UNIVERSITI PUTRA MALAYSIA

AN INTEGRATED ANOMALY INTRUSION DETECTION SCHEME USING STATISTICAL, HYBRIDIZED CLASSIFIERS AND SIGNATURE APPROACH

WARUSIA MOHAMED YASSIN

FSKTM 2015 43

AN INTEGRATED ANOMALY INTRUSION DETECTION SCHEME USING STATISTICAL, HYBRIDIZED CLASSIFIERS AND SIGNATURE APPROACH

By

WARUSIA MOHAMED YASSIN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

April 2015

DEDICATIONS

To My Family and Friends

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

AN INTEGRATED ANOMALY INTRUSION DETECTION SCHEME USING STATISTICAL, HYBRIDIZED CLASSIFIERS AND SIGNATURE APPROACH

By

WARUSIA MOHAMED YASSIN

April 2015

Chairman: **Nur Izura Udzir, Ph.D.**
Faculty: **Computer Science and Information Technology**

Intrusion detection systems (IDSs) effectively balance additional security in a computer system by identifying intrusive activities on a computer system, and their enhancements are developing at a surprising rate. Detection methods based on statistical and data mining techniques are widely deployed as anomaly-based detection system (ADS). Although the statistical-based anomaly detection (SAD) method fascinates researchers, the low attack detection rates (also known as the detection of true positive) that reflect the effectiveness of the detection system generally persist. Specifically, this is due to the packets affected by the outlier data points (i.e., the data points that have a huge dissimilarity with the common data points) and the defined threshold size that is usually performed without any further analysis on the observed packet. It provides a significant effect in the process to determine which packet is more likely attributes to the anomalous behaviour. In recent years, data mining based anomaly detection (DMAD), particularly classification methods, have been incessantly enhanced in differentiating normal and attack behaviour. Unfortunately, in such methods the outcomes, i.e., true positive, true negative, false positive and false negative detections that directly influence the rates of accuracy, detection, and false alarms are not much improved and thus raise a persistent problem in the employment of such systems. The specific drawback that causes this is the failure to differentiate the packets behaviour that resembles a similar behaviour more precisely, such as a normal behaviour having a similar anomalous content behaviour and vice versa. These inaccurate outcomes can compromise the reliability of IDSs and cause them to overlook the attacks. As ADS can process massive volumes of packets, the amount of processing time needed to discover the pattern of the packets is also increased accordingly and resulting in late detection of the attack packets. The main contributor for such a shortcoming is the need to re-compute every process for each packet despite the attack behaviour having been examined.

This study aims to improve the detection of an anomalous behaviour by identifying the outlier data points in the packets more precisely, maximizes the detection of packets with similar behaviours more accurately while reducing the detection time. An Integrated Anomaly Detection Scheme (IADS) is proposed to overcome the aforesaid

drawbacks. The proposed scheme integrates an ADS and signature-based detection system (SDS) approach for better and rapid intrusion detection. Therefore, Statistical-based Packet Header Anomaly Detection (SPHAD) and a hybridized Naive Bayes and Random Forest classifier (NB+RF) are considered for the ADS, and Signature-based Packet Header Intrusion Detection (SPHID) is proposed as the SDS. In SPHAD, statistical analysis is used to construct a normal profile using statistical formula, scoring the incoming packets, and computing the relationships between historic normal behaviour as a dependent variable against observable packet behaviours as the independent variable through linear regression. Then the threshold measurement (size) is defined based on R^2 and Cohen's-d values in order to improve the attack detection rate by identifying a set of outlier data points which are present inside the packets more precisely. Subsequently, NB+RF, a hybrid classification algorithm is used to distinguish similar and dissimilar content behaviours of a packet. The Naive Bayes (NB) classifier is employed to construct the values of the posterior and the prior probability of a packet, then this information as well as the header values and statistical analysis information are fed to the Random Forest (RF) classifier to improve the detection of actual attacks and normal packets. SPHID then extracts the distinct behaviour of the packets which are verified as attacks by NB+RF and compute it as attack signatures for faster future detections, as the detection time will be reduced for the attack whose signature is already included in the signature database.

The effectiveness of the IADS has been evaluated under different detection capabilities (i.e., false positive, false negative, true positive, true negative, false alarm, accuracy, detection rate, attack data detection rate, normal data detection rate) and detection times using the DARPA 1999 and ISCX 2012 intrusion detection benchmark datasets as well as with Live-data. Results from the experiments demonstrate that IADS could effectively detect attacks and normal packets more precisely compared to previous work and the ADS which performs intrusion detections without employing the SPHID method. In addition, the detection time of IADS is much improved as compared to ADS. Thus, IADS is a better solution for anomaly detection methods in detecting untrustworthy behaviour and to define attack and normal behaviours more accurately.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk Ijazah Doktor Falsafah

SKIM PENGESANAN PENCEROBOHAN ANOMALI BERSEPADU MENGGUNAKAN KAEDAH STATISTIK, PENGELAS HIBRID DAN PENDEKATAN TANDA KENAL

Oleh

WARUSIA MOHAMED YASSIN

April 2015

Pengerusi: **Nur Izura Udzir, Ph.D.**
Fakulti: **Sains Komputer dan Teknologi Maklumat**

Sistem pengesanan pencerobohan (IDS) memperseimbangkan alat tambahan keselamatan secara efektif dengan mengenal pasti aktiviti pencerobohan pada sistem komputer, dan penambahbaikan alat ini kerap berlaku pada kadar yang tidak dijangka. Kaedah-kaedah sistem pengesanan pencerobohan berasaskan anomali (ADS), yang menggunakan algoritma perlombongan data mampu mengenal pasti serangan-serangan yang tidak dikenali. Walaupun kaedah pengesanan anomali berasaskan statistik (SAD) memikat penyelidik, kadar pengesanan pencerobohan yang rendah yang juga dikenali sebagai pengesanan benar positif, mencerminkan keberkesanan sistem pengesanan umumnya berterusan. Khususnya, ia disebabkan oleh paket yang terjejas akibat titik-titik terpencil iaitu titik data yang mempunyai perbezaan besar dengan titik data biasa, dan saiz ambang yang biasanya ditakrifkan tanpa melakukan apa-apa analisa lanjutan terhadap paket yang diperhatikan. Ia memberi kesan yang ketara dalam proses untuk menentukan paket mana yang lebih cenderung kepada sifat-sifat tingkah laku yang beranomali. Sejak kebelakangan ini, pengesanan anomali berasaskan perlombongan data (DMAD), khususnya kaedah klasifikasi di tambah baik secara berterusan dalam membezakan tingkah laku normal dan pencerobohan. Malangnya, menerusi penggunaan kaedah ini, hasil output iaitu pengesanan packet normal dan pencerobohan yang secara langsung mempengaruhi kadar ketepatan, kadar pengesanan dan kadar *'false alarm'* tidak diperbaiki ke tahap yang lebih baik serta menimbulkan masalah dalam penggunaan sistem pengesanan anomali secara berterusan. Kelemahan khusus yang menyebabkan keadaan ini adalah akibat daripada kegagalan untuk membezakan tingkah laku kandungan paket yang menyerupai tingkah laku yang lain dengan lebih tepat, contohnya tingkah laku paket normal yang menyerupai tingkah laku paket beranomali dan sebaliknya. Hasil yang tidak tepat boleh menjejaskan kebolehpercayaan IDSs dan menyebabkan mereka terlepas pandang packet pencerobohan. Memandangkan ADS mampu memproses jumlah packets yang besar, jumlah masa pemprosesan yang diperlukan untuk menemui bentuk paket turut meningkat dan menyebabkan kelewatan dalam pengesanan paket pencerobohan. Penyumbang utama untuk kekurangan ini ialah keperluan untuk mengira semula setiap proses bagi setiap paket walaupun tingkah laku pencerobohan yang terlibat sudah diperiksa sebelum ini.
Kajian ini bertujuan untuk membaiki mahupun meningkatkan pengesanan tingkah laku

