HYBRID FUZZY TECHNIQUES FOR UNSUPERVISED INTRUSION DETECTION SYSTEM

WITCHA CHIMPHLEE

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia

JULY 2008

To my beloved father, mother, wife, daughter, and parents

ACKNOWLEDGMENTS

First, I am deeply grateful to my supervisors Prof. Dr.Abdul Hanan Abdullah and my co-advisor Associate Prof. Dr.Mohd Noor Md Sap for constant encouragement, thoughtful advice, and expert supervision.

Secondly, I wish to thank and express my deepest gratitude and appreciation to Prof.Dr.Mohd. Ishak bin Desa, Prof.Dr.Md.Yazid bin Mohd Saman, Associate.Dr.Mohd Aizaini Bin Maarof, Associate Dr. Naomie Salim, Associate Dr.Siti Mariyam Shamsuddin, Associate Dr.Abdul Samad Bin Haji Ismail, Assoc. Dr.Siti Zaiton Mohd Hashim and Dr.Kamalrulnizam Bin Abu Bakar for their valuable suggestions during my researches. I sincerely thank to all staff of our faculty who extended their best cooperation during my study and stay here.

Thirdly, special thanks to my wife, Siriporn Chimphlee, for her unwavering love, understanding, encouragement, and tolerance. I would like to dedicate this thesis to my darlings, Waratchaya Chimphlee, who have filled my life with joy, hope, and passion. I am also deeply indebted to my family members in Thailand, who provided their unconditional love and support throughout the years and gave me strength to complete my graduate studies. This thesis is dedicated in part to them.

Fourthly, I would like to thank Suan Dusit Rajabhat University and the Associate Prof.Dr. Sirote Phonpantin for his constant encouragement on my pursuing continues study and gave me a chance to have study in overseas. I would like to thank Thailand’s Commission on Higher Education, Ministry of Education for funding and its support for my studying.

Finally, I want to thank my colleagues: Dr.Anjum, Dr.Faisal, Dr.Mahmood, Dr.Wahub, Dr.Hany, Dr.Abdul rahman, Dr.Cahol, Dr.Adul Majid, Dr.Mohd Yazid, Dr.Anazida, Dr.Foad, Daliyoes, Satria, Shukor, Azlan, Chaliaw, Siriluck, Kittikhun, Nimitr, Ladda and Nano for their discussions and friendship. Without them, my life in UTM would not have been so enjoyable.

ABSTRACT

Network intrusion detection is a complex research problem especially when it deals with unknown patterns. Furthermore, if the amount of audit data instances is large, human labelling becomes tedious, time-consuming, and expensive. A technique which can enhance the learning capability of an anomaly intrusion detection system is required. Unsupervised anomaly detection methods have been deployed to address the weaknesses of both signature-based and supervised anomaly detection. These methods take a set of unlabelled data as input, in which the majority of data set is normal traffic, and attempt to find intrusion hidden in the data. Although the unsupervised anomaly detection has received a lot of attention from many researchers, it still has many drawbacks which can be improved. This thesis proposes a framework which comprises three components: feature selection, new clustering and novel cluster labelling. The task of feature selection is to choose relevant feature which is obtained through statistical testing. The new clustering technique is called F2ART which is a hybrid of Fuzzy c-means and Fuzzy Adaptive Resonance Theory. It incorporates a modified similarity measure and a new learning rule which also includes a fuzzy membership value in improving the detection rate. Finally this thesis also proposes a new cluster labelling algorithm called Normal Membership Factor (NMF). This algorithm introduces weighting degree of probability of clusters, which can decrease false positive rate. Based on the experimental results that have been carried out using the KDD Cup 1999 data set, it indicates that the framework provides the best performance in terms of detection rate compared to the current unsupervised anomaly detection approaches. Unlike traditional anomaly detection methods that require 98 percent of the unlabelled data to be in normal pattern, this framework can still work with only 80 percent of the normal pattern. In addition, it can also improve the analysis of new data over time without the need to retrain over all the previous and new data.

ABSTRAK

Pengesanan pencerobohan rangkaian merupakan bidang kajian yang kompleks terutamanya jika ianya melibatkan corak yang tidak dikenali. Di samping itu, jika data audit trafik menjadi besar, penglabelan mengambil masa yang lama, rumit serta mahal. Teknik baru yang boleh memberikan keupayaan pembelajaran yang lebih baik terhadap sistem pengesanan anomali adalah diperlukan. Kaedah pengesanan secara anomali tidak berselia dapat mengatasi masalah yang ada pada kaedah berasaskan tandatangan dan pengesanan anomali berselia. Kaedah-kaedah ini akan mengambil satu set data tanpa label sebagai input, di mana majoriti set data itu adalah trafik normal dan seterusnya akan mencuba untuk mengenalpasti pencerobohan tersembunyi di dalam data. Walau pun pengesanan anomali tidak berselia telah mendapat perhatian ramai penyelidik, masih terdapat kelemahan pada kaedah ini yang boleh diperbaiki. Objektif kajian ini adalah untuk mencadangkan satu rangka kerja yang mengandungi 3 komponen asas iaitu: pemilihan ciri, kaedah pengkelompokan dan pelabelan kelompok yang baru. Pemilihan ciri adalah bertujuan untuk menentukan ciri-ciri yang berkaitan sahaja yang diperolehi melalui ujian statistikal. Manakala teknik pengkelompokan baru yang dikenali sebagai F2ART iaitu gabungan *Fuzzy c-means* dan *Fuzzy Adaptive Resonance Theory* dicadangkan bagi mempercepatkan pengesanan terhadap serangan yang baru. Teknik ini menggunakan pengukuran persamaan yang telah diubahsuai dan peraturan pengetahuan termasuk nilai keahlian kabur. Kajian ini juga turut mencadangkan algoritma penglabelan kelompok yang baru yang dikenali sebagi *Normal Membership Factor* (NMF). Ia menggunakan pendekatan pemberat kepada kebarangkalian kelompok yang dapat mengurangkan kadar penggeraan palsu. Berdasarkan ujikaji yang menggunakan set data KDD Cup 1999, didapati rangka kerja cadangan memberi prestasi terbaik berbanding pengesanan anomali tidak berselia sedia ada. Berbeza dengan kaedah pengesanan anomali tradisional yang memerlukan 98 peratus data tidak berlabel bercorak normal, rangka kerja ini hanya memerlukan 80 peratus daripada data tidak berlabel bercorak normal. Di samping itu, ianya boleh memperbaiki analisis data baru tanpa perlu dilatih semula menggunakan data-data terdahulu dan data baru.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AID	Anomaly Intrusion Detection
ART	Adaptive Resonance Theory
AUC	Area Under the Curve
BN	Bayesian Networks
CART	Classification and Regression Trees
CE	Classification Error
DARPA	The Defence Advanced Research Projects Agency
DR	Detection Rate
DoS	Denial of Service
EM	Expectation Maximisation
F2ART	Hybrid Fuzzy c-means and Fuzzy Adaptive Resonance Theory
FAR	False alarm Rate
FART	Fuzzy Adaptive Resonance Theory
FCM	Fuzzy c-means
FNR	False Negative Rate
FPR	False Positive Rate
GCV	Generalised cross-validation
GrIDS	Graph-based Intrusion Detection System
HIDS	Host-based Intrusion Detection System
ICA	Independent Component Analysis
IDAMN	Intrusion Detection Architecture for Mobile Networks
ID	Intrusion detection
IDES	Intrusion Detection Expert System
IDP	Intrusion Detection Problem
IDS	Intrusion detection systems
IGDR	Intrusive Generalisation or Detection Rate
KDD	Knowledge Discovery in Databases

LGP	Linear Genetic Programming
LTM	Long Term Memory
MARS	Multivariate Adaptive Regression Spines
MCE	Mean Classification Error
MSE	Mean Square Error
NFR	Network Flight Recorder
NG	Normal Generalisation
NIDS	Network-based intrusion detection systems
NMF	Normal Membership Factor
NSM	Network Security Monitor
OG	Overall Generalisation
PC	Principal Component
PCA	Principal Component Analysis
PV	Principal Variable
R2L	Remote to Login
ROC	Receiver Operating Characteristic
SAINT	Security Analysis Integration tool
STM	Short-term memory
SVM	Support Vector Machine
TCP	Transmission Control Protocol
TPR	True Positive Rate
U2R	User to Root
UAD	Unsupervised Anomaly Detection
UADR	Unknown Attack Detection Rate

LIST OF SYMBOLS

z	Membership value is low
Π	Membership value is medium
S	Membership value is high
$F = \{f_1, f_2, ..., f_p\}$	List of feature
$\mathrm{cov}(y)$	Covariance matrix y
$Cor[\ ]$	Correlation matrix
λ_i	Eigenvalues
e^i	Eigenvectors
$tr(A)$	Determinant and trace of the matrix A
O_{ij}^{pred}	Predicted output
O_{ij}^{pred}	Desired output
ρ	Vigilance parameter value
α	Choice parameter
β	Learning rate value
$\wedge$	Fuzzy AND operator
w_j	Weight vector
t_j	Top-down weight
b_j	Bottom-up weight
$\|\ .\ \|$	The norm operator
$I=(a,a^c)$	Complement code
$U = (\mu_{ij})_{Nxc}$	Fuzzy partition matrix
$d_{ij} = \|\| x_i - v_j \|\|^2$	Euclidean distance
$m \in [1, \infty)$	The weighting exponent
μ_{ij}	Membership function value
$WC(c_i)$	Weight of clusters

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

The computer networks security plays a strategic role in modern computer systems. The continual increase of attacks against networks and their resources has created a necessity to protect these valuable assets. Attacks on computer networks are serious problem because most deployed computer systems are vulnerable to those attacks. Most attacks are composed of a series of anomaly events. Intrusion detection (ID) is a rapidly growing field and it is an important technology for the business sector in its effort to build systems for network security. It involves processing and learning of the large number of examples in order to detect intrusions. Such process becomes computationally costly and impractical when the number of records, to train against, grows dramatically. It is critical to develop methods for data dimension reduction, effective monitoring algorithms for intrusion detection, and means for their performance improvement [1]. Therefore, unsupervised learning is very beneficial for intrusion detection domain, since the labelled data is expensive while unlabeled data can be obtained very easily from log files and audit files.

