A Network Intrusion Detection System Using Supervised Learning Techniques

Shalini G, Jaya Kumar M, Abhishek P, Dhamodaran M

Department of Computer Science and Engineering
Dr. T. Thimmaiah Institute of Technology,
Kolar Gold Field, Karnataka, India

shalini@drttit.edu.in

Abstract. With rapid increase in the use of computer network in the fields of industries, education, commerce, social media etc., makes the data security a need of hour. And one of the major threats for data security is the network intrusions, where the attacker intruded the network to steal confidential data's like (passwords, account details, etc.,), tries to stop the services or take control of the user devices. In order to stop this intrusion, the Network Intrusion Detection System is proposed (NIDS). The NIDS monitors the network and if any attack occurs, this system detects the attack and will alert the user about respective attack that occurred. Thus,these systemshelp in preventing intrusions in our networks. And for developing the most accurate NIDS, four different Machine Learning (ML) algorithms are used.The four algorithms used to build a high accuracy system are Random forest, SVM, Naïve Bayes, KNN are used. By identifying the algorithm with highest accuracy, the exact attack can be detected hence the required preventive measures can be taken.

Keywords: Anomaly-Based Intrusion, KDD Cup99, Random Forest, SVM, KNN, Naïve Bayes.

1 Introduction

A Network Intrusion Detection System (NIDS) is asoftware that is developed and implemented for network security. This Network Intrusion Detection Systems continuously monitors network traffics and check for all inbound packets and patterns. These systems help us in preventing the attacker (Intruders) who tries to access or manipulating the user data or denying the services of the user. The NIDS Systems continuously monitors the user networks and checks for inbound network packets and data patterns. If any malicious pattern is discovered, the system checks for possible occurrences of attacks in the network and the alert is raised to user about the attacks occurred. The most common categories of intrusions that may occur in the network are dos, probe, r2l and u2r. In order to train the system to detect the above attacks, the KDDCup99 dataset is used. The KDDCup99 dataset contains 41 attributes, this da-

ICASISET 2020, May 16-17, Chennai, India

DOI 10.4108/eai.16-5-2020.2304044

taset contains both input and output labels. The system is trained using data with output label and tested using data without output labels.

On receiving the alert about the attack occurred, the user can take preventive measure to ensure the security of the network. The user can take preventive measures like barring attackersip address, mac address, etc., thus preventing the attacks in the future. The further enhance the NIDS, the system is combined Network Intrusion Prevention System (NIPS) to automate the detection and prevention of attacks. The major drawbacks of this NIDS systems are the accuracy of detection. The system should detect the attacks accurately so that the user can take respective preventive measures to obtain the network security and avoid future attacks.The main areas of applications of NIDS are Social Networking, E-Learning, E-Business etc., where these systems help the user in achieving network security. Since these NIDS has vast areas of applications the system should have more user-friendly interface and better Security Management Tools.

2 Related Work

A literature review is an objective, survey of the research works relevant to a topic that are under consideration. Here, is the replication of the literatures presented. The author B.A. Tama et.al [1] provides a system role for the computer security, where it relies on models trained using data to detect the malicious activities. NSL-KDD have been employed to improve performance and to process insecure information.Here it is been deployed in the transportation layer, where it copes the system from representing a further opportunity for the attackers. L.Atzori et.al [2] Rule based technique, which is time consuming due to the encoded rule. Here PCA & SVM are been proposed to monitor the network traffic, event occurring compared to Naïve Bayes algorithm. The objective is to increase the accuracy while avoiding false positive alert. Mohammed a. Ambusaidi et.al [3] As the development of the technology, many threats have emerged. Due to the serious threat of intruders, the algorithm performs balance detections and it tries to keep the false positive rate acceptable in the real time networking attacks. Jiong Zhang et.al [4] The solution to prevent intrusion is to use an effective tool to control the flow of network traffic. The overall efficiency is obtained by calculating the prediction accuracy, time taken to detect, false positive and false negative rates. [5][6]. Vijayakumar et.al [10][11] developed a network traffic restriction using MAC address.

3 Proposed Model

In this section detail approach of the proposed system called Network Intrusion Detection System is described.The Major problems faced by NIDS is accuracy, the detected attack should accurate so that the exact/relevant preventive measures can be taken, so accuracy of the system becomes the major objective in this work. In order to choose the algorithm that as highest accuracy we train four different algorithms and find an algorithm that as highest accuracy and it will be used. The historical records

(KDD Cup99 Dataset) is used as input. KDD Cup99 Dataset consists of 41 attributes and 125973 instances for determining four attacks in the network. The attacks that can be predicted using KDD Cup99 Dataset are DOS, Probe, r2l, u2r. The input is feed to the system for feature engineering where the dataset is divided into training data and test data. Four algorithms (Random Forest, SVM, Naïve Bayes, KNN) [7] are used in this work. The training data is processed to all the four algorithms to obtain the trained model and the model is validated using Test data and to predict the most accurate Result or Attack. [8] After detecting the attack, the system will send the user attack notification via Matplot library, text message, email etc., The architectural diagram of our proposed system is shown in the figure given below.

