APPROACH FOR SOLVING ACTIVE PERTURBATION ATTACK PROBLEM IN STEPPING STONE DETECTION

MOHD NIZAM BIN OMAR

UNIVERSITI SAINS MALAYSIA

2011

APPROACH FOR SOLVING ACTIVE PERTURBATION ATTACK PROBLEM IN STEPPING STONE DETECTION

by

MOHD NIZAM BIN OMAR

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

August 2011

ACKNOWLEDGEMENTS

Bismillahirrahmanirrahim

Alhamdulillah

I would like to thank Professor Dr. Rahmat Budiato as my supervisor. Without his support, this research would be not completed successfully. I also want to thank Universiti Utara Malaysia and the Ministry of Higher Education for having funded me during the three years of my Ph.D study. Not forgetting also, my colleagues who have given me motivation and support to make this research possible.

My gratitude also goes to my beloved wife, Zurianawati Ibrahim, my cheerful daughter, Nurin Afriena Batrisyia and my handsome son, Ahmad Khalish Zharfan for giving me their full support during my three-year study period. The same goes to my father, Omar Ahmad and my mother, Seliah Mat Sap for their constant prayers all the way from my hometown, Temerloh, Pahang. To my father-in-law, Ibrahim Lebai Bakar and my mother-in-law, Mahani Ahmad, I thank them for providing me with their assistance in my research, including the setting up of a complete office at their home.

Last but not least, I also appreciate all who have been involved in making this Ph.D journey possible and its final completion a reality. For my family members, in-laws and friends who have given me your support, thanks for everything.

TABLE OF CONTENTS

CHAPTER 2 - LITERATURE REVIEW

CHAPTER 5 – RESULT OF THE STUDY AND ANALYTICAL DISCUSSION

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

LIST OF ABREVIATIONS

APA	Active Perturbation Attack
AI	Artificial Intelligence
AR	Associate Rules
ANN	Artificial Neural Network
BMU	Best Matching Unit
BLN	Bayesian Learning Network
CC	Connection Chain
CERT	Computer Emergency Response Team
CIS	Caller Identification System
CPU	Central Processing Unit
DIDS	Distributed Intrusion Detection System
DM	Data Mining
DDOS	Distributed Denial of Service
DOS	Denial of Service
DSS	Direct Stepping Stone
DW	Differentiate Windows
FPR	False Positive Rate
FTP	File Transfer Protocol
H-SSD	Hybrid Stepping Stone Detection
HI-SSD	Hybrid Intelligent Stepping Stone Detection

HSSD	Host-based Stepping Stone Detection
I-HSSD	Intelligent Host-based Stepping Stone Detection
I-NSSD	Intelligent Network-based Stepping Stone
IDS	Intrusion Detection System
IDIP	Intrusion Identification and Isolation Protocol
IETF	Internet Engineering Task Force
IP	Internet Protocol
IP Sec	Internet Protocol Security
IPD	Inter-packet Delay
IRS	Intrusion Response System
ISS	Indirect Stepping Stone
LAN	Local Area Network
MMS	Mini/Max Sum Ration
MNA	Measurement and Network Analysis Group
NAM	Network Animator
NIC	Network Interface Card
NLANR	National Laboratory for Applied Network Research
NN	Neural Network
NPD	Normalized Dot Product
NSSD	Network-based Stepping Stone Detection
OS	Operating System
PC	Personal Computer
RBF	Radial Basis Function
RFC	Request for Comments
RTT	Round Trip Time

SDBA	Standard Deviation-based Clustering Approach
SOM	Self-Organizing Map
SSD	Stepping Stone Detection
SSH	Secure Shell
STAT	Statistical Correlation
SWT	Sleepy Watermark Tracing
TCP/IP	Transmission Control Protocol/Internet Protocol
TPR	True Positive Rate
VoIP	Voice Over Internet Protocol
VPN	Virtual Private Network
WAN	Wide Area Network

LIST OF SYMBOLS

$Host_n$	-	*n*-th Host
A	-	Flow A
B	-	Flow B
$SS_{A,B}$	-	A list of stepping stone host from host A to host B
$t_{A,E}$	-	Time processing for host A to host E
I	-	Set of item
T	-	Set of transaction
$\sigma(x)$	-	Support count
$s(x \rightarrow y)$	-	Support that x and y occur together
$c(x \rightarrow y)$	-	Confidence that transaction containing x that also containing y
Inp	-	Input
TPR	-	True Positive Rate
FPR	-	False Positive Rate
APA	-	Active Perturbation Attack
O	-	Output
H	-	Hybrid Approach
Int	-	Intelligent Approach
a	-	Arrival time

