A Hybrid Approach for Intrusion Detection using Misuse Detection and Genetic Algorithm

Thesis submitted in partial fulfilment of the requirements for the award of degree of

Master of Technology

in

Computer Science and Applications

Submitted By
Rohini Rajpal
(Roll No. 601303024)

Under the supervision of:
Dr. Sanmeet Kaur
Assistant Professor, CSED

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA – 147004

June 2015

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled, "*A Hybrid Approach for Intrusion Detection using Misuse Detection and Genetic Algorithm* ", in partial fulfillment of the requirements for the award of degree of Master of Technology in *Computer Science and Applications* submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of *Dr. Sanmeet Kaur* and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for award of any other degree of this or any other University.

(Rohini Rajpal)

601303024

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. Sanmeet Kaur)

Assistant Professor, CSED

Countersigned by

(Dr. Deepak Garg)

Head

Computer Science and Engineering Department

Thapar University

Patiala

(Dr. S. S. Bhatia)

Dean (Academic Affairs)

Thapar University

Patiala

ACKNOWLEDGEMENT

First of all I would like to thank the Almighty, who has always guided me to work on the right path of the life.

This work would not have been possible without the encouragement and able guidance of my supervisor and coordinator **Dr. Sanmeet Kaur**. I thank my supervisor for her time, patience, discussions and valuable comments. Her enthusiasm and optimism made this experience both rewarding and enjoyable.

I am also thankful to **Dr. Deepak Garg,** Head, Computer Science and Engineering Department.

I will be failing in my duty if I don't express my gratitude to **Dr. S. S. Bhatia**, Professor and Dean of Academic Affairs of University, for making provisions of infrastructure such as library facilities, immensely useful for the learners to equip themselves with the latest field.

I am also thankful to the entire faculty and staff members of Computer Science and Engineering Department for their direct-indirect help, cooperation, love and affection, which made my stay at Thapar University memorable.

Last but not least, I would like to thank my parents for their wonderful love and encouragement, without their blessings none of this would have been possible. I would also like to thank my close friends for their constant support.

ABSTRACT

Network Security has become the crucial issue for most of the organizations in the recent past. Mostly discussions on security include the tools and methods that can be deployed to protect and defend the networks. The use of network security tools have increased over the years due to increase in security threats. Many methods have been developed to secure computer networks and communication over the Internet. In today's fast-changing Information technology world, even the best available security is deficient for the latest vulnerabilities. In order to protect data and system integrity, Intrusion Detection has become central area for researchers. Intrusion detection method is one such method which has gained importance over the past few years.

In this dissertation, we have proposed an algorithm based on combination of Misuse Detection and Genetic Algorithm approach. We are using feature selection technique to extract important features from dataset. Genetic Algorithm is used for evolving best fit rules; this algorithm works on the principle of survival of fittest. For training and testing the rules, KDD Cup'99 datasets are used. The results of Misuse Detection and Proposed System are compared on various parameters like detection rates, false positive rates and number of attacks detected. Results prove that proposed approach has better detection rates and low false positive rates than Misuse Detection. Proposed System detects ten different types of attacks with high detection rates and low false positive rates. This System is also compared with existing systems which were described in research papers and results shows that our system gives less false positive rates than existing systems.

LIST OF FIGURES

LIST OF SNAPSHOTS

LIST OF PSEUDOCODE/ALGORITHM

LIST OF TABLES

TABLE OF CONTENTS

Chapter 1
Introduction

1.1 Introduction to Network Security

With the immense use of Internet, information has become valuable resource that needs to be protected from unauthorized access. Hackers make use of security breaches present in the system or network to attack it. Due to security breaches, individual users and organizations get affected. Internet has completely changed the way the things were done in the past but the sensitive information sent over it has become vulnerable to various types of attacks. Network is a collection of two or more computers linked together. Security refers to protection of information assets through the use of technology, process and training. Network security refers to activities designed to prevent and monitor unauthorized access, misuse, modification, or denial of a computer network. In today's fast-changing Information technology world, even the best available security is deficient for the latest vulnerabilities. Attack is any attempt to destroy, disable or make unauthorized use of an asset. Detection of attacks in network traffic is one of major goals of security. There are major three goals of security namely, confidentiality, integrity and availability. Confidentiality means that information should be hidden from unauthorized access. Integrity means that information should be protected from unauthorized change. Availability means that information should be available to an authorized user when it is needed. There are various security mechanisms recommended by X.800 to provide security services like encipherment, traffic padding, routing control, access control, authentic exchange and digital signatures [1]. Encipherment means hiding or covering data. The two techniques used for encipherment are cryptography and steganography. This mechanism helps in providing data confidentiality. Traffic padding is used to prevent the attacker from doing traffic analysis by inserting data that is not useful. The route between sender and receiver are continuously changed so as to prevent the attacker from checking a particular route. To check whether data integrity is preserved or not, a check value can be used. The check value is sent by sender along with data. The receiver also creates its own check value and then receiver can compare its created check value with the one received by the receiver to check the integrity of message.

1.2 Introduction to Intrusion Detection

Intrusion Detection is process of detecting attacks within computer or networks to identify security breaches. There are mainly two categories of intrusion detection techniques: namely, misuse detection and anomaly detection.

There are various ways for network security. Some of them are Firewall, Honeypot, Antivirus, and Intrusion Detection System. Firewall is a combination of hardware and software that protect any network from outside network. It allows some packets to pass and block some packets. Firewall applies some rules on each packet, based on that rule it decides whether to pass or block a particular packet. Honeypot is a trap set used by organizations to detect information about unauthorized user. It works by fooling attackers into believing that it is a legitimate system. The attackers attack the system without knowing that they are being observed. When attacker attacks the system, all information about the attacker such as IP address of attacker will be collected and this information can be used to trace back to source of an attack. An antivirus helps to detect and remove viruses from computer system. Intrusion Detection System gathers and analyzes information from various areas within a computer or a network to identify possible security breaches. Section 1.2.1 briefly explains intrusion detection system, its architecture and classification.

1.2.1 Intrusion Detection System

An intrusion detection system is a network security technology originally built for detecting vulnerability in a computer or network. Intrusion detection system can be a software or physical appliance that monitors network traffic in order to detect unwanted activity and malicious traffic. The advantage of IDS is, it is easier to deploy and detect many attacks just by checking packet header. The disadvantage of IDS's is, it gives many false positive alarms and false negatives.

1.2.1.1 Architecture of IDS

The various components of Intrusion Detection System are shown in Figure 1.1. The main components of Intrusion Detection System are Information Collection, Detection and Response.

- *Information Collection:* It is responsible for collecting data from system that is being monitored. It acts like an agent which continuously watches or monitors the network in real time. Example of input collection sources are network packets, log files and system logs. In this module, all input data is fed to event generator which then converted them to set of events and transferred them to sensor.

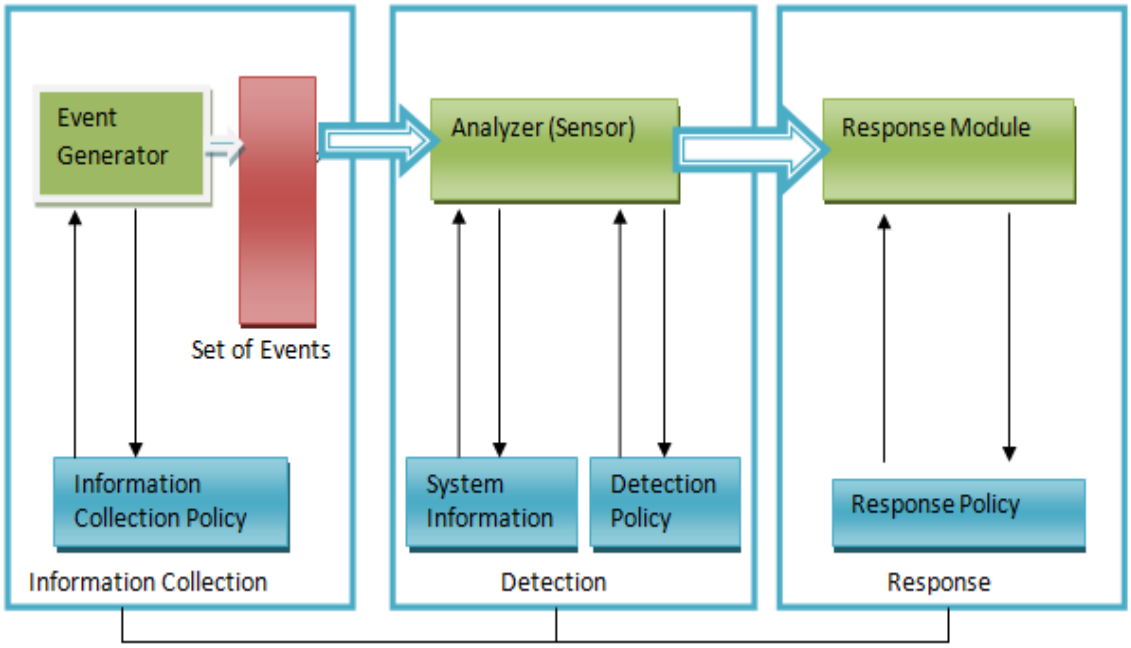


Figure 1.1. Architecture of Intrusion Detection System

- *Detection:* It processes the data collected from sensors to identify intrusive activity and use a system of rules to generate alarms from security events received. In this module, system information acts as knowledge base which contains information of all previously detected attack signatures. This information is usually provided by network and security experts. All the information about the previously detected attack signatures or patterns should be present in the database. When sensor detects some kind of malicious activity or signature, it matches it with current database and report to response component based on detection policy.

- *Response:* It is responsible for taking actions when an intrusion is detected. These responses can either be automatic or involve human interactions [26]. Response component can either send an alarm or an email notification about the intrusion detected to the administrator depending on the type of configuration.

1.2.1.2 Classification of IDS

Intrusion Detection Systems can be classified on the basis of approach, structure, data source, behaviour after an attack, analysis timing and protected system which are shown in Figure 1.2.

a) *Approach:* Based on the approach of intrusion detection, IDS can be divided into two types namely, signature detection and anomaly detection.

i) Signature Based Detection: In this approach, IDS matches known patterns with observed events and generates an alert if there is a match. This approach is very effective in detecting those intrusions whose signatures have been stored in database (known attacks) but it is very ineffective in detecting unknown attacks. The major disadvantage is that there is difficulty in updating information about new types of attacks. In misuse intrusion detection system, detectors analyze the incoming packets and try to find pattern matching; patterns were already stored in database. With this approach, only known intruders can be detected. There are various types of misuse intrusion detection methods like signature based, rule based, state transition, data mining [11]. The advantage of using this approach is to achieve less false positive rate but the main disadvantage is it cannot detect attacks which are unknown or never happened before.

ii) Anomaly Detection: In this approach, anomaly detector constructs normal user profile and some threshold value. It then uses current user behaviour to detect if there is any mismatch between profiles and detect intrusions based on difference between profiles. In order to match user profile, system is required to train to produce initial user profiles to train the system with regard to legitimate user behaviours. The disadvantage of this approach is high false positive rates and there is requirement of updating normal behaviour in the system. In anomaly based intrusion detection system, detectors detect behaviours on a computer or computer network [9]. Normal profiles are stored in database and some threshold value is set. If system is deviating from normal profile and its deviating value is greater than threshold value, then this system generates alerts. These detectors construct profile for connections, users and servers for their normal behaviour. Anomaly detection relies on being able to define desired behaviour of the system and then to distinguish between desired and anomalous behaviour [8]. There are

various types of anomaly intrusion detection methods like statistical, profiling, distance based, model based [13]. The advantage of anomaly detectors over misuse is that it can also detect intrusions which had not happened before and disadvantage is high false positive rates.

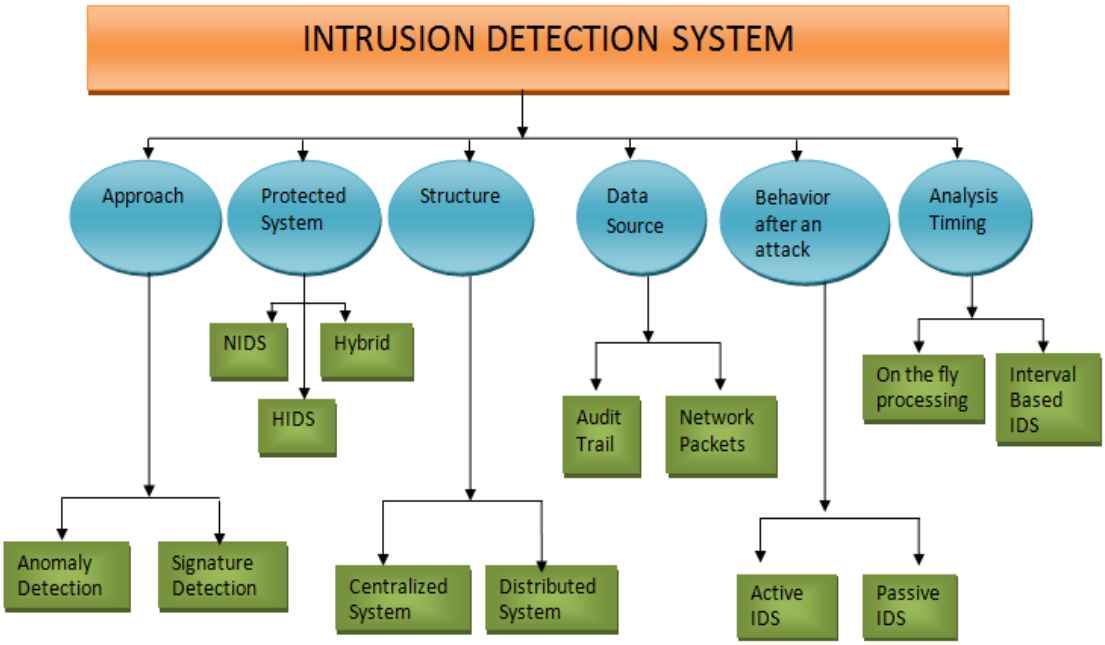


Figure 1.2. Classification of IDS

b) *Behavior after an Attack:* Based on the behavior after an attack, means how an intrusion detection system reacts after capturing alerts, can be divided into two types namely, active and passive IDS.

i) Active IDS: Active intrusion detection system not only detects attacks but also responds to them by logging out intruders or blocks some ports. This system also generates alerts and tries to patch software holes before getting hacked.

ii) Passive IDS: Passive intrusion detection systems do not react to any attacks. They simply generate alerts when suspicious traffic is detected and log network packets into log file. These systems are configured only to monitor and analyze network traffic.

c) *Protected System:* Intrusion detection System uses information that comes either from a single host or from a whole segment of a local network. IDS which collect information from single host are called Host based IDS are called host based IDS and those uses whole segment are called Network based IDS.

i) *NIDS:* Network based Intrusion Detection System produces data about local network usage. It examines traffic on all layers of OSI (Open Systems Interconnection) Model. NIDS analyze all network packets that reach the network interface card. They not only deal with a single host but all hosts in the network segment. NIDS can be installed on active network elements like routers. Figure 1.3 shows network based intrusion detection system.

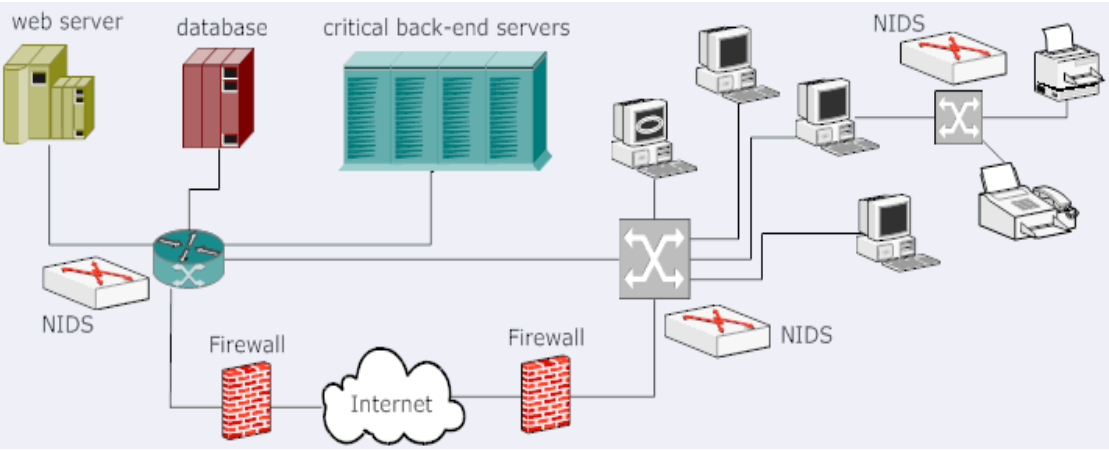


Figure 1.3 Network Intrusion Detection System

ii) *HIDS:* Host based intrusion detection system examines network packets that are attempts to access the host. This system only examines the network layer of the protected host. This system also monitor file system integrity and maintains log file for all activities and also monitor register state of the system and if there is any illogical change, then it generates alert. Figure 1.4 shows host based intrusion detection system.

d) *Structure:* Based on structure, IDS can be classified into two categories namely, centralized and distributed system.

i) *Centralized System:* In centralized system, data is collected from single or multiple hosts. All data is transferred to centre location for analysis. Examples are ARMD, Bro and ARMD.

ii) *Distributed System:* In distributed system, data is collected at each host. In this structure, there is distributed analysis of data. Examples are AAFID, GRIDS and CSM.

