Soft Computing and Classification Approach to Anomaly Based Intrusion Detection System: An Overview

Ms. Sayali R. Kshirsagar
Department of Computer Engineering
JSPM's
Rajarshi Shahu College of Engineering Tathawade, Pune, India.
sayalikshirsagar111@gmail.com

Dr. P. B. Kumbharkar
Department of Computer Engineering
JSPM's
Rajarshi Shahu College of Engineering Tathawade, Pune, India.
pbk.rscoe@gmail.com

ABSTRACT

System security is of essential part now days for huge organizations. The Intrusion Detection frameworks (IDS) are getting to be irreplaceable for successful assurance against assaults that are continually changing in size and complexity. In IDS complex information is being put away and handled in arranged frameworks. With extensive use of internet service, there is constant risk of interventions and harm. Thus Intrusion Detection System (IDS) is fundamental component of computer network security which is software based monitoring mechanism for detecting presence of malicious activity in the network. IDS has collected attention by keeping high security levels, providing loyal and safe announcement of the information between dissimilar organizations. IDS arranges computer behavior into two main categories: normal and anomaly activities. Many forms for intrusion detection have been proposed before but none shows honest results in this field. So we are inspecting for better outcome in this field. Our system takes various kinds of layout policies for IDS using machine learning approach. We are focusing here on intrusion detection rate and classification accuracy of the system.

Keywords— ***Classification Techniques Intrusion Detection System, Machine Learning, Soft Computing.***

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. 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 are being faced by industries. One of the solution to this problem is the use of 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 they 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 [7]. 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 for detection of new attacks more robust systems need to be designed.

Basically Intrusion Detection System (IDS) ordered into two distinctive types Host Base Intrusion Detection System (HIDS) and Network base Intrusion Detection System (NIDS). Today's system security foundation promisingly relies on Network intrusion detection Framework (NIDS) [8, 9, and 10]. NIDS gives security from known interruption assaults. It is unrealistic to stop interruption assaults, so associations should be prepared to handle them. IDS is a cautious component whose main role is to keep work continuing considering every conceivable assault on a framework. Interruption recognition is a procedure used to distinguish suspicious movement both at system and host level. Abnormality identification and abuse location are the two principles of intrusion detection techniques. The oddity identification model depicts the typical conduct of a client to recognize this current client's irregular or ignorant activity [13].

The wide use of computer system in today's general public, especially the sudden surge in hugeness of e-business to the world has made PC system safety as a global priority. Since it is not practicable to fabricate a plan without any vulnerabilities. For the most part intruder is characterized as a framework, project or person who tries to and may get to be unbeaten to break into a data framework or execute an activity not formally permitted. We imply interruption as any arrangement of procedures that attempt to trade off the honesty, privacy, or availability of a network asset. The demonstration of identifying procedures that attempt to trade off the honesty, attentiveness, or accessibility of a network asset can be implied as interruption discovery. An interruption location framework is a gadget or programming application that screens system and/or framework, exercises for irritated approach of violation and produces data to an administration position. Interruption identification is the procedure of observing the activities happening in a network framework. Fundamentally when an intruder try to break into a data framework or perform an activity not authoritatively permitted, we imply to this activity as an interruption. Interruption system may incorporate abusing programming bugs and plan misconfigurations, secret word incensed, sniffing unsecured exchange, or misusing the outline defect of express conventions. Intrusion Detection Systems are an important tool in overall implementation of an organization's

information security policy, which reflects an organization's statement by defining the rules and practices to provide security, to handle intrusions and to recover from damage caused by security violations. An Interruption Location Framework [12] is a plan for distinguishing interruptions and reporting them definitely to the best possible authority.

II. LITERATURE SURVEY

Naseer, Sheraz, et al. [1] proposed the system to investigate the suitability of Deep Learning approaches for anomaly-based intrusion detection system. For this study, they developed anomaly detection models based on different deep neural network structures including Convolutional Neural Networks, Auto encoders and Recurrent Neural Networks. These deep learning models were trained on NSLKDD training dataset and evaluated on both test datasets provided by NSLKDD namely NSLKDDTest+ and NSLKDDTest21. All experiments in this study are performed by authors on a GPU-based test bed. Conventional Machine Learning based intrusion detection models were implemented using well-known classification techniques including Extreme Learning Machine, Nearest Neighbor, Decision-Tree, Random-Forest, Support Vector Machine, Naive-Bays, and Quadratic Discriminant Analysis. Both deep and conventional Machine Learning models were evaluated using well-known classification metrics including, Receiver Operating Characteristics, Area under Curve, Precision-Recall Curve, mean average precision and accuracy of classification. Experimental results of deep IDS models showed promising results for real-world application in anomaly detection systems.

