Taxonomy of Feature selection in Intrusion Detection System

Vahid Kaviani Jabali[1*], mina rahbari[2], armin kashkouli[3]

[*1]Corresponding Author: Faculty of Engineering, Khorasgan Branch, Islamic Azad University, Isfahan, Iran.

[2]Computer Department, Faculty of Engineering, Islamic Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran.

[3]Computer Department, Faculty of Engineering, Islamic Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran.

Abstract:

Although, using Internet for daily life and business has raised significantly but this popularity has brought enormous amount of risk by network attacks. Intrusion detection techniques is one most interesting research area in network security. Using IDS systems in networks can help to identify abnormal activities or detect attacks patterns to secure internal assets. In this literature, intrusion detection methods have been used by various machine learning approaches. In this article reviews the importance of security countermeasures. It begins with a background review on computer security and the taxonomy of Intrusion Detection and current technique of feature selection and drawing the taxonomy of intrusion detection system. This paper covers details of IDS design and development issues. It is studied for dimensionality reduction to find which means achieved a better accuracy and reduce workload, followed by existing techniques to compare a classifier and classifiers' designs.

Key words

Taxonomy, Feature selection, Detection, System.

1. Introduction

By developing of networks and computers, in the same time, keeping data safe and secure in computers becomes one of most interesting and challenging area in Network and security. In spite of the fact that attackers try to achieve the sensitive and critical data to take advantage of them. Due to many motivations, there are plenty number of news about misusing information and attacking computers across the globe which have done by intruders. However, many studies and investigations have been conducted to increase the safety and security of networks and computers; there is various attack and most of them still new and opened scope for research. Today after passing a half of century from emerging computer to the world and growing a vast varieties of countermeasures and mitigation approaches against hackers but the necessity of developing new method for reducing exposure and penetration is undeniable due to arriving more novel attacks day by day.

The progress of computer technology has affected communication technology. From 1980s, many devices have been invented and developed. The progress in the network technology changes the way of communication and data distribution in the world because many businesses and companies use this technology for trading and marketing their products and contacting their partner and customers properly. Due to the completion and surviving in this generation among all organizations, the importance of safeguard and other countermeasures to stop penetration of intruders to their sensitive or critical information has been raising significantly. To begin with definition in terms of attack, intruder is somebody who can maliciously interrupt, captures, modify, steal or delete important information in the computers and applications by network access or by direct access like run executable code in PC. Attackers use different resources of victim to do the attack. Specifically, they misuse hardware vulnerabilities or software weakness to penetrate the system.

Nowadays security countermeasures such as access control [2] and authentication [3] have been developed to achieve Confidentiality, Integrity and Availability and to block unauthorized intruders from accessing and modifying information. These prevention methods are developed as a front line of defense system. The advantages of the Internet, namely the availability and amount of information, also it is apparent exposure method and the largest threat to the sensitive and critical security. [4] stated that the Intrusion Detection System is second line of defense or detection method against any kind of external threats. The aim of IDS is to identify and preserve computer system from penetrations of intrusions. In fact there are two techniques for detection in IDS systems which are anomaly detection and misuse detection. Different approaches purpose own different technique.

Some examples about intrusion concerns are [1]:

i. Unauthorized modifications in system files or user information.
ii. Illegal access or modification of user files or information.
iii. Unauthorized modifications of system information in network components

For instance: modifications of router tables in an Internet to deny use of the network. Some of the necessary features an intrusion detection system should possess include [1]:

i. Be able to protect them self or be a fault tolerant and run continually with minimum human control. The IDS must recover themselves from system crashes, either accidental or caused by malicious activity.
ii. Be able to work automatically which is preventing an attacker to manipulate the IDS easily. Moreover, the IDS must be able to track any modifications.

Manuscript received June 5, 2017
Manuscript revised June 20, 2017

iii. Enforce IDS with the optimized overhead on the system to avoid interfering with the normal operation of the system.
iv. The IDS have to be adaptable and configurable in order to changes in system and user behavior over time. In terms of accuracy easy to implement the security policies and user behavior of the systems that are being monitored.
v. Able to detect different types of attacks accurately and must not track any legitimate activity as an intrusion or false positives and conversely at the same time, the IDS must not fail to recognize any real attacks (false negatives).

2. Purpose of This Review

Based on the other researches done in Intrusion Detection System area, it is clear that the effectiveness of an IDS model relies on retraining of the reference models and enhancing the recognition of classifiers. Getting better accuracy result for distinguish abnormal behaviors is feature selection which is one of most essential issue in IDS. Feature selection is where a feature subset is selected to represent the data. The importance of feature selection can be studied in two dimensions. In the first aspect, noise separates from raw data. Secondly omit redundant and incoherent features that cause major accuracy loss and time consumption in detection. In the following it is reviewed the importance of security countermeasures. It begins with a background review on computer security and the taxonomy of Intrusion Detection and current technique of feature selection and drawing the taxonomy of intrusion detection system. This chapter covers details of IDS design and development issues. It starts with dimensionality reduction as a mean to achieve a better accuracy and reduce workload, follows by existing techniques to develop a classifier and classifiers' designs.

2.1 Computer Security

Three main elements of quantitative evaluation of networked environment attributes namely are, availability, confidentiality and integrity which were defined by National Institute of Standards and Technology [8].

a) Availability refers to services and information which can be flowed through network timely and reliably and exchanged between source and destination.
b) Confidentiality addresses the information and data that the disclosure and access which is restricted and critical to the authorized owners.
c) Integrity refers to the authenticity of information which is transmitted between source and destination. Integrity also includes non-repudiation information that truly originates from the declared (or expected) source, and the source sends information to the intended destination.

In order to illustrate all above security attributes, in following [figure 1] the significant branches of security was showed.

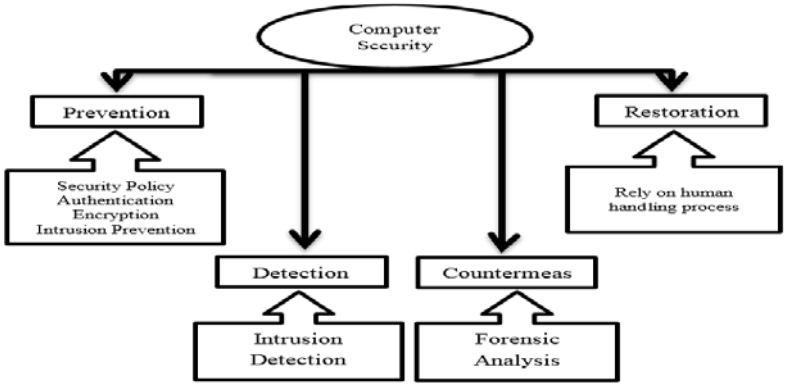

Figure 1: Classification of Computer Security Assurance

2.2 Intrusion Detection System

Intrusion detection systems are known as intrusion alarms or burglar alarms of the computer security field. The goal of this system is to defend a system by using a combination of an alarm and process to prevent intruders and malicious activities whenever the security and sensitive information have been compromised [9].Thus most often a security expert can respond to the alarm and take the appropriate action and mitigate the damage and disclosure for example by calling on the proper external authorities, and so on. It should take it to consideration that Intrusion can be found in different types of intrusion. As instance, a hacker might hijack an account by stealing user's password as result impersonate the identity to a system.