d	-	Departure time
A	-	Defined range of arrival time
D	-	Defined range of departure time
W	-	Windows size
e_a	-	Element of a (arrival time)
e_d	-	Element of d (departure time)
α	-	Defined delay for outbound
c	-	Count counter
C	-	Defined counter
In	-	Inbound flows of *n*-host
O_n	-	Outbound flows *n*-host
$F_{b,e}$	-	Flows from host origin, b to host destination, e
R	-	Defined range of capture packet
i	-	Packet data
d_i	-	Data with packet data I
C	-	Number of d_i
C_{new}	-	C if connection if not from the same connection
L	-	List of unique d_i and its connection
d_{new}	-	Detected L
J	-	The joint of d_i and d_{new}
J_{new}	-	List of joint that not just one list
k	-	Incoming stream
l	-	Outgoing stream
d_p	-	Dropped packet
c_t	-	Period of time where dropped packet occurred

n	-	Number of packet
m	-	Number of packekt
$k_{(n+1)}$	-	List of packet of k-host
$l_{(m+1)}$	-	List of packet of l-host
s	-	Size
DW_s	-	Differentiate windows with s size
H_n	-	*n*th Host
C_n	-	*n*th Connection chain
n	-	Number of host
m_c	-	Prototype vector
$d(x, m_c)$	-	Best Matching Unit
m_i	-	Node
t		Time
x	-	Input pattern
h_{ci}	-	Time-variable non-increasing
R	-	Range of input being capture
c	-	Number of layer used in SOM
[m,n]	-	Layer of neuron to be used to classify the vector
t	-	Input included in SOM
e	-	Number of training
a	-	Arrival time
d	-	Departure time
I	-	InSet
O	-	OutSet
w	-	Weight

Δ	-	Threshold
d	-	Delay
c	-	Chaff
d_p	-	Dropped packet
c_n	-	Connection chain of n-host
d	-	Segment of packet
c	-	Number of packet in d
s	-	Selected random packet
r_d	-	Random delay formula
c_t	-	Current time
r_t	-	Random time
b	-	Delay packet
p	-	Percentage of delay
P	-	Percentage of crafting
s_r	-	Selected random packet
r_c	-	Random chaff formula
$C_{n,m}$	-	Connection chain from host n to host m
t_{HSSD}	-	Time processing for HSSD without using NSSD
$t_{HSSD'}$	-	Time processing for HSSD with NSSD
l	-	First flow on simulation
k	-	Second flow on simulation
[a]	-	Element of support count a
[b]	-	Element of support count b
[a, b]	-	Element of support count of a and b
n	-	Number of straight line

l	-	List of stepping stoned host
e	-	Confidence level

PENDEKATAN PENYELESAIAN SERANGAN PENEMBUSAN AKTIF DI DALAM PENGESANAN BATU LONCATAN

ABSTRAK

Batu loncatan merupakan salah satu daripada teknik menyembunyikan jejak yang digunakan oleh penceroboh untuk menyembunyikan jejaknya. Untuk lebih daripada satu dekad, para penyelidik menumpukan usaha mereka untuk mempertingkatkan pendekatan Pengesanan Batu Loncatan (PBL) untuk mengidentifikasi secara tepatnya hos yang dipergunakan untuk melakukan serangan batu loncatan. Tambahan pula, Serangan Penembusan Aktif (SPA) seperti lengah, jatuhan paket dan chaf mengancam pendekatan PBL. Hari ini, di antara pelbagai jenis SPA, chaf, lengah dan jatuhan paket adalah sangat penting.

Sistem PBL semasa hanya menumpukan usaha kepada satu atau dua jenis PBL atau kombinasi di antaranya. Disebabkan ketepatan, ketagaran dan masa pelaksanaan PBL terdedah kepada serangan, PBL yang sangat tegar, jitu dan pantas diperlukan bagi mengelakkan serangan oleh PBL.

Pendekatan hybrid boleh meningkatkan ketepatan di dalam PBL. Oleh yang demikian, untuk menyelesaikan masalah PBL, penyelidikan ini memperkenalkan penyelesaian yang dipanggil Pengesanan Batu Loncatan Hibrid Pintar (PBLHP) melalui eksperimen yang rapi dengan menggunakan testbed dan sembilan kombinasi daripada tiga jenis

(lengah, jatuhan paket dan chaff). PBL. Keputusan daripada keputusan eksperiment kemudiannya dibandingkan menggunakan pendekatan tradisional juga teknik kepintaran buatan. Hasil daripada keputusan eksperiment memperlihatkan PBLHP menghasilkan 0% Kadar Positif Salah dan 100% Kadar Positif Benar untuk keseluruhan kes. Melalui keptusan yang diperolehi membuktikan bahawa PBLHP adalah lebih jitu dan juga tegar melawan PBL, sesuatu yang telah sediada tanpa penyelesaian di selesaikan lebih daripada satu dekat yang lalu.

