SIGNIFICANT FEATURE IDENTIFICATION MECHANISM FOR IPv6 IN ENHANCING INTRUSION DETECTION SYSTEM

ZULKIFLEE BIN MUSLIM

DOCTOR OF PHILOSOPHY

2017

Faculty of Information and Communication Technology

SIGNIFICANT FEATURE IDENTIFICATION MECHANISM FOR IPv6 IN ENHANCING INTRUSION DETECTION SYSTEM

Zulkiflee bin Muslim

Doctor of Philosophy

2017

SIGNIFICANT FEATURE IDENTIFICATION MECHANISM FOR IPv6 IN ENHANCING INTRUSION DETECTION SYSTEM

ZULKIFLEE BIN MUSLIM

A thesis submitted
in fulfillment of the requirements for the degree of Doctor of Philosophy

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2017

DECLARATION

I declared that this thesis entitled, “Significant Feature Identification Mechanism For IPv6 In Enhancing Intrusion Detection System” is the result of my own research work except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ……
Name : ZULKIFLEE BIN MUSLIM
Date : ……

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Doctor of Philosophy.

Signature	: ……..
Supervisor Name	: PROF. DR. SHAHRIN B. SAHIB
Date	: ……..

ABSTRACT

An *Intrusion Detection System* (IDS) is a security mechanism used to detect attack patterns that occur in a network. IDS has been adapted from an Internet Protocol version 4 (IPv4) onto an Internet Protocol version 6 (IPv6) environment for the same purpose. IPv6 security issue requires IDS that has the capability to ease new threats. However, the ineffectiveness of the existing native IPv6 detection techniques cause the network attack detection in IPv6 is unconvincing. This problem has emerged due to lack of feature analysis to identify the most significant features before the chosen features used in detection technique construction. Therefore, this study has propose a technique called *Significant Feature Identification Mechanism for IPv6* (SIMv6) as a solution for feature selection issue in IPv6 domain. The *SIMv6* model has a capability of self-learning and flexible to fit with any type of data which requires feature selection solution. In this study, *SIMv6* is applied on the IPv6 dataset to identify the most significant features which is named as *Significant Features in IPv6* (SigFeatv6). Then, *SigFeatv6* is tested and evaluated its performance to differentiate between normal and attack packets accurately. The performance of *SigFeatv6* then has been compared with the performance of other features used by existing native IPv6 detection techniques. *ANOVA* and *T-Test* are the statistical tests used to evaluate the significant difference for the accuracy score between different features. Next, as time feature is an important feature for future detection technique a derived feature called *TimeInterval* was introduced to enhance the *timestamp* feature. *SIMv6* again is applied on a new set of IPv6 dataset which includes *TimeInterval* as one of its feature. The result indicates that features proposed by *SIMv6* obtained 99.87% accuracy score in average to differentiate between normal and various IPv6 network attacks packet. From the findings, *SIMv6* is capable of determining the most significant features to distinguish IPv6 packet status more effective compares to other features used by prior studies. Furthermore, the introduction of *TimeInterval* feature has improved the *SigFeatv6* performance. A testbed based on IPv6 network environment was deployed to produce a reliable IPv6 dataset. In the future, a new detection technique can be formulated based on the features proposed in *SigFeatv6* while *SIMv6* can also be applied in other domains which require feature selection solution.