beranomali dengan mengenalpasti titik-titik data terpencil di dalam paket dan memaksimumkan pengesanan paket yang mempunyai tigkah laku yang sama dengan lebih tepat disamping mengurangkan masa pengesanan. Satu skim pengesanan anomali bersepadu (IADS) dicadangkan untuk mengatasi kelemahan-kelemahan di atas. Skim yang dicadangkan menyepadukan ADS dan pendekatan sistem pengesanan tanda kenal (SDS) untuk pengesanan pencerobohan yang lebih baik dan cepat. Oleh itu, pengesanan anomali pengepala paket berasaskan kaedah statistik (SPHAD) dan pengelas hibrid Naive Bayes dan Random Forest (NB+RF) yang dicadangkan dipertimbangkan sebagai sistem ADS, dan pengesanan intrusi pengepala paket berasaskan tanda kenal (SPHID) sebagai SDS. Analisa statistik digunakan untuk membina profil normal menerusi formula statistik, memberi skor kepada setiap paket yang masuk dan mengira perhubungan antara tingkah laku paket normal sejarah yang digunakan sebagai pembolehubah bersandar terhadap tingkah laku paket baharu yang boleh dicerap sebagai pembolehubah bebas melalui regresi linear di dalam SPHAD. Kemudian ukuran (saiz) ambang ditakrif berdasarkan nilai-nilai R^2 dan Cohen's-d untuk meningkatkan mahupun membaiki kadar pengesanan pencerobohan dengan mengenalpasti titik-titik data terpencil yang berada di dalam paket dengan lebih tepat. Selepas itu, NB+RF, algoritma pengelas hibrid digunakan untuk membezakan tingkah laku kandungan paket yang sama dan yang berbeza. Pengelas Naive Bayes (NB) digunakan untuk membina nilai-nilai kebarangkalian *'prior'* dan *'posterior'* sesuatu paket terlebih dahulu, kemudian nilai-nilai tersebut, kandungan nilai pengepala paket serta maklumat berkenaan analisa statistik disalurkan kepada pengelas Random Forest (RF) untuk meningkatkan mahupun membaiki pengesanan paket pencerobohan dan normal yang sebenar. SPHID mengekstrak tingkah laku paket yang unik yang ditentusahkan sebagai pencerobohan oleh NB+RF dan mengiranya sebagai tanda kenal pencerobohan untuk mengesan pencerobohan dengan lebih cepat pada masa akan datang, dimana masa pengesanan dapat dikurangkan sekiranya tanda kenal bagi sesuatu pencerobohan didapati wujud di dalam pangkalan data tanda kenal.

Keberkesanan IADS telah dinilai di bawah keupayaan pengesanan yang berbeza iaitu positif palsu, negatif palsu, positif benar, negatif benar,kadar *'false alarm'*, kadar ketepatan, kadar pengesanan, kadar pengesanan data pencerobohan dan kadar pengesanan data normal serta tempoh masa pengesanan menggunakan data-data penanda aras pengesanan pencerobohan seperti DARPA 1999, ISCX 2012 serta data hidup. Keputusan eksperimen menunjukkan bahawa IADS dapat mengesan paket-paket pencerobohan dan normal dengan lebih tepat berbanding dengan kajian sebelum ini serta ADS, yang merupakan skim yang melakukan pengesanan pencerobohan tanpa menggunakan kaedah SPHID. Tambahan pula, pengesanan masa IADS adalah baik berbanding dengan kaedah ADS. Oleh itu, IADS merupakan satu penyelesaian yang lebih memuaskan untuk kaedah ADS dalam mengesan tingkah laku yang tidak dipercayai dan mendefinisi paket pencerobohan dan normal dengan lebih tepat.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and deepest gratitude to my supervisor Associate Prof. Dr. Nur Izura Udzir and my committee members Dr. Azizol Abdullah, Dr. Taufik Abdullah, Dr. Hazura Zulzalil, and Madam Zaiton Muda for their continuous encouragement, valuable advice, and guidance throughout this research. I really appreciate the freedom they provided while I was working on my research and their openness to new ideas.

My special thanks go to my dearest friends who were always willing to help and share their ideas and knowledge even when busy with their own research. I will always treasure their friendship.

Most of all, I would like to express my sweetest appreciation to my family for their affectionate support, patience, and encouragement. Their prayers and good wishes constantly helped me to be strong, especially in difficult times. I am forever grateful and indebted to them.

