-replicate and/or lie dormant until summoned via its command features to extract data or erase files. Here we describe categories of malware software in, form of the table.

V. CONCLUSION AND FUTURE WORK

In this paper have reviewed and analyzed the existing malware detection techniques and compare it with the advantage and disadvantage and also discuss some current problem are still remain. From the analysis researcher has focused a new dynamic feature extraction of malware detection techniques. Some process of technique based on rule mining and some other are based self propagated feature extraction technique. This will contribute ideas in malware detection technique field by generating an optimize method for malware detection. Moreover, although crimeware and state-sponsored cyber attacks and campaigns are the most visible form of attack, FIs should recognize the increasing threat from both external and internal sources, and take practical measures to detect and defend against potential internal malware interference with business process. Is should evaluate their vulnerability to the malware described in this paper

REFERENCES

[1] Kai Huang, Yanfang Ye, Qinshan Jiang" ISMCS: An Intelligent Instruction Sequence based Malware Categorization System" in IEEE, 2012,PP 65-78.

[2] Jonghoon Kwon, Heejo Lee" BinGraph: Discovering Mutant Malware using Hierarchical Semantic Signatures" in IEEE, 2012 PP 123-130.

[3] P.R.Lakshmi Eswari, N.Sarat Chandra Babu" A Practical Business Security Framework to Combat Malware Threat" World Congress on Internet Security 2012 IEEE, Pp 57-65.

[4] Ahmed F.Shosha, Chen-Ching Liu, Pavel Gladyshev "Evasion-Resistant Malware Signature Based on Profiling Kernel Data Structure Objects" in 7th International Conference on Risks and Security of Internet and Systems 2012 , PP 78-86.

[5] Hira Agrawal, Lisa Bahler, Josephine Micallef, Shane Snyder, and Alexandr Virodov "Detection of Global, Metamorphic Malware Variants Using Control and Data Flow Analysis" in IEEE, 2013 PP 120-127.

[6] Vinod P., V.Laxmi, M.S.Gaur, Grijesh Chauhan" MOMENTUM:MetamOrphicMalware Exploration Techniques Using MSA Signatures" International Conference on Innovations in Information Technology (IIT) IEEE, 2012 PP 232-238..

[7] Robiah Y, Siti Rahayu S., Mohd Zaki M, Shahrin S., Faizal M. A., Marliza R. "A New Generic Taxonomy on Hybrid Malware Detection Technique" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 5, No. 1, 2009, PP 56-61.

[8] Aubrey Derrick Schmidt, Rainer Bye, Hans Gunther Schmidt, Jan Clausen "Static Analysis of Executables for Collaborative Malware Detection on Android" 2009 IEEE, Pp 230-238.

[9] Kevadia Kaushal, Prashant Swadas, Nilesh Prajapati "Metamorphic Malware Detection Using Statistical Analysis" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307 (Online), Volume-2, Issue-3, July 2012, PP 49-56.

[10] Yanfang Ye, Tao Li, Qingshan Jiang, and Youyu Wang "CIMDS: Adapting Postprocessing Techniques of Associative Classification for Malware Detection" IEEE Transactions On Systems, Man, And Cybernetics Part C: Applications And Reviews, Vol. 40, No. 3, May 2010 Ieee, Pp 298-309.

[11] Raman Singh,Harish Kumar, R.K. Singla"Review of Soft Computing in Malware Detection" in IJCA 2009, PP 54-61.

[12] Mihai Christodorescu, Somesh Jha "Semantics-Aware Malware Detection " in IEEE, 2009 PP 120-135.

[13] Pranith Kumar D, Anchal Nema, Rajeev Kumar "Hybrid Analysis of Executables to Detect Security Vulnerabilities "in IEEE, 2009 PP 56-67.

[14] Kent Griffin Scott Schneider Xin Hu Tzi cker Chiueh "Automatic Generation of String Signatures for Malware Detection "in Symantec Research Laboratories 2008, PP 1-29.

[15] Vinod P., V.Laxmi,M.S.Gaur "Survey on Malware Detection Methods" 2008, PP 56-64.

[16] Roland Cheung "Document Malware Attacks" HKCERT - Information Security Seminar (March) PP 1-44.

[17] X. Hu, T. Chiueh, and K. G. Shin, "Large-scale malware indexing using function-call graphs," ACM Conference on Computer and Communications Security, 2009, PP. 611-620.

[18] L. Yujian and L. Bo, "A Normalized Levenshtein Distance Metric," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol, 29, no.6, 2007, PP 1091-1095.

[19] G. R. Thompson and L. A. Flynn, "Polymorphic Malware Detection and Identification via Context-Free Grammar Homomorphism," Bell Labs Technical Journal, vol. 12, no. 3, 2007, PP 139-147.

[20] L. A. Gordon, M. P. Loeb, W. Lucyshyn, and R. Richardson. 2004 CSI/FBI computer crime and security survey. Technical report, Computer Security Institute, 2004.

[21] Guillaume Bonfante, Matthieu Kaczmarek, and Jean-Yves Marion. Architecture of a Morphological Malware Detector. Computer Virology, 2009, PP 263–270.

[22] Matthieu Kaczmarek Guillaume Bonfante and Jean-Yves Marion. Control Flow Graphs as Malware Signatures. 2007, PP 235-243.

[23] Christopher Kruegel, Engin Kirda, Darren Mutz, William Robertson, and Giovanni Vigna. Polymorphic Worm Detection using Structural Information of Executables. In RAID, Springer-Verlag, 2005, PP 207-226.

[24] Heejo Lee Kyoochang Jeong. Code Graph for Malware Detection. In International conference on Information Networking, ICOIN, in IEEE, 2008, PP 1-6.

[25] Lin, Da and Stamp, Mark, Hunting for undetectable metamorphic viruses, In Journal Computer Virology, volume (7), issue (3), August, 2011, PP. 201–214.