1.1 Overview

In the era of information society, as computer networks and related applications are becoming more and more popular, the potential threats to the global information infrastructure have increased tremendously. To defend various cyber attacks and computer viruses, lots of computer security techniques have been studied in the last decade, which include cryptography, firewalls and intrusion detection systems (IDSs). When an attack occurs, instead of taking preventive measures, intrusion detection mechanisms usually will only log or report the incident. It can be defined as the problem of identifying the activity of individuals who are using a computer system without authorisation or those who have legitimate access to the system but are abusing their privileges. Intrusion detection systems (IDS) have been actively investigated for about two decades. Despite the substantial research efforts and commercial investments, IDS are still immature and cannot be considered as a complete defence because of the low ability to detect new types of attack and high false alarms rates. Anomaly detection consists of analysing and reporting unusual behavioural patterns in computing systems. According to Axelsson, "the early anomaly detection systems were self-learning, that is they automatically formed an opinion of what the subject's normal behaviour was" [2]. This is due in part to uncertain situations, which come from the unknown characteristics of attacks and system vulnerabilities.

The implementation of early intrusion detection mechanisms was primarily based on the audit records generated by the host operating system. Audit data were manually inspected by system administrators or security experts in order to detect intrusions. This was expensive, time-consuming, and inaccurate due to the extremely large amount of audit data. As a result, the "misuse detection scheme" was then developed. In misuse detection, previous attack signatures are stored and attacks are detected by matching audit events with the stored signatures. Although misuse detection methods can find most known attacks if the signatures are well defined, they are useless for detecting unknown intrusions. Moreover, defining an attack signature is not an easy task at all. To address the weaknesses of misuse detection, the concept of anomaly detection was introduced to monitor systems by Anderson [3] in 1980 and was then improved by Denning [4] in 1987. Denning assumed that security violations could be detected by inspecting abnormal system usage patterns from the audit data. Deviations from normal behaviour patterns are flagged systematically as intrusions. The implementations of early anomaly detection techniques were based on self-learning. Knowledge about normal behaviours of subjects was automatically formed through training. The notion of anomaly detection did not only consider the normal profiles but it also took into account the abnormal behaviours that are extracted from known attacks. Thus, according to whether the learning process, the anomaly detection schemes are naturally classified into two categories: supervised and unsupervised. Regardless of the approach used, the intrusion detection problem has been formulated to classify system behaviour patterns into two categories: normal and abnormal.

Supervised anomaly detection schemes depend on labelled training datasets, making the intrusion detection process error-prone, costly and time consuming. It is concerned with a collection of labelled data that come in the form of ordered pairs namely a feature vector describing the data and its class assignments. Supervised learning methods build a model for rare events. Any mistake in labelling the training data may lead to decreased performance of the detector. On the other hand, unsupervised anomaly detection schemes allow training based on unlabeled datasets, facilitating online learning and improving detection accuracy. By facilitating online learning, unsupervised approaches provide a higher potential to find novel attacks, which are not always included in the training data. By removing the need to label the dataset, unsupervised approaches carry greater potential for detection accuracy. The clustering technique is a part of unsupervised learning for intrusion detection, whereby the task of determining the number of clusters is a difficult issue since the occurrence of intrusions is unknown [5]. It is require some techniques do not need labelled training data, which determines automatically the optimal number of clusters for a set of data. It also can learn a new pattern and it is not forgetting those learned previously, thereby significantly reducing false alarm rate when normal behaviour is changing.

In the following sections, this chapter briefly states the background of the problem, statement of the problem, objectives of study, scope of study, significant of the study, and outline of the thesis.

1.2 Background of the Problem

Most of the existing IDS use all the features in network packet to measure and look for intrusive patterns. Some of these features are irrelevant and redundant. The drawback of all the features may degrade the performance of an IDS [6]. There are many techniques for feature selection including Artificial Neural Network, Support Vector Machine, Genetic Algorithm, Principal Component Analysis, Rough Set and few others [7]. The Feature selection in IDS is finding best feature subset to represent the data for next processing. The significance of feature selection can be viewed as following. First is to filter out noise and remove redundant and irrelevant features. Second, feature selection can be implemented as an optimization procedure of search for an optimal subset of features that better satisfy a desired measure [6]. The output of an IDS can only be as accurate as its input [8]. For detecting a given type of attack the IDS needs to be capable of making the appropriate observations, i.e., it needs access to data that are relevant for detecting the attack. As new network attacks are emerging, the need for IDS to detect novel attacks becomes pressing [9]. Misuse detection by nature is unable to detect new attacks [10]. Due to that, the method is very efficient in reducing false alarms; it requires training data with labelled attacks [11]. Indeed, training data with labelled attacks is rarely available. The problem is further complicated by a limitation of classification algorithms that a classifier can only recognise the classes it has seen in the training data. Many researchers have highlighted the conventional way of making misuse rather than anomaly detection [12]. Clustering is one popular method that has been used to discriminate against the normal deviations (normal activity) from abnormal deviations (attacks) [13]. The clusters may easily miss new attacks even when they are captured by an anomaly detection module. So the problem of how to reduce an anomaly detection system's false alarm rate and meanwhile preserving its ability to detect new attacks poses a challenge [14].

All the existing clustering methods have some built-in shortages: (1) the result of detection is sensitive to the parameters that are difficult to be determined and (2) it is not reasonable to assume that the smaller size clusters of objects have more possibilities of being anomalous [15]. A clustering technique can be used as a classification one by assigning to each cluster the label of the class with more data samples in the cluster. If two or more class labels can be assigned, then a conflict resolution strategy can be applied. Clustering is used in anomaly detection systems to separate attack and normal samples. The most important advantage of using clustering to detect attacks is the ability to find new attacks that have not been seen before (i.e., no recorded pattern signatures associated with the new attacks) [15]. Traditional classification-based systems will have difficulty classifying such attack correctly. Clustering algorithms can group new data instances into coherent groups that can be used to augment the performance of existing classifiers. High quality ("pure") clusters can also assist an expert with labelling [14].

Traditional anomaly detection algorithms often require a set of purely normal traffic data from which models can be trained to represent normal traffics [16]. The labelled data or purely normal data is not readily available since it is time consuming and expensive to be manually classified. Purely normal data is also very hard to obtain in practice, since it is very hard to guarantee that there are no intrusions when it was collecting network traffic. The amount of available network audit data instances is usually large; human labelling is tedious, time-consuming, and expensive. Many methods of IDS totally depend on the training data sets, which should not only be "clean" data sets but also involve most normal behavioural patterns of the detected object. However, it is indeed very difficult and costly to meet both the requirements [17].

Applying unsupervised anomaly detection in network intrusion detection is a new research area that has already drawn interest in the academic community. Eskin *et al* [18] investigated the effectiveness of three algorithms in Intrusion Detection. Supervised anomaly detection in network intrusion detection, which uses purely normal instances as training data, has been studied extensively in the academic community. An approach for modelling normal traffic using self-organising maps is presented in [19], while another one uses principal component classifiers to obtain the model [20]. Another approach uses the normal data to generate abnormal data and uses it as input for a classification algorithm [21]. Though unsupervised intrusion detections in general look promising, it is believed that their approach has a few problems. First, they modified the data significantly by limiting the number of attacks to 1 ~ 1.5 % of the complete training dataset so that their hypothetical assumption is true. Second, each cluster is self-labelled as attacks or normal, based purely on the number of instances in it. This is also the primary reason they control the percentage of attacks in the whole dataset to be very small (< 1.5 %) [14]. Third, the results of detection are sensitive to the parameters which are difficult to be determined. Finally, it is not reasonable that the objects in the small clusters are labelled anomalous [17]. However, there have been insufficient discussions about the proportion ratio of normal pattern in data set.