APPROACH FOR SOLVING THE ACTIVE PERTURBATION ATTACK PROBLEM IN STEPPING STONE DETECTION

ABSTRACT

Stepping stone is one of the hidden tracking techniques used by an intruder to hide its tracks. For more than a decade, researchers have focused themselves in enhancing the Stepping Stone Detection (SSD) approaches in order to identify accurately a compromised host using stepping stones to attack. In addition, Active Perturbation Attacks (APA) such as delays, dropped packets and chaffs threaten the SSD approaches. Today, among the types of APAs, chaffs, delays and packet drops are very significant.

The current SSD systems only deal with one or two types of APA or a combination of them. Since SSD's accuracy, robustness and the execution time are prone to attacks by APAs a very robust, accurate and fast SSD is required to anticipate attacks by the APAs.

A hybrid approach can increase the accuracy in detecting the stepping stone attacks. Thus to overcome the APA problems, this research proposes a solution called Hybrid Intelligent SSD (HI-SSD) entailing extensive experiments using a testbed and nine combinations of three APA types (delays, dropped packets and chaffs). The experiment results are then compared using traditional approaches as well as artificial intelligent techniques. The results show that HI-SSD produced 0 % False Positive Rates

(FPR) and 100% True Positive Rates (TPR) for all cases. These results prove that HI-SSD is more accurate and is also robust against the APAs, something that had remained unsolvable for more than a decade.

CHAPTER 1

INTRODUCTION

This chapter acts an introduction to the entire research that was conducted for three years. Beginning with the introduction of the problem, it also discusses the background of the problems related to the active perturbation attacks in stepping stone detection. The description of the research problems then follows with the objectives, significance of research, scope of research, research contributions, the organization of the thesis and finally the summary of this chapter.

1.1 Introduction

Network security is a field designed to protect an organization from intruders or hackers (McClure et al., 2009). Network security can be divided into three categories such as prevention, detection and response (Cohen, 1997). Response capabilities must be the right kind of response because a response plays an important role after the detection has successfully identified an intrusion or attack. Response, according to Fessi et al., (2010) is a decision tool for the Intrusion Detection System (IDS) to prevent attacks and to ensure that the computer environment is safe.

One important response is tracing. Carver (2010) lists tracing as one of the possible response types besides other acts such as account termination, connection cut

and so forth. Tracing can be divided into two categories; i) traceback and ii) Stepping Stone Detection (SSD). Traceback is for solving IP-spoofing problems and SSD is for detecting stepping stones used by intruders to hide their original locations (Dong-li, 2010).

SSD is the research focus of this study. The usage of stepping stone detection as an application is important in the Botnet detection (Strayer et. al, 2008). SSD research has evolved from one stage to another and has led to a currently, matured SSD field. The evolution involved the usage of data payload by Staniford-Chen and Herberlein (1995), to the timing-based packet information by Zhang and Paxson (2000), and Yoda and Etoh (2000) to the latest evolution, conducted by Almulhem and Troare (2011) in the use of Artificial Intelligence (AI) techniques. Through this evolution some issues and problems appeared. Among the issues and problems include encrypted connection (Zhang and Paxson, 2000), (Yoda and Etoh, 2000), confidence bounce (Blum et al., 2004), and active perturbation (Wu and Huang, 2007).

The most salient issue or problem that needs to be solved is active perturbation as mentioned by Wang (2004) and Venkateshaiah (2006). Although there are efforts to solve the active perturbation problems, no research has attempted to solve the problem, totally. Some of them, for example, by Blum et al., (2004), Wu and Huan, (2007) and He and Tong, (2007) only solved the delay and chaff kind of APA. For this reason, to develop SSD a way of overcoming APA problems needs to be developed.

After having observed most SSD research, it is found that the enhancement or creation of a new SSD technique will not solve the active perturbation problems totally. In the other word, a new proposed SSD approach or technique will not promise that

active perturbation problem will be solved, totally. If there are any solutions, they would only overcome just one or two types of the active perturbation problem. What would happen if a new active perturbation technique appears? For that reason, after looking at the journey of the SSD research, the proposed research hopes to find another way to solve the SSD problem and ensure that it is robust against the APA problem.

As with other security tools such as IDS, the SSD can also be divided into two categories; host and network-based (Almulhem and Traore, 2007b). Based on the observation of the advantages of both categories, it is found that the combination of both host and network-based SSD will provide the combined advantages of SSD. In other words, the advantages of host-based SSD and network-based SSD can be combined to create a hybrid advantage of SSD. Called a hybrid and as known, a hybrid is a mixture of two very different things.

