Hybrid Approach towards IDS IPS IRS using Reinforcement Learning

Kirti Choudhary, Komal Dhing, Shivani Pacharne, Sweety Samanta
UG Students, Department of IT AISSMS IOIT Kneddy Road, Near RTO,411001
Prof Anuja phapale
Prof. Department of IT AISSMS IOIT Kneddy Road, Near RTO,411001

Abstract- System security is an 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 intricacy. With information honesty, privacy and accessibility, they must be solid, simple to oversee and with low upkeep cost. Different adjustments are being connected to IDS consistently to recognize new assaults and handle them. This paper proposes a semi-supervised model based on combination of ensemble classification for network traffic anomaly detection. As most IDS try to perform their task in real time but their performance hinders as they undergo different level of analysis or their reaction to limit the damage of some intrusions by terminating the network connection, a real time is not always achieved. In this research, we are going to implement intrusion detection system (IDS) using anomaly intrusion detection method for misuse as well anomaly detection. The proposed framework is used a classifier, whose information base is demonstrated as a administer, for example, "if-then" and enhanced by a hereditary calculation. The system is tried on the benchmark KDD'99, NSL KDD and ISCX intrusion dataset and contrasted and other existing methods accessible in the writing. The outcomes are empowering and show the advantages of the proposed approach. GA based rule creation system according to their feature selection method that worked. For the Classification purpose we use fuzzy logic it generates rules. Patternphase match each new observation with established normal profile to detect anomaly.

Keywords- Network traffic anomaly, anomaly detection, semi-supervised model, intrusion detection, network security, DARPA data set.

I. INTRODUCTION

Implementation of IDS for distributed architecture using online Adaboost based approach combined with weak classifiers [11] viz. decision stump and GMM (Gaussian Mixture Model) overcome the difficulty of handling multi attribute network connection data along with maintaining highest detection rate and accuracy. They proposed a distributed intrusion detection framework, in which a local parameterized detection model was constructed in each node using the online Adaboost algorithm. A global detection model was constructed in each node by combining the local parametric models using a small number of samples in the node. This combination is achieved using an algorithm based on Particle Swarm Optimization (PSO) and support vector machines. The global model in each node is used to detect intrusions. Experimental results show that the improved online Adaboost process with GMMs obtains a higher detection rate and a lower false alarm rate than the traditional online Adaboost process that uses decision stumps. It is also shown that PSO and SVM-based algorithm effectively combines the local detection models into the global model in each node; the global model in a node can handle the intrusion types that are found in other nodes, without sharing the samples of these intrusion types. Luigi Coppolino et. al [12] designed a hybrid, lightweight, distributed Intrusion Detection System (IDS) for wireless sensor networks. Implemented IDS uses both misuse-based and anomaly-based detection techniques. It is composed of a Central Agent, which performs highly accurate intrusion detection by using data mining techniques, and a number of Local Agents running lighter anomaly-based detection techniques on the notes. Decision trees have been adopted as classification algorithm in the detection process of the Central Agent and their behavior has been analyzed in selected attacks scenarios. The accuracy of the proposed IDS has been measured using CART, CHAID, C5.0 and Bayesian networks and achieves maximum accuracy of detection rate. System [13] developed layered approach to solve the delity problem using machine learning approach known as genetic algorithm using KDD dataset. Principal Component analysis is used for feature reduction. Layered approach gives better detection rate along with reduced computational and overall time required for detection of anomalous events.

A machine learning approach [14] known as Genetic algorithm, to identify such attack type of connections. Intrusion detection system used information in the form of audit trails or packet of the network. In layered model three layers used to identify DOS, probe, R2L and U2R attacks. Each layer is separately trained with a small set of relevant

features and then deployed sequentially. Feature reduction is achieved by Principal Component Analysis. Layered model is used for decrease computation and the overall time required detecting anomalous events. It is a powerful tool for analyzing data and found similar patterns in the data. S.Vijayarani, M.Divya analyzed the performance of the three classification rule algorithms, viz. PART, C4.5, RIPPER and algorithms [15] from the experimental results it is concluded that in the case of time factor number of rules generation, Part algorithm seems better than the other two algorithms for breast Cancer and heart disease Dataset. The performance of three well known data mining classifier algorithms viz., ID3, J48 and Naive Bayes [16] were evaluated based on the 10-fold cross validation test. Using the KDDCup'99 IDS data set experimental results express that Naive Bayes is one of the most efficient inductive learning algorithm and decision trees are more remarkable as far as the detection of new attacks is concerned.