As it called a masquerader and the detection of such intruders is an important problem for the field. Other important issues of intruders are people who are legitimate users of the system but who can easily abuse from their

privileges, and people who are inside network and whether unintentionally or intentionally exploit the web server application or software to do something maliciously, often found on the Internet to attack the system through a network. With this ongoing list with ongoing list of attacks and threats to computer and network, the IDS is still an active area for research and project. Audit data collection agents gather information about a system which is observed by intrusion detection system. The output of data which is stored and processed by IDS can use for further action. Normally security experts begin for further investigation to understand what causes the IDS's alarm. When James Anderson was working on the enhancement of forensic and surveillance abilities of computer system he invented the idea of Intrusion Detection System IDS in1980. Also Dorothy Denning in 1987 came up with a generic seminal framework of an IDS with five statistical models. The abuse of systems involves abnormal activities or known as intrusion behavior that can be detected by IDS. The core of the framework was based on training system where the knowledge was came from statistical understanding based on system audit trails or system resources. It took advantage of statistical properties (e.g., mean and variance) of normal activities to build a statistical to determine whether observed activities deviated significantly from the norm profile. It was a rule-based pattern-matching system. Generic intrusion detection model was known as abstract model or framework proposed by Denning.

2.3 Category of Intrusion Detection Systems

IDS can be classified either based on its monitoring scope or detection techniques as shown in [figure 2 And 3]. The monitoring scope can be further grouped into host-based and network-based. On the other hand, the detection techniques can be either grouped as anomaly or misuse detections.

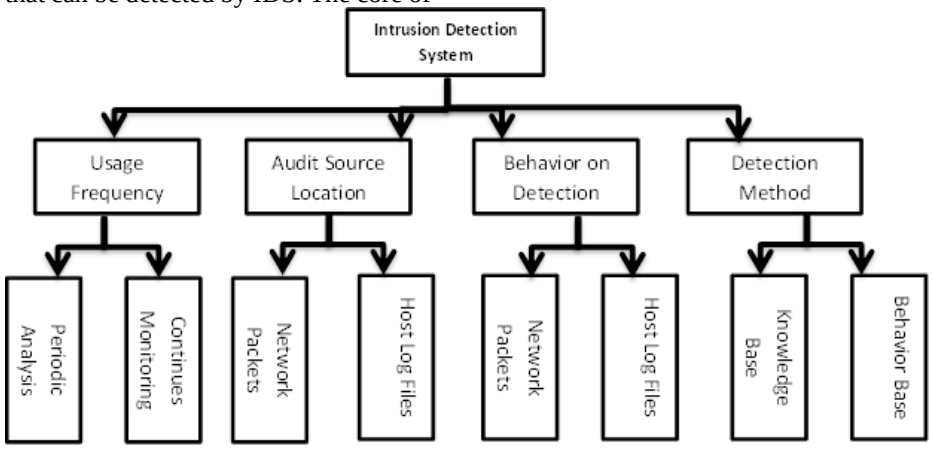

Figure 2: Taxonomy of Intrusion Detection System

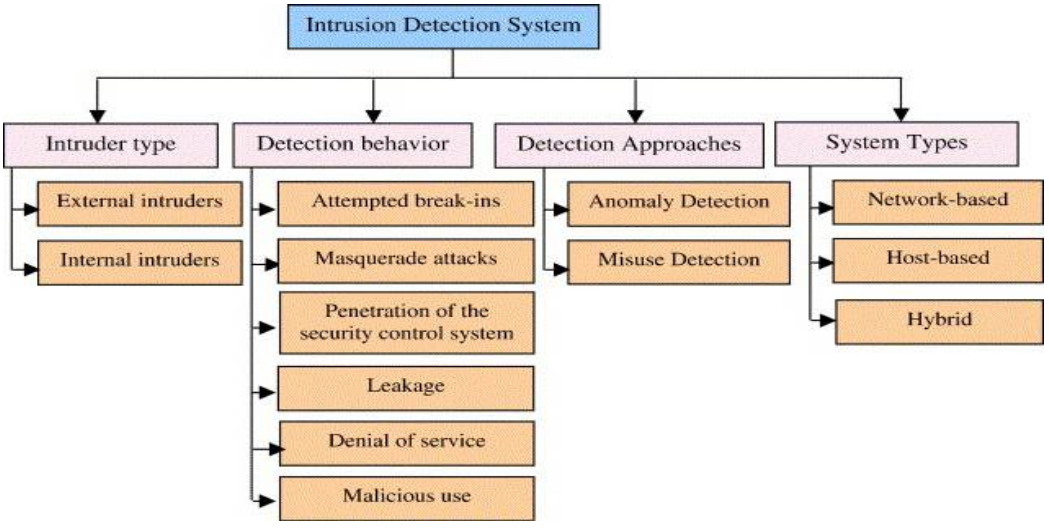

Figure 3: Taxonomy of Intrusion Detection Types

2.3.1 Types of Intrusion Detection Systems by Detection Techniques

a) Host-based IDS

Based on the sources of data, intrusion detection systems can be divided into two major classes, host-based and network-based. In the first kind of systems, the intrusion detection mechanism is installed on the local host/terminal. By examining the status of audit information on system's behavior, the system finds signs of intrusion and can then protect its own local machine. The audit information can be obtained from different sources such as system logs and activities, application logs, and target monitoring.
These logs could be Unix logs, NT/2000/XP logs, firewall logs, router logs, web server logs, and FTP logs. The intrusions can be critical file modifications, segmentation fault errors recorded in logs, crashed services or extensive usage of the processors [10]. From the system point of view, all users are considered as local clients to the target environment.

b) Network-based IDS

In the network base is not just the host-based intrusion detection system to protect its own host machine by examining audit trail, network-based intrusion detection system protects the entire environment of the network by monitoring all the activities from both incoming and outgoing packets of the network. By analyzing the traffic data that goes through the network, the potential and possible intrusions can be identified. In general, the network traffic that needs to be monitored is quite heavy and large even in small networks. With good location for sensors on the network, instead of central sensor in the network, deploying sensors in different locations to achieve better efficiency is more effective.
Network-based IDS monitors any number of hosts on a network by inspecting the audit trails of multiple hosts [11]. Since attempted intrusions can happen via the network, network-based IDS needs to monitor multiple events generated on several hosts to integrate sufficient evidence. Since most of the hosts are networked and attacks can also be launched from remote, this study focuses on network-based IDS. Mainly Anomaly and misuse are two detection methods which are used both in network-based and host-based IDSs.

2.3.2 Type of Intrusion Detection Systems by Monitoring Scope

There are different approaches to detect malicious and invasion activities by intrusion detection system based on the infustructure of system and the type of intrusion detection.

