ENSEMBLE METHODS IN INTRUSION DETECTION

KEKERE TEMITOPE JOSIAH

A dissertation submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

JANUARY 2015

This dissertation is dedicated to my family for their endless support and encouragement.

ACKNOWLEDGEMENT

First, I want to thank God for the gift of life without which I wouldn't be able to complete my study let alone this thesis. I want to thank my parents, Pastor & Mrs Kekere for their unending love, sacrificial support especially for sponsoring me to Malaysia to acquire advanced expertise. I sincerely thank my supervisor, **Dr Shukor Razak**, for his mentorship, guidance, support and confidence in me throughout my research in the faculty of computing, UTM.

I want to appreciate the efforts of my brothers and sister; Dr Victor Kekere, Engr Daniel Kekere and Mrs Ruth Umahi. Worthy of mention is the effort of Dr Clement Folorunsho for his response squad during my health crisis. I am not ungrateful to Dr Philip Achimugu for his enlightenment, time and support during my study in UTM. Special thanks to beloved people in The Redeemed Christian Church of God, Jalan Mutiara Emas, Johor Bahru, (a.k.a. PowerPalace) Pst Ruthie Gunasundari A/P, Pst Alabi Michael, Pst Olowoyo Austin, Pst Goke Oladokun, Pst Chidi Okpechi, Pst Charles Uti, Oluwagbemi Oluwatolani (Mummy Success), Augusta Nduka, Lilian Oladokun (Mummy Church), Bayo-Philip Patrick, Ayo Sanwoolu, Esther Tazamu, Bosede Edwards, Kemi Adedokun, Isaac Katuka, Kavitha, Carol, Previn, Daniel, Stanley, Paul, Pa and Mummy Esther. I also want to thank Tobi Akanbi, Marwa Alhazmi, Aminu Muhammed, Salisu Borodo, Nansukusa Hidaya Ngaya, Rabiu Idris, my friends in the faculty of computing, UTM. I also acknowledge the following Dr Olukayode Obasan, Dr Femi Folorunsho Ayinde, Dr Adebanjo Adekiigbe, Adeyemi Ajao.

ABSTRACT

As services are being deployed on the internet, there is the need to secure the infrastructure from malicious attacks. Intrusion detection serves as a second line of defense apart from firewall and cryptography. There are many techniques employed in intrusion detection which include signature detection, anomaly and specification based detection system. These techniques often trade off accuracy with false positive rate. In this study, anomaly detection using ensembles is used to automatically classify and detect attack patterns. It has been proven that ensembles of classifier outperform their base classifiers. Several multiples of classifiers have been combined to improve the performance of intrusion detection system. Commonly used classifiers include Support Vector Machines, Decision Trees, Genetic Algorithms, Fuzzy, Principal Component Analysis. The study employed KStar clustering and Instance Based classification algorithms to detect intrusions in NSL-KDD dataset. The results show that the ensemble we designed has a *1-error rate* of 99.67% and false positive 0.33%. The response time of the anomaly is 0.18seconds. The chosen ensemble outperformed the rest of the ensembles (rPART & SMO and J48) and the base classifiers. The performance of the combiners has showed that the study has built a model with high detection, and reduced error.

ABSTRAK

Sebagai perkhidmatan sedang diperluaskan di internet, terdapat keperluan untuk menjamin infrastruktur daripada serangan jahat. Pengesanan pencerobohan berfungsi sebagai pertahanan peringkat kedua selain dari "firewall" dan kriptografi. Terdapat pelbagai teknik yang digunakan dalam pengesanan pencerobohan iaitu pengesanan tandatangan, anomali dan spesifikasi berasaskan sistem pengesanan. Teknik tersebut mempertimbangkan ketepatan berdasarkan kadar kesalahan positif. Dalam kajian ini, pengesanan anomali berasaskan pengumpulan digunakan untuk mengkelaskan dan mengesan corak serangan secara automatik. Ia terbukti dapat mengumpul pengelasan yang melebihi pengelasannya. Beberapa pengelas digabungkan untuk meningkatkan prestasi sistem pengesanan pencerobohan. Pengelas yang selalu digunakan adalah Sokongan Mesin Vektor, Pokok Keputusan, Algoritma Genetik, Kabur, Analisis Komponen Utama. Kajian ini menggunakan pergkelasan KStar algoritma pengkelasan segera untuk mengeson pencerobohan dalam set data NSL-KDD. Kajian menunjukkan bahawa pengumpulan yang dibangunkan mempunyai kadar 1-kesilapan sebanyak 99.67% dan kesalahan positif 0.33%. Masa tindak balas daripada anomali adalah 0.18saat. Pengumpul yang dipilih telah mengatasi (rPART & SMO dan J48) dan Pengelas asas. Prestasi daripada penggambungan ini telah menunjukkan bahawa kajian telah membina sebuah model dengan pengesanan tinggi, dan kesilapan dikurangkan.

TABLE OF CONTENTS

2 LITERATURE REVIEW

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS		MEANING
DBIR	-	Data Breaches Investigation Report
DBMS	-	Database Management System
ID3	-	Iterative Dichotomiser
J48	-	Decision Trees
KNN	-	K-Nearest neighbor
OWASP	-	Open Web Application Security Project
rPART	-	Decision Trees
SMO	-	Function
SQL	-	Structured Query Language
WEKA	-	Waikato Environment for Knowledge Analysis

CHAPTER 1

INTRODUCTION

1.1 Introduction

Conventional or cybercrime is always one step ahead of security. Electronic crimes include phishing, email spoofing, denial of service, pornography, structured query language injection, data diddling. With services and application deployed over an infrastructural technology called the internet, crime or cybercrime may never stop. This is why cybercrime experts are gathering information about these attacks in order to find techniques to curb or reduce these attacks. In this research, the focus is on intrusion detection as well as the techniques that have been developed to detect attacks.

Computers have grown from desktop computers, super computers, to tablets, smart devices and high performance computers providing extreme computation and telephone services together with super processing abilities. These devices are often getting miniaturized and networked or connected either through a local area network, wireless local area network or a wide area network. Services are deployed over the internet to hundreds of billions of interconnected devices. Other platforms include Point of Sales (POS), Automated Teller Machines (ATM). These platforms provide a channel of communication suitable for electronic commerce, online retailing, online advertisement, email services, social networking, chatting, online banking services, massive open online courses (MOOC) (e.g. Coursera, EdX), teleconferencing, webinars, online radio, music and video streaming.

The sheer volume of data generated over the internet from click streams, crawled pages, social networking sites, internet of things, sensor networks, cloud, mobile apps, geo location sensors, is becoming huge and has provided many challenges for data management for big data thast cannot be handled by classic database software. 72 hours of video are uploaded to Youtube by the minute, Twitter generates 500million tweets per day, Facebook has more than 1.15 billion active users, Wal-Mart receives 10 million cash registers transactions in 2012, United Parcel Service receives on average 39.5million tracking requests from customers per day. Computer databases are now growing at an explosive rate that government and business organizations are now applying data crunching tools to make inference and useful analysis from the data. According to IDC report in 2011, data volume created and copied is 1.8ZB and it will increase nine times in every five years.