To solve the active perturbation problem, which in turn will ensure that the SSD problems will be solved, an intelligent technique will be used. Looking at research patterns in SSD, a Recent, research by Almulhem and Troare (2010) and Wu and Huang (2010) involved the Artificial Intelligence (AI) technique in SSD-based research. However, based on the tests which had been executed the results were very much compromised. However, AI-based SSD is one technique to overcome the active perturbation problems.

1.2 Research Background

Active perturbation problems are well-known problems in the Stepping Stone Detection (SSD) approach. Several researches such as those by Wang et al. (2001), Venkateshaiah (2006) and Wu and Huang (2007b) have attempted to solve active perturbation

problems. However, since the introduction of the active perturbation problems by Want et al., (2001) to the present time, most of the SSD researches have attempted to solve the problems (active perturbation) separately. By separately, in this case, it is meant that each researcher just attempted to solve one or two active perturbation problems identified by Wang (2006). For example, the research by Dohono et al. (2002) tried to solve the delay problem. Then, Blum et al. (2004) and He et al. (2006) tried to solve the chaff problem and Zhang et al. (2006) focused on delay and chaff. In other words, each research tried to solve only one or two of active perturbation problems. In addition to the dropped packet problem that has been left unsolved until now (Wang, 2004) there are other possible active perturbation types that can cause the SSD approach to be unstable at anywhere and anytime due to the existing or new kinds of active perturbation attacks. There are packet reorder, flow repacketization, flow split and merge, and so forth (Wang, 2004) and the combination of these is very disastrous to the SSD approaches. In the other word, the SSD approaches become very unstable when more kinds of active perturbation exist in the perturbation. As such the SSD needs to overcome all of these problems and not just one or two certain types of active perturbation so as to make sure that SSD can be effective. The 0% False Positive Rate (FPR) and 100% True Positive Rate (TPR) as used by Almulhem and Troare (2007b), Zhang and Paxson (2000) and Blum et al. (2004) have become the benchmarks of FPR and TPR.

1.3 Research Problem

This research intends to solve problems related to active perturbation attacks in the Stepping Stone Detection (SSD) approach. The main question is, “How to produce an

SSD approach that is robust to active perturbation problems while at the same time be able to maintains its TPR and FPR?"

Subsequent questions of the main research question are as follows:

i) What is the active perturbation problem that is influencing the SSD approach and how can this be proven to exist?

ii) Why is there still an active perturbation attack on the SSD approach although the SSD approach research has been running for more than a decade?

iii) How can a SSD approach be developed which can overcome the active perturbation problem and how can this be validated?

iv) What dataset can be used in the validation processes?

Sub-questions i) and ii) have been answered through the preliminary studies and preliminary literature review (Chapter 2.2) to show the importance of solving active perturbation problems and to prove the influence of active perturbation problems in current SSD approach. Sub-question iii) proposes a new approach and at the same time tries to verify the proposed approach. Lastly, sub-question iv) will be used to evaluate the proposed approach through the usage of real data.

1.4 The Goal of the Research

The goal of the research is:

i) to solve the active perturbation attack (APA) problems in Stepping Stone Detection (SSD) approach through the usage of Self-Organization Map (SOM) and Associative Rule (AR) techniques

1.5 Objectives

The objectives of the research are:

i) to show the existence of active perturbation problems in SSD approach by conducting a simulation

ii) to develop a SSD approach that can overcome active perturbation problems by using a hybrid and intelligent approach.

iii) to prove the capability of the proposed SSD approach by conducting a well-planned experiment.

1.6 Significance of the Research

Since the early days of SSD research (Staniford-Chen and Herberlein, 1995) to the recent AI-based SSD research (Almulhem and Troare, 2010), the SSD approach has evolved from different detection perspectives and not from the active perturbation attack perspective. This can be verified by looking at a research by Wu and Huang (2008a) that only focused on the detection of stepping stones and not the active perturbation attacks. Research by Jianhua and Huang (2007) also showed that the issue of active perturbation attacks was still left unsolved although they realized the existence of the problem. Moreover, from a preliminary research on one of the active perturbation problems, it was been proven active perturbation attacks can influence the SSD flows. The importance of active perturbation problems has also been recognized by researchers such as Wang et al. (2001), Venkateshaiah (2006) and Wu and Huang (2007).