I certify that a Thesis Examination Committee has met on 30 April 2015 to conduct the final examination of S.M.Warusia Mohamed Bin S.M.M Yassin on his thesis entitled "An Integrated Anomaly Intrusion Detection Scheme Using Statistical, Hybridized Classifiers and Signature Approach" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Dr. **Hamidah Ibrahim**
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Dr. Norwati Mustapha
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Dr. Azmi Jaafar
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Dr. Kwok Lam For
Associate Professor
City University of Hong Kong
Hong Kong
(External Examiner)

ZULKARNAIN ZAINAL, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 17 June 2015

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Nur Izura Udzir, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Hazura Zulzalil, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Azizol Abdullah, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Mohd Taufik Abdullah, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

_____

BUJANG BIN KIM HUAT, Ph.D.
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION

Declaration by Graduate Student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules, or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: _____

Name and Matric No: _____

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature:
Name of Chairman of Supervisory Committee:
Nur Izura Udzir, PhD

Signature:
Name of Member of Supervisory Committee:
Hazura Zulzalil, PhD

Signature:
Name of Member of Supervisory Committee:
Azizol Abdullah, PhD

Signature:
Name of Member of Supervisory Committee:
Mohd Taufik Abdullah, PhD

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AC	Accuracy
A-DR	Attack Detection Rate
ADM	Anomaly Detection Model
ADS	Anomaly-based Detection System
ALAD	Application Layer Anomaly Detector
ANN	Artificial Neural Network
CIA	Confidentiality, Integrity and Assurance
DARPA	Defence Advanced Research Projects Agency
DBMS	Database Management System
DM	Data Mining
DMAD	Data Mining-based Anomaly Detection
DR	Detection Rate
DS	Dynamic Score
DST	Dempster Shafer Theory
DT	Decision Tree
FA	False Alarm
FN	False Negative
FP	False Positive
NB+RF	Hybridized Naive Bayes and Random Forest Classifier
HIDS	Host-based Intrusion Detection Systems
HMM	Hidden Markov Models
IADS	Integrated Anomaly Detection Scheme
IDES	Intrusion Detection Expert System
IDS	Intrusion Detection System
ISCX	Information Security Center of Excellence
LNID	Lightweight Network Intrusion Detection System
LRA	Linear Regression Analysis
LVQ	Learning Vector Quantization
MCS	Multiple Classifier Systems
MIT-LL	MIT Lincoln Labs
MLP	Multi-Layer Perceptron
MRROC	Maximum Realizable Receiver Operating Characteristics
MyCERT	Malaysia Computer Emergency Response Team
NB	Naïve Bayes
N-DR	Normal Detection Rate
NETAD	Network Traffic Anomaly Detector
NIDS	Network-based Intrusion Detection Systems
NN	Neural Network
PAID	Packet Analysis for Intrusion Detection
PbPHAD	Protocol Based Packet Header Anomaly Detection
PHAD	Packet Header Anomaly Detector
PS	Packet Score
RF	Random Forest
ROC	Receiver Operating Characteristics
RP	Resilient Back Propagation
SA	Statistical Analysis
SAD	Statistical-based Anomaly Detection

SCG	Scaled Conjugate Gradient
SDS	Signature-based Detection System
SPHID	Signature-based Packet Header Intrusion Detection
SPHAD	Statistical-based Packet Header Anomaly Detection
SS	Static Score
SVM	Support Vector Machine
TN	True Negative
TP	True Positive

CHAPTER 1
INTRODUCTION

1.1 Background

Protecting an organization's assets against threats from the network has become a major challenge in the wake of increasing network-based attacks. In addition, the confidential assets and vulnerabilities of computer and network systems could be exposed to cyber attacks if not well protected with security defenders. Cyber attacks are invasive tactics or operations used by unethical parties either from corporations or individuals against vulnerable systems (i.e., computer systems, computer networks, computer infrastructures, and computer information) in an attempt to modify, steal and/or destroy them (Kuang, 2007). Denial-of-service, Web site defacement, password sniffing, web browser exploits, and breach of access are examples of the consequences which could result from cyber attacks. In addition, these attacks have become more sophisticated and harmful as the Stuxnet (Karnouskos, 2011; Vida et al., 2014) worm recently showed.

Consequently, it is extremely important to develop mechanisms for intrusion detection in view of the conviction that suspicious activities can be detectable by taking measures to avoid their further breeding against computer networks or systems. Intrusion detection is the process of monitoring the activities taking place in a computer or network system and scrutinizing them for indications of potential intrusions and in determining suspicious activities there. Thus, intrusion detection systems (IDSs) are formed to detect cyber attack activities attempting to compromise the confidentiality, integrity, and availability (CIA) of interconnected computing systems (Zhou, 2005). Nowadays, IDS are the most extensively applied and significant components in computer security.

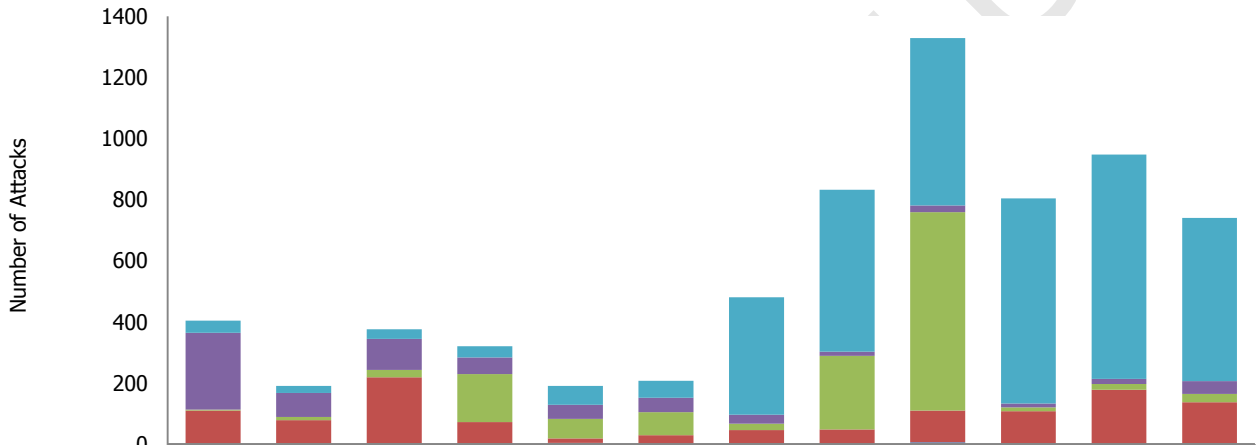
1.2 Motivation

Electronic transactions, online banking, hosting portals, etc., have raised Internet usage dramatically and cover almost the entire globe. Unfortunately, these trends also fuel hacking activities and dangerous cyber attacks that are able to breach even the strongest firewalls. Data from the Malaysia Computer Emergency Response Team (MyCERT)[1] show a significant growth in cyber attacks in 2014 (Figure 1.1). Total cyber incidents from 2000 to 2014 are presented in Figure 1.2.