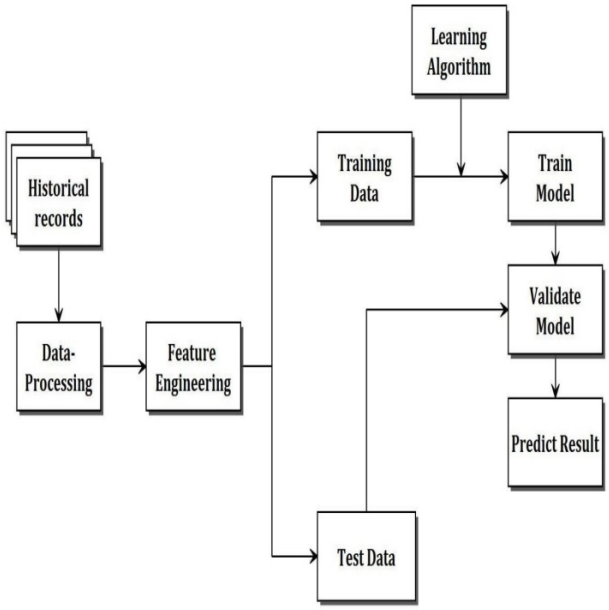

Fig 1: Architecture Diagram

3.1 Pre-Processing Stage

In this stage, the proposed NIDS system does three stages of pre-processing namely mapping, Nan data analysis, null analysis. (1) .Mapping, The KDD CUP99 dataset attributes are checked for their datatypes. The attributes of object/string datatype are mapped to integers value for computation. The attributes of object or string values are mapped to integer values for better computation and result. (2.)Nan analysis, missing values in dataset are identified and relevant value is added. In order to do Nan analysis Pandas Series. isna() function can be used which detect missing values in the given series object. It returns a Boolean same-sized object indicating if the values are NA. (3.) Null analysis,Null analysis is done to check for nullvalues in our input dataset. It

helps in finding null values of attributes and percentage of null values in dataset does helping us in computation.

3.2 Feature Engineering

After pre-processing the KDDCup99 dataset is splitted into training and test dataset. For achieving the greater accuracy of the system, the training dataset created should be larger than the test dataset. Since the system is implemented using supervised learning techniques and algorithms, the training dataset should contain both input and output label to train the learning models whereas the test should have only input data to validate the models so the output labels is dropped in test data.

3.3 Training Stage

In this stage, the all the four supervised learning models are trained using the training data and these models will extract the output labels i.e., xAttack attribute 0f KDDCup99 dataset keep it ready for prediction process. The model will be validate using test datasets, where the accuracy scores of models are obtained.Here large dataset is used to train our four algorithms. After training, our models will ready for prediction stage where only input data will be given and our model should check this input values and predict the most accurate output. The models after training are stored in user memory location for future use.

3.4 Prediction Process

In this process, the supervised learning models is provided with test data which doesn't contains output labels and the trained learning models should predicted this output labels. These predicted output labels are the attacks that may occur on the network. In real time application of NIDS, the system should read all inbound network packet and collect these values and provide them as the test data to the system so that the system checks for attacks or intrusions patterns in the networks. And if any malicious patterns are encountered the system should check for possible intrusion and if the intrusion is detected then the respective attack or intrusion should be reported to the user.

3.5 Output Visualization

After predicting the attacks, the system will send an alert message to the user about the attack that occurred with preventive messages of the attack. This alert notification can send to user using Matplot library, email, text message etc., On receiving the alert about the possible attacks or intrusions the user can take preventive measures, thus data security is attained.