The importance of solving active perturbation attacks can also be viewed from its capabilities as a response in the detection-response approach in Intrusion Detection System (IDS) and Intrusion Response System (IRS) techniques. Carver (2010) listed

tracing intruder (SSD as one of its techniques) as one response system. According to Cohen (1997), a late response to the intrusion can spell big disasters although the IDS had successfully detected the intrusion. Cohen (1997) becomes a chosen citation in this case because this is the basic reference that needs to be referred to when talking about IDS's response or IRS. Moreover, ineffective response techniques can be a big problem to IDS itself, such as giving a response to an innocent Internet user or host. Moreover, by using the proposed approach, not only can the present active perturbation be detected but a new similar problem that has just appeared can also be solved.

Therefore, an effective purpose of SSD in this research is not only to give a good response to IDS but also to ensure that SSD is robust towards the active perturbation problem which has not been fully been solved since it was first introduced nearly fifteen years ago (Staniford-Chen and Herberlein, 1995). In addition, solving the active perturbation problem can also redirect the SSD approach in the right direction in the Internet environment. The significance of the research can also be looked from the research contributions that have been identified.

Research contributions of this proposed research can be seen from its efforts to solve the active perturbation problem itself. Preliminary studies on active perturbations have shown that active perturbations can influence the SSD approach (Blum et al., 2004). Until today, there no research in SSD has tried to solve the active perturbation problem although researchers recognize its effect on the entire SSD approach. Therefore, through a preliminary study, the question of how big the influence active perturbations have on the SSD approach can be answered. This contribution not only can provide a clear picture of the problems of active perturbation attacks but at the same

time, present a proper solution to the active perturbation attack problem which has been unsolved for fifteen years.

Another contribution of this research is making the SSD approach ready for the Internet environment. As mentioned before, it has been proven through the preliminary research that active perturbation can make SSD approach function less effectively. The preliminary experiment used to prove this was within the Local Area Network (LAN). Imagine, what would happen to the SSD approach if it is known that Internet environment is very prone to these active perturbation problems. By making the SSD approach robust towards the active perturbation problems, it is now ready to work in the Internet environment.

In addition, the use of the intelligent technique in SSD not only provides a solution to the current active perturbation problem, but also it can be used on new active perturbations such as dropped packets. The contributions are in the form of development of the SSD approaches that accurate, fast and robust. All of these contributions will be explained later in this thesis.

The contribution of this proposed research also can be looked from several contributions such as HI-SSD, a solution to APA, SSD-based taxonomy, simulation on dropped packets in SSD, and the architecture of HI-SSD.

1.6.1 HI-SSD

HI-SSD is advantages from different perspectives such as the accuracy of detection which is robust against the APA, the existence of intelligent approaches towards new APA, and an active approach which is flexible either to be used as a detector or a response.

1.6.1 (a) Accurate SSD

The HI-SSD that has been produced in this research has the advantage of accurately detecting the stepping stones. This has been proven by a series of different experiments that have been executed during the research period. As described previously, before the HI-SSD was produced as the final research product, its generation was based on I-NSSD, I-HSSD and so forth. Through these experiments also, the proposed research showed that it produced accurate results as compared to current SSD approaches. Moreover, the experiments on HI-SSD also showed compromised results.

1.6.1 (b) Robust

Another advantage of the proposed research regards its robustness to APA. Through the experiment on two types of APA (chaff and delay) and a type that has not been tested before (dropped packet), it is shown that the proposed research is also robust towards APA. Furthermore, the combination of each type of APA also affects the overall function of the proposed approach.

1.6.1 (c) Intelligent Approach

The dropped packet can be categorized as a new APA type. By solving the dropped packet problem in the research, it has been proven that the proposed approach can also be used to solve a new APA.

1.6.1 (d) Active Approach

This research uses an SSD approach that applies active detection. This active approach means that the proposed approach does not need to capture network packets all of the time. Network packets can be captured when necessary. This approach gives the

advantage of saving properties such as network load, CPU processing and also memory space.

Most current approaches in SSD-based research are the passive-based approaches. For example a research by Staniford-Chen and Herberlein (1993), Yoda and Etoh (2000) and Zhang and Paxson (2000) chose passive approaches instead of the active approach.

1.6.1 (e) Flexible

The last advantage of the proposed SSD approach is its flexibility. Flexible in this research means that HI-SSD can be used either as detection or response.

The HI-SSD as a detection puts the HI-SSD as a detector. For that purpose, HI-SSD works as any other security-based tool such as antivirus, firewalls, IDS and so forth. HI-SSD works in the front line to detect the existence of stepping stones before it passes the work to the next tool such as the IDS to determine either the detected flow is an intrusion or not. In the other words, the stepping stone needs to be detected first before the next step of security can be executed.