a) Anomaly Detection

Anomaly detection assumes that intrusions will always reflect some deviations from normal patterns. Static and dynamic detection are two approaches for anomaly detection. By controlling and monitoring a portion of system for not being changed is the base of static anomaly detector idea. Normally when hardware is not needed to be checked, it is the base of static detector which only addresses the software portion. The correct functioning of the system relies on code which is the constant portion of data. For example, in kernel of the operating systems, data never changes from critical software to bootstrap. Static anomaly detectors concentrate on integrity if an error has took place or the static portion of the system has altered by an intruder then static portion of system deviates from pervious state [12]. Typically audit records or monitored data traffic network are used in dynamic anomaly detection. Audit records in operating systems have captured for only events that are important not all of them as result audit records will be observed in a sequence. In distributed systems, partial ordering of events is sufficient for detection. In other types, only combined information, as instance combined processor resource used during a time interval which is not directly represented is stored. In this case, in order to distinguish normal resource consumption from abnormal resource consumption, the thresholds are defined.
In this kind of detection, the system track and monitor the behavior of computer users to determine which one can be normal or abnormal and also behavior-based intrusion detection also known as anomaly detection models normal or expected behavior of computer users. If the pattern of behavior or data deviates from the learned normal behavior, an alarm is raised. Advantage of this approach is that novel and unseen attacks can be detected easily due to mismatching. Since it assumes any deviation from normal patterns is regarded as abnormal behaviors or activities, another benefit of this technique is that it is not required to continuously keep up with hackers' techniques [13]. Also, it is less dependent on target operating environments compared with the misuse detection technique. The main problem of this technique is it might have a high number of false alarms due to any deviations from the learned behaviors. Since not every deviation is a real intrusion, the security administrator may ignore some of these false alarms and ignore the real anomalous activities.
The drawbacks of this method are [14]:

i) During the profile construction and training phases, there is high possibility that some users' activities skipped if the users not properly monitored also known as false alarm rate.
ii) The database of normal behavior profile requires a constant update; that's why it is associated with more

false alarm rates because this needs to close the system for update frequently.

iii) During training phase learning system new behaviors can cause false alarm because system detects anomalies.

b) Misuse Detection

Misuse detection is based on the knowledge of system vulnerabilities and known attack patterns [15]. Misuse detection is concerned with finding intruders who are attempting to break into a system by exploiting some known vulnerability. Ideally, a system security administrator should be aware of all the known vulnerabilities and eliminate them. The term intrusion scenario is used as a description of a known type of intrusion; it is a series of events that would outcome in an intrusion without some outside preventive intervention. An intrusion detection system continually compares recent activities to find known intrusion scenarios. It is necessary to ensure that one or more attackers are not attempting to exploit known vulnerabilities. To perform this, each intrusion scenario must be described or modeled. The main difference between the misuse methods is in how they distinguish or model the behavior that constitutes an intrusion. The original misuse detection systems used rules to describe events indicative of intrusive actions that a security administrator looked for within the system. Large numbers of rules can be difficult to interpret. If-then rules are not grouped by intrusion scenarios as result making modifications to the rule set can be difficult whereas the affected rules are scattered across the rule set. To resolve these problems, new rule organized methods include model based rule organization and state transition diagrams. Misuse detection systems look for events that is matched in the rules. It might be possible to fit an intrusions scenario. The events can be used for later investigation by audit records and be monitored live by monitoring system calls [16].

Some disadvantages [16]:

i) One of difficulties of this technique is to keep the data base updated from recent attack signatures.

ii) The IDSs with method are inherently unable to detect new attacks because a constant updating of the attack signature database for correlation is necessary.

iii) Maintenance of an IDS is crucially based on with patching and analyzing of security exploits, which is a time-consuming process.

iv) The attack knowledge is operating environment–dependent, so it must be configured in strict compliance with the operating system (version, platform, applications used etc.).

v) Another disadvantage of this kind of attack is struggling with internal attacks. Typically, abuse of legitimate user privileges cannot be tracked or sensed by the system as a malicious activity.

In order to choice a proper IDS, based on the requirement and applicability of users and various of networks, many factors can affect the decision to which technique and ability is more efficient and proper to the owner [17]. Generally, there are four possible states of detection as illustrated in [figure 4].

I. Intrusive but not anomalous: These are false negatives. An intrusion detection system fails to detect this type of activity as the activity is not anomalous. These are called false negatives because the intrusion detection system falsely reports the absence of intrusions.

II. Not intrusive but anomalous: These are false positives. In other words, the activity is not intrusive, but because it is anomalous, an intrusion detection system reports it as intrusive. These are called false positives because an intrusion detection system falsely reports intrusions.

III. Not intrusive and not anomalous: These are true negatives; the activity is not intrusive and is not reported as intrusive.

IV. Intrusive and anomalous: These are true positives; the activity is intrusive and is reported as such.

There are two types of intrusion detection systems that employ one or both of the intrusion detection methods outlined above. Host-based systems base their decisions on information obtained from a single host (usually audit trails), while network-based intrusion detection systems obtain data by monitoring the traffic in the network to which the hosts are connected [53].

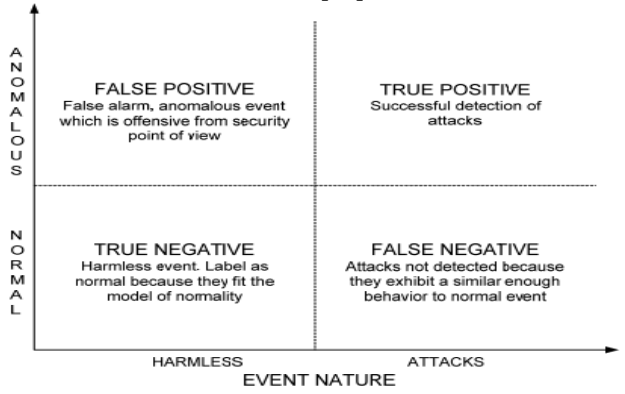

Figure 4: Four Possible Outcomes of Detection [53]

2.4 Network Traffic

A typical network in order to transmit the information via network is using several network packets. A packet includes two parts, header and payload. The information placed in the header are; date, start time, duration, service, source and destination ports, source and destination IP addresses. Meanwhile, payload consists of data or information intended for the recipient.

Some attacks or intrusions can misuse the irregularity in the header, also some others are exploit in the payload which make it difficult to be tracked. In typical and legitimate traffic, the services and communications have explicit model and pattern. According to [14], an intrusive and hostile traffic can be differentiated from these usual communications across the network because it exhibits suspicious characteristics. For example, DoS and Probe attacks usually generate network activity which can be potentially monitored and tracked by an IDS because they cause an significant activity in network in terms of total number of connections initiated during a given interval. Meanwhile land attacks (classified under Probe) exhibit irregularity by using bogus source address where both destination and source addresses are equal [14].

In the case of R2L attacks, they usually explore vulnerabilities on the authentication mechanisms of certain applications, such as ftp, telnet, http and smtp [20].An example is by sending invalid input which causes a buffer overflow or an input validation error in the code running the service. Sometimes, an attacker sends one (or a few) carefully crafted packets including shell-code which is executed at the remote machine to elevate his privileges. Usually, few packets are sufficient to successfully gain access to remote machine. [17] Noted that it is nearly impossible for systems that use traffic models to detect such anomalies.

2.5 Categories of Intrusions

The idea of abnormal behavior detection in computer users was introduced by Anderson in 1980 [22]. In this paper, it is studied and classified threat as an intentional unauthorized access to information, data modification, or exhibit a system that is useless or unreliable. After that many researches have been proposed a various different schemes to group attacks into the categories. For example, in 1987 Denning [23] classified abnormal patterns of system usage into eight categories. It is grouped in attempts for breaching access control, impersonating identity, successful break-in, exposing by legitimate user, modify by legitimate user, denial of service, internal attack, Trojan horse and virus.

