Adaptive Intrusion Detection Based on K-SVMeans Algorithm

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

Master of Engineering
in
Software Engineering

Submitted By
Parneet Kaur
(Roll No. 801131017)

Under the supervision of
Dr. V.P. Singh
Assistant Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR UNIVERSITY
PATIALA – 147004

July 2013

Certificate

I hereby certify that the work which is being presented in the thesis entitled, ***"Adaptive Distributed IDS based upon KSVM"***, in partial fulfilment of the requirements for the award of degree of Master of Engineering in *Software Engineering* 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 *VP Singh* 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.

Parneet Kaur

(Parneet Kaur)

(801131017)

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

(V P Singh)

Assistant Professor, CSED

Countersigned by

(Dr. Maninder Singh)

Head

Computer Science and Engineering Department

Thapar University

(Dr. S. K. Mohapatra)

Dean (Academic Affairs)

Thapar University

Patiala

Acknowledgement

I would like to express my deepest appreciation to Dr V P Singh, my mentor and thesis supervisor for his constant support and motivation. He had been instrumental in guiding me throughout the thesis with his valuable insights, constructive criticisms and interminable encouragement.

I am also thankful to Dr. Maninder Singh, Head, School of Mathematics and Computer Applications Department for his constant support and encouragement. I would like to thank all the faculty members and staff of the department who were always there at the need of the hour and provided all the help and facilities, which I required, for the completion of this work.

I offer my deepest gratitude to my family for their support and affection and for believing in me always. I also want to thank my colleagues, who have given me moral support and their relentless advice throughout the completion of this work.

At last but not the least I would like to thank ”The Creator of Destinies” for not letting me down at the time of crisis and showing me silver lining in the dark clouds.

Parneet Kaur
(801131017)

Abstract

Dependency of organisations and individuals on network based systems is growing day by day. The growth of complex computer networks augments the vulnerability of systems. This ever growing connectivity of systems gives more access to attackers and makes it even more difficult for security analysts to protect their system. Assuring secure and reliable operation of networks has become a priority research area these days. Protection techniques of network have not kept up with the increasing threat. Traditional defence mechanisms such as user authentication, data encryption, avoiding programming loopholes and firewalls are used as the first line of defence against attacks. Different types of counter measures are being devised every day. Intrusion detection system (IDS) is a relatively novel technology. Intrusion detection system identifies patterns of known intrusions (misuse detection) or differentiates anomalous network data from normal data (anomaly detection). The information collected by IDS is used for safeguarding the systems.

In this research work, a novel Intrusion Detection System (IDS) architecture is proposed. It includes both anomaly and misuse detection approaches. The framework of hybrid intrusion detection system has been proposed. The major emphasis is on the anomaly detection module of the IDS. This module implements a hybrid machine learning algorithm called *k*-support vector means clustering algorithm. The live network traffic used as an input for the algorithm is captured by Wireshark. The algorithm clusters the network traffic into normal and anomalous packets.

Table of Contents

List of Figures

Chapter 1
Introduction

1.1 Introduction to network security

Information stored on systems had to be protected since the very introduction of computers. The need for protecting files in computer systems became more evident with the advent of shared systems. The need became even more severe for systems being accessed over a public network, the Internet. Information in transit must be protected from unauthorized release and changes. The connection itself must be securely established and maintained. The rapid development of computer network technology and Internet has brought huge convenience to people. Internet does not conceal or protect the information completely as the Internet is open system for general public.

Due to recent advances in network technology, computer systems have become even more vulnerable to attacks. Novel attacks are appearing endlessly. In recent years, statistics have shown that number of reported intrusions in the Symantec Global Internet Security Threat Report is growing [1]. In Malaysia, the latest 6-month report for 2009 indicated a 100% increment on the number of reported cases. Moreover, in another continent, the Washington News revealed that a part of a $5.4 million contract was repaid to the Pentagon from a security company, Apptis Inc., after the company failed to provide adequate computer security services. In UK, a survey on UK businesses, conducted by PricewaterhouseCoopers between October 2007 and January 2008, reveals an increasing number of incidents, with 94% of very large companies encountered an incident and 76% of them had at least one serious incident. Above examples show that impact from attackers is inevitable.

Moreover, our dependency on network based systems is growing day by day. On the other hand, protection techniques of such systems have not kept up with the increasing threat. Traditional defence mechanisms such as user authentication, data encryption, avoiding programming loopholes and firewalls are used as the first line of defence against attacks. Till date no combination of technology can protect the system completely because systems face novel attacks every other day. Hence, researchers are now attempting to find better and safer methods to prevent, minimize and

overcome such incidents. In order to counter the problem of attacks, tools such as Firewall, Anti Virus, Intrusion Detection, Prevention, and Response Systems have emerged. Their aim is to monitor the system or network activity and detect, prevent, or counter suspicious incident. The technologies deployed by security managers to protect the enterprise, are useful for defending attacks to some extent only. They have their own limitations. For example, firewalls may be configured to block certain types of traffic, but attackers still find ways to exploit legitimate traffic types to mount their attacks. The following major security objectives of any application have paramount importance to ensure the security of network [2]:

- **Confidentiality:** It means that certain information is only accessible to those who have been authorized to access it.
- **Integrity:** It guarantees that a message being transferred is never corrupted.
- **Availability**: Availability guarantees the survivability of network services despite denial of service (DoS) attacks.
- **Authenticity:** Authenticity ensures that participants in communication are genuine and not impersonators.
- **Non-repudiation:** It ensures that the sender and the receiver of a message cannot deny that they have ever sent or received such a message.
- **Authorization**: It is a process in which a trusted certificate authority issues a credential to an entity.
- **Anonymity**: It means that all the information related to owner or current user entity identification should be kept secret and not distributed to other communicating parties [30].

1.2 Attacks in network

Attacks can be of two types:

1.2.1 Passive attacks

A passive attack [3] is the one in which the intruder eavesdrops or monitors the transmitted data but does not modify the message stream in any way. The goal of the

attacker is to get information in transit. Two types of passive attacks are release of message contents and traffic analysis. The release of message contents is simply reading the contents of a message. It can be troublesome if the message is carrying sensitive or confidential data. A second type of passive attack, traffic analysis is subtler. An intruder makes inferences by observing message patterns. It can be done even if messages are encrypted. In this attack, the eavesdropper analyzes the traffic, determines the location and identifies communicating hosts. Eavesdropper can also observe the frequency and length of message being exchanged. This information is used to predict the nature of communication. All incoming and outgoing traffic of network is analysed but not altered. Passive attacks are very difficult to detect because they do alter the data. Neither the sender nor the receiver is aware that a third party has read the messages during their exchange. This can be prevented by means of encryption of data. Encryption is the technique for masking the contents of message. If one had encryption protection done, an opponent can still observe the pattern of this message [6]. The opponent could determine the location and identity of communicating hosts and could observe the frequency and length of messages being exchanged. The information might be useful in guessing the nature of the communication that was taking place. Thus, the emphasis in dealing with passive attacks is on prevention rather than detection.

1.2.2 Active attacks

Active attacks [3] involve some modification of data stream or creation of a false stream. An active attack is one in which the intruder may send messages, replay old messages, change messages in wire, or delete selected messages in transit [7]. A typical active attack is one in which an intruder pretends to be one end of the conversation, or acts as a man-in-the-middle. They can be subdivided into four categories:

- **Masquerade:** A masquerade takes place when one entity pretends to be a different one. This attack generally includes one of the other forms of active attacks. For instance, authentication sequences can be captured and replayed after a valid authentication sequence has taken place. This enables an authorized entity with few privileges to obtain extra privileges by impersonating an entity that has those privileges [4].

- **Replay:** Replay [4] involves the passive capture of a data unit and its subsequent retransmission to produce an unauthorized access.

- **Modification of messages:** Modification of messages means that some part of a legitimate message is altered or that messages are delayed or reordered, to produce an unauthorized effect. For example, a message meaning "Allow John Smith to read confidential files account" is modified to mean "Allow Fred Brown to read confidential file accounts" [4].

- **Denial of service:** The denial of service prevents or inhibits the normal use of communication devices. This attack may have a specific target, for instance, an entity may suppress all messages meant for a particular destination. Another form of this attack is disruption of the entire network. This is done either by crippling the network or by overloading it with messages in order to degrade its performance [4].

1.3 Threats to Network Security

Threats can be defined as potential violations of security. They exist because of vulnerabilities or weaknesses in a system. Basically, there are two types of threats: accidental threats and intentional threats [5].

1.3.1 Accidental Threats

Accidental threats result in either the exposure of confidential information or occurrence of an illegal system state. Exposures can appear from both hardware and software failures as well as from user and operational mistakes. This results in the violation of confidentiality. It can also be demonstrated as modification of an object that is the violation of object integrity.

1.3.2 Intentional Threats

Intentional threats can be defined as an action deliberately performed by an entity with the intention of violating security. Examples of such attacks are modification, interception, interruption and fabrication of data as shown in Figure 1 [5].

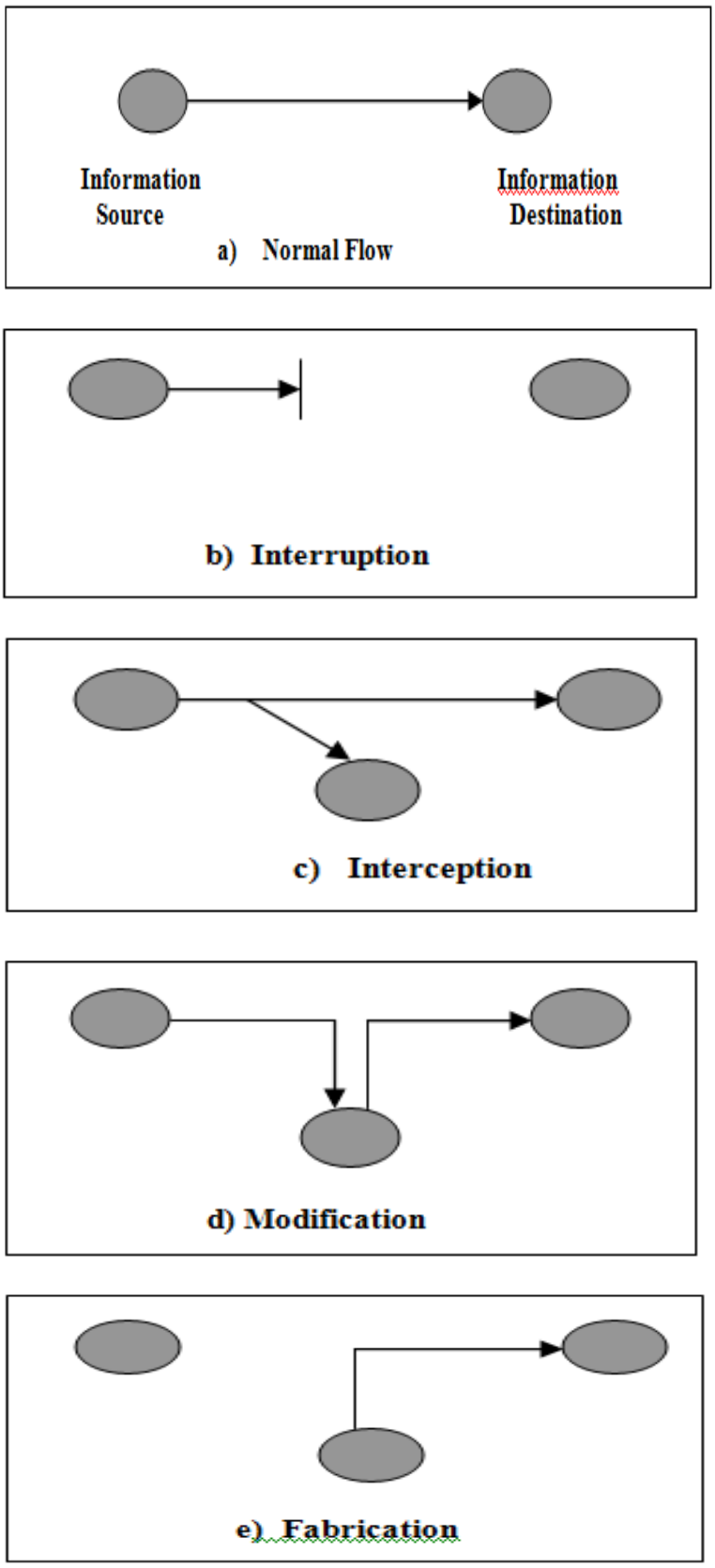

Figure1: Intentional Threats [5]

There are numerous threats to network security. Some of them are as listed below:

- **SYN flooding:** It is a denial-of-service attack in which a large amount of SYN packets are sent to a network or a server. The attack packets usually have spoofed source addresses to hide the real attacking sources and this makes defence much harder. The SYN flooding attacks exploit the TCP's three-way handshake mechanism and its limitation in maintaining half-open connections. When a server receives a SYN request, it returns a SYN/ACK packet to the client. Until the SYN/ACK packet is acknowledged by the client, the connection remains in half open state for a period of up to the TCP connection timeout, which is typically set to 75 seconds. The server has built in its system memory a backlog queue to maintain all half-open connections. Since this backlog queue is of finite size, once the backlog queue limit is reached, all connection requests will be dropped [6].

- **Packet sniffers: P**acket sniffing [7] is a method of tapping each packet flowing across the network. It is a technique in which a user sniffs data belonging to other users of the network. Packet sniffers can be used as an administrative tool or for malicious purposes. Network administrators operate them for monitoring and validating network traffic.

- **Viruses:** A small piece of code that recursively replicates a possibly evolved copy of itself on real programs. They run every time a program runs and multiply to form new generations. Most of them can reproduce and attack other programs. Trojan horse is a program containing hidden functions that can exploit the privileges of the user running the program. It can erase important information, send credit card numbers and password to the intruder [8].

- **Spyware**: Spyware is a new type of potentially unwanted program whose goal is to monitor users' online behaviour without user consent. Users infected with spyware generally experience highly degraded reliability and performance such as increased boot time, unresponsive system, and frequent application crashes [9].

1.4 Network security tools

Information security managers have used multiple technologies to keep their network safe from intrusions. However, as an effect of the improvements in technology, networks are now connected to other outside networks – including the Internet. So the corporations face a wide range of threats. So, security managers are under a lot of pressure to prevent any penetration to the network perimeter. Similar to the physical security, there are numerous security tools to help security managers in setting up complex protection strategy plans for their computer systems. Most common ones are listed below:

1.5.1 Firewalls

Firewall [10] is a hardware or software solution implemented within the network to enforce security policies by controlling network access. The original function of firewall was protecting a network from unauthorized external access. Now firewalls can also inspect and filter traffic arriving or departing a network by comparing packets to a set of rules and performing the matching rule action, which is accept or deny. A firewall is often seen as the first step toward a network security solution. Firewalls must be installed at the choke points to control network traffic and implement network security policy of the organization for its external network connections, especially for the Internet. Because many Internet-based services are inherently insecure, a firewall is needed to disable some services according to the organizational security policy. A firewall can act as a wall between the two networks. The person want access to the either network has to pass this wall before entering. A firewall is a system that is set up to control traffic flow between two networks. There are various firewall products but they are grouped into three major types based on their mechanisms:

- **Packet filtering**: Packet filtering [11] is a mechanism that controls the flow of packets in a network by examining their headers. No content-based decisions are made. The decision is exclusively based on the packet headers which include type of traffic (such as TCP, UDP, ICMP), characteristics of the transport layer communications sessions (such as source and destination ports), source and destination address. Packet filters are coupled with interfaces and the packet that flows through the interface can be restricted.

Packet filter firewall may have rules such as: allowing certain hosts send email via Simple Mail Transfer Protocol or not permitting any outside system to connect to an internal host via Telnet.