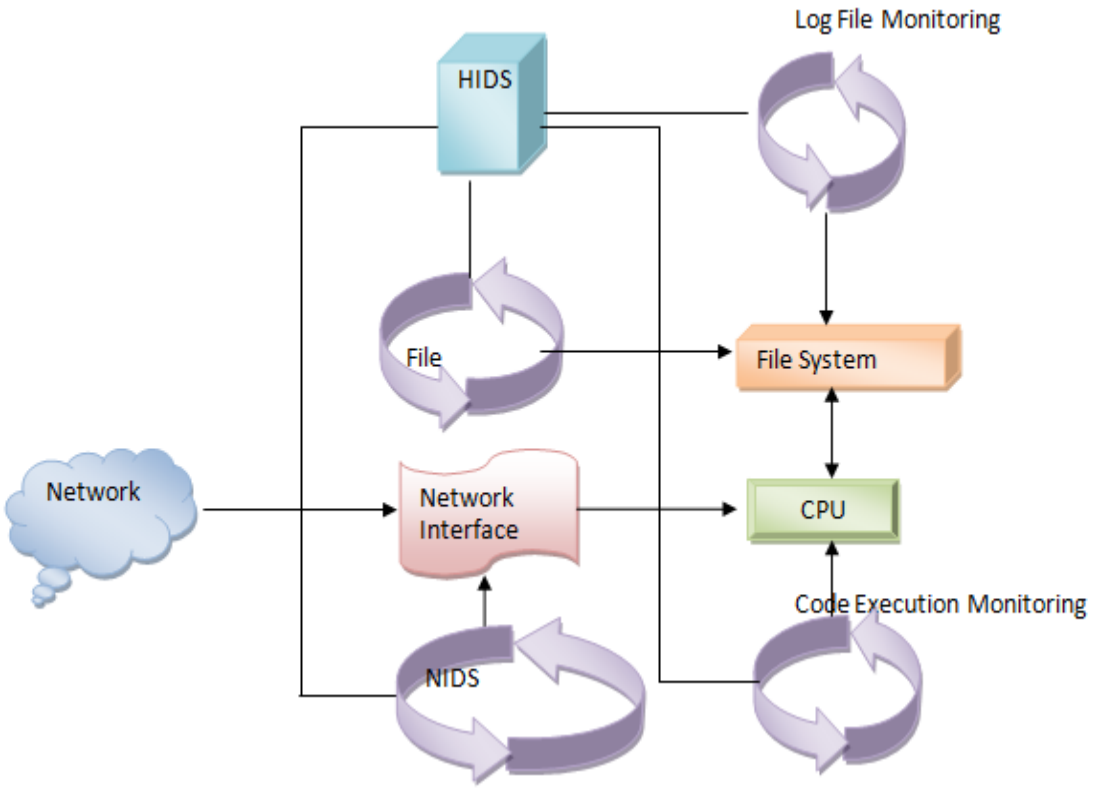


Figure 1.4. Host Intrusion Detection System

e) *Data Source:* Based on source of data, IDS can be classified into two types, namely audit trail and network packets.

i) *Audit Trail:* An audit trail is a set of records that provide documentary evidence of the sequence of activity that have affected at any time a specific procedure or event.

ii) *Network Packets:* Network packet is a data unit which contains control information and user data.

f) *Analysis Timing:* Based on analysis timing, IDS can be classified into two types namely, on the fly processing and interval based IDS.

i) *On the fly processing:* With on the fly processing, IDS performs online verification of events and respond them simultaneously. An ID using this processing requires more RAM because high data storage is required to trace all network packets online.

ii) *Interval Based IDS:* With interval based processing, IDS uses a process to check the status and content of log files at predefined intervals.

1.3 Introduction to Machine Learning

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning in artificial intelligence. Machine learning explores the construction and study of algorithms that can learn from and make predictions on data. Such algorithms operate by building a model from example inputs in order to make data driven predictions or decisions. Machine learning tasks are typically classified into three broad categories namely, supervised learning, unsupervised learning and reinforcement learning. There are various machine learning algorithms available for classification of data. Some of them are Decision Tree Learning, Association Rule Learning, Artificial Neural Network, Inductive Logic Programming, Support Vector Machine, Clustering, Bayesian Network, Feature Learning and Genetic Algorithm.

1.4 Introduction to Genetic Algorithm

Theoretical foundations of Genetic Algorithm were initially developed by Holland in 1970's. The inspiration of GA is based on the evolutionary process of biological organisms in nature. During the course of evolution, natural population evolves according to the principle of natural selection and survival of the fittest. Individuals who are easily adaptable to all environmental conditions and have higher fitness are more likely to reproduce and generate offspring while lower fitness individuals are eliminated from population [29]. A Genetic Algorithm stimulates these processes by taking an initial population of individuals and applying GA operators in each generation. Each individual is encoded as a chromosome which is a solution to the problem. A chromosome is a collection of genes, means an individual is made up of genes. The fitness of each individual is calculated by objective function. Highly fit individuals are given chances for reproduction, in crossover procedure. Mutation is optional for changing some of genes in individual to avoid duplicity. This evolution, selection, crossover process repeated until the condition is fulfilled.

1.5 Organization of Thesis

Chapter 1 describes the brief introduction of network security and various techniques used for intrusion detection. This chapter presents detailed information about

intrusion detection system, its architecture and classification. This chapter also discusses about machine learning and genetic algorithm in brief.

Chapter 2 provides literature survey of various techniques used for detecting intrusions using genetic algorithm. This chapter includes various researches for intrusion detection with more emphasis on genetic algorithm and hybrid approach using genetic algorithm. This chapter also discusses about machine learning algorithms and genetic algorithm in detail which is the topic of concern in this thesis.

Chapter 3 states the problem statement of the thesis, the objectives and goals to be achieved to carry out the thesis work.

Chapter 4 presents the methodology used to achieve the objectives, the experimental setup, implementation details of the hybrid approach.

Chapter 5 provides the results and discussion of the work carried out. This chapter also includes the main findings of the thesis. The proposed algorithm is compared with existing intrusion detection techniques in terms of detection rates and false positive rates.

Chapter 6 concludes with the findings of the proposed algorithm. Lastly, the direction for the future enhancement of the proposed algorithm has also been stated.

Chapter 2
Literature Survey

This chapter is the output of literature survey of techniques used to detect intrusions either by genetic algorithm or any hybrid approach using combination of genetic algorithm and some other machine learning algorithm. This chapter also discusses about machine learning algorithms, KDD dataset and genetic algorithm which is the topic of concern in this thesis.

2.1 Machine Learning

Machine learning is a type of artificial intelligence that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data. Machine learning tasks are typically classified into three broad categories, depending on the nature of the learning signal or feedback available to a learning system [30].

- *Supervised Learning:* The computer is presented with example inputs and their desired outputs and the goal is to learn a general rule that maps inputs to outputs.
- *Unsupervised Learning:* No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning is helpful in discovering hidden patterns in data.
- *Reinforcement Learning:* A computer program interacts with a dynamic environment in which it must perform a certain goal such as driving a vehicle, without a teacher explicitly telling it whether it has come close to its goal or not.

2.1.1 Machine Learning Algorithms

There are various machine learning algorithms available for classification of data. Some of them are explained below:

- *Decision Tree Learning:* In decision tree learning, decision trees are used for classification of data and it is used in data mining. Decision tree helps to create a

model which takes input and predict the value of target variable. All nodes except leaf nodes represent input variables and leaf nodes are targeted values. Any path from root to leaf represents solution of the problem [38].

- *Association Rule Learning:* This method is used to identify relations between large set of databases, based on those relation design some rule. The rule generation in this algorithm took two processes. In first process, minimum support is applied to all attributes in dataset and in second step minimum confidence is maintained to derive rules [39].

- *Artificial Neural Network:* This algorithm is inspired by human nervous system and it is used for approximating the solutions. Collection of neurons connected together to form biological neural network is used for deriving solutions and is called hidden layer. These systems are largely used in statistics [40].

- *Inductive Logic Programming:* This technique uses logic programming to derive hypothesis about the facts stored in database. The facts and the rules are knowledge base for inductive logic programming. Based on the knowledge base and some positive and negative examples from the facts conclude the hypothesis. Prolog language is used for this technique. This type of machine learning is useful in natural language processing field [41].

- *Support Vector Machine:* Support Vector Machine is supervised machine learning model which is used for classification of data. Suppose we have two categories of data in training dataset and SVM helps to categorize new data. SVM constructs a hyperplane in n dimensional space which can be used for classification of data [42].

- *Clustering:* Clustering is the process of grouping similar objects together in same group while some other objects in other groups. Each group is called cluster. This technique is used in image analysis, machine learning, data mining, statistical data analysis. Some of the clustering techniques are hierarchical clustering, centroid based clustering, distribution based clustering and density based clustering [43].

- *Bayesian Network:* Bayesian Network is a probabilistic graph model that represents set of random variable and their conditional independence with the help of directed acyclic graph. Edges in this network represent conditional dependency and nodes which are not connected are conditionally independent. Each node has probability function which takes input from parent variable and gives output as probability of variable represented as node [44].

- *Feature Learning:* Feature Learning is motivated by the fact that machine learning tasks such as classification requires input which is easy to process. So, it is necessary to discover relevant features from raw data. Examples of feature learning techniques are principal component analysis, dictionary learning [45].

- *Genetic Algorithm:* Genetic Algorithm works on principle of natural selection and evolution. Algorithm starts with population of individuals and evolves by selection, crossover and mutation process. Output of this algorithm is highly fit individuals which further reproduce to form offspring. In this way, best solution space is obtained from population of solution space. Detail of Genetic Algorithm is explained below [5].

2.2 Genetic Algorithm

Genetic Algorithm is an intelligent probabilistic search algorithm which can be applied to a variety of combinational optimization problems. Theoretical foundations of Genetic Algorithm were initially developed by Holland in 1970's. The inspiration of GA is based on the evolutionary process of biological organisms in nature. During the course of evolution, natural population evolves according to the principle of natural selection and survival of the fittest. Individuals who are easily adaptable to all environmental conditions and have higher fitness are more likely to reproduce and generate offspring while lower fitness individuals are eliminated from population [29]. The combination of good characteristics from highly adaptive ancestors may produce even more fit offspring. In this way, species evolve more and more to become well adapted on environment.

A Genetic Algorithm stimulates these processes by taking an initial population of individuals and applying GA operators in each generation. Each individual is encoded

as a chromosome which is a solution to the problem. A chromosome is a collection of genes, means an individual is made up of genes. The fitness of each individual is calculated by objective function. Highly fit individuals are given chances for reproduction, in crossover procedure. Mutation is optional for changing some of genes in individual to avoid duplicity. This evolution, selection, crossover process repeated until the condition is fulfilled.

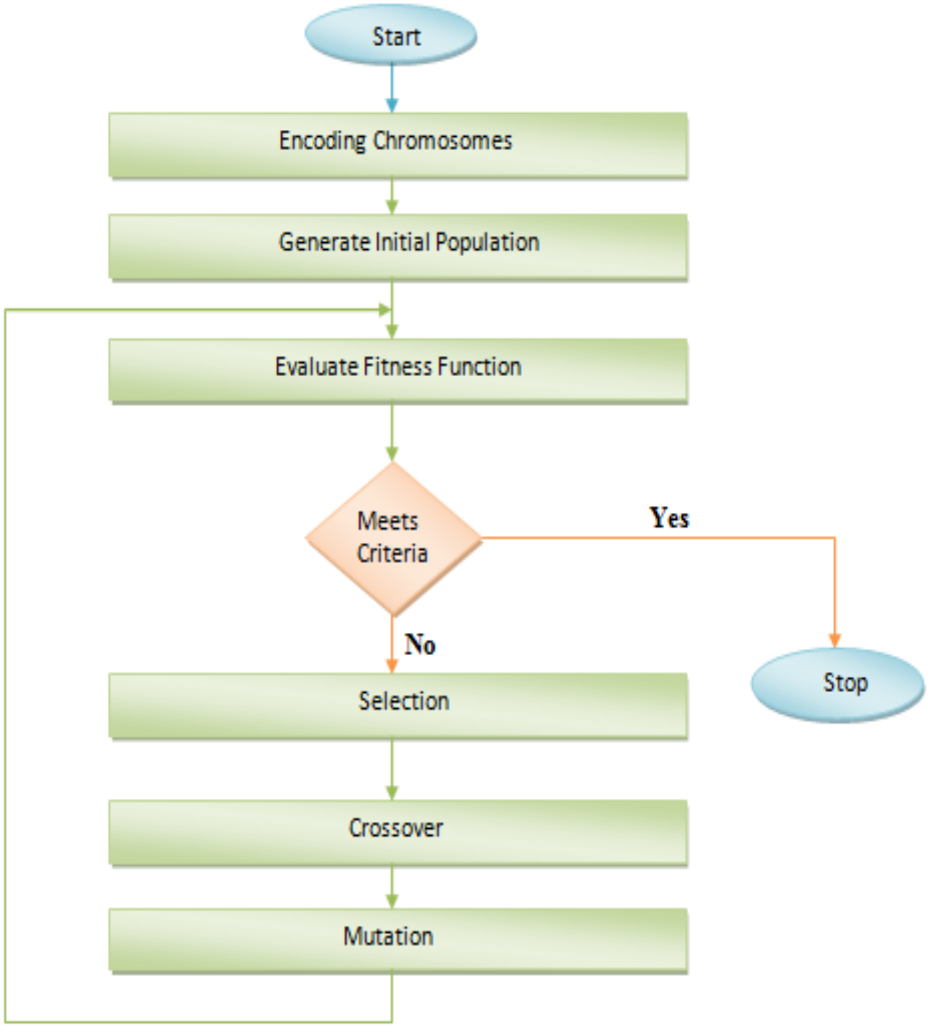


Figure 2.1 Genetic Algorithm Flow

2.2.1 Encoding of a Chromosome

The chromosome should be encoded in such a way that it must represent information about the solution. The most commonly used way to encode chromosome is in binary string. Each bit represents some information about solution. Every chromosome is a collection of genes where each gene represents each bit of chromosome.

Table 2.1. Representation of Chromosome

Chromosome 1	1110001101011010
Chromosome 2	1101010101110100

2.2.2 Pseudo Code of Genetic Algorithm

Pseudo Code 2.1

```
BEGIN
Initialize population with random candidate Solutions
Evaluate each candidate
Repeat until (Termination condition is satisfied)
DO
  Select parents
  Recombine pair of parents
  Mutate the resulting offspring
  Evaluate new candidate
  Select individuals for next generation
END DO
END
```

2.2.3 Genetic Algorithm Operators

The basic operations used in Genetic Algorithm are selection, crossover and mutation. Performance of Genetic Algorithm is dependent on these operators. Selection and crossover affects more on performance while mutation impact is light.

- *Selection:* In selection, chromosomes are given a probability of being selected that is directly proportional to their fitness. Higher the fitness, more the chances for generating offspring. For example, if we have four chromosomes of certain fitness and only two are allowed in next generation then chromosomes with highest fitness are allowed to mate to generate new offspring's. In Table 2.2.,

only chromosome 1 and chromosome 3 are allowed for crossover because they have higher fitness than Chromosome 2 and Chromosome 4.