ABSTRAK

Sistem Pengesanan Pencerobohan (IDS) merupakan sebuah mekanisme keselamatan untuk mengesan corak serangan yang berlaku di dalam rangkaian. Sistem ini telah diadaptasikan daripada persekitaran IPv4 ke IPv6 untuk tujuan yang sama. Isu keselamatan IPv6 memerlukan IDS yang mampu mengekang ancaman baru. Walaubagaimanapun, ketidakberkesanan teknik pengesanan IPv6 asli menyebabkan pengesan serangan rangkaian di persekitaran IPv6 tidak meyakinkan. Masalah ini timbul berpunca dari kekurangan analisa ciri-ciri untuk mengenalpasti ciri yang paling penting sebelum ciri-ciri terpilih digunakan untuk membangunkan teknik pengesanan. Oleh itu, kajian ini mencadangkan sebuah teknik bernama Significant Feature Identification Mechanism for IPv6 (SIMv6) sebagai penyelesaian untuk isu pemilihan ciri-ciri dalam domain IPv6. SIMv6 mempunyai kemampuan untuk belajar sendiri dan fleksibel untuk sesuai dengan sebarang jenis data yang memerlukan solusi pemilihan ciri-ciri. Dalam kajian ini, SIMv6 telah diaplikasikan ke atas set data IPv6 untuk mengenalpasti ciri-ciri terpenting yang dinamakan sebagai Significant Features in IPv6 (SigFeatv6). Kemudian, SigFeatv6 diuji dan dinilai prestasinya untuk membezakan antara paket normal dan serangan secara tepat. Prestasi SigFeatv6 kemudiannya dibandingkan dengan prestasi ciri-ciri yang gunakan dalam teknik pengesanan IPv6 asli sedia ada. ANOVA dan T-Test merupakan pengujian statistic yang digunakan untuk menilai perbezaan ketara untuk markah ketepatan antara ciri-ciri berbeza. Selepas itu, memandangkan ciri masa merupakan ciri utama untuk teknik pengesanan akan datang sebuah ciri dipanggil TimeInterval telah diperkenalkan untuk meningkatkan ciri timestamp. SIMv6 sekali lagi diaplikasikan ke atas set data IPv6 baru termasuk TimeInterval sebagai salah satu ciri-cirinya. Keputusan menunjukkan ciri-ciri yang dipilih of SIMv6 mendapat 99.*87% markah ketepatan secara purata untuk membezakan antara paket normal and pelbagai jenis serangan rangkaian IPv6. Dari hasil dapatan, SIMv6 mampu menentukan ciri-ciri paling penting untuk membezakan status paket IPv6 dengan lebih efektif berbanding dengan ciri-ciri yang digunakan oleh kajian-kajian sebelum ini. Tambahan lagi, pengenalan ciri TimeInterval telah meningkatkan prestasi SigFeatv6. Sebuah lapangan kajian berdasarkan persekitaran rangkaian IPv6 telah dibangunkan untuk menghasilkan set data IPv6 yang boleh dipercayai. Untuk masa depan, sebuah teknik pengesanan baru boleh dirumuskan dengan menggunakan ciri-ciri yang disarankan dalam SigFeatv6 sementara SIMv6 boleh diaplikasikan ke domain lain yang memerlukan penyelesaian pemilihan ciri-ciri.*

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful

Alhamdulillah, all praises to Allah for His strength and blessing in completing this thesis. I would like to express my gratitude to my respectful supervisor as well as my mentor, Professor Dr. Shahrin Sahib, whose expertise, understanding, and patience significantly enhanced my graduate experience. I appreciate his vast knowledge, skill, and genuine empathy throughout my study.

I would like to express my appreciation to my co-supervisor and all my colleagues and staff members of FTMK for their cooperation in making this study a success. Not to forget Universiti Teknikal Malaysia Melaka (UTeM) and Ministry of Higher Education Malaysia for sponsoring this study.

Last but not least, my deepest gratitude goes to my beloved parents: Hjh Halimah Md Isa and Hj Muslim B. Hussain, and also to my sisters, Noor Aishah, Noraini, Noor Azizah, and Noor Ashikin, for their endless love, prayers, and encouragement. Heartfelt thanks to my beloved wife, Haniza Nahar, and my precious children, Afiq, Wani, Anis, Wafi, Aniq, Ahmad and Warda, for their love, care, memories, and endless support of me.

To those who indirectly contributed to this research, your kindness means a lot to me. Thank you very much.

Zulkiflee Bin Muslim, 2017.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

IDS	Intrusion Detection System
Timestamp	The original *time* value extracted from a captured packet
TimeInterval	Duration between arrival *times* of two consecutive *timestamp*
SrcIP	Source of IP address
SrcPort	Source of Port address
DstIP	Destination of IP address
DstPort	Destination of Port address
Protocol	The protocol value extracted from a captured packet
Hlim	Hop Limit
Nlength	Next length
Npayload	Next payload
TN	True Negative
TP	True Positive
FN	False Negative
FP	False Positive
FPR	False Positive Rate
Recall	Credibility measurement
Precision	Correct identification measurement
SVM	Support Vector Machine
PSO	Particle Swarm Optimization
SigFeatv6	The Significant Features in IPv6 (Proposed in this study)
SigFeatTIv6	The Significant Features in IPv6 with TimeInterval feature
SIMv6	Significant Feature Identification Mechanism for IPv6
TNR	True Negative Rate
TPR	True Positive Rate
TA	Total Accuracy

LIST OF PUBLICATIONS

A. Publications

Zulkiflee Muslim, Mohd Faizal Abdollah, Mohd Fairuz Iskandar Othman, Nur Azman Abu, and Shahrin Sahib. (2011). Behavioral Analysis on IPv4 Malware in both IPv4 and IPv6 Network Environment. *International Journal of Computer Science and Information Security (IJCSIS),* 9 (2).

