# Study on Machine Intelligence Techniques for Intrusion Detection System

Thupakula Bhaskar
Research Scholar
SSSUTMS, M.P.
*shiridisaibaba22@gmail.com*

Dr. Tryambak Hiwarkatr
Research Guide
SSSUTMS, M.P
*tahiwarkar@gmail.com*

Dr. K. Ramanjaneyulu
Research Co-Guide
PVPSIT, A.P
*kongara.raman@gmail.com*

***Abstract*** -This survey paper describes a focused literature survey of machine learning (ML) techniques for intrusion detection.Cyber security is the term which defines the protection of internet connected systems include the hardware, data and software from various attacks. There are number of software solutions developed to improve cyber security over the year and Intrusion detection system (IDS) is one of the major thing in the last decade. Intrusion Detection Systems include approaches that support to detect and identify intrusive and non-intrusive network packets, this model shows the Machine learning concept on IDS. Machine learning techniques have been applied to intrusion detection systems which have an important role in detecting Intrusions. This paper also presents the system design of an Intrusion detection system to reduce false alarm rate and improve accuracy to detect intrusion.

***Index Terms*** – *Machine Learning, Intrusion Detection System,Cyber Security.*

*****

## I. INTRODUCTION

Progressions in computing and network technology have concerned in the pursuit of getting into the Internet as a vital part of our everyday life. Besides, the number of people using Internet seems to be grown rapidly.Intrusion is a termed as a cyber-attack that tries to evade the security appliance of a network or even host system. Attacker could be an unidentified stranger who tries to use the system, or an insider who attempts to gain and mistreat the authorized privileges [1]. Recognition of the Attack, feature selection and executing the layered approach suggestively decrease the time needed to train and test the model. For the network administrators and security specialists, Intrusion detection (ID) is considered as the chief and stimulating problem. The complicated security tools states is about the attackers whom occur with new and a progressive penetration approaches to overcome the security systems which are installed. The intrusion occurs over the server or system for the reason of the present system liabilities, such as user exploitation, misconfiguration of the system, or program flaws. Vast online services and lots of big servers are running in the system in a world-wide network. Simultaneously, such networks turn out to be susceptible to more attackers and thus need an intelligent intrusion detection aspect to protect their network system [2]. The intrusion detection systems actively safeguards the computer system by sensing an attack and probably preventing it. Noticing the hostile attacks is determined by the number and type of suitable actions (Fig.1).