The HI-SSD can also be used as a response. As explained previously, SSD is one of the response techniques. Moreover, most of SSD-based research assumed that stepping stone detection is used as response just like IDS. This can be viewed, for example, in the research by Yung (2002), Wu and Huang (2008b) and Yang and Huang (2005), which assumes the number to be more than three for it to be a stepping stone intrusion.

1.6.2 A Problem Solving on APA

As described in Chapter 2, APA was a kind of problem that just existed and no research attempted to solve the APA totally. Just a single or two kinds of APA was successfully solved. For example through the research that was conducted by Zhang et al. (2006), active perturbation problem on chaff and delay was focused. Ventakeshaiah and Wright (2007) and Padhye (2008), on the other hand, focused on the delay kind of active perturbation only. Another research conducted by He et al. (2006) only attempted to solve the chaff kind of active perturbation. From the three researches given here, it is shown that each research just attempted to solve one or two kinds of active perturbations, although according to Wang (2004), there are other types of active perturbation that need to be solved. Furthermore, dropped packets have not been solved until this thesis was written.

This proposed research will not only solve the delayed type of active perturbation, but also the chaff type of active perturbation. Not only that, the research will also successfully solve the combination between delay, chaff and perturbation. In other words, the research will successfully solve the combination between delay and chaff, delay and chaff together with dropped packet, chaff and delay, chaff and delay together with dropped packet and so forth. All of this information can be studied further in Section 6.9.

Delay and chaff represents the current problem of APA and dropped packet represents the unsolved APA. The combination of both parts shows the capabilities of the proposed HI-SSD against current and future kinds of APA. Moreover, in comparison to other approaches (content-based and AI-based SSD), the proposed HI-

SSD successfully detects the stepping stone when compared with other approaches. The results of the experiment are shown in Table 6.7.

By solving the APA, the attacker or intruder, now, cannot hide his or her tracks anymore. This will bring a step ahead, applications such as Intrusion Detection System (IDS) or Intrusion Response System (IRS) in providing more effective types of active responses. By doing this, the entire solution on detection-response will be more reliable. Not only both IDS and IRS can become active solutions, but at the same time reduce the time gap between detection and response that according to Cohen (1997) can influence the overall detection and response processes.

1.6.3 SSD-based Taxonomy

The extensive study of SSD has let the proposed research in producing a complete taxonomy on the overall SSD-based research. This taxonomy on SSD-based research is important to other researchers who might be interested in SSD-based research.

The full taxonomy on SSD-based research can be seen in Figure 2.2. The purpose of the taxonomy is to predict the future of SSD-based research, based on the past and present SSD-based research. By doing this, the research direction is more directed. Through this taxonomy also, the views of the overall SSD-based research also can be observed clearly.

By referring to Figure 2.2, it is shown that AI-based SSD, APA, Hybrid SSD, TPR and FPR, and timing-based are topics predicted to appear in the future. In fact, it is true for the time being as this thesis is written, most of the topics are valid and are used in this research although the taxonomy was created at the early stages of this research.

1.6.4 Simulation on Dropped Packet in SSD

Through the simulation of the dropped packet problem, it is shown that the problem really occurs in SSD. Section 4.4, 5.4 and 6.5 discuss the topics. Through a simulation that has been conducted, it can be shown that the dropped packet problem absolutely influences the overall SSD function.

The contribution can be looked from the execution of the simulation rather than the real test bed setting in the proposed research. The simulation executed at the early stages of the research lets the verification of dropped packet problem to be known faster. Therefore, the next stage or phase of the research can be conducted as soon the results are known.

1.6.5 Architecture of HI-SSD

The contribution of the proposed research also can be considered through the development of H-SSD and HI-SSD's architecture. From the observations on the related SSD-based research, there is no published architecture for either H-SSD or HI-SSD.

By proposing the architecture, it has made possible, to other researchers, to create other hybrid-based solution on SSD-based environment. The full architecture for H-SSD and HI-SSD can be referred to in Figure 4.7 and Figure 4.13 respectively. From both figures, the differences can be observed from the usage of I-HSSD and I-NSSD instead of HSSD and NSSD in HI-SSD.

1.7 Research Methodology

As the goal of this research is to solve the APA problem through the creation of the combination between SOM and AR approaches, the research methodology used during

the research period should be focused to the generation of the final approach (HI-SSD). Figure 1.1 shows the steps that have been taken.

