REVIEW INTRUSION DETECTION CLASSIFICATION BY METAHEURISTIC OPTIMIZATION WITH MACHINE LEARNING APPROACHES

Vibha Paneri[1], Mr. Girish Ameta[2]
Geetanjali Institute of Technical Studies Dabok, UDAIPUR INDIA

Abstract- The main objective of intrusion detection systems (IDS) is to discover the dynamic and the malicious form of network traffic that simply changes according to the characteristics of the network. The IDS methodology represents a prominent developing area in the field of computer network technology and its security. Different form of IDS has been developed working on distinctive approaches. One such kind of approach where it is used is the machine learning mechanism. In the proposed methodology an experiment is applied on the data-set named as KDD-99 including its subclasses such as denial of service (DOS), other types of attacks and the class without any form of attack. Depending upon the machine learning algorithms various distinct forms of IDS have been developed which further checks the optimization based potential features in connection with the neural network classifier for the various forms of IDS based attacks. This approach provides a comparative study between the ANN and the optimizer-based ANN technology. The experimental analysis shows the convolution neural network with SVM show effective analysis providing accurate forms of IDS thereby improving its detection based on individual class along with maintaining its results fundamentally.

Keywords-*Intrusion Detection System, Denial of service, Artificial Neural Network, Genetic Algorithms, Artificial Neural Networks.*

I. INTRODUCTION

In the present scenario the use of internet is growing at a large pace with is highly developed and emerging forms of ever growing network and its connectivity but the use of internet poses a great threat to cyber security. In order to maintain the high level of security there is an important need to overcome the cyber threats posing problems to various organizations, companies, and the firms. One of the major challenges among the cyber-security is to maintain the integrity of the intrusion detection system (IDS) thereby protecting it from major forms of attacks and to conquer the various form of risks of the intruded system [3]. The main function of the IDS is to identify a more precise form of intrusion. The illegal hackers of the security have found a large number of ways to break the security of the system whether it is a cloud network or the wireless-based network. Many researches have been performed by the technologists to curb the security threats from distinct forms of intrusions done to the cloud computing systems and the wireless system. So, the main objective of IDS is to protect the information whether it is governmental, public or private entity [5]. The use of IDS is mainly required in detecting the false and the poor detection rates. Whenever an attack is observed by the system or a harmful activity is done to the system, it automatically generates an alarm resulting in a false-positive alarm [1]. The research mainly focusses upon the enhanced capabilities of the intrusion detecting system and thereby reduces the occurrence of the false type alarms.

1.1 IDS: Overview

The term intrusion detection system i.e. IDS is a developing area having various forms of application in the computer technology and its inter-linked networks. Some of the important forms of IDS which identifies the traffic-data and its changing activities by using an algorithm (single class). But some of the single-class algorithms are not able to fetch a good detection rate and does not provide a low occurrence of the false alarms. So, the working methodology is based on using an intelligent hybrid technology comprising of different sets of classifiers which are helpful in enhancing the productivity of the system in an intelligent way. In IDS intelligent based mechanism various forms of data mining approaches such as Genetic Algorithms, Classification, Decision Trees, Artificial Neural Networks, and clustering have been used in the mining of data for the development in the field of IDS also the SVM i.e. support vector machines technology provides the best technique for classification of the clean as well as the intrusive form of data [2]. The main requirement of the IDS is not only to encounter the intruders in the data path but also to supervise the intruders of the data. The most important security aspects of an intrusion detection system consist of maintaining the following conditions.

- *Confidentiality:* Only an authorized user can detect the system.
- *Availability:* Here, the computer technology provides various forms of resources and the access to the legal users of the system without disturbing the working operation of the system.

- *Integrity:* The information must be protected from any kind of malicious act.

The following figure.1 explains the general structure of an intrusion dectection system.