`

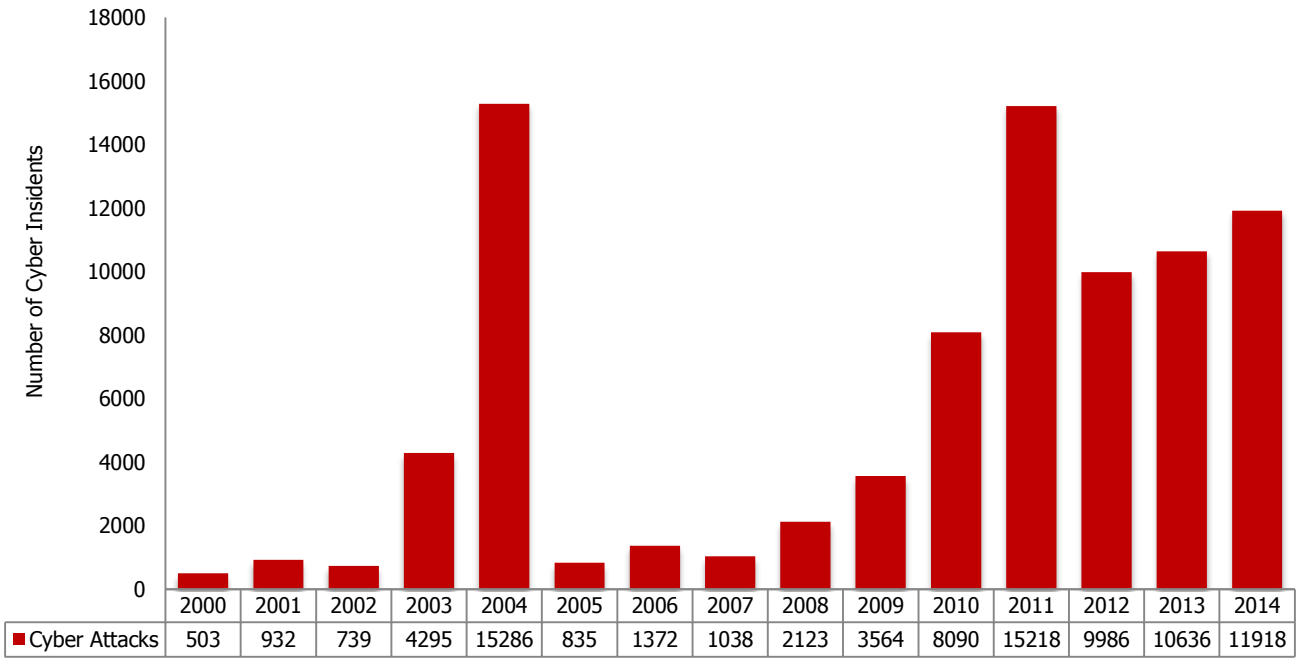
Cyber attacks have become an novel weapon of war around the world and their persistent growth against computer and network systems makes it critical to integrate more accurate IDS capable of maximizing correctly detectable data (i.e., true positives and negatives) and minimizing falsely detectable data (false positives and negatives) as

[1] http://www.mycert.org.my



	Jan	Feb	Mac	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dis
Spam	40	23	32	36	61	55	385	530	548	671	735	534
Malicious Codes	251	78	101	55	47	48	29	14	22	13	16	42
Intrusion Attempt	3	11	24	157	63	75	21	241	649	12	19	27
Intrusion	109	76	216	70	15	28	43	47	104	105	178	134
Denial of Service	1	2	3	2	4	1	3	1	6	3	0	3

Figure 1.1: Statistic of Reported Incidents, 2014



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Cyber Attacks	503	932	739	4295	15286	835	1372	1038	2123	3564	8090	15218	9986	10636	11918

Figure 1.2: Number of Reported Incidents, 2000-2014

as well as reducing the detection time to enable prompt identification of attacks.

Anomaly-based detection systems (ADS) which employ statistical analysis and data mining, particularly classification methods is a significant field to be explored for attaining the above mentioned capabilities. The necessity for continuous enhancement of intrusion detection capabilities, detection time, and its numerous approaches is the motivation for this research.

1.3 Problem Statement

Creating an anomaly-based detection system (ADS) model using statistical analysis and data mining approaches is demanding in a field of IDSs. Although various improved methods have been developed and introduced every year in statistical-based anomaly detection, the problem to identify the correct attack packet is still not satisfactory. Moreover, many such detection methods have a low attack detection rate (also referred as the detection rate of true positives) is an essential key indicator used to assess a statistical-based anomaly detection method. It is due to the use of anomaly scores in defining threshold measurement in identifying attack packet, which is affected from outlier data points (the data points that have a huge dissimilarity with the common data points called outlier data points) and the threshold size that usually defined without performing any further analysis on the observed packet. It gives a great impression in the process to determine the packet which is more likely to be anomalous. For example, such situation will get worse if there is more than one outlier data points in every single packet headers. Generally, this detection method generates maximum false alarms (false positives) due to the difficulty in accurately separating normal packet that is not visibly different from attack packet. Consequently, data mining approaches, particularly classification methods, are receiving growing interest within intrusion detection societies as they have proficiency for reducing false positives. The common challenge associated with classification methods is the performance of these detection systems in terms of detection rates, accuracy, and false alarm. However, the specific problem that causes this is a failure to differentiate the packets behaviour that resembles a similar behaviour more precisely. For example, an anomalous behaviour contains similar normal behaviours as the real normal packets and normal packet behaviours have similar anomalous content behaviours. This is the reason why the existing classification methods are less efficient in classifying attack and normal packet that contributes to false detections (false negatives and false positives) as well as fewer correct detections (true negatives and true positives). Thus, these inaccurate outcomes compromise the reliability of IDSs and cause them to overlook the attacks. Apart from detection capabilities, the detection time involved in using ADS methods are time consuming, resulting in delays in detecting whether a packet pattern is an attack or normal. For example, using these detection method procedures, each involved process need to be re-computed for each piece of packet despite the attack behaviour having been examined. In addition, time consuming issues can become worse if the packets relatively high.

Specifically, this thesis addresses the following issues:

1. A number of efforts offer statistical-based anomaly detections using packet header to identify abnormal behaviour such as Chen et al. (2010), Lee et al. (2008), Mahoney (2003), Mahoney and Chan (2001, 2002), Shamsuddin and Woodward, (2008), and Xiong et al. (2013). The major drawback of those detection methods is defining the threshold measurement in identifying the attack packets which is affected from outlier data points without performing any further analysis on the observed packets. Consequently, this statistical-based anomaly detection method is inadequate for identifying an attack packet more accurately and results in low attack detection rates (true positives).