- **Stateful inspection:** This technology has evolved from the need to accommodate some features of the TCP/IP protocol suite. Stateful inspection firewalls are packet filter firewalls with the ability of connection status awareness. This awareness is made by making a dynamic list of active connections between hosts which is called state table. Those packets are rejected by the firewall, which do not belong to an active connection or is not a connection request. A packet belonging to an active connection is allowed through bypassing the firewall rules and thus optimizing the investigation process [11].
- **Proxying:** It is a mechanism that provides all internal hosts the untrusted external network access while it appears that a single host is accessing outside. Since all connections to the external network is made by a single host, deep packet inspection can be done before passing packets to internal nodes. Proxying inspects source address, destination address, protocol, source port number, destination port number and payload of packets [12].

1.5.2 Intrusion Detection System

Intrusion detection system (IDS) is the second layer of the perimeter defence [13]. Its purpose is to detect both external attacks and internal misuse of computer and network resources. IDS protects the internal as well as external network from outside attack .In physical analogy, an IDS is equivalent to a video camera and motion sensor detecting unauthorized or suspicious activity and working with automated response [14]. These devices like firewalls inspect incoming and outgoing network traffic. Unlike firewalls they do not alter the traffic flow by dropping or passing certain packets. In fact they look for malicious traffic indicating an attack. IDS can be categorized into two general types known as anomaly detection and misuse detection. The main challenge in intrusion detection is that of separating anomalous events from normal events.

A perfect IDS must satisfy following two criteria:

- It must be able to identify intrusions correctly.

- It must not label legitimate action as an intrusion.

1.5.3 Honeypots

Honeypots [15] are highly flexible security tools with several applications. Honeypots do not solve a specific problem unlike Firewalls or IDS. Instead, they have manifold uses: prevention, detection and information collecting. Honeypots collect small data sets having high value. Theoretically, they should see no traffic because it has no legitimate activity. So is any interaction with a honeypot is most probably unauthorized or malicious activity. Similarly, any connection attempts made to a honeypot are mostly an attack or compromise. Honeypot is a security resource whose importance lies in being attacked or compromised. Honeypots in a network should not affect critical network services and applications. Those series of characteristics distinguish honeypots clearly of other solutions of security. Honeypots only gather bad activity. It is much easier to analyze the data a honeypots collects and even easier to derive a value from it [16]. They are designed for:

- Diverting an attacker from accessing crucial systems.
- Capturing information about activity of the attacker's.
- Encouraging the attacker to remain on the system long enough so that the security managers can respond back.

1.5.4 Network-based Antivirus systems

Network-based antivirus [17] systems are solutions that are installed on a gateway between two networks to prevent the spreading of viruses across the network. The restraints of host-based antivirus software demonstrate the need for properly implemented network-based antivirus systems that allow security managers to deploy comprehensive antivirus protection faster. They also guard against the rise of threats endangering the networks as they spread by exploiting known vulnerabilities. Network-based antivirus systems are installed in the Demilitarized Zone (DMZ) in order to capture incoming and outgoing packet and compare them with the contents to a database of known virus signatures. Network-based antivirus systems have to operate under much more difficult constraints in comparison to host-based antivirus systems. Files are transported over networks in the payload portions of packets. Each packet contains only a small part of the file. A characteristic packet payload on the

Internet consists of approximately 1,500 bytes. Howsoever, many viruses are substantially longer than 1500 bytes. They can exceed 100K bytes in length. Therefore it is not sufficient for network-based antivirus systems to simply scan each packet individually. A packet-by-packet scan will never detect it if a virus is longer than 1500 bytes and the signature for the virus depend on patterns occurring in portions of the packet that are separated by more than 1500 bytes.

1.6 Thesis Organisation

The first chapter briefly describes the background study and introduces the motivation behind the proposition of system and organisation of whole thesis.

The second chapter describes the literature survey of the topic under consideration. It deals with the basis of problem statement.

The third chapter defines the problem statement. It also analyses gaps present in the existing systems.

The fourth chapter gives the problem solution. It gives details of the proposed framework of Intrusion Detection System. A hybrid algorithm called k-means clustering Support Vector Machine is also proposed in this chapter.

The fifth chapter deals with implementation and results. The algorithm proposed in fourth chapter is implemented in this chapter. Its output is also provided.

The sixth chapter reflects upon the conclusion and future scope of the thesis report. And at last, list of references and papers published is given.

Chapter 2
Literature Survey

2.1 Intrusion Detection System

The propagation of e-commerce applications and the increasingly important role that networks play in modern business, has given new stimulus to the search for developing more secure systems. Assuring secure and reliable operation of networks has become a priority research area these days because of ever growing dependency on network technology. Intrusion detection products are gaining extensive recognition as important tools for enhancing the security of a computer network. Although firewalls have traditionally been seen as the "first line of defence" against would be intruders, intrusion detection software is rapidly becoming popular as a novel but effective approach for making networks more secure [50]. An IDS identifies patterns of known intrusions (misuse detection) or differentiates anomalous network data from normal data (anomaly detection Intrusion detection operates on the principle that any attempt to penetrate your systems can be detected and the operator alerted - rather than actually stopping them from happening. Breaking into a computer system is more often a computer user's hacker fantasy than a system administrator's reality. But, when asked, any system administrator will tell you that there's always a lingering uncertainty after all the security policies have been implemented and adhered to. So it is critical to protect the networks from attackers and the intrusion detection technology has become popular. Intrusion Detection System is a new safeguard technology for system security after traditional technologies, such as firewall, authentication, message encryption and so on. Intrusion detection is defined as the processes to identify the internal or external users who intend to do something unauthorized against the computer system. Intrusion detection also identifies the legal connected users who intend to misuse their privileges. Intrusion detection system can analyze and monitor customer and system activity, identify and reflect the activity patterns attacks activity patterns that have known by Management personnel. Intrusion detection systems (IDS) are based on the principle that malicious behaviours on computer or network systems will be noticeably different from normal behaviours.

The IDS receives and analyzes many data sources from computer systems or networks to detect abnormal patterns generated by the intruders who intend to attack or penetrate the computer and network system. The general IDSs should have the ability to detect unauthorized access/modification of system or user information/files, network component information and unauthorized use of system resources [18].

Intrusions in an information system are the activities that violate the security policy of the system. Intrusion detection is the process used to identify those intrusions. Intrusion detection has been studied for approximately twenty years. It is based on the belief that an intruder's behaviour will be noticeably different from that of a legitimate user and that many unauthorized actions will be detectable [19]. There are many reasons that make intrusion detection a necessary part of the entire defence system. First, many traditional systems and applications were developed without security in mind. In other cases, systems and applications were developed to work in a different environment and may become vulnerable when deployed in the current environment. For example, a system may be perfectly secure when it is isolated but become vulnerable when it is connected to the Internet. Second, due to the limitations of information security and software engineering practice, computer systems and applications may have design flaws or bugs that could be used by an intruder to attack the systems. As a consequence, some preventive mechanisms may not be as effective as expected. Intrusion detection complements these protective mechanisms and improves the system security [18].

Different IDS techniques and architectures have been developed. Most of the present IDS are centralised. According to Jianxiao and Lijuan, the individual monitors send intrusion data to the centralized controller component that performs analysis of the information received from each of the monitors. The problem is that single host is available to copy the collected information. So the monitored network is restricted and lots of data collection can overload the network [41]. According to Huang, Distributed Intrusion Detection System collects information on several key points of the computer network or computer system and analyzes this information. Distributed Intrusion Detection System based on multi-agent technology can effectively improve the detection accuracy and detection speed, and enhance the system's own security [42]. Li Tian made the system intelligent by gathering information using honeypot network and information processing using mobile agent [43]. Zhang combined the advantages of agent-based distributed analysis and clustering-based intrusion

detection technique. Reh´ak [44] introduced a prototype of agent-based IDS designed for deployment on high-speed backbone networks.

2.2 Structure and architecture of IDS:

Various components of general IDS as shown in figure 2.1 are [20]:

2.2.1 Data gathering device

This device is responsible for capturing data from system under monitoring. It acts as an agent continuously monitoring the network in real time. Inputs to a sensor are network packets, system call traces and log files. Sensors gather and forward this information to the analyzer.

2.2.2 Detector

Detector processes the data collected from sensors and identifies intrusive activities. It uses a database of rules to generate alarms from security events received.

2.2.3 Knowledge base

This module contains information collected from sensors in pre processed format. This kind of information is mostly provided by security experts. Database contains all the information related to the signatures or patterns of attacks previously detected. When the sensor detects some kind of malicious activity, it matches it with the current database and reports to the attack response module.

2.2.4 Configuration device

Configuration device supplies information about the present state of the Intrusion Detection System.

2.2.5 Response component

This component initiates actions when an intrusion is detected. These responses can either be automated or involve human interaction. Based upon the type of configuration the Response Module can either send an alarm or an email notification to the administrator about the intrusion detected.

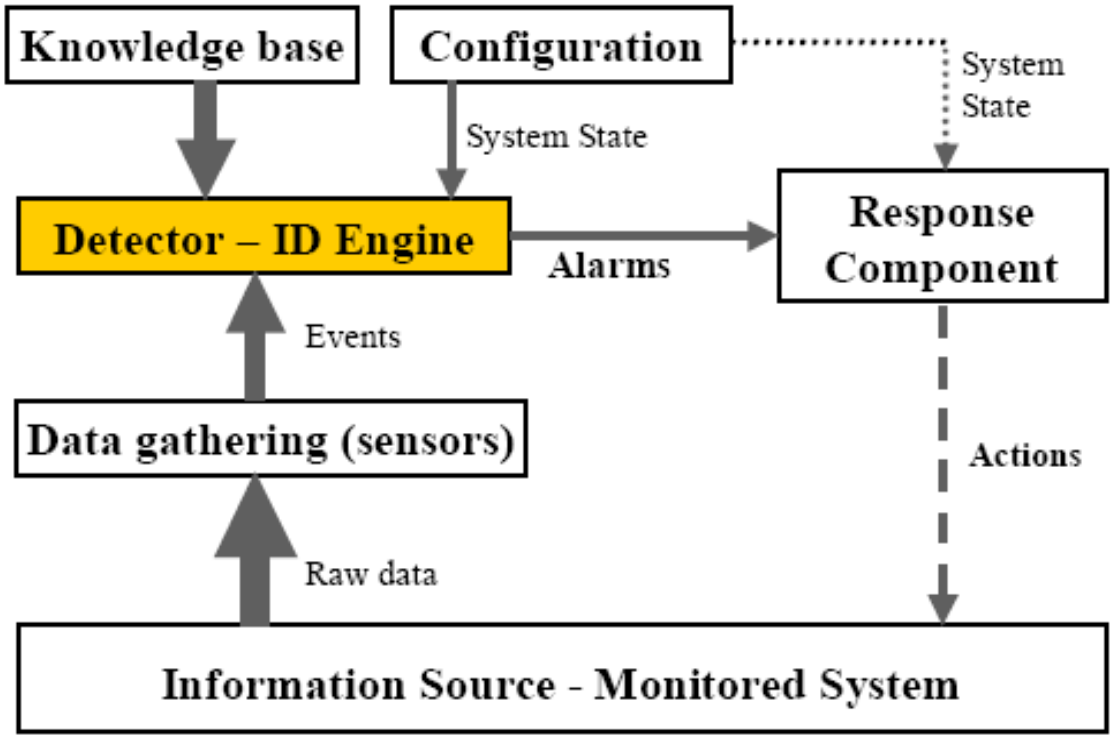

Figure 2.1: Basic Architecture of IDS [20]

2.3 Types of IDS

Many classifications of intrusion detection methods have been proposed in the past but no universally accepted taxonomy is available. A taxonomy [22] that is based on the synthesis of a number of existing ones is summarized in Fig. 2.2. The first criterion is information (data) source, which distinguishes IDSs based on the system that is monitored, i.e. source of input information. The source information can be

- audit trails (e.g. system logs) on a host,
- network connections/packets,
- application logs,
- wireless network traffic or
- Intrusion detection and/or sensor alerts produced by other intrusion detection systems.

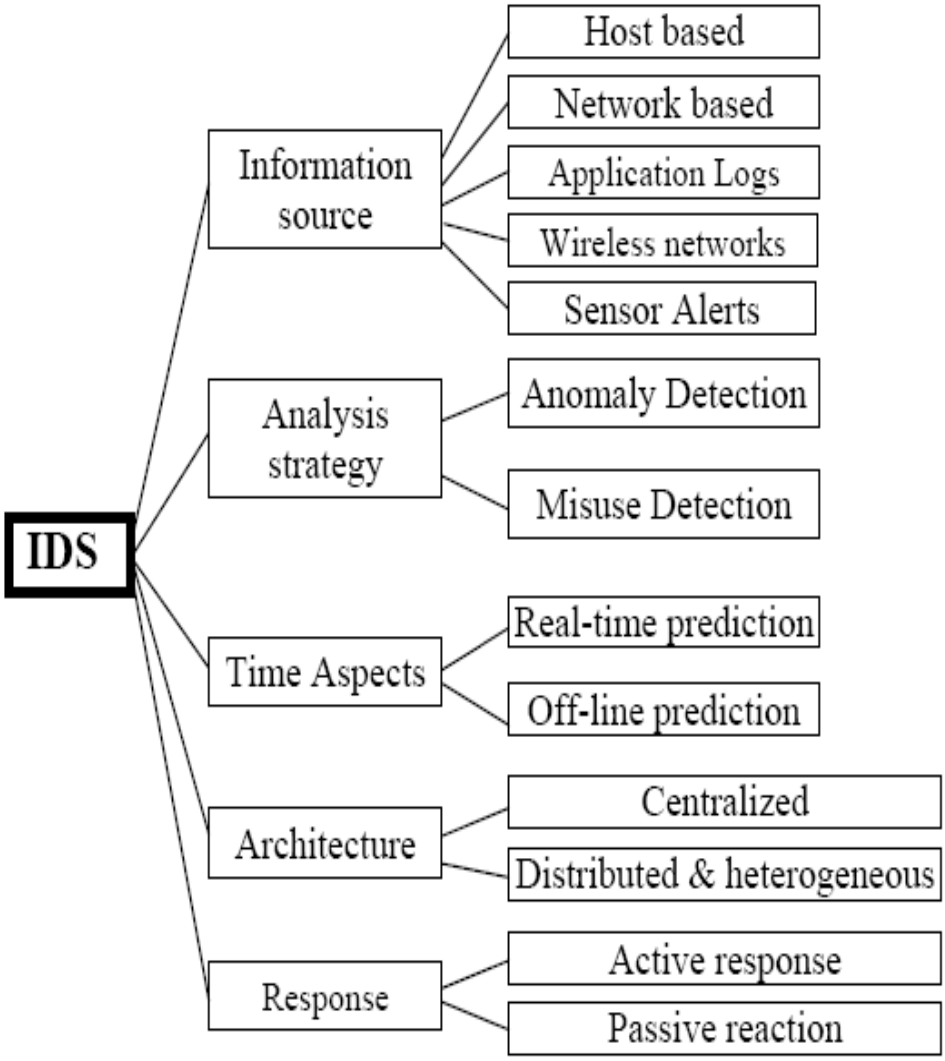

Figure 2.2: Taxonomy of IDS [21]

Traditionally IDS is classified into Misuse and Anomaly IDS. However there are other classifications as well.