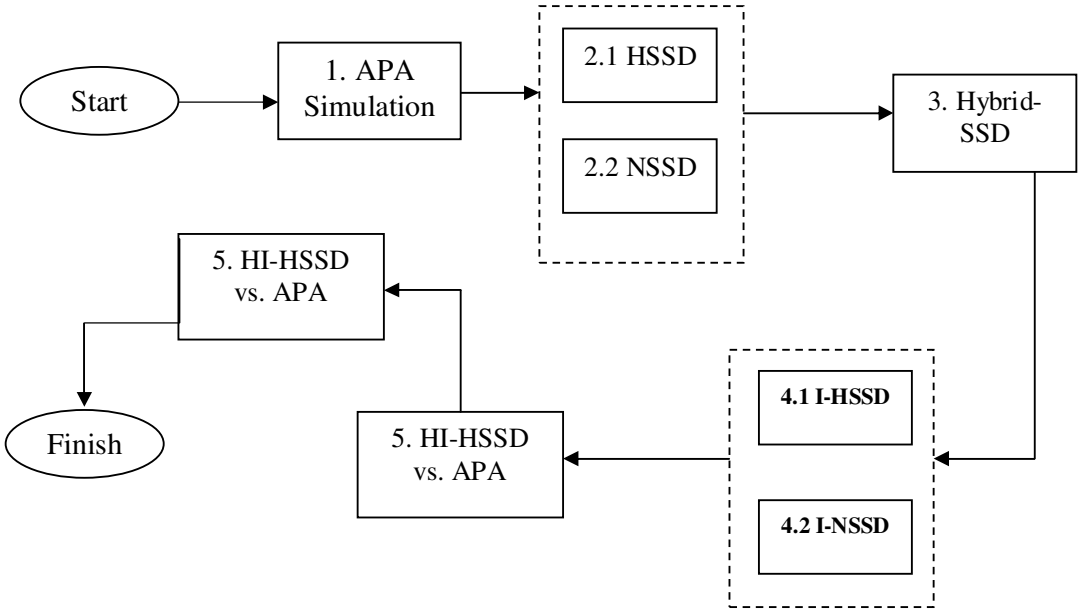

Figure 1.1: Steps in Research Methodology

From Figure 1.1 shows five main steps in this research methodology. The research methodology begins with the APA simulation so as to prove the existence of the APA problem in SSD before the real experiment begins. It is followed, next, by the Host-based SSD (H-SSD) and the Network-based SSD (N-SSD) experiments. The experiments intend to examine the host-based and network-based SSD algorithms that then will be used. Then, the experiment of the hybrid SSD continues. In this case, time that need to be used between the ordinary and hybrid SSD is focused. After that, intelligent host-based and intelligent network-based SSD take part. In this experiment, each host and the network-based intelligent SSD will be compared with the ordinary SSD respectively. Later, in the HI-SSD experiment, comparison will be made with other I-HSSDs and I-NSSDs. Finally, the experiment will close with the experiment on the proposed approach using the HI-SSD with APA dataset. In this case, another approach

that represents ordinary-based SSD and AI-based SSD will also take part in the experiment.

Each identified steps in the methodology described previously will be used in the operational framework that will be discussed in Section 3.2. In general, each experiment needs to achieve the goal of the experiment before the next experiment can be done. An explanation on the overall framework will be done in Chapter 3.

1.8 Scope of Research

The scope of the active perturbation attack here is limited to three types of active perturbation; i) delay, ii) chaff and iii) dropped packet because these three kinds of active perturbation are the most discussed and identified by previous researches in the SSD field. Previous researches such as Blum et al. (2004), Dohono et al. (2004) and Venkateshaiah (2006) have proven this and preliminary studies on this proposed research once again proves that dropped packets can influence the SSD processes.

When discussing SSD, most researches agreed to limit their scope to just detecting the stepping stones on hosts (Staniford-Chen and Herberlein, 1995), (Wang et al., 2001), (Yang and Lee, 2008). They did not bother to involve using the stepping stones. To identify, the person, who is really responsible for using the stepping stones, needs further investigation. This is more related to the computer forensics although it is possible to solve it by using a host-based SSD proposed in this research. For example, the authentication between the user and the host needs to be fixed first, so that all of the user's information can be recorded for the forensics, proposed.

Until today, most use of SSD is only valid in the interactive-based connections. For example, a research by Zhang and Paxson (2000) uses Telnet. He and Tong (2007)

uses SSH and according to Zhang et al. (2006), the attacker would usually employ the Telnet, SSH, IP Sec, VPN channels and anonymity system to establish stepping stone chains. For that reason, the proposed research also uses the interactive-based connections. In addition, most cases of intrusions or attacks use interactive-based connections as the medium. One hacker group known as 10pht agreed that interactive-based connections are most usable for attacking or hacking purposes (Mitnick and Simon, 2006). Another example was the intrusion into the Lawrence Berkeley Laboratory as mentioned in the book, 'The Cuckoo's Egg' by Clifford (2000). In the book, Clifford (2000) mentions that it is a difficult task for the administrator to track the intruder because the intruder usually uses a stepping stone technique to hide his or her tracks. In this case, although the administrator has a full capability control to his or her network, he still cannot detect the real location of the intruder.

