International Journal of Advance Research in Science and Engineering
Volume No.08, Issue No.07, July 2019
www.ijarse.com
IJARSE
ISSN: 2319-8354

INTRUSION DETECTION USING FEATURE SELECTION BY OPTIMIZATION WITH ARTIFICIAL NEURAL NETWORK

Ankita[1], Astha Gautam[2]

[1]*Computer Science and Engineering, L.R. Institute of Engineering and Technology, (India)*
[2] *Computer Science and Engineering, L.R. Institute of Engineering and Technology, (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 ANN_GWO 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 systems, Denial of Service, Artificial Neural Network.

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 [1-3]. 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 [8-10].

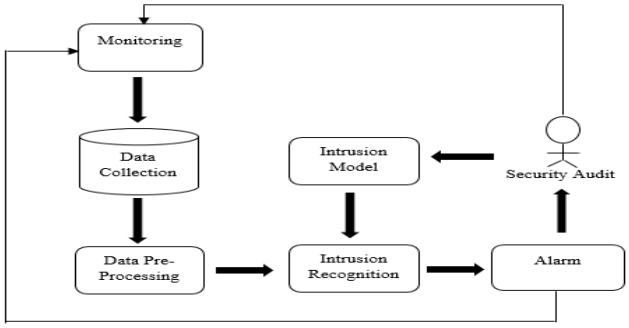

Fig.1 Basic structure of IDS

II. LITERATURE SURVEY

Wolfgang Banzhaf, et.al [1] researched on Intrusion detection based that relies on the computational-based 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 immune systems, ANN evolutionary computation, fuzzy systems, soft computing, and swarm-based 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. Zhou, Chenfeng Vincent, et.al [2] worked on attacks that are coordinated in nature such as DDoS attacks, worm-outbreaks, and large-scale scans, that occur in simultaneous way in case of multiple-networks. Such type of attacks are very hard to identify and with the use of intrusion detection systems (IDSs) i.e. isolated, the researchers has monitored only a limited part of the internet. This paper has summarized the current research in detecting using collaborative forms of intrusion detection systems (CIDSs). Specifically, two major challenges have been discussed. One was the architecture of CIDS and the other was alert correlation algorithms. The conclusion highlights on the opportunities for a collaborative study of intrusion detection systems. Tzong-Wann Kao, et.al [3] suggested a model upon SVM strategy that was predicated on intrusion detection program that joined an algorithm of clustering symbolizing a hierarchical assembly, a method of SVM, and the task of basic feature assortment. The algorithm based on hierarchical-based clustering provided with SVM with fewer, introspective, and highly-qualified training situations that were derived

from the training set KDD Cup 99. It had been in a position to greatly shorten working out time, but also enhance the resultant SVM-based performance. The procedure of feature-based selection process was put on eliminating insignificant features coming from the set of established therefore the acquired SVM unit could sort out the network traffic info more precisely. The popular dataset i.e. KDD Cup 1999 was mainly used to judge the recommended system. Weighed against other attack detection devices that were derived from the same dataset, this technique exhibited improved performance in the discovery of Probe and DoS attacks along with best overall performance in general accuracy. Modi, et.al [4] 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 [5] 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. Elshoush, et.al [6] focused on proper prevention of attacks that were linked to the computer-based systems. As the motive of complete prevention of attacks is not possible so the process of using the IDS plays a crucial role to overcome the harm that is done to the operating systems. Two most important forms of methods based on intrusion detection were used, the first one was misuse-based detection and the second was the anomaly-based detection. A CIDS i.e. collaborative, intelligent intrusion detection system was proposed to examine both the methods, as the individually obtained results from both the methods resulted in less form of accuracy. Specifically there are two major challenges in CIDSs research strategy. Both of them were reviewed and highlighted. The two challenges were the architecture of CIDSs and alert-correlation algorithms. Muamer N., et.al [7] conducted a study on using smart and intelligent form of data-mining methods to observe the incursion occurring in the local-networks. This paper suggested a better-quality strategy for IDS that combines the expert systems, the processes of data mining as employed in WEKA. The classification generally entails detection-based principle along with some of the phases of WEKA such as the processes of open-source data-mining. The joined methodology gives better performance of IDS based systems, and helps to maintain the process of detection more effectively. The experimental result was based on evaluating a novel strategy created 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 [8] 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 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 [9] 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 to existing algorithms.