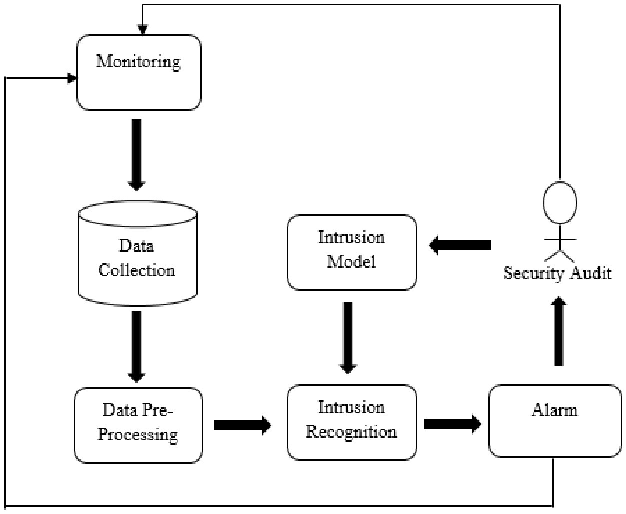

Figure 1: Basic structure of IDS

In order to increase the performance of the IDS, a method known as soft computation is done. The term “soft computing” refers to the process of different methods to get the best possible finite results. The eminent technology of Artificial Intelligence and the machine learning processes has resulted in accuracy and thereby providing the best suitable results as per the requirement. It has shown a great success in the IDS mechanism. The are various distinct forms of soft computing methods used in IDS detection such as Support Vector Machine [SVMs], Artificial Neural Network [ANNs], Genetic Algorithms [GA], Bayesian Networks, and Fuzzy Logic. In case of human eyes the researchers use the AI techniques to identify the intrusions that is the main reason why the researchers use the data mining processes and the artificial intelligent techniques to explore the feasible intrusions.

1.2 IDS: Architecture

The architecture of IDS comprises of its unique core element i.e. sensor popularly known as the analyzing engine to pin-point the intrusions occurring in the system. The sensor consists of a mechanism that helps in detecting the intrusions.

The sensor consists of a component known as event generator which performs the data collection shown in figure.2. It detects the way of collecting the data. The event generator consists of network, operating system and the network applications where it generates a set of events including audit (log) of the system or the packets of the network. This form of set events also involves the policy of information collection i.e. in or out of the system. Sometimes it is not necessary to store the data as it reaches simply to the analyser. So, basically the key role of the sensor is to extract or filter the data and remove the unwanted form of the data that is achieved from the event data set system [4]. Additionally, the database holds the configurational parameters of IDS that includes its mode of communication methods based on the response module. The sensor itself contains its own data observing all the historical multiplex forms of intrusions. Practically, the IDS may follow a structure based on an ‘agent’ principle where small modules (autonomous) are designed on ‘per-host’ basis approach. The agent mainly monitors and filters the activities scheduled within the area i.e. fully protected and further starting its initial analysis by undertaking a response action [10].

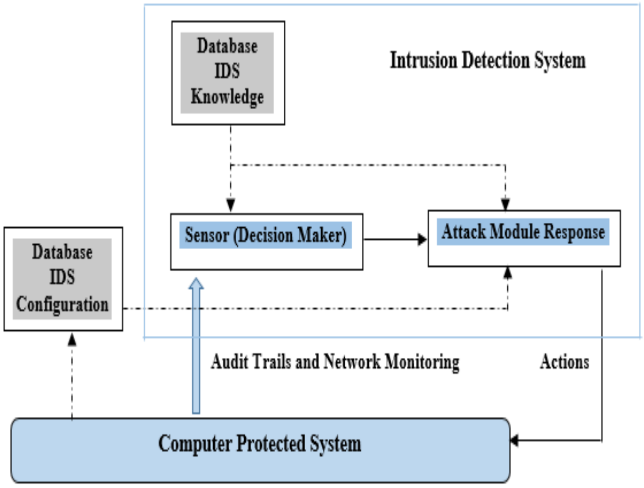
Database IDS Knowledge
Intrusion Detection System
Sensor (Decision Maker)
Attack Module Response
Database IDS Configuration
Audit Trails and Network Monitoring
Actions
Computer Protected System

Figure 2: Sample IDS (arrow width ∞ information between system components)

1.3 IDS: Classification and Types

There are various categories of IDS based on structure or detection. The IDS are classified based on characteristics as represented below in figure 3.

1.3.1 Based on Structure

The process of IDS is divided into three of its important categories based on its framework. These are Network based IDS, Host Based IDS and Application Based IDS.