Samrin, Rafath, D. Vasumathi et al. [2] presented the investigation of different techniques and intrusion classification on KDD Cup 99 dataset. So, by classifying the different network issues a new and effective technique is implemented which can categorize and identify intrusions in the KDD Cup 99 dataset. The IDS is majorly classified into two techniques namely misuse and anomaly detection. In misuse detection numerous existing techniques are used. The methods of anomaly detection include predictive pattern generation, neural network, sequence matching, statistics and supervising. These two techniques detect the network based issues. The presented information constitutes an important point to start for addressing Research & Development in the field of IDS. Based on this survey various limitations are addressed such as high false alarm rate, difficult to apply in massive datasets, high network traffic, time complexity in training and testing process etc.

Mehmood, Tahir and Helmi B. Md Rais [3] Compared different supervised algorithms for the anomaly-based detection technique. The algorithms have been applied on the KDD99 dataset, which is the benchmark dataset used for anomaly-based detection technique. The result shows that not a single algorithm has a high detection rate for each class of KDD99 dataset. The performance measures used in this comparison are true positive rate, false positive rate, and precision. This gives the comparative study of machine learning algorithms SVM, naive Bayes, J.48 and decision table for anomaly detection. The performance of the algorithms is tested on KDD99. For each class, each algorithm has different result and no single algorithm has high true positive rate for all 5 different classes. But the overall accuracy of J.48 decision tree is high among all other algorithms and it has low misclassification rate.

Levent Koc and Alan D. Carswell proposed work given in [4] utilizes data mining procedures in intrusion discovery frameworks for the grouping of the system occasions as either typical or assault. Naive Bayes (NB) strategy is a straightforward, effective and well known information mining technique that is based on probable freedom of characteristics likelyhood. Hidden Naive Bayes (HNB) is an expanded type of NB that keeps the NB's effortlessness and proficiency while unwinding its freedom presumption. Framework's exploratory research guarantees that the HNB paired classifier model can be connected to interruption discovery issue. Test comes about utilizing great KDD 1999 Cup interruption discovery dataset shows that HNB double classifier has better execution regarding location precision contrasted with the conventional NB classifier. Framework disclosed the expanding need to apply information mining strategies to arrange assault occasions. A straightforward and broadly utilized information mining strategy is checked which is Naive Bayes (NB) classifier. A paired classifier display in view of the Hidden Naive Bayes (HNB) technique as an augmentation to NB to decrease its naivety supposition. Framework's test ponder results demonstrate that the HNB twofold characterization display expanded with EMD discretization and CONS highlight determination channel strategies has better general outcomes in wording of detection accuracy, error rate and area under ROC curve than the traditional NB model.

Eman Abd EI Raoof Abas [5] used artificial immune system network based intrusion detection. In framework's structure authors proposed utilizing KDD Cup dataset for intrusion identification and applied R-piece calculation of counterfeit invulnerable framework system which is utilized for anomaly discovery. Around the productive execution of interruption location authors accomplished the exactness and less tedious attack discovery activities. Authors made similarity between interruption identification frameworks and fake invulnerable framework which assist us to accomplish framework's objective. AIS gives speculation, multilevel guard and collaboration between cells. RST solves

the issue of the unpredictability of KDD cup informational collection by diminishing 41 elements to six components. Enhanced RST likewise utilized for the execution of IDS by adding distinctive weights to the estimations of the six elements .The rate of true positive (TP) and true negative (TN) is relatively high.

Naila Belhadj Aissa, Mohamed Guerroumi [6] proposed A Genetic Clustering Technique for Anomaly-Based Intrusion Detection Systems (GC-AD). GC-AD uses a dissimilarity measure to form k clusters and genetic process is applied on these clusters. Authors proposed two stage fitness function calculation as 1) they introduced a confidence interval to refine the clusters in order to obtain partitions that are more homogeneous and then 2) they computed and maximized the inter-cluster variance over the generations. The accuracy of this technique is tested on different subset from KDD99 dataset.

Geethapriya Thamilarasu [7] proposed Genetic Algorithm based Intrusion Detection System for Wireless Body Area Networks (WBAN). The objective is to design and develop an intrusion detection framework to improve security in WBAN. For this author proposed a multi-objective genetic algorithm based intrusion detection system to provide optimal attack detection in these networks. Authors used only those features necessary for detecting a specific attack in the intrusion detection process, thereby decreasing the computational complexity.