III.PROPOSED METHODOLOGY

3.1 Proposed Methodology

To obtain the pre-thesis objective a methodology has been proposed which is further divided into three types of phases.

Phase 1: Collection and preprocessing

- Data-set collection
- Extraction of features through a data i.e."tcpdump"
- Converting the obtained features into binary representation
- Preparation of the input for its classification

Phase 2: Classification

- To find the best classifier from the available classifier.

- To test and train the tool of classification by the dataset-partitioning process.

Phase 3: Result analysis

- To compare the obtained results with their existing work.

The proposed working methodology is designed as below in figure.

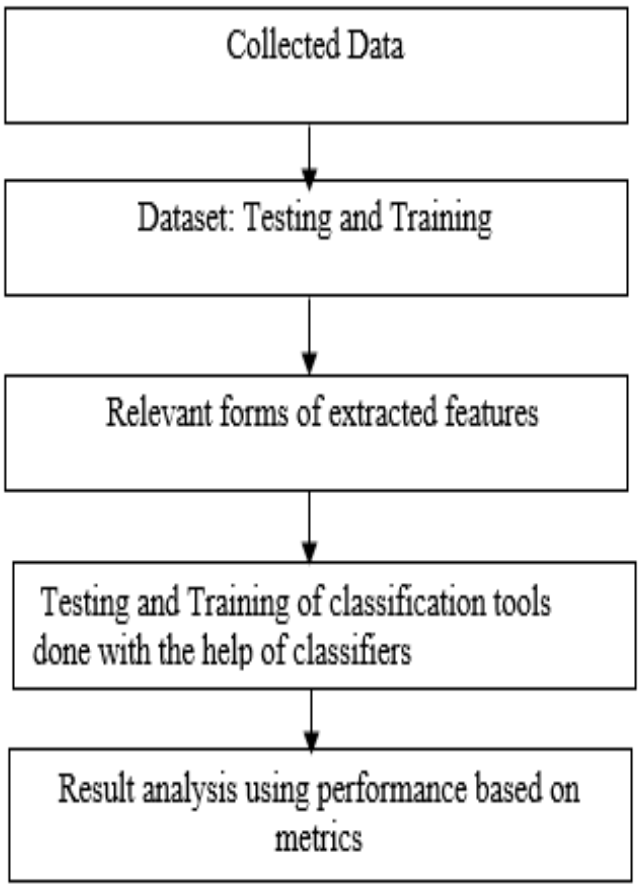

Fig.5 Proposed Methodology

3.2 Proposed methodology: Flowchart

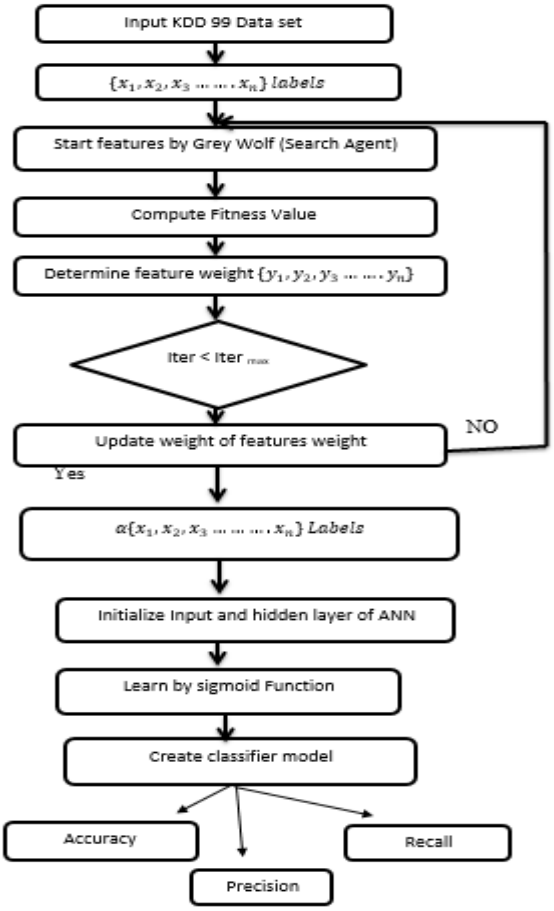

Fig.6 Proposed Flowchart

IV. RESULT ANALYSIS

Since discussions over experiments is usually implemented through the use of KDD-99 which usually having forty one feature units. These kind of features are utilized for learning and marketing and today they will accustomed to evaluate in conditions of assault. In this function we use to judge the rate of accuracy within an IDS. Inside the evaluation put into effect data based on number of attacks. Episodes are usually fall into 4 groups 1) Probe, 2) Dos, 3) U2R 4) R2L. Inside our evaluation all of us uses three classes 1) another attack that contains probe, U2R and R2L 2) DoS-attack 3) Regular attacks (non-attacks). In this function we measure the precision, accuracy, F-measure, and recall in a variety of case.

4.1 Result Analysis

Table.1 KDD CUP 99 dataset based attack type

	Artificial Neural Network	Artificial Neural Network with GA	Artificial Neural Network with PSO	ANN_GWO
Accuracy	89.32	93	90	96.23