II. LITERATURE SURVEY

According to Saeid Soheily Khah [1] proposed a system Intrusion detection in network systems through hybrid supervised and unsupervised mining process - a detailed case study on the ISCX benchmark dataset. This work proposes a K-Means base Randon Forest (kM-RF) which outperforms in overall the alternative methods through the accuracy, detection rate and false alarm rate. A benchmark intrusion detection dataset ISCX dataset has used to evaluate the efficiency of the kM-RF, and a deep analysis is conducted to study the impact of the importance of each feature defined in the pre-processing step. It also focuses on

- A dedicated pre-processing procedure to convert the categorical features to numerical ones and to build more isolated classes from the raw data,
- Some new features to consider payloads, IP scans and distributed attacks and
- A combination of k-means and random forest classifier

to detect intrusion more effectively. The efficiency of the suggested hybrid approach (kMRF) is analyzed on a dynamic, scalable and labeled benchmark dataset called as ISCX, which is the most up to date dataset compared to the other commonly explored ones for data intrusion benchmarking. The results show the benefits of the kM-RF, which outperforms the other state of the art methods through the high accuracy, high detection rate and low false alarm rate, overall. A Wilcoxon signed rank test is used to determine that the proposed kM-RF detection approach is significantly better than the other methods.

According to Parisa Alaei et. Al. [2] in this paper, a method is proposed to overcome this problem by performing online classification on datasets. In doing so, an incremental naive Bayesian classifier is employed. Furthermore, active learning enables solving the problem using a small set of labeled data points which are often very expensive to acquire. The proposed method includes two groups of actions i.e. offline and online. The former involves data preprocessing while the latter introduces the NADAL online method. The proposed method is compared to the incremental naive Bayesian classifier using the NSL-KDD standard dataset. There are three advantages with the proposed method: (1) overcoming the streaming data challenge; (2) reducing the high cost associated with instance labeling; and (3) improved accuracy and Kappa compared to the incremental naive Bayesian approach. Thus, the method is well-suited to IDS applications.

The fuzzy logic- based system can detect a malicious or intrusion behavior of a particular network, since it is rule based and it contains an improved set of rules. They have used automated approach for creation of fuzzy rules, which are obtained from the specific rules using recurrent items. The evaluations and experiment of the proposed intrusion detection (ID) system are performed with the well known KDD Cup 99 dataset. The proposed system achieved superior precision in identifying normal and intrusive records is shown clearly in the experimental result. The training kddcup dataset contains normal records as well as four types of master attacks. For creating rules, system uses 10 different features. In the testing phase, the testing dataset is given to the proposed system for classification of network normal or intrusive behavior. The final rules are then used for detecting accuracy of the system based on recall, definition, precision, F-measures for estimating rare class prediction. Given system only working on Training and testing dataset it can't work on real time benchmark dataset. Given system shows the very good detection rate for all attacks it will not work for new signatures or attack [3].

The use of intrusion detection systems in soft computing techniques like neural networks and neuro fuzzy networks is used to categorize network activities and specify what category of attack got generated. For the initial classification of the initial network traffic, neuro fuzzy classifiers are used. A system Fuzzy inference is later used to determine whether the activity is normal or abnormal. An IDS system is responsible for reducing false alarm rates. Human knowledge is used to create their fuzzy rule by Fuzzy inference systems. For classifying network traffic, the use of Genetic Algorithm is used in conjunction with ANFIS for obtaining best optimal solution. Genetic algorithms use a set of genetic parameters such as initialization of population, crossover, mutation rate, fitness and selection on current population to reproduce new optimal solution. There is a Poor detection rate for probe,

U2R and R2L.System can create only static rules; it cannot work on dynamically new generated rules [4].

The fuzzy logic- based system is used for detecting an intrusion or malicious behavior of a particular network. Automated strategy is used for generating fuzzy rules. The evaluations of the proposed intrusion detection system are performed for detecting intrusion with the use of KDD Cup 99 dataset. The proposed system achieved higher precision in identifying whether the records are normal or intrusive. The first component of the proposed system is to categorize input data or information into several classes depending upon different attacks involved in the intrusion detection. Second the designed strategy for automatic creation of fuzzy algorithm rules to give efficient learning. In general, the fuzzy rules created by fuzzy logic are given by the fuzzy system by analyzing intrusion behavior. The process of fuzzy generation is given in the following sub-section.