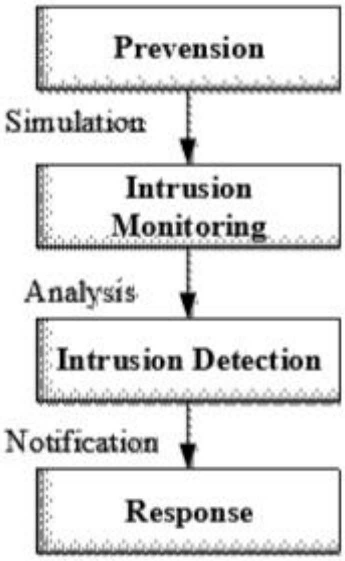

Fig 1: Intrusion detection system activities

IDSs supports in determine, regulate, and classify unauthorized usage, replication, modification, and devastation of information systems [3]. Discussing to the report made by the Internet Security Threat, shortened as the ISTR, about 430 million new malware variants, some of the internet threats such as 362 Crypto-ransom were found in 2015 [4]. IDS which use detailed logical method(s) to identify attacks, find their causes, and make aware the network administrators, have currently been improved to investigate the attempts regarding security violation [5]. There occurs two different types of IDS, based on where the intrusion detection performs: host-based IDS (HIDS) and network-based IDS (NIDS). Depending on how the intrusion detection occurs, an IDS can execute misuse detection (depend on signatures) and/or anomaly detection. IDS mainly used signature detection of the attacks caught in their signature records, they have an increased false alarm rates (FAR).New advanced methods comprising behavior-based modeling have been suggested to detect anomalies consist of data mining, statistical scrutiny, and artificial intelligence systems [6].IDSs can be characterised in numerous ways, however the most usual are misuse-based and anomaly based classes. Recognised attacks, such as Snort are actively identified by the Misuse-based IDS. This type of IDS having a low FAR, but it is unsuccessful to recognize new attacks that does not represent any rules in the database. A model of normal behavior is built by an Anomaly-based IDS and then differentiates any important abnormalities from this model considered as intrusions. New or unidentified attacks can be detected by this type of IDS but reports a high false alarm rate. To decrease the FAR of anomaly-based IDS, numerous machine learning methods, containing support vector machine (SVM) and extreme learning machine (ELM), have been used, along with models relating several systems [7]. SNORT restricts its responsive action to state the attack to an administrator console. Minnesota Intrusion Detection system (MINDS) is an alternative NIDS, which applies Anomaly Detection [8]. In the detection process like Data mining which uses statistical methods and AI techniques. Some Intrusion Detection Systems model the attacks using state transition diagrams [9]. New theories in Intrusion Detection are presented by this project like zero-copy-based packet capture method [10], which to reduce the memory overhead removes numerous copies of the packet.

**The subsequent ways are part of an intelligent intrusion or system attack** [11]:

- ***Information Collection:*** Collecting the target information includes gaining all the facts and information about the user under attack.
- ***Examining and perusing:*** Includes look over of the target host and examining the system's careless or defenseless areas as it examines for the confidential information.
- ***Remote to indigenous admittance:*** Mentions to the procedure of attaining the user system admission by R2L (remote to indigenous) attack types, such as password predicting, buffer run-off attack, and network snuffling. ***User to core access:*** system vulnerabilities are used by a normal system user to access the core of the system in this type of attack.
- ***Launch attacks:*** Example of these attacks are changing web pages, illegal access to confidential data, creating an entrances for upcoming attacks, or retrieving a new person's accounts.

The systems' characteristic problems can be characterised into distinct difficult sets based on competence, accuracy, and usage parameters [12]. In order to increase the detection strength and detection accuracy numerous IDS research studies have been made. In the initial stage, the research work attention lies by means of statistical methods and rule-based skilled systems [13].

Confident machine-learning representations having Linear Genetic Programming , neural networks, Bayesian networks, Support Vector Machines , Fuzzy Logic, Decision trees etc., are investigated for the design of IDS. Thus, one of the best common approaches in machine-learning prototypes is known as Neural Network (NN) that should be used for defining many tough real-world problems which has been excellently used into IDS.

## II. MACHINE LEARNING TECHNIQUES

In order to overcome the aforesaid challenges, many data mining methods were prepared [14]. Several machine learning-based techniques have been applied to IDS. Some of the most important techniques are explained in following subsections. Fig 2 shows the mostly used machine learning techniques used to classify intrusive and non-intrusive behavior.