2.3.1 Signature or Misuse detection

Misuse detection catches intrusions based on knowledge of known attack patterns, while anomaly detection detects intrusion based on deviation from normal patterns. They use patterns of already known attacks of the system to identify unknown intrusions. They compare audit data to attack patterns learned from the training data as shown in Figure 2.3. The observed data is considered intrusive if the sensor data matches the pattern of some known attack data. Misuse models are typically obtained

by training on a large set of data in which the attacks have been manually labelled. This kind of data is very expensive to produce because each piece of data must be labelled as either normal or some particular attack. Signature detection is the process of looking into network traffic for malicious bytes or packet sequences. The main advantage of this detection method is that signatures are easy to develop and understand if the network behaviour which has to be identified is already known. For instance, particular strings within an exploit payload to detect attacks that are attempting to exploit particular buffer-overflow vulnerability. A signature based IDS generate events which can communicate the cause of the alert. Pattern matching can also be performed very quickly on modern systems so the amount of power needed to perform these checks is very less for a confined rule set. Signature detection engines also have some disadvantages. A signature must be created for every attack, and novel attacks cannot be detected because they are able to detect known attacks only. While signature engines can detect attacks with a fixed behavioural pattern, they do not work well against the attack patterns created by a human or a worm with self-modifying behavioural characteristics. Detection is made even harder by advancing exploit technology that enables malicious users to hide their attacks behind nop generators, encrypted data channels and payload encoders. Another disadvantage is that signature engines often generate false positives since they are based on regular expressions and string matching.

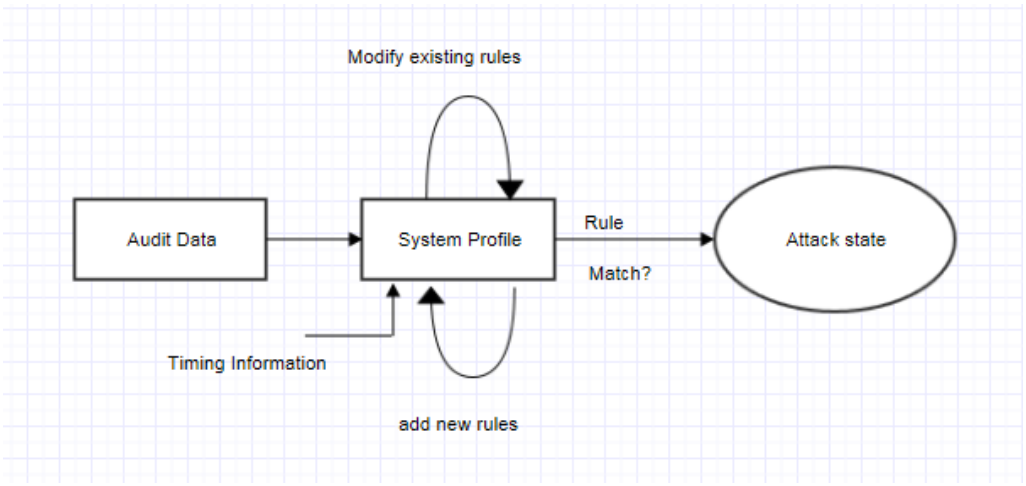

Figure2.3: Misuse Detection System [22]

2.3.2 Anomaly detection

The anomaly detection technique works on the basis of concept of a baseline for network behaviour. The baseline is description of accepted network behaviour learned or specified by the network administrators. Anomaly detection algorithms model normal behaviour. Anomaly detection models compare audit data to system profile learned from the training data as shown in Figure 2.4. If the sensor data deviates from normal behaviour, the anomaly detection model classifies the data as malicious. Events in an anomaly detection engine are generated by behaviour falling outside that predefined model of behaviour. One of the disadvantages of anomaly-detection engines is the difficultly of defining rules. The protocol must be defined, executed and tested for correctness. The process of rule development is further compounded by differences in vendor implementations of the various protocols. Moreover, detailed knowledge of normal network behaviour must be fed into the engine memory for attacks to be detected correctly. But once a protocol has been built and the behaviour has been defined, the engine can scale more quickly and easily than the signature-based model. This is because a new signature need not be created for every attack and potential variant. Another drawback of anomaly detection is that malicious activity falling within normal usage patterns is not detected. For instance, an activity of directory traversal on a targeted vulnerable server complying with network protocol can easily go unnoticed. However, Anomaly detection techniques are effective against unknown or novel attacks since no prior knowledge about specific intrusions is required. Misuse detection model generates less false positive alarms and introduce little overhead into the system by detecting only the intrusions having signatures.

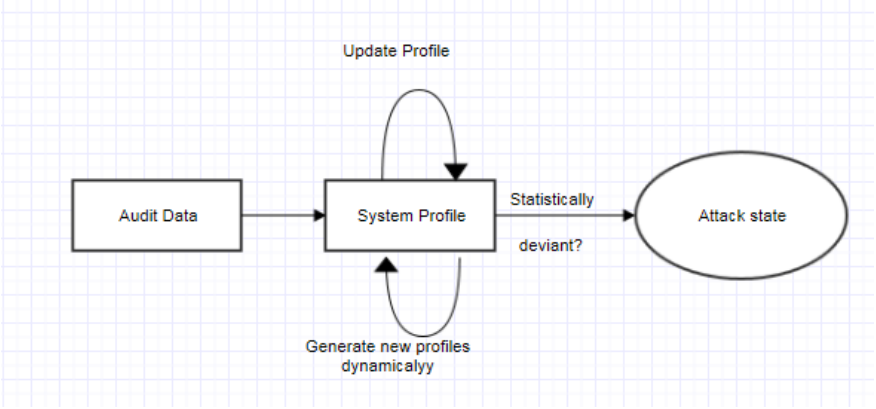

Figure2.4: Anomaly Detection System [22]

2.3.3 Host based IDS

Host based IDS examines the data from a single host. Host based intrusion detection tries to recognise unauthorized, illegitimate, and malicious behaviour on a specific device. HIDS usually involves an agent installed on each system which monitors and alerts on local OS and application activity. This installed agent uses signatures and rules to discover unauthorized activity. Host IDS only collects, identifies and alerts the system, so its role is only passive. Examples include ARMD, MIDAS, Tripwire [22].

2.3.4 Network based IDS

Network based IDS scrutinize network traffic and data from the connected hosts. Network based intrusion detection attempts to identify unauthorized, illegitimate and malicious behaviour based solely on network traffic. A network ID, using either a network Tap (Test Access Port), span port or hub collects packets that flow in a given network. Using the gathered data, the IDS system processes and flags any suspicious traffic. Examples include ASIM, Bro, CyberCop, EMERALD, GRIDS, INBOUNDS, NADIR, RealSecure, UNICORN.

2.3.5 Off-line IDS

Offline IDS examines system logs at fixed time intervals and reports any suspicious activity logged during inspection. It may not alert the administrator at the time of attack but can prevent future attacks by updating the rules in database. Examples are ASIM, NADIR, Stalker, Tripwire.

2.3.6 Real-time IDS

This type of IDS monitors the system continuously and reports suspicious activity as soon as it is detected. Real-time alerts the administrator at the time of attack so that a proper counter action can be taken. So it is beneficial than Offline IDS. Examples are AAFID, ARMD, Bro, CMDS, CyberCop, DIDS, EMERALD, INBOUNDS, NIDES, UNICORN.

2.3.7 Centralized IDS

IDS is centralized if intrusion data is collected from different hosts or networks and is passed on to a centralized controller component that scrutinise the information that it receives from each of the monitors. All the data is shipped to a central location for analysis. Examples are ARMD, Bro, CMDS, CSM, CyberCop, DIDS, NADIR, NIDES, Stalker and UNICORN.

2.3.8 Distributed IDS

Most of the current IDSs used are distributed ones because Host-based or network-based Intrusion Detection System is almost powerless for complex attacks. The main issue of this kind of system is that it can't identify novel attacks because it is signature based IDS which can only identify well known attack patterns. Data mining methods are used to automate the intrusion detection systems to identify novel attacks. Most popular way to identify intrusions is by studying the audit data produced by Operating System. Normal system activities are characterised with a profile, which is made by applying mining algorithms to audit data. Abnormal intrusive activities are identified by comparing the current activities with the profile. So another kind of IDS comes into picture called Adaptive IDSs which are based on data mining algorithms.

2.4 Adaptive IDS

In today's dynamic threat and networking environments, Signature based Intrusion Detection/Prevention Systems are unable to protect against ever-changing attacks and vulnerabilities. The reason is everyday new attacks are developed. We need to develop the system in such a way that it can learn new attacks itself. We have adaptive model for this purpose. It is a method for automatically building detection models for data-mining based intrusion detection systems. This significantly reduces the deployment cost of an intrusion detection system. In literature, a number of anomaly detection systems are developed based on many different machine learning techniques. For instance, some studies apply single learning techniques like neural networks, genetic algorithms, support vector machines, etc. Some systems are based on the combination of different learning techniques like hybrid or ensemble techniques. These techniques are developed as classifiers, which are used to classify

or identify whether the incoming Internet access is the normal access or an attack [22].

2.4.1 Neural Networks

The Multilayer Perceptions neural networks have been very successful in a variety of applications, producing results which often exceed other existing computational learning models. They are capable of approximating any continuous function to arbitrary precision, as long as they contain enough hidden units. This shows that such models can form any classification decision boundary in feature space and can act as non-linear discriminate function. When neural network is used for classification of pattern, there is one input node for each element of the feature vector. There is usually one output node for each class to which a feature may be assigned as shown in figure 2.5. The hidden nodes enable internal representation of the data to be developed by the NN during learning [14].

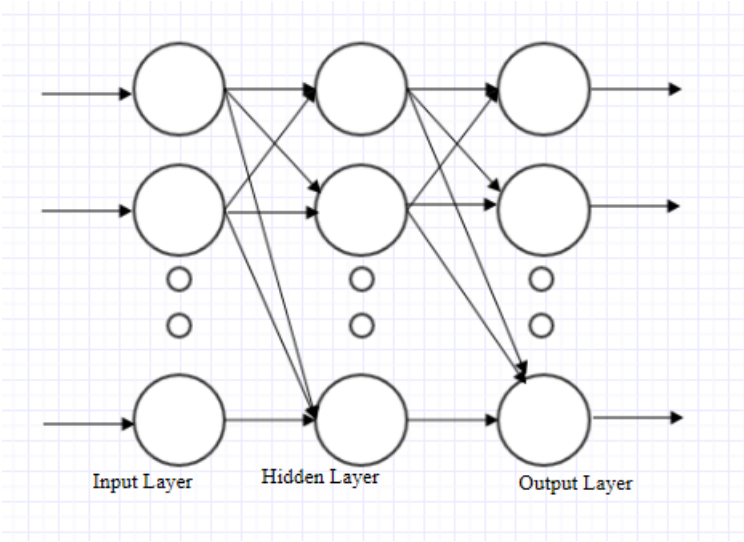

Figure 2.5: Neural Network Structure [14]

2.4.2 Support Vector Machines

The original SVM algorithm was proposed by Boser, Guyon & Vapnik in 1992. The present standard form (soft margin) was given by Vapnik and Corinna Cortes in 1995 [22].Support Vector Machines have become a very popular tool for machine learning tasks of classification and regression. Machine learning is about learning structure from data. Classifying data is a common task in machine learning. In machine

learning, support vector machines are supervised learning models that analyse the training data and recognize patterns and produces an inferred function known as classifier (for discrete output) or regression function (for continuous output). The basic SVM studies a set of input data and decides for each given input, which of two possible classes forms the output, which makes it a non-probabilistic binary linear classifier [22]. The classifier is a function which assigns labels to samples, even those samples which are completely new to the algorithm. Algorithm feeds on previously labelled samples and induces a classifier from them. The key idea in network security is to find useful patterns or features describing user behaviour on a system and a set of desired features to construct classifiers. These classifiers are then used to detect anomalies and intrusions from the new coming network traffic. An SVM model is a representation of the instances as points in space, mapped in such a way that the examples of the distinct types are divided by a clear and a wide gap. Mapping of new examples is done into that same space and they are predicted to belong to a category based on which side of the gap they fall into. Support Vector Machines (SVMs) construct a decision surface in the feature space. This feature space bisects the two categories and maximizes the margin of separation between two classes of points. The resulting decision surface can then be used as a basis for classifying points of unknown class. For instance, some data points belonging to one of two classes is given and the goal is to decide which class a *new* data point will belong to [23].

Support vector machines (SVM) [24] are the classifiers which were originally designed for binary classification and can also be used to classify the attacks. An SVM maps linear algorithms into non-linear space. It uses a feature called kernel function for this mapping. Kernel functions like polynomial function are used to divide the feature space by forming a hyper plane. The kernel functions can be used at the time of training of the classifiers which selects support vectors along the surface of this function. Support Vector machine classify data by using the support vectors which outline the hyper plane in the feature space as shown in Fig 2.6 . This process will involve a quadratic programming problem, and this will get a global optimal solution. Suppose we have N training data points:

{(x1, y1), (x2,y2), (x3, y3), ...,(xn , yn)},

where xi € R, d and yi € {+1,−1}.

Consider a hyper-plane defined by (w, b), where w is a weight vector and b is a bias. The basic idea is to seek classification surface, making the maximum margin of the classification boundary.

$$(w.x)+b=0$$

where x is a multidimensional vector. The reciprocal of classification margin is

$\frac{1}{2} \|w\|^2$. So, the optimization problem is expressed as:

$$\text{Min } \frac{1}{2} \|w\|^2$$

Optimal hyperplane H : $y = (w.x)+b = 0$

Supporting hyperplanes shown as lines in Figure 2.6:

H1 : $y = (w.x)+b = +1$

H2 : $y = (w.x)+b = -1$

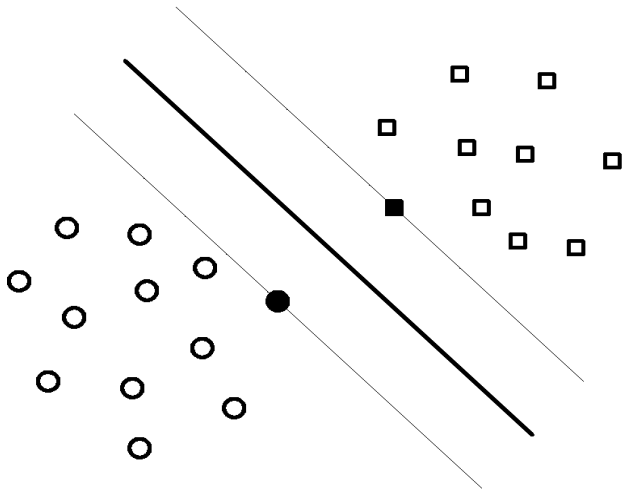

Figure 2.6 Hyper Plane in SVM [51]

The quality of generalization and ease of training of SVM is way too better than the traditional methods. SVM has a high generalisation accuracy but the response time of SVM classifiers is still a concern when applied into network intrusion detection. Its limitation is speed and size, both in training and testing [25].

Following are the steps of SVM Algorithm:

- Train SVM on new data set.

  $D=\{(a_i , c_i)| a_i \in R^n , c_i \in \{-1,1\}\}^m_{i=1}$

where a_i is an n-dimensional real vector and c_i is an indicator of the point a_i belongs to.

- Find the hyperplane separating negative and positive instances of dataset

  $wx-b=0$

  where w is a normal vector to the hyperplane.

  Find shortest distance separating hyperplane to closest positive(negative) data point.

- Find the margin of separating hyperplane

  $(d+d-)=2/||w||$.

- To get highest confidence classification, maximize the margin. Formulate the linear support vector problem as follows:

  $\text{Max } 1/||w||^2$

  $\text{s.t } c_i(a_i w-b) >= 1 \ \& \ i=[1,m]$

- For separable case when positive and negative data points are linearly separated, they satisfy the following constraints:

  $a_i w-b>=0$, for $c_i=1$,

  $a_i w-b<=0$, for $c_i=-1$

  or they can be combined into one set of inequalities:

  $c_i(a_i w-b)-1 >=0$ for all i.

- Solve for w and find the classification

2.4.2.1 Advantages of SVM

1. The quality of generalization and ease of training of SVM is far beyond the capacities of these more traditional methods. Its generalisation ability is controlled by changing the kernel. By developing a kernel specific to application, its generalisation ability can be extended to wide range of applications.
2. By choosing an appropriate generalization grade, SVMs can be robust, even when the training sample has some bias.
3. SVM can model complex and real-world problems such as text and image classification, network traffic analysis.

