Faculty of Computer Science and Information Technology

A Dynamic Approach using Indicators of Compromise to Detect Malicious Code

Chiadighikaobi Ikenna Rene

Master of Science
2018

A Dynamic Approach using Indicators of Compromise to Detect Malicious Code

Chiadighikaobi Ikenna Rene

A thesis submitted

In fulfilment of the requirements for the degree of Master of Science

(Computer Science)

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK
2018

UNIVERSITI MALAYSIA SARAWAK

Grade: _____

Please tick (√)

Final Year Project Report	☐
Masters	☑
PhD	☐

DECLARATION OF ORIGINAL WORK

This declaration is made on the ……..day of……..2016.

Student's Declaration:

I Chiadighikaobi Ikenna Rene, 1602003. Computer Science and Information Technology ----------

(PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled, A Dynamic Approach using Indicators of Compromise to Detect Malicious Code is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

_____ Chiadighikaobi Ikenna Rene (1602003)

Date submitted Name of the student (Matric No.)

Supervisor's Declaration:

I---------- (SUPERVISOR'S NAME) hereby certifies that the work entitled, ---------- (TITLE) was prepared by the above named student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of ---------- (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work

Received for examination by: _____ Date:_____

(Name of the supervisor)

I declare this Project/Thesis is classified as (Please tick (√)):

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*****
☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*****
☑ **OPEN ACCESS**

Validation of Project/Thesis

I therefore duly affirmed with free consent and willingness declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies for the purpose of academic and research only and not for other purpose.
- The Centre for Academic Information Services has the lawful right to digitise the content to for the Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic exchange between Higher Learning Institute.
- No dispute or any claim shall arise from the student itself neither third party on this Project/Thesis once it becomes sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student except with UNIMAS permission.

Student's signature lyke _____ Supervisor's signature: _____
(Date) (Date)

Current Address:
2nd Floor, Lot 13939, Sl.5 Jln Keranji Tabuan Square, 93350 Kuching Sarawak, Malaysia

Notes: ***** If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the period and reasons of confidentiality and restriction.

[The instrument was duly prepared by The Centre for Academic Information Services]

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citation, which have been duly acknowledged. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

_____

Name: Chiadighikaobi Ikenna Rene

Matric No. 16020033

Faculty of Computer Science and Information Technology

Universiti Malaysia Sarawak

ACKNOWLEDGEMENT

First, I would like to thank, praise and give glory to the Almighty God for the grace, love and favor He bestowed on me during this program. This journey was encouraged by the wonderful family God gave me, they showed their support through love, provision, financial assistant and most importantly they laid the foundation for the progress. I like to extend an appreciation to my Supervisor, Assoc. Prof. Dr. Johari Bin Abdullah for his guidance and his guidance. I would like to appreciate all my friends and everyone who showed love and care to me during the course of this journey.

ABSTRACT

Malicious activities (malcode) are self replicating malware and a major security threat in a network environment. Timely detection and system alert flags are very essential to prevent rapid spreading of malcode in the network. Automatic signature generation systems has likewise been use to address the issue of malcode, yet there are many works required for good detection. Based on the behavior way of malcode, a behavior approach is required for such detection. In this thesis a dynamic approach technique is proposed for malcode detection and rapid malcode behavior rules are automatically generated based on their Indicator of Compromise (IOC) behavior, as this approach is achieved using Weka system for clustering technique, T-Pot for intrusion data collection, Cuckoo Sandbox for malware data analysis and OpenIOC for IOC creation The experimental study in this thesis highlights the weakness in Signature-Based detection and static analysis of malcode data. The experimental study shows that the proposed approach using IOCRule achieved a detection rate of 87.50%, false negative of 12.50% when evaluated using CTU 2016/2017 Malware dataset. As the evaluation of CTU 2016/2017 Malware dataset achieved a detection rate of 1.18% and a false negative rate of 98.82%. This shows that the proposed approach achieved a much higher detection rate and lower false negative rate compared to the signature-based detection.