1. Network Based Intrusion Detection System [NIDS]: The Network based IDS represents a passive network analyzing the traffic related to the network and for finding out the evidence of various forms of attacks. When a NIDS detects an attack, it provides an instant report to the administrator. It basically checks the types of attack that are incoming and outgoing networks and is usually placed inside the router. But the NIDS is unable to find out the encrypted source of information and is not able to distinguish some forms of attacks. There is no effect of system-failure over the NIDS. Being autonomous in nature these systems are simple to run and easy to install [3]. The NIDS consists of some advantages and disadvantages explained as follows [8] [11]:

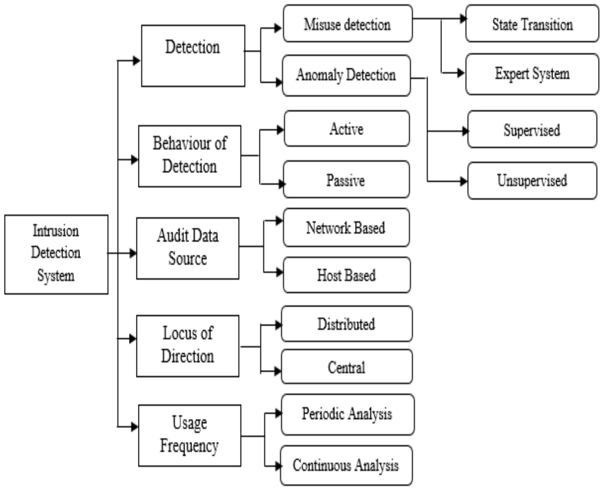

Figure 3: Classification of IDS based on its characteristics

Advantage of NIDS

- The passive network helps in maintaining the ongoing working operations of the system.
- With the use of this simple setup it becomes easy to monitor the network operations.
- These systems do not get exposed easily when a certain form of direct attack occurs.

Disadvantages of NIDS

- When the network becomes large these systems are not able to identify the type of attacks.
- The NIDS are not able to pin point the encrypted source of data that results in a degraded quality of its performance.
- Some forms of attack are not identified due to high level of malicious data content.

2. Host Based Intrusion Detection System [HIDS]: The type of detection that is placed in the computer server represents the host of the system usually called as HIDS. As the name suggests a mechanism that helps in analyzing the stored and the system files and further tells about the changes or the deletions done by the attacker in the system files. The HIDS simply detects the part i.e. not detected by the NIDS mechanism. These are more liable to the attacks that are direct in nature and are inclined to attacks based on DOS i.e. Denial of Service. These systems have low type of false positive rates as the command is implemented on the host (definite) which are more influencing than the types of attacks done across the network [14]. When the system file is interrupted, the system gets active and generates an alarm. Some of the examples of HIDS includes the Tripwire, CISCO HIDS and Symantec ESM.

Advantages of HIDS

- Host-based IDS detects the deformity present in the network.
- If the switched network gets exploited then it does affect the HIDS in any form.
- It helps to solve the confusing attacks present in the NIDS methodology.

Disadvantages of HIDS

- The HIDS is a sensitive system based on DOS type attacks.
- These are time consuming.
- When the attack is done against the host or if it is a direct form of attack, then the problem of data loss and the loss of its functionality occurs.
- Large amount of disk space is required that degrades the system's quality or performance.
- It is not able to pin-point the non-host or the multi-host devices of the network.

3. Application based IDS: The application-based IDS is another development of HIDS which monitors the different types of events such as the inspection of the files, checking the abnormal functions like exceeded permission, void-file execution, etc. It helps in analyzing the communication between the user and the application and monitors the traffic of the network i.e. encrypted [9].

1.3.2 Based on Detection Method

In-order to build a smart IDS, the main aim of the system is to reduce the various types of false alarms i.e. the positive and the negative alarms. Based on the detection method the IDS is divided into two of its main categories that includes the Misuse detection and the Anomaly based detection.

1. Misuse detection: It is also known as the signature-based IDS designed to compare the signatures or the patterns that are made over the incoming path of the traffic network. These signatures help in detecting the attacks in a very accurate manner. The main aim of misuse detection is that it helps in finding the eminent forms of attack. But one major disadvantage of the system is that it cannot detect the new forms of attack with the changed form of signatures resulting in negative false alarms due to which it deals with a vast number of negative false alarms. The typical misuse-detection is given as follows:

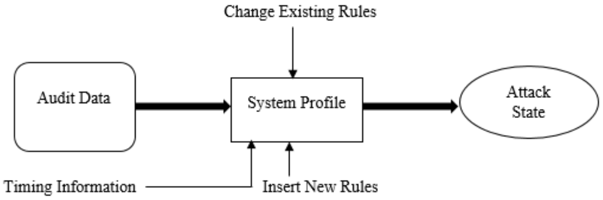

Figure 4: Typical Misuse Detection

2. Anomaly IDS: This term anomaly IDS is also known as statistical IDS i.e. it monitors the network traffic identified as a normal method deriving a potential base-line. In order to determine the intrusion activities of the system, each network is observed at regular intervals and further matched with the base-line of the system. The process basically requires statistical as well as the behavioral models that are used for detecting the attacks that allows the false-negative rate whereas the presence of an attack is determined by the patterns of the programs or the users depending upon the normal or the abnormal activities of the system. The anomaly detection is represented as below in figure 1.4

