REVIEW ON IDS DETECTION AND CLASSIFICATION BY OPTIMIZATION OF FEATURES

Ankita[1] Astha Gautam [2]
[1, 2] Computer Science and engineering
L.R. Institute of engineering and Technology
Solan(H.p.), 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.**

Keywords:* *Intrusion detection systems, denial of service, convolution neural network

I. INTRODUCTION

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 [1] 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 SVM technology deals with high class accuracy in detecting the data intrusions. To avoid redundancy, inadequacy and the noisy data forms there is an urgent need to go for selection i.e. feature based [3]. The basic operation of an intruder to search the faulty operative conditions in the network or the systems. So, an intruder helps to find out the best optimized solutions to identify the intrusions in the data. 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. The main function of IDS is to inspect the various types of attacks done on the system and thereby providing a defence mechanism to fight against these attacks in such a way that it also provides information about the intrusions. So, an IDS provides a mechanism that deals with the safety of current network security system [3]. The following figure.1 explains the general structure of an intrusion dectection system.

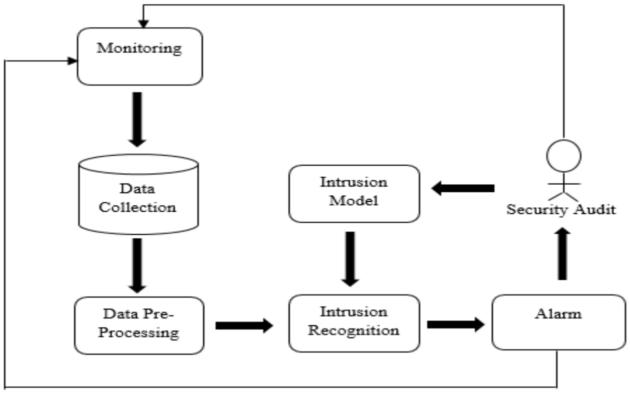

Figure 1: Basic structure of IDS

Examples of Intrusion: compromising remote root, defacement of web server, cracking/ guessing passwords copying viewing databases/ sensitive data, packet sniffer running, pirated software distribution, using an unsafe modem to net access, imitating a consumer for resetting password using an abandoned workstation.

1.2 Types of Intrusion Detection

1. Host-based IDS

- Acquire audit info from sponsor audit paths.
- Detection of attack against an individual host.
- 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.

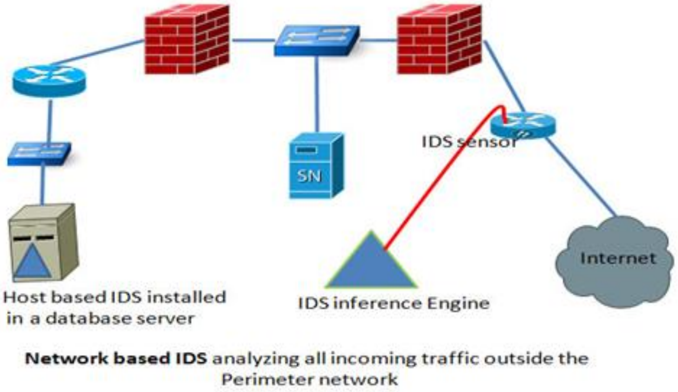
IDS sensor
SN
Internet
Host based IDS installed in a database server
IDS inference Engine
Network based IDS analyzing all incoming traffic outside the Perimeter network

Figure 2: Host-based network

Drawbacks 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.

2. Network-Based IDS

- The traffic network is used as the audit databases, releasing the responsibility on the hosts that always offer common services of computing.
- Detect attacks coming from network.
- 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.

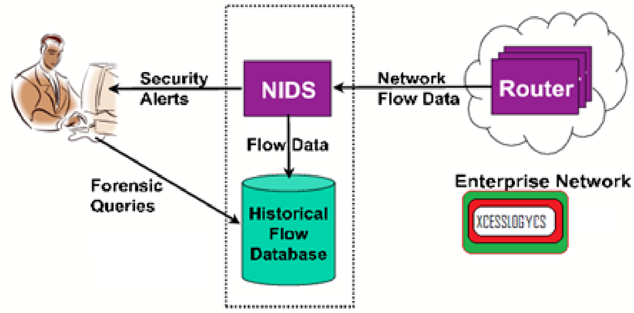

Figure 3: Host-based network

Drawbacks 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.