Keywords: Malware, indicators of compromise, IDS, malcode, dataset, honeypot.

Pendekatan Dinamik Menggunakan Petunjuk Kompromi untuk Mengesan Kod Berbahaya

ABSTRAK

Aktiviti berniat jahat (malkod) adalah malware yang menduplikasikan diri dan merupakan ancaman kepada keselamatan utama dalam rangkaian persekitaran. Pengesanan yang tepat pada masanya dan sistem amaran sangat penting untuk mencegah penyebaran malkod yang pantas dalam rangkaian. Sistem penjanaan tandatangan automatik juga telah digunakan untuk menangani isu malkod, namun terdapat banyak langkah yang diperlukan untuk pengesanan yang baik. Berdasarkan tingkah laku malkod, pendekatan tingkah laku diperlukan bagi pengesanan sedemikian. Di dalam tesis ini, teknik pendekatan yang dinamik dicadangkan untuk pengesanan malkod dan peraturan tingkah laku malkod yang cepat dijana secara automatik berdasarkan tingkah laku Indikator Kompromi (IOC), kerana pendekatan ini dicapai dengan menggunakan sistem Weka untuk teknik Kluster, T-Pot untuk pengumpulan data pencerobohan, Cuckoo Sandbox untuk analisis data malware dan OpenIOC untuk penciptaan IOC. Kajian eksperimen di dalam tesis ini menunjukkan kelemahan dalam pengesanan berasakan Tandatangan dan analisis statik data malkod. Kajian eksperimen menunjukkan bahawa pendekatan yang dicadangkan menggunakan IOCRule mencapai kadar pengesanan 87.50%, negatif palsu pada kadar 12.50% ketika dinilai menggunakan dataset CTU 2016/2017 Malware. Sebagai penilaian CTU 2016/2017, set data Malware mencapai kadar negatif palsu 98.82%. Ini menunjukkan pendekatan yang dicadangkan mencapai kadar pengesahan yang jauh lebih tinggi dan kadar negatif palsu yang lebih rendah berbanding pengesanan berasaskan tandatangan.

Kata kunci: *Malware, petunjuk kompromi, IDS, malkod, dataset, honeypot.*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

IOA	Indicators of Attack
IOC	Indicators of Compromise
IDS	Intrusion Detection System
IOCRule	Indicators of Compromise Rule
Malcode	Malicious Code
ML	Machine Learning
NIDS	Network Intrusion Detection System

CHAPTER 1

INTRODUCTION

1.1 Research Background

Networking occurs when there is an interconnection of two or more computers for communication purposes. Communication can be in many forms which includes the distribution of data from one device to another. This data or information is distributed within networks or from one network environment to another. The Internet, also called "the Net", is a connection of networks, where users at any networked device, if permitted, can receive and transmit information from any network point. Due to the amount of data and information available over the network, data theft has become part of the network activity, as network intrusion occurs on daily basis [1].

1.1.1 Network Security

Network security is an essential part of computer networks, as intrusion attack is a threat to network communication. Network intrusion happens when an unauthorized system or user gain access into a network system and manipulates data or information. "There are a total of 184 billion exploited detections, 1.8 billion average daily attack volume, 6,298 unique exploited detections and exploited volume per firm averaged at 2.5 million, with a median of 456, and 69% of firms saw severe attacks" [2]. Network security is an important aspect of networking, because a malicious code (malcode) attack is a threat to the network user.

Malcode is an application security threat that can be in the form of software system and cannot be efficiently controlled by traditional anti-virus software's and systems.

Network intrusion has emerged from known attack to unknown attack (Malcode), as there has been an exponential development in the Malcode family that the present security approach is not able to detect [3, 4]. Over the years, detection of Malcode or malware attack has been done using signature-based approach, whereby different techniques have been applied such as:-

Automating the generation of malware signatures using honeypot, Signature generator and by clustering or classifying the generated signatures and applying the honey pot data with machine learning techniques to generate signatures for intrusion detection [5-9].