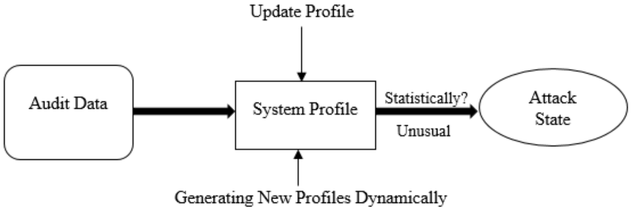

Figure 5: Typical anomaly detection

1.3.3 Based on attack

The attack-based IDS are categorized as follows [4]

1. Normal Category: This category does not have any form of data attack. It consists of a state where the system has no alteration and no such kind of abnormality occurs in the state of the system.

2. DOS-Attack: This is generally known as the denial of service attack. Here, in this form of attack, the hacker or the attacker of the system perform various form of illegal activities such as illicit calculations, makes the data memory typically jammed by sending the malignant data sources or the data packets in such that it is not able to maintain the authentic activity of the system. The attacker performs a Botnet attack and takes the advantage of the distance. The process of DoS consists of various types of attacks [9]:

- *Operating DoS Attack:* The operating form of attack deals with bugs created by the system and fixes them with the help of patches. This form of attack is also known as "Teardrop" where the attacker of the system takes the advantage of TCP/IP fragmentation creating risk to codes of re-assembly.
- *Networking Attacks:* This type of attacks puts certain restrictions on the protocols of the network. One such example of example of this kind of networking is "SYN flood" attack. Here, the attacker creates an IP spoofing with excessive links that are half-open. Further, the attacker implements the operation by sending the data to the victim in the form of SYN-packets with IP addresses that are spoofed. At the end, the victim checks the data sent and generates the SYN/ACK response to the spoofed (fooled) IP address. The attacker in response blocks the computer of the victim.

3. Probe Attack: This type of attack consists of collecting the information, analyzes the network operation in order to extract the valid form of IP address to pin-point the distinct services used for the network for performing a smart and wise attack on these services [24]. Various forms of probe-attack include the following:

- IPsweep: Identifies the service on a particular (specific) port.
- Portsweep: Detects and monitors the port services that are hosted by the single host).
- Nmap: It represents a form of tool for network mapping

4. Remote to Local (R2L) Attack: This type of attack is done when the user gets the access of system or it finds a root/link through the system (remote) to perform an attack. Sometimes in case of R2L attacks, the most common pathway to enter into a system is the internet.

5. User to Root (U2R) Attack: The process of user to root attack defines the activity where an attacker opens a fake account, make the system weak or creates the bugs into the system by squandering the authorization processes. The most commonly used U2R attack is the flow of the buffer where an attacker takes the advantage of the fault occurring in the program and congregate the additional information into a buffer i.e. kept on an execution stack.

II DATA MINING TECHNIQUES

The processing of data from the different sources results in gigantic data-sets that cannot be analysed properly. So, by analysing the sources of data-set, the data-mining techniques plays a significant role in revealing the hidden data source and the normal or abnormal forms of patterns. This particular section states the different forms of data-mining techniques in order to detect the various forms of attack observed in the network [12].

2.1 Association rules

This is method which identifies the connection or association between the variables in large amount of data-sets, association among the data attributes and helps in determining the system values. As the nature of this rule is based on pattern discovery so, we cannot rectify the problems related to classification and prediction. In association rule mining process two of the threshold values are considered. One is the maximum support and the other is the minimum confidence

2.2 Classification

When each sample of data set is assigned to a unique form of class then it is termed as the process of classification. Generally, it is used for signature-based technique but it can also be used for anomaly-based detection technique. In this type of technique, firstly, the datasets which are available are predefined. There are various types of classification techniques as explained below:

1. Decision Tree: It is well known recursive method forming a structure like a tree. Here, the divide and conquer methods are adapted for segregating the attribute value. The process of classification starts from the root-node towards the path of the leaf node. The root-node denotes the values of the attribute whereas the leaf node denotes the class-label. A large set of data tree gives the excellent performance rate.

