International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488|

||Volume 8, Issue 6, June 2020||

Intrusion Detection System and Vulnerability Identification using Machine Learning Algorithms

Gauri Vilas Rasane[1], Dr. Sunil Rathod[2]

P.G. Student, Department of Computer Engineering, Dr. D. Y. Patil School of Engg, Pune, India[1]

Assistant Professor, Department of Computer Engineering, Dr. D. Y. Patil School of Engg, Pune, India [2]

ABSTRACT: Network security is very essential in today's environment in data security, cloud security as well as all the resources security which is shared in network environment. Basically IDS is the such kind of program which takes unauthorized access of vulnerable resources. It has categorized into Network base IDS and Host base IDS. Intrusions and abuse are constantly threatening to comprehensive internet service use. Therefore, the system for intrusion detection is the most important component of the machine and its network security. Intrusion Detection System (IDS) is an algorithm-focused computer network surveillance system that detects the presence of malevolent interference in the network. The IDS system has been recognized for maintaining high standards of safety, meaning that information is exchanged with confidence and security amongst dissimilar organizations. Systems for intrusion detection divide user activity into two main categories: regular, and distrustful. This paper system proposed an approach with machine learning algorithms for GA-FLN base IDS program. Several intrusion detection opportunities have been suggested before, but none shows acceptable results so systems are investigating for a better outcome in this region. The research suggested even takes a description of different kinds of structure techniques for Intrusion Detection System. System additionally research in these extraordinary methodologies, their exactness and also false positive proportions..

KEYWORDS: Intrusion Detection system, Soft computing , classification techniques.

I.INTRODUCTION

The security of computer networks has been in the focus of research for years. The organization has come to realize that information & network security technology has become very important in protecting its information. The massive rise in the use of computer systems in the general public today, and in particular the sudden increase in e-business hugeness towards the wealthy community, has rendered computer asylum a global goal. Since creating a system without any defects is not actually feasible, an important field of study has been noted for interference. The three most commonly applied innovations in defense are safety at night, network-based security and host-based support. Host-based security is difficult to implement, given the reality that the majority utilize it often. The hard part is to force it on every single entity. While this is quite convenient in a small network of the same machines as the network expands and becomes heterogeneous, from an administrative point of view it becomes a big headache. Any successful attempt or unsuccessful attempt to compromise the integrity, confidentiality, and availability of any information resource or the information itself is considered a security attack or an intrusion. Every day new kind of attacks is being faced by industries. One of the solutions to this problem is by using Intrusion Detection System (IDS). The wide use of computer networks , and the increase in web based business has made security of the host and network an important issue as these are vulnerable to attacks. These attacks can be passive that just reads confidential data or it can be active attack that also modifies or fabricates the data [1]. Since it is not possible to avoid these vulnerabilities and design a completely secure system . Intrusion detection has become a major challenge. The primary objective of Intrusion detection system is to identify the attack and in some cases analyze it. Various techniques and approaches have been developed . But with the evolution of new attacks more robust systems need to be designed.

II.LITERATURE SURVEY

ParisaAlaeiet. Al.(2018) [1] , in this paper ,the approach implies a technique to solve malicious attack detection problem by reviewing data sets online. This is done through the use of an incrementally naïve Bayesian classifier. In comparison, active learning allows the problem to be answered through the use of a small set of defined data points which are often quite expensive to acquire. The proposed technique comprises of two acting classes, i.e., offline and

online. The former includes results pre-processing, while the latter uses the NADAL electronics. The proposed solution is similar to the formal, naïve classification using Bayesian the NSL-KDD standard dataset.

Alsughayyir, Bayan et al.(2019)[2], according to this paper, a deep learning (DL) is used which can create a better and more efficient architecture for intrusion detection (ID). The purpose approach focuses on classifying normal behavior from anomalous network activity. The Intrusion Detection System (IDS) is one of the tools used to detect unauthorized network or system operation and protect the system from attacks on the network. Attacks are observed in the device by discriminating between common and irregular network behavior and functionality. This work also defines numerous methods to produce IDS which has they utilized in experiment analysis.