Fuzzy logic techniques and theorems can deal with vagueness and imprecision in the real world. It has been widely used in control systems, decision-making, and information retrieval, but has not yet made substantial inroads into computer security. To use fuzzy systems to identify malicious network activity that combines simple network traffic metrics with fuzzy rules to determine the likelihood of specific or general network attacks [22]. The advantage of using fuzzy logic is that it allows one to represent concepts that could be considered to be in more than one category (or from another point of view – it allows representation of overlapping categories) [23]. In standard set theory, each element is either completely a member of a category or not a member at all. In contrast, fuzzy set theory allows partial membership in sets or categories. Fuzzy logic has been combined with data mining techniques for solving the intrusion detection problem (IDP) [24]. The purpose of introducing fuzzy logic is to deal with the fuzzy boundary between the normal and abnormal classes. Fuzzy rules allow us to easily construct if-then rules that reflect common ways of describing security attacks. The types of attacks that can be described may be of a general nature or very specific, depending on the granularity of the data feeds used in the rules [25].

Some early research on IDS attempted to use neural nets for intrusion detection. Such systems were trained on normal or attack behaviour information and then detect intrusions or attacks. Supervised and unsupervised nets have been used in IDSs. Most supervised neural net architectures require retraining, in order to improve analysis capability due to changes in the input data. On the contrary, unsupervised nets offer an increased level of adaptability to neural nets, and have been used in intrusion detection systems. An Adaptive Resonance Theory (ART) networks cluster inputs by unsupervised learning [26]. Each time a pattern is presented, an appropriate cluster unit is chosen, and the cluster's weights are adjusted to let the cluster unit learn the pattern. The degree of similarity of patterns placed in the same cluster is controlled by a reset mechanism via a *vigilance parameter* [27]. A new pattern presented to the nets is associated with one of the existing clusters, only if the feature is similar to the members of the cluster. Otherwise, the nets create a new cluster. ART is used for classifying network traffic into normal and intrusive/attack [28].

The performance of intrusion detection system is measured by how well the system can accurately predict intrusion and low false positive rate [29]. There are numerous methods that discuss the evaluations of intrusion detection systems. Some methods emphasise on the importance of *detection rate (DR)* and *false positive rates (FPR)*; while others look into the novel pattern detection rate [30]. The performance of classifiers is evaluated with respect to their classification of unseen normal and intrusive patterns. The metrics embraced here are the generalisation abilities of the classifiers because they are the most important aspects of an anomaly detection scheme. Evaluation of the generalisation capability of any intrusion detection should consider the ability of the system to recognise new normal as well as intrusive behaviours [31].

Many researchers design the framework by integrates another component for increasing the accuracy detection and decreasing false alarm rate. It considers the issues involved in standardising formats, protocols, and architectures to co-manage intrusion detection and response systems [32]. Most current intrusion detection systems employ signature-based methods or data mining methods which rely on labelled training data [18]. Intrusion detection (ID) is an important component of infrastructure protection mechanisms. Intrusion detection systems (IDSs) need to be accurate, adaptive, and extensible. Given these requirements and the complexities of today's network environments, it needs a more systematic and automated IDS development process rather that the pure knowledge encoding and engineering approaches [33]. Currently anomaly intrusion detection framework has disadvantage. First, based on the practical assumption that normal instances dominate attack instances, the authors simplify self-labelling heuristic by find the largest cluster and label it *normal*; sort the remaining clusters in ascending order of their distances to the larger cluster; label all the other clusters as *attacks* [13]. Secondly, the clustering groups of the data require a number of clusters before processing [14]. Thirdly, retraining over all data includes the previous and new data. It takes much time for this task [5]. Lastly, some framework was not handling unseen patterns [34].

The normal behaviour is profiled based on normal data for anomaly detection and the behaviour of each type of attack are built based on attack data for intrusion identification. Given a data set with possible unlabelled attacks, it desires an algorithm that learns a model for anomaly detection. In this case, it does not assume the training data to be free of attacks. However, it assumes that the majority of the training data is normal; otherwise, the attacks are said to constitute "normal behaviour." It also desires the algorithm and the learned models to achieve relatively high detection rates with low false alarm rates. Three main issues need to be addressed here. Firstly, determining the number of clusters and secondly without the requirement of retraining over all the previous and new data. Thirdly, the ART models solve the so-called *stability-plasticity* dilemma where new patterns are learned without forgetting those learned previously.

1.3 Statement of the Problem

Many intrusion detection systems attempt to design the framework for increasing the accuracy detection and decreasing false alarm rate. In the complexity of today's network environments, it needs the framework that cooperates with connected and related several component for accurate, adaptive, and extensible. Given these requirements it needs a more systematic and automated IDS development process rather that the pure knowledge encoding. A framework consists of components. Supervised anomaly detection is the one of component can be tackle IDS problem. It establishes normal profiles of systems or networks by training using a labelled dataset. It has drawback with incapability to the analysis of new data over time without the requirement of retraining over all the previous and new data. The biggest problem of supervised anomaly detection is the need to label the training data. Otherwise, unsupervised anomaly detection uses unlabelled or noisy data to identify intrusions. It allows training based on unlabelled datasets, which is easy to obtain from a real world system, facilitating online learning and improving detection accuracy. Clustering analysis is the most widely used learning technique in unsupervised anomaly detection schemes [18]. When applying clustering techniques for intrusion detection, determining the number of clusters is a difficult issue since the occurrence of intrusions is unknown. The general approach and current practice assume that data instances always belong to two categories: normal clusters and intrusive clusters, and that the number of normal data instances largely outnumbers the number of intrusions [18, 35, 36]. However, if data instances are impurity, these assumptions unavoidably lead to a high false alert rate. It is to require the technique that to deal with the blur line between the normal and abnormal classes to deal with the fuzzy boundary between the normal and abnormal classes. It is not required to be determined the number of clusters previously and it also can improve the analysis of new data over time without the requirement of retraining over all the previous and new data. In addition, feature selection process in the intrusion detection systems for increase the accuracy of performance of the detection rate. However, *to hybrid more than two techniques it is a challenging task to develop a clustering method that should handle the clustering problem in adaptive learning environment. It is incorporate with feature selection and labeling clusters for producing better results with high detection and low false alarm rate.*

After studying the background of the problem, there are several issues that should be addressed.

1. How many components in a framework that can cover and can solve the IDS problem?
2. Which are important component in a framework?
3. How is the current anomaly intrusion detection being done?
4. What are the existing techniques available?
5. What are their strengths and weaknesses?
6. What are the relevant attributes to be considered in anomaly intrusion detection?
7. How to improve the detection rate and reduce the false alarm rate of anomaly intrusion detection?

8. How to efficiently and effectively design and implement an intrusion detection system to detect known and novel attacks?
9. Current techniques used in computer security are not able to cope with the changing environment and increasingly complex nature of computer systems and their security. How can these be solved?
10. How can PCA feature selection, F2ART, and NMF solve complex IDS problems, and new pattern detection?
11. Is it possible for PCA to select feature without losing information?
12. Is there any ability of a neural network to learn a new pattern and the ability for the new learning not to be affected by the previous learning?

1.4 Objectives of the Study

The main goal of this research is to improve versions of fuzzy techniques to cluster the attacks type of data. Therefore, this thesis is carried out in order to fulfil the following objectives:

1. To propose a framework that comprise of feature selection, fuzzy clustering and labelling clusters for network anomaly detection with solving complex intrusion detection system problems, i.e. uncertain data, and handle about false alarm rate.
2. To develop a clustering algorithm that hybrid benefit of two techniques together to increase the performance accuracy of detection rate.
3. To develop a clusters labelling algorithms to decrease false positive rate by weighting clusters with a degree of probability of clusters.

Intrusion detection systems attempt to design the framework for increasing the accuracy detection and decreasing false alarm rate. Various techniques were studied in intrusion detection field and still nowadays researchers are still focusing on implementing the latest techniques in order to improve the intrusion detection model. This has raised recent interest in anomaly detection, in which a model is built of normal behaviour and significant deviations from the model are flagged anomalously. Most of the anomaly detection algorithms require the training datasets to be free of attacks. However, the intrusion models that all these methods adopt to totally depend on the instances of the training data sets, so clean data sets (attack free) are crucial for building applied anomaly detection. In fact, collecting clean data sets is very difficult and costly, so it is essential to study the unsupervised intrusion detection methods. It needs to improve the analysis of new data over time without the requirement of retraining over all the previous and new data.

1.5 Scope of the Study

The objectives of this study have been stated in the previous section. In order to achieve these objectives, it is decided to follow the scope, which covers the following aspects:

1. The study focuses only on secondary data, i.e., available from published, authoritative sources. It should be noted that this research is not concerned with real time detection systems but only proposes them.
2. Performance benchmark on KDD Cup 1999 data sets, in measuring the performance and ability of the proposed method by dividing the data into two groups, Group 1 has 88,911 instances and Group 2 has 49,547 records.
3. Using the several evaluations for the performance of the classifiers that calculated based on the testing patterns.

1.6 Significance of the Study

Generally, anomaly intrusion detection approaches build normal profiles from labelled training data. However, labelled training data for intrusion detection is expensive and not easy to be obtained. This thesis addresses the unsupervised anomaly intrusion detection accuracy problem involved in false alarm rate. Towards the conclusion of this thesis it will portray a clearer view regarding the classification of the blurred line between normal behaviours and anomalous. It would be useful to examine the attack with impurity of data set in dynamic environment.