Fatemeh Barani [8] proposed an approach based on genetic algorithm (GA) and artificial immune system (AIS), called GAAIS, for dynamic intrusion detection in AODV-based Mobile ad hoc networks (MANET). The performance of GAAIS is evaluated by authors for detecting several types of routing attacks simulated using the NS2 simulator. These attacks includes Flooding, Blackhole, Neighbor, Rushing, and Wormhole. Experimental results show that the proposed approach using GAAIS is more efficient in comparison with similar approaches.

Alka Chaudhary, et al. [9] proposed a novel intrusion detection system based on first order of sugeno-type fuzzy inference system for ad hoc flooding attack. In the proposed intrusion detection system authors used fuzzy logic because fuzzy logic can able to handle the uncertainty and give the decisions within the range 0 to 1. The simulation results of proposed intrusion detection system show the good performance in terms of high true positive and low false positive rates. This system is developed for mobile ad hoc networks.

Salah Eddine Benaicha, Lalia Saoudi [10] presented a Genetic Algorithm (GA) approach with an improved initial population and selection operator. This new GA approach efficiently detect various types of network intrusions. GA is used to optimize the search of attack scenarios in audit files because of its good balance exploration / exploitation. It gives the subset of potential attacks which are present in the audit file in a acceptable processing time. In testing phase they have used NSL-KDD99 dataset. By combining the IDS with Genetic algorithm authors found that there is increase in the performance of the detection rate of the Network Intrusion Detection Model and reduction in the false positive rate.

III. PROPOSED SYSTEM

The proposed figure 1 shows the overall execution of proposed system. System first takes input from different external sources like KDD CUP, NSL KDD, ISCX and real time network packets. The whole system consist three different modules these are below.

Firstly apply the discretization and normalization methods on KDD dataset before attribute selection. Then apply ensemble approach with multiple soft computing classifiers and finally test the data whether it is normal or anomaly.

System Modules

Basically there are two phase in the proposed system, we have taken KDD dataset for system training as well testing purpose.

Module 1: Intrusion Detection System (IDS)

This phase executes the first Genetic Algorithm and Fuzzy Algorithm for features extraction and for creating background rules. Once background rules have created, testing phase has done with the help of Decision Tree (DT) for master and sub attack classification.

Module 2: Intrusion Prevention System (IPS)

The prevention system work for prevent the known attacks which is already generated by remote sources. The system having the feature which automatically block when any attack has generated. Here some pattern classification algorithms work for finding the same network flow as well as packet signatures.

Module 3: Intrusion Response System (IRS)

Basically it work for provide the security from different type of unknown attacks. The system holds ensemble modules for detecting malicious activity. Different classification approaches work for find such kind of unknown attacks. This module also provides the forensic features for creating the attacks log reports.

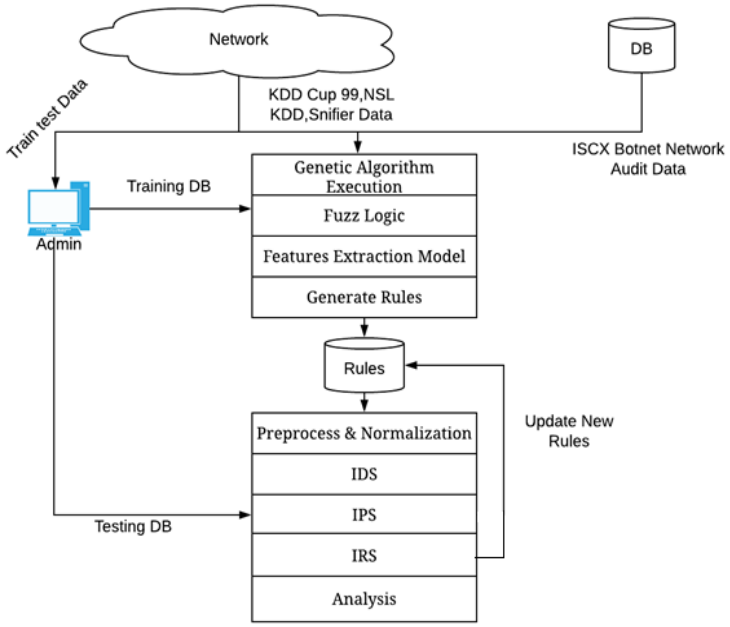

Figure 1: Proposed System architecture

Methodology

In our system there are basically two phases.