In 1988 Smaha [24] categorized intrusion in six major types in following: exposure, denial of service, exploit, penetration, impersonation and break-in. In 1997 Dekker [25] addressed network security incident as threats which can violate security policy. It is categorized them into the compromising account, root exposure, sniffing packets, denial of service, probing, attacking internet infrastructure and injecting malicious code. In 1999, “DARPA Intrusion Detection Evaluation Data Set” which is created at Lincoln Laboratory in MIT called KDD99 dataset [26]. This dataset contains thirty-nine types of attacks that are defined into four major categories. Those are Probe, U2R, R2L, DoS attacks. One of the types is DoS attacks. In this category of attacks, attackers try to interrupt a network orhost resources in order to stop access from legitimate users from the computer service. The victim can be mail server, web server, DNS server, and so on. In the DARPA KDD99 category, there are variety common forms of DoS attacks that are included. Such as, over 70% portions of the attacks in the DoS category are smurf. The attackers take advantage of vulnerability of ICMP (Internet Control Message Protocol), and that attack can cause a crash of target system. Sending a large number of ICMP “echo request” packets to the victim’s address potentially can help attackers to compromised and spoofed source address of the intended target system. Any machine in the subnets of network will reply by sending ICMP “echo reply” packets back to the target. If the number of the packets is more than the ability of the system for handling responses, the result is the spoofed system will crash and no longer be able to service to the real ICMP requests. Another common approach to crash a system is neptune attacks. Over 25% portion of DoS attacks are neptune in the data set. It is a flood attack which is known as is SYN (Synchronize) that exists in TCP/IP (Transmission Control Protocol/Internet Protocol). Attacker attempts to send a large number of requests to establish connections but in this way never responds to target systems. While the attacker still sends request new connections, because requests are faster than the system’s replies thus the system cannot carry out all requests and the legitimate requests can never be satisfied. In the meantime, the system may dace with buffer over flow and run out of memory and even crash [24].

The second type of attacks is Probe attacks. By using applications to automatically search and scan a large number network IP addresses then the attacker can explore what are the vulnerabilities of the targeted computers. when any vulnerability is found in system, the attacker can breach to the system as a result obtain the access and start to collect information without authorization. The DARPA KDD99 data set introduce six modes of scanning from Probe attack category. those are ipsweep, mscan, nmap, portsweep, saint, and satan [25].

The third category of attacks is U2R attacks. In order to get root access of the system, the attacker masquerade or pretends as a legitimate user without authorization hence exploits the system’s vulnerabilities. The DARPA KDD99 data set consists of eight different types of U2R attacks. Among them, buffer_overflow attack is the most ordinary one that starts with by feeding many data into a fix length buffer. When the volume of data exceeds the size of the buffer that can hold, the extra information will overflow into other buffers and overwrite the instructions that supposed to be executed. As result system may crash or the system executes the attacker’s even if it is part of the original programs [26].

The forth category of attacks is R2L attacks. This type of attacks is an unauthorized access attack which can gain an access in network as a user of local computer therefore exploits the computer's vulnerabilities. Totally fifteen types of R2L attacks are included in the DARPA KDD99 data set. For instance, the ftp_write attack is one means that the attacker produces rhost file to make anonymous (File Transfer Protocol) FTP writable directory and finally gain local login to the system. The guess_passwd is another way that the attacker attempts to obtain access to the account of user by guessing the possible passwords repeatedly. Any service which requires password to get access potentially becomes a target, for example, rlogin, ssh, ftp, telnet, pop, and imap in [Table 1].

Table 1: Four Classes of Attacks [24]

DoS	R2L	U2R	Probe
apache2, back, land, mailbomb, netpune, pod, processtable, smurf, teardrop, udpstorm.	ftp_write, guess_passwd, imap, multihop, named, phf, sendmail, snmpgetattack, snmpguess, spy, warezclient, warezmaster, worm, xlock, xsnoop	buffer_overflow, httptunnel, loadmodule, perl, ps, rootkit, sqlattack, xterm.	ipsweep, mscan, nmap, portsweep, saint, satan.

2.6 Techniques Used in Anomaly Detection

In this subsection, we review a number of different architectures and methods that have been proposed for anomaly detection. These include statistical anomaly detection, data-mining based methods, and machine learning based techniques.

The IDS technology has progressed in parallel to the complexity of the computer security issues. Initially, the research focus during 1980's to 1990's was to automate the detection of intrusions. Then, it progressed by including intelligent features in IDS where Artificial Intelligent techniques were used. As intrusions become more complex, better performance IDS based on hybrid approach were proposed beginning of 2000. The search for better IDS continues until to date where holistic and adaptability is needed as the new attacks are evolved and normal traffic patterns are changing. As the computer network becoming more complex and sophisticated, the network traffics become more vulnerable to attacks.

2.6.1 Statistical Anomaly Detection

In statistical methods for anomaly detection, the system observes the activity of subjects and generates profiles to represent their behavior. The profile typically includes such measures as activity intensity measure, audit record distribution measure, categorical measures (the distribution of an activity over categories) and ordinal measure (such as CPU usage). Typically, two profiles are maintained for each subject: the current profile and the stored profile. As the system/network events (viz. audit log records, incoming packets, etc.) are processed, the intrusion detection system updates the current profile and periodically calculates an anomaly score (indicating the degree of irregularity for the specific event) by comparing the current profile with the stored profile using a function of abnormality of all measures within the profile. If the anomaly score is higher than a certain threshold, the intrusion detection system generates an alert. Statistical approaches to anomaly detection have a number of advantages. Firstly, these systems, like most anomaly detection systems, do not require prior knowledge of security flaws and/or the attacks themselves. As a result, such systems have the capability of detecting ''zero day'' or the very latest attacks. In addition, statistical approaches can provide accurate notification of malicious activities that typically occur over extended periods of time and are good indicators of impending denial-of-service (DoS) attacks. A very common example of such an activity is a portscan. Typically, the distribution of portscans is highly anomalous in comparison to the usual trace distribution. This is particularly true when a packet has unusual features (e.g., a crafted packet). With this in mind, even portscans that are distributed over a lengthy time frame will be recorded because they will be inherently anomalous.

2.6.2 Hybrid Systems

The most recent development in outlier detection technology is hybrid systems. The hybrid systems discussed in this section incorporate algorithms from at least two of the preceding sections (statistical, neural or machine learning methods). Hybridization is used variously to overcome deficiencies with one particular classification algorithm, to exploit the advantages of multiple approaches while overcoming their weaknesses or using a meta-classifier to reconcile the outputs from multiple classifiers to handle all situations. We describe approaches where an additional algorithm is incorporated to overcome weaknesses with the primary algorithm next.

It has been suggested in the literature [27,28,29] that the monitoring capability of current intrusion detection systems can be improved by taking a hybrid approach that consists of both anomaly as well as signature detection strategies. In such a hybrid system, the anomaly detection technique aids in the detection of new or unknown attacks while the signature detection technique detects known attacks. The signature detection technique will also be able to detect attacks launched by a patient attacker who attempts to change the behavior patterns with the objective of retraining the anomaly detection module so that it will accept attack behavior as normal. Tombini et al. [30] used an approach wherein the anomaly detection technique is used to produce a list of suspicious items. The classifier module which uses a signature detection technique then classified the suspicious items into false alarms, attacks, and unknown attacks. This approach works on the premise that the anomaly detection component would have a high detection rate, since missed intrusions cannot be detected by the follow-up signature detection component. In addition, it also assumed that the signature detection component will be able to identify false alarms. While the hybrid system can still miss certain types of attacks, its reduced false alarm rate increases the likelihood of examining most of the alerts.

2.7 Machine Learning Techniques

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data. Machine Lerner can capture characteristics of examples to take advantage of them to get unknown underlying possible distributions.