Borkar, Amol et Al.(2017),According to [3] Consists of the literature review of the Inner Intrusion Detection System (IIDS) and the Intrusion Detection System (IDS), which uses various data collection techniques and testing approaches for the system to function in real time. Data mining techniques are being developed for cyber analytics to allow intrusion prevention. Techniques used before like firewall, and IDS failed to detect the real-time attackers that happened in the absence of the manager, without his knowledge. Recognizing the attacker in real time is challenging, because it can produce duplicate IP and attack packets. A computer network is a Software and Hardware mix. Each part carries risks, Poor security, and deficiencies. The assault against the Ransomware leaves data vulnerable. Those who learn programming and programs can find out about the various activities conducted on the systems quickly from the log files. We'll help to ensure security.

Bhosale, Karuna S. et al.(2018),According to [4] deep neural network (DNN), It is studied as a type of deep learning framework to build scalable and powerful IDSs to detect and recognize accidental and unexpected cyber attacks. The continuous change in network activity and the rapid development of attacks make it necessary to examine multiple databases that are created over the years via static and dynamic approaches. Such type of study makes it easier to identify the right algorithm which can work effectively to predict potential cyber-attacks. On numerous publicly available test malware databases, a detailed assessment of DNN studies and other classical machine learning classifiers is shown. The Optimum network parameters and topologies for DNNs are chosen using hyper parameter filtering methods KDDCup 99 dataset.

Chamou et al. (2019) [5], A large number of businesses around the world are being targeted and threatened by the daily emergence of new and evolving threats. It is for this reason that the scientific community has drawn attention to the nature and improvement of the performance of Intrusion Detection Systems. It is an innovative method to track malicious activity using deep learning techniques in terms of DDoS and ransomware cyber-threats. Cyber security accomplishment, Data protection and secure communication are considered essential owing to the rapid growth and use by most Internet users of Web apps. At the same period, there was increased exposure to more sophisticated cyber threats over the Internet and computer networks, in the digital world of academia and industry, especially in small and medium-sized enterprises (SMEs), with financial costs.

Christos, et al.(2019), according to [6] Self-using Novel Network Intrusion Prevention Program Organizing with Support an enhanced neural Vector system network. The proposed system, because of its design, does not provide a security solution that is neither signed nor dependent on rules, and is capable of mitigating known and unknown risks with high accuracy. Based on our experimental results The suggested design can use the NSL KDD dataset to access specialized online education, so it fits for efficient and scalable industrial applications. Machine-based NIDS / NIPS programming partially Signature-and address the above bugs Suffering from Rule-based Industrial NIDS / NIPS. Not to be lost .

Liang, Wei, et al. (2019), according to [7] for multifunctional performance, effectively increase The pace of identification and the actual output of detection of abnormal behavior in industrial networks. The novel apps are dual to easily pick a node with a high security coefficient as the cluster's center and coordinate the multi-function data in a cluster around the middle. Experimental results show that the algorithm suggested is of high quality in terms of detection rate and time relative to other algorithms. In the networking sector, the sensitivity of identification of suspicious data exceeds 97.8%, and the incorrect detection result falls by 8.8%. Intrusion detection systems can successfully identify and monitor intruder incidents, although difficult with network security technology. So, the use of intrusion detection systems in industrial networks will overcome the limitations of conventional network security techniques, so perfecting the entire industrial safety system networks.