4. SVM performs well on data sets having many attributes. There is no upper limit on the number of attributes. The only constraints are those imposed by hardware.
5. SVMs deliver a unique solution, since the optimality problem is convex. It is formulated as quadratic programming problem, so there is global optimum solution [38].

2.4.2.2 Disadvantages of SVM

1. Because training of SVM is done by solving the associated dual problem, the number of variables is equal to the number of training data. Thus for large training data, solving the dual problem becomes difficult from both the memory size and training size.
2. The biggest limitation of the support vector approach lies in choice of the kernel.SVM classifiers have only one user-chosen parameter, once the kernel is fixed. The best choice of kernel for a given problem is still a research issue [25].
3. There is lack of transparency of results.
4. SVM uses direct decision functions. Thus an extension to multiclass problems is not straight forward and there are several formulations.
5. A second limitation is speed and size, both in training and testing. Training for very large datasets (millions of support vectors) is an unsolved problem.
6. The response time of SVM is still a problem when applied to real-time business intelligence systems such as stock market surveillance and network traffic analysis [27].

2.4.3 K-Means Algorithm

The term "*k*-means" was first used by James MacQueen in 1967. The idea though goes back to Hugo Steinhaus in 1957. The standard algorithm was first introduced by Stuart Lloyd in 1957 but it wasn't published outside Bell labs until 1982 [27].

In data mining, *k*-means clustering is a method of cluster analysis which aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster having the nearest mean. This process partitions the data space into Voronoi cells. In mathematics, a Voronoi diagram simply divides the space into a number of

regions. The regions are called Voronoi cells [31].Simply speaking it is an algorithm to group your objects based on attributes/features into K number of groups. The clustering is done by minimizing the sum of squares of distances between data and the cluster centroids. Aim of K-mean clustering is simply to classify the data into k different clusters through the iteration and to converge it to a local minimum. In this process, data objects are grouped into disjoint clusters in such a way that the data in the same cluster is similar and data belonging to a different cluster is different. So the generated clusters are compact and independent. Euclidean distance is usually considered to determine the distance between data object and the cluster centroids [28].

Following steps shows the demonstration of *k*-means algorithm [29]:

- k initial "means" are randomly generated within the data domain.
- k clusters are created by connecting every observation with the closest mean.
- The centroid of each cluster becomes the new mean.
- Steps 2 and 3 are repeated until centroids don't change their position anymore.

This is a very simple and reasonably fast algorithm. It is also efficient in processing large data sets like network traffic. The only difficulty is in comparing the quality of the clusters produced. Another limitation of k-means is that k should be specified in advance. But in Intrusion detection k is set to be two since there are two clusters for normal and anomalous data.

2.4.3.1 The K-Means Algorithm Process

- The dataset is grouped into K clusters and the data points are randomly associated to the clusters resulting in clusters having almost same number of data points.
- For each data point:

  Find the distance between each data point and its respective cluster.

  Leave the data point if it is nearest to its own cluster. Otherwise move it into the closest cluster.
- Repeat the above step until centroids don't change their position anymore and no data point moving from one cluster to another. Now the clusters are stable and it marks the end of clustering process.

- The choice of initial partitioning can greatly influence the final clusters that result [35].

2.4.3.1 Advantages of k-means clustering algorithm:

1. This is a very simple algorithm.
2. It is reasonably fast algorithm although its "worst case" behavior is poor.
3. K-Means may form tighter clusters than hierarchical clustering, especially in case of globular clusters are [34].

2.4.3.3 Disadvantages of k-means clustering:

1. It is tough to compare the quality of the clusters produced. For instance, for different initial partitions or different values of K affect the final outcome [28].
2. Fixed number of clusters can make it problematic to predict the value of K.
3. The main limitation of k-means is its cluster model. This concept is based on spherical clusters which are separable in such a way that the mean value converges towards the center of cluster. The clusters are expected to be of same size, so that their assignment to the nearest cluster center is the correct assignment [34].
4. The k-means result depends upon the data set. It works fine on some data sets, while fails on others [28].
5. Different initial clusters can result in different final clusters. The program should be rerun using the same as well as different values of K to compare the results achieved [35].

2.4.4 Comparison of SVM and K-means clustering algorithm:

- In contrast to SVM there are no target output labels in training and testing datasets in k means clustering. The machine simply receives inputs and the task is to learn and differentiate them [36].
- SVM is machine learning task of inferring a function from labelled training data. In k means it is possible to find and learn the hidden structures inside the unlabeled data [37].

- In SVM predetermined classes are provided. Machine learner's task is to look for patterns and construct mathematical models. Patterns are learnt from training data to simulate given output. In k clustering, no classification is provided. Machine learner's task is to look for patterns in data and seek out similarity between pieces of data in order to determine whether they can be characterised as forming a group [39].

Chapter 3
Problem Statement

During literature review, some research gaps have been encountered. A problem has been formulated in an attempt to overcome some of those gaps.

3.1 Research Gaps

The research gaps observed during literature review are as follows:

- Most of the intrusion detection systems models are ineffective for novel attacks.
- There is a problem of lack of transparency in results.
- Intrusion detection systems take a lot of time to identify attacks. Ample time is not left in order to respond back.
- Most of the distributed intrusion detection systems are signature based. They have to be made adaptive to novel attacks.

3.2 Problem Formulation

Due to recent advances in network technology, computer systems have become more vulnerable to attacks. The growth of complex computer networks provides added complicities for the intrusion detection problem. The ever growing connectivity of systems gives more access to attackers and makes it even more difficult for security analysts to protect their system. Our dependency on network based systems is growing day by day, so does the threat. But protection techniques of such systems have not kept up with the increasing threat. Traditional defence mechanisms such as user authentication, data encryption, avoiding programming loopholes and firewalls are used as the first line of defence against attacks. Till date, no combination of technology is able to protect the system completely as systems are facing novel attacks every other day.

Intrusion detection systems (IDS) are used as the last line of defence. IDS identifies patterns of known intrusions (misuse detection) or differentiates anomalous network data from normal data (anomaly detection). Intrusion detection is defined as the

process of diagnosing the system for activities running without authorization and those having legitimate access to the system but overstepping their privileges. The existing counter measures have their own drawbacks, so assuring secure and reliable operation of networks has become a priority research area. Efforts are being made to improve the existing protection technology. The main objective of thesis is to provide a framework for improved IDS, which will overcome the drawbacks of existing IDS.

3.3 Objectives of Thesis

To fulfil the above aim, the following goals are set for the thesis work:

- To study the existing technology of Intrusion detection system and to discover their drawbacks.
- To propose a novel architecture for Intrusion Detection System based on machine learning/adaptive module to detect novel attacks.
- To implement the proposed algorithm in Java using Netbeans.

3.4 Methodology used in Proposed Solution

Following are the steps performed to implement the above objectives:

- Literature Survey is carried out. Various Research papers referenced below are explored to formulate a problem.
- Hybrid Intrusion Detection System architecture is proposed.
- A hybrid algorithm of SVM and k-means is proposed which couples the benefits of both the algorithms.
- The anomaly detection module of the proposed framework is implemented. The presented algorithm is implemented in Java using Netbeans.
- The packets which are used as input are captured using a sniffer called Wireshark. These packets are stored in a dump file and passed to Java program as an input.
- The program is run to classify the traffic into normal and anomalous packets.

Chapter 4

Proposed Solution

In this chapter, hybrid algorithm is presented along with its class diagram. The hybrid framework and its UML diagrams are also provided.

4.1 Hybrid approach: K-means SVM (KMSVM) algorithm

As compared in literature survey, both k-means clustering and SVM have their own advantages and drawbacks. In the proposed solution, a hybrid of both is used instead of choosing one algorithm. It is a challenging task to improve the efficiency of the clustering algorithm without reducing the generalization performance of Support Vector Machine. To face this challenge, a new hybrid algorithm based on the integration of SVM and *k*-means clustering is developed. Support vector machines have been used to build classifiers that can help users make intelligent business decisions. Despite high generalisation accuracy of SVM, their response time is still a concern when applied into real-time business intelligence systems like stock market surveillance and network traffic analysis. The K-means SVM shortens the response of SVM classifiers by reducing the number of support vectors [39]. The KSVM algorithm blends the *k*-means clustering technique with SVM and needs another input parameter: the number of clusters. Response time of SVM classifiers can be accelerated by lowering the number of support vectors. *k*-means clustering method is used to gather a data set smaller than the original one to train SVM, which further lowers the number of SVs while maintaining the training accuracy. With decrease in the number of training examples, computational time of the algorithm falls greatly. There are two approaches for taking advantage of *k*–means clustering algorithm to reduce the number of support vectors used for training the support vector machine. The first approach applies *k*–means clustering to compose a dataset of much smaller size than the actual one. The second approach lowers the number of support vectors by which SVM classifier's decision function is spanned through *k* –means clustering [37].

E[Pr(Error)]<= *E[number of support vectors]/number of training vectors .....(1)*

From inequality (1), it can be deduced that a small number of support vectors will generate a small testing error and also leads to better generalization capability in SVM [66].

Successful use of *k*-means requires a cautiously selected distance measure that demonstrates the properties of the clustering task. Supervised data is used for training *k*-means since designing this distance measure by hand is a tough job. SVM method that finds a distance measure is used so that *k*-means generates the desired clustering, given the training data sets with desired partitioning. Our approach couples the benefits of both the algorithms. This is done by selecting the initial clusters with the help of SVM which are otherwise generated randomly in k-means. This way, a more precise result is obtained as it is a known fact that the end result of k-means depends upon the initial clusters chosen [68].

4.2 Adaptive Distributed Intrusion Detection Algorithm based on hybrid k-SVM clustering technique:

1. Gather Packets.
2. Apply Misuse Detection Algorithm at nodes.
3. Send remaining packets to centralised anomaly detection agent. Apply k-means clustering Support Vector algorithm.
   - Given a training dataset D containing m data points:

     $D = \{(a_i , c_i)| a_i \in R^n , c_i \in \{-1,1\}\}^m_{i=1}$

     where a_i is an n-dimensional real vector and c_i is an indicator of the class where the point a_i belongs to.
   - Separate the dataset into positive (c=1) and negative (c=2) instances with a hyperplane

     wx-b=0,

     where w is normal vector to the hyperplane, x is point of the hyperplane,

     b is real value, 1/||w|| is perpendicular distance from hyperplane to origin.

- For separable case when positive and negative data points are linearly separated, they satisfy the following constraints:

$a_i w-b>=0$, for $c_i=1$,

$a_i w-b<=0$, for $c_i=-1$

or they can be combined into one set of inequalities:

$c_i(a_i w-b)-1 >=0$ for all i.

- Choose w and b to maximize the margin to get highest confidence classification.

Formulate the linear support vector problem as follows:

$$\text{Max } 1/||w||^2$$

$$\text{s.t } c_i(a_i w-b) >= 1 \ \& \ i=[1,m]$$

- The resulting two clusters will be assumed as the initial clusters of k clustering algorithm.
- Set k=2 (for normal and anomalous traffic in training data) initial cluster centres.
- Assign each packet x_i € S to the group that has closest centroid s.t $|| x_i - c_k|| <= ||x_i - c_j||$. Assign x_i to c_k.
- To get optimum cluster, subset P of the set S should have maximal value of total distance between all instances in the set S.
- Recalculate the positions of k centroids.
- Repeat steps 2 & 3 until centroids no longer move.

4. Nodes sending anomalous packets are informed by centralised node.
5. New anomalous information is updated in Rule Mining Agent which feeds Misuse Detection Agent next time.

4.3 Framework of Adaptive Distributed Intrusion Detection System

The conventional approaches to intrusion detection involving a central unit to monitor an entire system have several drawbacks. To overcome them research witnessed a wealthy number of works heading towards a distributed framework of monitors that carry out local detection and provide information to perform global detection of intrusions. The volume of data acquisition is very enormous for the common intrusion detection system. Only a small number of these data is meaningful for intrusion

detection. It will be easy for powerful systems to use more of their computing resources. But the common computers might not have sufficient resources to run the system smoothly. Because of this, intrusion detecting system should have distributed structure and characteristics so that different detecting programs at different levels can run at different computers with different function. Only in this way, can computers collaborate with each other and we can make use of computer resources reasonably, and run detecting work efficiently. The individual monitors send intrusion data to the centralized controller component that performs analysis of the information that it receives from each of the monitors [40].

The main issues of this kind of existing centralized systems are [40]:

- The real-time of the intrusion detection and response is not good.
- Single host to copy with the collected information, it means that the monitored network is restricted. Lots of data collection can overload the network.
- Add new hosts cause the load on the centralized controller to increase significantly. As a result, it makes the IDS non-scalable.
- The flexibility of the system is not good and it lacks of dynamic configuration ability.
- Lack of cooperation between different IDSs. It needs to use the combination of Host-based IDS (HIDS) and Network-based IDS (NIDS).

So distributed system architecture is used in our approach. Most of the current distributed IDSs are signature based. A major shortcoming of such IDSs is that they can't identify novel attacks but only well known attack patterns for which signatures are available. To overcome this limitation, IDS is made capable of adapting to the changing attack atmosphere. Data mining methods are used to automate the intrusion detection systems making it anomaly based IDS as well [45]. Short-comings of anomaly based IDS, namely a high false positive rate and the ability to be fooled by a correctly delivered attack are overcome by signature based Distributed architecture [46]. Feature of adaptation is introduced in Distributed IDS with the help of machine learning algorithm. Two algorithms are compared: SVM and k-means clustering and a hybrid of the two algorithms is used. A general framework for a hybrid IDS that is both adaptive and distributed is also provided. A novel Intrusion Detection System (IDS) architecture is proposed which includes both anomaly and misuse detection

approaches. This hybrid Intrusion Detection System architecture consists of a centralised anomaly detection module and distributed signature detection modules. The proposed anomaly detection module uses hybrid machine learning algorithm called k-support vector machine clustering (KSVM) [49]. This hybrid system couples the benefits of low false-positive rate of signature-based intrusion detection system and anomaly detection system's ability to detect new unknown attacks.

The proposed framework is based on the network-based intrusion detection techniques. It is extended from architecture of distributed and adaptive IDS. It is a distributed IDS having sniffer agents and signature detection agents at nodes [47]. It has a star type of architecture as shown in Figure1. The architecture proposed is basically composed of five components: Sniffer Agent, Signature based Intrusion Detection Agent, Anomaly based Intrusion Detection Agent, Rule Mining Agent, Signature based Rules Database. Network traffic is captured from different nodes using Sniffer agents such as Wireshark. These packets are then stored in a dump file from where they can be passed onto Anomaly Detection Agent. The IDS is made adaptive by adding a machine learning component at a centralized node. This centralized node applies KSVM algorithm to the incoming packets. Signature based intrusions are detected at nodes having Signature Detection Agent and rest of the suspicious data is passed on to the centralized node having Anomaly Detection Agent. This information is passed on to Signature based Intrusion Detection Agent which matches the patterns with available rules in Signature based Rules Database. The patterns which correspond to available signatures are declared as intrusion. So, all the known attacks are detected at individual nodes itself. This reduces the burden on centralised node which will now focus on detecting novel attacks. Any suspicious data left is further passed on to centralised node's Anomaly based Intrusion Detection Agent. This agent applies KSVM algorithm which distributes the data into two different clusters: normal and intrusion. The Rule Mining Agent summarises this information of anomalous data declared by Anomaly Detection Agent and updates Signature based Rules Database with profile of new encountered attack. Misuse Detection Agent is fed from time to time by Signature Database with association rules to update its signatures. If the same attack is faced in future, it will be detected by Signature based Intrusion Detection Agent available at nodes. So, this architecture

helps in reducing the burden of resources on one hand while identifying novel attacks on the other.

