Network Intrusion Detection using Linear Regression

Ms. Vinitha Nalmas, M.Tech. Student, Department of CSE,

Ms. A Madhavi, Assistant Professor, Department of CSE,

VNR Vignana Jyothi Institute of Engineering and Technology, Telangana, India.

Abstract – Detecting intrusions can identify unknown attacks in a network and has been one of the successful ways to enhance network security. The current methods for identifying network anomalies are largely based on traditional machine learning models, such as KNN, SVM, etc. While these methods can accomplish excellent functionality, their accuracy is comparatively low and rely heavily on manual identification of network threats. BAT is a network anomaly detection model that has been developed combining Liner Regression, 3 Layer Neural Network and attention mechanism. Attention mechanism helps in filtering the network flow vector containing packet vectors, which can acquire the significant features for classifying the network traffic. We have also adopted multiple convolutional layers to identify the local features of network traffic. As multiple convolutional layers are used to analyze the data samples, the BAT model is referred to as BAT-MC. The softmax classifier is used for classifying the network activity. We assessed the proposed model on a public standard dataset, and the initial findings suggest that our model has better efficiency than other methods.

1. INTRODUCTION

1.1 Introduction

With the growth and advancement of digital technology, the Web offers a range of useful resources for users. However, we are now experiencing a range of security challenges. Network malware, data leakage, and disruptive threats are on the rise, making network protection the target of the public's best interest and government entities. Luckily, these issues are quite well fixed. Even so, mostly with the exponential expansion of Internet businesses, the forms of traffic load are growing regularly, and the features of network activity will become increasingly nuanced, which poses major problems in the identification of intrusions. How to classify numerous potentially malicious amounts of traffic, particularly unpredictable suspicious network traffics, is a key issue that cannot be avoided. Data transmission can, in particular, be split into two types (normal traffics and malicious traffics). Besides, network traffic can indeed be divided into several categories: Standard, DoS (Denial of Service Attacks), R2L (Root to Local Attacks), U2R (User to Root Attack), and Probe (Probing attacks). Therefore, detecting intrusions can be called a classification issue. By improving efficiency & recognizing suspicious traffic, the performance of network security can be significantly increased. Neural network methods have been commonly used in network security to track malicious activity. These approaches, though, come within the framework of supervised learning and also prioritize function optimization and variety, there have been difficulties in selecting features and cannot efficiently address huge intrusion data. Leading to low identification quality and high probability of false alarms. Subsequently, intrusion prevention techniques focused on Machine Learning are being proposed over the years.

A network classification model based on a convolutional layer has been proposed where the traffic data is stored as a graphic. This approach does not require manually identifying features, and also can take the raw traffic data as input data for the classifier. In another research, the investigators also provide an overview of the feasibility of Convolutional Neural Networks (CNNs) to monitor the behavior of connected devices by describing it as a series of states that shift over time. In another paper, the writers monitor the availability of the Long Short-Term Memory (LSTM) network in the classification of intruder traffic. Authors proved that LSTM will learn all the attack categories embedded in the training data. Many of these techniques consider entire network data that have a of series of bytes of information. They do not make good use of existing traffic data. For illustration, CNN transforms ongoing network activity to graphics. This is similar to considering traffic as autonomous and dismissing the internal connections of network traffic data. Network traffic is a hierarchical assembly of data. Specifically, network traffic is a data component comprising of several data components. A packet is a network unit made up of several bytes. Also, the traffic characteristics of those in the same and separate packets are substantially different. The packets must be processed individually. In several other terms, not all network data is similarly relevant for traffic in the extraction phase of certain network activity data.

1.2 Problem Statement

Intrusion prevention plays a major role in maintaining internet data stability. Even so, with the exponential growth of the internet industry, the forms of network transmission are growing regularly, and the features of network behavior are becoming more complicated, which presents huge challenges for detecting intrusions. How to distinguish different malicious internet traffic, in particular, Unusual potentially malicious access is a crucial concern that can then be stopped.

2. Literature Survey

2.1 A Review of the technique used

Intrusion prevention technology can be categorized into three main categories: pattern matching approaches, conventional artificial intelligence methods, and recurrent neural networks. Just at outset, people primarily use pattern recognition techniques to analyze intrusion. Pattern matching methodology is a key function of intrusion detection and prevention algorithms. In a publication, the researchers make a list of patterns matching methodologies in Intrusion Detection System: KMP calculation, BM calculation, BMH calculation, BMHS calculation, AC calculation, and AC-BM calculation. Analyses show that the improved methodologies can fast-track the calculation speeds and also reduce the time. In another research, Knuth-MorrisPratt Algorithm, and Rabin-Karp Algorithm are contrasted to check which of them is faster in detecting intrusions. Pcap documents have been utilized as datasets to decide the proficiency of the calculations.

