Faculty of Information and Communication Technology

ENHANCED FAST ATTACK DETECTION TECHNIQUE FOR NETWORK INTRUSION DETECTION SYSTEM

MOHD FAIZAL BIN ABDOLLAH

PhD

2009

ENHANCED FAST ATTACK DETECTION TECHNIQUE FOR NETWORK INTRUSION DETECTION SYSTEM

MOHD FAIZAL BIN ABDOLLAH

A thesis submitted

in fulfillment of the requirements for the degree of Doctor of Philosophy

in Information and Communication Technology

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2009

DECLARATION

I declare that this thesis entitle "Enhanced Fast Attack Detection Technique For Network Intrusion Detection System" is the result of my own research 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 : **MOHD FAIZAL BIN ABDOLLAH**

Date : 20/9/09

DEDICATION

Dear Allah

I devoted my life and death to You, Allah. May my life is within Your guidance.

Dear My Parent

Thank you for your sacrifice and love. No such compensate except from Allah.

Dear My Beloved Wife

Your support, patience and encouragement give me strength to finish this study. May Allah bless us.

Dear Teachers

Thank you for all the knowledge. May your knowledge are beneficial and useful for all humanity.

Dear My Siblings

Thank you for your support and love. May Allah forgive us.

Dear My Children

May Allah guide and protect us to be good Muslims.

ACKNOWLEDGEMENT

I would like to express full of my sincere gratitude to my advisors, Professor Dr. Hj. Shahrin Sahib of Faculty of Information Technology and Communication, University of Technical Malaysia for his guidance, patience and support through out the year of my Ph.D study in University Technical Malaysia, Melaka. Furthermore, his constant encouragement during the course of my research and graduate studies is duly acknowledgement with a deep gratitude.

I would also like to express my deep appreciation to Associate Professor Dr Ismail Mohammad, Professor Dr Nanna Suryana Herman and Mr Asrul Hadi Yaacob for their critical feedback and valuable suggestion during the course of my study. I extend all of my gratitude to my friends in University Technical Malaysia Melaka and Multimedia University Melaka who provide a lot of help and support during my research and made my life colourful. Although I could not list all of their names here, their warmth would live in my heart all over the time.

Last, but not least, a big thank you to my wife, my children, my father, my mother and all my family, who supported me all along and took such a great patience and interest in my Phd work.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In the last decade, the network has grown both in size and importance. In particular TCP/IP network and most notably the world wide Internet have become the main infrastructure to exchange data and carry out transaction. They have also become the main mean to attack host. The popularity of intrusion tools and script are the main contribution of the attack inside the network. Gathering valuable information from vulnerable machine such as IP address and vulnerable application is the first step for the attackers to launch an attack to the vulnerable machine. There are numerous techniques to get this information such as sweeping, scanning, probing and so on. These information gathering techniques can be divided into two categories which are Fast Attack and Slow Attack. Fast attack can be defined as an attack that uses a large amount of packets or connections within a short period in few seconds. Meanwhile the Slow Attack can be defined as an attack which takes much longer time in the sense of few minutes to few hours to complete. In order to detect these attacks, introducing intrusion detection system (IDS) inside the network is necessary. An IDS has the capability to analyze the network traffic and recognize incoming and on-going intrusion. IDS has several weaknesses which need to be tackled to improve the accuracy of detection. The current weakness is on selecting the suitable threshold for detecting the intrusion activity. Selecting too high of value may generate excessive false alarm while too low may miss the malicious activity. Hence, this research introduces a new technique in selecting a suitable threshold for detecting the intrusion activity especially for Fast Attack. The threshold selected in this research has been analyzed, examined, tested and proven that it is able to increase the accuracy of detection to 99.5% using statistical approach and decrease the speed of detection. Besides introducing a new technique to identify and select the threshold, this research also revealed the feature influence and reason behind the selection of the feature. Selecting unnecessary features may cause computational issues and decrease the accuracy of detection. Furthermore, current research more concentrates more on technique of detection rather than feature selection. Most research uses the features without highlighting the influence of the feature inside the system itself. Thus this research will reveal the influence of the features in predicting the result of the detection. The results show that the selection of features and the threshold selected using the new technique has a strong potential to detect the fast attack and significantly reduce the false alarm generated by the intrusion detection system.