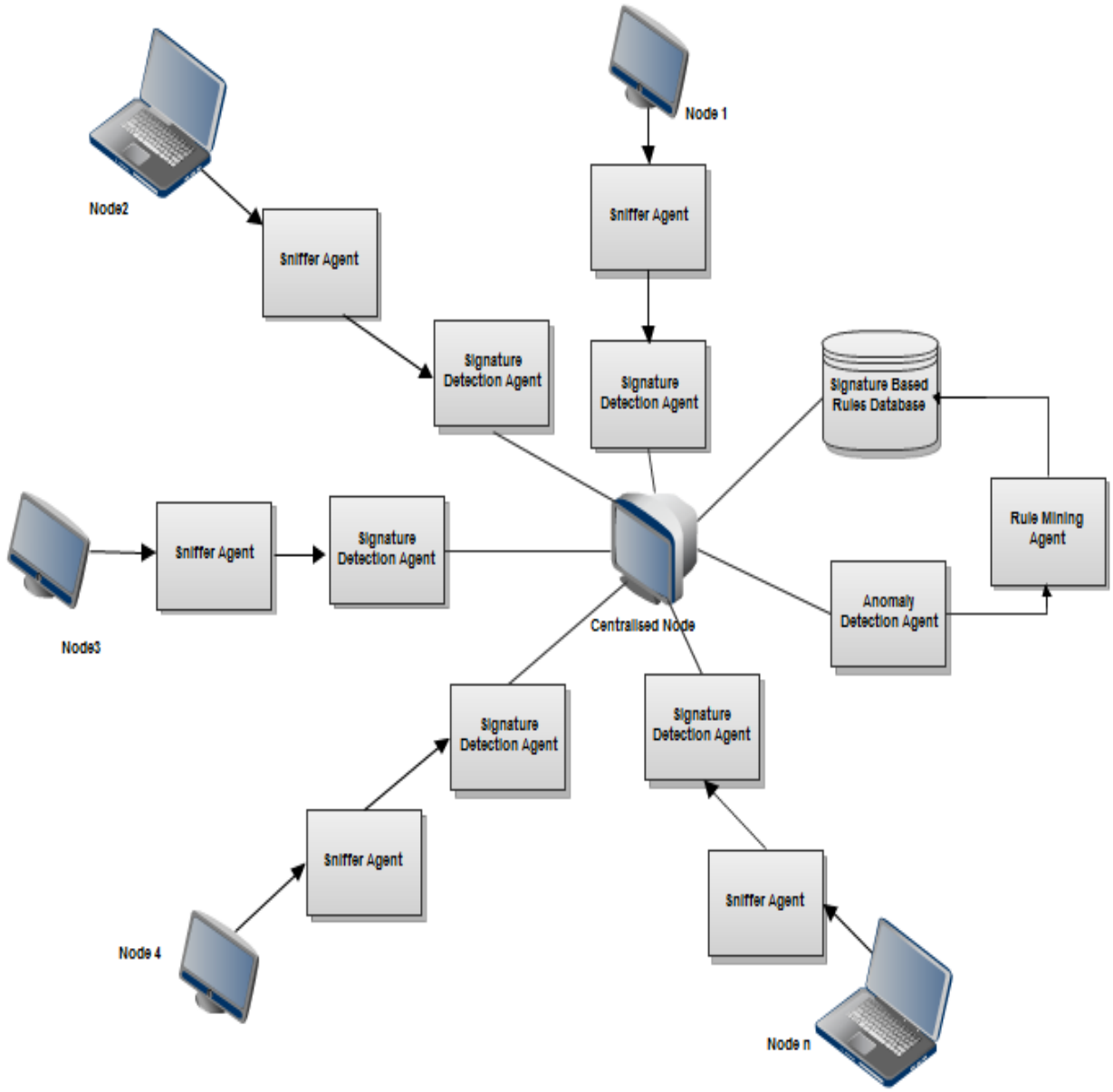

Figure 4.2: Framework of Adaptive Distributed IDS

Chapter 5

Implementation and Results

In this chapter packet capturing by Wireshark is shown. Packets are studied to explore various fields. These packets are later used in implementation of KSVM algorithm. Hybrid k-means Support Vector Machine is implemented in Java using Netbeans Integrated Development Environment.

5.1 Implementation Setup

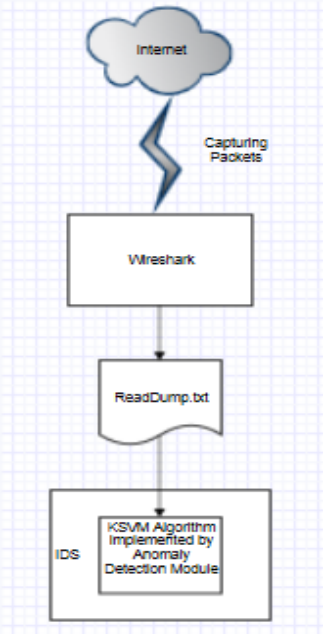

Figure 5.1: Implementation Setup Diagram

5.2 Basic steps performed in implementation

1. First the network packets are captured using a GUI sniffer called Wireshark. Various fields of packets are studied to choose one of them as an attribute to be used in the algorithm. In this implementation, total length of the packet is chosen.

2. All the packets under study are stored in ReadDump.txt file. Respective length of each packet is calculated in this step.

3. The input ReadDump.txt file is provided to the Java code of the k-means Support Vector Machine. This algorithm divides the traffic into normal and anomalous by clustering them into two. Output of anomalous packets is provided so that we can discard them.

5.3 Capturing Packets

Every data transfer on the Internet involves packets. For instance, every Web page received comes as a group of packets and every e-mail sent leaves as a cluster of packets. The network breaks the message into bytes of certain size. These parts are called packets. Each packet contains the information to reach its destination: the sender IP address, the intended receiver IP address, frame number, length of packet, version etc. The packets carry the data according to different protocols like TCP, SSDP, ICMP etc. A typical packet contains perhaps 1,000 or 1,500 bytes. If it contains data more than 1500 bytes it is considered suspicious. The frames may belong to different protocols like TCP, SSDP, DHCP, ARP, HTTP. The snapshot 4.3 shows the TCP Frame. In similar fashion different types of frames are captured by Wireshark.

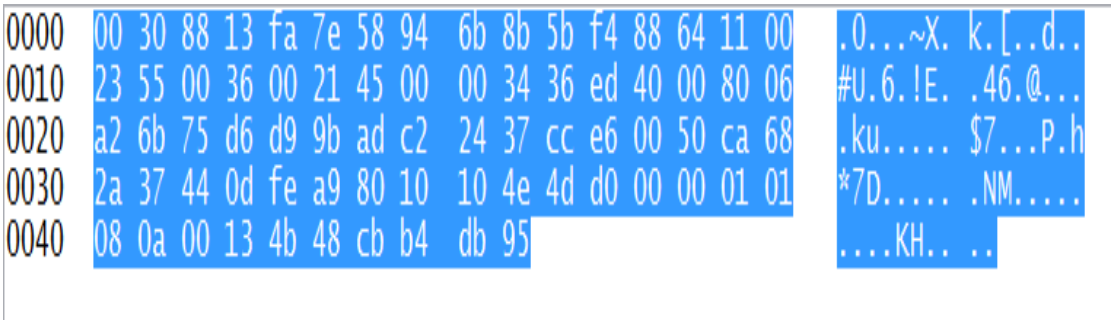

Figure 5.2: Captured Frame

```
⊞ PPP-over-Ethernet Session
⊞ Point-to-Point Protocol
⊞ Internet Protocol Version 4, Src: 117.214.217.155 (117.214.217.155), Dst: 173.194.36.55 (173.194.36.55)
⊟ Transmission Control Protocol, Src Port: 52454 (52454), Dst Port: http (80), Seq: 1137, Ack: 199, Len: 0
    Source port: 52454 (52454)
    Destination port: http (80)
    [Stream index: 4]
    Sequence number: 1137    (relative sequence number)
    Acknowledgment number: 199    (relative ack number)
    Header length: 32 bytes
  ⊞ Flags: 0x010 (ACK)
    Window size value: 4174
    [Calculated window size: 16696]
    [Window size scaling factor: 4]
  ⊞ Checksum: 0x4dd0 [validation disabled]
  ⊞ Options: (12 bytes), No-Operation (NOP), No-Operation (NOP), Timestamps
  ⊞ [SEQ/ACK analysis]
```

Figure 5.3: Different Fields of Frame

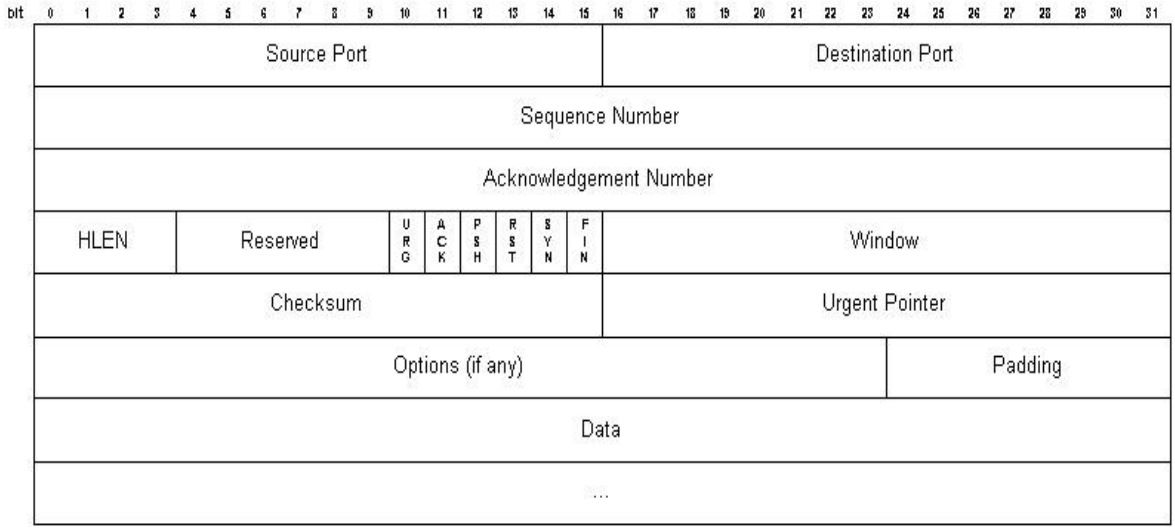

Figure 5.4: TCP Frame Format [57]

Network packets are captured using Wireshark as shown in Figure 2. Interface of Wireshark shows the captured packets which can be displayed in decimal or hexadecimal form. It displays destination host address, source host address, version, total length, flags, time to live and other fields. Any of the above attributes of packets can be extracted to use in KSVM algorithm. KSVM needs atleast one of these attributes for comparison performed during clustering. For instance, total length of packet is used in this implementation. A rule can be formed that packets having length more than 1500 bytes is discarded. This type of rules can be applied to different

attributes of packets and a policy for Intrusion Detection System can be made in such a way.

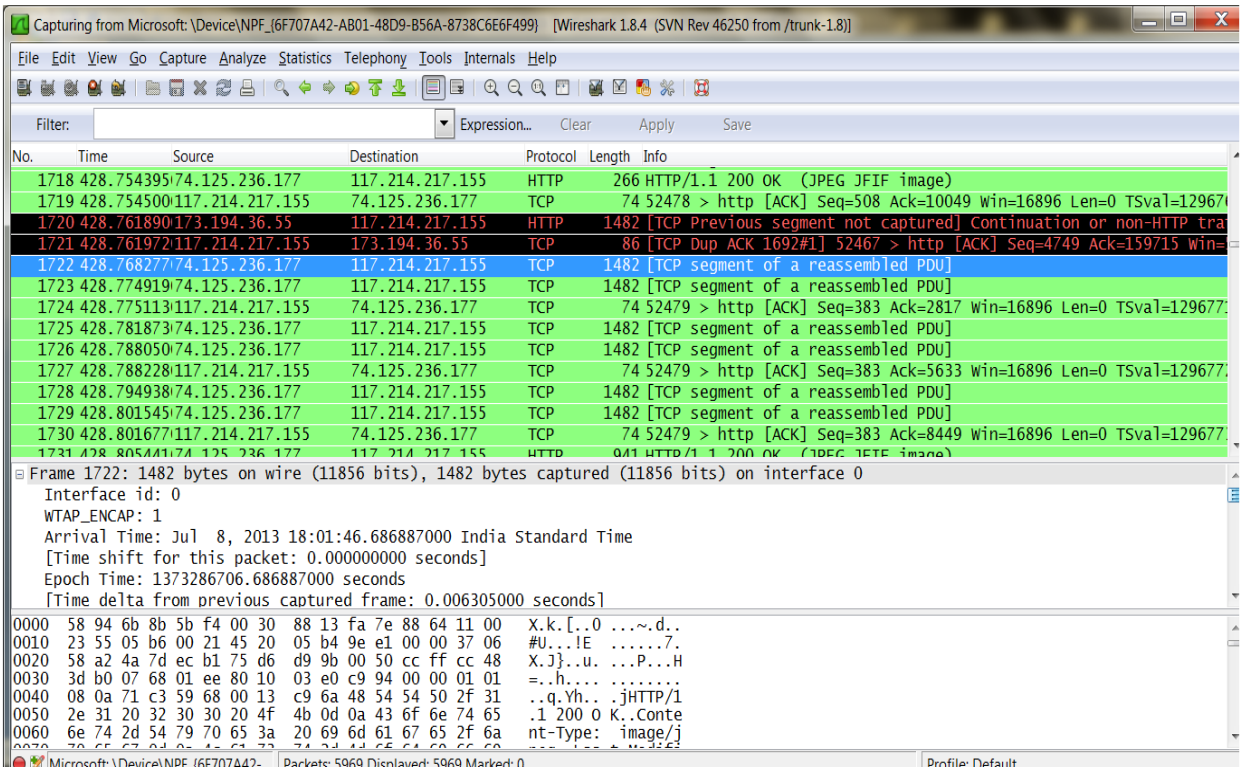

Figure 5.5: Wireshark Interface

The packets captured by Wireshark are stored in a dump file and are used afterwards at the time of implementation of clustering algorithm. The file called ReadDump.txt is used as an input to the code. The code picks up the packets from the file shown in Figure3. The ReadDump.txt file also stores the size of each packet along with each packet. This total length attribute of the packet is used later at the time of clustering.

In general k-means clustering, initial clusters are created by randomly generated kmeans. Here, the number of clusters is fixed to two, for normal and anomalous packets. As we are using a hybrid algorithm in this paper, two initial clusters of packets are generated using SVM instead of simply assuming them. The centroids of these two clusters are the means for k clustering now. Clusters are created by associating every observation with the nearest mean. The process of clustering is repeated until convergence has been reached and centroids don't change their position anymore. In the end we get two clusters one for normal data and one for anomalous data. As total length is the attribute used in k clustering, the packets showing

abnormal length are separated from the rest of the packets. Those are the anomalous packets and can be discarded.