These conventional pattern recognition calculations have genuine imperfections, which can't accomplish the impact of identifying intrusions. Finding an effective calculation that has high proficiency and low false positive rates remains the focal point of current work. With the advancement of AI/ ML, the use of fuzzy calculations for identifying intrusions has become another area of interest for researchers. The anomaly detection methodologies dependent on AI have made a ton of progress. In a research, the researchers propose a strategy for classifying network data that is dependent on Support Vector Machine (SVM). Test results on NSL-KDD cup 99 of network data indicated that the cataloguing accuracy of this technique was close to 99%. In a publication, the authors join k-mean clustering based on the KNN classifier. The trial results on the NSL-KDD dataset show that this technique significantly improves the speed of the KNN classifier.

In a research, the researchers propose another method to join the mismanagement and identifying anomalies where they apply the random forests calculation. Exploratory outcomes show that the discovery rate of the proposed framework is 94.7% and the flase positive rate is 2%. In another research, the NSL-KDD dataset is assessed through Artificial Neural Network (ANN). The discovery rate for identifying intrusions is 81.2% and for segregating the types of attacks it was 79.9%. In another research, a new method based on the Decision Tree (DT) is proposed to identify intrusions. Test results using the applicable feature selection (CFS) subset portray that the DT-based system to identify intrusions has higher precision. As portrayed above, AI techniques have been proposed and have made significant progress in successfully identifying intrusions. Nonetheless, these techniques require enormous preprocessing and complex classification of network traffic information. It is very difficult & complex to provide a solution to the colossal network data classification problem using AI/ ML approaches.

3. OVERVIEW OF THE SYSTEM

3.1 Existing System

- Study of the feasibility of recurrent neural networks (RNN) in existing approaches to identify the activity of internet traffic by predicting it as a series of sequences that evolve.
- Monitor the adequacy of the Long Short-Term Memory (LSTM) system in current techniques to assess invasion traffic. Findings further suggest that LSTM can acquire all classes of attacks concealed in preparation data.

3.1.1 Disadvantages of Existing System

- ✓ Many of the strategies consider the whole data traffic in general that consists of series of bytes of information. They cannot make good use of the internet traffic database.
- ✓ Current approaches consider traffic as separate and neglect the internal connections of internet traffic.

3.2 Proposed System

- ✓ We suggest an end-to-end BAT machine learning algorithm consisting of a linear regression and classification model. Linear regression can well address the issue of network security and offer a modern form of analysis.
- ✓ Compared to standard deep learning approaches the BAT-MC model will gather data in each packet. By maximizing the use of network activity mapping methods, the BAT design can extract features more thoroughly.
- ✓ We test our proposed model with an NSL-KDD data collection.

3.2.1 Advantages of Proposed System

- ✓ The proposed approach is used to evaluate the essential level of traffic variables to achieve fine-grained characteristics that are more helpful in malicious traffic detection.
- ✓ Also, at hidden layers, the features created by the softmax function are then imported into a fully connected layer for the fusion function, which possesses the major characteristics that correctly quantify the traffic behavior of the system.

3.3 System Modules

In this project work, I used two modules and each module has its functions, such as:

1. **Employee**
2. **Admin**

3.3.1 Employee module

Using the Employee module, employees of the organization can get themselves registered. Upon registration, the employee can login to their portal and raise a request to the Admin to check for anomalies in the network or to check whether any intrusion occurred in the network. Once the Admin verifies the network data as requested by the employee and send a response, the employee can view the response & take appropriate action as required. The employee can also view his profile.

3.3.2 Admin module

Administrator of the network can login to the Admin module & view the requests for validating the network traffic. The admin can log the network data based on the company's network traffic, then upload the data to admin portal, preprocess it, divide the dataset into train & test data & then pass it on to the model to predict from the provided data & obtain the results. These results can then be shared with the employee who raised the request to check for anomalies or intrusions the network & necessary action can be taken based on the data. The admin can view his profile, as well as manage other users from the portal.