Malcode utilizes sophisticated ways to make itself hidden from intrusion detection systems such as anti-malware software and infrastructures. Some malware activities remain undiscovered for years as they steal confidential data and also damage the system [10]. Malcode uses the Internet to call-back-home to communicate with the attack initiating server to receive new tasks and updates. When malcode like Bot-net tries to communicate with its Command and Control (C&C) centre, it uses a known network protocol to pass through network defenses' measures such as firewalls and anti-virus. Malicious programs are capable of hiding themselves or disabling their activity when they detect an attempt to discover them. Therefore, there is a need to use techniques that can detect malicious activities on systems. Some previous works have been done using the static approach of data analysis, as some of the research centered on analyzing network traffic as they focused on network layers and protocol, system files, file structures or on certain malware family [11, 12].

1.1.2 Malware Analysis

There are two general ways to deal with malware analysis, namely static analysis that studies the malware without executing it and dynamic analysis where the behavior of the malware is observed. As malware quickly advances the need for a liable detection system is crucial. Malware may include scanning abilities to the point that each infected host can additionally extend the botnet by exploiting obscure vulnerabilities in operating frameworks. In fact, even with an installed and newly refreshed anti-infection programming, the normal client could remain unnoticed since malware utilizes strategies to remain undetected. The test with obfuscation techniques was displayed in a Black Hat gathering [13], where it was expressed that detecting these sorts of malware is exceptionally troublesome continuous or posthumous analysis. Despite the fact that if the original malware is identified by hostile to infection application, an alternate variation will sidestep the regular example matching system since it yields an alternate example. It is moreover essential to gain information about their behavior to create exact identification plans.

Malware analysis is a technique that helps obtain information about a malware's behavior [14]. Regularly utilized approaches are static analysis that reviews the malware without executing it, and dynamic (behavioral) analysis which examine malware as they execute. Despite the fact that the two strategies may fulfill a similar objective of studying how malware functions, the tools, and aptitudes required are diverse [15]. Static analysis approach is done by analyzing the source code of the malware to understand how it works. Normally, static analysis utilizes string examining tools, such as disassemblers, debuggers and compilers to study source codes. Subsequent to applying these tools on the malware's executable, the investigator or malware examiner goes through the source code to gain

information on how the malware works for instance how it infects systems and transmits. The most common method for conducting a dynamic analysis is to run the malware and see what happens. Note that this approach is not without issues, since you may wind up destroying all information on your framework or let the malware spread if the relinquished host is associated with the Internet.

This thesis focus on malware data analysis, Indicators of Compromise (IOC) feature extraction, clustering and dynamic rules development for malicious code detection. This research is an improved version of the traditional intrusion-detection method, which deals with a static based signature created with String techniques [5, 16], as this research applies Machine learning techniques to cluster malcode behavior together for better IOC creation for intrusion detection. Indicators of compromise referred to as IOCs, consist of one or more artifacts that relate to a particular security incident or attack [17, 18]. The intent of assembling IOCs for a specific item of malware or malicious code is to state, with a relatively high degree of confidence, whether or not such items are present in a given environment. Based on the proposed research, IOC will be used to create a behavior rule for Network Intrusion Detection System (NIDS) for the detection of malcode activities. Over the years different type of systems such as Intrusion Detection System (IDS), Intrusion Prevention System (IPS) and Firewall have been implemented and established to address the issue of a security threat. Most of the system addresses known attack, using misuse techniques or (signature) pattern matching techniques. To lure intrusion attack for data collection in this research, honey-pot system will be implemented [19, 20]. In this research, the honey-pot will be deployed in an isolated environment, (a virtual machine), that consists of software components that constantly analyze the system events and record intrusion behavior.