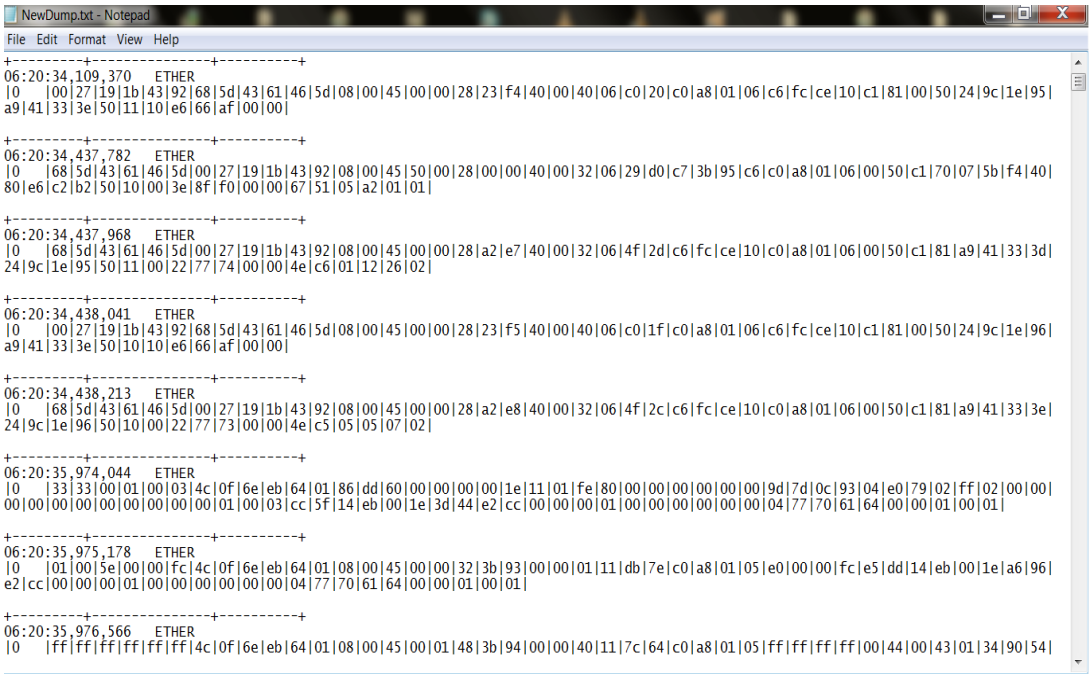

Figure 5.6 Dump File

5.4 Output

The output of the KSVM code is copied from the output screen of SVM.

run:

Total number of the packets 92

Initial clusters generated by SVM:

K1{ 54 54 42 54 42 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 }

K2{ 60 60 60 84 64 342 84 64 109 60 92 87 124 379 87 124 192 66 66 66 66 417 1494 1051 91 60 91 91 66 60 84 64 342 84 64 92 92 91 92 91 91 60 91 60 60 60 60 60 60 60 60 1192 60 113 183 60 91 60 1194 95 60 109 202 60 109 }

At this step

Value of clusters

K1{ 54 54 42 54 42 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 54 }

K2{ 60 60 60 84 64 342 84 64 109 60 92 87 124 379 87 124 192 66 66 66 66 417 1494 1051 91 60 91 91 66 60 84 64 342 84 64 92 92 91 92 91 91 60 91 60 60 60 60 60 60 60 60 1192 60 113 183 60 91 60 1194 95 60 109 202 60 109 }

Value of m

m1=53.111111111111114 m2=171.73846153846154

At this step

Value of clusters

K1{ 54 60 60 54 60 84 64 84 64 109 42 60 54 92 42 87 87 66 66 54 66 54 66 54 54 91 54 60 91 54 91 66 60 54 84 64 84 64 92 92 91 92 91 54 91 60 91 54 60 54 54 60 60 54 54 60 54 60 54 54 54 60 60 60 60 54 60 54 54 91 54 60 95 60 109 54 60 109 }

K2{ 342 124 379 124 192 417 1494 1051 342 1192 113 183 1194 202 }

Value of m

m1=67.28205128205128 m2=524.9285714285714

At this step

Value of clusters

K1{ 54 60 60 54 60 84 64 84 64 109 42 60 54 92 42 87 124 87 124 192 66 66 54 66 54 66 54 54 91 54 60 91 54 91 66 60 54 84 64 84 64 92 92 91 92 91 54 91 60 91 54 60 54 54 60 60 54 54 60 54 60 54 54 54 60 60 60 60 113 183 54 60 54 54 91 54 60 95 60 109 202 54 60 109 }

K2{ 342 379 417 1494 1051 342 1192 1194 }

Value of m

m1=73.64285714285714 m2=801.375

At this step

Value of clusters

K1{ 54 60 60 54 60 84 64 342 84 64 109 42 60 54 92 42 87 124 379 87 124 192 66 66 54 66 54 66 54 417 54 91 54 60 91 54 91 66 60 54 84 64 342 84 64 92 92 91 92 91 54 91 60 91 54 60 54 54 60 60 54 54 60 54 60 54 54 54 60 60 60 60 113 183 54 60 54 54 91 54 60 95 60 109 202 54 60 109 }

K2{ 1494 1051 1192 1194 }

Value of m

m1=87.11363636363636 m2=1232.75

At this step

Value of clusters

K1{ 54 60 60 54 60 84 64 342 84 64 109 42 60 54 92 42 87 124 379 87 124 192 66 66 54 66 54 66 54 417 54 91 54 60 91 54 91 66 60 54 84 64 342 84 64 92 92 91 92 91 54 91 60 91 54 60 54 54 60 60 54 54 60 54 60 54 54 54 60 60 60 60 113 183 54 60 54 54 91 54 60 95 60 109 202 54 60 109 }

K2{ 1494 1051 1192 1194 }

Value of m

m1=87.11363636363636 m2=1232.75

The Final Clusters By Kmeans are as follows:

K1 Normal data{ 54 60 60 54 60 84 64 342 84 64 109 42 60 54 92 42 87 124 379 87 124 192 66 66 54 66 54 66 54 417 54 91 54 60 91 54 91 66 60 54 84 64 342 84 64 92 92 91 92 91 54 91 60 91 54 60 54 54 60 60 54 54 60 54 60 54 54 54 60 60 60 60 113 183 54 60 54 54 91 54 60 95 60 109 202 54 60 109 }

K2 Anomalous data{ 1494 1051 1192 1194 }

Enter the cluster number from which you want to retrieve packets.

2

Size of the list 4

Size of the final Packet list4

Packet: 68 5d 43 61 46 5d 00 27 19 1b 43 92 08 00 45 00 05 c8 57 1b 40 00 2f 06 6c
62 cd fb f3 08 c0 a8 01 06 00 50 c1 c5 0f f7 ec d8 46 21 82 17 50 10 00 7f e8 9b
00 00 48 54 54 50 2f 31 2e 31 20 32 30 30 20 4f 4b 0d 0a 78 2d 61 6d 7a 2d 69 64
2d 32 3a 20 59 6c 6d 69 68 53 37 73 49 75 67 6f 37 6d 73 4d 67 74 47 4b 51 66
79 34 2f 45 52 59 4f 39 67 4c 4b 41 4b 6d 45 47 54 48 6e 33 34 32 78 65 65 44 6f
43 52 45 66 70 59 53 76 76 46 53 51 61 68 6e 0d 0a 78 2d 61 6d 7a 2d 72 65 71
75 65 73 74 2d 69 64 3a 20 32 35 38 42 38 42 32 36 34 43 30 42 36 36 36 43 0d
0a 44 61 74 65 3a 20 54 75 65 2c 20 30 32 20 4a 75 6c 20 32 30 31 33 20 30 36 3a
32 30 3a 34 37 20 47 4d 54 0d 0a 4c 61 73 74 2d 4d 6f 64 69 66 69 65 64 3a 20 57
65 64 2c 20 32 36 20 53 65 70 20 32 30 31 32 20 31 39 3a 32 34 3a 30 37 20 47
4d 54 0d 0a 45 54 61 67 3a 20 22 35 62 32 34 36 38 66 34 38 36 61 62 32 66 33
32 64 33 66 36 38 30 31 32 66 33 62 65 30 36 65 66 22 0d 0a 41 63 63 65 70 74
2d 52 61 6e 67 65 73 3a 20 62 79 74 65 73 0d 0a 43 6f 6e 74 65 6e 74 2d 54 79 70
65 3a 20 69 6d 61 67 65 2f 70 6e 67 0d 0a 43 6f 6e 74 65 6e 74 2d 4c 65 6e 67 74
68 3a 20 32 30 39 32 0d 0a 53 65 72 76 65 72 3a 20 41 6d 61 7a 6f 6e 53 33 0d 0a
0d 0a 89 50 4e 47 0d 0a 1a 0a 00 00 00 0d 49 48 44 52 00 00 00 18 00 00 00 18
08 06 00 00 00 e0 77 3d f8 00 00 00 19 74 45 58 74 53 6f 66 74 77 61 72 65 00 41
64 6f 62 65 20 49 6d 61 67 65 52 65 61 64 79 71 c9 65 3c 00 00 03 24 69 54 58
74 58 4d 4c 3a 63 6f 6d 2e 61 64 6f 62 65 2e 78 6d 70 00 00 00 00 00 3c 3f 78 70
61 63 6b 65 74 20 62 65 67 69 6e 3d 22 ef bb bf 22 20 69 64 3d 22 57 35 4d 30 4d
70 43 65 68 69 48 7a 72 65 53 7a 4e 54 63 7a 6b 63 39 64 22 3f 3e 20 3c 78 3a 78
6d 70 6d 65 74 61 20 78 6d 6c 6e 73 3a 78 3d 22 61 64 6f 62 65 3a 6e 73 3a 6d 65
74 61 2f 22 20 78 3a 78 6d 70 74 6b 3d 22 41 64 6f 62 65 20 58 4d 50 20 43 6f 72
65 20 35 2e 30 2d 63 30 36 31 20 36 34 2e 31 34 30 39 34 39 2c 20 32 30 31 30 2f
31 32 2f 30 37 2d 31 30 3a 35 37 3a 30 31 20 20 20 20 20 20 20 20 20 22 3e 20 3c 72

64 66 3a 52 44 46 20 78 6d 6c 6e 73 3a 72 64 66 3d 22 68 74 74 70 3a 2f 2f 77 77 77 2e 77 33 2e 6f 72 67 2f 31 39 39 39 2f 30 32 2f 32 32 2d 72 64 66 2d 73 79 6e 74 61 78 2d 6e 73 23 22 3e 20 3c 72 64 66 3a 44 65 73 63 72 69 70 74 69 6f 6e 20 72 64 66 3a 61 62 6f 75 74 3d 22 22 20 78 6d 6c 6e 73 3a 78 6d 70 3d 22 68 74 74 70 3a 2f 2f 6e 73 2e 61 64 6f 62 65 2e 63 6f 6d 2f 78 61 70 2f 31 2e 30 2f 22 20 78 6d 6c 6e 73 3a 78 6d 70 4d 4d 3d 22 68 74 74 70 3a 2f 2f 6e 73 2e 61 64 6f 62 65 2e 63 6f 6d 2f 78 61 70 2f 31 2e 30 2f 6d 6d 2f 22 20 78 6d 6c 6e 73 3a 73 74 52 65 66 3d 22 68 74 74 70 3a 2f 2f 6e 73 2e 61 64 6f 62 65 2e 63 6f 6d 2f 78 61 70 2f 31 2e 30 2f 73 54 79 70 65 2f 52 65 73 6f 75 72 63 65 52 65 66 23 22 20 78 6d 70 3a 43 72 65 61 74 6f 72 54 6f 6f 6c 3d 22 41 64 6f 62 65 20 50 68 6f 74 6f 73 68 6f 70 20 43 53 35 2e 31 20 4d 61 63 69 6e 74 6f 73 68 22 20 78 6d 70 4d 4d 3a 49 6e 73 74 61 6e 63 65 49 44 3d 22 78 6d 70 2e 69 69 64 3a 46 44 43 30 42 36 37 43 30 30 32 41 31 31 45 32 41 41 43 42 45 45 32 41 41 46 33 37 32 38 45 45 22 20 78 6d 70 4d 4d 3a 44 6f 63 75 6d 65 6e 74 49 44 3d 22 78 6d 70 2e 64 69 64 3a 46 44 43 30 42 36 37 44 30 30 32 41 31 31 45 32 41 41 43 42 45 45 32 41 41 46 33 37 32 38 45 45 22 3e 20 3c 78 6d 70 4d 4d 3a 44 65 72 69 76 65 64 46 72 6f 6d 20 73 74 52 65 66 3a 69 6e 73 74 61 6e 63 65 49 44 3d 22 78 6d 70 2e 69 69 64 3a 37 37 39 44 45 38 36 31 30 30 32 41 31 31 45 32 41 41 43 42 45 45 32 41 41 46 33 37 32 38 45 45 22 20 73 74 52 65 66 3a 64 6f 63 75 6d 65 6e 74 49 44 3d 22 78 6d 70 2e 64 69 64 3a 37 37 39 44 45 38 36 32 30 30 32 41 31 31 45 32 41 41 43 42 45 45 32 41 41 46 33 37 32 38 45 45 22 2f 3e 20 3c 2f 72 64 66 3a 44 65 73 63 72 69 70 74 69 6f 6e 3e 20 3c 2f 72 64 66 3a 52 44 46 3e 20 3c 2f 78 3a 78 6d 70 6d 65 74 61 3e 20 3c 3f 78 70 61 63 6b 65 74 20 65 6e 64 3d 22 72 22 3f 3e 60 90 5f 2a 00 00 04 9e 49 44 41 54 78 da 7c 55 4d 68 54 57 14 fe ee 7d 6f de 4c 66 a6 69 82 d5 26 12 88 c1 b4 38 c5 80 6d 1a 77 d5 55 37 d9 28 ba ca a2 a0 e2 20 8a d0 d0 d4 b6 9b 06 a4 0b db d4 94 29 48 c0 a6 66 e7 36 98 8d eb 06 a9 48 c5 45 88 18 29 8a 16 c4 84 a4 69 a2 99 cc cf fb bb fd ee 7d f3 67 32 f1 c1 7b f7 be fb be fb 9d 73 be 7b ce 79 42 29 85 66 d7 e4 9f 57 0e c1 52 d3 b0 ec 7b 10 d6 e9 ec a7 23 65 bc ed 5a 78 79 08 71 e2 13 c9 7b b0 9c d3 78 3f 65 f0 76 53 f2 bb 3f 76 c3 c2 ef 68 71 7a 10 8a 3c 42 24 b9 bc b3 81 bf 97 bb 91 20 7e 77 6b 0f 02 99 87 57 c7 cb ed 9e ff e4 40 60 14 09 bb 1f 9c 40 21 97 ed 1f 59 db d9 f3 55