This thesis handles fuzzy attack data to encourage development of systems and algorithms with KDD Cup 1999 dataset by producing the new framework of hybrid Fuzzy c-means, Fuzzy ART, and labelling clusters, for network anomaly detection with solving complex intrusion detection system problems. In addition, it is intended to investigate the importance of pre-processing phase which includes data cleaning and feature selection process in the intrusion detection systems for increase the accuracy of performance of the detection rate. Moreover, this thesis also compared the performance of F2ART framework for network anomaly detection with previously proposed methods in finding the strengths and weaknesses of the proposed method.

In addition, other researchers can take advantage of this research based on the following aspects. First of all, the study contributes to researchers to encouraging that more works should explore the advantages of F2ART through improved theories. Secondly, practitioner can enhance their understanding of this technique by looking at the exposure of another promising technique of intrusion detection system as the existing techniques. Thirdly, the findings from this research are also useful for researchers who are interested in applying F2ART algorithm in fundamental data due to the fact that historical data will be beneficial both in the commercial and academic sectors. Finally, the results of this research will be useful for practitioners who intend to further their study.

1.7 Thesis Outline

The outline of the thesis is provided in Figure 1.1. This thesis concerned with the clustering methods in the computer network intrusion detection areas. It stresses on the interest in detecting known and novel network intrusion attacks that can be detected with activity monitoring schemes. Below is an outline of the thesis. Chapter 1 introduces the problem of computer security and the need for intrusion detection systems that will be further elaborated in the thesis. Chapter 2 reviews literatures dealing with intrusion detection systems and research. It introduces concepts of clustering, soft computing, fuzzy logic, etc. It also includes a brief introduction to data mining, particularly to classification and clustering. A survey of the supervised and unsupervised learning that have been applied to intrusion detection is presented at the end of the chapter.

Chapter 3 presents and discusses the research methodology. Chapter 4 presents the feature selection for effective anomaly detection - some techniques of feature selection methods that have been widely used in this area and presents PCA feature selection with experiments this algorithm. Chapter 5 details the F2ART framework and the Hybrid Fuzzy c-means and Fuzzy Adaptive Resonance Theory (F2ART) clustering approach for intrusion detection. The procedures of FCM, Fuzzy ART are presented as well. A hybrid F2ART is constructed in order to improve the detection rate of attacks.

Chapter 6 describes experimental setup and results for F2ART methods and description of the data was also presented for experimental. This chapter also illustrates the results of applying the FCM, Fuzzy ART, and F2ART methods that the data have used in the case study. The performance of the clustering on intrusion data is also studied. Chapter 7 explains statements on the research achievements, discussions and conclusions of this thesis are presented in this chapter. This is included by the research findings and discussions directions will be made regarding the directions for future research. Appendix B shows the sample of data set in pre-processing step. Appendix C shows list of the presentations and publications.

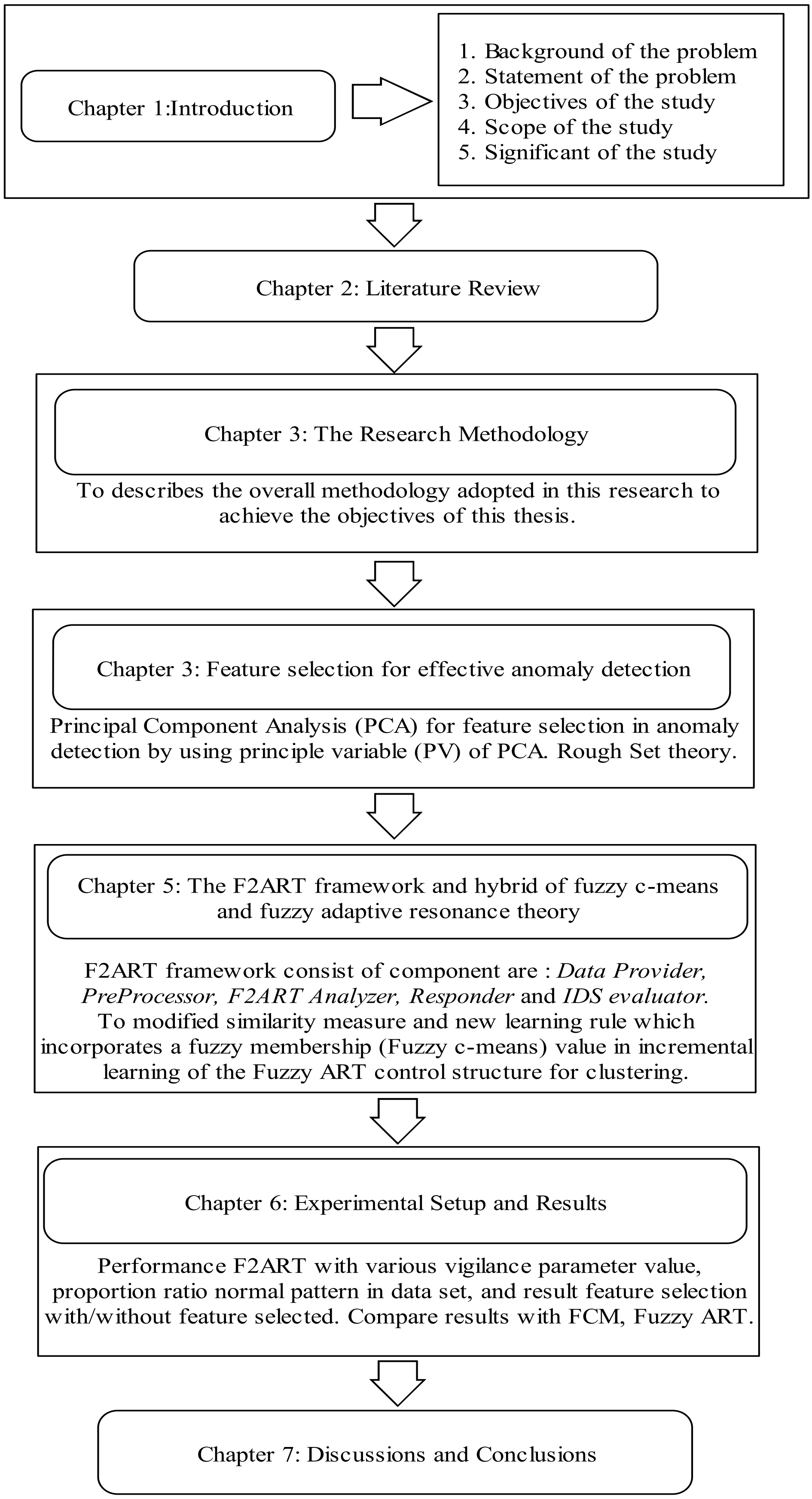

Figure 1.1 Outline of thesis

REFERENCES

1. Zhu X: Anomaly Detection Through Statistics-Based Machine Learning for Computer Networks. The University of Arizona, Department of Systems and Industrial Engineering; 2006.
2. Axelsson S: Intrusion Detection Systems: A Taxonomy and Survey. In *Technical Report No* 99-15: Department of Computer Engineering, Chalmers University of Technology, Sweden; 2000.
3. Anderson JP: Computer Security Threat Monitoring and Surveillance. (79F296400 TR ed. Fort Washington, PA: James P. Anderson Co.,; 1980.
4. Denning DE: An intrusion-detection model. *IEEE Transactions on Software Engineering* 1987, 13.
5. Chimphlee W, Sap MNM, Abdullah AH, Chimphlee S, Srinoy S: Unsupervised Anomaly Detection without Prior Knowledge Using Clustering. In *International workshop on information Technology* 2005 (*IAIT*2005). Bangkok, Thailand; 2005.
6. Zainal A, Maarof MA, Shamsuddin SMH: Feature Selection Using Rough-DPSO in Anomaly Intrusion Detection. In *International Conference Computational Science and Its Applications (ICCSA* 2007). pp. 512-524. Kuala Lumpur, Malaysia; 2007:512-524.
7. Mukkamala S, Sung AH: Feature selection for intrusion detection using neural networks and support vector machines. *Journal of the Transportation Research Board of the National Academies* 2003:33-39.
8. McAuliffe N, Wolcott D, Schaefer L, Hubbard B, Haley T: Is your computer being misused? A survey of current intrusion detection system technology. In *Sixth Computer Security Applications Conference*. 1990: 260-272.
9. Lee W, Stolfo SJ: Data Mining Approaches for Intrusion Detection. In *Proceeding of the 7th USENIX Security Symposium* 2000
10. Lee W, Stolfo SJ, Mok KW: A Data Mining Framework for Building Intrusion Detection Models. In *IEEE Symposium on Security and Privacy*. 1999: 120-132.

11. Giacinto G, Roli F, Didaci L: Fusion of multiple classifiers for intrusion detection in computer networks. *Pattern Recognition Letters* 2003, 24:1795-1803.

12. Verwoerd T, Hunt R: Intrusion Detection Techniques and Approaches. *Computer Communications* 2002, 27:128-140.

13. Li X: Clustering and Classification Algorithm for Computer Intrusion Detection. Arizona State University, Arizona State University; 2001.

14. Zhong S, Khoshgoftaar T, Seliya N: Evaluating Clustering Techniques for Network Intrusion Detection. In 10*th ISSAT International Conference on Reliability and Quality Design*; *August,* 2004*; Las Vegas, Nevada, USA*. 2004: 149-155.

15. Katos V: Network intrusion detection: Evaluating cluster, discriminant, and logit analysis. *Information Sciences* 2007, 2007.