2.7.1 Pattern Classification

Pattern recognition is the action to take raw data and activity on data category [31]. The methods of supervised and unsupervised learning can be used to solve different pattern recognition problems. In supervised learning, it is based on using the training data to create a function, in which each of the training data contains a pair of the input vector and output (i.e. the class label).The learning (training) task is to compute the approximate distance between the input-output examples to create a classifier model. When the model is created, it can classify unknown examples into a learned class labels.

2.7.2 Neural Networks

Neural network approaches are generally non-parametric and model based, they generalize well to unseen patterns and are capable of learning complex class boundaries. After training the neural network forms a classifier. However, the entire data set has to be traversed numerous times to allow the network to settle and model the data correctly. They also require both training and testing to fine tune the network and determine threshold settings before they are ready for the classification of new data. Many neural networks are susceptible to the curse of dimensionality though less so than the statistical techniques. The neural networks attempt to fit a surface over the data and there must be sufficient data density to discern the surface. Most neural networks automatically reduce the input features to focus on the key attributes. But nevertheless, they still benefit from feature selection or lower dimensionality data projections.

The neural network is information processing units which to mimic the neurons of human brain [32]. Multilayer perceptron (MLP) is the widely used neural network architecture in many pattern recognition problems. A MLP network consists of an input layer including a set of sensory nodes as input nodes, one or more hidden layers of computation nodes, and an output layer of computation nodes. Each interconnection has associated with it a scalar weight which is adjusted during the training phase. In addition, the back propagation learning algorithm is usually used to train a MLP, which are also called as back propagation neural networks. First of all, random weights are given at the beginning of training. Then, the algorithm performs weights tuning to define whatever hidden unit representation is most effective at minimizing the error of misclassification.

2.7.3 K-Nearest Neighbor

K-nearest neighbor (k-NN) is one of the most simple and traditional nonparametric technique to classify samples [33][34]. It computes the approximate distances between different points on the input vectors, and then as signs the unlabeled point to the class of its K-nearest neighbors. In the process of create k-NN classifier, k is an important parameter and different k values will cause different performances. If k is considerably huge, the neighbors which used for prediction will make large classification time and influence the accuracy of prediction.

k-NN is called instance based learning, and it is different from the inductive learning approach [34]. Thus, it does not contain the model training stage, but only searches the examples of input vectors and classifies new instances. Therefore, k-NN‘‘on-line” trains the examples and finds out k-nearest neighbor of the new instance.

2.7.4 Support Vector Machines

Support vector machines (SVM) is proposed by [35]. SVM first maps the input vector into a higher dimensional feature space and then obtain the optimal separating hyper-plane in the higher dimensional feature space. Moreover, a decision boundary, i.e. the separating hyper-plane, is determined by

support vectors rather than the whole training samples and thus is extremely robust to outliers.
In particular, an SVM classifier is designed for binary classification. That is, to separate a set of training vectors which belong to two different classes. Note that the support vectors are the training samples close to a decision boundary. The SVM also provides a user specified parameter called penalty factor. It allows users to make a tradeoff between the number of misclassified samples and the width of a decision boundary.

2.7.5 Self-Organizing Maps

Self-organizing map (SOM) [37] is trained by an unsupervised competitive learning algorithm, a process of self-organization. The aim of SOM is to reduce the dimension of data visualization. That is, SOM projects and clusters high-dimensional input vectors onto a low-dimensional visualized map, usually 2 for visualization. It usually consists of an input layer and the Kohonen layer which is designed as two-dimensional arrangement of neurons that maps n dimensional input to two dimensions. Kohonen's SOM associates each of the input vectors to a representative output. The network finds the node closest to each training case and moves the winning node, which is the closest neuron (i.e. the neuron with minimum distance) to the training case. That is, SOM maps similar input vectors onto the same or similar output units on such a two-dimensional map. Therefore, output units will self-organize to an ordered map and those output units with similar weights are also placed nearby after training.

2.7.6 Decision Trees

A decision tree classifies a sample through a sequence of decisions, in which the current decision helps to make the subsequent decision. Such a sequence of decisions is represented in a tree structure. The classification of a sample proceeds from the rootnode to a suitable end leaf node, where each end leaf node represents a classification category. The attributes of the samples are assigned to each node, and the value of each branch is corresponding to the attributes [38].
A well-known program for constructing decision trees is CART (Classification and Regressing Tree) [39]. A decision tree with a range of discrete (symbolic) class labels is called a classification tree, whereas a decision tree with a range of continuous (numeric) values is called a regression tree.

2.7.7 Naive Bayes Networks

There are many cases where we know the statistical dependencies or the causal relationships between system variables. However, it might be difficult to precisely express the probabilistic relationships among these variables. In other words, the prior knowledge about the system is simply that some variable might influence others. To exploit this structural relationship or casual dependencies between the random variables of a problem, one can use a probabilistic graph model called Naïve Baysian Networks (NB).
The model provides an answer to questions like ‘‘What is the probability that it is a certain type of attack, given some observed system events?" by using conditional probability formula. The structure of a NB is typically represented by a directed acyclicgraph (DAG), where each node represents one of system variables and each link encodes the influence of one node upon another [40]. Thus, if there is a link from node A to node B, A directly influences B.

2.7.8 Genetic Algorithms

Genetic algorithms (GA) use the computer to implement the natural selection and evolution [41]. This concept comes from the ‘‘adaptive survival in natural organisms". The algorithm can generate a large population of candidate programs. Some type of fitness measure to evaluate the performance of each individual in a population is used. A large number of iterations is then performed that low performing programs are replaced by genetic recombinations of high performing programs. That is, a program with a low fitness measure is deleted and does not survive for the next computer iteration.

2.7.9 Fuzzy Logic

Fuzzy logic (or fuzzy set theory) is based on the concept of the fuzzy phenomenon to occur frequently in real world. Fuzzy set theory considers the set membership values for reasoning and the values range between 0 and 1. That is, in fuzzy logic the degree of truth of a statement can range between 0 and 1 and it is not constrained to the two truth values (i.e. true, false). For examples, ‘‘rain" is a commonly natural phenomenon, and it may have very fierce change. Raining may be able to convert the circumstances from slight to violent [42].

2.7.10 Hybrid Classifiers

In the development of an IDS, the ultimate goal is to achieve the best possible accuracy for the task at hand. This objective naturally leads to the design of hybrid approaches for the problem to be solved. The idea behind a hybrid classifier is to combine several machine learning techniques so that the system performance can be significantly improved. More specifically, a hybrid approach typically consists of two functional components. The first one takes raw data as input and generates intermediate results. The second one will then take the intermediate results as the input and produce the final results [43].

In particular, hybrid classifiers can be based on cascading different classifiers, such as neuro-fuzzy techniques. On the other hand, hybrid classifiers can use some clustering-based approach to preprocess the input samples in order to eliminate unrepresentative training examples from each class. Then, the clustering results are used as training examples for classifier design. Therefore, the first level of hybrid classifiers can be based on either supervised or unsupervised learning techniques.
Finally, hybrid classifiers can also be based on the integration of two different techniques in which the first one aims at optimizing the learning performance (i.e. parameter tuning) of the second model for prediction.

2.7.11 Ensemble Classifiers

Ensemble classifiers were proposed to improve the classification performance of a single classifier [44]. The term ‘‘ensemble” refers to the combination of multiple weak learning algorithms or weak learners. The weak learners are trained on different training samples so that the overall performance can be effectively improved.
Among the strategies for combining weak learners, the ‘‘majority vote” is arguably the most commonly used one in the literature. Other combination methods, such as boosting and bagging, are based on training data resampling and then taking a majority vote of the resulting weak learners in [Table].