2. ID3 Algorithm: It is an algorithm based on attributes creating a decision-tree on the basis of trained data-sets. It is used in natural as well as the machine leaning methodologies. The mechanism of ID3 helps in constructing the information and the entropy gains to design a decision tree.

3. J48 Algorithm: This is a form of C4.5 algorithm which constructs a decision tree based on the information gain of an attribute denoting the high level gain. But the disadvantage of using this algorithm is that it require more time for central processing unit to run and needs a huge space for memory [6]. In, J48-algorithm, set of rules are produced by analyzing decision- based tree

4. Random Forest: This technique is based on random analysis where each tree is designed by distinct data-sets on random based selection. A high-quality dimensional data can be handled easily in this form of method [12, 13].

5. Naive Bayes classifier: This a probability-based classifier method with the assumption based on the membership probability. It works typically on the relation among variables i.e. dependent and independent variables that derives the probable conditions:

$$P\ (H/X) = \{(P(X/H).\ P(X))\}/\ P\ (H)$$

where, X = recorded data, H = hypothesis, P (H) = prior probability, P (H/X) and P(X/H) = posterior probability. The Naive-Bayes classifier can be easily designed without the use of iterative complex parameters.

2.3 Genetic Algorithm

Genetic algorithm represents a best technology for data-mining technology that selects can hold the information from a vast collection of data or a data-box, further finding the different operating modes to gather the accurate results. This is based on the theory of natural evolution. The fitness function evaluates the quality of each and every rule [8]. The main properties of this genetically based algorithm is that it depends upon the self-learning and the robustness properties. So, these are very helpful is detecting high rates, wide space for solutions and the low-false positive rates.

2.4 Neural Network

The term neural network represents a paradigm for the process of information system i.e. based on working of the biological nervous systems. It represents a set of elements that are processed highly consisting of linked or interconnected nodes which produces an alteration to the input-nodes creating the desired form of output, where every node is connected such that it forms an adequate connection in its neighboring-layers. It consists of an input-layer, hidden -layer, and the output-layer [11] [14]. The input-layer carries the input, the hidden-layer focusses upon data processing obtained from the input-layer, and the output-layer denotes the output of the system. There are two types of learning done through the neural networks i.e. the supervised and the un-supervised learning. Thus for maintaining high accuracy the Multilayer Perception (MLP) is used.

2.5 Markov Model

This method is based on the approaches of learning techniques. Here, the states that are definite in nature in HMM i.e. Hidden Markov Model are controlled by the transition-probability sets. After, the probability-distribution mechanism, and output gets generated and this process repeated again and again till the desired results are not achieved. The HMM uses it calling methodology to detect the intrusions of the system. Hidden Markov Model (HMM) is also used to detect intrusions using the system calls.