Loganathan, Gobinath et al.(2018), according to [8] Present a new multi-attribute method to estimate a network packet sequence based on past packets using the Sequence-to-Sequence (Seq2Seq) encoder-decoder algorithm. This The model is used to learn the standard sequence of packets in TCP communications in an attack-free dataset, and is then used to classify anomalous packets in TCP traffic. We demonstrate that the experimental multi-attribute model Seq2Seq identifies anomalous raw TCP packets in the DARPA 1999 dataset which are part of 97 percent intrusions

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488|

||Volume 8, Issue 6, June 2020||

through precision. It can also detect selected intrusions with 100% real-time precision and surpass existing algorithms based on replicated neural network models, like LSTM. The Detecting Irregularities in raw TCP packets via a Seq2Seq algorithm designed specifically for sequences with different attributes. Packets in connections apart from regular network traffic are used to train the model system. Anomalies are known to science as actual packets which deviate significantly from the packets planned. Training the model on normal traffic rather than intrusion traffic gives access to extensive training data and enables the model to detect even new unknown threats which deviate from regular traffic pattern.

Mayank Agarwal et al.(2017),According to [9], This paper describes a system for intrusion detection of PS-Poll DOS infiltration in 802.11 networks, using a distinct case structure in real time. This methodology utilizes RTDES to track DOS attack on a single Event System in real-time. High detection rate and accuracy rate are one of the significant advantages, but shortage of frames is one of the major drawbacks. This system also able to detect software as well as hardware attacks simultaneously.

Saeid Soheily Khah et al. (2018)[10] this System, network intrusion detection(ID) is tackled by unattended and unattended hybrid mining-a comprehensive case study on the ISCX dataset. This proposes a hybrid intrusion detection (k M-RF) which generally outperforms an alternative technique in terms of false alarm rate, precision, and detection time. ISCX(A standard intrusion detection dataset) is used to assess the efficacy of kM-RF, and an in-depth analysis is performed to check the effects of any pre-processing characteristics or characteristics detected. It also uses a special pre-treatment method for categorical transformation Tools or attributes for numerical data and to create more segregated groups from raw data. Some new features or applications to find payloads, clustered attacks and IP scans and a mix of k-means and random forest classifiers to prevent further interference effectively.

III.PROPOSED SYSTEM

The goal of proposed anomaly network intrusion detection system is to maximize the detection accuracy, to minimize false positive rate and detector generation time. Basically there are two phase in the proposed system, system have taken NSLKDD dataset for system training as well testing purpose.

The proposed System worked with an ensemble configuration. When two or more combinations form a new model commonly called an ensemble model. This ensemble model incorporates input from multiple classifiers and has produced a single composite classification. Our conceptual structure consists of numbers for the classifiers. First, the software receives data from various outlets, both online and offline. Once the data is collected by software, other data mining strategies will be applied in different classifications approaches.

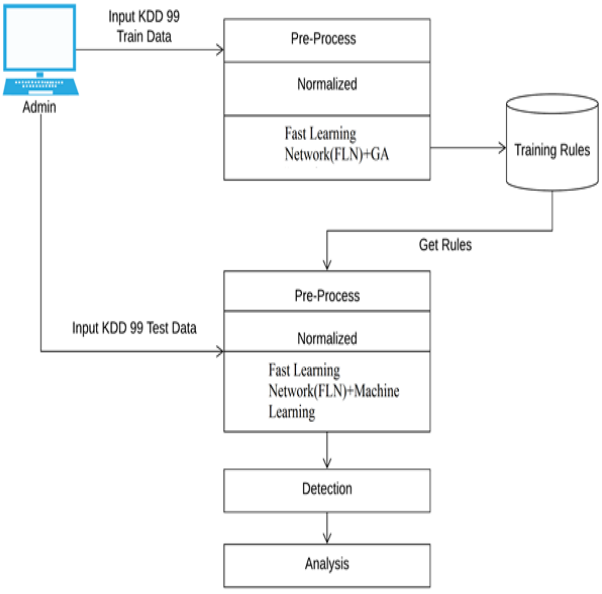

Figure 1: Proposed System architecture

System Modules

Training Phase:

1. Upload training data for feature extraction.
2. Apply PSO for rule creation
3. Create rules set as normal pool as well as intrusion pools set.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488|

||Volume 8, Issue 6, June 2020||

Testing Phase:
1. Upload Testing data or any packet which is collected from network environment.
2. Extract all features using attribute selection.
3. Apply Normalization approach on dataset.
4. Apply ensemble approach on all train as test features.
5. Show results with classification accuracy.
6. Classify all attacks.
7. Show detection results.