ABSTRAK

Beberapa dekad yang lepas, rangkaian telah berkembang dan menjadi lebih penting. Rangkaian TCP/IP dan khususnya Internet telah menjadi keutamaan untuk pertukaran data dan melaksanakan transaksi. Ia juga telah menjadi tumpuan utama untuk menyerang hos. Kemudahcapian perkakasan serangan dan skrip merupakan penyumbang utama kepada serangan di dalam rangkaian. Pengumpulan maklumat berharga daripada komputer yang mudah diserang seperti alamat IP dan aplikasi yang mudah diserang merupakan langkah awal untuk penyerang melancarkan serangan ke atas komputer yang mudah di serang. Terdapat pelbagai teknik untuk mendapatkan maklumat ini seperti penemuan, pengimbasan dan penyiasatan. Teknik pengumpulan maklumat ini boleh dibahagikan kepada dua kategori iaitu serangan pantas dan serangan lambat. Serangan pantas boleh ditakrifkan sebagai serangan yang menggunakan paket atau perhubungan yang banyak di dalam suatu masa yang singkat iaitu beberapa saat. Manakala serangan lambat boleh ditakrifkan sebagai serangan yang memerlukan masa yang lama iaitu beberapa minit ke beberapa jam untuk selesai. Bagi membolehkan pengesanan serangan-serangan ini, pengenalan kepada sistem pengesanan pencerobohan di dalam rangkaian adalah diperlukan. Sistem pengecaman pencerobohan ini mempunyai keupayaan untuk membuat analisa trafik rangkaian and mengecam penceroboh yang masuk dan keluar. Sistem pengecaman pencerobohan ini mempunyai beberapa kelemahan yang perlu diatasi untuk meningkatkan ketepatan pengesanan. Kelemahan semasa sistem pengecaman pencerobohan ini adalah di dalam memilih aras yang sesuai untuk mengesan aktiviti pencerobohan. Pemilihan nilai yang terlalu besar akan menyebabkan lambakan kesalahan log manakala pemilihan terlalu kecil menyebabkan serangan tidak dapat dikesan. Justeru itu, kajian ini memperkenalkan teknik baru di dalam memilih aras yang sesuai untuk mengesan aktiviti pencerobohan terutamanya serangan pantas. Aras yang dipilih di dalam kajian ini akan dianalisa, diperiksa, diuji dan disahkan berupaya untuk meningkatkan ketepatan pengesanan sehingga 99.5% dengan menggunakan teknik statistik dan mengurangkan masa pengecaman. Di sebalik memperkenalkan teknik baru untuk mengenalpasti and memilih aras, kajian ini juga mendedahkan pengaruh atribut and sebab di sebalik pemilihan atribut tersebut. Tambahan pula, kajian semasa lebih tertumpu kepada teknik pengesanan daripada pemilihan atribut. Penyelidik hanya menggunakan atribut tanpa mendedahkan pengaruh atribut tersebut di dalam sistem yang digunakan. Oleh itu, kajian ini akan mendedahkan pengaruh atribut di dalam meramalkan keputusan pengecaman pencerobohan. Keputusan yang diperolehi menunjukkan atribut yang dipilih dan aras yang dipilih menggunakan teknik baru ini mempunyai potensi yang amat baik untuk mengesan serangan pantas dan seterusnya mengurangkan kesalahan amaran yang dihasilkan oleh sistem pengecaman pencerobohan.

CHAPTER 1

INTRODUCTION

1.1 Introduction

We now live in the information age; hence it is nearly impossible to imagine our lives without the Internet and information systems. Nowadays the Internet plays an important role in stock market, access to weather forecast, E-medicine, E-commerce and even daily newspapers. The networking revolution has fully come to age in the last decade. As the network grows in size and complexity and computer services expands, vulnerabilities within local area and wide area network has become mammoth albeit problematic. The problems occur due to the increasing number of intrusion tools and exploiting scripts which can entice anyone to launch an attack on any vulnerable machines. The attack can be launched in term of fast attack or slow attack. Fast attack can be defined as an attack that uses a large amount of packet or connection within a few second (Lazarevic et al, 2003). Meanwhile, slow attack can be defined as an attack that takes a few minutes or a few hours to complete (Wenke et al, 1999). Both of the attack gives a great impact to the network environment due to the security breach. Referring to the statistic produced by CERT (*Computer Emergency Response Team*) from 1995 till 2007 as depicted in Figure 1.1-1, it is shown that there is a growth on the number of intruders in the computing environment.