II. RELATED WORK

Kemmerer et.al [1] presented a study by framing a simple question of why there is a need of intrusion detection system. Suppose, the owner of a house is out of town and he has locked his home with all the windows and doors closed. But, there is someone outside his home who wants to enter. Firstly, he rings the bell and checks the main door if it is locked or not then after sometime he checks the windows of the house that too are locked which makes sure that the house is safe. So, the question is why an alarming bell is installed. This question particularly sticks to the IDS. Why there is a need to plant the detection systems if the security is tight and secure. The reason to install these detective systems is that the intrusions still exist because sometimes the people may forget to lock their doors or windows, the same case occurs with the computer based networks which do not provide us 100% security of the system to work accurately. So, based on this study the researchers has tried to explain the techniques based on IDS to deal with these kind of intrusions present in the network. Steven T et.al [2] proposed a study on an application of STATL that represents a descriptive language based on a transition-based attacking system that is constructed to support the IDS. This form of descriptive language describes a process of penetration done to the computer network implemented by a hacker. These type of penetrations includes attacking activities performed by the hacker. The STATL description is used by the IDS to extract the stream events and the ongoing intrusions occurring in the system. As the IDS works under distinct environments such as Windows NT, Linux etc. and the domains like the host or the network. So, this extensible form of language helps in dealing with different targets as required. This language basically describes both the host and the network attacks. Here, in this paper an IDS based tool-set i.e. based on the descriptive language has been executed. This tool-set depicts various favorable and the desires results. There is a deep study of syntax based on the STATL language. Common real examples of both the network and the host are also described in the paper. Pi-Cheng et.al [3] conducted a research based on two of its issues related to the IDS designs. The two issues include the selection based on optimization of rule-based selection and the discovery in case of attack. This type of approach provides a connection between the junked packets. An algorithm is implemented for the attack identification and the rule based selection. The study is performed on the threats and describes the relationship for an application based web-server and the gateway. The algorithm is implemented over a signature-based IDS for having the better form of results. Cavusoglu et.al [4] conducted a research on security systems of IT. The information technology firms rely on various forms of technologies such as IDS and the firewalls to manage the risks of the organizations. There exists some most interesting facts related to security alerts in IT industries. This paper presented a study to demonstrate the values of IDS adopted in an IT company. The configuration of IT was represented by the true-positive and the false-positive rates which further consists of determining the negative or the positive rates of an organization. It was shown specifically that an organization or a firm experiences a positive-rate from an IDS based on one of the condition that the rate of detection is more than the critical value. When a firm experiences a positive or a negative value, an IDS prevents the occurrence of hackers that means an IDS targets the hacker's activity whether the alarm is positive or negative as the rate of detection is same. The results so obtained showed that the positive rate detected by an IDS is the result of increased amount of deterrence enabled by its improved detection. The use of optimized form of IDS indicates that the firm experiences a value i.e. non-negative in nature. Kim, Jungwon, et.al [5] conducted a research on the use of artificial immune systems in IDS which is an interesting concept that relied on two main reasons. Firstly, the immune system of a human provides the best protection. Secondly, the present techniques used for maintaining the computer security are less reliable and complex in nature. Here, the researchers have used various distinct algorithms for the development of the systems and the best possible outcomes. The analysis has been done based on the important developments within this area of research, in addition to forming suggestions for future research options. Panda, et.al [6] worked on the mining techniques if the data that are applied in designing the IDS in order to secure computational resources against access i.e. unwanted. This paper has shown unique performance of well-defined data-mining classifier-algorithms such as ID3, J48 and Naïve Bayes that have been evaluated based upon 10-fold-cross validating test.The data that has been used is KDDCup'99 IDS which further shown that the Naïve Bayes method is the most effective algorithm of learning based process, and the mechanism adopted for decision trees is more interesting for the purpose of detection. Zhang, J., et.al [7] proposed new frameworks that involved the use of a data mining algorithms such as the hybrid-network-based IDSs, random-forests in misuse, and an anomaly based detection. The hybrid mechanism has improved the performance of detection with the combination of misuse advantages. Here, the detection analysis was done on KDD'99 data-set the Knowledge Discovery and Data Mining. In case of misuse-detection, automatic intrusions based patterns are built using random-forests algorithm over trained data-sets. After this approach, the intrusions are detected by network-based matching activities against the patterns. Whereas in anomaly detection approach, novel forms of intrusions are detected by the outlier detection of the random-forests algorithm. In the end the patterns are built by the random forest algorithmic approach, the pattern relating outliers are obtained. The results demonstrate that the use of misuse detection approach was much better than the best KDD'99 data-set approach that provided low false rate, high amount of detection rate that resulted in an overall increased performance of the IDS system. Aydın, M. Ali, et.al [8] proposed a hybrid-IDS by joining the two types of approaches in single system. These two approaches include and network traffic anomaly detection (NETAD) and packet header anomaly detection (PHAD) that are basically an anomaly-based IDSs with use of the misuse form of IDS Snort comprised of an open-source project. The so called hybrid-IDS is examined by the using of data-base lab, MIT Lincoln Laboratories network traffic data (IDEVAL). The results represented a comparative behavior of the attacks detected by misuse-based IDS, with the hybrid-IDS and showed that the hybrid based IDS are more powerful as compared to misuse-based detection. Wolfgang Banzhaf, et.al [9] researched on Intrusion detection based that are based on the computational intelligence In order to build a good model of IDS, it should include the important features of computational intelligence (CI) systems that consists of high computational speed, fault tolerance, adaptation, and error resilience properties. Here, the study has provided an overview to the problem of intrusion detection based on CI systems. The scope has encompassed CI core-method, including evolutionary computation, artificial neural networks, evolutionary computation, artificial immune systems, soft computing, fuzzy systems, and swarm intelligence. The research has summarized that allowed us to clarify the research challenges that are existed already, and highlights the methods by promising new research solutions. The findings survey has provided useful methods to conduct the research in the current IDS technology.