Cyber-attacks is traceable to the mischievous act that happened in 1903 when Nevil Maskelyne, an inventor and magician cracked the wireless telegraphy of Guglielmo Marconi as John Ambrose Fleming, a physicist was about to demonstrate that confidential messages could be sent through the radio system wirelessly. Nevil sent Morse code messages through the projector being used for demonstration.

In June 1982, the Central Intelligence Agency (CIA) subverted an industrial software controlling Trans-Siberian pipeline, causing the pipeline to explode because the Soviet Union was planning to steal the software from the Canadian developers. A cracker in Germany broke into the computer at the Lawrence Berkeley National Laboratory, a U.S. Department of Energy facility, and other military computers in the U.S were traced by a physicist, Clifford Stoll, in August 1986.

Morris worm infected 60,000 computers on Wednesday November 2, 1988 across a 400 connected local area network. The worm which reportedly did not cause any software or hardware damage was developed by Robert Morris who exploited the weakness of Berkeley UNIX version; it however slowed down internet usage across major computer centers like NASA Ames Laboratory, Lawrence Livermore National Laboratory, SRI, MIT, University of California at both Berkeley and San Diego campuses, University of Maryland, Purdue and the Rand Corporation (Branscomb, 1989).

On March 2, 1988, Richard Brandow infected thousands of Macintosh computers in the US and Canada with the Aldus Peace Virus by transferring an embedded game to a commercial software which contained the virus. Like the Morris worm, it did not cause any damage. It closes after displaying the message below:

“Richard Brandow, the publisher of MacMag, and its entire staff would like to take this opportunity to convey their universal message of peace to all Macintosh users around the world” (Branscomb, 1989, 1990; Spafford, Heaphy, & Ferbrache, 1989).

Melissa virus is a macro virus that began with an attachment to an email note with the subject line “Important Message from [the name of someone],” and the body text reads “Here is the document you asked for …don’t show anyone else;-)”. The document is foten named list.doc. This virus spread quickly through distributed email attachments disabling a number of safeguards in Microsoft Word ’97 or 2000 and sends mails to 50 contacts if Microsoft Outlook is present. Melissa disabled a large number of corporate and other mail servers (Chen, 2004). Melissa is not just a virus and worm but also a Trojan (Berghel, 2001).

Red code worm in July 19, 2001 infected 250,000 systems in nine hours by finding vulnerable systems and installing itself on to it. The malicious code was deployed from a university in China and carried out an “index-server flaw” a vulnerability in Microsoft Internet Information Services deployed on Windows 2000, NT and beta version of Windows XP servers. ISAPI is an indexing tool that assigns data files to executables automatically but does not check for buffer overflow which red code exploited. Other variants are Code Red v1 and Code Red v2 (Berghel, 2001;

Moore & Shannon, 2002; Naik, Ajgaonkar, Nadarge, & Agawane, 2014; Zou, Gong, & Towsley, 2002).

Blaster worm infected about 100,000 Microsoft XP, 2000 and NT4 systems on Wednesday July 16, 2003. In August 11, a variant of the worm called Lovsan also struck. The worm copied directly from the dcom.c exploit, added its own code, and launched a coordinated denial of service (DoS) attack to exhaust Windowsupdate.com resources using a transmission control protocol port 80 SYN flood (Bailey, Cooke, Jahanian, & Watson, 2005). Welchia or Nachi and SDBot, variants of the Blaster worm also appeared on the scene. Though the author of the worm was never caught, the authors of the variants have been apprehended (Bailey et al., 2005; Chen, 2004).

Amjad Farooq Alvi and Basit Farooq from Pakistan infected more than 100,000 IBM PC disks of university students and journalist in 1988. Froma Joselow, a reporter could not print her work on receiving a blank screen with the message from the two Pakistani brothers displayed on her computer monitor. A Phd thesis was also destroyed by the Pakistani Brain Virus. Like the Aldus Peace Virus it was embedded with commercially distributed software but was targeted at boot up disks (Branscomb, 1989, 1990; Highland, 1988; Schmidt & Arnett, 2005).

Donald Gene Burleson in an attempt to revenge after being sacked from brokerage and insurance firm in Forth Worth, Texas wiped out the sales records of the company until the MIS staff came to the rescue by rebuilding the system from scratch and reinstalling a new operating system from IBM (Branscomb, 1989, 1990; Tavani, 1999).

The electronic conglomerate, Sony PlayStation lost names, addresses and about 77 million credit card details to cyber-attacks on 17 and 19 April 2011. The Japanese company did not tell the public about the attack until Tuesday 26 April, 2011 that obtained people’s names, email addresses, birth dates, usernames, passwords, logins, security questions. Allen Paller, a research director of the SANS Institute noted that the attack is the largest internet ever security break-in.

Robert Philip Hanssen stole and sold US classified documents to the Soviet Union from 1979 to 2001 using cyber espionage. He was sentenced to life imprisonment (Programs, 2002; Vise, 2002). Between 2007 and 2009, 71 governmental bodies and US military has been hacked several times. The Department of Defence (DOD) admits that some 24,000 files were lost due to cyber espionage. In 2011, Cyworld subscribers, a social networking site in South Korea, were divulged to the public in an attack. The attackers also hit government organizations and 1.8 million customer data was stolen from Hyundai Capital. In 2012, two crackers were arrested for having access to 8.7 subscribers of KT Mobile. Hanjuan Jin was in possession of 1,000 documents of Motorola, a telecoms company where she worked formerly. She was sentenced to four years in prison.

Estonian government experienced a denial of service (DOS) attack in 2007 by unknown attackers which disrupt government and banking services. The database of both parties of presidential campaigns were hacked by anonymous attackers. In 2008, the webpage of Georgian government was defaced by intruders and “Graffitti” appeared on their webpage. China Aerospace Science & Industry Corporation (CASIC) found spywares on the computers.

Also, Conficker worm targeted at Microsoft operating system (OS) in November, 2008 exploited the flaw the vendor OS and added dictionary attacks in cracking administrator passwords to form botnets. It is the largest known computer worm (Dittmann, Karpuschewski, Fruth, Petzel, & Munder, 2010).

Israeli government in 2009 experienced the crackers activities with over 5,000,000 computers affected. Baidu, a popular Chinese search engine and Twitter, an online social networking service was disrupted by Iranian cyber army in 2010. Stuxnet, is a complex malware targeted at Siemens industrial plant in Indonesia, Iran in 2010 (Falliere, Murchu, & Chien, 2011; Farwell & Rohozinski, 2011; Langner, 2011).

Disconnecting from the internet may be a safe way to avoid attack like the Finance department and Treasury board of the Canadian government did in 2010. In September 2011, duqu worm, a reconnaissance attack collected digital certificated from infected systems (Bencsáth, Pék, Buttyán, & Félegyházi, 2012; Chien, OMurchu, & Falliere, 2012; Jain & Sardana, 2012).