2. Classification methods have been introduced and widely employed by various researchers in the field of ADS with the aim to reduce false detection rates as well as increase correct detection rates. Unfortunately, existing classification methods are less efficient in classifying an attack and normal packet and contribute to increases in false negatives and false positives with lower rates of true negatives and true positives. The major reason causes those limitations have been a failure to differentiate the packets behaviour that resembles a similar behaviour more precisely. There have been a number of earlier researches performing intrusion detection using the classification approach and these had more than 1% false positive or false alarm rates. These include Decision Tree (Kosamkar et al., 2014), Support Vector Machine (Kosamkar et al., 2014), and Naive Bayes (Sagale et al., 2014) with 9.79%, 4.94%, and 1.48% as false positive rates, respectively.

3. In most regular practices the ADS method only focuses on improving the detection performance by overlooking its capability in terms of detection time. Thus, the detection time for an intrusion detection process using ADS method is time consuming. An example of previous work are Tribak et al., (2012).

1.4 Research Questions

This thesis proposes an Integrated Anomaly Detection Scheme (IADS) based on a number of integrated methods, namely, statistical-based packet header anomaly detection (SPHAD), hybridized classifiers (NB+RF), and signature-based packet header intrusion detection (SPHID) that use attack signatures in examining packet header behaviours to address the following questions:

1. Do the statistical analyses applied to different measurements express the dissimilar and similar behaviours of the packet headers?

2. Does the usage of a threshold mechanism increase actual attack detections by overcoming the suspected outlier data points drawbacks?

3. Do the features derived from the statistical approach provide a clear picture on the data and assist the integrated classifiers to minimize false positives and

false negatives and to maximize true positives and true negatives?

4. Does the transformation of unique attack behaviour into a signature structure minimize the detection time in ADS as well as increase the number of packets processed in a second?

1.5 Objectives of Research

The main objective of this research is to propose an Integrated Anomaly Detection Scheme (IADS) which integrates anomaly-based detection system (ADS) and signature-based detection system (SDS) approach for better and more rapid intrusion detection. As such, three different kinds of detection methods have been proposed in this thesis.

The specific objectives are to:

1. Propose a normal scoring approach, linear regression analysis and Cohen's-d measurement to identify the outlier data points which able to differentiate attack behaviours more precisely as statistical-based anomaly detection.

2. Propose a hybridized Naive Bayes and Random Forest classifier to differentiate and identify a similar behaviour of an attack and normal more accurately.

3. Propose a signature-based packet header intrusion detection method to reduce detection times in the ADS method.

1.6 Scope of Research

This research focuses on the ADS method which utilizes statistical analysis and hybridized classifiers between Naive Bayes and Random Forest to accurately identify intrusive and non-intrusive packet header behaviour with minimum false positives and false negatives as well as maximum true positives and true negatives. In addition, the detection method is designed such that it could operate accurately in identifying intrusion packet behaviours on various machines (multiple host network-based intrusion detection system, NIDS) and on a single machine (host-based intrusion detection system, HIDS). The scope is also on reducing detection time in the ADS method by creating known attack signature behaviours. The DARPA 1999 and ISCX 2012 intrusion detection benchmark dataset as well as Live-Data are used to assess the proposed, individual, and existing detection methods.

1.7 Research Contributions

The major contribution of this research is the creation of an Integrated Anomaly Detection Scheme (IADS) that could identify a number of intrusive and non-intrusive behaviours (false positive, false negative, true positive and true negative) more accurately and to minimize detection times via a signature-based packet header intrusion detection method by producing attack signatures for observable behaviour in contrast to ADS methods (without employing signatures).

The following are the contributions of this research:

1. Formulating a statistical method that could score packets, appraise the degree of the observed packet relationship through linear regression analysis, and Cohen's-d as a threshold measurement to improve the detection rate of intrusion or attack by overcoming the outliers limitations. Experiments show that the proposed model is capable of maximizing actual attack-detectable data (true positives) more accurately compared to previous work.

2. Creating a hybridized classifier of Naive Bayes and Random Forest to differentiate and identify the similar actual behaviours of an attack and normal more accurately, particularly which able to decrease false negatives and false positives, and increase true negatives and true positives. These methods have shown remarkable outcomes and improvements for all aforesaid factors which directly improved the accuracy, detection, and false alarm rates as compared to the individual and existing methods.

3. Developing a Signature-based Packet Header Intrusion Detection method where signatures are created based on distinct attack behaviours after being classified by hybridized classifiers from the detection file for future detection and to decrease the detection time. Thus, the detection time is reduced upon utilizing signatures for detection purpose as compared to the Anomaly Detection Scheme (ADS) which performs intrusion detections without employing signatures.

1.8 Organization of Thesis

This section presents an outline of the entire thesis which is organized as follows:

Chapter 1 presents the introduction and includes among others the background, problem statement, research objectives and questions and contributions of the thesis.

Chapter 2 reviews related studies of the subject matter which includes intrusion detection systems (IDSs), statistical-based anomaly detection (SAD), and data mining-based anomaly detection (DMAD). The end of the chapter discusses the

related work within this field which employs statistical analysis and hybridized classifiers.

Chapter 3 provides a brief explanation of the research methodologies adopted in this research. The requirement analysis involved in the process of identification and investigation of the research requirement is detailed out. This chapter also describes how the proposed IADS is designed and implemented. In addition, the experimental design and experimental setup involving the amount of data applied and selection of specific applications to perform the research and evaluation criteria used to evaluate the performance is also highlighted.

Chapter 4 describes the proposed Integrated Anomaly Detection Scheme (IADS). A comprehensive discussion is provided on the components of IADS which is designed based on the Statistical-based Packet Header Anomaly Detection (SPHAD), Hybridized Naive Bayes and Random Forest Classifiers (NB+RF) and Signature-based Packet Header Intrusion Detection Method (SPHID). Each analysis involved in SPHAD and the NB+RF as well as the SPHID for formation of attack behaviour signatures is briefly explained in this chapter.

Chapter 5 presents the implementation of different detection methods in the proposed detection scheme using a MySql database, Matlab programming, and SQL script. The procedure for implementation is clearly explained by giving examples for each step which needs to be performed in this detection scheme.

Chapter 6 presents a performance evaluation of the IADS. The effectiveness of the proposed SPHAD, NB+RF and SPHID are assessed using a number of datasets and the detection results based on different criteria are illustrated and discussed.

Chapter 7 summarizes the entire thesis and recommendations on possible extensions of this research as future work.

REFERENCES

Abad, C., Taylor, J., Sengul, C., Yurcik, W., & Rowe, K. (2003). Log correlation for intrusion detection: a proof of concept. In *19th Annual Computer Security Applications Conference, 2003. Proceedings.* (pp. 255–264). IEEE.

Aberson, C. L. (2011). *Applied Power Analysis for the Behavioral Sciences*. Taylor & Francis.