3. Distributed IDS

- Collect audit data from multiple type of hosts or perhaps the network which helps in connecting the hosts.
- Detection of attacks containing multiple type of hosts.

1.2 Techniques of intrusion detection system

1.2.1 Statistical anomaly detection

- Collect the data related to activity of appropriate users over a given time period.
- Apply arithmetical tests to define the confidence level.

It consists of the following:

1. Threshold detection

- It is crude & ineffective detector
- It involves calculating no. of specific event occurrences over a given time interval.
- Disadvantage: generate lots of false negatives and positives.

2. Profile based

- A profile contain group of parameters or guidelines i.e. audit records which are insight to intrusion detection.
- On the basis of audit records analysis

➔ Gather metrics: guage, counter, utilization of resource, interval time

➔ Analyze: standard and mean deviation, multivariate, period series, functional model, markov process.

1.2.2 Rule-based detection

- A rules-based set was defined that is used to find that the given behavior comprises of an intruder or not.
- In this history of audits records, the records are usually investigated to recognize pattern usage and to make automatically built rules that describes such kind of used patterns
- A set of rules is applied to see whether a given behavior is suspicious.

It falls into two categories:

1. Anomaly Detection: Usage patterns are collected to analyze deviation from the past patterns, with the help of certain rules.

2. Penetration Detection: This is an expert system that looks for illegitimate behavior.

1.3 Analysis Approaches

The two main categories through which network can be analysed for the detection of intrusion are Misuse and Anomaly detection.

1 Misuse detection: It is a perspective where the detection of intrusions on the basis of paradigm matching. Here the abnormal structure behavior is defined at first by collecting the paradigm of attack, and any other behavior is defines as normal behavior by matching them opposing the already recorded attacks. In short, whatever we don't recognize is ordinary. By means of using IDS based attack signatures, represents an instance of this advancement. Signature based systems can only detect and identified, prior been established. It stands opposing deviation apprehension access which employ the reverse perspective, defining the behavior of a normal system and defining other activity as unusual (abnormal). The disadvantage of this perspective is that intrusion detection is accurate only for the known attacks. We needed to update the database of attacks to recognizing the new unseen attacks.

2. Anomaly detection: A great Anomaly-Based Invasion Detection Strategy is a framework for determining computer attacks and wrong use by the activity based on supervisory system and ranked as regular or anomalous. The categorization is founded on guidelines, instead of signatures or paradigm, and can identify any kind of wrong use that comes out of normal functional system.

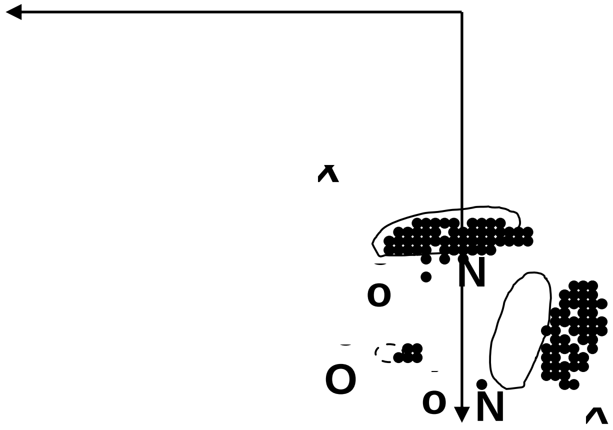

Figure 4: Anomaly based Intrusion detection

In inconsistency recognition, the starting (base) point is usually defined by the machine system, or perhaps normal condition of the systems, load traffic, process, breakdown, and common size of packet. The identifier of the system administers network-based segments to associate state to the standard baseline to check out anomalies

II. DATA MINING TECHNIQUES

The processing of data from the different sources results in gigantic data-sets that cannot be analysed properly [8]. 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 [10].