Cataloged vulnerabilities

Year	Total vulnerabilities cataloged
Q1-Q3, 2008	6,058
2007	7,236
2006	8,064
2005	5,990
2004	3,780
2003	3,784
2002	4,129
2001	2,437
2000	1,090
1999	417
1998	262
1997	311
1996	345
1995	171
Totals	**44,074**

Source by Computer Emergency Response Team

Figure 1.1-1 : Number of Computer Incident from 1995 until 2007

The same scenario was prevalent in Malaysia as well, in which the number of host that had been compromised by the attacker also increased as reported by CyberSecurity Malaysia in Figure 1.1-2 (CyberSecurity, Malaysia, 2007). The increasing number of host infected will cause a threat to the national security which needs to be protected.

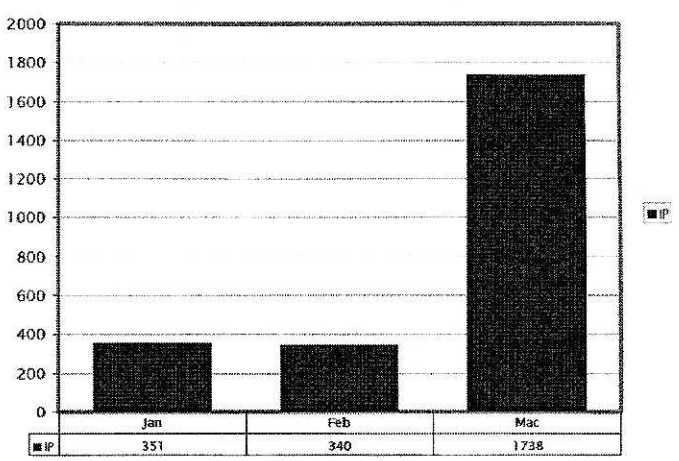

Figure 1.1-2 : Number of Infected Host

Besides the statistical information regarding the security breach, there are some the press reports and white paper as well regarding the security breaches in some of the remote workstation.

1. October 7, 1999: Hackers apparently working from Russia have systematically broken into American Defense Department computers for more than a year and took vast amounts of unclassified but nonetheless sensitive information, commented one U.S. officials. Besides penetrating the Pentagon's defenses, the hackers had raided unclassified computer networks at Energy Department nuclear weapons and research labs, at the National Aeronautics and Space Administration and many university research facilities and defense contractors (Allen et al, 2000)

2. June 1, 1999: After NATO jets hit the Chinese Embassy in Belgrade in May; hackers from China attacked a handful of U.S. government sites, including one maintained by the Energy Department. In an unrelated incident, the official White House site was shut down briefly because of an attempt to tamper with it by unidentified hackers (Allen et al, 2000).

3. A rapidly spreading worm had infected an estimated 1.1 million PCs in a 24 hours and bringing the total number of infected computers to 3.5 million (Cybersecurity, 2009).

The above illustration shows the seriousness and sophisticated nature of recent cyber attacks which is evident. The growth of the incident on the computing environment has reflected on the growth of the Internet itself. The entire incidents that occur will cause a major loss for many organizations. From a survey conducted by American Society For Industrial Security and Pricewaterhouse-Cooper, there was a lost about USD 45 million for 1000 companies due to the security breach (John E. Conovan, 2001). According to marketing networking group CMO Council, a data breach could cost an average of $14 million US dollar on a recovery cost (Boonbox, 2008). A survey made by CSI Computer Crime and Security reported that the lost due to security breach was USD288 by 618 per respondent. The total number of respondent involved was 144 security practitioners (Robert Richardson, 2008). Therefore, it is very important that the security mechanism of a

system is designed so as to prevent unauthorized access to system resources and data. This will definitely minimize losses face by the organization. However, completely preventing breaches of security appear, at present, unrealistic. However, trying to detect these intrusion can be attempted so that action may be taken to repair the damage. Furthermore, detecting the intrusion especially fast attack as quickly as possible may reduce the possibility of damage occurring inside the network of the organization. The literature in this field of research is called **Intrusion Detection**.

1.2 Intrusion Detection System

Intrusion detection can be divided into three types which are host based intrusion detection system, network based intrusion detection system and hybrid based intrusion detection system. Although the intrusion detection can be divided into three different types, the main goal for each of them is the same which is intrusion detection. Intruder detection system is a system to detect attacks, or to classify them as unwanted authorized login, regardless of their success (Allen et al, 2000). The detection method used by intrusion detection system can be classified as anomaly based detection and signature based detection. This system is responsible to identify intrusion, which is defined as unauthorized use, misuse or abuse of the computer system by either the unauthorized user or external perpetrators (Puketza et al, 1996);(Vokorokos et al, 2006). Additionally, intrusion detection system is used to help computer system to prepare to deal with many kinds of attack such as scanning, worm and virus attack (Verword and Hunt, 2003). One of the initial goals to accomplish intrusion detection is by collecting information from a variety of system and network sources and analyze the batch information, search for symptoms which means security problems (S. Razak et al, 2002). The success of analyzing the information may help to detect the intrusion activity in the network. Although the intrusion detection system has the capability to detect the intrusive behavior, unfortunately it also has several weaknesses. The next section will discuss further the weaknesses of the intrusion detection system.