24,000 files were stolen from a defence contractor in the US in July 2011. Kaspersky discovered “Red October”, a virus that collects information from government agencies, research firm, military installations, energy providers and other critical infrastructures by exploiting vulnerabilities in Microsoft Word and Excel.

In 2013, Russian crackers had access to 54 million citizens ID data. British Broadcasting Corporation (BBC) server was also cracked on Christmas. Chinese also targeted the Federal Election Commisiion in the US in December 2013. In 2014, dropbox was hacked.

A contractor stole names, credit card details, and social security number of half the population of South Korea in January 2014 by copying it on a flash drive and sold it to marketing firms. In October 2014, a gang of cyber criminals from Latin America was able to crack seventeen (17) Automated Teller Machine (ATM) and stole $1.2 million belonging to United Overseas Bank, Affin Bank, Al Rajhi Bank and Bank of Islam. The closed circuit television (CCTV) footage from the banks showed that 2-3 Latin American men entered and withdrew money from these targeted ATM. A cybercrime expert reported that a RM100 chip, specific technical knowledge, and a free malware on the internet was what was required to crack the ATM. It is also reported that the attack would not have been successful without insider information.

In the light of these attacks, it is germane that security be incorporated into the computer networks. When TCP/IP model was built, the developers did not have security in mind in terms of confidentiality, integrity and availability of data.

One of the defenses against these attacks apart from cryptography and firewall is intrusion detection system. Intrusion detection is the process of monitoring computer activities for security violation in terms of keeping the confidentiality of data private, making sure the data is unaltered and kept available for use whenever.

Different intrusion detection exist depending on their use. There is host based intrusion detection system that monitors for security violation on host systems. This is achieved by installing the application software on the host computer or device and it flags for intrusion whenever there is any. The second type of intrusion detection is the network intrusion detection system placed inline of the network. It monitors network packets that are mischievous.

One of the approaches to intrusion detection systems is signature detection. It checks for intrusion by searching the database for recognizable patterns of attacks. If similar attack pattern is found, it flags for intrusion. This flagging is reported to the Security Analyst that cross-checks if an attack occurred. Therefore, signature detection is a database of attack patterns stored over time which the detection engine uses to match attack signatures. The strength of signature detection or misuse detection system is that it captures all attack pattern that are previously stored in the system. However, same attack patterns can be easily altered by attackers and missed by misuse detection system.

Another approach to intrusion detection is anomaly based detection. Anomaly detection looks out for violation of security in a system by first profiling normal usage of the system and any pattern that deviates from this norm is flagged as intrusion. Anomaly system uses statistical techniques to profile or model the normal usage of the system and builds a model with it over time. These statistical techniques can detect variation from the model and reports deviation as intrusion. Anomaly detection can detect novel attacks. It has been proven to detect zero-day attacks.

Some scholars have combined anomaly detection with misuse detection system to combine the features of both approaches. This is because misuse detection come down with false positives and false negatives. Anomaly detection comes down with false positives. By combining anomaly detection and signature detection systems, the false positives are notably reduced and false negatives are eliminated.

The methods employed in intrusion detection could vary from single, to hybrid to ensemble. Single methods refers to the use of a classifier or a technique used in the detection engine of either a host based, network based, misuse or anomaly based detection system. Hybrid methods refers to the combination of two techniques or classifiers in the detection engine. Ensemble methods refers to three or more classifiers used to detect intrusion in an intrusion detection system. It is proven that ensemble methods have yielded better result in lowering false alarm and high accuracy (Chebrolu, Abraham, & Thomas, 2005; Gogoi, Bhattacharyya, Borah, & Kalita, 2014; Mukkamala, Sung, & Abraham, 2005; Reddy, Ramadevi, & Sunitha, 2014).

Misuse and anomaly detection systems are measured in terms of accuracy and false alarm. Other measures of performance include response time, f1 measure, precision and recall. Accuracy is the measure of correctness of the detection model. It is the proportion of true results in a population. Accuracy can also be measured in terms of how efficient the system is.

Having explored intrusion detection, the study now takes a cursory look at the dataset available for intrusion detection. They include KDD dataset, the first public dataset available to cybercrime expert from MIT laboratory (1998,1999, and 2002 versions), NSL-KDD dataset, which is an improved version of KDD data, DEFCON 9 capture the flag (CTF) dataset, UNM audit dataset, McPAD dataset, ADFA intrusion detection dataset (Linux and Windows), CSIC 2010 HTTP dataset, ITOC 2009 dataset, ECML-PKDD 2007 HTTP dataset (recommender system), Industrial System Control (ISC) Attack dataset (SCADA), Botnet Malware, IDS Bag dataset, Netflow intrusion detection dataset, Tezpur University intrusion detection system (TUIDS), Acer 2007 dataset, Kyoto University benchmark dataset, Greenberg dataset, Ozone dataset, Windows-Users and Intruder-simulation Log (WUIL) dataset, Schonlau et. al. (SEA) dataset (masquerading user data) and ISCX 2012 dataset. In this research, NSL-KDD dataset has been selected for use.

1.2 Problem Background

In a rapidly growing world of ours, we are faced with overwhelmingly large volumes of data which contains patterns that can be mined or extracted to find interesting details. Because these data is big in the very sense of the word, data analysts need tools and techniques capable of mining features relevant to the field of study. In this research, the interest is in patterns of attacks in NSL-KDD dataset as well as the various classifiers that have been used to identify features or attributes that can be used to trace attacks in intrusion detection system.

Extensive research exists in anomalous detection using machine learning techniques. Some have used classifiers such as KNN, Random Forest, J48, Decision Table, Bayes Networks, and SMO to improve the accuracy of anomalous detection. Most of these researchers trained their algorithm on the publicly available NSL-KDD dataset suitable for anomalous detection. Some evaluated using accuracy while others combined accuracy with precision, recall, F1 score all of which are standard benchmark for evaluation.

KNN algorithm is a method for classifying patterns in data that have similarity to others usually known as neighbors. It uses a value of k to determine its neighbors. K can be 1, 2… 5. It is a parametric algorithm that does not assume the distribution of background data and is useful for anomalous detection when the boundary is irregular. It assumes that data exist in a feature space and uses distance to find other patterns that are similar to each other. The works of Naoum and Al-Sultani (2012) combined linear vector quantization and kNN to improve the accuracy of detection of anomalous events with 89% and 0.09s learning rate.

Bayes Net or Bayesian Network is another classifier used to find interesting features in data. While KNN uses neighbors for attack detection, Bayes Net uses node of similar patterns or variables. It is a probabilistic classifier that is useful for full representation of any dataset of any complexity. It provides a graphical representation of nodes that are mutually independent and allows system analyst to view intermediate variables that can be used for detection. Unlike KNN that does not use parameter Bayes Net does. Bayes Net is capable of showing the sequence of events with its directed graph a characteristic that differs it from Markov's. It learns the structure or domain of the data, as well as its parameter. Kumaravel and Niraisha (2013) made an attempt to reduce false alarm rate by administering an ensemble of Bayes Net, Naive Bayes, rule Jtrip, Decision Stump classifiers and achieved an incredible accuracy of 99.54% and false alarm of 0.46%. In their work, rule Jtrip was the best of the classifiers with 99.98% accuracy and 0.02% false alarm.