IV.SYSTEM DESIGN

This Phase, various algorithm is used where, we first initialize the chromosomes and group of chromosomes we say as population is created. Once the population is created crossover is applied to obtain new generation of chromosomes. Mutation is applied for updating bit value of attributes of chromosomes randomly. The fitness function will define the fitness value of each chromosome and a selection criterion is applied for selected optimal rules. When variation is completed then Genetic algorithm will get terminated. The outputs of genetic algorithm are genetic rules. The output of genetic algorithm that is genetic rules is given as an input to fuzzy logic. In this phase probability of each attribute is calculated which is used for classification of data as attack or normal

Step 1: System first collect network traffic from network audit data using packetX Lib and Wincap driver or some synthetic dataset like KDDCup99, NSLKDD, ISCX and WSNTrace etc.
Step 2: Select features of each connection and apply Genetic Algorithm (GA) for rule creation.
Step 3: Once rule created store it into local database directory called as BK rules.

Testing Phase:
In this Phase, Fuzzy rules are given as an input to the Neural Network algorithm for the classification of sub attack. Here system collect the network traffic data using PacketXLib and Wincap Driver. On each instance neural network algorithm will be applied. Transfer function will be used for calculating each node weight .Using Defined threshold , sub attacks can be classified.

Step 1: System collect the network traffic data using PacketX Lib and Wincap driver or NSLKDD
Step 2: Read each instance and apply ensemble (J48, ANN, NB) algorithm.
Step 3: Calculate the weight using given functions for each connection.
Step 4: Finally classify the each attack with sub attack type using define threshold (e.g. DoS, PROBE, U2R, R2L, Network attacks, Active Attack, Passive Attack, Advance attack etc)

V.DATASET DESCRIPTION

The inherent flaw in the KDD cup 99 dataset [9] has been publicized by several statistical analyses has affected the detection accuracy of many IDS modeled by researchers. NSL-KDD data set [3] is a refined version of its precursor. It contains vital records of the complete KDD data set. There are a group of downloadable files at the disposal for the researchers. They are listed in the table 1.

TABLE 1 : LIST OF NSL-KDD DATASET FILES AND THEIR DESCRIPTION

S. No.	Name of the file	Description
1	KDDTrain+.ARFF	The full NSL-KDD train set with binary labels in ARFF format
2	KDDTrain+.TXT	The full NSL-KDD train set including attack-type labels and difficulty level in CSV format
3	KDDTrain+_20Perce nt.ARFF	A 20% subset of the KDDTrain+.arff file
4	KDDTrain+_20Perce nt.TXT	A 20% subset of the KDDTrain+.txt file
5	KDDTest+.ARFF	The full NSL-KDD test set with binary labels in ARFF format
6	KDDTest+.TXT	The full NSL-KDD test set including attack-type labels and difficulty level in CSV format
7	KDDTest-21.ARFF	A subset of the KDDTest+.arff file which does not include records with difficulty level of 21 out of 21
8	KDDTest-21.TXT	A subset of the KDDTest+.txt file which does not include records with difficulty level of 21 out of 21

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488|

||Volume 8, Issue 6, June 2020||

In each record there are 41 attributes unfolding different features of the flow and a label assigned to each either as an attack type or as normal. The details of the attributes namely the attribute name, their description and sample data are listed in the tables. Table 2 contains type information of all the 41 attributes available in the NSL-KDD data set. The 42nd attribute contains data about the various 5 classes of network connection vectors and they are categorized as one normal class and four attack class. The 4 attack classes are further grouped as DoS, Probe, R2L and U2R. The description of the attack classes.