Precision	88.45	91	90.38	97.33
Recall	87.12	92	92	98.33
F-measure	85.89	94	87	93.13

Table.2 Data used to study the effectiveness of the four approaches for above discussed attack

Normal	**Dos**	**R2L**	**U2R**	**Probe**
	Smurf	PHF	Root-kit	Portsweep
	Process table	Xlock	Eject	Satan
	Pod	Send-mail	Perl	Saint
	Land	Guess_password	Buffer overflow	M-scan

Table.3 Algorithm Types vs. Types of attack in terms of precision, F-measure, accuracy and recall

Algorithm type	**Types of attack**	**Accuracy**	**Precision**	**Recall**	**F-measure**
ANN	Other attack	87	84	87	83
	Dos Attack	88	87	89	84
	Normal Attack	90	87	86	86
ANN with GA	Other attack	94	88	85	87.23
	Dos Attack	89	91	86	86.23
	Normal	90	90	83	89.13

	Attack				
ANN with PSO	Other attack	92	90	85	84.23
	Dos Attack	95	89	87	83.23
	Normal Attack	91	89	90	87.34
ANN_GWO	Other attack	98.62	96.23	97.33	95.23
	Dos Attack	92.23	90.23	96.33	95.13
	Normal Attack	97.23	96.13	99.56	92.23

1. Simulation is used to study the statistical data in this section. Result of table.1 and 3 is given by simulation process is presented in graphical form.

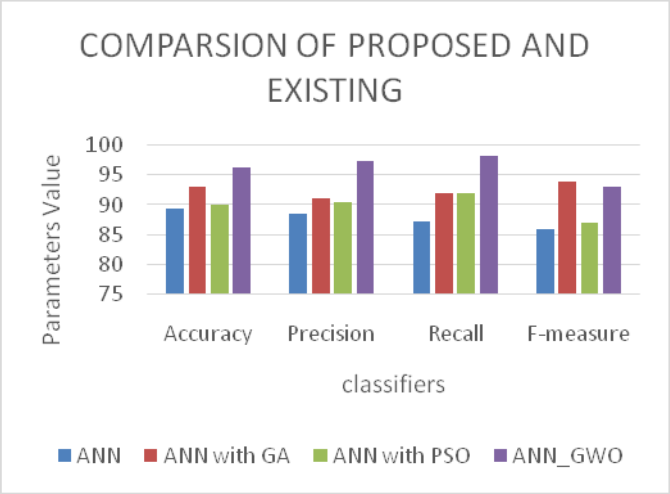

Fig.8 Simulated graph of table.1

Figure 8 displays the simulated evaluation of table.1 particular in conditions of precision, accuracy, f-measure and recall. This figure evaluates the effectiveness i.e. is usually revealed coming from all of the four algorithms which are Artificial Neural Network displayed by green colour , Artificial Neural Network along with PSO presented simply by purple colour, Artificial Neural Network in conjunction with GA presented by red colour and Artificial Neural

Network with both GA and PSO represented by blue colour. Evaluation exhibits that the ANN with GA and PSO provides better effect when it comes to all of the four guidelines (accuracy, accuracy, recall, F-measure).