- Mining of only length recurrent items.
- Rule generation
- Rule filtering
- Generating fuzzy rules

The old fitness function for Genetic Algorithm is used. The apriori algorithm is used for association rules for increasing the system time complexity and execution time [5].

Intrusion Detection (ID) with Genetic Algorithm (GA) and fuzzy Logic describes two different behavior of training intrusion detection (ID) system to recognize possible attacks in a particular network or computer system: genetic algorithms and fuzzy logic. It will describe an approach by using fuzzy genetic algorithms and compare those records with rules obtained using a decision tree. Genetic algorithm uses genetic parameters like crossover, mutation, selection for obtaining optimal solution. The GA rules which are generated by Genetic algorithm are given as an input to Fuzzy logic for classification of master a

Section 1: described the measures that are used in determining the accuracy rate of IDS.

Section 2: described the use of a fuzzy genetic algorithm in IDS.

Section 3: described the results of using a traditional genetic algorithm.

The detection rate for a proposed system is approximate 98.00%.There are numerous restrictions to the prevention-based approach for network and computer security. It is almost certainly not possible to construct a totally secure system. The prevention-based security viewpoint constrains the user's productivity and activity [6].

Intrusion Detection (ID) System using fuzzy genetic algorithm (FGA) is used to classify network attack. The proposed approach evaluates intrusion detection (ID) system in terms of false positive alarm, detection rate and detection speed. Fuzzy genetic algorithm (FGA) can classify two activities i.e. malicious and normal behavior. For each generation, population size of 10 is considered. Mutation rate of 30% and single point crossover is applied. Online network dataset can be detected by fuzzy genetic algorithm, within 2 to 3 seconds. Preprocessing requires 2 seconds and fraction of second is required for detection. With low false positive rate and high accuracy, fuzzy genetic algorithm can detect recent network activities using online dataset and KDDcup dataset. The rate of detection is over 97.5% .System cannot detect unknown attack or the attacks whose signature is not predefined. System cannot generate dynamic rules [7].

According to [8], a system to accurately detect potential attack has developed by using various techniques like decision free, Random forest and KNN. To overcome the limitation of the previous system that was not able to detect the IPV6 attacks, a new method are proposed. The developed system produce the impressive and efficient result in identifying IPV4-based attack keeping in mind the future scope. The effectiveness of various algorithm evaluated. Detection accuracy, precision, recall percentage were measured.

According to [9] Has stated that clustering and KDD can be efficiently used to detect novel anomaly called NEC. An unsupervised anomaly is used to produce high detection rate and less false passive rate. It is an appropriate way to solve the problem and find the anomaly which does not need a labelled data set. The system is verified over NSL-KDD 2009 dataset. The preprocessing model transforms all features into the real number and normalised dataset at the end the evaluation component will compare predicate result an accurate result.

Concerning A Survey of Data Mining and Machine Learning for CSID [10], a survey of data mining and machine learning for cybersecurity instruction detection is performed to ensure cyber security. Packet-header and net flow packet header are used for the instruction detection system to be able to reach networks and kernel level data. The future scope that is kept in mind is that data mining and machine learning cannot ware without representative data and also it's very time-consuming. The complexity of different machine learning and data mining algorithm is discussed, The paper also provides a set of comparison criteria for machine learning/data mining methods Intrusion Detection System help discovered, determine and identify unauthorized used, duplication, alteration and destruction of the information system.

III. RESEARCH METHODOLOGY

The proposed system worked with ensemble approach, system first collect data from different online as well as offline sources. Once data has collected by system it will

apply some data mining strategies with different classification approaches. The attribute selection done with the help of weka 3.7 tool. The below figure show the proposed system architecture and execution. The proposed research work our aim to generate strong rules and increase the detection rate for DOS, PROBE, U2R and R2L for NIDS and HIDS