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 [5]. In, J48-algorithm, set of rules are produced by analyzing decision- based tree.

4. *NB Algorithm:* It uses both the classifier methods i.e. the Naïve Bayes and the decision tree methods. Naïve Bayes is used in leaf nodes and the root-node uses a classifier based on decision tree.

5. *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 [6, 11].

6. *K-Nearest Neighbor:* This represents a simple form of classification technique where it describes the distance among different data points and locates the data points that are not labelled. Ro its nearest neighboring class. It is based on the some of the important conditions i.e. if ‘m’ denotes the value equal to one, then object gets simply assigned to its neighboring value. But if the value of ‘m’ is large then its prediction is very difficult in such case.

7. 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

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

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.

8. *Support Vector Machine (SVM):* It is generally used for the process of classification and prediction. It represents the two main classes of data-points using the method of hyperplane which denotes the +1(normal-data) and -1(suspicious-data) values [17, 29]. The hyperplane condition is stated as below:

$$(W^*\ X) + b = 0 \qquad \text{(ii)}$$

Where,
W (weight vector) = w1, w2…………… wn
X (attribute values) = x1, x2… xn
b = a scalar

Here, the main objective of SVM is to use some part of data to train the system and to identify linear-optimal hyper-plane for maximizing the gap between the margins of separation [12].

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 generally based on the natural evolution theory. The fitness function evaluates the quality of each and every rule [4]. 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 comprises of a hidden, input and the output layer [9] [13]. 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. HMM was also used to detect intrusions using the system calls.

2.6 Hierarchical Clustering

The most commonly used algorithm for hierarchical clustering is the BIRCH hierarchical-clustering which works on some of the data points instead of caring about full form of data-set. Every point of data-sets that are abstracted-points represents centroid of data points clusters. The main advantage of using such kind of clustering is that it is very helpful in dealing with noise based applications. It possesses efficient memory and provides a high standard quality of clustering at a minimal cost.

2.7 K-Mean Clustering

This is most extensive form of clustering algorithm and depicts an easy and simple way to deal with different processes. The first step is the identification of number of clusters ‘k’ that are stated to distribute the samples or instances into a number of clusters that are pre-defined. The first method is to select the 'k' samples denoted as clustering center. Secondly, each and every instance gets assigned to its nearest cluster. The distance of separation between an instance and the center is obtained by using Euclidean distance for the assigning mechanism of the instances based samples.

III. RELATED WORK

Dias GV et.al [1] conducted a study indicated an intrusion detection system based on SVM methodology that combines an algorithm (hierarchical clustering), feature selection method and the technique of SVM. The algorithm i.e. used helps in providing the support vector machine with maintaining an abstracted form of high level of trained examples obtained from the trained set-up of KDD Cup 1999. The study indicates high level performance of SVM based technology which further resulted in a reduced form of training-time. The method of feature-based selection was adopted to remove the un-necessary features of the training set in order to maintain the levels of accuracy. The dataset of KDD cup-1999 was used to analyze the proposed system. When the system was compared with the other forms of data set, the experimental analysis showed that the result based on the performance analysis was not so good as compared to KDD Cup-1999 dataset. So, the methodology based on this dataset showed better analysis in detection of probe and DoS based attacks, maintaining accuracy globally. 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. Chebrolu, Srilatha et.al [3] conducted a research on IDS that examined all the features of data to detect misuse or intrusion patterns. Some of its features may be of redundant nature or donate small quantity to the detection process. The study purpose was to classify unique input features in building an IDS i.e. efficient and effective computationally. An investigated was done based on the performance of algorithms based on feature-selection. The first one was the Bayesian networks (BN) and the other was the CART i.e. classification and regression trees including both the BN and CART in the form of an ensemble. The results showed that input feature-selection was mainly required to design an IDS i.e. light in weight, effective and, efficient for real scenario detection techniques. In the end, the researchers proposed an architecture i.e. hybrid in nature for joining the different algorithms of feature-selection for current scenario of IDS.