Table 2: Fundamentals of the A-NIDS techniques

Technique: basics- Pros, - Cons		Subtypes
A)Statistical-based: stochastic behavior	• Prior knowledge about normal activity not required. Accurate notification of malicious activities. • Susceptible to be trained by attackers. • Difficult setting for parameters and metrics. • Unrealistic quasi-stationary process assumption.	• A.1) Univariate models (independent Gaussian random variables) • A.2) Multivariate models (correlations among several metrics) • A.3) Time series (interval timers, counters and some • other time-related metrics)
B)Knowledge-based: availability of prior knowledge/data	• Robustness. Flexibility and scalability. • Difficult and time-consuming availability for high-quality knowledge/data.	• B.1) Finite state machines (states and transitions) • B.2) Description languages (N-grams, UML,) • B.3) Expert systems (rules-based classification) • C.1) Bayesian networks (probabilistic relationships among variables)
C)Machine learning-based: categorization of patterns	• Flexibility and adaptability. • Capture of interdependencies. • High dependency on the assumption about the behavior accepted for the system. • High resource consuming.	• C.2) Markov models (stochastic Markov theory) • C.3) Neural networks (human brain foundations) • C.4) Fuzzy logic (approximation and uncertainty) • C.5) Genetic algorithms (evolutionary biology inspired) • C.6) Clustering and outlier detection

2.7.12 Single Classifiers

The intrusion detection problem can be approached by using one single machine learning algorithm. In literature, machine learning techniques (e.g. k-nearest neighbor, support vector machines, artificial neural network, decision trees, self-organizing maps, etc.) have been used to solve these problems.

2.7.13 Feature Selection

IDS can be a combination of software and hardware. Most of IDSs perform their task in real time. However, there are also IDSs that do not work in real time, because of performing analysis for forensic audits. There are some IDSs that react to intrusions in real time manner. This reaction usually imposes to reducing the loss and damage

by terminating a network connection and other approaches. For those IDSs need to do auditing data, it is difficult even by using computer's power because detecting suspicious behavior for incoming data even in small networks are complicated process [45].

Audit data grabs various features of the connections such as the source and destination bytes of a TCP connection, the audit data would show the source and destination bytes of a TCP connection, or the number of unsuccessful login or duration of a connection. There are some complex feature relationships, which are not easy for humans to find. As consequence, The IDS must reduce the amount of data to be processed. Some data may not be useful to the IDS and thus can be eliminated before processing. In complex classification domains, features may contain false correlations, which cause the process of detection intrusions with some struggle. Also some features may be redundant so the extra features can increase computation time, and can have an impact on the accuracy of the IDS. Feature selection improves classification by searching for the subset of features, which best classifies the training data [46].

In complex classification domains, some data may hinder the classification process. Features may contain false correlations, which hinder the process of detecting intrusions. Further, some features may be redundant since the information they add is contained in other features. Extra features can increase computation time, and can impact the accuracy of IDS. Feature selection improves classification by searching for the subset of features, which best classifies the training data. The features under consideration depend on the type of IDS, for example, network-based IDS will analyze network related information such as packet destination IP address, logged in time of a user, type of protocol, duration of connection etc. It is not known which of these features are redundant or irrelevant for IDS and which ones are relevant or essential for IDS. There does not exist any model or function that captures the relationship between different features or between the different at tacks and features. If such a model did exist, the intrusion detection process would be simple and straightforward. In this paper we use data mining techniques for feather selection. The subset of selected features is then used to detect intrusions.

In general, two different approaches for feature selection can be distinguished: filter and wrapper approaches. Using a filter approach, the selection of appropriate features is based on distance and information measures in the feature space and is carried out completely independent from the classifier deployed. In contrast, with a wrapper approach the selection of features is based on the classifiers accuracy. Although a filler approach might be faster, we apply a wrapper approach as better classification results are generally achievable.

The wrapper approach is addressed by means of an evolutionary algorithm (EA) for feature selection and structure optimization for radial basis function (RBF) networks. RBF networks are chosen because of their excellent classification capability and the fact that different hard-computing techniques (e.g., conventional clustering techniques or singular value decomposition) can be applied for network training. The TCPOP data of the 1998 Defense Advanced Research Projects Agency (DARPA) IDS evaluation [47], [48] is used for our experiments and seven attacks (Back, Dict, Guest, Ipsweep, Nmap, Portsweep, and Warezclient) are detected.

2.7.14 Wrapper Approach

Wrapper methods apply the unsupervised-learning algorithm to each candidate feature subset and then evaluate the feature subset by criterion functions that use the clustering result. Wrapper methods directly incorporate the clustering algorithm's bias in search and selection. The basic components are the feature search method, the clustering algorithm, and the feature selection criterion.

In paper [49] provided a survey of wrapper methods for unsupervised learning. Most wrapper methods in that survey apply a feature-selection criterion similar to the one that the clustering algorithm optimizes. The clustering criterion deals with defining similarity metric or defines what natural means. The feature-selection criterion defines interestingness. These two criteria need not be the same. There are two feature selection criteria—maximum likelihood and scatter separability for a wrapper method that applies sequential forward search wrapped around Gaussian mixture model clustering. To cluster data, it is needed to make assumptions and define natural grouping. With this model, it is assumed that each of our natural groups is Gaussian. Here, [50] investigated that two ways Maximum likelihood is the same criterion that they used in their clustering algorithm. Maximum likelihood prefers the feature subspace that can be modeled best as a Gaussian mixture. We also explored scatter separability because many can use it with many clustering algorithms. Scatter separability is similar to the criterion function used in discriminant analysis. It measures how far apart the clusters are from each other normalized by their within-cluster distance. High values of maximum likelihood and scatter separability are desirable. In [50] concluded that no one criterion is best for all applications.

2.7.15 Subspace Clustering

Rakesh Agrawal and his colleagues introduced CLIQUE (Clustering in Quest), a subspace-clustering algorithm that proceeds level-by-level from one feature to the highest dimension or until it generates no more feature subspaces with clusters (regions with high density points).The idea is that dense clusters in dimensionality d should remain dense

in d-1. Subspace clustering also lets you discover different clusters from various subspaces and combine the results. Several new subspace clustering methods were developed after CLIQUE and summarized in a review by Lance Parson, Ehtesham Haque, and Huan Liu.

2.7.16 Probabilistic Model

Law et al (2003) incorporate feature saliency as a missing variable in a finite-mixture model that assumes relevant features to be conditionally independent given the cluster component label and assumes irrelevant features to have a probability density identical for all components. So, it can perform feature selection and clustering simultaneously in single expectation-maximization iteration.

2.7.17 Clustering

As it is mentioned earlier, you can perform feature selection by clustering in the feature space to reduce redundancy. Clustering has recently become popular because of research in microarray analysis. Clustering is simply clustering the row (sample space) and column (feature space) simultaneously. Inderjit S. Dhillon, SubramanyamMallela, and Dharmendra S. Modha provide a review of the literature in [Table 3].