Decisions are made every day and researchers are motivated to make formal decisions from a body of knowledge sorting the important from the irrelevant. Decision Table is a hierarchical and tabular representation of inference process in modeling a knowledge system. It is useful for data acquisition, verification and validation processes. It offers a legible way of representing complex knowledge systems to comprehend and solve the problem at hand. It is similar to Bayes Net because it uses *premises and conclusion* as nodes. Apart from the fact that the *decisions* in the decision table are represented in a table, it is a tree-like representation that connects the premises and concludes with the use of branches. This method is complete, correct and consistent because input data can be verified and analysis follows logical rules. Experimental implementation of ten (10) classifier was done (Sengupta & Sil, 2011) to improve the accuracy of machine learning on the

KDD dataset, the first dataset used for intrusion detection. Accuracy of Decision Table provided by the authors was 95.3% with false alarm of 4.7%. Other classifiers examined include Cognitive Rule, OneR, PART, JRip, NNge, Zero, Bayes Net, Ridor and Rough Set Theory (RST). RST, a similar technique to Decision Table was the best of the machine learning with an accuracy of 98.5% with 1.5% false alarm. Future work includes increasing the learning rate of RST and its optimization. This shows that for either Decision Table or RST, optimization is necessary.

J48 is an improvement over Iterative Dichotomiser (ID3) invented by Quinlan in 1986 (Quinlan, 1986) for generating decision tree in a dataset. ID3 is an iterative classifier that generates a simple decision tree from all possible decision trees after an accurate classification of the *attributes* in a dataset containing *objects*. J48 or C4.5 is a supervised method that uses the same concept of information entropy like ID3. C4.5 can handle continuous and discrete attributes simultaneously, classify training data with missing values, works on attribute values of varying costs, prunes trees after they are being generated. Evaluation of about eight (8) algorithms was done by Thaseen and Kumar (2013) to classify NSL-KDD data for intrusion detection. The algorithms include Random Tree with 99.74% accuracy, NB Tree with 99.62%, J48 with 99.57%, C4.5 with 99.55%, RepTree with 99.54%, Random Forest with 99.5%, AD Tree with 98.13%, and LAD Tree with 97.7%. Error rates and learning rates were also reported.

Random Forest are an ensemble of learning classifiers introduced by (Ho, 1995) for automatic variable selection that handles big data or *predictors* in no time. In other words, it is used to predict data when the response is not known from a subset predictable with known response. Though early development is traceable to the scholarly works of Amit and Geman (1997), random forest is one of the best classifier that ranks its estimates in a natural way without a need for tuning or pruning like C4.5. It has been widely deployed on bioinformatics data and biomarker data as well as UCI data. The works of (Eid, Azar, and Hassanien (2013)) showed that discretization increases the speed of Random Forest (99.1%; 2.87s) amongst other classifiers using F-measure metrics. Other classifiers evaluated includes Rep

Tree (98.1%; 3.75s), C4.5 (99.0%; 3.05s), Decision Table (96.3%; 132.0s), and Naïve Bayes (93.6%; 0.21s).

1.3 Problem Statement

From network-based to host-based IDS, misuse to anomaly, detection methods have proven to be relatively accurate except for false alarms. Existing anomaly detection for intrusion detection using machine learning have identified the need to reduce false positives (Medhane 2013; Scarfone & Mell 2010; Sun & Beznosov 2010; Gander *et al* 2013; Choras *et al* 2013; Valeur *et al* 2005) and overhead generated (Khalkhali, Iman, *et al* 2011; Kemalis 2008; Keromytis 2009; Huihui & Tonnge 2013; Chuan-Xiang 2009).

Existing machine learning algorithms have showed a high degree of accuracy in detecting intrusions with reduced false positive. It is therefore necessary to explore an ensemble of these algorithms to see which combination give higher accuracy and reduced runtime overhead while addressing the following:

1. How to process raw dataset for intrusion detection?
2. How to increase the accuracy of learning algorithms used for intrusion detection?
3. How to reduce the false positives in anomaly detection?
4. What are the best aggregates of classifiers that provide higher accuracy and reduced runtime?

1.4 Purpose of Study

This study explores the performance of these learners in terms of accuracy and speed. At the end a comparative analysis shall be examined to see which composition of classifier performed better. The study shall evaluate a set of classifiers that has high detection rate, response time and low false positive rate.

1.5 Objectives of the Study

Four objectives examined in this research include:

1. To carry out dataset processing, segmentation, feature extraction, and evaluate classifiers accuracy on varying proportions of the dataset.
2. To investigate potential ensemble and select the best ensemble classifier.
3. To do a comparative analysis of the ensemble against each classifier.
4. To increase detection rate and reduce false positive rate of anomaly system.

1.6 Scope of the Study

The scope of the research borders on the following:

1. NSL-KDD is one of the standard dataset used for anomaly detection (http://nsl.cs.unb.ca/NSL-KDD/)
2. 75% of the dataset is used to train and 25% is used to test the algorithms; Instance based learning (IBk), K-Nearest Neighbor (KNN), Decision Tree (C4.5), Sequential Minimal Optimization (SMO), Rules set (ID3).
3. The benchmark used for evaluation includes accuracy, speed, precision, recall, and f1-score.
4. Attacks considered in this research is limited to Probe, Remote to Local (R2L), User to Root (U2R), and Denial of Service
5. Simulation of this research is done using WEKA.

1.7 Significance of the Study

With increasing attacks targeted at the internet, it is necessary to curb both insider and malicious attacks from violating the security policies of the network. Machine learning is a branch of artificial intelligence that has been used by researchers to classify normal queries from anomalous queries. These classifiers can be combined and evaluated based on the accuracy and speed that they provide to investigate which individual or ensemble classifier performs best.

1.8 Organization of Thesis

This dissertation consist of six (6) chapters. Chapter one introduces the study of intrusion detection, research objectives and questions, scope of the study and its primary objectives. The second chapter is a survey and summary on existing techniques in intrusion detection. Chapter three is a description of the methodology that this research employed. Dataset collection and division, preprocessing, feature extraction is the focus of chapter four. In chapter five, simulation, result and analysis is been discussed and chapter six concludes the dissertation report with a summary of the research objectives, contribution and future work.

REFERENCES

Ahmad, I., Abdullah, A. B., & Alghamdi, A. S. (2009, 4-6 Oct. 2009). *Application of artificial neural network in detection of probing attacks.* Paper presented at the Industrial Electronics & Applications, 2009. ISIEA 2009. IEEE Symposium on.

Ahmad, I., Abdullah, A. B., Alghamdi, A. S., Baykara, N., & Mastorakis, N. (2009). *Artificial neural network approaches to intrusion detection: a review.* Paper presented at the WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering.