Zulkiflee Muslim, Siti Azirah Asmai, Haniza Nahar, Zakiah Ayop, and Shahrin Sahib. (2011). Behavioral analysis on IPv4 malware on different platforms in IPv6 network environment. In: *IEEE International Conference Open Systems*. IEEE.

Zulkiflee Muslim, Robiah Yusof, Nur Azman Abu, and Shahrin Sahib. (2012). Improvising Intrusion Detection for Malware Activities on Dual Stack Network Environment. *World Academy of Science, Engineering and Technology (WASET),* 67 (International Science Index 67, 2012), pp. 536 544.

Zulkiflee Muslim, Haniza Nahar, Shahrin Sahib, and Mohd Khanapi Abd. Ghani. (2014). A Framework of IPv6 Network Attack Dataset Construction by Using Testbed Environment. *International Review on Computers and Software (IRECOS),* 9 (8), pp. 1434 1441.

Zulkiflee Muslim, Mohd Sanusi Azmi, Sharifah Sakinah Syed Ahmad, Shahrin Sahib, Mohd Khanapi Abd Ghani. (2015). A Framework of Features Selection for IPv6 Network Attacks Detection. *World Scientific and Engineering Academy Society (WSEAS) Transactions on Communications*.

CHAPTER 1

INTRODUCTION

1.1 Introduction

The objective of this chapter is to provide an overall view of this study. In this chapter, there will be a detailed discussion about the research problem, which will be solved toward the end of this study. Then, the research problem will be elaborated into the main research question. From the research question, the objectives of this study will be identified. After that, various limitations of this study will be underlined in the research scope. Next, the research design of this study will be explicitly explained. Subsequently, the research contribution of this study will be elaborated. Finally, a chapter summary is presented.

1.2 Research Overview

Internet Protocol version 6 (IPv6) was invented in 1998, as stated in the RFC (Request for Comments) 2460 (Deering and Hinden 1998). IPv6 is a new technology that is considered a successor to IPv4 technology. IPv6 was invented to overcome certain issues that were present in the previous technology. The main issue taken into serious consideration was the fact that IPv4 addresses are facing total depletion. Although a technique called Network Address Translation (NAT) has been proposed in the IPv4

network to overcome the address depletion issue, the solution is only a temporary one. Users are still demanding unique IP (Internet Protocol) addresses to be assigned to their nodes. The emergence of new technologies, such as the Internet of Things (IOT) (Xia et al. 2012), cloud computing (Zissis and Lekkas 2012), and wireless technology applications (Al Ameen et al. 2012), makes the need for IP addresses even more severe. Due to this issue, the Internet Engineering Task Force (IETF) drafted the first specification for IPv6 technology in RFC 2460 in 1998.

As the number of Internet users continues to increase, IANA is rarely able to allocate IP addresses to new users. As an alternative, the new IPv6 technology has been introduced to overcome the IP address space allocation issues. June 8, 2011 was World IPv6 Day, which was intended to encourage IPv4 Internet users to migrate to IPv6 Internet. There are some new features offered in IPv6 that are better than IPv4 (Reddy et al. 2012). Unfortunately, IPv6 is not backward compatible to IPv4. The IPv4 Internet and the IPv6 Internet are considered two different worlds. Users have to decide whether they want to use IPv4 or IPv6 to connect to the Internet.

Although IPv6 was officially launched in June 2011, many IPv4 users have been reluctant to migrate because they feel comfortable with the existing IPv4 protocol (Huston 2013). Based on a survey, only large companies participated in World IPv6 Day. Many other companies still lack adequate resources to adopt IPv6. Some of the limitations faced by these companies in 2011 included not having enough of a budget to purchase new devices that support IPv6. Furthermore, some companies did not have enough awareness of IPv6 technology to drive its adoption; they did not have the required knowledge or skills to implement IPv6 (Alhassoun and Alghunaim 2016). Finally, and most importantly, most of these companies remained unconvinced by the lure of IPv6 technology since it was still

new and might have had unexpected hiccups. IPv6 had not been fully tested on a large network scale prior to its launch (Trinh et al. 2010; Caicedo et al. 2009).