Abhaya, K., Jha, R., & Afroz, S. (2014). Data Mining Techniques for Intrusion Detection: A Review. *International Journal of Advanced Research in Computer and Communication Engineering*, *3*(6), 6938–6942.

AL-Nabi, D., & Ahmed, S. (2013). Survey on Classification Algorithms for Data Mining:(Comparison and Evaluation). *Computer Engineering and Intelligent Systems*, *4*(8), 18–25.

Amor, N. Ben, Benferhat, S., & Elouedi, Z. (2004). Naive Bayes vs decision trees in intrusion detection systems. In *Proceedings of the 2004 ACM symposium on Applied computing - SAC '04* (p. 420). New York, New York, USA: ACM Press.

Anderson, D., Frivold, T., Valdes, A., & Tamaru, A. (1995). *Next-generation Intrusion Detection Expert System (NIDES) - a summary*. Menlo Park, CA 94025-3493.

Atefi, K., Yahya, S., Dak, A. Y., & Atefi, A. (2013). A Hybrid Intrusion Detection System Based On Different Machine Learning Algorithms. In *Proceedings of the 4th International Conference on Computing and Informatics, ICOCI 2013* (pp. 312–320). Sarawak: Universiti Utara Malaysia.

Baum, L. E., & Petrie, T. (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics*, *37*(6), 1554–1563.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Bronstein, A., Das, J., Duro, M., Friedrich, R., Kleyner, G., Mueller, M., Cohen, I. (2001). Self-aware services: using Bayesian networks for detecting anomalies in Internet-based services. In *2001 IEEE/IFIP International Symposium on Integrated Network Management Proceedings. Integrated Network Management VII. Integrated Management Strategies for the New Millennium (Cat. No.01EX470)* (pp. 623–638). IEEE.

Brumley, D., Newsome, J., Song, D., & Jha, S. (2008). Theory and Techniques for Automatic Generation of Vulnerability-Based Signatures. *IEEE Transactions on Dependable and Secure Computing*, *5*(4), 224–241.

Burgess, M., Haugerud, H., Straumsnes, S., & Reitan, T. (2002). Measuring System Normality. *ACM Trans. Comput. Syst.*, *20*(2), 125–160.

Chen, C.-M., Chen, Y.-L., & Lin, H.-C. (2010). An efficient network intrusion detection. *Computer Communications*, *33*(4), 477–484.

Cho, Y., Kang, K., Kim, I., & Jeong, K. (2009). Baseline Traffic Modeling for Anomalous Traffic Detection on Network Transit Points. In *Proceeding APNOMS'09 Proceedings of the 12th Asia-Pacific network operations and management conference on Management enabling the future internet for changing business and new computing services* (pp. 385–394). Berlin, Heidelberg: Springer-Verlag.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297.

Denning, D. E. (1987). An Intrusion-Detection Model. *IEEE Transactions on Software Engineering*, *SE-13*(2), 222–232.

Ektefa, M., Memar, S., Sidi, F., & Affendey, L. S. (2010). Intrusion detection using data mining techniques. In *2010 International Conference on Information Retrieval & Knowledge Management (CAMP)* (pp. 200–203). IEEE.

Ellis, P. D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge University Press.

Eskin, E., Arnold, A., Prerau, M., Portnoy, L., & Stolfo, S. (2002). A Geometric Framework for Unsupervised Anomaly Detection: Detecting Intrusions in Unlabeled Data. In *Applications of Data Mining in Computer Security*. Kluwer.

Estévez-Tapiador, J. M., García-Teodoro, P., & Díaz-Verdejo, J. E. (2004). Measuring normality in {HTTP} traffic for anomaly-based intrusion detection. *Computer Networks*, *45*(2), 175–193.

Farid, D. M., Zhang, L., Rahman, C. M., Hossain, M. A., & Strachan, R. (2014). Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems with Applications*, *41*(4, Part 2), 1937–1946.

Farmer, J. D., Packard, N. H., & Perelson, A. S. (1986). The immune system, adaptation, and machine learning. *Physica D: Nonlinear Phenomena*, *22*(1-3), 187–204.

Faysel, M. A., & Haque, S. S. (2010). Towards Cyber Defense : Research in Intrusion Detection and Intrusion Prevention Systems, *10*(7), 316–325.

Fernández-Blanco, E., Aguiar-Pulido, V., Munteanu, C. R., & Dorado, J. (2013). Random Forest classification based on star graph topological indices for antioxidant proteins. *Journal of Theoretical Biology*, *317*, 331–7.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *The British Journal of Mathematical and Statistical Psychology*, *63*(Pt 3), 665–94.

Gaffney, J. E., & Ulvila, J. W. (2001). Evaluation of intrusion detectors: a decision theory approach. In *Proceedings 2001 IEEE Symposium on Security and Privacy. S&P 2001* (pp. 50–61). IEEE Comput. Soc.

García-Teodoro, P., Díaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, *28*(1–2), 18–28.

Gargiulo, F., Mazzariello, C., & Sansone, C. (2013). Multiple Classifier Systems: Theory, Applications and Tools. In M. Bianchini, M. Maggini, & L. C. Jain (Eds.), *Handbook on Neural Information Processing SE - 10* (Vol. 49, pp. 335–378). Springer Berlin Heidelberg.

Gates, C., & Taylor, C. (2007). Challenging the Anomaly Detection Paradigm: A Provocative Discussion. In *Proceedings of the 2006 Workshop on New Security Paradigms* (pp. 21–29). New York, NY, USA: ACM.

Golmah, V. (2014). An Efficient Hybrid Intrusion Detection System based on C5. 0 and SVM. *International Journal of Database Theory & Application*, *7*(2), 59–70.

Hasan, M., Nasser, M., Pal, B., & Ahmad, S. (2014). Support Vector Machine and Random Forest Modeling for Intrusion Detection System (IDS). *Journal of Intelligent Learning Systems and Applications*, *2014*(February), 45–52.

Hosseinpour, F., Vahdani Amoli, P., Farahnakian, F., Plosila, J., & Hamalainen, T. (2014). Artificial Immune System Based Intrusion Detection: Innate Immunity Using an Unsupervised Learning Approach. *International Journal of Digital Content Technology and Its Applications*, *8*(5), 1–12.

Ingham, K. L., & III. (2007). Anomaly Detection for HTTP Intrusion Detection: Algorithm Comparisons and the Effect of Generalization on Accuracy.

Ippoliti, D. (2014). Automated network anomaly detection with learning, control and mitigation.

Jain, N., & Srivastava, V. (2013). DATA MINING TECHNIQUES: A SURVEY PAPER. *IJRET: International Journal of Research in …*, *2*(11), 116–119.