1.3 Weakness of Network Intrusion Detection System (NIDS)

The detection method used by the network intrusion detection system can be classified as anomaly based system and signature based system. Both these detection methods are used by the intrusion detection system which have their own drawback in detecting the intrusion

activity. Therefore, the weakness of the NIDS based on the anomaly based NIDS and signature based NIDS will be discussed.

Signature based NIDS have a major drawback in identifying the new intruder in the network. It is because signatures based NIDS depend on the intrusion pattern that have been declared inside a database of the intrusion detection system. If there is a new attack inside the network and the signature of the new attack has not been stated inside the database, thus the system will not be able to detect this new attack (Kingsly Leung & Christopher Leckie, 2005). This problem arises because signature based network intrusion detection system relies on sets of predefined rules that are provided by administrator, automatically created by the system, or both (James Cannady, 1998), (Kang et all, 2005). Generating new rules need a deep knowledge and skills from administrator. However, not many administrators have such deep knowledge and skills to generate new signature (Koike & Ohno, 2004)

Meanwhile, in anomaly based approach, the intruder detection attempts to model the expected behavior of objects (users, processes, network hosts and the like). Any action that does not correspond to expectation is considered suspicious. The strength of these methods lies in their ability to differentiate normal user behavior, anomalous acceptable behavior, and intrusive behavior (James Cannady, 1998), (Zhang et al, 2007). The anomaly based detection has difficulties to determine the threshold value by which behavior must deviate from a profile in order to be considered as possible attack (Chris Herringshaw, 1997);(Robertson et al, 2003);(Xin et al, 2004). Selecting low value of threshold may result in generating excessive false positive, while high value of threshold may result undetected malicious behavior. Furthermore, incomplete profiling due to weaknesses of threshold selection may lead to false alarm (Chris Herringshaw, 1997). Derrick et al, (2007) also claim that setting of threshold level to an appropriated value to minimize the false positive still becomes a central issue which need to be solved. Therefore, introducing a new technique that can identify an appropriate value of threshold is necessary to minimize the false positive (Derrick et al, 2007) especially for fast attack detection. Based on the explanation above, there is a significant contribution in introducing a new technique to identify the static threshold value for detecting the fast attack intrusion activity. By introducing a new technique, an appropriate static threshold value can be identified to help the intrusion detection make a good decision in recognizing the fast attack intrusion

activity. This claim is also supported by Stamford et al, 2002 in their research to select the useful threshold value in detecting the attacker.

Although both system have their own drawbacks, anomaly based detection has capabilities to recognize new attack inside the network without a need to update new rules (Gaurav Tandon & Philip K. Chan, 2007). This capability requires appropriate value of threshold to distinguish between the normal and abnormal behavior of the fast attack activity. Hence, introducing a new technique to select an appropriate threshold is required to reduce the false alarm generated by the anomaly based detection for the fast attack detection.

1.4 Problem Statement

Currently anomaly based detection faces a problem to identify the accurate and appropriate threshold value to differentiate between the normal network traffic and abnormal network traffic. Selecting too low of a threshold value may cause the system to generate many false alarms, while selecting too high of threshold may miss the malicious activity. Therefore it is necessary to introduce a new technique to identify the appropriate threshold value for the anomaly based detection so that it can build a complete profile to distinguish between the normal network traffic and abnormal network traffic in term of fast attack intrusive behavior.

Besides selecting a suitable threshold value, selecting an important feature is also important in developing the intrusion detection system. It is because the success of intrusion detection system depends on the set of feature selected inside the intrusion system. There are numerous features inside the network traffic and most of the research just used the feature without revealing the influence of the feature itself (Onut & Ghorbani, 2006). Furthermore, previous researcher only concentrated on technique of detection rather than the feature itself (Onut & Ghorbani, 2006). Revealing the influence of the feature in the detection model may lead to the success of the intrusion detection system especially for fast attack detection. It is because the significant contributions of the feature can be identified and the unnecessary features can be eliminated. Therefore, it is vital to reveal the feature influence inside the detection model using a statistical approach.

Thus, this research will focus on introducing a new technique to identify a threshold for the intrusion detection especially fast attack. Besides that, the feature influence will also be