Packet Size:1494

Packet: 68 5d 43 61 46 5d 00 27 19 1b 43 92 08 00 45 00 04 0d 57 1c 40 00 2f 06 6e 1c cd fb f3 08 c0 a8 01 06 00 50 c1 c5 0f f7 f2 78 46 21 82 17 50 18 00 7f 43 be 00 00 87 91 8e 62 57 ba 1f d2 02 0d e4 f0 5e aa 86 b7 b7 c8 92 22 e9 45 c4 e5 10 a5 01 fc 70 81 cb cb 93 f3 b9 5e 04 e1 0b 84 aa 9c fd 64 a4 ae e9 c2 62 0a 42 5e 44 9b 33 84 96 04 50 f0 88 17 cb f0 82 5e 94 fc 17 70 c3 b2 f8 ed ee 95 18 bd 14 50 a2 07 76 78 1d 31 fb 28 ef 88 40 53 85 21 47 b5 c6 f9 0c 37 df cf b6 7e 31 49 c9 04 7c e2 93 c1 75 b4 b6 1c 45 aa 85 5a 50 8c 80 58 2f a0 63 fe 1a a5 9d a1 f1 fb 64 52 63 84 ef 47 4c 64 e0 38 bd b0 64 44 5c b5 20 85 de dc 4e ef 4f d1 d0 00 44 f0 01 2c b1 1f 29 2b 83 b6 77 7a 11 77 1a f0 34 10 23 be 25 d1 4e 43 a7 10 a8 01 9b 22 1d 87 65 ed 43 2c 56 f7 5a 5f a2 f2 d0 59 e6 07 7a 7c c9 f7 79 24 c5 71 92 ee 43 3a cd 57 f1 26 5e 47 a1 23 2e 96 b4 bc c4 cb 79 69 16 35 50 34 1c 86 a8 3d 22 03 3a 0a 21 6f 71 32 41 67 60 ce 47 ca 3a 58 54 6e 54 24 b5 35 c6 b9 05 e9 4c c8 6d 84 8d 73 33 15 91 b6 4a fc c5 97 c7 35 67 0c a1 d8 b2 57 44 86 5d af 82 b7 1e db f4 ac 2e 05 b6 18 0a b9 16 50 1e 69 dd 46 28 6f 66 3f 1e f6 b1 ba 1e 65 b7 39 7c 8d 6b 50 40 63 cb 94 c7 6a b9 8d 20 76 13 ef 4a 5f 1b f8 81 c0 0e 7a d9 49 d0 67 04 7f c4 91 99 65 c2 f6 21 ed 69 7e 1f 36 e4 26 b1 e3 c4 ab 0e 94 dc 08 6f db 11 5e 7b 2e 2d 1f 76 72 9a b5 30 ac c9 8d 9f 8d ad 62 72 6e bc 83 6e 75 11 c8 7a 40 3b b5 5f a4 a1 b9 ec c1 2f 4b 4d 8b 2c bf 19 e1 e3 f1 14 a3 6d 67 7a 2f 32 3d e7 60 8b 52 d3 42 1b fa c3 d6 f1 77 33 fc 76 7a bf c6 90 19 33 da 70 10 4b 4d 0d 24 92 eb 90 aa 1b 9e af 9d 59 43 a1 40 bc 68 43 3a 55 c3 d7 22 c8 8f e7 18 b2 ca 11 78 c2 18 8e d6 3d 1a 7b c4 f1 0e 23 59 e4 b8 94 fe fa ab 29 b3 e1 b5 df 09 2b cc 31 ef 4f 50 22 bb 72 26 1e 8b ec 11 c7 3b 8c 84 78 b1 64 0c e4 af fe ca b3 08 67 a0 fc 41 93 86 ba a7 54 8b a2 aa a0 0e 26 0c 9f a7 2f 5d ea c1 ab d0 86 e5 cd 20 28 0e 22 ce 16 e1 54 8a 4d 3b a5 0d e9 03 2f b2 d7 79 fe 73 3b 7f 35 b7 9b ab 83 50 61 44 6e b2 6a 4b 0b 17 55 63 bc 5f f9 c4 d3 91 a0 1c 91 eb 3d aa ea 88 aa a4 aa 88 aa 3a f4 75 be a9 03 24 38 5c 6f 11 0d e4 8d 76 aa 1e 22 20 5e 1d 86 13 ab f6 a9 2d d1 aa ba 41 8e d4 3a b8 c0 97 23 06 23 e5 0e e4 7a 43 85 2c 74 2f f0 71 c4 44 5a 6b 2f aa c1 50 a5 9a 03 df c8 6a 93 b4 8f 71 ec ad f5 11 03 6c 28 36 53 3c fe 13 36 2e dd ba 9f 32 15 3f 67 b3 db 8b 98 15 f5 28 a3 b9 ac e3 cb 2e b0 5e 78 82 52 40 bc 7a ca d3 c7 04 73 78 80 e8 63 a6 6b ca 86 1e a4 5b af eb cf b2 35 9f e3 fa 33 dd d4 e1 aa 2c a3 1e 80 f4 8e b1 c8 da 8d 21 54 aa 7e b3 c8 ec

2a ce a2 60 9d 83 2d 9f d1 71 25 36 c6 c6 05 37 c7 e9 55 17 bc 50 eb 3b c6 8f 19 66 40 11 c5 e0 32 b7 5e 4b 8f 7e b7 59 0b 7f b5 28 e0 c8 38 ff 78 5d d8 70 0f 30 55 c7 90 8c 65 f0 9a ec 2b ee 65 4a 79 0d 99 ce cd 6d 75 60 6a e1 e7 5f ce f0 5f 7c c3 e8 5d f4 a7 d8 7f ce a7 47 bf 71 77 fc 5d fe 57 3c 83 84 ba c1 dc 07 56 8a 53 70 ed f3 c8 ec 7a 03 5f 2f b4 b1 71 ca 83 59 ae f4 a1 e4 3d 20 f9 c9 f4 f7 df fe b3 23 f9 bf 9b 94 87 78 2b ec c3 ca c6 03 78 f6 49 7c b8 67 1b de 6e c8 9a 02 7f 12 0f 79 fa 69 04 38 9b 1e 7d 0b b9 29 3c e2 5d f7 21 4a 85 34 ca e2 2c 32 7b 9a e2 ff 17 60 00 ba c0 09 c2 ce 7a 87 80 00 00 00 00 49 45 4e 44 ae 42 60 82

Packet Size:1051

Packet: 00 27 19 1b 43 92 68 5d 43 61 46 5d 08 00 45 00 04 9a 24 12 40 00 40 06 7e 8d c0 a8 01 06 4a 7d 87 93 c1 4f 01 bb 32 09 ce 41 5f 6a 58 a9 50 18 0f 6e 43 dc 00 00 17 03 01 04 6d c3 74 9f 8b b5 e4 e0 60 b0 ff ea 7b b9 a6 45 2a dd dd 28 52 fe 4f 1f 97 9a c5 8a 8e ac d2 72 64 1a fa 7f aa 60 62 dd 68 ab 72 8c e9 fc 14 c4 8f 66 21 4c 0b fa a8 03 b1 ef 4b fd ea 08 58 14 5b 99 9a ce 09 38 f2 c8 21 11 d9 96 8d 71 c7 b2 ae 2c dc 53 35 d0 10 69 05 7d c3 d6 c8 f1 67 16 00 cf ac 09 01 cd 8f 08 19 47 ae 6f 6c 8b 13 0f 2f 2d 11 47 e0 2e 2b 6c 0a 8f 36 41 12 1f a5 92 55 f9 bd 15 71 6e 4e b8 2c 76 a4 aa ef 20 91 17 76 f0 b4 59 ec f7 2e 2f 95 35 5d bc 38 e0 24 30 e6 46 b0 e7 a8 af d5 39 0f 33 7e f7 49 b7 88 f4 fb aa 60 db 57 5b 54 8a 33 c2 5c b1 06 20 fe 40 a3 94 86 15 80 4c 10 1f 5e 1e 90 5a fb 0b 76 ca 4f cb e6 dc 35 2e be bc 6f ec f0 1e 1b 0d 7d b9 6c 1d 01 cf 44 db f9 4b b7 5f 37 8e b7 ca b4 a5 00 2f 27 07 9b e0 e7 e3 fb 87 01 81 a1 e3 3f b7 96 99 3e 28 d8 75 28 b2 dd 12 ad 9f a0 e7 3e 17 a1 59 e5 e5 55 3b c1 6b 32 0a d0 fa a3 16 14 f8 70 e9 90 a3 ce ab 30 45 20 2e 40 7c 11 04 4e 85 0b b1 ff fa 15 51 4f b4 36 ba 27 f1 f2 09 93 40 91 56 21 48 b6 7c 67 27 4f 32 d9 f5 88 98 bf 1d f1 dc 73 ef eb c8 81 2f 2c f5 d6 78 b0 9c 06 00 45 0b b4 e6 dc 14 7b ef 53 55 3b 5b be a8 c8 28 6f 27 ce 3d 46 a8 ac 6f 62 b3 07 ef 18 d4 35 db 72 0c a7 57 d2 ba 86 03 1b 40 ff 07 61 c5 ce dc 2d 2c db 20 4c b5 d7 84 c9 26 eb fc 48 36 36 0a bd 14 ed dc 05 f9 96 e6 aa 6b c1 6b 2d e5 e5 bc af 6d 15 78 04 29 e1 f2 ac be 66 6f a0 2a 72 55 9e dd be 26 04 d5 01 5a 38 7f 9b 47 9b 84 3c 6e 70 8c 52 85 b8 cd a3 12 66 2d e9 5d 88 bb 16 95 f9 ce a0 e5 b0 1e 38 ca 88 e2 82 dd 92 16 1d 7b 67 c8 92 61 4b 94 2d b6 55 91 bd 85

6e 0a 00 8e 37 af f9 82 d2 69 51 4b 51 83 4e f1 05 69 17 70 e1 84 09 ac 9b 85 c6 23 34 0f 3e c4 e9 a4 86 9f 56 4d 23 10 4d 54 c4 72 03 74 81 6e 65 10 9f e6 7c f4 db c1 84 14 29 dc 02 e9 a4 21 3a c4 91 8d 55 f9 fe 95 39 86 7a 8c e9 e1 4e 7d 40 dc 6c 4b e2 41 5d 8a ea 01 64 ba a0 5c d6 93 29 4d 56 9c 52 63 bf cc 3f fe b2 7d 9d dd 2d f4 20 51 3a fa 90 dd 41 0e 6e 6f 76 44 ee fd b7 0a df 4f 49 4a 5e 0b 12 f7 58 8c 78 19 da 5e aa e7 65 28 9f 3b 3b e1 a3 3d 20 16 ee 50 88 f0 d6 6d 69 b2 85 01 a9 09 be fc 27 19 9b cd ee 34 8e 8c 19 cc d8 32 99 6e a6 61 1e da 8e 43 2e de 4c ac 73 16 df b3 fb 70 97 50 e1 c6 65 94 53 21 d2 c5 d0 67 e0 8a 95 d0 15 00 2d 7e d1 8a a3 c0 ee 61 01 e0 e0 eb 9c 4c 8b 89 15 da c3 75 2c 7f 88 72 84 b3 70 61 88 e9 47 57 5d e2 4d 88 ff 11 5e e8 35 73 37 af d3 49 84 cc c9 b4 2e 8b 33 39 fe 58 60 60 8e f4 cb 7f 0b fd b6 6b c9 fa 70 2b be 4c 9f cb 44 78 4a 30 e8 f8 0b b5 aa 28 42 1f 3e e2 28 10 8a 4c 0a b6 14 59 b6 bf b0 ac a6 94 cb 89 58 18 b1 ca c9 2d 87 ae 57 b0 18 31 87 62 03 1f e7 ab d7 42 ab bb 4b 27 bf e4 1a 9d e9 5b 3d ca 21 82 db 0c 82 12 7d 48 b0 45 4c b0 a6 1a 99 8a 2e 1a 69 a9 4c 4d 33 35 69 08 19 08 a1 6c d1 89 bf 27 7c 34 58 de c4 a3 b2 c2 ed 42 48 11 cc b5 c1 fe 39 6b 98 73 78 c7 95 6c bb 2f ed e6 54 ca 57 db cb 18 da db d8 49 63 12 76 d8 88 cd af c6 8d ab fc 00 54 29 04 60 ff 92 e4 48 ad d2 1e 20 76 a9 e3 9a 95 1d b1 a1 25 71 7e 6e 23 1c 31 d4 69 9a 0f 46 a8 fe 81 5c 9d 75 c1 c0 6b 22 58 04 a4 11 4d 04 24 fd 81 10 00 63 db 25 b2 4d f7 2a 98 5c 85 38 89 3e ee bb 1f ab f6 a5 ef fa 39 bd 2f 7d c8 c6 91 e8 fd 97 94 18 c9 eb 8d 48 c3 8b 4c a4 c8 b5 73 f8 65 1b 16 0c fb e3 fc 12 e0 b9 6d b4 ba 8d ce 67 be d9 99 dd ff ef 23 7e 1d e1 af a2 8c ca 9e 29 f8 e3 82 cc cf 45 8d ea bb e4 5d f6 61 4d ae 21 a1 1c 6c 95 92 16 61 65 2b 1f 4d 39 4f a1 a7 07 22 2e 97 9c df cf d2 a0 b7 3b 7c d7 35 0f e5 ce ae f1 a3 1d 17 d7 84 9e 8f 09 c9 64

Packet Size:1192

Packet: 00 27 19 1b 43 92 68 5d 43 61 46 5d 08 00 45 00 04 9c 24 17 40 00 40 06 7e 86 c0 a8 01 06 4a 7d 87 93 c1 4f 01 bb 32 09 d2 b3 5f 6a 59 65 50 18 10 c2 de a8 00 00 17 03 01 04 6f d5 0a 0a ea 98 56 a7 3f b3 7f 7c 7e a3 83 bd dc 35 ae ac 29 76 48 73 ed a9 a9 7a 0e a3 36 8f 1d 90 4c a9 19 ed d9 c4 fa cf bb f6 24 5a 40 da 7b be a4 08 65 35 08 53 62 2d e8 00 d0 91 97 80 db 64 1d b2 a6 46 ea ad f2 89

7f 13 af f8 8b c1 3d 40 bc 40 c4 7b f3 c7 05 a7 02 80 c3 ab 5a ed d1 a5 38 cd fc b8 6d e6 59 f8 c9 e5 45 2f c7 be 22 36 cf 08 09 e7 a0 63 2e c4 63 fd fc c8 78 dd 4e 71 f8 cb 53 a4 c7 49 b5 b1 ac 2b e5 c1 3b 14 2f fd fa f8 5c ea b4 02 c0 39 50 c6 8c e1 69 03 41 19 8b 57 cc 1f e1 cf 28 53 24 f1 20 b7 7d 98 bb 75 77 91 87 4a b1 ec 41 85 dd a3 6a 14 71 58 24 47 7f c6 09 83 59 cc 23 db 3e 68 f5 36 ac 42 01 5e c7 9c 2a 66 17 8b ab f1 52 3c 95 46 11 b0 ac 55 8d 13 74 92 e5 76 cc d3 37 ed 92 60 b1 4e 34 11 a5 7e 25 81 ab ab 6d ed 25 e3 62 91 c0 50 03 95 d0 ff d2 49 d8 f4 22 d8 12 64 89 2f 6b e6 ff 27 21 0a 27 95 f7 31 48 f6 e3 b6 c0 e6 db c5 07 ba 40 e4 37 24 64 42 fe 8e d7 b8 a9 fa fb 7e ef 7e 76 7a 64 6f 2c 05 8d 22 57 42 c8 12 88 a4 c0 a8 b7 7f bd d2 f0 f0 1d 2d ff 6f b5 ac 28 7f d3 54 60 99 98 64 93 66 15 0d b6 10 e8 ec 6f df 78 3d 1b 84 ad 15 43 8b de a5 19 a8 b1 cd 1d 3f 14 65 03 35 7b f3 33 15 99 e8 d9 f9 c0 31 21 8b 44 71 8d 8b a4 d2 e9 4b 7d 38 ca 9c 1e 79 25 45 a6 e5 ec ee 87 b7 dd 1d 44 17 09 7e 97 f7 a6 32 31 68 c1 3d 00 53 4e 16 43 65 4c e5 af bd 29 b1 ce 90 04 d5 b7 20 00 df a4 99 86 95 27 ab e5 ca 01 dc d9 50 0a 39 6b 0a 59 44 69 67 0b c9 fa 20 81 40 0f 99 0f 1e 80 71 46 4c 68 46 2d 4f 2e c1 76 c3 68 4f ef 96 ee 03 cd 9e bd 9c 4d c1 01 91 1e 53 ab e1 15 7d 63 d3 55 6b a8 d2 db 8e 23 46 90 c9 98 97 0b 8c 2a ad 8e 9c 49 a6 e3 50 4e c6 fb 83 1d b4 0e 02 74 51 ec dc 7f bf 3d 24 dd bd ec 7a cd 8d a9 f4 7c 0f 77 34 b4 fc ba 33 e6 4f 3b 0c 2b ce 4d 4b 50 59 4c b8 6f bc 9b 0c d3 03 dd c6 86 4c db 55 28 55 28 d7 93 6b d4 5f 92 1d ac d2 bd db 5d f2 bc 09 5b c3 e3 b4 a1 4e 9a 87 eb ba 82 20 63 9b d3 5c e8 31 78 b2 92 85 b5 98 4d fa f2 23 12 04 72 ff 7d 9d 88 1c 75 2b 66 d2 09 09 56 5e 36 e1 8f b9 51 cd 99 e5 03 3c 68 1b 09 5d 18 16 50 4e e6 19 ff 35 f3 be 26 27 d3 30 6c b9 53 32 bf 32 bf a1 ff e0 1d 4a cb b8 ed 56 5a 1a fa 94 4e 90 d5 ba 13 4f 32 c1 86 de cb 37 b7 19 c2 5b 41 00 6a 28 e9 5a ea 76 58 93 66 9f 30 b7 85 25 fb 2d 95 5f 2b 51 a1 19 4f a1 df ce 66 dc d3 fb d9 e3 d2 3b b2 68 3a 93 60 b9 5d 5d de 0f bb 88 ac d8 5f 2a a1 c6 5a 71 d1 87 21 49 41 3e 81 50 30 a2 0b 6a 2f 30 fb c9 82 df 37 ab 7f 75 03 f4 d1 e4 c3 b8 e1 76 fb 9b 34 14 63 1d 09 ba 58 17 77 d0 6d c4 ca 10 cd b0 8b 6f 3f ca da 08 04 ee 7d c3 af 51 66 05 d7 16 e9 a9 2a 0d cf ce 7e 44 f0 95 d2 b8 68 f4 cb 0d 3c 6d 9a ca 0c 77 8d 08 ca 25 69 93 ad 5b 43 4f 80 bf b1 5d c1 2e d0 0a 45 e9 ce db 1d f1 17 e8 42 50 e0 74 fd 59 1b f2 1b 15 7b 81 3a 3d 8d d3 9e 5f d8 ed ea 13 1a 23 04 0a c0 9c 4d 3d 50 cf 0b 23 5e 64 dc 50 9f af 67 ec 60 22 3b 5d 03 6c 2d dc 9e dd fc cf c9 65 8e af c6 a8 89 c5 45 b4 fc 25 6c 4c 2d 8c 0d a2 4c 6a 47 a1 89 23 d5 25 cb e7 18 69 e3 8c 01 e3 d0 eb db 92 01 ff 58 96 ae 07 05 0c db 44 af 7d 90 26 bb dd 67 55 ae 7d 1c 8a 3a 14 4a e4 71 03 47