Jashan, J., & Bag, M. (2012). Cascading of C4.5 Decision Tree and Support Vector Machine for Rule Based Intrusion Detection System. *International Journal of Computer Network and Information Security*, *4*(8), 8–20.

Javitz, H. S., & Valdes, A. (1991). The SRI IDES statistical anomaly detector. In *Proceedings. 1991 IEEE Computer Society Symposium on Research in Security and Privacy* (pp. 316–326). IEEE Comput. Soc. Press.

Jiawei Han, M. K. (2006). *Data Mining concepts and techniques* (Second., p. 800). USA: Morgan Kaufmann.

John, G. H., & Langley, P. (1995). Estimating Continuous Distributions in Bayesian Classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (pp. 338–345). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Joseph F. Hair Jr, William C. Black, Barry J. Babin, R. E. A. (2009). *Multivariate Data Analysis* (7th ed., p. 816). Prentice Hall.

Julock, G. (2013). *The effectiveness of a random forests model in detecting network-based buffer overflow attacks*.

Karnouskos, S. (2011). Stuxnet worm impact on industrial cyber-physical system security. In *IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society* (pp. 4490–4494). IEEE.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, *17*(2), 137–52.

Kind, A., Stoecklin, M., & Dimitropoulos, X. (2009). Histogram-based traffic anomaly detection. *IEEE Transactions on Network and Service Management*, *6*(2), 110–121.

Koc, L., Mazzuchi, T. A., & Sarkani, S. (2012). A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier. *Expert Systems with Applications*, *39*(18), 13492–13500.

Kosamkar, V., & Chaudhari, S. S. (2014). Improved Intrusion Detection System using C4 . 5 Decision Tree and Support Vector Machine. *International Journal of Computer Science and Information Technologies*, *5*(2), 1463–1467.

Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA, USA: MIT Press.

Kuang, L. vivian. (2007). *DNIDS: A Dependable Network Intrusion Detection System Using the CSI-KNN Algorithm*.

Kumar, P., & Gupta, N. (2014). OPEN ACCESS A Hybrid Intrusion Detection System Using Genetic-Neural Network. *International Journal of Engineering Research and Applications (IJERA)*, (March), 59–63.

Kumari, N., Sunita, & Smita. (2013). Comparison of ANNs, Fuzzy Logic and NeuroFuzzy Integrated Approach for Diagnosis of Coronary Heart Disease: A Survey. *International Journal of Computer Science and Mobile Computing*, *2*(6), 216–224.

Lakhina, A., Crovella, M., & Diot, C. (2005). Mining Anomalies Using Traffic Feature Distributions. In *Proceedings of the 2005 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (pp. 217–228). New York, NY, USA: ACM.

Lee, K.-C., Chang, J., & Chen, M.-S. (2008). PAID: Packet Analysis for Anomaly Intrusion Detection. In T. Washio, E. Suzuki, K. Ting, & A. Inokuchi (Eds.), *Advances in Knowledge Discovery and Data Mining SE - 58* (Vol. 5012, pp. 626–633). Springer Berlin Heidelberg.

Liao, H.-J., Richard Lin, C.-H., Lin, Y.-C., & Tung, K.-Y. (2013). Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications*, *36*(1), 16–24.

Lippmann, R., Haines, J. W., Fried, D. J., Korba, J., & Das, K. (2000). The 1999 DARPA off-line intrusion detection evaluation. *Computer Networks*, *34*(4), 579–595.

Louvieris, P., Clewley, N., & Liu, X. (2013). Effects-based feature identification for network intrusion detection. *Neurocomputing*, *121*(0), 265–273.

Lu, Y. (1996). Knowledge integration in a multiple classifier system. *Applied Intelligence*, *6*(2), 75–86.

Mahoney, M. V. (2003). Network Traffic Anomaly Detection Based on Packet Bytes. In *Proceedings of the 2003 ACM Symposium on Applied Computing* (pp. 346–350). New York, NY, USA: ACM.

Mahoney, M. V, & Chan, P. K. (2001). *PHAD: Packet Header Anomaly Detection for Identifying Hostile Network Traffic*.

Mahoney, M. V, & Chan, P. K. (2002). Learning Nonstationary Models of Normal Network Traffic for Detecting Novel Attacks. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 376–385). New York, NY, USA: ACM.

Mahoney, M. V, & Chan, P. K. (2003). Learning Rules for Anomaly Detection of Hostile Network Traffic. In *Proceedings of the Third IEEE International Conference on Data Mining* (p. 601–). Washington, DC, USA: IEEE Computer Society.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133.

Muda, Z., & Yassin, W. (2011). A K-Means and Naive Bayes learning approach for better intrusion detection. *Information Technology Journal*, *10*(3), 648–655.

Muda, Z., Yassin, W., Sulaiman, M., & Udzir, N. (2014). K-Means Clustering and Naive Bayes Classification for Intrusion Detection. *Journal of IT in Asia*, *4*.

Mukkamala, S., Janoski, G., & Sung, A. (2002). Intrusion detection using neural networks and support vector machines. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)* (pp. 1702–1707). IEEE.

Ouivirach, K., Gharti, S., & Dailey, M. N. (2013). Incremental behavior modeling and suspicious activity detection. *Pattern Recognition*, *46*(3), 671–680.

Panda, M., Abraham, A., & Patra, M. R. (2012). A Hybrid Intelligent Approach for Network Intrusion Detection. *Procedia Engineering*, *30*, 1–9.

Panda, M., & Patra, M. (2007). Network intrusion detection using naive bayes. *International Journal of Computer Science and Network Security*, *7*(12), 258–263.

Patcha, A., & Park, J.-M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks*, *51*(12), 3448–3470.

Patel, R., Thakkar, A., & Ganatra, A. (2012). A Survey and Comparative Analysis of Data Mining Techniques for Network Intrusion Detection Systems. *International Journal of Soft Computing Journal*, *2*(1), 265–271.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106.

Roli, F., Kittler, J., & Windeatt, T. (Eds.). (2004). *Multiple Classifier Systems* (Vol. 3077). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ryszard S. Choras. (2015). *Image Processing & Communications Challenges 6*. (R. S. Choraś, Ed.) (Vol. 313). Cham: Springer International Publishing.

S. Fugate. (2012). *Methods for Speculatively Bootstrapping Better Intrusion Detection System Performance*. University of New Mexico.

S. Juma, Muda, Z., & Yassin, W. (2014). Reducing False Alarm Using Hybrid Intrusion Detection Based On X-Means Clustering and Random Forest Classification. *Journal of Theoretical and Applied Information Technology*, *68*(2), 249–254.