Kim, Dong Seong, et.al [4] projected a technique based on Genetic Algorithm to revamp SVM i.e. Support Vector Machines based IDS. The SVM denotes a novel technique of classification that has revealed a high class performance in various applications. The security-based scholars have proposed SVM based IDS. Here, they have used the fusion of SVM and GA to boost the global performance. This type of inter-mixing resulted in a model based on "optimal detection" for SVM classifier where this method not only represented the "optimized-parameters" for SVM but also resulted in an "optimized-feature set" among the data-set. A demonstration was done to check the feasibility of the method by performing experiments on data-set named KDD 1999 for detection of intrusions in the system. Panda, et.al [5] 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 algorithms based on the concept of data-mining classifier such as Naïve Bayes, ID3, and J48 that have been estimated 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 [6] proposed new frameworks that involved the use of a data mining algorithms such as the hybrid-network-based IDSs, an anomaly based detection, and random-forests in misuse. The hybrid mechanism has improved the performance of detection with the combination of misuse advantages. Here, the detection analysis was done on the Knowledge Discovery and Data Mining (KDD'99) data-set. In case of misuse-detection, automatic intrusions based patterns are built using the algorithm based on random-forests over trained data-sets. After this approach, the incursions are usually detected by network-based matching actions in contradiction of the patterns. Whereas in case of anomaly-based detection approach, new forms of intrusions are noticed with the help of outlier detection of the algorithm i.e. random-forests algorithm. In the end the designs/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 well than that of 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. Modi, et.al [7] conducted a survey on different intrusions that affected the integrity of cloud- resources, confidentiality, availability, and the services linked. The proposals of subsuming the IPS i.e. Intrusion Prevention Systems and IDS i.e. Intrusion Detection Systems in cloud technology are examined. The researcher's recommended the positioning of IDS/IPS in Clouded environment to acquire the needed security in the next generation future-based network developments. Muhammad Hilmi Kamarudin, et.al [8] proposed their study on technology of network security that has become a supreme method for the protection of information or the data. With the excessive growth of internet technology, various forms of attack cases are observed in a day to day life. So, to tackle such kind of attacks, a methodology of IDS is adopted and the process of Machine Learning is the most used technology in the IDS. The study based on recent years has shown that the Machine Learning Intrusion Detection system provides a good detection rate and a high accuracy. Thus this paper includes performance analysis based on Machine Learning algorithm known as Decision Tree (J48) where a comparison has been done with two of the other machine learning algorithms named as the NN and the SVM's. These algorithms were tested on the strategy of false alarming rate, rate of detection, and accuracy of four classes of attacks. From the experimental analysis it was detected that the J48 (Decision-tree) algorithm performed well as compared to the other two machine learning algorithms. Deepika P Vinchurkar, et.al [9] directed a research on Intrusion Detection Systems that consists 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 IDS 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

Table.1 Existing Scheduling Model

Author’s Name	Year	Methodology Used	Proposed Work
Nadiammai, *et.al.*	2014	Hybrid IDS model, Semi-Supervised Method, EDADT algorithm, and HOPERAA Varied Algorithm	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.
Chebrolu, Srilatha *et.al.*	2005	Bayesian networks (BN) classification and regression trees (CART)	Conducted a research on IDS that examined all the features of data to detect misuse or intrusion patterns.
.Modi, *et.al.*	2013	IDS/IPS in Clouded environment	Conducted a survey on different intrusions that affected the integrity of cloud- resources, confidentiality, availability, and the services linked.
Aafreen K. *et.al.*	2017	Rule based or signature-based method	Proposed a work using the IDS tool for anomaly detection that provides network security to the system
Wang, Huiwen *et.al.*	2017	Support Vector Machine (SVM)	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.