1. Training Phase:

In this Phase, Genetic 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.

2. 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 andWincap 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.

IV. Dataset Details

The inherent drawbacks in the KDD cup 99 dataset [14] has been revealed by various statistical analyses has affected the detection accuracy of many IDS modeled by researchers. It contains essential records of the complete KDD data set. There are a collection of downloadable files at the disposal for the researchers. They are listed in the table 1.

TABLE 1: LIST OF KDD DATASET FILES AND THEIR DESCRIPTION

S. No.	Name of the file	Description
1	KDDTrain+.ARFF	The full 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

V. CONCLUSION

Since the study of intrusion detection began to gain momentum in the security community roughly ten years ago, a number of diverse ideas have emerged for solving these security problems. Intrusion detection systems vary in the sources they use to obtain data and in the specific techniques they employ to analyze data. Most systems today classify data either by misuse detection or anomaly detection. Each approach has its relative merits and is accompanied by a set of limitations. It is likely not realistic to expect that an intrusion detection system be capable of correctly classifying every event that occurs on a given system. Perfect detection, like perfect security, is simply not an attainable goal given the complexity and rapid evolution of modern systems.

After the completion of this survey we can conclude there are different techniques that can used for detection, some soft computing as well as some classification approaches are effective for detect the different attacks. Some system has work on signature base anomaly detection with creation of different rules. KDD cup dataset has used for training and testing purposed. Finally every system shows the maximum accuracy for attack detection, but none of these are has focused on unknown attack detection or misuse detection.

VI. FUTURE WORK

We propose to embed the multi-classification approach (ensemble) with different algorithms on NIDS as well as HIDS as future enhancement. The second challenging task for future enhancement is to detect and prevent the attacks in both online and offline environment with forensic approach.

REFERENCES

[1] Nasser, Sheraz, et al. "Enhanced Network Anomaly Detection Based on Deep Neural Networks." IEEE Access 6 (2018): 48231-48246.
[2] Samrin, Rafath, and D. Vasumathi. "Review on anomaly based network intrusion detection system." *Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), 2017 International Conference on*. IEEE, 2017.
[3] Mehmood, Tahir, and Helmi B. Md Rais. "Machine learning algorithms in context of intrusion detection." *Computer and Information Sciences (ICCOINS), 2016 3rd International Conference on*. IEEE, 2016.
[4] Levent Koc and Alan D. Carswell, Network Intrusion Detection Using a HNB Binary Classifier in IEEE 2015
[5] Eman Abd EI Raoof Abas Artificial immune system based intrusion detection: anomaly detection and feature selection IEEE 2015.
[6] Naila Belhadj Aissa, Mohamed Guerroumi, A Genetic Clustering Technique for Anomaly-Based Intrusion Detection Systems IEEE 2015.
[7] Geethapriya Thamilarasu, Genetic Algorithm based Intrusion Detection System for Wireless Body Area Networks, 3rd IEEE International Workshop on Security and Forensics in Communication Systems 2015.
[8] Fatemeh Barani , A Hybrid Approach for Dynamic Intrusion Detection in Ad Hoc Networks Using Genetic Algorithm and Artificial Immune System in IEEE 2014.
[9] Alka Chaudhary, Vivekananda Tiwari, Anil Kumar, A Novel Intrusion Detection System for Ad Hoc Flooding Attack(Using Fuzzy Logic in Mobile AdHoc Networks IEEE 2014
[10] Salah Eddine Benaicha, Lalia Saoudi, Salah Eddine Bouhouita Guermeche, Ouarda Lounis , Intrusion Detection System Using Genetic Algorithm Science and Information Conference 2014
[11] Mohammed A. Ambusaidi et al. Building an intrusion detection system using a filter-based feature selection algorithm IEEE TRANSACTIONS ON COMPUTERS, VOL., NO NOVEMBER 2014.
[12] Hachmi Fatma and Limam Mohamed, A two-stage technique to improve intrusion detection systems based on data mining algorithms IEEE 2013.
[13] Saeed Khazaee and Maryam Sharifi Rad, Using fuzzy c-means algorithm for improving intrusion detection performance IEEE 2013
[14] Wagh SK, Pachghare VK, Kolhe SR. Survey on intrusion detection system using machine learning techniques. International Journal of Computer Applications. 2013 Jan 1;78(16).
[15] Shadi Aljawarneha,, Monther Aldwairi, Muneer Bani Yassein , Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model