4 Results and Discussions

This section describes the detail of experimental result of proposed NIDS system that tested over the test dataset. For the experimentation purpose, more than 100 test instances for attacks are checked for four different algorithms. It is found that Random Forest Classifier as more accuracy than other algorithms used in our work. The graph plotting the accuracy of four algorithms used in model is shown below:

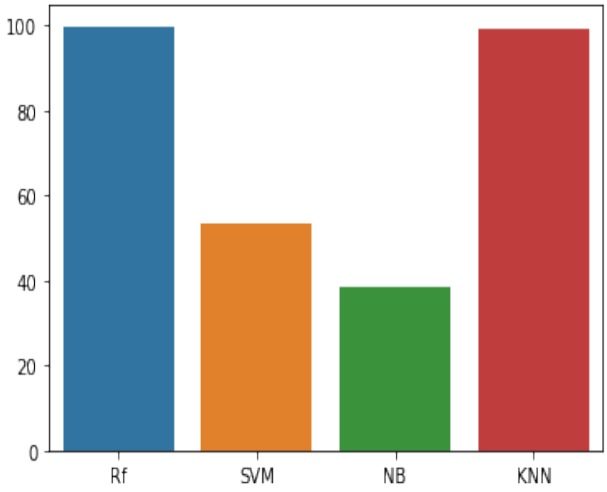

Fig 2: Algorithms Accuracy Plot

The Random Forest Algorithm as the accuracy of 99.86664126500285 and it's the most accurate algorithm that can be used to build NIDS using KDD Cup99 Dataset. The reason for random forest algorithm having high accuracy is that it created n decision trees and all these decision trees contains an individual outputs and to obtain the final output the Random Forest Algorithms uses majority voting techniques where average outputs of all decision tress are taken thus removing the bias in output. And another reason for high accuracy is the more trees in the forest, the more robust would be the prediction and thus higher accuracy.

5 Conclusion

This paper we conclude that, on using NIDS, the malicious activities can be predicted and an alert is raised to the user reducing the attacks on system or network. on using Random Forest algorithm, the most accurate attack in the network can be predicted thus correct preventive method can take to safeguard user data and information. For future work, After Detecting the attack, The Information and Data is sent to Network Intrusion Prevention System (NIPS). For resolving the current attack and Preventing Future attacks in the Networks

A Network intrusions prevention system (NIPS) is a network security system that helps the user in preventing the malicious attacks or intrusions in the networks. This system takes information from NIDS and performs the mechanism to prevent the attacks or intrusions in the networks. A Network intrusion prevention system (NIPS) prevents the attack by using information given by the A Network intrusion detection system (NIDS), the information given by NIDS i.e., (i.) Attack Logs such as type of attack, attack motive, attack pattern, etc. (ii.) Attacker Logs such as attacker IP address IMEI number etc. Using the Attack Logs and Attacker Logs the NIPS prevents The Attacks.

References

1. B. A. Tama and K.-H. Rhee, ``An extensive empirical evaluation of classier ensembles for intrusion detection task," Compute. System. Sci. Eng., vol. 32, no. 2, pp. 149_158, Nov. 2017.
2. L. Atzori, A. Iera, and G. Morabito, ``The Internet of Things: A survey," Compute. Network., vol. 54, no. 15, pp. 2787_2805, Oct. 2016.
3. Mohammed a. Ambusaidi, Member, Priyadarsi Nanda and Zhiyun Tan, “Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm", IEEE transactions on computers, vol. 65, no. 10, October 2016.
4. Jiong Zhang, Mohammad Zulkemine, and Anwar Haque, “Random-Forest-Based Network Intrusion Detection System", IEEE Transactions in Systems, Man, and Cybernetics, Part C (Applications and Reviews) (Volume: 38, Issue: 5, Sept. 2018).
5. Prof. D.P. Gaikwad and Dr.R.C. Thool, Architecture Taxonomy and Product of IDS, International Conference on Computer Applications, Computer Application-II, doi:10.3850/978-981-0873042_0382.
6. Ahmed, S.T., Sandhya, M. & Sankar, S. TelMED: Dynamic User Clustering Resource Allocation Technique for MooM Datasets Under Optimizing Telemedicine Network. Wireless PersCommun 112, 1061–1077 (2020). https://doi.org/10.1007/s11277-020-07091-x
7. S. T. Ahmed and S. Sankar, "Investigative Protocol Design of Layer Optimized Image Compression in Telemedicine Environment", Procedia Computer Science, vol. 167, pp. 2617-2622, 2020, [online] Available: https://doi.org/10.1016/j.procs.2020.03.323
8. Gunashree, M., Ahmed, S. T., Sindhuja, M., Bhumika, P., Anusha, B., &Ishwarya, B. (2020). A New Approach of Multilevel Unsupervised Clustering for Detecting Replication Level in Large Image Set. *Procedia Computer Science*, *171*, 1624-1633.https://doi.org/10.1016/j.procs.2020.04.174
9. M. Anathi, K. Vijayakumar , “An intelligent approach for dynamic network traffic restriction using MAC address verification”, Computer Communications,Elesvier,5 February 2020.
10. K. Pradeep Mohan Kumar, M. Saravanan, M. Thenmozhi ,K. Vijayakumar, “ Intrusion detection system based on GA-fuzzy classifier for detecting malicious attacks”, wiley, Feb 2019.