Al-Jarrah, O., & Arafat, A. (2014). *Network Intrusion Detection System using attack behavior classification.* Paper presented at the Information and Communication Systems (ICICS), 2014 5th International Conference on.

Amaral, J. P., Oliveira, L. M., Rodrigues, J. J., Han, G., & Shu, L. (2014). *Policy and network-based intrusion detection system for IPv6-enabled wireless sensor networks.* Paper presented at the Communications (ICC), 2014 IEEE International Conference on.

Amit, Y., & Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural computation, 9*(7), 1545-1588.

Anderson, D., Frivold, T., & Valdes, A. (1995). *Next-generation intrusion detection expert system (NIDES): A summary*: SRI International, Computer Science Laboratory.

Anderson, J. P. (1980). Computer security threat monitoring and surveillance: Technical report, James P. Anderson Company, Fort Washington, Pennsylvania.

Ashoor, A. S., & Gore, S. (2011). *Intrusion Detection System (IDS): Case Study.* Paper presented at the Proceedings of 2011 International Conference on Advanced Materials Engineering (ICAME 2011).

Bailey, M., Cooke, E., Jahanian, F., & Watson, D. (2005). The blaster worm: Then and now. *Security & Privacy, IEEE, 3*(4), 26-31.

Baraka, H. B., & Tianfield, H. (2014). *Intrusion Detection System for Cloud Environment.* Paper presented at the Proceedings of the 7th International Conference on Security of Information and Networks.

Barot, V., & Toshniwal, D. (2012). *A new data mining based hybrid network Intrusion Detection model.* Paper presented at the Data Science & Engineering (ICDSE), 2012 International Conference on.

Beigh, B. M. (2014, 5-7 March 2014). *One-stop: A novel hybrid model for intrusion detection system.* Paper presented at the Computing for Sustainable Global Development (INDIACom), 2014 International Conference on.

Bencsáth, B., Pék, G., Buttyán, L., & Félegyházi, M. (2012). *Duqu: Analysis, detection, and lessons learned.* Paper presented at the ACM European Workshop on System Security (EuroSec).

Berger, M., Erlacher, F., Sommer, C., & Dressler, F. (2014). *Adaptive load allocation for combining Anomaly Detectors using controlled skips.* Paper presented at the Computing, Networking and Communications (ICNC), 2014 International Conference on.

Berghel, H. (2001). The code red worm. *Communications of the ACM, 44*(12), 15-19.

Besson, L., & Leleu, P. (2009). *A Distributed Intrusion Detection System for Ad-Hoc Wireless Sensor Networks: The AWISSENET Distributed Intrusion Detection System.* Paper presented at the Systems, Signals and Image Processing, 2009. IWSSIP 2009. 16th International Conference on.

Bi, Y. (2012). The impact of diversity on the accuracy of evidential classifier ensembles. *International Journal of Approximate Reasoning, 53*(4), 584-607.

Branscomb, A. W. (1989). *Rogue Computer Programs--Viruses, Worms, Trojan Horses, and Time Bombs: Prank, Prowess, Protection Or Prosecution?* : Program on Information Resources Policy, Harvard University, Center for Information Policy Research.

Branscomb, A. W. (1990). Rogue computer programs and computer rogues: Tailoring the punishment to fit the crime. *Rutgers Computer & Tech. LJ, 16*, 1.

Breiman, L. (1996). Bias, variance, and arcing classifiers.

Burguera, I., Zurutuza, U., & Nadjm-Tehrani, S. (2011). *Crowdroid: behavior-based malware detection system for android.* Paper presented at the Proceedings of the 1st ACM workshop on Security and privacy in smartphones and mobile devices.

Butun, I., Morgera, S., & Sankar, R. (2014). A Survey of Intrusion Detection Systems in Wireless Sensor Networks.

Camiña, J. B., Rodríguez, J., & Monroy, R. (2014). Towards a Masquerade Detection System Based on User's Tasks *Research in Attacks, Intrusions and Defenses* (pp. 447-465): Springer.

Chebrolu, S., Abraham, A., & Thomas, J. P. (2005). Feature deduction and ensemble design of intrusion detection systems. *Computers & Security, 24*(4), 295-307.

Chen, Q., & Aickelin, U. (2008). Dempster-shafer for anomaly detection. *arXiv preprint arXiv:0803.1568*.

Chen, T. M. (2004). Intrusion detection for viruses and worms. *IEC Annual Review of Communications, 57*.

Cheung, S., Dutertre, B., Fong, M., Lindqvist, U., Skinner, K., & Valdes, A. (2007). *Using model-based intrusion detection for SCADA networks.* Paper presented at the Proceedings of the SCADA Security Scientific Symposium.

Chien, E., OMurchu, L., & Falliere, N. (2012). *W32. Duqu: the precursor to the next stuxnet.* Paper presented at the Proc. of the 5th USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET).

Chouhan, P. K., Hagan, M., McWilliams, G., & Sezer, S. (2014). *Network Based Malware Detection within Virtualised Environments.* Paper presented at the Euro-Par 2014: Parallel Processing Workshops.

Chowdhary, M., Suri, S., & Bhutani, M. (2014). Comparative Study of Intrusion Detection System.

Creech, G., & Hu, J. (2013a). *Generation of a new IDS test dataset: Time to retire the KDD collection.* Paper presented at the Wireless Communications and Networking Conference (WCNC), 2013 IEEE.

Creech, G., & Hu, J. (2013b). A Semantic Approach to Host-based Intrusion Detection Systems Using Contiguous and Discontiguous System Call Patterns.

Creech, G., & Jiankun, H. (2013, 7-10 April 2013). *Generation of a new IDS test dataset: Time to retire the KDD collection.* Paper presented at the Wireless Communications and Networking Conference (WCNC), 2013 IEEE.

de la Hoz, E., Ortiz, A., Ortega, J., & de la Hoz, E. (2013). Network Anomaly Classification by Support Vector Classifiers Ensemble and Non-linear Projection Techniques *Hybrid Artificial Intelligent Systems* (pp. 103-111): Springer.

Denning, D. E. (1987). An intrusion-detection model. *Software Engineering, IEEE Transactions on*(2), 222-232.

Denning, D. E. (1997). *Cyberspace attacks and countermeasures.* Paper presented at the Internet besieged.

Didaci, L., Giacinto, G., & Roli, F. (2002). *Ensemble learning for intrusion detection in computer networks.* Paper presented at the Proceedings of the Workshop on Machine Learning, Methods and Applications, held in the context of the 8th Meeting of the Italian Association of Artificial Intelligence (AI* IA).

Dietterich, T. G. (2000). Ensemble methods in machine learning *Multiple classifier systems* (pp. 1-15): Springer.

Dittmann, J., Karpuschewski, B., Fruth, J., Petzel, M., & Munder, R. (2010). *An exemplary attack scenario: threats to production engineering inspired by the Conficker worm.* Paper presented at the Proceedings of the First International Workshop on Digital Engineering.

Eid, H. F., Azar, A. T., & Hassanien, A. E. (2013). *Improved Real-Time Discretize Network Intrusion Detection System.* Paper presented at the Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012).