Table 2.2. Individuals with Their Fitness

Individual	Encoding	Fitness Value
Chromosome 1	1110001101011010	0.6
Chromosome 2	1101010101110100	0.5
Chromosome 3	1010110101111001	0.7
Chromosome 4	1011011011010101	0.2

- *Crossover:* Crossover selects genes from parent chromosome and generates a new offspring [27]. We can select any crossover point. In example given below, we have two parents Parent 1 and Parent 2 and crossover point is 3rd bit. In the child's chromosome representation, we can see that Child 1 has 3 bits of Parent1 and 5 bits of Parent 2 .Similarly, Child2 has 3 bits of Parent 2 and 5 bits of Parent 1. Crossover point affects the performance of Genetic Algorithm.

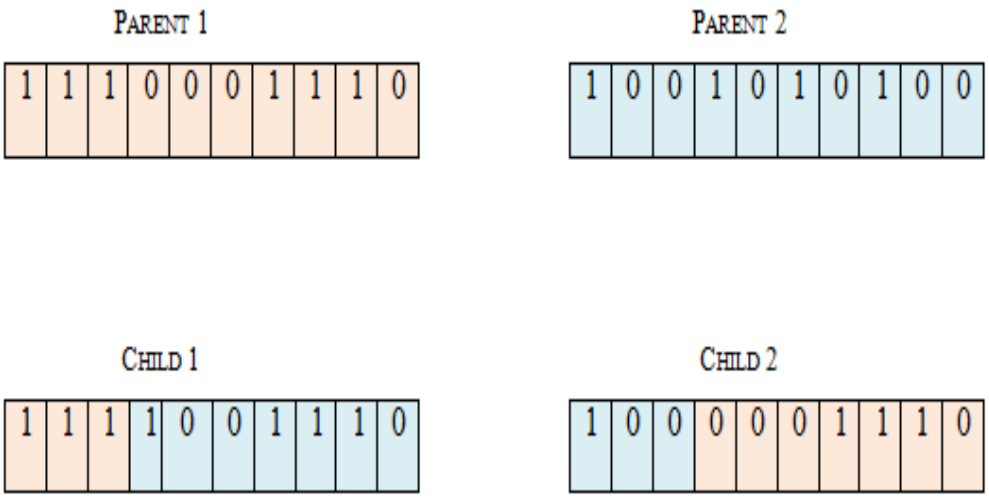


Figure 2.2. Crossover at 3rd point

- *Mutation:* After selection and crossover, we have a set of new individuals [28]. In order to ensure that all individuals are not exactly the same, mutation is done on individuals. Mutation is a slight change in encoding of chromosome. It just changes one or two bits in the chromosome. Random bit position is chosen for

mutation, if bit is 1, set it to 0 and vice versa. In the example below, mutation is done on 5th bit position. Mutation rate should be 0.01 to 0.02.

Figure 2.3 Mutation at 5th point

2.2.4 Genetic Algorithm Parameters

There are three basic parameters of Genetic Algorithm namely, crossover probability, mutation probability and population size.

- *Crossover Probability:* Crossover Probability tells about the expectation of crossover to be performed. If there is 0% crossover probability then in next generation all individuals are exactly the same as parents. If there is 100% probability then all chromosomes are new in next generation means all parent undergo crossover.

- *Mutation Probability:* Mutation Probability tells about the change in chromosome. If mutation probability of a chromosome is 100% then all bits of chromosome have been changed and if it is 0% then no bit has been changed.

- *Population Size:* Population size tells us about the number of chromosomes in the population. If there is less number of chromosomes then possibility of crossover and mutation is less and if numbers of chromosomes are more, then possibility also increases but after a certain population, it does not affect the possibility and slows down the system.

2.2.5 Designing Rules for Intrusion Detection

Every rule for intrusion detection is simple if-then clause. Suppose we have n features namely, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9…….a_n in a dataset and this dataset contains classes namely, c_1, c_2, c_3…. c_n. For rule designing we can use n features or subset of n features according to our requirement. Suppose all attributes contain numeric values and if we are using three features of this dataset then rule can be designed as:

if (a_1 =1 && a_2 =2 && a3=3) then c_1

where, a_1, a_2, a_3 are attributes of dataset and c_1 is class of the record.

2.2.6 Calculating Fitness Function of a Rule

To determine a fitness value of each rule, the following fitness functions can be used.

$$fitness = \frac{\alpha}{A} - \frac{\beta}{B} \tag{2.1}$$

$$fitness = w1 * support + w2 * confidence \tag{2.2}$$

where, support=|A and B|/N and confidence =|A and B|/|A|

In fitness function (1), α is the number of correctly detected attacks, A is the total number of attacks in the training dataset, β is the number of normal connections incorrectly characterized as attacks, i.e. false-positives, and B is the total number of normal connections in the training dataset[15]. Scale of fitness values is [-1, 1], where -1 is the lowest and 1 the highest value. High detection rate and low rate of false-positives result in a high fitness value. On the other side, low detection rate and high rate of false-positives result in a low fitness value.

In fitness function (2) of each rule , where N is the total number of network connections in the training dataset, |A| stands for the number of connections matching the condition A, and |A and B| stands for the number of connections that matches the rule if A then B . The weights w1and w2 are used to control the balance between the two terms of a rule [10].

2.3 KDD Dataset Description

Since 1999, KDD'99 has been the most wildly used data set for the evaluation of anomaly detection methods. This data set is prepared by Stolfo *et al.* [32] and is built based on the data captured in DARPA'98 IDS evaluation program [33]. DARPA'98 is about 4 gigabytes of compressed raw (binary) tcpdump data of 7 weeks of network traffic, which can be processed into about 5 million connection records. KDD training dataset consists of approximately 4,900,000 single connection vectors each of which contains 41 features and is labeled as either normal or an attack, with exactly one

specific attack type. There are four major categories of attack. All attacks fall under these four categories.

- *Denial of Service Attack (DoS):* is an attack in which the attacker makes some computing or memory resource too busy or too full to handle legitimate requests.

- *User to Root Attack (U2R):* is a class of exploit in which the attacker starts out with access to a normal user account on the system and is able to exploit some vulnerability to gain root access to the system.

- *Remote to Local Attack (R2L):* occurs when an attacker who has the ability to send packets to a machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access as a user of that machine.

- *Probing Attack:* is an attempt to gather information about a network of computers for the apparent purpose of circumventing its security controls.

Table 2.3. Attacks in KDD training and testing data

Attack Category	Attacks in KDD training dataset	Attacks in KDD testing dataset
DoS	back neptune land smurf teardrop pod	apache2 mailbomb processtable
Probe	satan nmap ipsweep portsweep	mscan saint
R2L	warezmaster warezclient ftpwrite guesspassword imap multihop phf spy	sendmail named snmpgetattack snmpguess xlock xsnoop worm
U2R	rootkit bufferoverflow loadmodule perl	httptunnel ps sqlattack xterm

In 1999, the original TCP dump files were preprocessed for utilization in the Intrusion Detection System benchmark of the International Knowledge Discovery and Data Mining Tools Competition. To do so, packet information in the TCP dump file is summarized into connections. Specifically, "a connection is a sequence of TCP packets starting and ending at some well defined times, between which data flows from a source IP address to a target IP address under some well defined protocol" . This process is completed using the Bro IDS, resulting in 41 features for each connection; Features are grouped into four categories namely, basic features, content features, time based traffic features, host based traffic features.

2.4 Survey of Existing Intrusion Detection Techniques using Genetic Algorithm Approach

Various researchers have proposed intrusion detection using genetic algorithm. Intrusion detection has been discussed by the following authors in their research work.

Salah *et al.* (2014), presents a Genetic Algorithm approach with an improved initial population and selection operator to improve intrusion detection. They have used KDD dataset for training and testing. For rule generation, they selected six features namely, duration, protocol, service, flag, src_bytes, and dst_bytes. To determine fitness of rule, they have used support-confidence framework. For evolving new generation, fitness value of rule should be greater than 0.6 in their approach. They have found five different types of attack namely, neptune, smurf, teardrop, pod and back [31].

Fatemeh (2014), presented a hybrid approach for dynamic intrusion detection in MANET's. In his work, he was using Genetic Algorithm and artificial immune system to classify attacks such as flooding, wormhole, rushing, neighbour and blackhole. Negative Selection algorithm was used in this approach which is a part of Artificial Immune Systems. This is adaptive to network topology, detectors in this approach uses partial or total updating method. Anomalous behaviour of a feature is detected by spherical detectors. This approach increases the average detection rate and running time of the system [16].

Padmadas *et al.* (2013), proposed layered based approach to detect four groups of attacks namely, R2L, DoS, U2R and probe. In their approach, they have used genetic algorithm for intrusion detection. They proposed four layered approach, each layer correspond to each type of attack. The main disadvantage of this approach is that they didn't provide any mathematical model for detecting attacks of each type. There is no parameter to calculate each layer attacks. The advantage of this approach is that they have detected R2L attacks with 90% accuracy [18].

Jongsuebsuk *et al.* (2013), proposed real time intrusion detection using fuzzy Genetic Algorithm to classify attack. They used Fuzzy rule to classify attacks while genetic algorithm is used for finding appropriate fuzzy rule and give optimal solution. In their approach, they have used two datasets KDD dataset and their own network data for intrusion detection. They have detected two different types of attacks namely, DoS and Probe. Their results proved that they have improved detection time using FuzzyGA algorithm [2].

Senthilnayaki *et al.* (2013), proposed a system in which Genetic Algorithm was used for feature selection and advanced J48 classifier was used for classifying attacks. They have used KDD dataset for detecting intrusions. Genetic Algorithm was used for selecting nine features from dataset namely protocol_type, service, src_bytes, dst_bytes, flag, diff_srv_rate, dst_host_srv_count, dst_host_rerror_rate, dst_host_srv_rerror_rate. In their work, they have used J48 classifier for increasing accuracy and speed while decreasing the error rate of the system. They have detected Probe and DoS attacks only [19].

Fan Li (2010), proposed combination of neural network and genetic algorithm to improve detection rates. The IDS presented in this paper is a combination of both misuse and anomalous detection and it is adaptive and flexible. In his approach, hybrid evolutionary algorithm was used and KDD Cup'99 dataset was used. Attacks detected are U2R, R2L, DoS, PRB. The result in this paper shows the maximum detection rate of 91.51% while false positive rate was 1.31 %. This approach is better than RWNN, BMPNN and ENN. Using other approaches detection rates of R2L attacks was very low, so this approach is better [20].

Wang (2009), presented expert fuzzy system based on Genetic Algorithm and fuzzy logic to improve detection rates with comparatively using less fuzzy rules. He used fourteen attributes out of forty one described in KDD Cup'99 dataset. Fuzzy rules were used to classify data and Genetic Algorithm was used to optimize membership function. Fuzzy Expert System presented in his work is very flexible and adaptive. In his paper, he designed the system in such a way that less number of fuzzy rules was required to classify attacks and attaining high detection rates. When this algorithm was applied to Intrusion Detection System, better results are obtained [21].

Chang *et al.* (2009), proposed an algorithm which combines wavelet neural network with Genetic Algorithm to achieve network efficiency and low false positive rates. In this paper, wavelet neural network was used to reduce localization problems and Genetic Algorithm was used for optimization of network structure and network weights. In their paper, they have experimented two approaches, wavelet neural network and wavelet neural network with genetic algorithm. The results of Genetic Algorithm with wavelet neural network are better than wavelet neural network in terms of detection rates and false positive rates [17].

Zorana *et al.* (2007), proposed a misuse detection system based on Genetic Algorithm approach. In this approach, KDD Cup'99 dataset was used for both training and testing. Support-confidence framework was used to calculate fitness of each rule. They have used three features namely, duration, src_bytes, dst_host_srv_rerror_rate. In their approach, they have used principle component analysis for feature selection and Genetic Algorithm for deriving best rules in order to increase detection rate. With this approach, they have detected three types of attacks neptune, smurf and portsweep but they have improved detection rates and reducing false positive rates to 1.6%. They not only classify the attack but also the type of attack which is also important to know for recovery of attacks [4].

Hui *et al.* (2005), presents a software implementation of Genetic Algorithm based approach to network intrusion detection. The genetic algorithm is employed to derive a set of classification rules from network audit data and the support-confidence framework with weights are utilized as fitness function to evaluate the quality of each rule. In their approach, they have used DARPA dataset. Six features were selected for

intrusion detection namely, duration, protocol, source_port, destination_port, source_ip, destination_ip. The attacks detected were pod and portsweep only [5].

Sadiq Ali Khan (2011), presents rule based Network Intrusion Detection using Genetic Algorithm. KDD Cup'99 dataset was used and only eight features are taken for classifying attacks. Features selected were service, land, flag, logged_in, root_shell, su_attempted, is_host_login, is_ guest_login. He proposed efficient approach to classify DoS or Probing attack [6].

Amira *et al.* (2012), proposed an approach for detecting intrusions using detectors generated by genetic algorithm. The approach presented in this paper uses negative selection technique, means negative selection technique of immune system to detect anomalies in the negative search space or complement or non self space. They have used NSL-KDD data set for testing. Their results shows that detection rates are better than other machine learning techniques and they also have detected some attacks which were not present in training dataset [34].

Balajinath *et al.* (2001), proposed an algorithm based on Genetic Algorithm to learn individual user behavior. This algorithm uses the past learning to calculate new individual user behavior. Fitness function of normal behaviour is stored in three tuple format. These three tuple value of every command is compared with non intrusive behavior value to find attacks. It has ability to find new intrusions in the system by learning from past. This algorithm gives detection rates of 96.80% and false alarm rate of 3.2%. The advantage of this system is that it can be extended in other environments [7].

Dong *et al.* (2005), proposed SVM + Genetic Algorithm based network intrusion detection system. They actually improved SVM based intrusion detection system by fusion of Genetic Algorithm for optimization of features. KDD Cup'99 dataset was used. When generations are increased, detection rates also increases. Although, SVM based intrusion detection system is better than other like neural network based intrusion detection system but as the number of features in dataset increases, detection rates decreases. So, when Genetic Algorithm is used with SVM based IDS, it gives higher detection rates and low false positive rates [22].

Anup *et al.* (2008), implemented Genetic Algorithm approach for classifying attacks. They designed a rule set consists of six different types of attacks which are majorly classified into two categories namely, denial of service and probing attacks. In this paper, KDD dataset was used for classification and detected majorly smurf and probe attacks. The results concludes that they have achieved high detection rate of 100% for denial of service attack , low false positive rate of approximately 0.2% and accuracy rate is greater than 95% in this approach [23].

Mohammad *et al.* (2012), implemented intrusion detection system using Genetic Algorithm. In this paper, KDD dataset was used to classify DoS, Probe, R2L and U2R attacks. They have made chromosomes of 23 different groups of attacks and then by using Genetic Algorithm classify all categories of attacks but detection rate of R2L attack is only 5.4% and U2R is 18.9%. Their approach has given satisfactory results for DoS attacks and detection rate of these attacks are 99.4% while probe has detection rate of 71.1% [24].

Amira *et al.* (2013), proposed a multilayer hybrid machine learning technique for anomaly detection and classification. In this paper, they have used principle component analysis in first layer for feature selection. In second layer, Genetic Algorithm was used as anomaly detector and in third layer they label the attacks using machine learning classifiers. Classifiers used in this approach are Naïve Bayes, Decision tree and artificial neural networks. They also concluded that best results are obtained by applying Euclidean distance measure and with population size of 200. They have detected all types of attacks like DoS, probe, R2L and U2R [25].

Adetokunbo *et al.*, proposed a comparison between snort wireless and genetic programming based intrusion detection. They have detected data link layer attack namely deauthentication attack which usually occurs in data link layer. Results show that genetic programming based detection is more effective than snort wireless. They have achieved detection rate of 100 percent and false positive rate of 0.1 percent [36].

Anil *et al.* (2013), proposed a hybrid method based on SVM, Genetic Algorithm and self organized feature map for anomaly detection. Support vector machine was used to classify input as normal or anomalous. Genetic Algorithm was used to select the most prominent features and self organized feature map was used for grouping of

similar data using similarity matrix from dataset. They have used KDD dataset for detection of anomalous intrusions. Their approach have improved detection rates up to 10 percent and reduction in false positive rates up to 50 percent compared to support vector machine [35].

The below table compares the intrusion detection approaches by various researchers using genetic algorithm and their results respectively.