1.9 Organization of the Thesis

There are six chapters in this thesis: Chapter 1: Introduction, Chapter 2: Literature Review, Chapter 3: Research Methodology, Chapter 4: Hybrid Intelligent SSD Approach, Chapter 5: Results of the study and discussion of the analysis and Chapter 6: Conclusion and Future work. A brief explanation of each chapter is given below.

Chapter 1 provides a general introduction of the stepping stone detection and the background of the problems that have to be solved in this research. It also describes in detail the problems, objectives, framework and lastly its importance and scope.

In Chapter 2 there is the literature review on the topics related to this research. Beginning with the introduction of the literature review, each topic is then described further in detail. The first topic in the literature review is about network security. It is

then, followed by topics such as stepping stones, hybrid, intelligent, active perturbation and lastly, the efforts to solve active perturbation through the use of hybrid intelligent SSD.

The following topic (Chapter 3), discusses the research methodology used in this research. This topic, as usual, begins with introduction. Then, it is followed by the operational framework or research procedures. Next, it continues with a discussion on the formulation of the research problems. After that, the research design and hypothesis are discussed. The topic then continues with an explanation on research design and hypothesis, experiment testbed, data analysis, limitations of the research, list of assumptions and lastly, the summary of the research methodology itself.

The bulk, of the contents, of this thesis is located in Chapter 4. Generally, this chapter explains how the proposed approach, known as Hybrid Intelligent Stepping Stone Detection (HI-SSD), is formed. Beginning with the architecture of the Hybrid Stepping Stone Detection (H-SSD), this chapter proceeds to explain the intelligent concept that is embedded in the HSSD. The final and complete HISSD to be used to solve the active perturbation problem is discussed before the summary of the chapter is given.

In Chapter 5, all related experiments, results and analyses are collected here. Beginning with an explanation of the simulation experiment that runs especially for the active perturbation problem, this chapter then goes to describe the experiments and results of HSSD, Intelligent Host-based Stepping Stone Detection (I-HSSD), Intelligent Network-based Stepping Stone Detection (I-NSSD) and the most important part, HI-SSD.

In the final chapter (Chapter 6), the summary of the thesis is discussed. The limitations of the research are also discussed. This chapter concludes with possible future works that can be done.

CHAPTER 2

LITERATURE REVIEW

Beginning with the introduction which gives a brief explanation of the literature to be reviewed, the chapter continues with the description of the terminology used in this research. After that, the stepping stone detection is explained in detail beginning with the past, present and future of SSD. The explanation also includes the issues that need to be handled. The discussion then turns to SSD problems. From the problems, a few suggestions are given. These include the hybrid approach to increasing accuracy in detection and to overcome the problems of processing time and the intelligent techniques for robustness and APA problem-solving. Two kinds of AI techniques have been successfully identified the based on host-and network-based SSD properties.

2.1 Introduction

The Internet has become more important than ever before, but at the same time, Internet attacks have increased significantly (CERT, 2010). Attackers can use intermediate hosts as their stepping stones before attacking the real targets (Zhang and Paxson, 2000). The compromised hosts enable the attackers to hide their tracks. Appendix A shows an

illustration of the overview concept of the Stepping Stone Detection (SSD) environment. In the illustration, SSD involves more than two hosts and each host is connected by a chain.

According to Zhang and Paxson (2000), SSD is a process to find a chain of stepping stones. Since the first research on SSD by Staniford-Chen and Heberlein (1995) to the latest by Wu and Huang (2010), many related issues have appeared. For example, research by Wang (2004) provided an SSD system that is active. Active in the SSD system means that the system only needs to execute the SSD process when necessary instead of passive that requires SSD system to execute the SSD process all of time. Research by Yoda and Etoh (2000) introduced an SSD that is robust towards encrypted connections. Research by Zhang et al. (2006), on the other hand, focused on solving active perturbation problems such as chaff and delay. Blum et al. (2004), in their research, tried to detect stepping stones by introducing the confidence bound. After careful observation of previous research topics, the focus in this thesis is to find solutions to the address of accuracy, the time-consuming nature of SSD approach and its robustness.

The Active Perturbation Problem (APA) is the problem given the most focus by researchers in SSD-based research. For example, a study conducted by Wang and Reeves (2003) was the first to focus on this problem. This was followed by He et al. (2006), Venkateshaiah and Wright (2007) and Zhang et al. (2006) which proposed various solutions to APA. However, their solutions have not totally solved the APA problem. In short, there have been no solutions to the combinations of different types of APA and to newly-identified APAs.