Table 2 : MAPPING OF ATTACK CLASS WITH ATTACK TYPE

Attack Class	Attack Type
DoS	Back, Land, Neptune, Pod, Smurf,Teardrop,Apache2, Udpstorm, Processtable, Worm (10)
Probe	Satan, Ipsweep, Nmap, Portsweep, Mscan, Saint (6)
R2L	Guess_Password, Ftp_write, Imap, Phf, Multihop, Warezmaster, Warezclient, Spy, Xlock, Xsnoop, Snmpguess, Snmpgetattack, Httptunnel, Sendmail, Named (16)
U2R	Buffer_overflow, Loadmodule, Rootkit, Perl, Sqlattack, Xterm, Ps (7)

VI.RESULTS AND DISCUSSION

The proposed research basically focuses on the soft computing approach and classification-based detection, basically both methods having the good detection rate but generating sometimes more false positive ratio. In real-time environments, some systems are also not applicable, and some may not focus on misclassified anomalies. As noted, the mark is still missing in most applications because there is no program that currently provides a discovery rate of 100% and the sky is the limit.

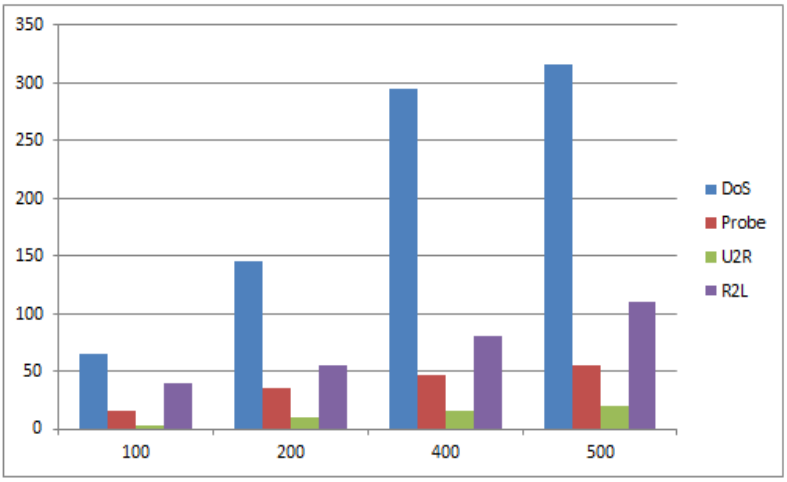

Figure 2 : overall attack found ratio in ensemble approach

We have two tests. In the first investigation, we utilized our fluffy hereditary calculation to group typical system information and assault. At that point, we indicate identification rate acquired for KDD99 dataset. I characterize them into two classes which are ordinary and assault. In the second analysis, we utilized the fluffy hereditary calculation to arrange sorts of assaults in the online continuous sniffer dataset.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488|