Table 2.4 Comparison of intrusion detection techniques by various researchers

S.No.	**Year of Publication**	**Authors**	**Approach**	**Results**
1.	2014	Fatemeh Barani	Genetic + Artificial Immune System	Dynamic Intrusion Detection in MANET's with high detection rate.
2.	2014	Salah *et al.*	Genetic Algorithm	Improved search time without losing performance of the system.
3.	2013	Jongsuebsuk *et al.*	Fuzzy + Genetic Algorithm	High Detection Rate.
4.	2013	Anil *et al.*	SVM + Genetic Algorithm + SOFM	Increased detection rates up to 10% and reduction in false positive up to 50%
5.	2013	Amira *et al.*	Multilayer Hybrid Approach using Genetic Algorithm	Euclidean distance method gives better results in terms of true positive rates compared to Minkowski.
6.	2013	Senthilnayaki *et al.*	Genetic Algorithm + Modified J48 Decision Tree	Less error rate and time for classification is reduced.
7.	2013	Padmadas *et al.*	Layered Approach based on Genetic Algorithm	Implemented four layered system for four types of attack groups.
8.	2012	Mohammad *et al.*	Genetic Algorithm	Improved detection rate of DoS attacks to 99.4%
9.	2012	Amira *et al.*	Genetic Algorithm + Immune System	Better results than machine learning algorithms and found anomalies.
10.	2011	Sadiq Ali Khan	Genetic Algorithm	Classify DoS or Probing attack

11.	2010	Fan Li	H E NN + Genetic Algorithm	Increased detection rate.
12.	2009	Chang *et al.*	Wavelet Neural Network + Genetic Algorithm	Reducing false alarm and increased network efficiency
13.	2009	Wang Yunwu	Fuzzy + Genetic Algorithm	Less fuzzy rules required to achieve high detection rates.
14.	2008	Anup *et al.*	Genetic Algorithm	High detection rate of denial of service attacks
15.	2007	Zorana *et al.*	PCA + Genetic Algorithm	Intrusion Detection process becomes faster.
16.	2005	Hui *et al.*	Genetic Algorithm	Proposed system updates new rules.
17.	2005	Dong *et al.*	SVM + Genetic Algorithm	Gives better result than SVM based IDS in terms of detection rates.
18.	2001	Balajinath *et al.*	GBID(Genetic Algorithm Based Intrusion Detector)	Detect intrusions with an accuracy of 96.8%

Chapter Summary: This chapter describes about various machine learning techniques. Detailed analysis of KDD dataset is presented. Genetic Algorithm is explained in detail along with rule designing procedure and method to calculate fitness of each rule. This chapter also presents survey of research work carried out by various researchers and analysis of their results.

Chapter 3

Problem Statement

The current scenario of network security is very complicated. The statistics show that cyber crime has risen over the years. Enterprises are targeted by hackers either to steal the confidential information or just harm the organization by disrupting their services. Network security has really become important over the recent years. The organizations need to protect their network from attacks which occur from outside as well as within the organization. For the purpose of providing the solution to the problem of cyber attacks, many security solutions have been developed like antivirus, firewall, Intrusion Detection System, honeypots etc.

Intrusion Detection Systems plays vital role in detecting various kinds of attacks but their detection rates are comparatively less. Various techniques have been deployed to improve detection rates of intrusion detection system like embedding machine learning algorithms in their existing systems. The problem here is to enhance detection rates of network attacks.

Classification of attacks is very important in network forensics because a good recovery after the damage can be made by an attack can be done by knowing the exact type of an attack and its mechanism. The problem here is not only to detect attack but also to find exact type of an attack and classify them into four major categories namely, DoS, Probe, R2L and U2R.

Since Intrusion Detection System has to guard against more types of attacks, the number of false positive is likely to increase. Any Intrusion Detection System is considered to be good if it has ability to generate less false positive rates. Reducing false positive rates in Intrusion Detection System is a major topic of research these days. The problem here is to suppress false positive rates which improving detection rates of network attacks.

The topic of concern in this thesis is improving detection rates of network attacks; classify the type of attack into their respective categories and diminishing false positive rates of attacks.

The objectives of the thesis include the following:

- To design a hybrid approach which is a combination of genetic algorithm and misuse detection for detection and classification of network attacks.
- To make a testbed for testing the above model.
- To implement the proposed approach resulting in high detection rates and minimizing false positive rates.
- To validate the performance of misuse detection with hybrid approach based on detection rates, number of attacks classified, false positive rates.

Chapter 4
Implementation Details

The solution to the problem discussed in previous chapter is to develop a hybrid intrusion detection system which is based on combined approach of misuse detection and genetic algorithm. This chapter provides implementation details of the hybrid approach. The methodology used to implement the objectives and experiments performed are explained in this chapter. This chapter describes about the algorithmic details of all the experiments carried out in this thesis.

4.1 Methodology

This section presents the methodology used to carry out the thesis work. Figure 4.1 describes the schematic diagram of the system combining misuse detection and genetic algorithm. The datasets used in this approach are taken from http://nsl.cs.unb.ca/NSL-KDD/. KDD dataset is derived from DARPA dataset. KDD dataset consist of 10% of original dataset that is approximately 494020, single connection vectors each of which contains 41 features and is labelled with exact one attack type, i.e. either normal or attack. The proposed approach uses training dataset to collect information about connections and class of every connection is stored in database. Training and testing datasets are loaded into the system. The proposed approach contains two stages namely, misuse detection and genetic algorithm for classifying attacks. In first stage, i.e. misuse detection, the training data attributes are compared with testing data attributes and results are stored in database. In this stage, every testing connection is matched with training connection and if there is exact match, then system assigns some class to that testing connection. The process continues until all testing connections are compared with training connections. In second stage, i.e. Genetic Algorithm, rules with the highest fitness are inserted in training dataset and these rules are used for classifying attacks. Genetic Algorithm is used to derive best fit rules from the set of rules. Genetic Algorithm is used for obtaining optimized solutions from solution space. The methodology of proposed approach consists of six major steps namely, *Preprocessing, Feature Selection, Misuse Detection, Genetic Algorithm, Pattern Matching and Results Interpretation.* The details of each process are explained below.

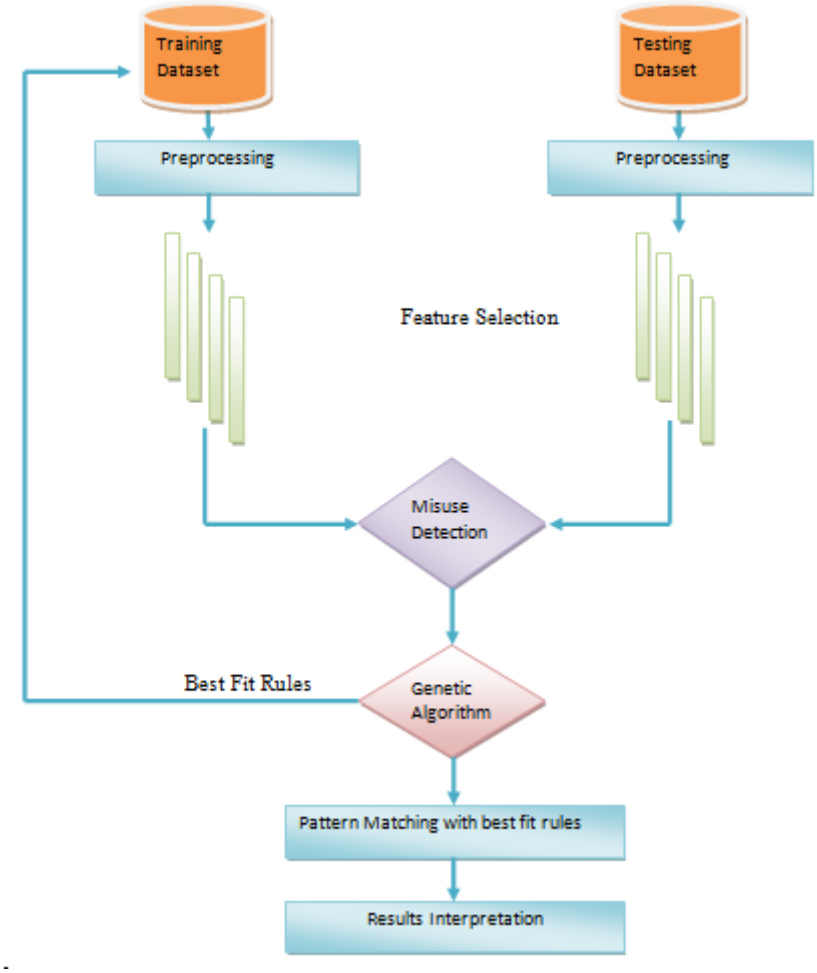


Figure 4.1 Schematic diagram of the system

- **Preprocessing**

KDD Cup'99 Training and KDD Cup'99 Testing Datasets are used for both training and testing of connections. This step involves removal of redundant records from both datasets. If we do not remove redundant data then result will be biased towards only few types of attacks. This step also includes conversion of categorical data to numerical data. The attributes in KDD data set are either in string or integer format. For ease of programming, we need to convert all these attributes to a single format, so we converted these three attributes to integer data type. Any random integer value can be assigned to these string attributes. Like service attribute has sixty two different string values, so we converted them to integer values ranging from 1 to 62. Similarly we converted protocol type and classname, we can assign any value, it is not mandatory to assign sequentially, it gives same result with any value but two strings

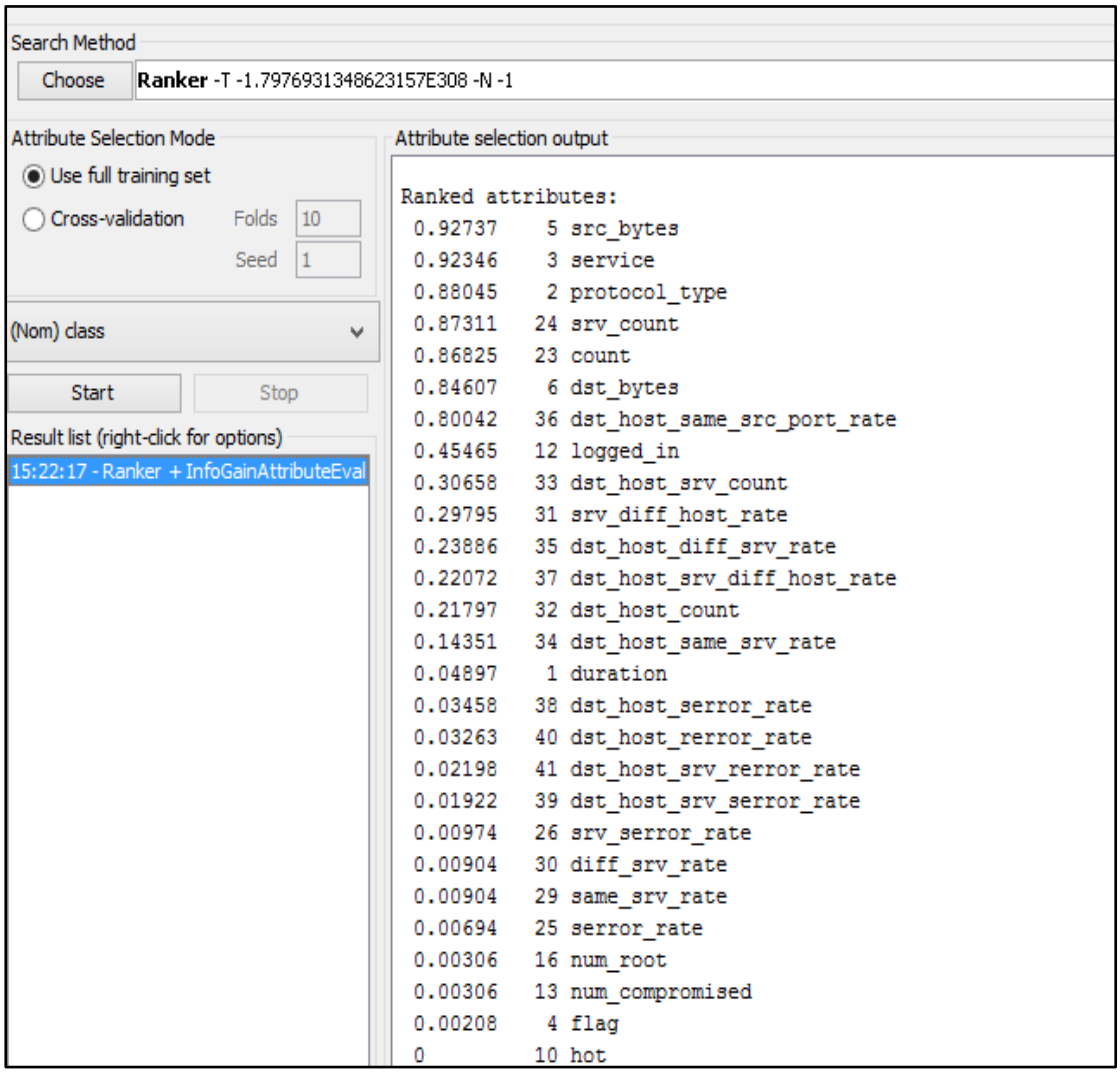
can’t have same value. Snapshot 4.1 shows the integer values assigned to some of the attributes in KDD dataset.

Protocol	Value		service	value	service	value		classname	value
icmp	1		ecr_i	1	nntp	32		smurf	1
tcp	2		private	2	netbios_ns	33		normal	2
udp	3		http	3	netbios_dgm	34		neptune	3
			smtp	4	netbios_ssn	35		snmpgetattack	4
			domain_u	5	uucp	36		portsweep	5
			ftp_data	6	courier	37		ipsweep	6
			eco_i	7	mtp	38		nmap	7
			telnet	8	gopher	39		xlock	8
			discard	9	remote_job	40		multihop	9
			name	10	ctf	41		worm	10
			whois	11	ssh	42		xterm	11
			ftp	12	nnsp	43		teardrop	12
			echo	13	IRC	44		sqlattack	13
			daytime	14	imap4	45		apache2	14
			time	15	urp_i	46		satan	15
			netstat	16	vmnet	47		pod	16
			finger	17	klogin	48		warezclient	17
			other	18	kshell	49		buffer_overflow	18
			auth	19	exec	50		guess_passwd	19
			domain	20	login	51		warezmaster	20
			systat	21	bgp	52		back	21
			ctf	22	ldap	53			
			link	23	printer	54			
			supdup	24	shell	55			
			host10s	25	efs	56			
			iso_tsap	26	Z39_50	57			
			csnet_ns	27	rje	58			
			pop_2	28	ntp_u	59			
			pop_3	29	sql_net	60			
			sunrpc	30	3_443	61			
			uucp_path	31	s38	62			

Snapshot 4.1 Preprocessing of various attributes of dataset

- **Feature Selection**

To improve performance, we are using attribute selection technique described in WEKA (Waikato Environment for Knowledge Analysis) i.e. InfoGainAttributeEval. InfoGainAttributeEval class is defined under weka.attributeSelection package. InfoGainAttributeEval evaluates the worth of an attribute by measuring the information gain with respect to the class. There are total forty one features in KDD Cup'99 data set. As we can see from snapshot 4.2, info gain value of attributes like src_bytes, service and protocol are higher than other attributes. Out of forty one, only three features are selected for further evaluation to classify attacks. Features selected are src_bytes, service, and protocol_type. Time complexity is very less with three attributes as compared to forty one attributes.



Snapshot 4.2 Feature Selection using InfoGainAttributeEval in WEKA.

- **Misuse Detection**

In this step, selected attributes from testing dataset are matched with same selected attributes of training dataset. If there is a match, then system will find the corresponding class of the training data which is already stored in database and assign it to testing data record.

- **Genetic Algorithm**

Genetic Algorithm is basically deployed to improve the performance of the system. Genetic Algorithms are stimulated by Darwin's theory about evolution. This algorithm is used to input initial rules (population) and output best fit rules (best individuals). The algorithm is explained in Chapter 2.2. This algorithm generates the best fit rules which are also added to training database. In this approach, we found all types of attacks with improved detection rate and low false positive rates.

- **Pattern Matching**

The rules obtained from genetic algorithm are added to training dataset. After applying genetic algorithm on random population of rules, again system is allowed to match the selected features of training and testing data. In this step, system is more capable to classify attacks because of best fit rules in training dataset.