16. Zhong S, Khoshgoftaar T, Seliya N: Clustering-based Network Intrusion Detection. *International Journal of Reliability, Quality and Safety Engineering* 2005.

17. Jiang S, Li Q, Wang H: A Novel Intrusion Detection Method. In *Network and Parallel Computing, IFIP International Conference; Wuhan, China*. Edited by Jin H, Gao GR, Xu Z, Chen H. NPC 2004, LNCS 3222; 2004: 459-462.

18. Eskin E, Arnold A, Prerau M, Portnoy L, Stolfo S: *A Geometric Framework for Unsupervised Anomaly Detection: Detecting Intrusions in Unlabeled Data.* Kluwer Academic Publisher; 2002.

19. Gonzalez F, Dasgupta D: Neuro-immune and self-organizing map approaches to anomaly detection: A comparison. In *In Proceedings of the* 1*st International Conference on Artificial Immune Systems*. 2002: 203-211.

20. Shyu M-L, Chen S-C, Sarinnapakorn K, Chang L: A Novel Anomaly Detection Scheme Based on Principal Component Classifier. In *Proceedings of the IEEE Foundations and New Directions of Data Mining Workshop, in conjunction with Third IEEE International Conference on Data Mining (ICDM'*03); *November* 19*,* 2003*; Melbourne, Florida, USA*. 2003: 172-179.

21. Gonzalez F, Dasgupta D: Anomaly detection using real-valued negative selection. *Genetic Programming and Evolvable Machines* 2003, 4:383-403.

22. Vanderbilt RE: Anomaly Detection in Computer Networks Using Type-2 Fuzzy Logic. Florida Institute of Technology, Computer Engineering; 2005.

23. Guan J, Liu DX, Wang T: Applications of Fuzzy Data Mining Methods for Intrusion Detection Systems. In *Computational Science and Its Applications – ICCSA* 2004: *International Conference*; *May* 14-17,2004*; Assisi, Italy*. Edited by Laganà A, Gavrilova ML, Kumar V, Mun Y, Tan CJK, Gervas O. Springer Berlin / Heidelberg; 2004: 706-714.

24. Gomez J: Soft Computing Techniques for Intrusion Detection. The University of Memphis, 2004.

25. Shah H, Undercoffer J, Joshi A: Fuzzy Clustering for Intrusion Detection. In *Fuzzy Systems,* 2003*, FUZZ* '03*, The* 12*th IEEE International Conference on*; 25-28 *May* 2003. 2003: 1274-1278.

26. Baraldi A, Alpaydm E: Simplified ART: a new class of ART algorithms. (TR-98-004 ed.: International Computer Science Institute, Berkeley CA; 1998.

27. Carpenter GA, Grossberg S: ART2: self-organization of stable category recognition codes for analog input patterns. *Applied Optics* 1987, 26:4919-4930.

28. Xiang G, Min W, Rongchun Z: Application of Fuzzy ART for Unsupervised Anomaly Detection System. In *Computational Intelligence and Security,* 2006 *International Conference on*; *Nov,* 2006. 2006: 621-624.

29. Anderson D, Lunt TF, Javitz H, Tamaru A, Valdes A: Detecting unusual program behavior using the statistical component of the next generation intrusion detection expert system (NIDES). SRI International technical report; 1995.

30. Dasgupta D, Gonzalez F: An immunity-based technique to characterize intrusions in computer networks. *IEEE Transactions on Evolutionary Computation* 2002, 6:281-291.

31. Mori Y, Iizuka M: Study of Variable Selection Methods in Data Analysis and its Interactive System. In *Proceedings of ISM Symposium - Recent Advances in Statistical Research and Data Analysis*. 2000: pp.109-114.

32. Kahn C, Porras PA, Staniford-Chen S, Tung B: A common intrusion detection framework (CIDF). In *Proceedings of the information surviability workshop, Orlanda FL*; *October*. 1998

33. Lee W, Stolfo SJ: A framework for constructing features and models for intrusion detection systems. *ACM Transactions on Information and System Security* 2000, 3.

34. Liu Y, Chen K, Liao X, Zhang W: A genetic clustering method for intrusion detection. *Pattern Recognition* 2004, 37:927-942.

35. Eskin E: Anomaly Detection over Noisy Data using Learned Probability Distributions. In *In Proceedings of the* 17*th International Conference on Machine Learning*; *July* 2000.*; Palo Alto, USA*. 2000

36. Portnoy L, Eskin E, Stofo SJ: Intrusion detection with unlabeled data using clustering. In *Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA*-2001); *November* 5-8,2001*; Philadelphia*. 2001

37. Luo S: Creating Models of Internet Background Traffic Suitable for use in Evaluating Network Intrusion Detection Systems. University of Central Florida, School of Computer Science in the College of Engineering and Computer Science; 2005.

38. Valeur F: Real-Time Intrusion Detection Alert Correlation. University of California Santa Barbara, 2006.

39. Zhang J, Zulkernine M: A Hybrid Network Intrusion Detection Technique Using Random Forests. In *Proceedings of the First International Conference on Availability, Reliability and Security (ARES'*06); 20-22 *April* 2006. IEEE Computer Society; 2006

40. Maiwald E: Fundamentals of Network Security. In. Edited by Woobury B: Mc Graw Hill Technology Education; 2004: *Information Security Series*].

41. NSTAC: Intrusion Detection Subgroup: Report on the NS/EP implications of Intrusion Detection Technology Research and Development.

42. Qureshi AA: Network Intrusion Detection Using An Innovative Statistical Approach. Florida Institute of Technology, 2006.

43. Park Y: A statistical process control approach for network intrusion detection. Georgia Institute of Technology, School of Industrial and Systems Engineering; 2005.

44. Kendall K: *A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems.* Computer Science, Massachusetts Institute of Technology: Boston; 1998.

45. Lazarević A, Ozgur A, Ertoz L, Srivastava J, Kumar V: A comparative study of anomaly detection schemes in network intrusion detection. In *In SIAM International Conference on Data Mining*. 2003

46. Jones AK, Sielken RS: Computer System Intrusion Detection: A Survey. Computer Science Department, University of Virginia; 2000.

47. Baiju S: How to Choose Intrusion Detection Solution. *SANS Institute* 2001.

48. Fan W, Miller M, Stolfo SJ, Lee W, Chan PK: Using artificial anomalies to detect unknown and known network intrusions. *Knowledge and Information Systems* 2004, 6:507-527.

49. Wang Y: A hybrid intrusion detection system. Iowa State University, computer science; 2004.

50. Proctor PE: *The Practical Intrusion Detection Handbook.* Pearson Education; 2000.

51. Amoroso EG: *Intrusion Detection.* Intrusion.Net Books; 1999.

52. Lu W: An Unsupervised Anomaly Detection Framework for Multiple-connection Based Network Intrusions. University of Victoria, Department of Electrical and Computer Engineering; 2005.

53. Mitchell T: *Machine Learning.* McGraw-Hill; 1997.

54. Smith R: Correlating Intrusion Alerts with Unsupervised Learning. University of Ottawa, School of Information Technology and Engineering; 2006.

55. Tjaden B, Welch L, Ostermann S, Chelberg D, Balupari R, Bykova M, Delaney M, Mitchell A, Li S, Lissitsyn D, Tong L: INBOUNDS: The Integrated Network Based Ohio University Network Detective Service. In *In* 4*th World Multiconference on Systemics, Cybernetics, and Informatics (SCI'*2000)*; Orlando, Florida*. 2000

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

57. H.Witten I, Frank E: *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* San Francisco: CA: Morgan Kaufmann Publishers; 2000.

58. Bonifacio JM, Cansian AM, Carvalho ACPLFd, Moreira ES: Neural networks applied in intrusion detection systems. In *in Proceedings of the*

1998 *IEEE Internal Joint Conference on Neural Networks*; *May*. 1998: 205-210.

59. Ghosh AK, Wanken J, Charron F: Detecting anomalous and unknown intrusions against programs. In *in Proceedings of IEEE* 14*th Annual Computer Security Applications Conference*. 1998: 259-267.
60. Lichodzijewski P, Zincir-Heywood AN, Heywood MI: Host-based intrusion detection using self-organizing maps. In *in Proceedings of the* 2002 *International Joint Conference on Neural Networks*; *May; Honolulu, HI*. 2002: 1714-1719.
61. Sung AH, Mukkamala S: Identifying important features for intrusion detection using support vector machines and neural networks. In *in Proceedings of the* 2003 *Symposium on Applications and the Internet (SAINT'*03). 2003: 209-216.
62. Frawley WJ, Piatetsky-Shapiro G, Matheus CJ: Knowledge discovery in databases: an overview. In *AI Magazine*, vol. 13. pp. 57-70; 1992:57-70.
63. Gunetti D, Ruffo G: Intrusion detection through behavioral data. In *in Proceedings of Intelligent Data Analysis (IDA'*99); *August; Amsterdam, Netherlands*. 1999
64. Chimphlee W, Sap MNM, Abdullah AH, Chimphlee S, Srinoy S: An Integrated Model of Intrusion Detection Based on fuzzy clustering and rule learning for identify attacks classes. *International Journal of Computer Science and Network Security (IJCSNS)* 2005, 5:82.
65. Zhang Z: Statistical anomaly denial of service and reconnaissance intrusion detection. New Jersey Institute of Technology, Department of Electrical and Computer Engineering; 2004.
66. NFR Security
67. Beale J, Foster JC, Posluns J, Caswell B: *Snort* 2.0 *Intrusion Detection* Syngress Publishing, Inc.; 2003.
68. Paxson V: *Bro: a system for detection network intruders in real-time.* 1999.
69. Computer Immune Systems, http://www.cs.unm.edu/~immsec
70. Abraham A, Grosan C, Chen Y: Cyber Security and the Evolution of Intrusion Detection Systems. *Journal of Educational Technology, Special Issue in Knowledge Management* 2005.