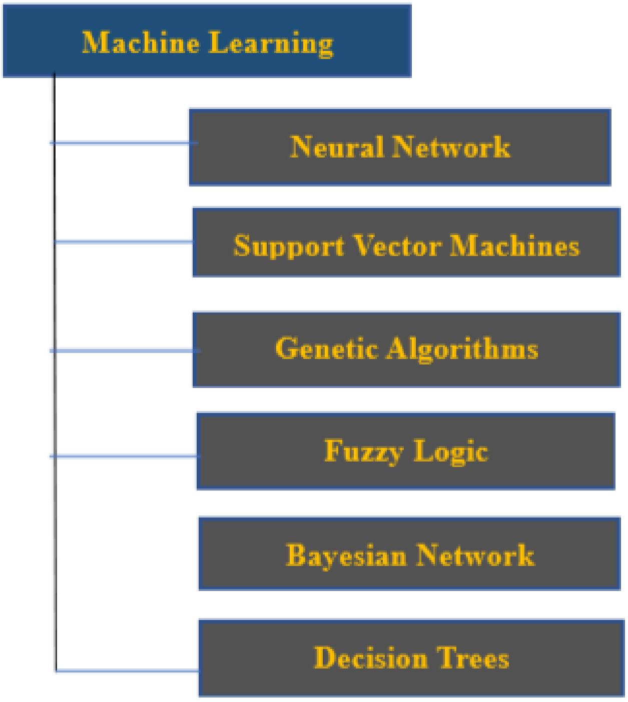

Fig 2: Classification of machine learning techniques

***Neural Networks*** :Artificial neural networks covers of a collection of processing elements that are very organized and convert a set of inputs to a set of desired outputs that is inspired by the way common nervous systems, such as the brain, process information. The technique of Neural Networks NN tracks the similar philosophies of how the human brains works. The Multilayer Perceptrons MLP has been extensively used neural network for intrusion detection. They are capable of approximating to arbitrary exactness, any unceasing function as long as they contain enough hidden units. This means that such representations can method any classification result border in feature space and hence act as non-linear distinguish function. When the NN is used for packets classification, there is one input node for each element of the feature vector. There is usually one output node for each class to which a feature may be assigned. The concealed nodes are connected to input nodes and some early weight assigned to these connection. These weights are familiar during the training process. The error function is defined, based on the Mean Squared Error MSE. The connection weights, between the nodes, are familiar according to the back-propagated error so that the error is reduced and the network learns. The input, output, and hidden layers neurons are variable. Neural network based intrusion detection system, intended to categorize the normal and attack forms and the kind of the attack. It is experimental that the long training time of the neural network was frequently due to the huge number of training vectors of computation facilities. .[14][17]

***Support Vector Machine***: In classification and regression, Support Vector Machines SVM is the utmost common and popular method for machine learning tasks. In this method, a set of training examples is given with each example is marked belonging into one of two categories. Then, by using the Support Vector Machines algorithm, a model that can predict whether a new example falls into one categories or other is built[15]. Support Vector Machine is a classification method possessing better learning capability for small samples, which has been widely applied in many fields such as Network Intrusion Detection, web page identification and face identification. Support Vector Machine applied in intrusion detection possesses such advantages as high training rate and decision rate, insensitiveness to dimension of input data continuous correction of various parameters with increase in training data which endows the system with self-learning ability, and so on. Besides, it is also capable of resolving many practical classification problems, such as problem involving small samples and non-linear problem. So, Support Vector Machine will become increasingly popular in network security.

***Decision Tree:*** Decision Tree algorithm is usually used for classification problem. In this algorithm, the data set is learnt and displayed. Therefore, whenever a new data item is given for classification, it will be classified accordingly learned from the previous dataset. Decision Tree algorithm can also be used for Intrusion Detection. For this reason, the algorithm will also learn and models data based on the training data. As a outcome, the model can categorize which attack categories does a upcoming data fits to base on the model constructed. One of the asset of Decision Tree is it can works well with huge data sets. This is important as large amount of data flow can be initiate across the computer networks. In real-time Intrusion Detection, it works well because Decision Tree gives the highest detection performance and can construct and interpret model easily. Another useful property of Decision Tree in Intrusion Detection model is its generalization accuracy. This is due to the trends in the future where there will always be almost novel attacks, and by having generalized accuracy provided by Decision Tree, these attacks can be detected [16].

***Genetic Algorithm:*** Genetic algorithm GA is a search method that finds an estimated solution to an optimization task - inspired from biological, evolution process and natural genetics and proposed by Holland (1975). GA practices hill climbing technique from an subjective selected number of genes. GA has four operators: initialization, selection, crossover and mutation. There are numerous scholars that used evolutionary procedures and especially GAs in IDS to sense malicious intrusion from normal use. Genetic algorithm based intrusion detection system is used to detect intrusion based on past behavior. A profile is created for the normal conduct based on that genetic algorithm learns and takes the decision for the unseen patterns. Genetic algorithms also used to develop rules for network intrusion detection. A chromosome in a distinct covers genes conforming to attributes such as the service, flags, noted in or not, and super-user efforts. These attacks that are common can be observed more exactly compared to uncommon attributes. Genetic algorithms are capable of deriving classification rules and selecting optimal parameters for detection process. The application of Genetic Algorithm to the network data consist primarily of the following steps:

- The Intrusion Detection System gathers the information about the traffic passing through a particular network.
- The Intrusion Detection System then applies Genetic Algorithms which is trained with the classification rules learned from the information collected from the network analysis done by the Intrusion Detection System.
- The Intrusion Detection System then practices the set of instructions to categorize the inward traffic as anomalous or normal founded on their pattern.

- GA as evolutionary procedures was successfully used in different types of IDS. With GA returned remarkable outcomes, the best suitability value was very closely to the ideal suitability value.

GA is a randomization search technique frequently used for optimization problem. GA was successfully able to generate a model with the desired characteristics of high correct detection rate and low false positive rate for IDS. [20][21] [22] [23].

***Fuzzy Logic***: Fuzzy logic is derivative from fuzzy set theory under which reasoning is approximate rather than exactly derived from classical base logic. Fuzzy methods are thus used in the arena of anomaly detection mainly because the structures to be measured can be seen as fuzzy variables. With fuzzy spaces, fuzzy logic allows an object to belong to dissimilar modules at the same time. This concept is helpful when the difference between classes is not well defined. This is the situation in the intrusion detection task, where the changes between the normal and abnormal modules are not well defined.

The changed fuzzy rules are not complex as no more than five characteristics are used in each rule. It allows description of the normal and abnormal behaviors in a simple way. Fuzzy rules have a clear benefit in real applications. Foremost, they produce rules that are easier to understand, later score high on interpretability. Subsequent, they produce a classifier instruction that is faster in deployment. This is especially crucial for data relating a large number of features. Though uncertain logic has proved to be effective, particularly against port scans and probes, its main difficulty is the high resource consumption and large time spent during the training. [18][19][22].

***Bayesian Networks:*** A Bayesian network is a model that encodes probabilistic relationships among the variables of interest. This method is usually used for intrusion detection in combination with statistical systems. The naïve Bayesian procedure is used for learning task, where a training set with target class is provided. Bayes' Theorem provides a way that we can compute the probability of a premise specified our earlier knowledge. Bayes' Theorem is stated as:

P(h|d) = (P(d|h) * P(h)) / P(d)

Where P(h|d) is the probability of hypothesis h given the data d. This is called the posterior probability.

P(d|h) is the probability of data d given that the hypothesis h was true.

P(h) is the probability of hypothesis h being true (regardless of the data). This is called the prior probability of h.

P(d) is the probability of the data (regardless of the hypothesis).

## III. COMPARISON OF MACHINE LEARNING TECHNIQUES

**TABLE I**

| Machine Learning Technique | Advantages | Disadvantages |
|---|---|---|
| Neural Networks | Competence to simplify from limited, noisy and inadequate data. Does not need skilled knowledge and it can find unknown or new intrusions. | Slow training procedure so not suitable for Realtime detection. Over-fitting may happen during neural network training. |
| Support Vector Machine | Better knowledge capacity for small samples. High training rate and decision rate, insensitiveness to dimension of input data. | Training takes a long time. Mostly used binary classifier which cannot give additional information about detected type of attack |
| Decision Tree | Decision Tree works well with massive data sets. High detection accuracy. | Building a decision tree is computationally intensive task. |
| Genetic Algorithm | Capable of arising best classification rules and Selecting optimal parameters. Biologically inspired and employs evolutionary algorithm. | This method cannot assurance. continual optimization comeback times. Over-fitting. |
| Fuzzy Logic | Reasoning is Estimated rather than precise. Effective, especially against port scans and probes. | High resource consumption Involved. Reduced, relevant rule subset identification and dynamic rule informing at runtime is a difficult task. |
| Bayesian Network | Encodes probabilistic relationships among the variables of interest. Ability to incorporate both Prior information and data. | Harder to handle continuous features. May not cover any good classifiers if prior knowledge is wron |