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 [10] 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 on technical basis was expensive and difficult to resolve. Hence intrusion can be used as a significant factor to compromise the confidentiality, availability, and integrity of a computer-based resources. IDS performs an essential part in discovering attacks and anomalies inside the network. In this ongoing working method, data exploration notion was usually combined with an IDS to recognize the kind of, hidden and relevant interested data for an individual efficiently and with a smaller amount execution time. Four type of problems such as for example: Classification of Data, Conversation based on High Level Human, Insufficient Tagged (labelled) Data, and Efficiency of Distributed DoS (Denial of Service Attack) attack was being resolved using the suggested algorithms just like Hybrid IDS model, Semi-Supervised Method, EDADT algorithm, and HOPERAA Varied Algorithm. Our recommended algorithm continues to be tested applying dataset (KDD Cup). All of the proposed protocol (algorithms) showed improved accuracy and reduced rate of fake alarm in comparison toexisting algorithms. M. A. Jabbar, et.al [11] proposed the research based on the IDS 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 IDS that is a necessary requirement. In this paper, a model has been designed for 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 and efficient 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. Aafreen K. et.al [12] proposed a work using the IDS tool for anomaly detection that provides network security to the system. The IDS represents a method to detect the processes of cyber-attacks and this process of detection is based on the amount of distinct forms of intrusive activities occurring in the operation of the system as the detection of an intrusion denotes a very complicated process. Some of the attacks are known while some of them are not known. The detection process of a known attack is not a difficult task as it can use a rule based or signature-based method but to pin-point an unknown attack is very challenging process. Earlier, the ensemble method was adopted that was a major development in machine-learning process which found a highly accurate form of classifier that was a combination of certain components of the classifier. So, this paper conducted a study based on classifier named cascaded support vector or it might be called as an improved version of ensemble classifier using a function i.e. kernel function. This kernel function represents a Gaussian function. A neural-network technique has been used for collecting its features based on different and distinct forms of attacks and this algorithm is more effective than the earlier method used. Wang, Huiwen et.al [13] 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.

IV. CONCLUSION

Intrusion can be characterized in terms of confidentiality, integrity, and availability. An event or action causes a breach of confidentiality if it allows to access resources, residing in a computer in an unauthorized manner. An event or action causes a breach of integrity if it allows to change the states of resources, residing in a computer in an unauthorized manner. Similarly, an event or action causes a breach of availability if it prohibits legitimate users to access resources or services, residing in a computer. Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusions. An intrusion detection system is a software or hardware that automates the process of monitoring and analyzing of events. The present scenario experiences various forms of developments and huge growth in advanced processing technologies consisting of connectivity among different networks but the 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.

V. REFERENCES

[1] Snapp, Steven R., James Brentano, Gihan V. Dias, Terrance L. Goan, L. Todd Heberlein, Che-Lin Ho, Karl N. Levitt et al. "DIDS (distributed intrusion detection system)-motivation, architecture, and an early prototype." In *Proceedings of the 14th national computer security conference*, vol. 1, pp. 167-176. 1991.

[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] Chebrolu, Srilatha, Ajith Abraham, and Johnson P. Thomas. "Feature deduction and ensemble design of intrusion detection systems." *Computers & security* 24, no. 4 (2005): 295-307.

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

[5] 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.

[6] 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.

[7] Modi, Chirag, Dhiren Patel, Bhavesh Borisaniya, Hiren Patel, Avi Patel, and Muttukrishnan Rajarajan. "A survey of intrusion detection techniques in cloud." *Journal of network and computer applications* 36, no. 1 (2013): 42-57.

[8] Jalil, Kamarularifin Abd, Muhammad Hilmi Kamarudin, and Mohamad Noorman Masrek. "Comparison of machine learning algorithms performance in detecting network intrusion." In *Networking and Information Technology (ICNIT), 2010 International Conference on*, pp. 221-226. IEEE, 2010.

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

[10] 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.

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

[12] Siddiqui, Aafreen K., and Tanveer Farooqui. "Improved Ensemble Technique based on Support Vector Machine and Neural Network for Intrusion Detection System." *INTERNATIONAL JOURNAL ONLINE OF SCIENCE* 3, no. 11 (2017).

[13] 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.