Elhag, S., Fernández, A., Bawakid, A., Alshomrani, S., & Herrera, F. (2015). On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on Intrusion Detection Systems. *Expert Systems with Applications, 42*(1), 193-202.

Esmalifalak, M., Liu, L., Nguyen, N., Zheng, R., & Han, Z. (2014). Detecting Stealthy False Data Injection Using Machine Learning in Smart Grid. *Systems Journal, IEEE, PP*(99), 1-9. doi: 10.1109/JSYST.2014.2341597

Falliere, N., Murchu, L. O., & Chien, E. (2011). W32. stuxnet dossier. *White paper, Symantec Corp., Security Response*.

Fan, W., & Stolfo, S. J. (2002). *Ensemble-based Adaptive Intrusion Detection.* Paper presented at the SDM.

Farwell, J. P., & Rohozinski, R. (2011). Stuxnet and the future of cyber war. *Survival, 53*(1), 23-40.

Flior, E., Anaya, T., Moody, C., Beheshti, M., Jianchao, H., & Kowalski, K. (2010, 12-14 April 2010). *A Knowledge-Based System Implementation of Intrusion Detection Rules.* Paper presented at the Information Technology: New Generations (ITNG), 2010 Seventh International Conference on.

Forrest, S., Hofmeyr, S. A., & Somayaji, A. (1997). Computer immunology. *Communications of the ACM, 40*(10), 88-96.

Forrest, S., Hofmeyr, S. A., Somayaji, A., & Longstaff, T. A. (1996). *A sense of self for unix processes.* Paper presented at the Security and Privacy, 1996. Proceedings., 1996 IEEE Symposium on.

Freund, Y., & Schapire, R. E. (1996). *Experiments with a new boosting algorithm.* Paper presented at the ICML.

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

Garg, S., Garg, A., Kandpal, A., Joshi, K., Chauhan, R., & Goudar, R. H. (2013, 26-27 Sept. 2013). *Ontology and specification-based intrusion detection and prevention system.* Paper presented at the Confluence 2013: The Next Generation Information Technology Summit (4th International Conference).

Giacinto, G., Perdisci, R., Del Rio, M., & Roli, F. (2008). Intrusion detection in computer networks by a modular ensemble of one-class classifiers. *Information Fusion, 9*(1), 69-82.

Gogoi, P., Bhattacharyya, D., Borah, B., & Kalita, J. K. (2014). MLH-IDS: A Multi-Level Hybrid Intrusion Detection Method. *The Computer Journal, 57*(4), 602-623.

Hareesh, I., Prasanna, S., Vijayalakshmi, M., & Shalinie, S. M. (2011, 3-5 June 2011). *Anomaly detection system based on analysis of packet header and payload histograms.* Paper presented at the Recent Trends in Information Technology (ICRTIT), 2011 International Conference on.

Highland, H. (1988). The brain virus—fact and fantasy. *Computer Fraud & Security Bulletin, 10*(11), 4-9.

Ho, T. K. (1995). *Random decision forests.* Paper presented at the Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on.

Hofmeyr, S. A., & Forrest, S. (2000). Architecture for an artificial immune system. *Evolutionary computation, 8*(4), 443-473.

Hofmeyr, S. A., Forrest, S., & Somayaji, A. (1998). Intrusion detection using sequences of system calls. *Journal of computer security, 6*(3), 151-180.

Holm, H. (2014). *Signature Based Intrusion Detection for Zero-Day Attacks:(Not) A Closed Chapter?* Paper presented at the System Sciences (HICSS), 2014 47th Hawaii International Conference on.

Huang, C.-T., & Gouda, M. G. (2006). Denial-of-Service Attacks. *Hop Integrity in the Internet*, 25-30.

Hubballi, N., & Suryanarayanan, V. (2014). False alarm minimization techniques in signature-based intrusion detection systems: A survey. *Computer Communications*.

Ilgun, K., Kemmerer, R. A., & Porras, P. A. (1995). State transition analysis: A rule-based intrusion detection approach. *Software Engineering, IEEE Transactions on, 21*(3), 181-199.

Jackson, K. A. (1999). Intrusion detection system (IDS) product survey. *Los Alamos National Laboratory, Los Alamos, NM, LA-UR-99-3883 Ver, 2*, 1-103.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22*(1), 4-37.

Jain, P., & Sardana, A. (2012). *Defending against internet worms using honeyfarm.* Paper presented at the Proceedings of the CUBE International Information Technology Conference.

Javitz, H. S., & Valdes, A. (1991). *The SRI IDES statistical anomaly detector.* Paper presented at the Research in Security and Privacy, 1991. Proceedings., 1991 IEEE Computer Society Symposium on.

Javitz, H. S., Valdes, A., & NRaD, C. (1993). The NIDES statistical component: Description and justification. *Contract, 39*(92-C), 0015.

JianLiang, M., & Yang, Y. (2012). The Research and Contrast of the Hybrid Intrusion Detection. In Y. Wu (Ed.), *Software Engineering and Knowledge*

Engineering: Theory and Practice (Vol. 115, pp. 915-919): Springer Berlin Heidelberg.

Jokar, P., Nicanfar, H., & Leung, V. C. M. (2011, 17-20 Oct. 2011). *Specification-based Intrusion Detection for home area networks in smart grids.* Paper presented at the Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on.

Jou, Y., Gong, F., Sargor, C., Wu, X., Wu, S., Chang, H., & Wang, F.-y. (2000). *Design and implementation of a scalable intrusion detection system for the protection of network infrastructure.* Paper presented at the DARPA Information Survivability Conference and Exposition, 2000. DISCEX'00. Proceedings.

Kang, D.-K., Fuller, D., & Honavar, V. (2005). *Learning classifiers for misuse and anomaly detection using a bag of system calls representation.* Paper presented at the Information Assurance Workshop, 2005. IAW'05. Proceedings from the Sixth Annual IEEE SMC.

Khamphakdee, N., Benjamas, N., & Saiyod, S. (2014, 28-30 May 2014). *Improving Intrusion Detection System based on Snort rules for network probe attack detection.* Paper presented at the Information and Communication Technology (ICoICT), 2014 2nd International Conference on.

Kim, G., Lee, S., & Kim, S. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. *Expert Systems with Applications, 41*(4), 1690-1700.

Ko, C., Ruschitzka, M., & Levitt, K. (1997). *Execution monitoring of security-critical programs in distributed systems: A specification-based approach.* Paper presented at the Security and Privacy, 1997. Proceedings., 1997 IEEE Symposium on.

Kolesnikov, O., & Lee, W. (2005). Advanced polymorphic worms: Evading ids by blending in with normal traffic.

Kumar, G., & Kumar, K. (2012). The use of artificial-intelligence-based ensembles for intrusion detection: a review. *Applied Computational Intelligence and Soft Computing, 2012*, 21.

Kumaravel, A., & Niraisha, M. (2013). *Multi-classification approach for detecting network attacks.* Paper presented at the Information & Communication Technologies (ICT), 2013 IEEE Conference on.