- **Results Interpretation**

The results of misuse detection and hybrid detection are compared. Results of hybrid (Genetic Algorithm + Misuse Detection) detection are far better than misuse detection. In our approach, we have detected ten different types of attacks with only three features. Along with this we have also improved the detection rates of all types of attacks compared to misuse detection.

4.2 Implementation of the System

The system proposed here is implemented in NetBeans using JAVA programming language and MySql database. Two tables are created in database namely, training dataset and testing dataset. These tables are linked with java main program using hibernate. The classes in java program are Crossover, Individual, FitnessofPopulation,

Training, Testing and Java Main Class. The abstract working of all classes is described below.

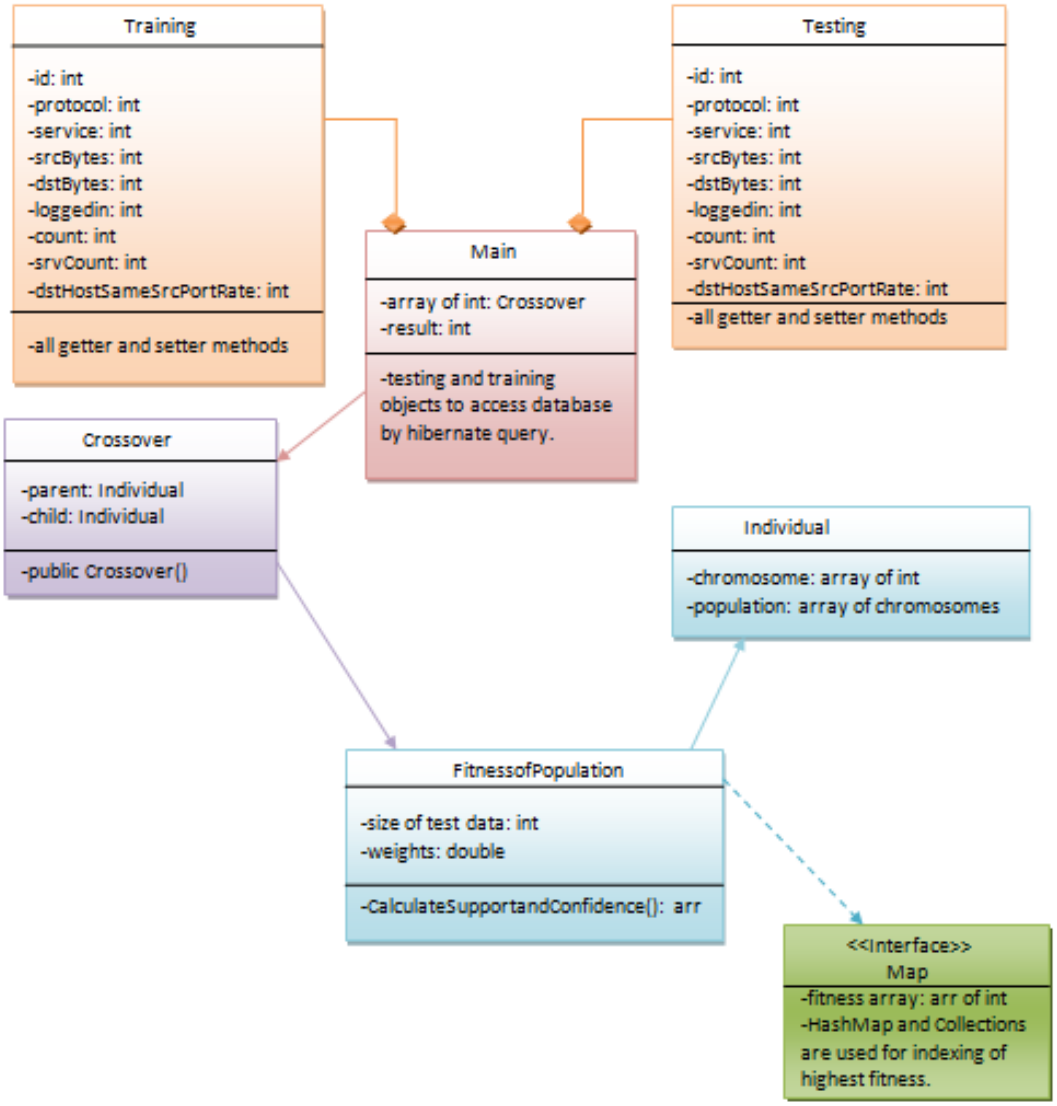


Figure 4.2 Class Diagram of the System

- *Training:* This is java bean class responsible for getting the values from database and storing it back to training database.

- *Testing:* This is also java bean class responsible for getting the values from database and storing it back to testing database. These two classes are mainly used for misuse detection.

- *Individual:* Every attribute is considered as a gene and a chromosome is collection of genes. So, every individual is a collection of attributes. Since we are

considering only three attributes out of forty one, so every individual array is formed of four attributes (three from feature selection and one is class attribute). Collection of these individuals is population. In intrusion detection these individuals are collection of rules.

- *FitnessofPopulation:* In this class fitness of every individual is calculated using support and confidence formula. After evaluating fitness of every individual (rule), we prioritize them using Map interface and Hashmap class. Sorting of fitness with their corresponding indexes are important to keep track of rules. Then higher fitness rules are extracted from this Hashmap using sublist function.

- *Crossover:* The higher fit rules are transferred to crossover and then these rules are extracted in pairs and then crossover with random function is applied on those pairs until all rules are covered. Random function is used because we didn't specify any crossover point, so every time it will consider a new point for crossover. The result of crossover is best fit rules (individuals) which are transferred to main program.

- *Main:* Main class sets the best fit rules in training dataset table and also collects values from testing and training dataset for pattern-matching and also for inserting the best rules in the training database so that in further evaluation, better results will be obtained.

4.3 Experimentation

Two experiments have been carried out for evaluation of detection rates with two different approaches. Subset of KDD Cup'99 dataset is used for both training and testing of data. In first experiment, misuse detection, is used for detecting intrusions from dataset using pattern matching. In second experiment, hybrid approach which is combination of misuse based and genetic algorithm is used for detecting intrusions with best fit rules and gives better detection rates. The goal of carrying out these experiments is to find detection rates and false positive rates of both approaches and compare them. Figure 4.3 shows the flow diagram of both the experiments. Below two sections explain experimentation details of misuse approach and hybrid approach. Algorithm of misuse detection and rule generation is also explained in detail.

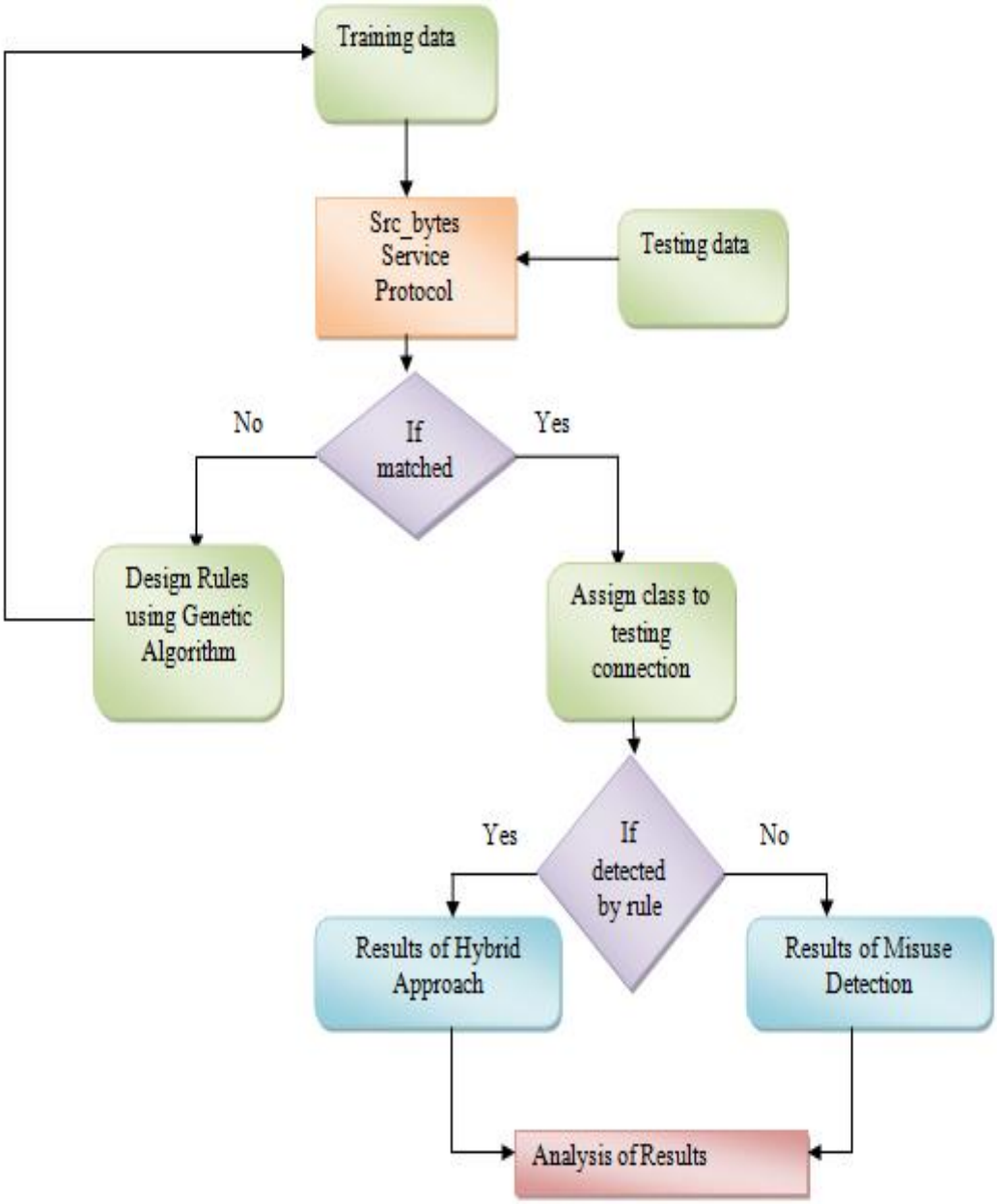


Figure 4.3 Flow Diagram of Misuse detection and Hybrid Approach

4.3.1 Experiment 1: Misuse Detection

In this experiment, training data attributes are matched with testing data attributes and if there is a complete match, means all attributes which are selected for intrusion detection from training dataset are matched with same attributes of testing data, then training data class is assigned to testing connection. Three attributes are selected based on their info gain value which is described in Section 4.1. Selected attributes are src_bytes, service, and protocol_type. The algorithm used in this thesis for misuse detection is explained below.

Algorithm 4.1: Misuse Detection:

Step 1. Remove redundant data from training dataset.

Step 2. Select three features based on rank of their Info Gain Value.

Step 3. Load testing dataset.

Step 4. For each connection in testing dataset

Step 5. Match selected features of training data with same features of testing data.

Step 6. If(class is not assigned for this connection)

6.1. if (all three attributes matched)

6.2. assign training class attribute value to testing connection.

6.3. else

6.4. don't assign any class to testing connection.

Step 7. else

Step 8. Break(exit for outer loop, do these steps for next connection).

Step 9. End For.

4.3.2 Experiment 2: Hybrid Approach

This experiment is a combination of misuse detection and genetic algorithm. In first stage, system is trained with 1500 connections of training dataset and then tested on 1400 connections of testing dataset. Subset of KDD Cup'99 dataset is used for both training and testing of data. Three attributes are selected based on their info gain value which is described in Section 4.1. Selected attributes are src_bytes, service, and protocol_type. Support confidence framework is used for designing rules using genetic algorithm which was explained in Chapter 2. This experiment includes algorithm for misuse detection which is described above and algorithm for rule generation using genetic algorithm which is described below.

Algorithm 4.2: For Rule Generation using Genetic Algorithm

Input: Population, Population Size, Crossover Point, Training dataset

Step 1. Initialize 350 Rules.

Step 2. Initialize w1 and w2 [range should be in between 0 to 1 such that w1+w2=1.

Step 3. N= total number of connections in training dataset.

Step 4. For each individual in the population

Step 5. a=0, ab=0

Step 6. For each connection in training dataset

6.1. if connection matches the individual(if-then both parts)

6.2. ab=ab+1;

6.3. end if

6.4. if connection matches the individual(if part only)

6.5. a=a+1;

6.6. end if

Step 7. End for

Step 8. Support (individual) =ab/N;

Step 9. Confidence (individual) =ab/a;

Step 10. Fitness (individual) =w1*support + w2*confidence;

Step 11. Calculate fitness of all individuals and select top 100 individuals out of them.

Step 12. End for.

Step 13. For each chromosome in new population

13.1. Apply crossover operator to form new offspring and crossover point is 0.5.

Step 14. End for.

Chapter Summary: This chapter presents the methodology of work carried out in thesis. This chapter also describes the implementation details of the proposed system. Class diagram of the system is also presented. Flow Diagram of Misuse Approach and Proposed Approach is presented. Experimentation details of misuse detection and hybrid approach is explained in detail with their algorithm.

Chapter 5

Results and Discussion

This chapter discusses the results of experimentation illustrated in previous chapter. Firstly, the results of misuse detection and hybrid approach are analyzed. Secondly, results of misuse detection are compared with results of hybrid detection. The performance is evaluated on detection rate, false positive rates and number of attacks classified by both approaches.

5.1 Analysis of Misuse Detection

Table 5.1 shows detection rates and false positive rates of misuse approach described in previous chapter. From Table 5.1, we can analyze that five attacks has been detected with this approach namely, smurf, neptune, snmpgetattack, ipsweep, teardrop. KDD dataset was used for training and testing of data. These attacks fall under three categories namely DoS, probe and R2L. Smurf, neptune and teardrop fall under DoS category while snmpgetattack and ipsweep fall under R2L and probe respectively. Misuse detection has achieved false positive rate of 0.35. Figure 5.1 shows detection rates of attacks using misuse approach.

Table 5.1. Detection rates and false positive rates of Misuse Approach

Attack Names	Type of Attack	Detection Rate (%)	False Positive Rate
Smurf	DoS	100	
Normal	No Attack	74.3	0.35
Neptune	DoS	66.3	
Snmpgetattack	R2L	2	
Portsweep	Probe	0	
Ipsweep	Probe	2.8	
Nmap	Probe	0	
Xlock	R2L	0	
Multihop	R2L	0	
Worm	R2L	0	
Xterm	U2R	0	
Teardrop	DoS	100	
Sqlattack	U2R	0	
apache2	DoS	0	

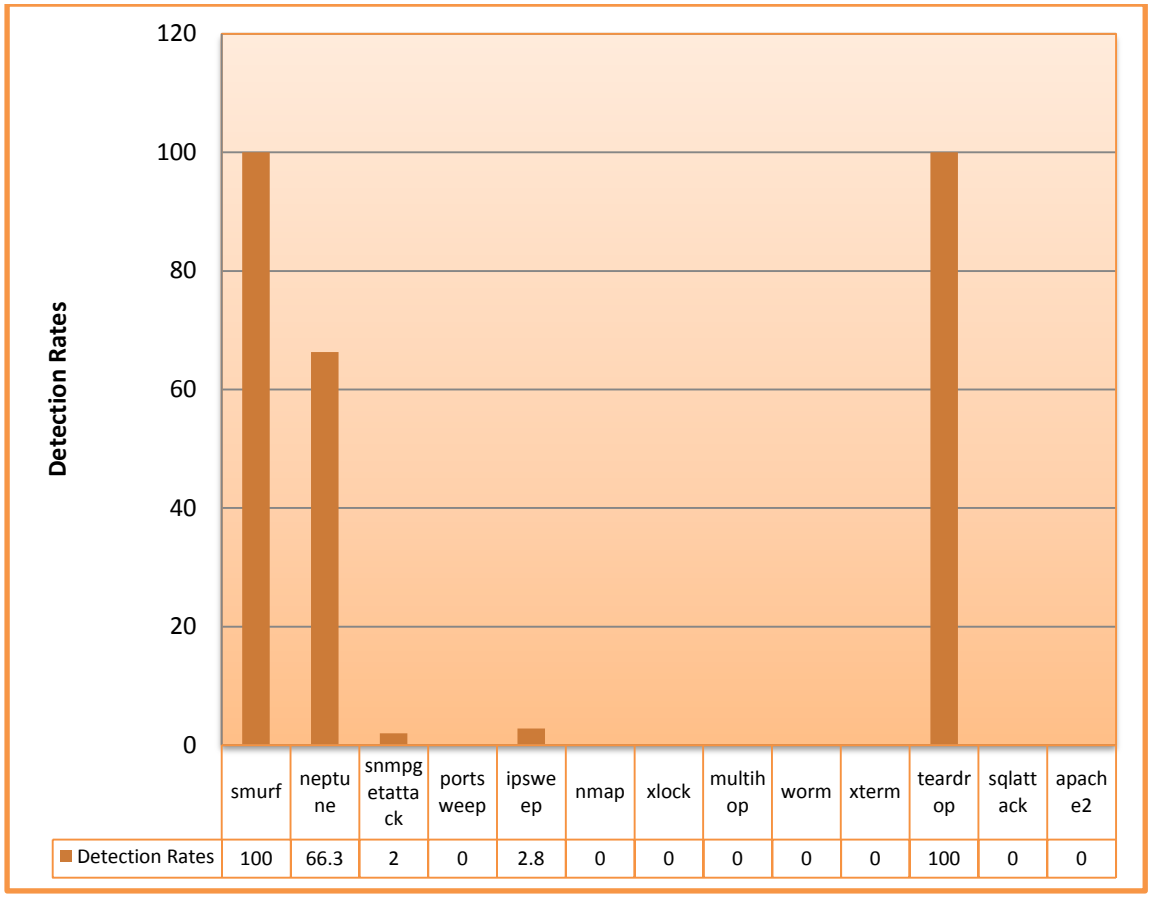


Fig. 5.1. Detection Rates of various attacks using Misuse Approach

5.2 Analysis of Hybrid Approach

This section describes and analyzes the results of hybrid approach discussed in previous chapter. Hybrid approach used here is combination of genetic and misuse detection. This approach is test with 200 rules and 350 rules with different weights. KDD dataset is used for training and testing of data. The following parameters are chosen and their detection rates and false positive rates are compared.