Table 3: Taxonomy of Feature Selection Model

Model search	Advantages	Disadvantages	Examples
Filter (Univariate)	Fast Scalable Independent of the classifier	Ignores feature dependencies Ignores interaction with the classifier	Euclidean distance i-test Information gain, Gain ratio (Ben-Bassat, 1982)
Filter (Multivariate)	Models feature dependencies Independent of the classifier Bettercomputational complexity than wrappermethods	Slower than univariate techniques Less scalable than univariate TechniquesIgnores interaction with the classifier	Correlation-based feature selection (CFS) (Hall, 1999) Markov blanket filter (MBF) (Koller and Sahami, 1996)Fast correlation-basedfeature selection (FCBF) (Yu and Liu, 2004)
Wrapper (Deterministic)	SimpleInteracts with the classifierModels feature dependenciesLess computationallyintensive than randomized methods	Risk of over fitting More prone than randomized algorithms to getting stuck in a local optimum (greedy search) Classifier dependent selection	Sequential forward selection (SFS) (Kittler, 1978) Sequential backward elimination (SBE) (Kittler, 1978) Plus q take-away r (Ferri et al., 1994) Beam search (Siedelecky and Sklansky, 1988)
Wrapper(Randomized)	Less prone to local optima Interacts with the classifier Models feature dependencies	Computationally intensive Classifier dependent selection Higher risk of overfitting than deterministic algorithms	Simulated annealing Randomized hill climbing (Skalak, 1994) Genetic algorithms (Holland, 1975) Estimation of distribution algorithms (Inza et al., 2000)
Wrapper (Embedded)	Interacts with the classifier Better computational complexity than wrapper methodsModels feature dependencies	Classifier dependent selection	Decision trees Weighted naive Bayes (Duda et al., 2001) Feature selection using the weight vector of SVM (Guyon et al., 2002; Weston et al., 2003)

2.8 Taxonomy of Feature Selection Algorithms

In general, wrapper and filter method have been proposed for feature selection. Wrapper method adopts classification algorithms and performs cross validation to identify important features. Filter method utilizes correlation based approach. Wrapper method demands heavy computational resource for training and cross validation while filter method lacks the capability of minimization of generalization error. In order to improve these problems, several studies have proposed hybrid approaches which combine wrapper and filter approach. In this section, we explain in detail the three key models with some famous feature selection algorithms in [Table 4].

Table 4: Key References for each Type of Feature Selection Technique

Filter Methods	Wrapper Methods	Other Methods
Bivariate (Jonassen, 2002) CFS (Wang *et al*., 2005; Yeoh *et al*., 2002) MRMR (Ding and Peng, 2003) USC (Yeung and Bumgarner, 2003) Markov blanket (Gevaert *et al*.,2006; Mamitsuka, 2006;Xing *et al*., 2001)	Sequential search(Inza *et al*., 2004;Xiong *et al*., 2001) Genetic algorithms (Jirapech-Umpai and Aitken, 2005;Li *et al*., 2001;Ooi and Tan, 2003) Estimation of distribution algorithms(Blanco *et al*., 2004)	Random forest(Dıaz-Uriarte andAlvarez de Andres, 2006; Jiang *et al*., 2004) Weight vector ofSVM (Guyon *et al*., 2002) Weights of logisticregression (Ma andHuang, 2005)

2.9 Importance of Data Reduction for Intrusion Detection Systems

IDSs have become important and widely used tools for ensuring network security. It is impossible to do classification by human when a small network has stored very large amount of audit hence IDS deploys to examine data. When it is come to analysis of data, even with computer assistance, it is difficult to detect abnormal behavior patterns due to large amount of features in data. There is complex relationship between features that is difficult for human to comprehend. If real time detection is required, this is extremely necessary to reduce the amount of data for an IDS to process efficiently. Reduction can occur in one of several ways. Data that are not considered useful can be filtered, keeping only the potentially useful data. Data should be categorized in group or cluster to distinguish hidden patterns. Overhead can be significantly reduced by storing the characteristics of the clusters instead of the individual data. Therefore, some redundant data can be removed by using feature selection.

3. Conclusion

This paper has reviewed the feature selection and the whole taxonomy of tools and methods used in various IDS systems. Feature selection plays an essential role for IDS in order to reduce redundant pattern and data. The challenging problems issues in are accuracy, overhead process and effectiveness in terms of time consumption. Thus after review of many trends and approaches in IDS, it is assumed that the necessity of an efficient algorithm in feature selection is undeniable. The purpose of this review is to investigate in feature selection of IDS which is selectively choose significant features which represents categories of classification of attacks are based on four established dominant categories which are Denial of Service (DoS), Probe, User to Root (U2R) and Remote to Local (R2L) as widely used in other studies in the field of IDS (Abraham et al., 2007; Shafi and Abbas, 2009; Tajbakhsh et al., 2009; Farid et al., 2010; Teng et al., 2010).to learn the pattern in network traffic. The KDD Cup 1999 Intrusion Detection data set is widely used by other researchers in this field to investigate to accuracy and correctness of their proposed classifications (Abraham et al., 2007; Jemili et al., 2007; Shafi an Abbas, 2009; Tajbakhsh et al., 2009; Farid et al., 2010).

References:

[1] Bishop Matt. Computer security e art and science: Addison Wesley; 2003.
[2] Russel, D. & Gangemi, G.T. 1992. Computer security basics.CA: O=Reilly & Associates Inc. 448p.
[3] Caelli,W., Dennis, L. & Shain, M. 1994. Information Security Handbook. First edition.Wilthire:Macmillan Press Ltd. 833p.
[4] Anderson, J.P. 1980. Computer Threat Monitoring and Surveillance. (In Anderson, J.P. Technical report, Fort

Washington, Pennsylvania.) Bishop Matt. Computer security e art and science: Addison Wesley; 2003.

[5] Lee W, Stolfo S, Mok K. A data mining framework for building intrusion detection models. proceedings of the IEEE symposium on security and privacy; 1999a.

[6] Sung AH, Mukkamala S. Identifying important features for intrusion detection using support vector machines and neural networks. In: Proceedings of International Symposium on Applications and the Internet (SAINT 2003); 2003. p.209e17.

[7] Forrest S, Perelson AS, Allen L, Cherukuri R. Self-nonself discrimination in a computer. In: Proceedings of the 1994 IEEE symposium on research in security and privacy. Los Alamitos, CA: IEEE Computer Society Press; 1994.

[8] I. Corona, G. Giacinto, F. Roli ,Intrusion detection in computer systems using multiple classifier systems O. Okun, G. Valentini (Eds.), Supervised and Unsupervised Ensemble Methods and Their Applications, Springer-Verlag, Berling/Heidelber (2008), pp. 91–104

[9] Technical Report James P Anderson Co Fort Washington (1980): Technical report, James P. Anderson Company, Fort Washington, Pennsylvania, Pages: 56

[10] T. Crothers, Implementing Intrusion Detection Systems, A Hands-On Guide for Securing the Network, Wiley Publishing, Inc., 2003.

[11] Mykerjee, B., et al, 1994, "Network Intrusion Detection", IEEE Network, Vol.8, No.3, pp.26-41.

[12] Forrest, S. et al, (1994) “Self-Nonself Discrimination in aComputer”, Proceeding of 1994 IEEE Symposium onResearch in Security and Privacy, Los Alamos, CA: IEEEComputer Society Press.