4. RESULTS

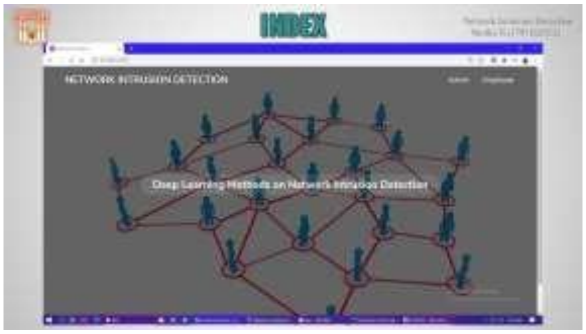

INDEX
NETWORK INTRUSION DETECTION
Deep Learning Methods on Network Intrusion Detection

Fig 4.1: Home page

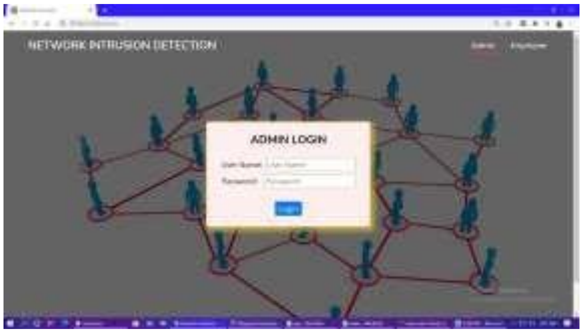


Fig 4.2: Admin Login Page

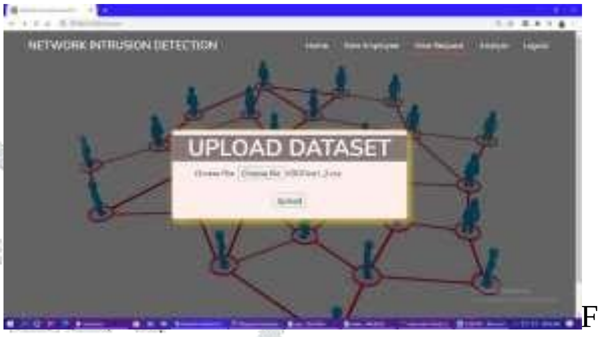


Fig 4.3: upload dataset

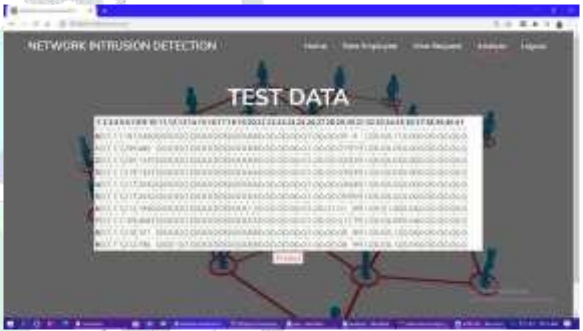
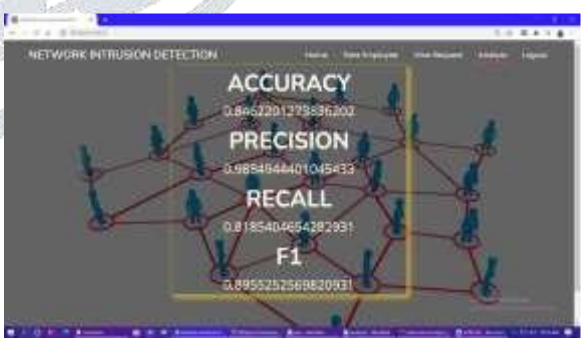


Fig 4.4: Test data



ACCURACY
0.8462201279836202
PRECISION
0.9894944401045433
RECALL
0.8185404654282931
F1
0.8955252569820931

Fig 4.5: Accuracy check

Fig 4.6: Employee Request page

Algorithm	Accuracy
Linear Regression	0.80
Random Forest	0.71
SVM	0.50
Naïve Bayes	0.69
KNN	0.50

Table 4.1: Accuracy comparison of Algorithms

Algorithm	Precision	
	0	1
Liner Regression	0.87	0.96
Random Forest	0.77	0.82
SVM	0.59	0.60
Naïve Bayes	0.70	0.71
KNN	0.40	0.62

Table 4.2: Precision comparison of Algorithms

Algorithm	Recall	
	0	1
Liner Regression	0.81	0.45
Random Forest	0.99	0.68
SVM	0.98	0.75
Naïve Bayes	0.98	0.69
KNN	0.99	0.62

Table 4.3: Recall comparison of Algorithms

Algorithm	F-Score	
	0	1
Liner Regression	0.93	0.91
Random Forest	0.73	0.63
SVM	0.64	0.69
Naïve Bayes	0.84	0.72
KNN	0.44	0.74

