ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

Ensemble Learning Network Traffic model for misuse and anomaly detection

[1] Jayashree Patil, [2] Sarita Patil
[1][2] Department of Computer Engineering, GHRCEM, Wagholi, Pune University, Pune.

***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 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.**

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

1. INTRODUCTION

Essentially Intrusion Detection System (IDS) ordered into two distinctive arranged Host Base Intrusion Detection System (HIDS) and Network Base Intrusion Detection System (NIDS). Today's system security foundation promisingly relies on System Interruption Recognition Framework (NIDS). 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. Two principle ID methods accessible are abnormality identification and abuse location. The oddity identification model depicts the typical conduct of a client to recognize this current client's irregular or unaccustomed activity. The wide augment utilization of PC systems in today's general public, especially the sudden surge in hugeness of e-business to the world riches, has made PC system asylum a global priority. Since it is not in fact practicable to fabricate a plan without any vulnerabilities, interruption recognition has occurred for an essential range of analyze. For the most part a gatecrasher 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 allude interruption as any arrangement of procedures that endeavor to trade off the honesty, privacy, or availability of a PC asset. The demonstration of identifying procedures that endeavor to trade off the honesty, attentiveness, or accessibility of a PC asset can be alluded as interruption discovery. An interruption location framework is a gadget or programming application that screens system and/or framework exercises for resentful exercises or approach infringement and produces data to an Administration position. Interruption identification is the procedure of observing the activities happening in a PC framework or organize and breaking down them for indications of likely occurrence, which are infringement or looming dangers of infringement of PC security arrangements, adequate use strategies, or normal security hones. Fundamentally when an interloper endeavor to break into a data framework or perform an activity not authoritatively permitted, we allude 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. An Interruption Location Framework is a plan for distinguishing interruptions and reporting them definitely to the best possible power

II. LITERATURE REVIEW

Implementation of IDS for distributed architecture using online Adaboost based approach combined with weak

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, March 2018

classifiers [1] 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 [2] 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.

Vikas Sharma, Aditi Nema developed layered approach to solve the delity problem using machine learning approach known as genetic algorithm [3] 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 [4] 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 [5] 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 [6] 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.

III. PROBLEM DEFINITION

In given research work, to develop the Intrusion Detection System for Distributed Architecture which will be able to detect and classify the incoming attacks using ensemble approach and fuzzy classifier use for drastic supervision as detection. Classification of sub attacks.

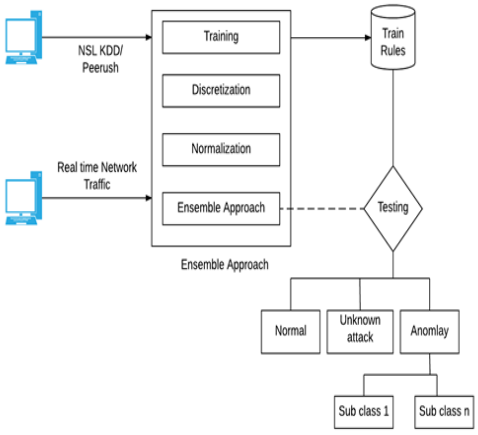

Figure 1: Proposed System Architecture

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, March 2018

IV. PROPOSED METHODOLOGY

In the proposed analyze work beneath is the structural planning graph with flowchart. To begin with we relate the hereditary calculation for signature base Rules.

Firstly, system generate the BK Rules base on NSL KDD dataset with different attribute selection. Then apply ensemble approach with multiple soft computing classifiers and finally test the data whether it is normal or anomaly.

V. SYSTEM MODULES

Basically there are two phase in the proposed system, we have taken NSL KDD dataset for system training as well testing purpose.
Module 1: Training Phase
Step 1: Upload training data for feature extraction.
Step 2: Apply Genetic algorithm for rule creation
Step 3: Create rules set as normal pool as well as intrusion pools set

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

VI. ALGORITHMS

1. Algorithm 1: GA for Rules Creation
Input: Set of network packet which consist 41 attributes with class label
Output: Set of normal as well intrusion rules
Step 1: Initialize the population randomly with 41 chromosomes.
Step 2: Initialize N (total number of records in the training set).
Step 3: For each chromosome in the new population
Step 4: Apply Crossover to best selected chromosome.
Step 5: Apply Mutation for each chromosome to generate new population.
Step 6: Calculate fitness= F(x) /sum (F(x)).
Step 7: Select 50% best fit chromosome and remove worse fit chromosome
Step 8: End for

2. Algorithm 2: NB
Step 1: Read all training rules from DB for each (Rec R into Train [])! = Null

Step 2 : items [] ←split(R)

Step 3: items1 [] ←split(TestF)

Step 4: Calculate Weight **(**DB [i], items1)

Step5: Return w;

3. Algorithm 3: ANN

Step 1: for all (T in **Hidden Layer** []! =null) do

Step 2: items [] ←split(T)

Step 3: items1 [] ←split(Input Nueron)

Step 4: w= Calculate Weight **(**Hidden Layer [i], Input Nueron **)**

Step5: Return w;

4. Algorithm 3: J48

Input: Feature of BK rules Train F [], features if test record Test F []

Output: highest Similarity weight for class label

Step 1: for all (T in **Train F** []! =null) do

Step 2: items [] ←split(T)

Step 3: items1 [] ←split(Test F)

Step 4: w = classify To All (Train, Test F[], Label)

Step5: Return w;

VII. RESULTS AND DISCUSSION

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.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, March 2018

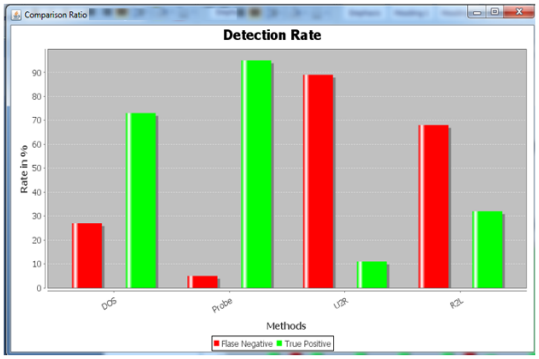

Fig 2 Existing System Performance

Table 1 Proposed System Overall Performance

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

VIII. 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.

IX. FUTURE WORK

Proposed research work also perform the better detection, On the basis ensemble approach implementation we got ideas system can achieve better detection rate for all attacks as well as unknown attacks. In future work we can have minimized the computation time consuming by the different algorithm.

X. REFERENCES

[1] 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.

[2] 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.

[3] 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).

[4] 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.

[5] 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).

[6] 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.