55 7f 2c 0c 08 ba c1 98 af 66 1d 65 9d 05 c8 8f d9 0b cb 89 9c dc 4d 29 35 48 42 b0 3f ba 04 b6 0e 5e 10 4f 74 a9 69 72 c9 57 06 f8 51 50 02 a2 aa a1 66 bd 78 a4 df ea 15 40 90 fe ca ba 15 09 67 d6 13 27 38 6b ba d4 b2 4e 8d 9e 2d 73 3c b3 17 d8 da 25 f9 17 40 00 00 26 ae fb e0 32 7b 3b ef 22 c6 e8 f9 b5 23 18 85 52 a9 88 cc 1f 09 28 b9 65 eb f5 5d 4c 48 e3 0f a7 16 10 b1 de 26 17 a9 f3 46 70 e4 ac df 1d 32 8a 79

Packet Size:1194

BUILD SUCCESSFUL (total time: 9 seconds)

5.5 Result

The packets displayed in the K2 cluster are anomalous packets since they belong to cluster two. They are the ones having total length greater than length set in the policy for Intrusion Detection System. The module Anomaly Detection of proposed IDS can identify anomalous data by classifying network traffic into two groups: One having normal behaviour and the other diverting from expected behaviour.

From the ninety two packets sent as input to the KSVM code, four are anomalous packets, the ones diverting from normal behaviour. The result was checked by giving different input packets and code was able to detect anomalous packets for every input. There is no restraint on the size of packet or the number of packets to be input.

Chapter 6
Conclusion and Future Work

6.1 Conclusion

The objective of this thesis is to propose a novel hybrid approach for Intrusion Detection System. To accomplish this objective, first the stages of an attack, types of attacks are analyzed to understand at which stages an attack can be detected. Then counter attack technologies such as firewall, network-based antivirus system are examined to determine when, how and why they are used. Also, the limitations of those technologies is defined which are the cause to the new IDS technology. Then a detailed study on intrusion detection systems was carried out. Finally an adaptive and distributed model for Intrusion Detection System is presented. The model uses the data collected by the sniffer agents of host nodes to detect signature attacks. Novel attacks are detected at the next level by anomaly based centralised node. For this purpose different machine learning algorithms were compared and a hybrid algorithm of the existing k-means and SVM is proposed. This hybrid machine learning algorithm called KSVM makes the model adaptive. The algorithm clusters the network traffic into normal and anomalous data. Compared with previous works, this solution has several advantages. First and foremost, our model detects novel attacks. Second, this model significantly reduces the overall load of an IDS system because it distributes the load on different nodes. Lastly, high false negative rate of adaptive IDS is taken care of by making the some components anomaly based.

6.2 Future Scope

In this thesis we showed an adaptive distributed IDS and its machine learning algorithm. Although the system proposed has distributed architecture but task of identifying attacks is confined to centre.

So our future direction will therefore focus on extending the anomaly based component to individual nodes like the signature based agent. This will make the system complex but will make the system respond in real time and independently. There can be drawback of increased overhead and overuse of resources. In future, an

approach can be devised which will control the load of resources as well as respond in real time. Also a feature of exchanging suspicious activity among different nodes can be devised so that they communicate directly instead of through centralised node.

References

[1] Fossi, Marc, Gerry Egan, Kevin Haley, Eric Johnson, Trevor Mack, Téo Adams, Joseph Blackbird "Symantec internet security threat report trends for 2010", vol. 16, 2011.

[2] Redwan, Hassen, and Ki-Hyung Kim, "Survey of security requirements, attacks and network integration in wireless mesh networks", In New Technologies, Mobility and Security, IEEE, 2008, pp. 1-5.

[3] Wang, Yong, Garhan Attebury, and Byrav Ramamurthy. "A survey of security issues in wireless sensor networks", 2006.

[4] William, Stallings, Cryptography and Network Security, 4/E. Pearson Education India, 2006.

[5] Stallings, William. Network Security Essentials: Applications and Standards (For VTU). Pearson Education India, 1982.

[6] Wang, Haining, Danlu Zhang, and Kang G. Shin, "Detecting SYN flooding attacks", INFOCOM 2002, Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, Proceedings, vol. 3, 2002, pp. 1530-1539.

[7] Ansari, Sabeel, S. G. Rajeev, and H. S. Chandrasheka, "Packet sniffing: a brief introduction", Potentials, IEEE 21, vol. 5, 2002, pp.17-19.

[8] Tso, Michael Man-Hak, and Bikram Singh Bakshi, "System for virus-checking network data during download to a client device", U.S. Patent 6,088,803, July 11, 2000.

[9] Wang, Yi-Min, Roussi Roussev, Chad Verbowski, Aaron Johnson, Ming-Wei Wu, Yennun Huang, and Sy-Yen Kuo, "Gatekeeper: Monitoring Auto-Start Extensibility Points (ASEPs) for Spyware Management", vol. 4, 2004, pp. 33-46.

[10] S. Nassar, A.E. Sayed, N. Aiad, “Improve the Network Performance By using Parallel Firewalls,” Proc. of 6th International Conference on Networked Computing, May 2010, pp. 1-5.

[11] Zalenski, Robert, "Firewall technologies." Potentials, IEEE 21, vol. 1, 2002, pp. 24-29.

[12] Acharya, Subrata, Jia Wang, Zihui Ge, Taieb F. Znati, and Albert Greenberg, "Traffic-aware firewall optimization strategies," IEEE International Conference, IEEE, vol. 5, 2006, pp. 2225-2230.

[13] A. S. Ashoor and S. Gore, “Importance of Intrusion Detection system (IDS)”, International Journal of Scientific and Engineering Research, vol. 2, Jan-2011, pp.1-4.

[14] Choudhary, Amit Kumar, and Akhilesh Swarup, "Neural network approach for intrusion detection", Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, ACM, 2009, pp. 1297-1301.

[15] V.Visoottiviseth, U. Jaralrungroj, E. Phoomrungraungsuk, P. Kultanon, “Distributed Honeypot log management and visualization of attacker geographical distribution”, Proc. of Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE), May 2011, pp. 23-28.

[16] Pearson, Malcolm E., Leon R. Warman, Robert G. Atkinson, David R. Reed, and Steven D. White, "Automatic elimination of viruses and spam", U.S. Patent Application 10/942,632, September 16, 2004.

[17] Fortinet, Inc. “Improving Network Protection and Performance with Network-Based Antivirus Technology,” White paper, Oct. 2002.

[18] Grace, Clive, and I. T. Journalist. "Understanding intrusion detection systems", PC Network Advisor 122, 2000, pp. 11-15.

[19] Cisar, P., and S. Maravic Cisar, "Intrusion detection-one of the security methods", Intelligent Systems and Informatics, 2008, 6th International Symposium, IEEE, 2008, pp. 1-6.

[20] A. Lazarevic, V. Kumar and J. Srivastava, “Managing Cyber Threats: Issues, Approaches and Challenges, chapter: A survey of Intrusion Detection techniques”, Kluwer Academic Publishers, 2005.

[21] Debar, Hervé, Marc Dacier, and Andreas Wespi. "A revised taxonomy for intrusion-detection systems", In Annales des télécommunications, Springer-Verlag, vol. 55, 2000, pp. 361-378.

[22] Osareh, Alireza, and Bita Shadgar. "Intrusion detection in computer networks based on machine learning algorithms", International Journal of Computer Science and Network Security (IJCSNS), vol. 11, 2008, pp. 15-23.

[23] Burges, Christopher JC, "A tutorial on support vector machines for pattern recognition", Data mining and knowledge discovery, vol. 2, 1998, pp. 121-167.

[24] Mulay, Snehal A., P. R. Devale, and G. V. Garje, "Intrusion detection system using support vector machine and decision tree", International Journal of Computer Applications 3, vol. 3, 2010, pp. 0975-8887.

[25] Xie, Lixia, Dan Zhu, and Hongyu Yang, "Research on SVM based network intrusion detection classification", In Fuzzy Systems and Knowledge Discovery, Sixth International Conference, IEEE, vol.7, 2009, pp. 362-366.

[26] Wang, Jiaqi, Xindong Wu, and Chengqi Zhang, "Support vector machines based on K-means clustering for real-time business intelligence systems", International Journal of Business Intelligence and Data Mining 1, vol. 1, 2005, pp. 54-64.

[27] Wang, Jiaqi, Xindong Wu, and Chengqi Zhang, "Support vector machines based on K-means clustering for real-time business intelligence systems." International Journal of Business Intelligence and Data Mining, vol.1, 2005, pp. 54-64.

[28] Na, Shi, Liu Xumin, and Guan Yong, "Research on k-means clustering algorithm: an improved k-means clustering algorithm", In Intelligent Information Technology and Security Informatics (IITSI), Third International Symposium, IEEE, 2010, pp. 63-67.

[29] Sarma, T. Hitendra, P. Viswanath, and B. Eswara Reddy, "A hybrid approach to speed-up the k-means clustering method", International Journal of Machine Learning and Cybernetics , 2013, pp. 1-11.

[30] Mishra, Bikram Keshari, Amiya Rath, Nihar Ranjan Nayak, and Sagarika Swain, "Far efficient K-means clustering algorithm", Proceedings of the International Conference on Advances in Computing, Communications and Informatics, ACM, 2012, pp. 106-110.

[31] Agarwal, Shalove, Shashank Yadav, and Kanchan Singh, "K-means versus k-means++ clustering technique", Engineering and Systems (SCES), 2012 Students Conference, IEEE, 2012, pp. 1-6.

[32] Finley, Thomas, and Thorsten Joachims, "Supervised k-means clustering", 2008.

[33] Fraley, Chris, and Adrian E. Raftery, "How many clusters? Which clustering method? Answers via model-based cluster analysis", The computer journal, 1998, pp. 578-588.

[34] Reddy, PC Chenna, and R. Siva Sankara Reddy, "K-Means Algorithm with Different Measurements in Clustering Approach", 2012.

[35] Tang, Jun, "Improved K-means Clustering Algorithm Based on User Tag", Journal of Convergence Information Technology, 2010, pp. 124-130.

[36] Ben-Hur, Asa, David Horn, Hava T. Siegelmann, and Vladimir Vapnik, "A support vector clustering method", Pattern Recognition, 15th International Conference on, IEEE, vol. 2, 2000, pp. 724-727.

[37] Xia, Xiao-Lei, Michael R. Lyu, Tat-Ming Lok, and Guang-Bin Huang, "Methods of decreasing the number of support vectors via K-mean clustering", In Advances in Intelligent Computing, Springer Berlin Heidelberg, 2005, pp. 717-726.

[38] Vishwanathan, S. V. M, and M. Narasimha Murty, "SSVM: a simple SVM algorithm." Neural Networks, Proceedings of the 2002 International Joint Conference, IEEE, vol. 3, 2002.

[39] Finley, Thomas, and Thorsten Joachims, "Supervised clustering with support vector machines." Proceedings of the 22nd international conference on Machine learning, ACM, 2005.

[40] Liu, Jianxiao, and Lijuan Li, "A Distributed Intrusion Detection System Based on Agents", Computational Intelligence and Industrial Application, Pacific-Asia Workshop, IEEE, vol. 1, 2008, pp. 553-557.

[41] Huang, Weijian, Yan An, and Wei Du, "A Multi-Agent-based Distributed Intrusion Detection System", Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on, IEEE, 2010.

[42] Tian Li, "Design and Implementation of Distributed Intelligent system", International Conference on Electrical and Control Engineering, 2010.

[43] Y.F. Zhang, Z.Y. Xiong, and X.Q. Wang. Distributed Intrusion Detection based on Clustering, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 2005, pp. 2379–2383.

[44] M. Reh´ak, M. Pechoucek, P. Celeda, J. Novotny, and P. Minarik, CAMNEP:Agent-Based Network Intrusion Detection System. In Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems, Portugal, 2008, pp.133–136.

[45] Lee, Wenke, Salvatore J. Stolfo, and Kui W. Mok, "Adaptive intrusion detection: A data mining approach", Artificial Intelligence Review 14, vol. 6, 2000, pp. 533-567.

[46] Huang, Ming-Yuh, Robert J. Jasper, and Thomas M. Wicks, "A large scale distributed intrusion detection framework based on attack strategy analysis", Computer Networks 31, vol. 23, 1999, pp. 2465-2475.

[47] Botía, Juan A., Jorge J. Gómez-Sanz, and Juan Pavón, "Intelligent data analysis for the verification of multi-agent systems interactions", In Intelligent Data Engineering and Automated Learning–IDEAL, Springer Berlin Heidelberg, 2006, pp. 1207-1214.

[48] Liu, Jianxiao, and Lijuan Li, "A Distributed Intrusion Detection System based on Agents." Computational Intelligence and Industrial Application, 2008, Pacific-Asia Workshop, IEEE, vol.1, 2008.

[49] Jaisankar, N., Swetha Balaji, S. Lalita, and D. Sruthi, "Intrusion Detection System Using K-SVMeans Clustering Algorithm".

[50] Denning, Dorothy E, "An intrusion-detection model", Software Engineering, IEEE Transactions, 1987, pp. 222-232.

[51] Pontil, Massimiliano, and Alessandro Verri, "Support vector machines for 3D object recognition", Pattern Analysis and Machine Intelligence, IEEE Transactions on 20, vol.6, 1998, pp. 637-646.

List of Publications

[1] Parneet Kaur, V P Singh, "Distributed IDS having machine learning module using k-means SVM", accepted in International Journal in Enhanced Research in Management and Computer Applications, ISSN No: 2319-7471, June 2013.

[2] Parneet Kaur, V P Singh, "Applying k-means SVM on live traffic to differentiate the anomalous data", communicated to International Journal of Computer Science and Software Engineering, July 2012.