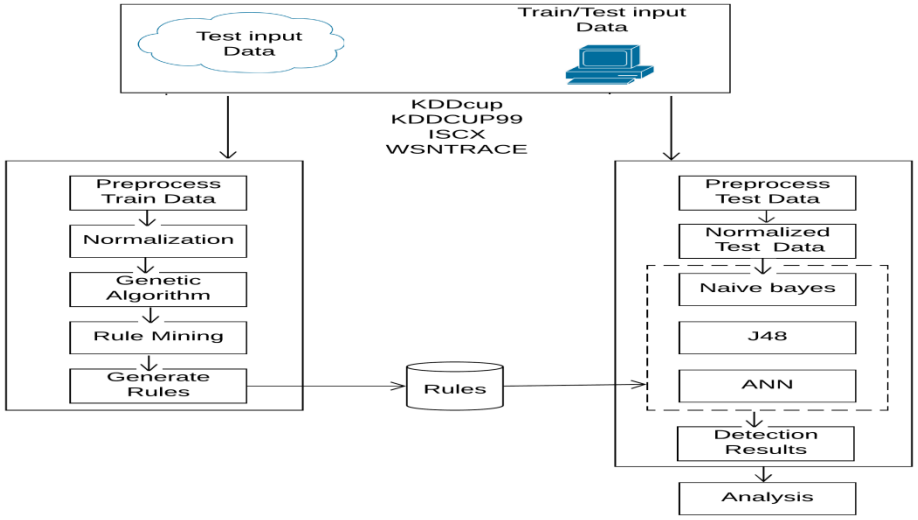

Fig.1: System Architecture

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, we have taken NSLKDD dataset for system training as well testing purpose.

System Modules

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

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.

Step 4: System collect the network traffic data using PacketX Lib and Wincap driver or NSLKDD

Step 5: Read each instance and apply ensemble (J48, ANN, NB) algorithm.

Step 6: Calculate the weight using given functions for each connection.

Step 7: 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)

IV. DATASET DESCRIPTION

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.

- NSL KDD
- KDD CUP 99 from DARPA organization.
- ISCX dataset
- Network real traffic dataset. (using packetX and winpcap)
- WSNTrace Dataset (WSN data generated using NS2 simulation environment)

V. SYSTEM ANALYSIS

Algorithm 1 : Rule policies generation algorithm

Input: Training set from network log or data packets, Attribute validation policies Background Knowledge (BK) policies, Threshold ‘th’.

Output: Rule set as policies or signatures.

Step 1: a. Read values from data file header fields to get Feature Vector.

b. Read data from a network connection to append in Feature []

Step 2: Validate each attribute for the preprocessing phase

Step 3: Normalized irrelevant attribute, and get normalized set NormSet[] ← {Att[i……n]}

Step 4 : for each (Feature into NormSet !=Null)
Step 5 : calculate weight w= (Feature, Bk)
Step 6 : if (w>=th)
Ruleset.add ← {Feature,Label}
End if
End for
Step 7: return Ruleset.

2. Pattern Matching Algorithm for sub attack classification

In detection phase we use the sub-score of proposed method to detect each new point from the distribution of the trained data point. This phase match each new observation with established normal profile to detect anomaly. This include following

Step 1: Standardize data with means and variances from sample training dataset.

Step 2: Compute similarity score of each observation with trained eigenvectors which map observed data to subspace.

Step 3: Compute distances of each observation as in (5), (6). A new connection will have 1 or 2 distance values depend on 1-threshold method or 2-thresholds method.

Step 4: Compare thresholds and detection decision: If new connection's distance is greater than any of the established threshold, it marks as anomaly connection. Otherwise, it is normal connection.

Step 5 : Return w;

VI. RESULTS AND DISCUSSIONS

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.