[13] Kim and Spafford, 1995 Gene H. Kim, Eugene H. Spafford Experiences with tripwire: using integrity checkers for intrusion detection <http://citeseer.ist.psu.edu/kim95experiences.html> (1995)

[14] Estévez-Tapiadoret al., 2004, J.M. Estévez-Tapiador, P. García-Teodoro, J.E. Díaz-Verdejo Anomalydetection methods in wired networks: a survey and taxonomy Computer Networks, 27 (16) (2004), pp. 1569–1584

[15] Mukkamala et al., 2004b Srinivas Mukkamala, Andrew H. Sung, Ajith Abraham, Vitorino Ramos Intrusion detection systems using adaptive regression splines ,in: I. Seruca, J. Filipe, S. Hammoudi, J. Cordeiro (Eds.), Sixth international conference on enterprise information systems, ICEIS'04, Portugal, vol. 3 (2004), pp. 26–33 ISBN 972-8865-00-7

[16] Giacintoetal., 2003 G. Giacinto, F. Roli, L. Didaci Fusion of multiple classifiers for intrusion detection in computer networks Pattern Recognition Lett., 24 (12) (2003), p. 1795

[17] Krugeletal., 2002 Krugel Christopher, Toth Thomas, Kirda Engin. Service specific anomalydetection for network intrusion detection. In: Proceedings of Symposium on Applied Computing; 2002.

[18] JoIo B. D. Cabrera, et al., “Statistical Traffic Modelin g For Network Intrusion Detection”, Proc of the 8th international Symposium on Modeling, Analysis and Sirnulation of Computer and Telecommunication Systems, San Fra ncisco, CA, 2000

[19] Anderson, J. P. (1980). Computer security threat monitoring and surveillance, (Technical Report), Washington, PA, James P. Anderson Co.

[20] Denning, 1987] Denning,D. E. (1987).An intrusion-detectionmodel. IEEE Transactions on Soft-ware Engineering, 13, 222-232.

[21] Smaha, S. E. (1988). Haystack: An intrusion detection system. Proceedings of the Fourth Aerospace Computer Security Applications Conference (pp. 37-44).

[22] M. Dekker, “Security of the Internet,” The Froehlich/Kent Encyclopedia of Telecommunications, Volume 15, pp. 231-255, New York, 1997.

[23] Jolliffe IT. Principal component analysis: Springer-Verlag; 1986. KDD cup 99 intrusion detection data set, http://kdd.ics.uci.edu/databases/kddcup99/kddcup.data_10_percent.gzO.

[24] T.F. Lunt, A. Tamaru, F. Gilham, R. Jagannathm, C. Jalali, P.G. Neumann, H.S. Javitz, A. Valdes, T.D. Garvey, A Real-time Intrusion Detection Expert System (IDES), Computer Science Laboratory, SRI International, Menlo Park, CA, USA, Final Technical Report, February 1992.

[25] D. Anderson, T. Frivold, A. Tamaru, A. Valdes, Next-generation intrusion detection expert system (NIDES), Software Users Manual, Beta-Update release, Computer Science Laboratory, SRI International, Menlo Park, CA, USA, Technical Report SRI-CSL-95-0, May 1994

[26] P.A. Porras, P.G. Neumann, EMERALD: event monitor- ing enabling responses to anomalous live disturbances, in: Proceedings of the 20th NIST-NCSC National Information Systems Security Conference, Baltimore, MD, USA, 1997, pp. 353–365.

[27] E. Tombini, H. Debar, L. Me ´, M. Ducasse ´, A serial combination of anomaly and misuse IDSes applied to HTTP traffic, in: Proceedings of the 20th Annual Computer Security Applications Conference, Tucson, AZ, USA, 2004.

[28] Te-Shun Chou, Kang K. Yen, Liwei An, Niki Pissinou, and Kia Makki. (October, 2007). Fuzzy Belief Pattern Classification of Incomplete Data. IEEE International Conference on Systems, Man and Cybernetics, pp. 535-540, Montreal, Quebec, Canada.

[29] Haykin, S. (1999). “Neural Networks: A Comprehensive Foundation.” 2nd. Ed. Upper Saddle River, N.J: Prentice Hall.

[30] Bishop, C. M. (1995). Neural networks for pattern recognition. England: Oxford University

[31] Manocha, S., & Girolami, M. A. (2007). An empirical analysis of the probabilistic K- nearest neighbour classifier. Pattern Recognition Letters, 28, 1818–1824.

[32] Vapnik, V. (1998). Statistical learning theory. New York: John Wiley. Wang, W., & Battiti, R. (2006). Identifying intrusions in computer networks with principal component analysis. In Paper presented at the proceedings of the first international conference on availability, reliability and security (ARES’06).

[33] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological Cybernetics, 43, 59–69.

[34] Mitchell, T. (1997). Machine learning. New york: McGraw Hill. Moradi, M., & Zulkernine, M. (2004). A neural network based system for intrusion detection and classification of attacks. In Paper presented at the proceeding of the 2004 IEEE international conference on advances in intelligent systems – Theory and applications. Luxembourg.

[35] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, P. J. (1984). Classification and regressing trees. California: Wadsworth International Group.
[36] Pearl, J. (1988). Probabilistic reasoning in intelligent systems. Morgan Kaufmann. Peddabachigari, S., Abraham, A., Grosan, C., & Thomas, J. (2007). Modeling intrusion detection system using hybrid intelligent systems. Journal of Network and Computer Applications, 30, 114–132.
[37] Koza, J. R. (1992). Genetic programming: On the programming of computers by means of natural selection. Massachusetts: MIT.
[38] Zimmermann, H. (2001). Fuzzy set theory and its applications. Kluwer Academic Publishers.
[39] Jang, J.-S., Sun, C.-T., & Mizutani, E. (1996). Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence. New Jersey: Prentice Hall.
[40] Kittler, J., Hatef, M., Duin, R. P. W., & Matas, J. (1998). On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(3), 226–239.
[41] L. Breiman, "Bagging Predictors," Machine Learning, vol. 24, no. 2, 1996, pp. 123–140.
[42] L. Breiman,"Random Forests,"Machine Learn- ing, vol. 45, no. 1, 2001, pp. 5–32.
[43] Chebrolu Srilatha, Abraham Ajith, Thomas Johnson. Hybrid feature selection for modeling intrusion detection systems. In: Pal NR, et al, editor. 11th International conference on neural information processing, ICONIP'04. Lecture Notes in Computer Science. vol. 3316. Germany: Springer Verlag; 2004. p. 1020e5. ISBN 3-540-23931-6.
[44] KDD'99 archive: The Fifth International Conference on Knowledge Discovery and Data Mining. URL: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (Last browsed in December 2007)
[45] DARPA Intrusion Detection Evaluation, MIT Lincoln Laboratory. URL: http://www.ll.mit.edu/IST/ideval/ (Last browsed in December 2007)
[46] W. Lee, S.J. Stolfo, K.W. Mok, Adaptive intrusion detection: a data mining approach, Artif. Intell. Rev. 14 (6) (2000) 533–567.
[47] Shelly Xiaonan Wu, Wolfgang Banzhaf The use of computational intelligence in intrusion detection systems: A review Memorial University of Newfoundland, St John's, NL A1B 3X5, Canada
[48] http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/
[49] A. Zainal, M. A. Maarof, and S. M. Shamsuddin, "Using Rough-PSO in Anomaly Intrusion Detection."
[50] Srilatha Chebrolua, Ajith Abrahama, Johnson P. Thomas (2004) Feature deduction and ensemble design of intrusion detection systems, Oklahoma State University
[51] Shailendra K. Preeti J. ,Effective Anomaly based Intrusion Detection using Rough Set Theory and Support Vector Machine, International Journal of Computer Applications (0975 – 8887), Volume 18– No.3, March 2011
[52] Esmat Rashedi, Hossein Nezamabadi-pour, and Saeid Saryazdi, "Gsa: A gravitational search algorithm," Information Sciences, vol. 179, no. 13, pp. 2232–2248, 2009.