After both the IPv4 and IPv6 networks had been implemented, many users felt that the IPv6 services were not as good as those services offered in IPv4 (Han et al. 2014). Many researchers have invested a great deal of effort to enhance IPv6 implementation in order to offer better services. The main goal is to offer IPv6 services that are at least on par with the IPv4 network. Hence, research topics in the IPv6 domain are heavily focused on IPv6 services (Bagnulo et al. 2012; Gu et al. 2013; Mrugalski et al. 2013). Unfortunately, researchers may have focused too much on IPv6 services and implementations; consequently, security issues are currently overlooked or neglected. What is more, the new threats emerge for IPv6 network are need to be focused and the current security solution for IPv6 is needed to be improved (Hendriks et al. 2015).

Table 1.1: The Impact of Intrusion Activities

Company (Intrusion Activity)	Year	Affected Items	Lost Information	Type of Loss
CheckFree Corp. (Hacked Web Server)	2009	Personal information was stolen.	5,000,000 users	Intangible
Google.com (Stolen Documents)	2009	Documents were stolen.	< 0.05% of documents	Tangible
New York Mellon Corp. (Employee Theft)	2009	Company money was stolen by an employee.	> 1 million USD	Tangible
Spain (Plane Crash)	2008	The maintenance system malfunctioned.	Human lives	Intangible

Table 1.1 above shows some of the impacts from intrusion attacks based on several reports. According to Patel et al. (2010), the intrusion activities caused tangible losses by affecting millions of database records as well as by creating millions of victims from

different organizations and companies. These intrusion attacks also generated financial losses that amounted to more than 1.5 million USD from several affected organizations in 2009. Meanwhile, some of the intangible losses incurred from the intrusion activity that came in the form of people dying. Based on a claim made by Bellovin (2010), an intrusion attack caused a plane crash back in 2008. The plane crashed as a result of a maintenance system malfunctioning after it had been compromised by malware attacks. In yet another example, more than five million users were affected when a personal information data server was hacked. A study found that more than twenty new vulnerabilities are detected in computer networking products every month (Patcha and Park 2007). In 2005, a survey completed by several companies indicated that total financial losses due to intrusion attacks amounted to around 130 million USD (C.S. Institute and F.B.O. Investigation 2005). What is more, without proper monitoring mechanism some sensitive data for companies and employees also can be misused for bad purposes (Gupta et al. 2017). It is clear from all these facts that intrusion activities cannot be treated carelessly, as their impact might not only involve money but also human lives. Hence, an IDS should be implemented as one countermeasure mechanism that can be used to deal with intrusion activities.

1.3 Research Problem

The impact of an intrusion attack can be quite severe if the launched attacks are not treated seriously (CyberSecurity Malaysia 2013). Cyber Security Malaysia, in a joint venture with the Information Telecommunications Authority of Oman (ITA), organized a conference to discuss emerging threats with regard to the cyber world. This demonstrates the fact that cyber threats cannot be treated locally; instead, they need to be addressed on a global level. What is more, intrusion activities tend to gradually increase alongside the

rapid development of the information society (Zhang 2009). Unfortunately, current control mechanisms proved insufficient to content intrusive activities (Levitt and Dias 2017).

IPv6 has been designed to overcome several issues that occurred in IPv4, especially in terms of security. However, IPv6 per se is not a panacea for all the security issues that transpired with IPv4 (Alangar and Swaminathan 2013). Other aspects, such as network design, application design, and users' policies, also contributed to the network security issues as a whole. Some of the knowledge discovered in IPv4 can still be applied in the IPv6 network environment. However, some of the knowledge is considered inappropriate for use in IPv6 due to its new features. In these cases, the use of an Intrusion Detection System (IDS) as a detection mechanism in IPv4 is still an alternative solution that can be implemented in an IPv6 environment. Nevertheless, the detection techniques used in IPv4 do need to be verified before being deployed in the IPv6 network environment.

The process of implementing IDS in IPv6 is almost identical to the process for IPv4. The detection techniques used in IPv4 can be applied in IPv6, but the detection techniques constructed in the IPv4 environment cannot be transposed directly to the IDS in the IPv6 network environment. This is simply because the data pattern discovered was based on the IPv4 network environment and the network pattern in IPv6 differs from that of its predecessor (Peng et al. 2013). What is more, the detection techniques invented for IPv6 network were never being tested in full scale of real network (Barbhuiya et al. 2013). As a result, the detection techniques are insufficient to ease threats in the IPv6 network environment.