71. Chimphlee W, Sap MNM, Abdullah AH, Chimphlee S, Srinoy S: To Identify Suspicious Activity in Anomaly Detection Based On Soft Computing. In *24th IASTED International Multi-Conference on APPLIED INFORMATICS*. pp. 359-364. Innsbruck, Austria: IASTED; 2005:359-364.

72. Chimphlee W, Sap MNM, Abdullah AH, Srinoy S, Chimphlee S: Network Intrusion Detection Based on Fuzzy Rough Clustering Methods. In *Joint 3rd International on Soft Computing and Intelligent Systems and 7th International Symposium on advanced Intelligent Systems*. Tokyo, Japan; 2006.

73. Zadeh LA: Role of soft computing and fuzzy logic in the conception, design and development of information/intelligent systems. *Lecture Notes in Computer Science* 1998, 695:1-9.

74. H.A. H: Security is fuzzy! applying the fuzzy logic paradigm to the multipolicy paradigm. In *Proceddings of the 1992-93 workshop on New Security Paradigms*; *August; Little Compton, RI, USA*. 1993: 175-184.

75. El-Semary A, Edmonds J, Gonzalez J, Papa M: A framework for hybrid fuzzy logic intrusion detection systems. In *The 14th IEEE International Conference on Fuzzy Systems*; *May; Tulsa University*. 2005: 325-330.

76. Orchard R: *FuzzyCLIPS version 6.04 user's guide.* National Research Council Canada; 1995.

77. Freedman D, Pisani R, Purves R: *Statistics.* W. W. Norton & Company; 1997.

78. Petrovskiy M: Outlier Detection Algorithms in Data Mining Systems. *Programming and Computer Software* 2003, 29:228-237.

79. Han J, Kamber M: *Data Mining: Concepts and Techniques.* Morgan Kaufmann; 2000.

80. Oh SH, Lee WS: Optimized Clustering for Anomaly Intrusion Detection. In *PAKDD 2003, LNAI 2637*. Edited by K.-Y. Whang JJ, K. Shim, J. Srivatana. Springer-Verlag Berlin Heidelberg; 2003: 576-581.

81. Carpenter GA, Grossberg S: A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing* 1987, 37:54-115.

82. Amini M, Jalili R: Network-Based Intrusion Detection Using Unsupervised Adaptive Resonance Theory (ART). In *The 4th Conference on Engineering of Intelligent Systems (EIS* 2004)*; Madeira, Portugal*. 2004

83. Carpenter GA: Neural network models for pattern recognition and associative memory. *Neural Networks* 1989, 2:243-257.

84. Rao MA, Srinivas J: *Neural networks algorithms and applications.* Alpha Science International Ltd.; 2003.

85. Tsoukalas LH, Uhig RE: *Fuzzy and neural approaches in engineering.* A Wiley-Interscience publication; 1996.

86. Kusiak A, Chung Y: GT/ART: Using neural networks to form machine cells. *Manufacturing Review* 1991, 4:293-301.

87. Pacella M, Semeraro Q, Anglani A: Manufacturing quality control by means of a Fuzzy ART network trained on natural process data. *Engineering Applications of Artificial Intelligence* 2004, 17:83-96.

88. Cao Y, Zhu Z, Wang C: Application of a Modified Fuzzy ART Network to User Classification for Internet Content Provider. In *APWeb Workshops* 2006*, LNCS* 3842. Edited by al. HTSe. Springer-Verlag Berlin Heidelberg; 2006: 725-732.

89. Bezdek JC: *Partition structures: A tutorial.* CRC Press, Boca Raton, FL.; 1987.

90. Albayrak S, Amasyalı F: fuzzy c-means clustering on medical diagnostic systems. In *Proceedings of the 12th Turkish Symposium on Artificial Intelligence and Artificial Neural Networks (TAINN'* 2003)*; Turkey*. 2003

91. Bezdek JC, Ehrlich R, Full W: FCM: Fuzzy C-Means Clustering Algorithm. *Computers and Geosciences* 1984, Vol. 10 (2-3):pp. 191–203.

92. Jang J-S, Sun CT, Mizutani E: *Neuro- Fuzzy and Soft Computing.* Prentice-Hall, USA.; 1997.

93. chimphlee W, Abdullah AH, Sap MNM, Chimphlee S, Srinoy S: A Rough-Fuzzy Hybrid Algorithm for Computer Intrusion Detection. *The International Arab Journal of Information Technology* 2007, 4:247-254.

94. Chimphlee W: Fuzzy Rough C-Means framework. 2006.

95. Ohrm A: ROSETTA Technical Reference Manual. Trondheim, Norway: Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU); 2000.

96. Chimphlee W, Sap MNM, Abdullah AH, Chimphlee S, Srinoy S: Anomaly-Based Intrusion Detection Using Fuzzy Rough Clustering. In 2006 *International Conference on Hybrid Information Technology (ICHIT* 2006). Cheju Island, Korea: IEEE, Springer; 2006.

97. Ye N, Li X: A Scalable Clustering Technique for Intrusion Signature Recognition. In *Proceedings of the IEEE Man, Systems and Cybernetics Information Assurance Workshop June* 5-6*,* 2001*; United States Military Academy West Point, New York*. 2001

98. Guan Y, Ghorbani AA, Belacel N: Y-means: a clustering method for intrusion detection. In *In Canadian Conference on Electrical and Computer Engineering*; *May; Montreal, Qubec, Canada*. 2003: 1-4.

99. Bace RG: *Intrusion Detection.* Pearson Education; 1999.

100. Debar H, Dacier M, Wespi A: Towards a taxonomy of intrusion detection systems. *Computer Networks: The International Journal of Computer and Telecommunications Networking* 1999, 31:805-822.

101. Halme LR, Bauer K: AINT misbehaving - a taxonomy of anti-intrusion techniques. In *in Proceedings of the* 18*th National Information System Security Conference; Baltimore, MD, USA*. 1995: 163-172.

102. Lee W: *A data mining framework for constructing features and models for intrusion detection systems.* Columbia University; 1999.

103. Northcutt S: *Intrusion signatures and analysis.* Indianapolis: New Riders; 2001.

104. Fang L, Le-Ping L: Unsupervised Anomaly Detection Based on an Evolutionary Artificial Immune Network. In *Evo Workshops* 2005*, LNCS* 3449. Edited by al. FRe. Springer-Verlag Berlin Heidelberg; 2005: 166-174.

105. Powers ST, He J: A hybrid artificial immune system and Self Organising Map for network intrusion detection. *Information Sciences*, In Press, Corrected Proof.

106. Anderson D, Frivold T, Valdes A: Next-generation Intrusion Detection Expert System (NIDES) A Summary. SRI Computer Science Laboratory Technical Report SRI-CSL-95-07; 1995.

107. Chimphlee W, Abdullah AH, Sap MNM, Chimphlee S: Unsupervised Anomaly Detection with Unlabeled Data Using Clustering. In *International*

conference on information and communication technology (ICCT-UMB 2005). pp. 42. Jakarta, Indonesia; 2005:42.

108. Chimphlee W, Sap MNM, Abdullah AH, Chimphlee S: Anomaly Intrusion Detection Using Fuzzy Clustering Methods. *Journal of Information Technology* 2006, 18:25-31.

109. Wang Q, Megaloikonomou V: A clustering algorithm for intrusion detection. In *proceedings of the SPIE Conference on Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security; Orlando, Florida, USA*. SPIE; 2005: 31-38.

110. Xiang C, Yong PC, Meng LS: Design of multiple-level hybrid classifier for intrusion detection system using Bayesian clustering and decision trees. *Pattern Recognition Letters* 2008, 29:918-924.

111. Lee K, Kim J, Kwon KH, Han Y, Kim S: DDoS attack detection method using cluster analysis. *Expert Systems with Applications* 2008, 34:1659-1665.

112. Lee W, Stolfo S, Chan P, Eskin E, Fan W, Miller M, Hershkop S, Zhang J: Real Time Data Mining-based Intrusion Detection. In *Information Survivability Conference and EXposition II*; *June*. 2001

113. Parras PA, Neumann PG: EMERALD: Event Monitoring Enabling Responses to Anomalous Live Disturbances. In *In Proceedings of the National Information Systems Security Conference*; *October* 1997*; Baltimore, USA*. 1997: 353-365.

114. Ilgun K, Kemmerer RA, Porras PA: State Transition Analysis: A Rule-Based Intrusion Detection Approach. *IEEE Transactions on Software Engineering* 1995, 21.