Langner, R. (2011). Stuxnet: Dissecting a cyberwarfare weapon. *Security & Privacy, IEEE, 9*(3), 49-51.

Lazarevic, A., Kumar, V., & Srivastava, J. (2005). Intrusion detection: A survey *Managing Cyber Threats* (pp. 19-78): Springer.

Limmer, T., & Dressler, F. (2010). *Dialog-based payload aggregation for intrusion detection.* Paper presented at the Proceedings of the 17th ACM conference on Computer and communications security.

Lippmann, R. P., Fried, D. J., Graf, I., Haines, J. W., Kendall, K. R., McClung, D., . . . Cunningham, R. K. (2000). *Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation.* Paper presented at the DARPA Information Survivability Conference and Exposition, 2000. DISCEX'00. Proceedings.

Locasto, M. E., Wang, K., Keromytis, A. D., & Stolfo, S. J. (2006). *Flips: Hybrid adaptive intrusion prevention.* Paper presented at the Recent Advances in Intrusion Detection.

Lunt, T. F., Jagannathan, R., Lee, R., Whitehurst, A., & Listgarten, S. (1989, 27-31 Mar 1989). *Knowledge-based intrusion detection.* Paper presented at the AI Systems in Government Conference, 1989.,Proceedings of the Annual.

Maglaras, L. A., & Jianmin, J. (2014, 27-29 Aug. 2014). *Intrusion detection in SCADA systems using machine learning techniques.* Paper presented at the Science and Information Conference (SAI), 2014.

Marchette, D. J. (2001). *Computer intrusion detection and network monitoring: a statistical viewpoint*: Springer.

Masarat, S., Taheri, H., & Sharifian, S. (2014). *A novel framework, based on fuzzy ensemble of classifiers for intrusion detection systems.* Paper presented at the Computer and Knowledge Engineering (ICCKE), 2014 4th International eConference on.

Massicotte, F., & Labiche, Y. (2012, 27-30 Nov. 2012). *On the Verification and Validation of Signature-Based, Network Intrusion Detection Systems.* Paper presented at the Software Reliability Engineering (ISSRE), 2012 IEEE 23rd International Symposium on.

Maxion, R. A., & Townsend, T. N. (2002). *Masquerade detection using truncated command lines.* Paper presented at the Dependable Systems and Networks, 2002. DSN 2002. Proceedings. International Conference on.

May, M. (2004). Specification-based Intrusion Detection.

McHugh, J. (2001). Intrusion and intrusion detection. *International Journal of Information Security, 1*(1), 14-35.

Ming-Yang, S., Chun-Yuen, L., Sheng-Wei, C., & Han-Chung, H. (2011, 27-30 June 2011). *Genetic-fuzzy association rules for network intrusion detection systems.* Paper presented at the Fuzzy Systems (FUZZ), 2011 IEEE International Conference on.

Mitchell, R., & Chen, I.-R. (2014). A survey of intrusion detection techniques for cyber-physical systems. *ACM Comput. Surv., 46*(4), 1-29. doi: 10.1145/2542049

Mitchell, R., & Chen, I. R. (2014). Behavior Rule Specification-based Intrusion Detection for Safety Critical Medical Cyber Physical Systems. *Dependable and Secure Computing, IEEE Transactions on, PP*(99), 1-1. doi: 10.1109/TDSC.2014.2312327

Mitchell, R., & Ing-Ray, C. (2014). Adaptive Intrusion Detection of Malicious Unmanned Air Vehicles Using Behavior Rule Specifications. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on, 44*(5), 593-604. doi: 10.1109/TSMC.2013.2265083

Modi, C., Patel, D., Borisaniya, B., Patel, H., Patel, A., & Rajarajan, M. (2013). A survey of intrusion detection techniques in cloud. *Journal of network and computer applications, 36*(1), 42-57.

Moffie, M., Kaeli, D., Cohen, A., Aslam, J., Alshawabkeh, M., Dy, J., & Azmandian, F. (2014). VMM-based intrusion detection system: Google Patents.

Moore, D., & Shannon, C. (2002). *Code-Red: a case study on the spread and victims of an Internet worm.* Paper presented at the Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment.

More, S., Matthews, M., Joshi, A., & Finin, T. (2012, 24-25 May 2012). *A Knowledge-Based Approach to Intrusion Detection Modeling.* Paper presented at the Security and Privacy Workshops (SPW), 2012 IEEE Symposium on.

Mukkamala, S., Sung, A. H., & Abraham, A. (2005). Intrusion detection using an ensemble of intelligent paradigms. *Journal of network and computer applications, 28*(2), 167-182.

Naik, M., Ajgaonkar, K., Nadarge, S., & Agawane, M. K. M. R. (2014). A Survey on Modeling & Detection of Camouflaging Worm. *International Journal Of Scientific Research And Education, 2*(04).

Naoum, R. S., & Al-Sultani, Z. N. (2012). Learning Vector Quantization (LVQ) and k-Nearest Neighbor for Intrusion Classification. *Learning, 2*(3), 105-109.

Nikolai, J., & Wang, Y. (2014). *Hypervisor-based cloud intrusion detection system.* Paper presented at the Computing, Networking and Communications (ICNC), 2014 International Conference on.

Om, H., & Gupta, A. K. (2014). Feature Selection and Decision Tree: A Combinational Approach for Intrusion Detection. *Case Studies in Secure Computing: Achievements and Trends*, 27.

Oza, N. C., & Tumer, K. (2008). Classifier ensembles: Select real-world applications. *Information Fusion, 9*(1), 4-20.

Panda, M., Abraham, A., & Patra, M. R. (2012). A hybrid intelligent approach for network intrusion detection. *Procedia Engineering, 30*, 1-9.

Park, K., Lin, Y., Metsis, V., Le, Z., & Makedon, F. (2010). *Abnormal human behavioral pattern detection in assisted living environments.* Paper presented at the Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments.

Park, Y., Baek, S. H., Kim, S.-H., & Tsui, K.-L. (2014). Statistical Process Control-Based Intrusion Detection and Monitoring. *Quality and Reliability Engineering International, 30*(2), 257-273. doi: 10.1002/qre.1494

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, K. (2013). *SECURITY SURVEY FOR CLOUD COMPUTING: THREATS & EXISTING IDS/IPS TECHNIQUES.* Paper presented at the International Conference on Control, Communication and Computer Technology, 24th.

Paxson, V. (1999). Bro: a system for detecting network intruders in real-time. *Computer networks, 31*(23), 2435-2463.

Pitropakis, N., Pikrakis, A., & Lambrinoudakis, C. (2014). Behaviour reflects personality: detecting co-residence attacks on Xen-based cloud environments. *International Journal of Information Security*, 1-7.

Programs, U. S. C. f. R. o. F. S. (2002). *A review of FBI security programs*: William S. Hein & Co., Inc.

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

Rao, U. P., Singh, N. K., Amin, A. R., & Sahu, K. (2014). *Enhancing detection rate in database intrusion detection system.* Paper presented at the Science and Information Conference (SAI), 2014.