Security should not be an out-of-the-box solution, as careful analysis of the environment at hand is needed before a solution can be provided. It is a step by step process and a thorough understanding of the system and constraints is needed. Intrusion Detection has been a research focus for long and there are still many issues that need to be addressed. As most IDS use static detection method, which is based on string signature [21], malicious code detection have been a challenging issue since it can incorporate a wide range of dynamic techniques, for example, indirect accesses and Trojan steeds. To distinguish conceivable vindictive practices of such malware can be a monotonous and testing assignment.

In other words, they are produced by impersonation or adjustment rather than advancement. A large portion of them had similar characteristics in relation to specific activities such as [21, 22, 23]:- A server connection programmed start-up, Framework registry access, Hostile to infection software turn off, and Bot software overhauls or uninstall. Furthermore, to perform specific behaviors, they have tendency to use the same function calls with a different structure. From the characteristics, the mining of similarity among system calls from many bot binaries derives the common semantic behavior to represent entire malicious code family for malware detection. The goal of this work is to identify a (metamorphic) Malcode using IOC and IOC-based rules detection method, incorporated with behavior analysis.

1.2 Problems Statement

According to Kaur and Singh [24], "the best intrusion attacks that keep away from discovery are Malcode as they do not demonstrate particular behaviors". Signature IDS lack the capability to detect Malcode (Unknown attack). The aim of this thesis is to detect

Malcode attack and identify attack features for intrusion IOCRule generation. Malcode attacks are vulnerabilities that are unpredictable at the time of attack, and it has high negative impact to the network environment [25, 26, 27].

Current research focus on improvement of intrusion signature generation for Intrusion Detection, as Signature-based Intrusion detection technique faces a challenge in high accuracy detection, due to its approach of detection [38, 67, 68, 71]. This technique uses static approach and intrusion signatures must exist in the Detection System database in other to enable detection of attack.

Signature-based Intrusion Detection System (IDS), which can be referred as pattern matching technique has been used by many organizations in detecting network threat, but it is unable to detect unknown (Malicious Code) attack [28]. Therefore there is a need for dynamic approach of detection rule to address the issue of unknown attack detection using Signature-based IDS [29, 30, 31].

Signature-based IDS focus its detection techniques on signatures, derived from static data analysis, which makes it difficult for malcode detection [5]. In order to have a good detection mechanism or outcome, the data analysis process plays an important role. [32, 33, 34]. This research uses dynamic data analysis to overcome the problem of static analysis. Malware analysis is a procedure to perform analysis of malware and study the behavior of malware [35, 36]. Many works have been done in malware analysis as many uses static methods, which is a method of malware analysis done without running the malware [13, 28, 37, 39].

1.3 Objectives

The main objective of the research is to design and develop a high accuracy Network Intrusion Detection method for malicious code activities. Other objectives of the research are as follows:

a. To design an algorithm to detect unknown network attack.

b. To analyze intrusion data and extract IOC from analyzed data.

c. To develop IOCRule for malicious code detection.

1.4 Scope and Limitation

The research is designed to aid existing intrusion detection methods, with the additional capability to detect variations of network intrusions in the form of malicious codes. To enable behavior analysis, IOC techniques are used to investigate and detect variants of known and unknown attacks while they are being transferred over a network. The similarity between familiar malicious code samples and unfamiliar incoming traffic is calculated based on related information or related symbols [14]. In this thesis, the research scope is restricted to intrusion-detection schemes whereby intrusion log can be used to perform packet analysis and prepare a result for future malware detection based on IOCRule. This research applies behavioral approach in intrusion rule's generation and to reduce false alarm it uses machine learning for feature clustering used for rules generation, which is a dynamic algorithm and improve detection rate.

1.5 Significance of the Research

This research explored methods and techniques of existing measures for honeypot based Malcode data collection, behavior analysis using Cuckoo Sandbox and identifying