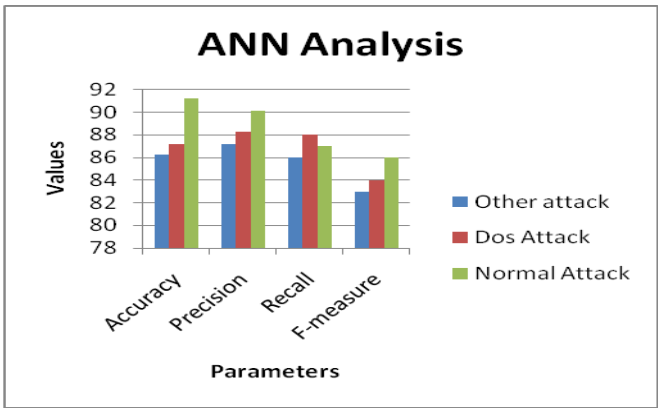

Fig.9 Analysis of ANN

Figure 9 and 10 parameters investigation of various proposed and classifier approach. In case of investigation parameters such as exactness, review, precision and f measure fluctuate as indicated by classifier yet one examination clear about suggested approach (PSO with GA in neural system) demonstrate huge enhance all parameters. In the event that examination just proposed methodology, review demonstrate huge enhancement then different parameters so it will clear sign.

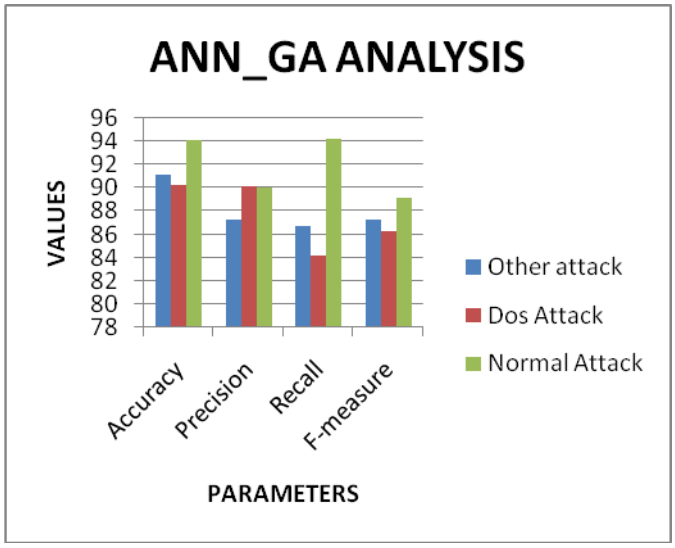

Fig.10 Examination with ANN _GA and ANN

In figure 11 insight investigation of every one of the three classes in Artificial Neural Network and Artificial Neural

Network with GA. In this examination we endeavor to indicate what the criticalness of our methodology is. This exchange we proceed in perception (3) moreover. So first point which examination by typical class 'n' with no assault working and in the two cases Artificial Neural Network and Artificial Neural Network with GA perform well contrast with other parameter like exactness, review and f-measure yet ANN_GA still preferable precision over ANN so include weighted by enhancement by one way or another perform in view of decreasing covering data learning. On the off chance that examination through DOS assault it additionally indicate higher exactness in ANN with GA.so we can close Feature advance weight is better methodology so by what means can enhance improvement these perception talk about in next standard.