Reddy, R. R., Ramadevi, Y., & Sunitha, K. (2014). *Real Time Anomaly Detection Using Ensembles.* Paper presented at the Information Science and Applications (ICISA), 2014 International Conference on.

Rehman, A., & Saba, T. (2012). Evaluation of artificial intelligent techniques to secure information in enterprises. *Artificial Intelligence Review*, 1-16.

Revathi, S., & Malathi, A. (2014). Detecting User-To-Root (U2R) Attacks Based on Various Machine Learning Techniques.

Rokach, L. (2010). Ensemble methods in supervised learning *Data mining and knowledge discovery handbook* (pp. 959-979): Springer.

Rowland, C. H. (2002). Intrusion detection system: Google Patents.

Ryan, J., Lin, M.-J., & Miikkulainen, R. (1998). Intrusion detection with neural networks. *Advances in neural information processing systems*, 943-949.

Sabhnani, M., & Serpen, G. (2003). *KDD Feature Set Complaint Heuristic Rules for R2L Attack Detection.* Paper presented at the Security and Management.

Schmidt, M. B., & Arnett, K. P. (2005). Spyware: a little knowledge is a wonderful thing. *Communications of the ACM, 48*(8), 67-70.

Schonlau, M., DuMouchel, W., Ju, W.-H., Karr, A. F., Theusan, M., & Vardi, Y. (2001). Computer Intrusion: Detecting Masquerades. 58-74. doi: 10.1214/ss/998929476

Sebring, M. M., Shellhouse, E., Hanna, M., & Whitehurst, R. (1988). *Expert systems in intrusion detection: A case study.* Paper presented at the Proceedings of the 11th National Computer Security Conference.

Sengupta, N., & Sil, J. (2011). Comparison of Performance for Intrusion Detection System Using Different Rules of Classification *Computer Networks and Intelligent Computing* (pp. 87-92): Springer.

Seresht, N. A., & Azmi, R. (2014). MAIS-IDS: A distributed intrusion detection system using multi-agent AIS approach. *Engineering Applications of Artificial Intelligence, 35*, 286-298.

Shiri, F. I., Shanmugam, B., & Idris, N. B. (2011, 27-29 May 2011). *A parallel technique for improving the performance of signature-based network intrusion detection system.* Paper presented at the Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on.

Shuang-can, Z., Chen-jun, H., & Wei-ming, Z. (2014). Multi-Agent Distributed Intrusion Detection System Model Based on BP Neural Network. *International Journal of Security & Its Applications, 8*(2).

Singhal, P., & Singh, G. (2014). Enhanced Intrusion Detection System using Hybrid Machine Learning Approach. *International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), 3*(7), pp: 384-388.

Snapp, S. R., Brentano, J., Dias, G. V., Goan, T. L., Heberlein, L. T., Ho, C.-L., . . . Grance, T. (1991). *DIDS (distributed intrusion detection system)-motivation, architecture, and an early prototype.* Paper presented at the Proceedings of the 14th national computer security conference.

Sonawane, S., Pardeshi, S., & Prasad, G. (2012). A survey on intrusion detection techniques. *World journal of science and technology, 2*(3), 127-113ISSN.

Spafford, E. H., Heaphy, K. A., & Ferbrache, D. J. (1989). A computer virus primer.

Subbulakshmi, T., Shalinie, S. M., GanapathiSubramanian, V., BalaKrishnan, K., AnandKumar, D., & Kannathal, K. (2011, 14-16 Dec. 2011). *Detection of DDoS attacks using Enhanced Support Vector Machines with real time generated dataset.* Paper presented at the Advanced Computing (ICoAC), 2011 Third International Conference on.

Taub, L. (2013). *Applying Conditional Random Fields to payload anomaly detection with CRFPAD.* Paper presented at the Southeastcon, 2013 Proceedings of IEEE.

Tavani, H. T. (1999). Social and ethical aspects of information technology. *Wiley Encyclopedia of Electrical and Electronics Engineering*.

Thaseen, S., & Kumar, C. (2013). *An analysis of supervised tree based classifiers for intrusion detection system.* Paper presented at the Pattern Recognition, Informatics and Medical Engineering (PRIME), 2013 International Conference on.

Tsang, C.-H., & Kwong, S. (2005). *Multi-agent intrusion detection system in industrial network using ant colony clustering approach and unsupervised* *feature extraction.* Paper presented at the Industrial Technology, 2005. ICIT 2005. IEEE International Conference on.

Undercoffer, J., Joshi, A., Finin, T., & Pinkston, J. (2003). Using DAML+ OIL to classify intrusive behaviours. *The Knowledge Engineering Review, 18*(03), 221-241.

Undercoffer, J., Joshi, A., & Pinkston, J. (2003). *Modeling computer attacks: An ontology for intrusion detection.* Paper presented at the Recent Advances in Intrusion Detection.

Vise, D. A. (2002). *The bureau and the mole: the unmasking of Robert Philip Hanssen, the most dangerous double agent in FBI history*: Atlantic Monthly Press.

Vokorokos, L., Balaz, A., & Trelova, J. (2012, 13-15 June 2012). *Distributed intrusion detection system using self organizing map.* Paper presented at the Intelligent Engineering Systems (INES), 2012 IEEE 16th International Conference on.

Vokorokos, L., Chovanec, M., Latka, O., & Kleinova, A. (2008, 21-22 Jan. 2008). *Security of distributed intrusion detection system based on multisensor fusion.* Paper presented at the Applied Machine Intelligence and Informatics, 2008. SAMI 2008. 6th International Symposium on.

Wang, K., Parekh, J. J., & Stolfo, S. J. (2006). *Anagram: A content anomaly detector resistant to mimicry attack.* Paper presented at the Recent Advances in Intrusion Detection.

Warrender, C., Forrest, S., & Pearlmutter, B. (1999). *Detecting intrusions using system calls: Alternative data models.* Paper presented at the Security and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on.

Wilson, R., & Obimbo, C. (2011). *Self-organizing feature maps for user-to-root and remote-to-local network intrusion detection on the KDD cup 1999 dataset.* Paper presented at the Internet Security (WorldCIS), 2011 World Congress on.

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

Xiao, L., Chen, Y., & Chang, C. K. (2014). *Bayesian Model Averaging of Bayesian Network Classifiers for Intrusion Detection.* Paper presented at the Computer

Software and Applications Conference Workshops (COMPSACW), 2014 IEEE 38th International.

Zainal, A., Maarof, M. A., Shamsuddin, S. M., & Abraham, A. (2008). *Ensemble of one-class classifiers for network intrusion detection system.* Paper presented at the Information Assurance and Security, 2008. ISIAS'08. Fourth International Conference on.

Zalavadia, M. B. S. (2014). Network Security Issues and Solutions.

Zou, C. C., Gong, W., & Towsley, D. (2002). *Code red worm propagation modeling and analysis.* Paper presented at the Proceedings of the 9th ACM conference on Computer and communications security.