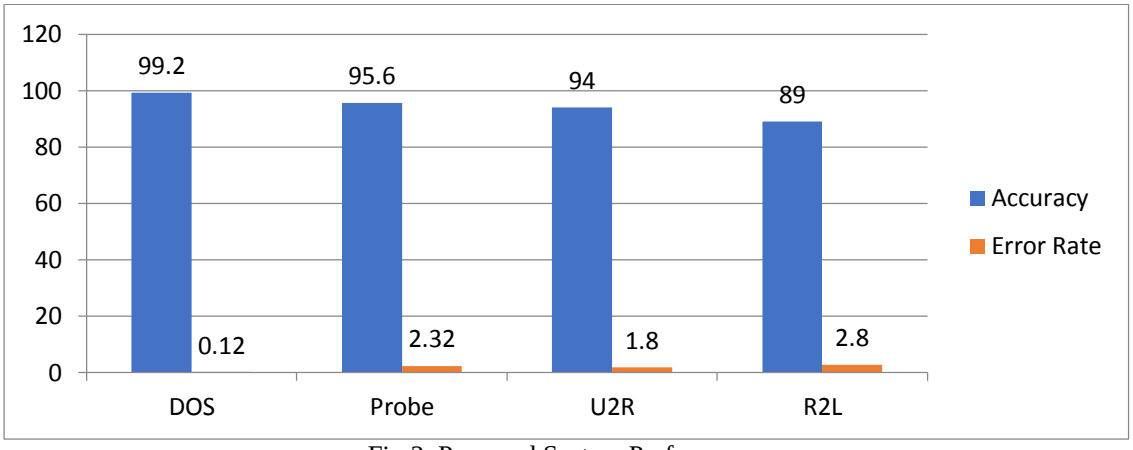

Fig.2: Proposed System Performance

Table 1 Proposed System Overall Performance

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

VII. CONCLUSION

In this research work we proposed ensemble method for network traffic anomaly detection. Our approach concentrated on building normal traffic profile of the anomaly detection model. Through experiments we also showed that some features of NSL-KDD and ISCX dataset are efficient with the normal profile. We propose a K-means clustering algorithm to reduce noise with input training data. The experiments showed that even with small training dataset (less than 1000 points), our approach has good performance including detection accuracy. We also proposed a new model integrates anomaly detection system with signature-based detection system along with some enhancements of building quality normal profile. In our future plan, we will develop and experiment the proposed model with an open source IDS in real network.

Future Work

Proposed research work also perform the better detection, On the basis Genetic algorithm implementation we got a ides we can achieve better detection rate for all attacks as well as unknown attacks. In future work we can minimized the computation time consuming by the genetic algorithm.

VIII. REFERENCES

[1]. 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.

[2]. 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.

[3]. R. Shanmugavadivu, "Network Intrusion Detection system using Fuzzy logic", ACM Digital Library, Volume 30 Issue 1, January 2007.

[4]. Emma Ireland, "Intrusion Detection with Genetic Algorithms and Fuzzy Logic", UMM CSci Senior Seminar Conference, Morris, MN, December 2013.

[5]. Rupesh B. Jadhav and Mr. Balasaheb B. Gite, "Real Time Intrusion Detection With Fuzzy, Genetic and Apriori Algorithm", International Journal of Advance Foundation and Research in Computer (IJAFRC), Vol 1, Nov 2014.

[6]. S. N. Pawar, "Intrusion detection in computer network using FGA", IEEE journal on parallel and distribute systems, Vol.23, No.3, March 2012.

[7]. P. Jongsuebsuk, N. Wattanapongsakorn and C. Charnsripinyo,"Real-Time Intrusion Detection with Fuzzy Genetic Algorithm", IEEE 2013.

[8]. Mohammed Anbar, Rosni Abdulah, Izan H. Hasbullah, Yung-Wey Chong; Omar E. Elejla, "Comparative Performance Analysis of classification algorithm for Internal Intrusion Detection ", 2016 14th Annual Conference on Privacy Security and Trust (PCT), Dec 12-14,2016, Penang, Malaysia.

[9]. Weiwei Chen, Fangang Kong, Feng Mei, GuiginYuan, Bo Li, "a novel unsupervised Anamoly detection Approach for Intrusion Detection System", 2017 IEEE 3rd International Conference on big data security on cloud, May 16-18,2017, Zhejiang, China.

[10]. Anna L. Buczak, Erha n Guven, "A Survey of Data Mining and Machine Learning methods for cybersecurity intrusion detection", IEEE communication surveys and tutorials, vol. 18, Issue 2,2016.

[11]. Salem Benferhat, Abdelhamid Boudjelida and Karim Tabia, Revising the outputs of a decision tree with expert knowledge: Application to intrusion detection and alert correlation", 2012 IEEE 24th International Conference on Tools with Artificial Intelligence.

[12]. Mr. V. K. Pachghare, Parag Kulkarni, \Pattern Based Network Security using Decision Trees and Support Vector Machine", IEEE International Conference on Tools with Artificial Intelligence.

[13]. Salem Benferhat, Karim Tabia, On the combination of naive Bayes and decision trees for intrusion detection",Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCAIAWTIC'05).

[14]. Jinhua Huang and Jiqing Liu, \Intrusion Detection System Based on Improved BP Neural Network and Decision Tree", 2012 IEEE fifth International Conference on Advanced Computational Intelligence(ICACI) October 18-20, 2012 Nanjing, Jiangsu, China..

[15]. Manish Kumar, Dr. M. Hanumanthappa, Dr. T. V. Suresh Kumar , Intrusion Detection System Using Decision Tree Algorithm", 2012 IEEE International Conference on Advanced Computational Intelligence(ICACI).

[16]. Mrutyunjaya Panda, Manas Ranjan Patra \A Comparative study of Data mining algorithms for network Intrusion Detection", 2008 IEEE first International Conference on Advanced Computational Intelligence.