Sagale, A., & Kale, S. (2014). Combining Naive Bayesian and Support Vector Machine for Intrusion Detection System. *International Journal of Computing and Technology*, *1*(3), 61–65.

Sapate, P., & A.Raut, S. (2014). Survey on Classification Techniques for Intrusion Detection. In *Computer Science & Information Technology (CS & IT)* (pp. 223–231). Academy & Industry Research Collaboration Center (AIRCC).

Schear, N., Albrecht, D. R., & Borisov, N. (2008). High-Speed Matching of Vulnerability Signatures. In *Proceedings of the 11th International Symposium on Recent Advances in Intrusion Detection* (pp. 155–174). Berlin, Heidelberg: Springer-Verlag.

Shakouri G., H., & Nadimi, R. (2013). Outlier detection in fuzzy linear regression with crisp input–output by linguistic variable view. *Applied Soft Computing*, *13*(1), 734–742.

Shamsuddin, S. B., & Woodward, M. E. (2008). Applying Knowledge Discovery in Database Techniques in Modeling Packet Header Anomaly Intrusion Detection Systems. *JSW*, *3*(9), 68–76.

Shamsuddin, S., & Woodward, M. (2007). Modeling protocol based packet header anomaly detector for network and host intrusion detection systems. *Cryptology and Network Security*, 209–227.

Shamsuddin, S., & Woodward, M. (2008). Applying Knowledge Discovery in Database Techniques in Modeling Packet Header Anomaly Intrusion Detection Systems. *Journal of Software (…*, *3*(9), 68–76.

Shiravi, A., Shiravi, H., Tavallaee, M., & Ghorbani, A. A. (2012). Toward developing a systematic approach to generate benchmark datasets for intrusion detection. *Computers & Security*, *31*(3), 357–374.

Sravani, K., & Srinivasu, P. (2014). Comparative Study of Machine Learning Algorithm for Intrusion Detection System. In S. C. Satapathy, S. K. Udgata, & B. N. Biswal (Eds.), *Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2013* (Vol. 247, pp. 189–196). Cham: Springer International Publishing.

Suen, C., & Lam, L. (2000). Multiple Classifier Combination Methodologies for Different Output Levels. In *Multiple Classifier Systems SE - 5* (Vol. 1857, pp. 52–66). Springer Berlin Heidelberg.

Sujatha, M., Prabhakar, S., & Devi, G. (2013). A Survey of Classification Techniques in Data Mining. *Ijiet.com*, *2*(4), 86–92.

Sulaimam, S., & Anitha, P. (2013). An Efficient Classification Mechanism for Network Intrusion Detection System based on Data Mining Techniques: A Survey. *International Journal of Computer Science and Business Informatics*, *6*(1), 1–12.

Taylor, C., & Alves-Foss, J. (2001). NATE: Network Analysis of Anomalous Traffic Events, a Low-cost Approach. In *Proceedings of the 2001 Workshop on New Security Paradigms* (pp. 89–96). New York, NY, USA: ACM.

Thaseen, S., & Kumar, C. A. (2013). An analysis of supervised tree based classifiers for intrusion detection system. In *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering* (pp. 294–299). IEEE.

Tribak, H., Delgado-Marquez, B. L., Rojas, P., Valenzuela, O., Pomares, H., & Rojas, I. (2012). Statistical analysis of different artificial intelligent techniques applied to Intrusion Detection System. In *2012 International Conference on Multimedia Computing and Systems* (pp. 434–440). IEEE.

Urtubia, A., Pérez-Correa, J. R., Soto, A., & Pszczólkowski, P. (2007). Using data mining techniques to predict industrial wine problem fermentations. *Food Control*, *18*(12), 1512–1517.

Vida, R., Galeano, J., & Cuenda, S. (2014). Vulnerability of state-interdependent networks under malware spreading. *Physica A: Statistical Mechanics and Its Applications*.

Waizumi, Y., Sato, Y., & Nemoto, Y. (2012). A Network-Based Anomaly Detection System Based on Three Different Network Traffic Characteristics. *Journal of Communication & Computer*, *9*(7), 805.

Wang, K. (2007). *Network Payload-based Anomaly Detection and Content-based Alert Correlation*. Columbia University, New York, NY, USA.

Wang, Y. (2004). *A hybrid intrusion detection system*. Iowa State University.

Woźniak, M., Graña, M., & Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion*, *16*(0), 3–17.

Wu, S. X., & Banzhaf, W. (2010). The use of computational intelligence in intrusion detection systems: A review. *Applied Soft Computing*, *10*(1), 1–35.

Xie, Y., Tang, S., Huang, X., Tang, C., & Liu, X. (2013). Detecting Latent Attack Behavior from Aggregated Web Traffic. *Comput. Commun.*, *36*(8), 895–907.

Xiong, W., Xiong, N., Yang, L. T., Park, J. H., Hu, H., & Wang, Q. (2013). An Anomaly-based Detection in Ubiquitous Network Using the Equilibrium State of the Catastrophe Theory. *J. Supercomput.*, *64*(2), 274–294.

Yassin, W., Udzir, N., Abdullah, A., Abdullah, M., Muda, Z., & Zulzalil, H. (2014). Packet Header Anomaly Detection Using Statistical Analysis. In J. G. de la Puerta, I. G. Ferreira, P. G. Bringas, F. Klett, A. Abraham, A. C. P. L. F. de

Carvalho, … E. Corchado (Eds.), *International Joint Conference SOCO'14-CISIS'14-ICEUTE'14 SE - 47* (Vol. 299, pp. 473–482). Springer International Publishing.

Yassin, W., Udzir, N. I., & Muda, Z. (2013). Anomaly-based Intrusion Detection Through K- Means Clustering and Naive Bayes Classification. In *Proceedings of the 4th International Conference on Computing and Informatics, ICOCI 2013* (pp. 298–303). Universiti Utara Malaysia.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.

Zhang, Y., Lee, W., & Huang, Y.-A. (2003). Intrusion Detection Techniques for Mobile Wireless Networks. *Wirel. Netw.*, *9*(5), 545–556.

Zhang, Z. (2004). *Statistical anomaly denial of service and reconnaissance intrusion detection*. New Jersey Institute of Technology Newark, NJ, USA.

Zhou, M. (2005). *Network Intrusion Detection: Monitoring, Simulation and Visualization*. University of Central Florida Orlando, Florida.

Zingg, D. W., Nemec, M., & Pulliam, T. H. (2008). A comparative evaluation of genetic and gradient-based algorithms applied to aerodynamic optimization. *Revue Européenne de Mécanique Numérique*, *17*(1-2), 103–126.