## IV. PROPOSED SYSTEM ARCHITECTURE

The proposed work is concentrated towards a novel deep learning based cyber security. First the training and the testing samples are extracted separately from the dataset. The dimensionality of the feature space is usually much larger than the size of training set. So, the selection of the most relevant features of the dataset for efficient and effective identification of the cyber-attacks. After extracting the features, the extracted features are clustered using clustering algorithm. Similarly the testing data is also feature extracted and clustered separately. Then a new Deep learning based classifier is proposed to classify the different types of cyber security attacks. The flow of proposed work is as shown in figure 3.

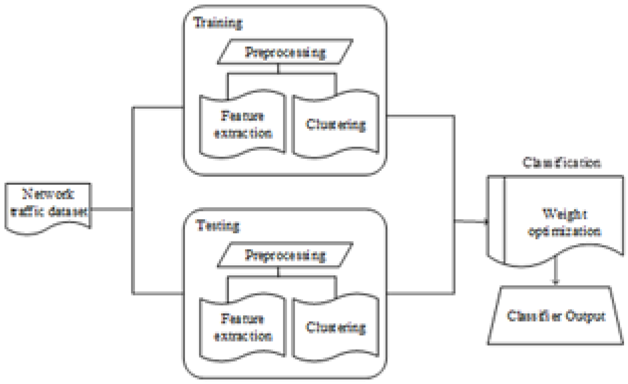

Fig 3: Proposed System Architecture

## V. CONCLUSION

- The proposed model will be developed and executed and the performances are evaluated in terms of training time, detection rate, false alarm rate and efficiency.
- The features are learned from the databases using deep learning algorithm.
- The optimization algorithm speedup the computational process and improve the performance metrics.

## REFERENCES

[1]. Buczak, Anna L., and Erhan Guven. "A survey of data mining and machine learning methods for cyber security intrusion detection." IEEE Communications Surveys & Tutorials 18, no. 2 (2016): 1153-1176.

[2]. Lin, Wei-Chao, Shih-Wen Ke, and Chih-Fong Tsai. "CANN: An intrusion detection system based on combining cluster centers and nearest neighbors." Knowledge-based systems 78 (2015): 13-21.

[3]. Narudin, Fairuz Amalina, Ali Feizollah, Nor Badrul Anuar, and Abdullah Gani. "Evaluation of machine learning classifiers for mobile malware detection." Soft Computing 20, no. 1 (2016): 343-357.

[4]. Kwon, Donghwoon, Hyunjoo Kim, Jinoh Kim, Sang C. Suh, Ikkyun Kim, and Kuinam J. Kim. "A survey of deep learningbased network anomaly detection." Cluster Computing (2017): 1-13.

[5]. Ambusaidi, Mohammed A., Xiangjian He, Priyadarsi Nanda, and Zhiyuan Tan. "Building an intrusion detection system using a filter-based feature selection algorithm." IEEE transactions on computers 65, no. 10 (2016): 2986-2998.

[6]. Al-Yaseen, Wathiq Laftah, Zulaiha Ali Othman, and Mohd Zakree Ahmad Nazri. "Multi-level hybrid support vector machine and extreme learning machine based on modified K means for intrusion detection system." Expert Systems with Applications 67 (2017): 296-303.

[7]. Kevric, Jasmin, Samed Jukic, and Abdulhamit Subasi. "An effective combining classifier approach using tree algorithms for network intrusion detection." Neural Computing and Applications 28, no. 1 (2017): 1051-1058.

[8]. Aslahi-Shahri, B. M., Rasoul Rahmani, M. Chizari, A. Maralani, M. Eslami, M. J. Golkar, and A. Ebrahimi. "A hybrid method consisting of GA and SVM for intrusion detection system." Neural computing and applications 27, no. 6 (2016): 1669-1676.