115. Chimphlee W, Abdullah AH, Sap MNM, Chimphlee S, Srinoy S: Integrating Genetic Algorithms and Fuzzy c-Means for Anomaly Detection. In *IEEE Indicon* 2005 *Conference*. pp. 575-579. Chennai, India; 2005:575-579.

116. Cox E: *Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration.* ELSEVIER; 2005.

117. Marchette DJ: *Computer Intrusion Detection and Network Monitoring: A Statistical Viewpoint.* Springer-Verlag New York, Inc. Secaucus, NJ, USA; 2001.

118. Smaha SE: An Intrusion Detection System. In *In Fourth Aerospace Computer Security Applications Conference*; *December*. Tractor Applied Science Inc., Texas; 1988

119. Wagner D, Soto P: Mimicry attacks on host-based intrusion detection systems. In *Conference on Computer and Communications Security Proceedings of the* 9*th ACM conference on Computer and communications security Washington, DC, USA*. 2002: 255-264.

120. Kurosawa S, Nakayama H, Kato N, Jamalipour A, Nemoto Y: A Self-adaptive Intrusion Detection Method for AODV-based Mobile Ad Hoc Networks. In *Mobile Adhoc and Sensor Systems Conference,* 2005*, IEEE International Conference on*; 7-10 *Nov.* 2005. 2005: 773-780.

121. Lane TD: Machine Learning techniques for the Computer Security of Anomaly Detection. Purdue University, 2000.

122. Ye N: A Markov chain model of temporal behavior for anomaly detection. In *Proceedings of the* 2000 *IEEE Systems, Man, and Cybernetics; Information Assurance and Security Workshop*. 2000

123. Escamilla T: *Intrusion Detection: Network security beyond the firewall.* New York: Wiley; 1998.

124. Lee W, Stolfo SJ, Mok KW: Mining in a data-flow environment: Experience in network intrusion detection. *In Knowledge Discovery and Data Mining* 1999:114-124.

125. Mukkamala S, Janoski G, Sung AH: Intrusion Detection Using Neural Networks and Support Vector Machine. In *Neural Networks,* 2002*, IJCNN* '02 *Proceedings of the* 2002 *International Joint Conference on* 12-17 *May* 2002. 2002: 1702-1707.

126. Wang L, Yu G, Wang G, Wang D: Method of Evolutionary Neural Network-based Intrusion Detection. *Journal NorthEastern University Natural Science* 2002, 23:107-110.

127. Kim W, Oh S-C, Yoon K: Intrusion Detection Based on Feature Transform Using Neural Network. In *ICCS* 2004*, LNCS* 3037. Edited by al. MBe. Springer-Verlag Berlin Heidelberg 2004; 2004: 212-219.

128. Beale R, Jackson T: Neural Computing: An Introduction. 1990.

129. Ilgun K: USTAT: A Real-Time Intrusion Detection System for UNIX. In *Proceedings of the* 1993 *IEEE Symposium on Security and Privacy*. IEEE Symposium on Security and Privacy; 1993: 16.

130. Jonsson E: An integrated framework for security and dependability. In *in Proceedings of the New Security Paradigms Workshop; Charlottesville, VA, USA*. 1998: 22-29.

131. Snort: The Open Source Network Intrusion Detection System

132. Laskov P, Rieck K, Schäfer C, Müller K-R: Visualization of anomaly detection using prediction sensitivity. In *Proceeding of Sicherheit*; *April,* 2005. Edited by Federrath H. GI; 2005: 197-208.

133. Xin J, Dickerson JE, Dickerson JA: Feature Analysis and Visualization for Network Anomaly Intrusion Detection. pp. 1-13: Electrical and Computer Engineering Department, Iowa State University; 2001:1-13.

134. Lippmann R: The 1999 DARPA Off-Line Intrusion Detection Evaluation. *Computer Networks* 2000, 34:579-595.

135. Lippmann R, Haines JW: Analysis and Results of the 1999 DARPA Off-Line Intrusion Detection Evaluation. In *Third International Workshop RAID*. Recent Advances in Intrusion Detection; 2000

136. MIT Lincoln Laboratory. Intrusion Detection Datasets.

137. Chimphlee W, Abdullah AH, Sap MNM, Chimphlee S, Srinoy S: Unsupervised Clustering Methods for Identifying Rare Events in Anomaly Detection. In 6*th International Enformatika Conference (IEC*2005), vol. 8. pp. 253-258. Budapest, Hungary: Transactions on Engineering, Computing and Technology; 2005:253-258.

138. Chimphlee W, Sap MNM, Abdullah AH, Chimphlee S, Srinoy S: Anomaly Detection of Intrusion Based on Integration of Rough Sets and Fuzzy c-means. *Journal of Information Technology* 2005, 17:1-14.

139. Estevez-Tapiador JM, Garcia-Teodoro P, Diaz-Verdejo JE: Anomaly detection methods in wired networks: a survey and taxonomy. *Computer Communications* 2004, 27:1569-1584.

140. Lee W, Xiang D: Information-theoretic measures for anomaly Detection. In *The* 2001 *IEEE Symposium on Security and Privacy*. 2001

141. Wang Y, Wong J, Miner A: Anomaly intrusion detection using one class SVM. In *5th Annual IEEE Information Assurance Workshop*; *June* 2004*; West Point, New York*. 2004: 358-364.

142. Ye N, Chen Q: An Anomaly Detection Technique Based On A Chi-Square Statistic For Detecting Intrusions Into Information Systems *Quality and Reliability Engineering International* 2001, 17:15-24.

143. Chimphlee W, Abdullah AH, Sap MNM, Chimphlee S, Srinoy S: To Misuse and Anomaly Attacks through Induction Analysis and Fuzzy Methods. *WSEAS Transactions on Computers* 2006, 5:49.

144. Patwardhan A, Parker J, Iorga M, Joshi A, Karygiannis T, Yesha Y: Threshold-based intrusion detection in ad hoc networks and secure AODV. *Ad Hoc Networks* 2008, 6:578-599.

145. John GH, Kohavi R, Pfleger K: Irrelevant Features and the Subset Selection Problem. In *In International Conference on Machine Learning*. 1994: 121-129.

146. Dash M, Liu H: Feature selection for classification. *Intelligent Data Analysis* 1997, 1:131-156.

147. Jain A, Zongker D: Feature selection: Evaluation, application, and small performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1997, 19:153-158.

148. Hall MA: Correlation-based feature selection for machine learning. Waikato University, Department of Computer Science; 1999.

149. Zhang M, Yao J: A Rough Sets Based Approach to Feature Selection. In *Proceedings of The 23rd International Conference of NAFIPS*; 27-30 *June* 2004*; Banff, Canada*. 2004: 434-439.

150. Jensen R, Shen Q: Fuzzy-rough data reduction with ant colony optimization. *Fuzzy Sets and Systems* 2005, 149:5-20.

151. Leung K, Leckie C: Unsupervised Anomaly Detection in Network Intrusion Detection Using Clusters. In *In Proc of 28th Australasian Computer Science Conference (ACSC); Newcastle, Australia*. 2005: 333-342.

152. Hung S-S, Shing-Min Liu D: A user-oriented ontology-based approach for network intrusion detection. *Computer Standards & Interfaces* 2008, 30:78-88.

153. Chimphlee W, Abdullah AH, Sap MNM, Chimphlee S, Srinoy S: Hybrid Model for Computer Intrusion Detection. In 4*th WSEAS International Conference on INFORMATION SECURITY, COMMUNICATIONS and COMPUTERS (ISCOCO* 2005). Canary Islands, Spain; 2005.

154. Leu F-Y, Li M-C, Lin J-C, Yang C-T: Detection workload in a dynamic grid-based intrusion detection environment. *Journal of Parallel and Distributed Computing* 2008, 68:427-442.

155. Stolfo S, Fan W, Lee W, Prodromidis A, Chan P: Cost-based modeling for fraud and intrusion detection: results from the JAM project. In *In Proceeding of the* 2000 *DARPA Information Survivability Conference and Exposition (DISCEX* '00); *January* 25-27 2000.*; Hilton Head, South Carolina*. 2000: 1130.

156. Lee W, Miller M, Stolfo S: Toward cost-sensitive modeling for intrusion detection.: Columbia University; 2000.

157. Cakanyildirim M, Yue WT, Ryu YU: The management of intrusion detection: Configuration, inspection, and investment. *European Journal of Operational Research*, In Press, Corrected Proof.

158. Noh S, Jung G, Choi K, Lee C: Compiling network traffic into rules using soft computing methods for the detection of flooding attacks. *Applied Soft Computing* 2008, 8:1200-1210.

159. Sherif JS, Ayers R, Dearmond TG: Intrusion detection: the art and the practice. Part I. *Information Management & Computer Security* 2003, 11:175-186.

160. Jiang, M.F., Tseng, S.S., Su, C.M.: Two-phase clustering process for outliers detection *Computer Statistical Data Analysis* 2001, 36:351-382.

161. Arshad MH, Chan PK: Identifying Outliers via Clustering for Anomaly Detection. Florida Institute of Technology, Department of Computer Sciences; 2003.

162. Jiang S, Song X, Wang H, Han J-J, Li Q-H: A clustering-based method for unsupervised intrusion detections. *Pattern Recognition Letters* 2006, 27:802-810.