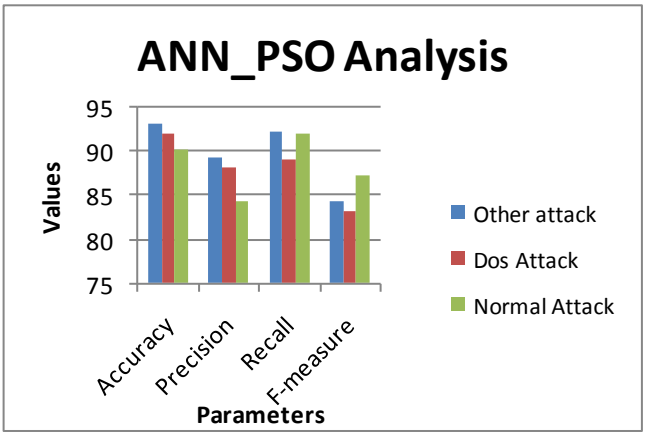

Fig.11 Analysis of ANN _PSO

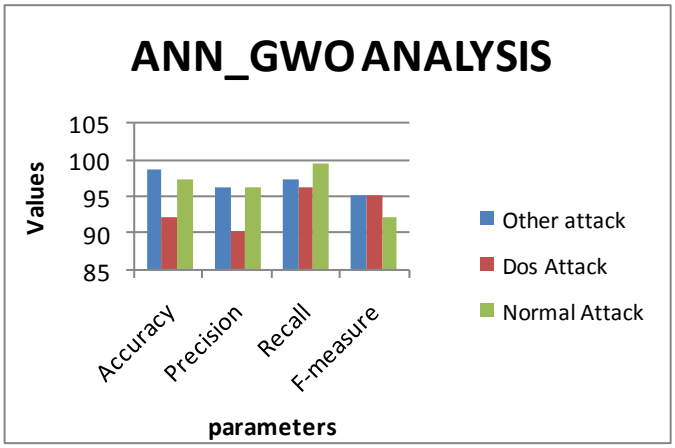

Fig.12 Analysis with ANN _PSO and ANN_GWO

Finally comparing the all four algorithms it can be analysed that algorithm ANN_GWO gives better results among all the algorithms for all the attacks we considered in our work. In figure 12 examination proceed from perception (2) and attempt to discovering centrality of streamlining enhancement impact on various classes' recognition by

characterization. On the off chance that investigation the both chart demonstrate the compelling review however for ordinary class so decrease the false positive rate this enhancement occurring with all classes like DOS assault and different assaults yet the viable outcome appear in other assault which increment fundamentally in proposed approach. So PSO streamlining is great however PSO with GA more enhance in other assault and ordinary class.

V. 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. 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 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.

REFERENCES

[1] 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.

[2] Zhou, Chenfeng Vincent, Christopher Leckie, and Shanika Karunasekera. "A survey of coordinated attacks and collaborative intrusion detection." *Computers & Security* 29, no. 1 (2010): 124-140.

[3] Horng, Shi-Jinn, Ming-Yang Su, Yuan-Hsin Chen, Tzong-Wann Kao, Rong-Jian Chen, Jui-Lin Lai, and Citra Dwi Perkasa. "A novel intrusion detection system based on hierarchical clustering and support vector machines." *Expert systems with Applications* 38, no. 1 (2011): 306-313.

[4] 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.

[5] 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.

[6] Elshoush, Huwaida Tagelsir, and Izzeldin Mohamed Osman. "Alert correlation in collaborative intelligent intrusion detection systems—A survey." *Applied Soft Computing* 11, no. 7 (2011): 4349-4365.

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

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

[9] 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.

[10] Agrawal, Shikha, and Jitendra Agrawal. "Survey on anomaly detection using data mining techniques." *Procedia Computer Science* 60 (2015): 708-713.

[11] J. Jabez, B. Muthukumar, "Intrusion Detection System (IDS): Anomaly Detection Using Outlier Detection Approach", Procedia Computer Science, Volume 48, 2015, Pages 338-346, ISSN 1877-0509, http://dx.doi.org/10.1016/j.procs.2015.04.191.

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

[13] Gupta, Neha, Komal Srivastava, and Ashish Sharma. "Reducing False Positive in Intrusion Detection System: A Survey." *International Journal of Computer Science and Information Technologies* 7, no. 3 (2016): 1600-1603.