[9]. Maglaras, Leandros A., and Jianmin Jiang. "Intrusion detection in scada systems using machine learning techniques." In Science and Information Conference (SAI), 2014, pp. 626631. IEEE, 2014.

[10]. Cohen, Aviad, Nir Nissim, Lior Rokach, and Yuval Elovici. "SFEM: Structural feature extraction methodology for the detection of malicious office documents using machine learning methods." Expert Systems with Applications 63 (2016): 324343.

[11]. Aljawarneh, Shadi, Monther Aldwairi, and Muneer Bani Yassein. "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model." Journal of Computational Science 25 (2018): 152-160.

[12]. Aminanto, Muhamad Erza, Rakyong Choi, Harry Chandra Tanuwidjaja, Paul D. Yoo, and Kwangjo Kim. "Deep abstraction and weighted feature selection for Wi-Fi impersonation detection." IEEE Transactions on Information Forensics and Security 13, no. 3 (2018): 621-636.

[13]. Maglaras, Leandros A., Jianmin Jiang, and Tiago J. Cruz. "Combining ensemble methods and social network metrics for improving accuracy of OCSVM on intrusion detection in SCADA systems." Journal of Information Security and Applications 30 (2016): 15-26.

[14]. Hua TANG, Zhuolin CAO "Machine Learning-based Intrusion Detection Algorithms" Binary Information Press, December, 2009.

[15]. J. Burges, "A tutorial on support vector machines for pattern recognition" Data Mining and Knowledge Discovery, vol. 2, pp. 12 1- 167, 1998.

[16]. Kamarularifin Abd Jalil, Muhammad Hilmi Kamarudin, Mohamad Noorman Masrek "Comparison of Machine Learning Algorithms Performance in Detecting Network Intrusion", 2010

[17]. D. Rumelhart, G. Hinton and R Williams, "Learning internal representations by back-propagating errors," Parallel Distributed Processing: Explorations in the

Microstructure of Cognition, D. Rumelhart and 1. McClelland editors, vol. I, pp. 3 18-362, MIT Press, 1986.

[18]. M. S. A. Khan, “Rule based Network Intrusion Detection using Genetic Algorithm,” International J. Computer Applications, vol. 18, no. 8, pp. 26–29, March 2011.

[19]. Rajdeep Borgohain, " FuGeIDS : Fuzzy Genetic paradigms in Intrusion Detection Systems," International Journal of Advanced Networking and Applications, vol. 3, no. 6, pp. 14091415, 2012

[20]. Dewan Md. Farid, Mohammad Zahidur Rahman “Learning Intrusion Detection Based on Adaptive Bayesian Algorithm” 14244-2136-7/2008

[21]. Jonatan Gomez and Dipankar Dasgupta “Evolving Fuzzy Classifiers for Intrusion Detection” Workshop on Information Assurance United States Military Academy, West Point, NY June 2001

[22]. A.A. Ojugo, A.O. Eboka, O.E. Okonta, R.E Yoro, F.O. Aghware “Genetic Algorithm Rule-Based Intrusion Detection System” (GAIDS), ISSN 2079-8407 VOL. 3, NO. 8 Aug, 2012

[23]. J. L. Zhao, J. F. Zhao, and J. J. Li, ―Intrusion Detection Based on Clustering Genetic Algorithm‖, International Conference on Machine Learning and Cybernetics IEEE, Guangzhou, 2005, pp. 3911-3914.

## AUTHOR INFORMATION

**Thupakula Bhaskar**, Research Scholar, Department of Computer Science and Engineering, (Research Scholar, Sri Satya Sai University of Technology and Medical Sciences, India.

**Dr.Trayambak Hiwarkar**, Research Guide, Department of Computer Science and Engineering, Research Guide, Sri Satya Sai University of Technology and Medical Sciences, Sehore,M.P,India.

**Dr.K.Ramanjaneyulu**, Research Co-Guide, Professor,Prasad V. Potluri Siddhartha Institue of technology,Kanuru – 520 007 Vijayawada. ,A.PIndia.