||Volume 8, Issue 6, June 2020||

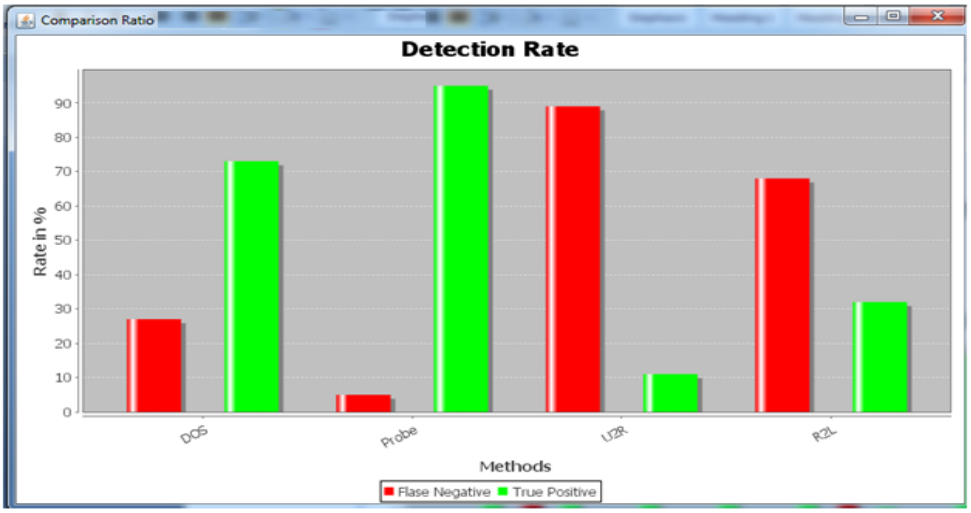

Figure 3 : Existing System Performance

Table 3: Proposed System Overall Performance

Detection Rate	DOS	Probe	U2R	R2L
Existing	86%	82%	76%	72%
Proposed	96%	89%	79%	81%

VII.CONCLUSION AND FUTURE WORK

System suggested ensemble method for network traffic anomaly detection in this research work. Our approach focused on building the model of anomaly detection normal traffic profile. System also showed through experiments that some features of the NSL-KDD and ISCX dataset with the normal profile are efficient. With input training data, system proposes a multiple machine learning algorithm to reduce noise. The experiments showed that our Approach works good, even with a small training sample, including precise recognition. They often call for a new architecture to merge anomaly detection system with signature-dependent detection system, along with some improvements to the usual performance profile of buildings. In our future plan, system will use an open source IDS to build and play with the proposed model in actual network. Several different ideas have emerged to confront this problem since about ten years ago the intrusion detection concept began to gain momentum in the security community. Detection systems for intrusion vary in the approaches used to collect the data and in the specific techniques used to analyze these data. To evaluate the system with some combination of network and synthetic dataset and generate the dynamic rules for strongly unknown attack detection in vulnerable environment.

REFERENCES

[1] Alaei P, Noorbehbahani F. Incremental anomaly-based intrusion detection system using limited labeled data. In Web Research (ICWR), 2017 3th International Conference on 2017 Apr 19 (pp. 178-184). IEEE.
[2] Alsughayyir, Bayan, Ali Mustafa Qamar, and Rehanullah Khan. "Developing a Network Attack Detection System Using Deep Learning." 2019 International Conference on Computer and Information Sciences (ICCIS). IEEE, 2019.
[3] Borkar, Amol, AkshayDonode, and Anjali Kumari. "A survey on Intrusion Detection System (IDS) and Internal Intrusion Detection and protection system (IIDPS)." 2017 International Conference on Inventive Computing and Informatics (ICICI). IEEE, 2017.
[4] Bhosale, Karuna S., Maria Nenova, and GeorgiIliev. "Modified Naive Bayes Intrusion Detection System (MNBIDS)." 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS). IEEE, 2018.
[5] Chamou, Dimitra, et al. "Intrusion Detection System Based on Network Traffic Using Deep Neural Networks." 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD). IEEE, 2019.
[6] Constantinides, Christos, et al. "A novel online incremental learning intrusion prevention system." 2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS). IEEE, 2019.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | Impact Factor: 7.488|

||Volume 8, Issue 6, June 2020||

[7] Liang, Wei, et al. "An Industrial Network Intrusion Detection Algorithm based on Multi-Feature Data Clustering Optimization Model." IEEE Transactions on Industrial Informatics (2019).
[8] Loganathan, Gobinath, JagathSamarabandu, and Xianbin Wang. "Sequence to sequence pattern learning algorithm for real-time anomaly detection in network traffic." 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE). IEEE, 2018.
[9] Mayank Agarwal, SankethPurwar, Santosh Biswas, Sukumar Nandi, "Internal Detection System for PS-Poll DOS attack in 802.11 networks using real-time discrete event system",IEEE,vol.4,issue4,2017.
[10]Sedjelmaci H, Senouci SM, Ansari N. A hierarchical detection and response system to enhance security against lethal cyber-attacks in UAV networks. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2018 Sep;48(9):1594-606.