Table.1 Existing Scheduling Model

Author's Name	Year	Methodology Used	Proposed Work

Zhang, J., *et.al*	2008	Data Mining Algorithms	Proposed new frameworks that involved the use of a data mining algorithms such as the hybrid-network-based IDSs, random-forests in misuse, and an anomaly based detection.
Jungwon, *et.al*	2007	Artificial Immune Systems	Conducted a research on the use of artificial immune systems in IDS which is an interesting concept that relied on two main reasons.
Aydın, M. Ali, *et.al*	2009	Network Traffic Anomaly Detection (NETAD) and Packet Header Anomaly Detection (PHAD)	Proposed a hybrid-IDS by joining the two types of approaches in single system.
Deepika P Vinchurkar, *et.al*	2012	Support Vector Machine	Conducted a research on Intrusion Detection Systems that consisted of high-level security of networks and thus provides the system dealing with security of network and the intrusion based attacks.
Wang, Huiwen *et.al* [14]	2017	SVM model	Proposed a methodology that focused on the fact that the security of the network has been increased at a very large pace for all the organizations, firms, and the most important is the security of an individual

Muamer N., et.al [10] conducted a study on using smart and intelligent form of data-mining approaches to observe the intrusion occurring in the local-networks. This paper suggested an improved strategy for Intrusion Detection System (IDS) that combines the expert systems, the processes of data mining as implemented in WEKA. The classification generally consists of the detection principle as well as some of the aspects of WEKA such as open-source data-mining processes. The combining methodology gives better performance of IDS based systems, and helps to maintain the detection more effectively. The result was based on evaluating a new design produced a better form of detection based on efficiency. So, the study presented a good approach to analyse the experiments on behalf of intrusion detection. Deepika P Vinchurkar, et.al [11] conducted a research on Intrusion Detection Systems that consisted of high-level security of networks and thus provides the system dealing with security of network and the intrusion based attacks. The ideal features of IDS includes a monitoring activity of network and the threats. The Intrusion Detection System is generally classified on the basis of the model and the data-source. But some of IDS techniques are more challenging in nature. The anomaly based IDS can be detected easily using various anomaly detection techniques. The process of dimension reduction is based on the analysis of principle component. The problem of construction classifier can be identified using a Support Vector Machine methodology. Nadiammai, et.al [12] focused upon the security issue of the networks and various developments in applications running on distinct platforms capturing an attention towards security of the network. This type of paradigm exploited the vulnerabilities of security that are technically difficult and expensive to solve. Hence intrusion is used as a key to compromise the integrity, availability and confidentiality of a computer resource. The Intrusion Detection System (IDS) plays a vital role in detecting anomalies and attacks in the network. In this work, data mining concept is integrated with an IDS to identify the relevant, hidden data of interest for the user effectively and with less execution time. Four issues such as Classification of Data, High Level of Human Interaction, Lack of Labeled Data, and Effectiveness of Distributed Denial of Service Attack are being solved using the proposed algorithms like EDADT algorithm, Hybrid IDS model, Semi-Supervised Approach and Varying HOPERAA Algorithm respectively. Our proposed algorithm has been tested using KDD Cup dataset. All the proposed algorithm shows better accuracy and reduced false alarm rate when compared with existing algorithms. M. A. Jabbar, et.al [13] proposed the research based on the intrusion detection system to notify and identify the type of activities or normal users or the hackers performing malicious operations. The IDS represents complicated and a linear problem dealing with traffic-data of the network. Many forms of IDS classes have been developed and proposed which further produced distinct levels of accuracy with the aim to maintain a robust and effective Intrusion detection system that is a necessary requirement. In this paper, a model has been designed for intrusion detection system (IDS) using a classifier based on random forest where, the Random Forest (RF) denoted an ensemble classifier and that performed very well as compared to the other classifiers that worked traditionally for an effective classification of different forms of attacks. The experiments were conducted on a data-set named NSL-KDD in order to calculate and analyse the performance of the system and the empirical form of result showed that the proposed model is more efficient for high rate detection and the detection of false alarm. Wang, Huiwen et.al [14] proposed a methodology that focused on the fact that the security of the network has been increased at a very large pace for all the organizations, firms, and the most important is the security of an individual. The use of IDS helps to prevent the data compromised behavior and to follow various forms of machine learning techniques to boost the performance of IDS. To main aim was to obtain high quality improvement in detection for the trained data-set. As the ratio of marginal-density denotes a powerful classifier i.e. univariate in its nature, the study adopted for the obtaining the results is based on framing an IDS based on SVM method entailing its augmented features. Uniquely, a method has been implemented based on logarithmic values of the ratios of the marginal density in order to obtain a good quality of its transformed features which improved the rate of detection based on SVM model. The set of data named NSL-KDD is basically used for the proposed method and the experimental results showed that the results are far much better than the existing forms or methods specifically targeting the rate of accuracy, its training speed, and the false alarm rate.