5.2.1 With 200 Rules

Numbers of rules affect the performance of detection in genetic approach. As number of rules increases, detection rates also increases. In this work, 200 and 350 rules are designed for classification of attacks. Along with these rules, weights also impact performance of the system. We have detected attacks with three combination of weights, w1=0, w2=1 and w1=0.4, w2= 0.6 and w1=1, w2=0.

- **w1=0, w2=1:** Table 5.2 shows detection rates of hybrid approach using 200 rules and weights w1=0, w2=1. Only four attacks namely, smurf, neptune, ipsweep, teardrop are detected by using these weights. Figure 5.2 shows detection rate of various attacks using hybrid approach with 200 rules and w1=0, w2=1.

Table 5.2. Detection rates of Hybrid Approach Using 200 Rules and weights, w1=0 & w2=1

200 Rules		w1=0,w2=1	
Attack Names	Type of Attack	Detection Rate (%)	False Positive Rate (%)
Smurf	DoS	100	0.35
Normal	No Attack	89.6	
Neptune	DoS	97.9	
Snmpgetattack	R2L	0	
Portsweep	Probe	0	
Ipsweep	Probe	100	
Nmap	Probe	0	
Xlock	R2L	0	
Multihop	R2L	0	
Worm	R2L	0	
Xterm	U2R	0	
Teardrop	DoS	100	
Sqlattack	U2R	0	
apache2	DoS	0	

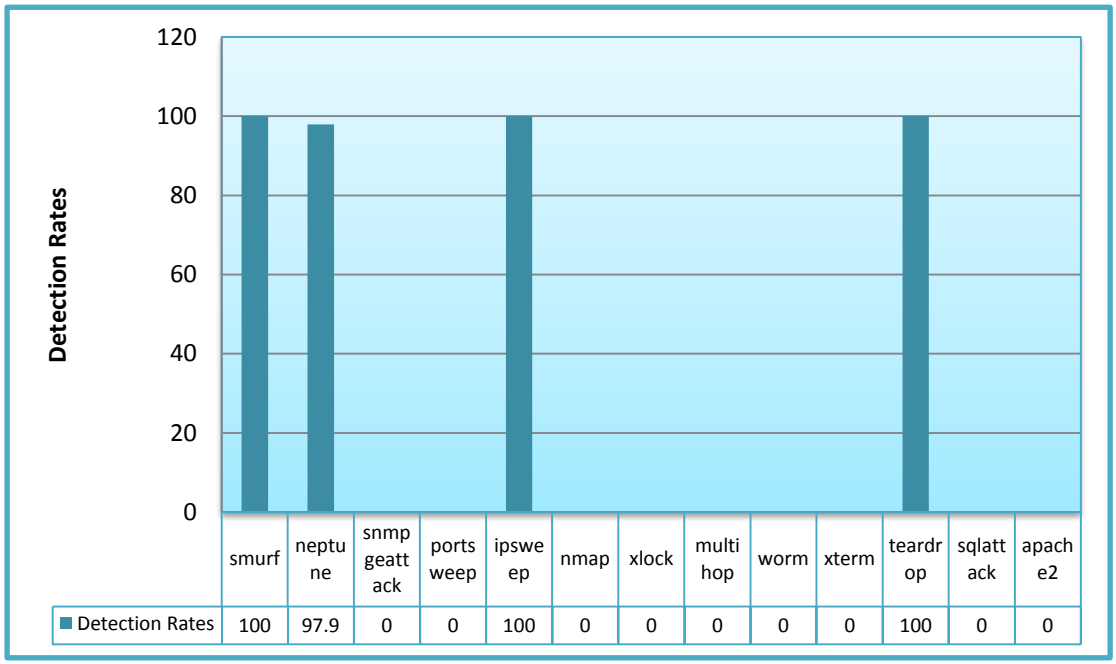


Fig. 5.2. Detection Rates of various attacks using Hybrid Approach and 200 rules with w1=0, w2=1

- **w1=0.4, w2=0.6:** Table 5.3 shows detection rates of hybrid approach using 200 rules and weights w1=0.4, w2=0.6. Only five attacks namely, smurf, neptune, portsweep, ipsweep, teardrop are detected by using these weights. Figure 5.3 shows detection rate of various attacks using hybrid approach with 200 rules and w1=0.4, w2=0.6.

Table 5.3. Detection rates of Hybrid Approach Using 200 Rules and weights, w1=0.4 & w2=0.6

200 Rules		w1=0.4,w2=0.6	
Attack Names	Type of Attack	Detection Rate (%)	False Positive Rate (%)
smurf	DoS	100	0.35
normal	No Attack	89.4	
neptune	DoS	23.3	
snmpgetattack	R2L	0	
portsweep	Probe	12.6	
ipsweep	Probe	100	
nmap	Probe	0	
xlock	R2L	0	
multihop	R2L	0	
worm	R2L	0	
xterm	U2R	0	
teardrop	DoS	100	
sqlattack	U2R	0	
apache2	DoS	0	

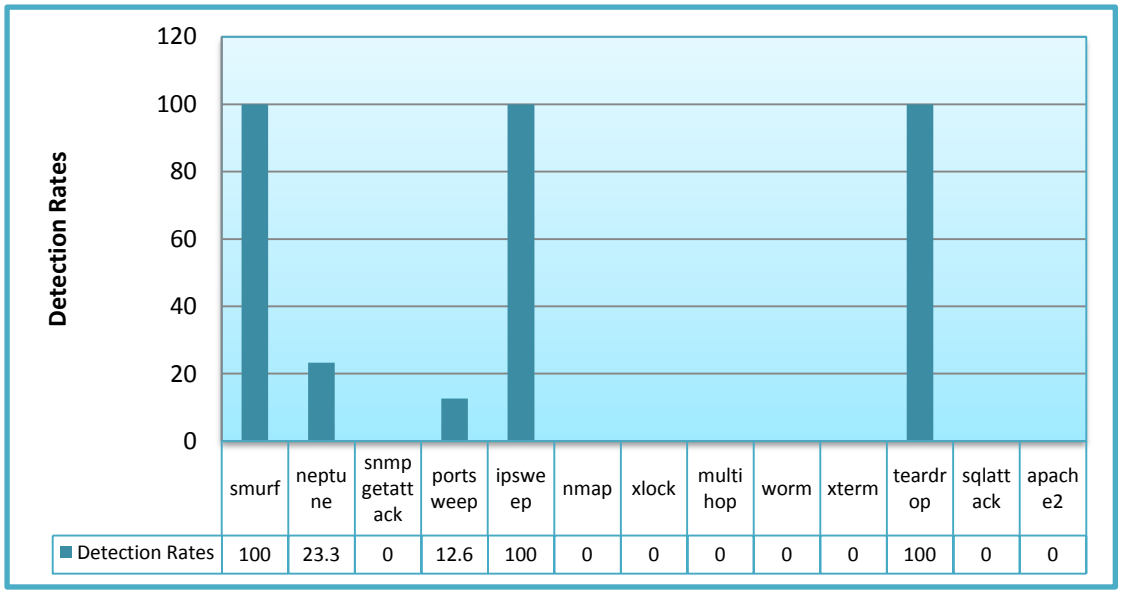


Fig. 5.3. Detection Rates of various attacks using Hybrid Approach and 200 rules with w1=0.4, w2=0.6

- **w1=1, w2=0:** Table 5.4 shows detection rates of hybrid approach using 200 rules and weights w1=1, w2=0. Only five attacks namely, smurf, neptune, portsweep, ipsweep, teardrop are detected by using these weights. Figure 5.3 shows detection rate of various attacks using hybrid approach with 200 rules and w1=1, w2=0.

Table 5.4. Detection rates of Hybrid Approach Using 200 Rules and weights, w1=1 & w2=0

200 Rules		w1=1,w2=0	
Attack Names	Type of Attack	Detection Rate (%)	False Positive Rate (%)
smurf	DoS	100	0.35
normal	No Attack	89.8	
neptune	DoS	54.9	
snmpgetattack	R2L	0	
portsweep	Probe	84.4	
ipsweep	Probe	100	
nmap	Probe	0	
xlock	R2L	0	
multihop	R2L	0	
worm	R2L	0	
xterm	U2R	0	
teardrop	DoS	100	
sqlattack	U2R	0	
apache2	DoS	0	

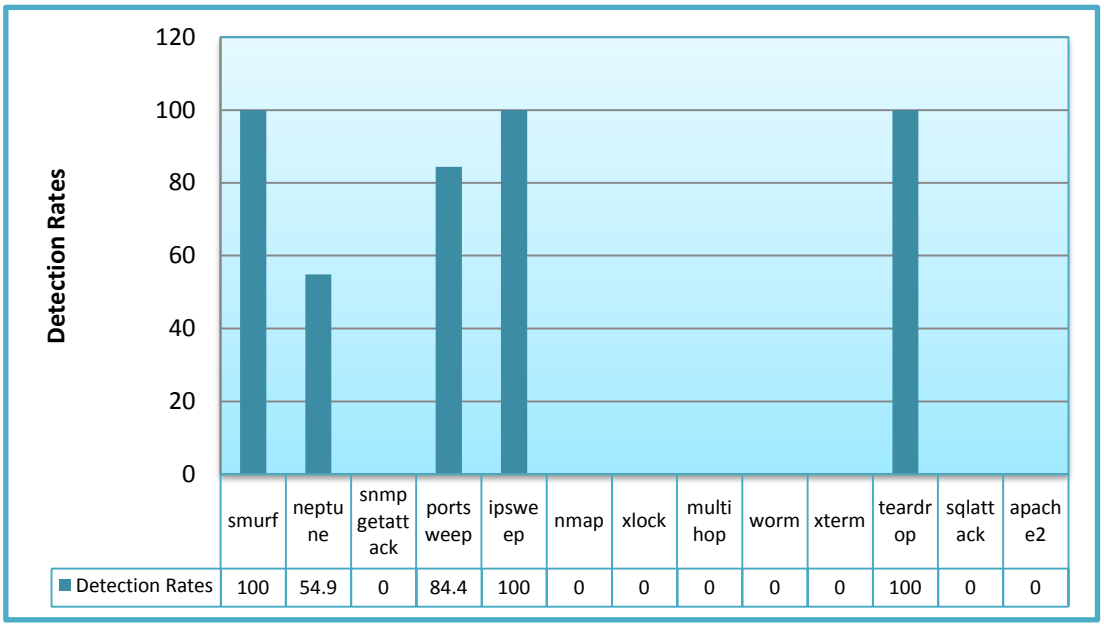


Fig. 5.4. Detection Rates of various attacks using Hybrid Approach and 200 rules with w1=1, w2=0

5.2.2 With 350 Rules

As number of rules increases, detection rates also increases but there is sudden increase in false positive rates. These 350 rules are again tested with three different combinations of weights which are described below.

- **w1=0, w2=1:** Table 5.5 shows detection rates of hybrid approach using 350 rules and weights w1=0, w2=1. This combination gives highest detection rates of ten different types of attacks which fall under four major categories of attacks. Attacks detected are smurf, neptune, snmpgetattack, portsweep, ipsweep, nmap, multihop, xterm, teardrop and apache2. Figure 5.3 shows detection rate of various attacks using hybrid approach with 200 rules and w1=0, w2=1. Four DoS, two R2L, three probe and one U2R attack types are detected with this approach with a false positive rates of 1.6 percent. As we can see from results in Table 5.5, we have achieved detection rates of greater than 95 percent for five different types of attacks. So, this approach is good for intrusion detection. Also, time taken is very less in this approach, so one can use this system in areas where high speed is required. This approach also detects every category of attack including U2R which were not detected by misuse detection.

Table 5.5. Detection rates of Hybrid Approach using 350 Rules and weights, w1=0 & w2=1

350 Rules		w1=0,w2=1	
Attack Names	Type of Attack	Detection Rate (%)	False Positive Rate (%)
smurf	DoS	100	1.6
normal	No Attack	73.4	
neptune	DoS	95.5	
snmpgetattack	R2L	100	
portsweep	Probe	2.9	
ipsweep	Probe	100	
nmap	Probe	2.08	
xlock	R2L	0	
multihop	R2L	33.3	
worm	R2L	0	
xterm	U2R	33.3	
teardrop	DoS	100	
sqlattack	U2R	0	
apache2	DoS	81.06	

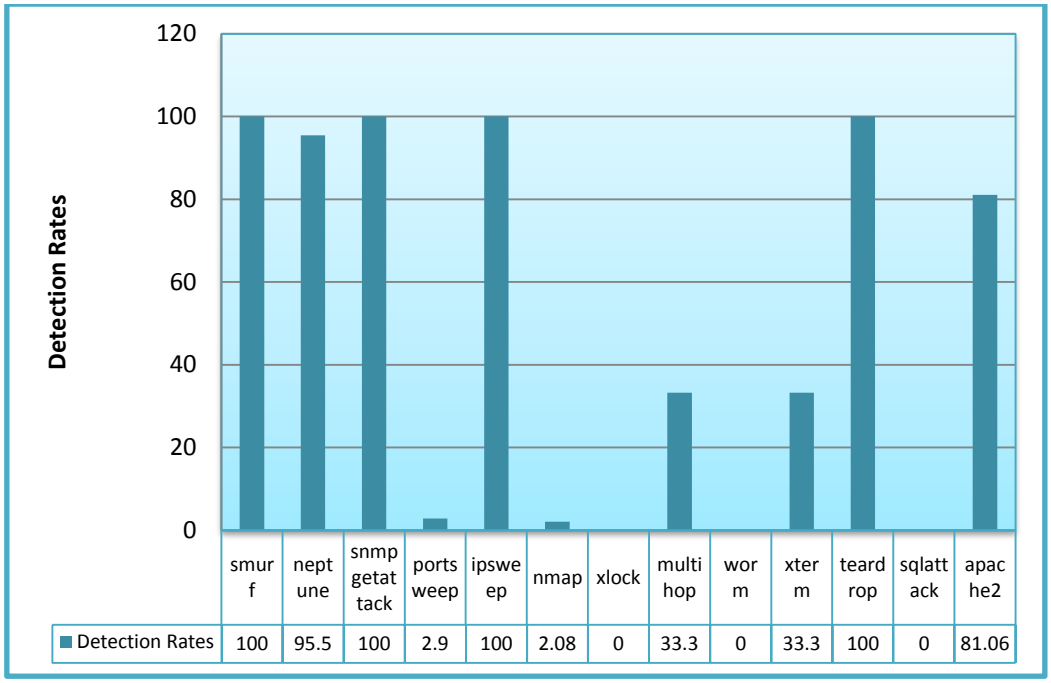


Fig. 5.5. Detection Rates of various attacks using Hybrid Approach and 350 rules with w1=0, w2=1

- **w1=0.4, w2=0.6:** Table 5.6 shows detection rates of hybrid approach using 350 rules and weights w1=0.4, w2=0.6. Only six attacks namely, smurf, neptune, portsweep, ipsweep, teardrop, apache2 are detected by using these weights. Figure 5.3 shows detection rate of various attacks using hybrid approach with 200 rules and w1=0.4, w2=0.6.