163. Ramaswamy S, Rastogi R, Shim K: Efficient Algorithms for Mining Outliers from Large Data Sets In *Proc of the ACM SIGMOD Conference*. 2000: 427-438.

164. Novikov D, Yampolskiy RV, Reznik L: Anomaly Detection Based Intrusion Detection. In *Proceedings of the Third International Conference on Information Technology: New Generations (ITNG'*06). 2006

165. Jolliffe IT: *Discarding Variables in a Principal Component Analysis I. Artificial Data, Applied Statistics,* 21. 1972.

166. Jolliffe IT: *Discarding Variables in a Principal Component Analysis.,II: Real data. Applied Statistics,* 22. 1973.

167. McCabe GP: *Principal variables, Technometrics* 1984.

168. Beale EM, Kendall MG, Mann DW: The Discarding of Variables in Multivariate Analysis, Biometrika. 1967, vol.54:pp. 357–366.

169. Krzanowski WJ: *Selection of Variables to Preserve Multivariate Data Structure,Using Principal Components, Applies Statististics.* 1987.

170. Tanaka Y, Mori Y: Principal component analysis based on a subset of variables: Variable selection and sensitivity analysis. *Journal of Mathematics and Management Sciences* 1997, Vol.17:pp.61-89.

171. Sung AH, Mukkamala S: The Feature Selection and Intrusion Detection Problems. In *Proceedings of the* 9*th Asian Computing Science Conference, Lecture Notes in Computer Science (ASIAN* 2004). Edited by Maher MJ. Springer-Verlag Berlin Heidelberg 2004; 2004: 468-483.

172. Venkatachalam V, Selvan S: An Approach for Reducing the Computational Complexity of LAMSTAR Intrusion Detection System using Principal Component Analysis. *International Journal of Computer Science* 2007, Vol. 2, Number 1:pp. 76-84.

173. Ma P: Log Analysis-Based Intrusion Detection via Unsupervised Learning. *Master thesis, School of Informatics University of Edinburgh* 2003.

174. Chebrolu S, Abraham A, Thomas JP: Hybrid Feature Selection for Modelling Intrusion Detection Systems. In *Neural Information Processing,* 11*th International Conference, ICONIP* 2004; *November; Calcutta, India*. 2004: 1020-1025.

175. Kittler J: *Feature Selection and Extraction.* Academic Press; 1986.

176. Langley P, Sage S: Selection of Relevant Features in Machine Learning. In *In Proceeding of the AAAI Fall Symposium on Relevance*. AAAI Press; 1994: pp. 140-144.

177. Crosbie M, Spafford G: Active defense of a computer system using autonomous agents. In *Technical Report* 95-008*, COAST Group*: Department of Computer Sciences, Purdue University, West Lafayette, Indiana; 1995.

178. Gao J, Cheng H, Tan P-N: A Novel Framework for Incorporating Labeled Examples into Anomaly Detection. In 2006 *Siam Conference on Data Mining*; *April* 20-22*,* 2006*; Bethesda, Maryland, USA* 2006

179. Chimphlee W, Sap MNM, Abdullah AH, Chimphlee S: Uncertainty Measurement for Identify Suspicious Activity in Anomaly Detection. In *International Symposium on Bio-inspired Computing (BIC*05). Johor Bahru, Malaysia; 2005.

180. Chimphlee W, Sap MNM, Abdullah AH, Chimphlee S, Srinoy S: Semi-Supervised Learning to Identify Suspicious Activity for Anomaly Detection. In 3*rd International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS*2005). Singapore; 2005.

181. Chimphlee W, Abdullah AH, Sap MNM: Unsupervised Anomaly Intrusion Detection in Computer Security. In *Post Graduate Annual Research Seminar (PARS'*07) *Universiti Teknologi Malaysia (Best Paper Award)*. 2007

182. Peddabachigari S, Abraham A, Grosan C, Thomas J: Modeling intrusion detection system using hybrid intelligent systems. *Journal of Network and Computer Applications* 2007, 30:114-132.

183. Wang Y: A comparative study of classification algorithms for network intrusion detection Florida Atlantic University, 2004.

184. Al-Subaie M: The Power of Sequential Learning in Anomaly Intrusion Detection. Queen's University, School of Computing; 2006.

185. Martin A, Doddington G, Kamm T, Ordowski M, Przybocki M: The DET Curve in Assessment of Detection Task Performance. In *Proceedings EuroSpeech* 1998: 1895-1898.

186. Chen Q: Computer Intrusion Detection through Noise Cancellation. Arizona State University, 2001.

187. Lazarevic A, Kumar V: Feature Bagging for Outlier Detection. In *Proceeding of the* 11*th ACM SIGKDD international conference on Knowledge Discovery in data mining*; *August* 21-24*; Chicago, Illinois, USA*. 2005: 157-166.

188. Zhang G: Intrusion detection in networks using singular value decomposition. Queen's University, Department of Computing and Information Science; 2000.

189. Kantardzic M: *Data Mining - Concepts, Models, Methods, and Algorithms.* IEEE Press, Wiley-Interscience; 2001.

190. Pan L, Zheng H, Nahavandi S: The application of Rough set and Kohonen network to feature selection for object extraction. In *The* 2*nd International Conference on Machine Learning and Cybernetics*; 2-5 *November* 2003. 2003: 1185-1189.

191. Kohavi R, John GH: Wrappers for Feature Subset Selection. *Artificial Intelligence* 1997, 97:273-324.

192. Chouchoulas A, Shen Q: Rough set-aided keyword reduction for text categorization. *Application Artificial Intelligence* 2001, 15:843-873.

193. Jensen R, Shen Q: rough and fuzzy sets for dimensionality reduction. In *Proceedings of the* 2001 *UK Workshop on Computational Intelligence* 2001: 69-74.

194. Pawlak Z: *Rough sets: Theoretical Aspects of Reasoning About Data.* Kluwer academic publishers; 1992.

195. Shen Q, Chouchoulas A: A modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems. *Engineering Applications of Artificial Intelligence* 2000, 13:263-278.

196. Smithson M: Ignorance and Uncertainty: Emerging Paradigms. *Springer-Verlag, Berlin* 1989.

197. Beyon MJ: Stability of continuous value discretisation: an application within rough set theory. *International Journal of Approximate Reasoning* 2004, 35:29-53.

198. Zadeh LA: Fuzzy Sets. In *Inf Control*, vol. 8. pp. 338-353; 1965:338-353.

199. Jensen R: Combining rough and fuzzy sets for feature selection. University of Edinburgh, School of Informatics; 2005.

200. Jinsong F, Tingjian F: Chinese Character Classification Based on Rough Set and SVM Algorithm. In *MVA* 2000 *IAPR Workshop on Machine Vision Applications*. The University of Tokyo, Japan; 2000.

201. Diamantaras KI, Kung SY: *Principal Component Neural Networks: Theory and Applications.* John Wiley & Sons, Inc., New York, USA; 1996.

202. Wang J, Plataniotis KN, Venetsanopoulos AN: Feature selection for subject indentification in surveillance photos. In 2004 *International Conference on Image Processing (ICIP)*. 2004: 71-74.

203. Kasabov N, Futschik M, Middlemiss M: Knowledge based neural networks for online and off-line modeling and rule extraction in bioinformatics. In; *March*. Edited by Systems AiNIP. 2001

204. Fausett L: *Fundamentals of Neural Networks: Architectures, Algorithms and Applications.* Prentice-Hall, Englewood Cliffs, NJ, pp. 218-287.; 1994.

205. Kumar S: Classification and Detection of Computer Intrusions. Purdue University, USA, 1995.

206. Hoang XD: E-Commerce Security Enhancement and Anomaly Intrusion Detection Using Machine Learning Techniques. RMIT University, School of Computer Science and Information Technology, Science, Engineering, and Technology Portfolio; 2006.

207. Michael CC, Ghosh AK: Simple, state-based approaches to program-based anomaly detection. *ACM Transactions on Information and System Security* 2002, 5:203-237.

208. Petrovic S, Alvarez G, Orfila A, Carbo J: Labelling Clusters in an Intrusion Detection System Using a Combination of Clustering Evaluation Techniques. In *in Proceedings of the* 39*th Hawaii International Conference on System Sciences*; 4-7 *January; Kauai, HI, USA*. 2006

209. Cao Y, Li Y, Liao X: Applying Modified Fuzzy Neural Network to Customer Classification of E-Business. *X Deng and Y Ye (Eds): WINE* 2005*, LNCS* 3828© *Springer-Verlag Berlin Heidelberg* 2005 2005:pp. 356 – 365.

210. Bezdek JC, Tsao EC, Pal NR: Fuzzy Kohonen Clustering Networks. In *Proceedings of the First IEEE Conference on Fuzzy Systems*; *March,* 1992*; San Diego*. 1992

211. Carpenter GA, Grossberg S, Rosen DB: Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance Theory. *Neural Networks* 1991, 4:759-771.

212. Cinque L, Foresti GL, Gumina A, Levialdi S: A modified fuzzy ART for image segmentation. In *In* 11*th International Conference on Image Analysis and Processing (ICIAP'*01). 2001

213. KDD Cup 1999 Intrusion Detection Datasets, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html, Access Year : 2007