III. CONCLUSION

The present scenario experiences various forms of developments and a huge growth in advanced processing technologies consisting of connectivity among different networks but methodology is vulnerable by the activities of the intruders or the attackers of the system. These specifically smart attackers interrupt the operation with new and fascinating methods of data-breaching among large networks. Though there are various forms of available intrusion of intrusion detection systems that can detect the intrusions occurring in the network i.e. based on the false positive detection rate and the alert rates but with the detection rate of intrusions, they also have a high false-positive rate resulting in an adequate system comprising of low accuracy level of the system and are generally more prone to different kinds of attack. This usually helps the intruder to enter into the system and perform a pre-planned attack. So, this pre-thesis will propose a hybrid approach to reduce the false positive alarms. The experimental analysis consists of a specified particular form of data-set and the process of feature-based selection will be done to improve the analysis. These features obtained will be used for the classification-tool training and testing the performance of the system. Finally, the result obtained will be compared with the results that already exist.

IV. REFERENCES

[1] Kemmerer, Richard A., and Giovanni Vigna. "Intrusion detection: a brief history and overview." *Computer* 35, no. 4 (2002): supl27-supl30.

[2] Eckmann, Steven T., Giovanni Vigna, and Richard A. Kemmerer. "STATL: An attack language for state-based intrusion detection." *Journal of computer security* 10, no. 1-2 (2002): 71-103.

[3] Hsiu, Pi-Cheng, Chin-Fu Kuo, Tei-Wei Kuo, and Eric YT Juan. "Scenario based threat detection and attack analysis." In *Security Technology, 2005. CCST'05. 39th Annual 2005 International Carnahan Conference on*, pp. 279-282. IEEE, 2005.

[4] Cavusoglu, Huseyin, Birendra Mishra, and Srinivasan Raghunathan. "The value of intrusion detection systems in information technology security architecture." *Information Systems Research* 16, no. 1 (2005): 28-46.

[5] Kim, Jungwon, Peter J. Bentley, Uwe Aickelin, Julie Greensmith, Gianni Tedesco, and Jamie Twycross. "Immune system approaches to intrusion detection–a review." *Natural computing* 6, no. 4 (2007): 413-466.

[6] Panda, Mrutyunjaya, and Manas Ranjan Patra. "Network intrusion detection using naive bayes." *International journal of computer science and network security* 7, no. 12 (2007): 258-263.

[7] Zhang, Jiong, Mohammad Zulkernine, and Anwar Haque. "Random-forests-based network intrusion detection systems." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 38, no. 5 (2008): 649-659.

[8] Aydın, M. Ali, A. Halim Zaim, and K. Gökhan Ceylan. "A hybrid intrusion detection system design for computer network security." *Computers & Electrical Engineering* 35, no. 3 (2009): 517-526.

[9] Wu, Shelly Xiaonan, and Wolfgang Banzhaf. "The use of computational intelligence in intrusion detection systems: A review." *Applied soft computing* 10, no. 1 (2010): 1-35.

[10] Mohammed, Muamer N., and Norrozila Sulaiman. "Intrusion detection system based on SVM for WLAN." *Procedia Technology* 1 (2012): 313-317.

[11] Vinchurkar, Deepika P., and Alpa Reshamwala. "A Review of Intrusion Detection System Using Neural Network and Machine Learning." (2012).

[12] Nadiammai, G. V., and M. Hemalatha. "Effective approach toward Intrusion Detection System using data mining techniques." *Egyptian Informatics Journal* 15, no. 1 (2014): 37-50.

[13] Farnaaz, Nabila, and M. A. Jabbar. "Random forest modeling for network intrusion detection system." *Procedia Computer Science* 89 (2016): 213-217.

[14] Wang, Huiwen, Jie Gu, and Shanshan Wang. "An effective intrusion detection framework based on SVM with feature augmentation." *Knowledge-Based Systems* 136 (2017): 130-139.