Table 5.6. Detection rates of Hybrid Approach Using 350 Rules and weights, w1=0.4 & w2=0.6

350 Rules		w1=0.4,w2=0.6	
Attack Names	Type of Attack	Detection Rate (%)	False Positive Rate (%)
smurf	DoS	100	0.30
normal	No Attack	89.6	
neptune	DoS	95.8	
snmpgetattack	R2L	0	
portsweep	Probe	2.9	
ipsweep	Probe	100	
nmap	Probe	0	
xlock	R2L	0	
multihop	R2L	0	
worm	R2L	0	
xterm	U2R	0	
teardrop	DoS	100	
sqlattack	U2R	0	
apache2	DoS	78.51	

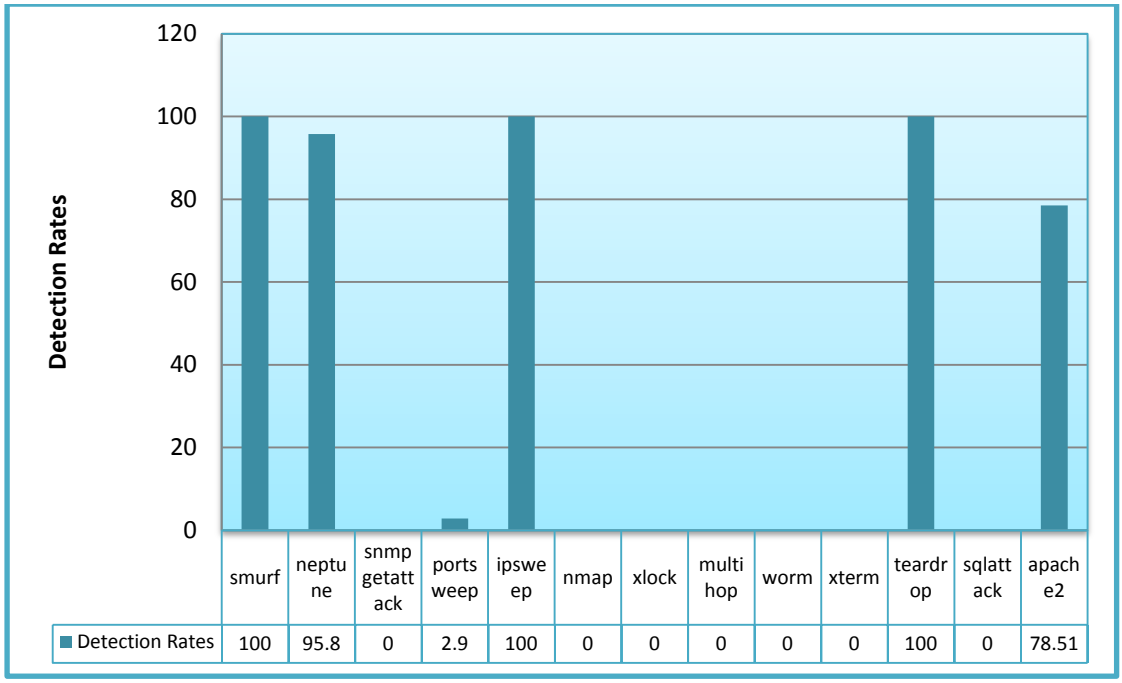


Fig. 5.6. Detection Rates of various attacks using Hybrid Approach and 350 rules with w1=0.4, w2=0.6

- **w1=1, w2=0:** Table 5.7 shows detection rates of hybrid approach using 350 rules and weights w1=1, w2=0. Ten different types of attacks namely, smurf, neptune, snmpgetattack, portsweep, ipsweep, nmap, multihop, xterm, teardrop and apache2 are detected by using these weights. Figure 5.7 shows detection rate of various attacks using hybrid approach with 200 rules and w1=1, w2=0.

Table 5.7. Detection rates of Hybrid Approach Using 350 Rules and weights, w1=1 & w2=0

350 Rules		w1=1,w2=0	
Attack Names	Type of Attack	Detection Rate (%)	False Positive Rate (%)
smurf	DoS	100	1.5
normal	No Attack	73.5	
neptune	DoS	19.9	
snmpgetattack	R2L	100	
portsweep	Probe	99.09	
ipsweep	Probe	100	
nmap	Probe	2.08	
xlock	R2L	0	
multihop	R2L	33.3	
worm	R2L	0	
xterm	U2R	33.3	
teardrop	DoS	100	
sqlattack	U2R	0	
apache2	DoS	81.07	

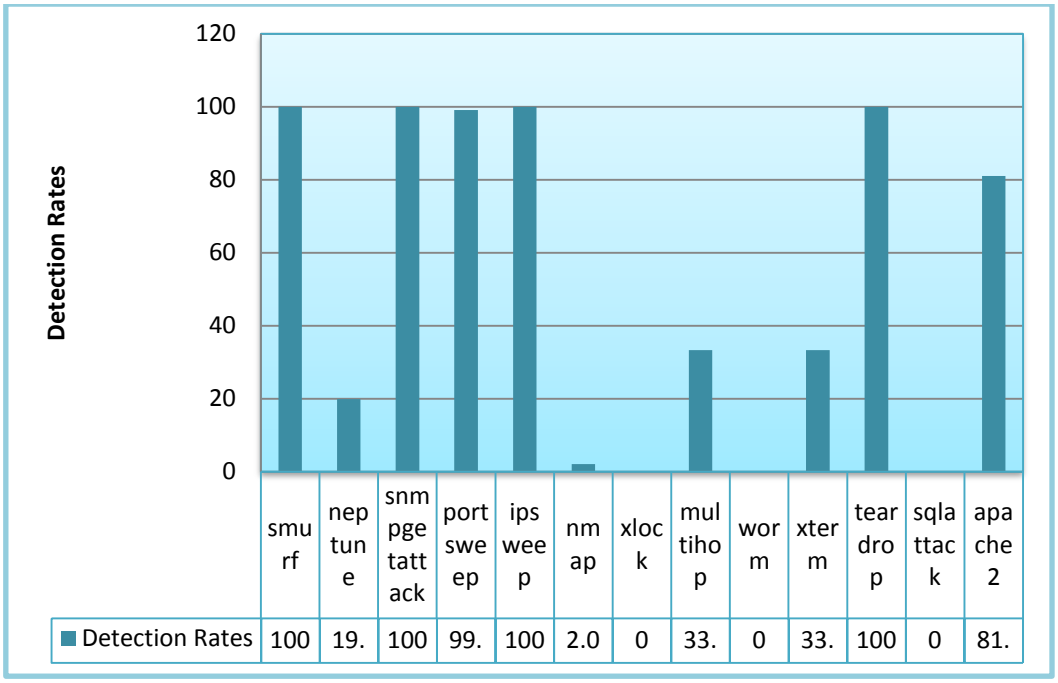


Fig. 5.7. Detection Rates of various attacks using Hybrid Approach and 350 rules with w1=1, w2=0

5.2.3 Comparison of False Positive Rates of 200 Rules with 350 Rules

As number of rules increases, detection rates increases but false positive rate also increases with increase in number of rules. Figure 5.8 shows false positive rates of 250 rules and 350 rules with different weights. As shown in figure, when number of rules increases, false positive rates also increases. After a certain limit, if we increase number of rules, it will have no impact on detection rates. So, it is better to have less number of rules for lower false positive rates.

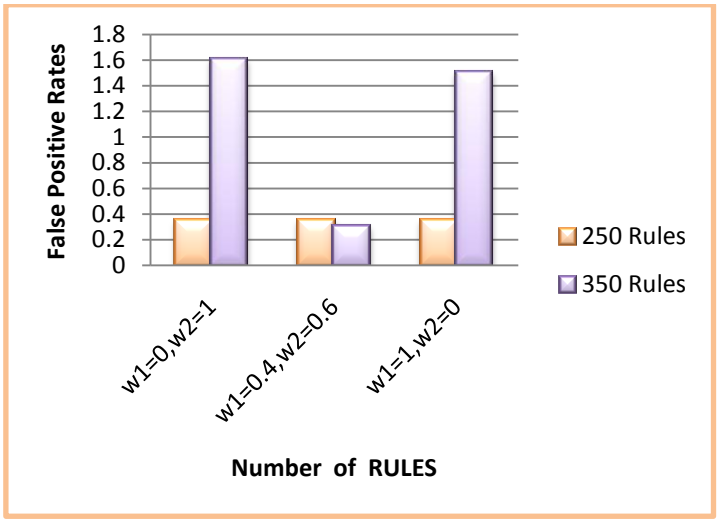


Fig 5.8 Comparison of false positive rates of 200 rules v/s 350 rules

5.3 Evaluation of Proposed Method

This section discusses the comparison of proposed method with misuse detection. The following parameters are chosen and their performance is evaluated.

- Detection rates of various attacks
- Number of attacks detected
- False Positive Rates
- Comparison of proposed method with related papers

- **Detection rates of various attacks**

Detection Rate of an attack is calculated as ratio between the number of correctly detected intrusions and total number of intrusions [37], as given in equation 5.1

$$Detection\ Rate = \frac{\#True\ Positive}{\#Intrusion\ in\ testing\ dataset} \tag{5.1}$$

Figure 5.9 shows comparison of detection rates of misuse approach and proposed approach. As we can easily analyze from figure 5.9 detection rates of proposed approach are very higher than misuse detection. Detection rates of neptune, snmpgetattack and ipsweep are higher in proposed approach. Along with these misuse approach is not able to detect portsweep, nmap, multihop, xterm and apache2.

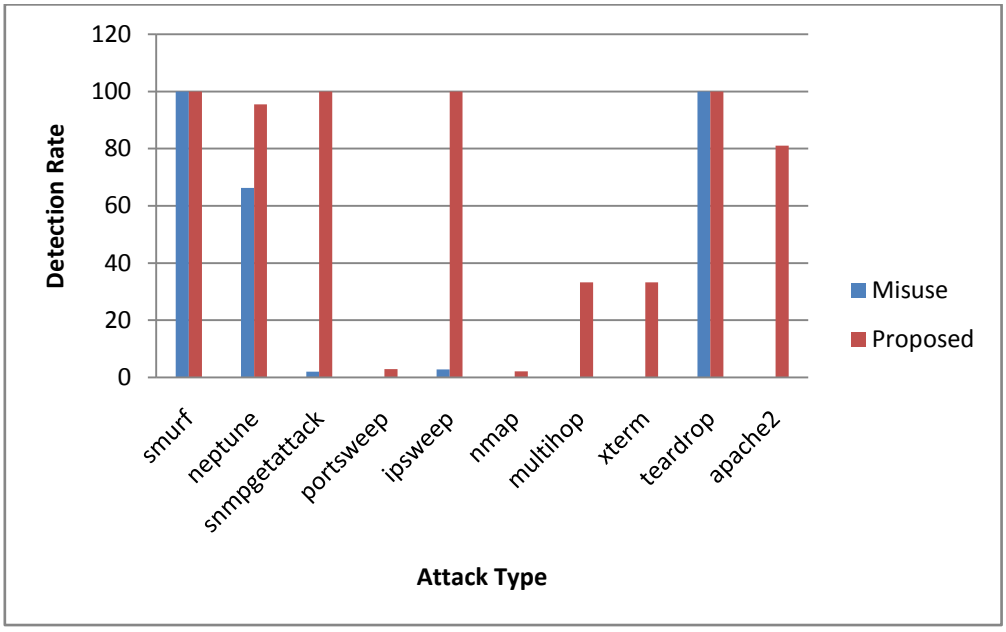


Fig. 5.9. Detection Rates of various attacks detected by misuse v/s proposed approach

- **Number of attacks detected by misuse v/s proposed**

The attacks detected by any intrusion detection system fall under four major categories namely, DoS, Probe, R2L, U2R. Number of attacks detected by misuse approach is five while proposed approach is able to detect ten types of attacks. Testing dataset contains fourteen types of classes in which thirteen classes are attacks and one class is normal. Table 5.8 shows the attacks detected by misuse and proposed approach. Figure 5.10 represents number of attacks detected by misuse and proposed approach. Proposed approach is able to find each type of attack while misuse is not able to find U2R attack. So, our approach is better than misuse approach.

Table 5.8. Attacks detected by misuse v/s proposed approach

	MISUSE	Proposed
DoS	**neptune, smurf, teardrop**	**neptune, smurf, teardrop, apache2**
Probe	**ipsweep**	**ipsweep, nmap, portsweep**
R2L	**snmpgetattack**	**snmpgetattack, multihop**
U2R	**--**	**xterm**

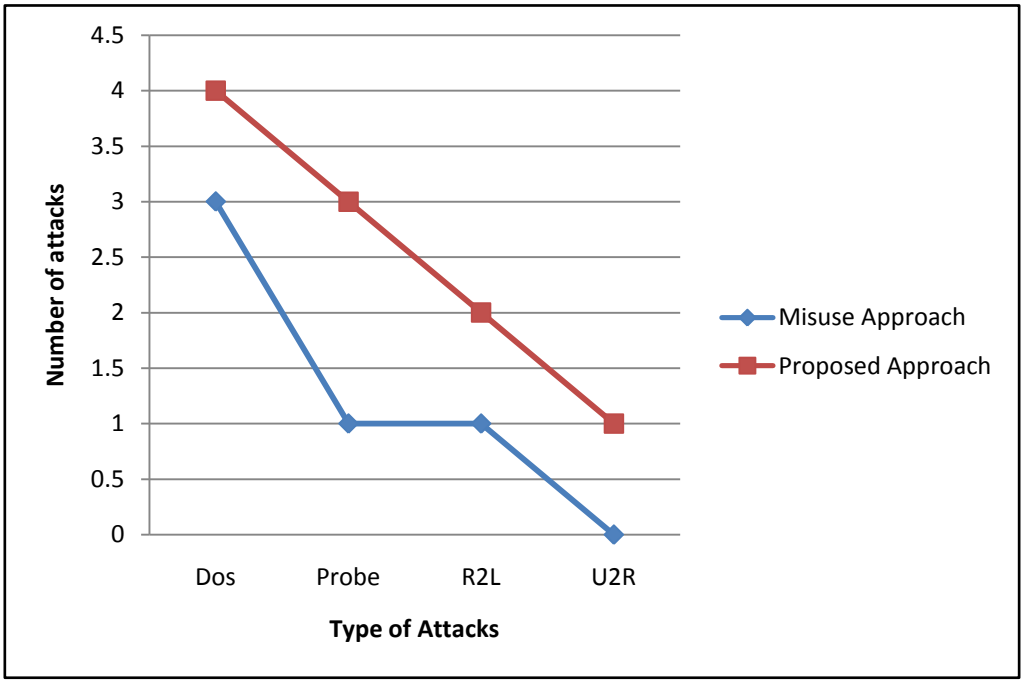


Fig. 5.10. Number of attacks detected by misuse v/s proposed approach

- **False positive rates of misuse v/s proposed**

False positive ratio is defined the proportion of normal connections which is falsely detected and labelled as an attack.

$$False\ positive\ rate = \frac{\#of\ normal\ connections\ detected\ as\ attack}{\#\ of\ normal\ connections} \quad (5.2)$$

False positive rate of misuse approach is 0.35 and our proposed system is able to achieve false positive rate of 0.30 with 350 rules and weights as w1=0, w2=1. Figure 5.11 shows false positive rates of misuse approach and proposed approach. Hence, this system can be used where high detection rates as well as low false positive rate is required. Generally, misuse approach has less false positive rate but our proposed system has relatively lesser false positive rates.