Table 4.4: F-Score comparison of Algorithms

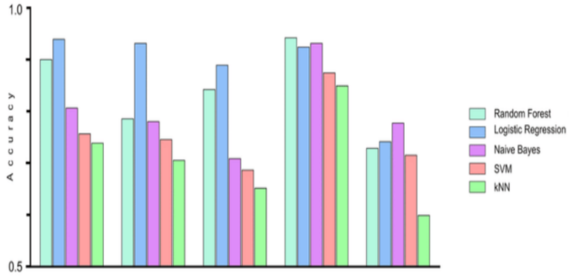


Fig 4.7: Accuracy comparison of Algorithms – Bar Graph

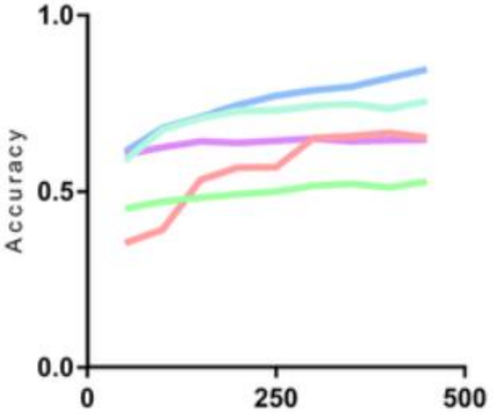


Fig 4.8: Accuracy comparison of Algorithms – Line Graph

5. CONCLUSION

The current deep learning methods in the network traffic classification research don't make full use of the network traffic structured information. Drawing on the application methods of deep learning in the field of natural language processing, we propose a novel model BAT-MC via the two phase's learning of Linear Regression & 3 Layer Neural Network and attention on the time series features for intrusion detection using NSL-KDD dataset. Each data packet can produce a packet vector. These packet vectors are arranged to form a network flow vector. Attention layer is used to perform feature learning on the network flow vector composed of packet vectors. The above feature learning process is automatically completed by deep neural network without any feature engineering technology.

Future Enhancement

- ✓ Many changes are in store for intrusion detection and intrusion prevention. Some of these changes could actually be negative — at least from the perspective of intrusion detection. For example, the Gartner Group, a technology research and consulting organization, asserts that IDSs will soon be relics of the past. Gartner says that IDSs have not established themselves in the IT marketplace, that they produce too low a return on investment (ROI) for all the resources expended, and that excessive false alarms and misses have greatly impaired their usefulness. Gartner predicts that intrusion prevention technology will prevail in the belief that shutting off intrusions altogether is better than allowing intrusions to occur and just monitoring them. Accordingly, Gartner recommends that IT organizations turn to firewalls, not IDSs. Many IT security experts denounced Gartner's prediction, though, saying that Gartner does not really understand how intrusion detection fits in with a layered defense approach (of which many believe that intrusion detection is a critical part) and that intrusion detection technology is still growing and improving.
- ✓ Regardless of whatever sliver of truth there may or may not be in Gartner's prediction, two things are certain — intrusion detection is still a long way from being mature, and intrusion prevention technology is in its infancy. Massive changes are in store for both areas. There are some of the areas within intrusion detection and intrusion prevention in which substantial and beneficial progress is likely to occur. These areas include the following:
- ✓ The continued reduction in reliance on signatures in intrusion detection
- ✓ The growth of intrusion prevention
- ✓ Advances in data correlation and alert correlation methods
- ✓ Advances in source determination
- ✓ Inclusion of integrated forensics functionality in IDSs and IPSs
- ✓ Greater use of honeypots

REFERENCES

- B. B. Zarpelo, R. S Miani, C. T. Kawakani, and S. C. de Alvarenga, ''A survey of intrusion detection in Internet of Things,'' J. Netw. Comput. Appl., vol. 84, pp. 25–37, Apr. 2017.
  B. Mukherjee, L. T. Heberlein, and K. N. Levitt, ''Network intrusion detection,'' IEEE Netw., vol. 8, no. 3, pp. 26–41, May 1994.
- S. Kishorwagh, V. K. Pachghare, and S. R. Kolhe, ''Survey on intrusion detection system using machine learning techniques,'' Int. J. Control Automat., vol. 78, no. 16, pp. 30–37, Sep. 2013.
- N. Sultana, N. Chilamkurti, W. Peng, and R. Alhadad, ''Survey on SDN based network intrusion detection system using machine learning approaches,'' Peer-to-Peer Netw. Appl., vol. 12, no. 2, pp. 493–501, Mar. 2019.
- M. Panda, A. Abraham, S. Das, and M. R. Patra, ''Network intrusion detection system: A machine learning approach,'' Intell. Decis. Technol., vol. 5, no. 4, pp. 347–356, 2011.
- W. Li, P. Yi, Y. Wu, L. Pan, and J. Li, ''A new intrusion detection system based on KNN classification algorithm in wireless sensor network,'' J. Electr. Comput. Eng., vol. 2014, pp. 1–8, Jun. 2014.