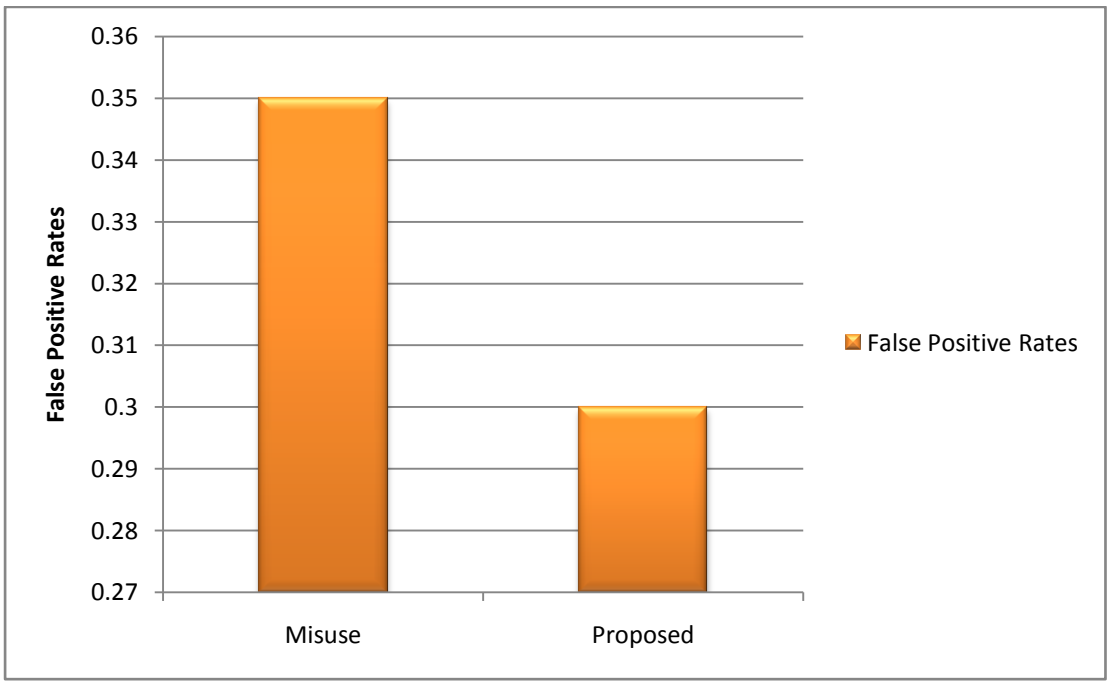


Fig. 5.11. False positive rates of misuse v/s proposed approach

- **Comparison of proposed method with existing techniques**

Various researchers have worked on genetic algorithm for improving detection rates and diminishing false positive rates. Fig 5.12 shows the false positive rates of proposed approach and approaches used by various researchers. Results of their work and proposed approach are compared and below graph shows that proposed approach gives lesser false positive rates.

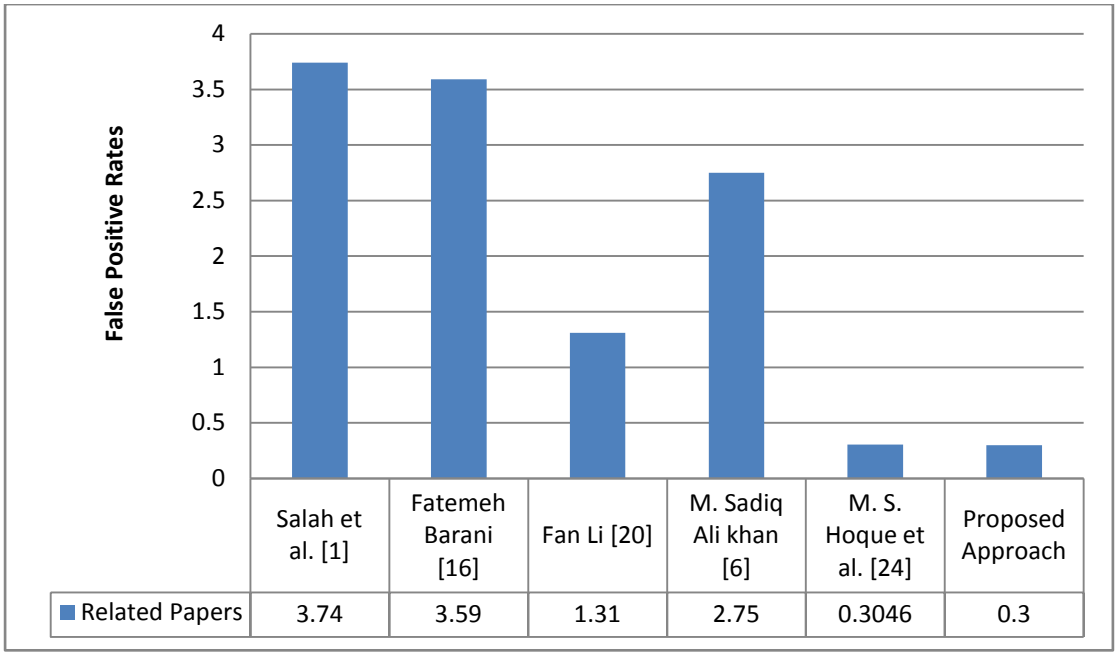


Fig. 5.12. False positive rates of various related papers and our approach

Chapter Summary: This chapter presents the analysis of results obtained by misuse detection approach as well as hybrid approach which is combination of misuse detection and Genetic Algorithm. This chapter also evaluates results of misuse detection and proposed approach on basis of detection rates, false positive rates and number of attacks detected by each approach.

Chapter 6
Conclusion and Future Scope

The research carried out in this thesis mainly emphasize on the need to protect the information systems from various kinds of threats. The increase use of internet and complex architecture of the networks made the system vulnerable to intrusions. The network can be secured by deploying security solutions like Antivirus, Firewall, Honeypot and IDS. In this thesis, hybrid approach with feature selection is implemented to detect intrusions. In this approach, combination of misuse detection and Genetic Algorithm is used. Software implementation of the system and class diagram of this approach is presented. Feature Selection was used to identify the most important features of network connections and Genetic Algorithm was used to derive best fit rules from a large population of rules.

Proposed system has ability to update new rules and it is easy to maintain. This system not only classifies connections as normal or intrusive but also finds the type of attack which is also very important. Classification of attack is also very important because recovery can't be made until we know the exact type of attack. Our system detects ten different types of attacks with only three features out of forty one. So, time complexity of proposed system is also less as compared to others. As one can see from results, proposed system has high detection rates than misuse detection. This system also gives less false positive rates compared to previous techniques used by researchers. Because of feature selection, time taken is less and speed is relatively high. So, it can be used for intrusion detection where high speed is required.

Improving detection rates in any intrusion detection system is a challenging task. In future, detection rates of proposed system may be enhanced by merging machine learning algorithm in this system.

REFERENCES

[1] Forouzan, Behrouz A., and Debdeep Mukhopadhyay-Cryptography. "Network Security." Data Communications and Networking: 961-962

[2] Jongsuebsuk, P., N. Wattanapongsakorn, and C. Charnsripinyo. "Real time intrusion detection with fuzzy genetic algorithm." In Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 10th International Conference on, pp. 1-6, IEEE, 2013.

[3] Aleksandar L., Vipin K., and Jaideep S., Massive Computing Managing Cyber Threats, Issues, Approaches, and Challenges. Chapter 2. Intrusion Detection: A Survey. Computers/General Information. Springer. 2005.

[4] Bankovic, Zorana, Dusan Stepanovic, Slobodan Bojanic, and Octavio Nieto Taladriz. "Improving network security using genetic algorithm approach."Computers & Electrical Engineering 33, no. 5 (2007): 438-451.

[5] Gong, Ren Hui, Mohammad Zulkernine, and Purang Abolmaesumi. "A software implementation of a genetic algorithm based approach to network intrusion detection." In Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS International Workshop on Self-Assembling Wireless Networks. SNPD/SAWN 2005. Sixth International Conference on, pp. 246-253. IEEE, 2005.

[6] Khan, M. Sadiq Ali. "Rule based network intrusion detection using genetic algorithm." International Journal of Computer Applications 18, no. 8 (2011): 26-29.

[7] Balajinath, B., and S. V. Raghavan. "Intrusion detection through learning behavior model." Computer Communications 24, no. 12 (2001): 1202-1212.

[8] Axelsson, Stefan. Intrusion detection systems: A survey and taxonomy. Vol. 99. Technical report, 2000.

[9] Kumar, Sandeep. "Classification and detection of computer intrusions." PhD diss., Purdue University, 1995.

[10] Lu, Wei, and Issa Traore. "Detecting new forms of network intrusion using genetic programming." Computational Intelligence 20, no. 3 (2004): 475-494.

[11] Kumar, Sandeep, and Eugene H. Spafford. "A software architecture to support misuse intrusion detection." (1995).

[12] John Henry Holland. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press, 1992.

[13] Garcia-Teodoro, Pedro, J. Diaz-Verdejo, Gabriel Macia-Fernandez, and Enrique Vazquez. "Anomaly-based network intrusion detection: Techniques, systems and challenges." Computers & Security 28, no. 1 (2009): 18-28.

[14] Pohlheim, Hartmut. "Genetic and evolutionary algorithms: Principles, methods and algorithms." (2006).

[15] Hashemi, V. Moraveji, Z. Muda, and W. Yassin. "Improving Intrusion Detection Using Genetic Algorithm." Information Technology Journal 12, no. 5 (2013).

[16] Barani, Fatemeh. "A hybrid approach for dynamic intrusion detection in ad hoc networks using genetic algorithm and artificial immune system." In Intelligent Systems (ICIS), 2014 Iranian Conference on, pp. 1-6. IEEE, 2014.

[17] Chang, Ning, Yujing He, Li Huifang, and Hui Ren. "A Study on GA-Based WWN Intrusion Detection." In Management and Service Science, 2009. MASS'09. International Conference on, pp. 1-4. IEEE, 2009.

[18] Padmadas, M., Nikhil Krishnan, J. Kanchana, and Madurakavi Karthikeyan. "Layered approach for intrusion detection systems based genetic algorithm." In Computational Intelligence and Computing Research (ICCIC), 2013 IEEE International Conference on, pp. 1-4. IEEE, 2013.

[19] Senthilnayaki, B., K. Venkatalakshmi, and Ajaykumar Kannan. "An intelligent intrusion detection system using genetic based feature selection and Modified J48 decision tree classifier." In Advanced Computing (ICoAC), 2013 Fifth International Conference on, pp. 1-7. IEEE, 2013.

[20] Fan, Li. "Hybrid neural network intrusion detection system using genetic algorithm." In Multimedia Technology (ICMT), 2010 International Conference on, pp. 1-4. IEEE, 2010.

[21] Wang, Yunwu. "Using Fuzzy Expert System Based on Genetic Algorithms for Intrusion Detection System." In Information Technology and Applications, 2009. IFITA'09. International Forum on, vol. 2, pp. 221-224. IEEE, 2009

[22] Kim, D.Seong, Nguyen, Ha-Nam and Park, Jong Sou. "Genetic algorithm to improve SVM based network intrusion detection system." In Advanced Information Networking and Applications, 2005. AINA 2005. 19th International Conference on, vol. 2, pp. 155-158. IEEE, 2005.

[23] Goyal, Anup and Kumar, Chetan. "GA-NIDS: a genetic algorithm based network intrusion detection system." Northwestern university (2008).

[24] Hoque, M. Sazzadul, Mukit, Bikas and Naser, Abu. "An implementation of intrusion detection system using genetic algorithm." arXiv preprint arXiv: 1204.1336 (2012).

[25] Aziz, Amira Sayed A., Mostafa Salama, Aboul Ella Hassanien, Sanaa EL-Ola Hanafi and M. F. Tolba. "Multi-layer hybrid machine learning techniques for anomalies detection and classification approach." In Hybrid Intelligent Systems (HIS), 2013 13th International Conference on, pp. 215-220. IEEE, 2013

[26] Lazarevic, Aleksandar, Vipin Kumar, and Jaideep Srivastava. "Intrusion detection: A survey." In *Managing Cyber Threats*, pp. 19-78. Springer US, 2005.

[27] Tutorial available on http://www.obitko.com/tutorials/genetic-algorithms/operators.php

[28] Tutorial available on http://geneticalgorithms.ai-depot.com/Tutorial/Overview.html

[29] Beasley, John E., and Paul C. Chu. "A genetic algorithm for the set covering problem." European Journal of Operational Research 94, no. 2 (1996): 392-404.

[30] Bishop, Christopher M. Pattern recognition and machine learning. Springer, 2006.

[31] Benaicha, Salah Eddine, Lalia Saoudi, Bouhouita Guermeche, and Ouarda Lounis. "Intrusion detection system using genetic algorithm." InScience and Information Conference (SAI), 2014, pp. 564-568. IEEE, 2014.

[32] Stolfo, Salvatore J., Wei Fan, Wenke Lee, Andreas Prodromidis, and Philip K. Chan. "Cost-based modeling for fraud and intrusion detection: Results from the JAM project." In DARPA Information Survivability Conference and Exposition, 2000. DISCEX'00. Proceedings, vol. 2, pp. 130-144. IEEE, 2000.

[33] Lippmann, Richard P., David J. Fried, Isaac Graf, Joshua W. Haines, Kristopher R. Kendall, David McClung, Dan Weber et al. "Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation." In DARPA Information Survivability Conference and Exposition, 2000. DISCEX'00. Proceedings, vol. 2, pp. 12-26. IEEE, 2000.

[34] Aziz, Amira Sayed A., Mostafa Salama, Aboul Ella Hassanien, and Sanaa EL-Ola Hanafi. "Detectors generation using genetic algorithm for a negative selection inspired anomaly network intrusion detection system." In Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on, pp. 597-602. IEEE, 2012.

[35] Anil, S., and R. Remya. "A hybrid method based on genetic algorithm, self-organised feature map, and support vector machine for better network anomaly detection." In Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on, pp. 1-5. IEEE, 2013.

[36] Makanju, Adetokunbo, Patrick LaRoche, and A. Nur Zincir-Heywood. "A Comparison Between Signature and Machine Learning Based Detectors."

[37] Folino, Gianluigi, Clara Pizzuti, and Giandomenico Spezzano. "GP ensemble for distributed intrusion detection systems." In Pattern Recognition and Data Mining, pp. 54-62. Springer Berlin Heidelberg, 2005.

[38] Kwok, Suk Wah, and Chris Carter. "Multiple decision trees." arXiv preprint arXiv:1304.2363 (2013).

[39] Aher, Sunita B., and L. M. R. J. Lobo. "A comparative study of association rule algorithms for course recommender system in e-learning." International Journal of Computer Applications 39, no. 1 (2012).

[40] Naoum, Reyadh Shaker, Namh Abdula Abid, and Zainab Namh Al-Sultani. "An Enhanced Resilient Backpropagation Artificial Neural Network for Intrusion Detection System." International Journal of Computer Science and Network Security 12, no. 3 (2012): 11-16.

[41] Costantini, Stefania. "Towards active logic programming." arXiv preprint arXiv:1403.5508 (2014).

[42] Mukkamala, Srinivas, Guadalupe Janoski, and Andrew Sung. "Intrusion detection using neural networks and support vector machines." In Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on, vol. 2, pp. 1702-1707. IEEE, 2002.

[43] Turgut, Damla, Sajal K. Das, Ramez Elmasri, and Begumhan Turgut. "Optimizing clustering algorithm in mobile ad hoc networks using genetic algorithmic approach." In Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE, vol. 1, pp. 62-66. IEEE, 2002.

[44] Sebyala, Abdallah Abbey, Temitope Olukemi, Lionel Sacks, and Dr Lionel Sacks. "Active platform security through intrusion detection using naive bayesian network for anomaly detection." In London Communications Symposium. 2002.

[45] Kotsiantis, Sotiris B., Ioannis D. Zaharakis, and Panayiotis E. Pintelas. "Machine learning: a review of classification and combining techniques."Artificial Intelligence Review 26, no. 3 (2006): 159-190.

Video Presentation

Video has been uploaded on YouTube and is available at the following link

https://youtu.be/Ud2WWQeDmgk

Research Publications

- Rohini Rajpal and Sanmeet Kaur, "A Hybrid Intrusion Detection Approach using Misuse Detection and Genetic Algorithm", International Conference on Signal Processing, 2015. [Status- Accepted]

- Rohini Rajpal, Ramandeep Kaur and Sanmeet Kaur, "Improving Detection Rate using Misuse Detection and Machine Learning", IEEE Security & Privacy. [Status- Communicated]