BAT: DEEP LEARNING METHODS ON NETWORK INTRUSION DETECTION USING NSL-KDD DATASET

MOHAMMAD ABDUL WAHEED FAROOQUI[1], BELLAMKONDA UPENDER[2]

ITHAGONI TEJASWINI[3], ADE PAVAN4

1 & 2, Associate Professor, CSE department, Brilliant Grammar School Educational Society's Group of Institutions, Hyderabad, TS.
3&4 UG SCHOLARS, CSE department, Brilliant Grammar School Educational Society's Group of Institutions, Hyderabad, TS.

ABSTRACT

Intrusion detection has shown to be a successful method of network security since it can detect unknown attacks from network traffic. These days, most network anomaly detection techniques are based on conventional machine learning models, such KNN, SVM, etc. Despite the fact that these techniques can yield some exceptional features, they have a low accuracy rate and mostly rely on human traffic feature design, which is out of date in the big data era. A traffic anomaly detection model, or BAT, is suggested as a solution to the issues of low accuracy and feature engineering in intrusion detection. Bidirectional Long Short-Term Memory (BLSTM) and attention mechanisms are included in the BAT model. The network flow vector made up of packet vectors produced by the BLSTM model is screened using an attention mechanism in order to extract the essential characteristics for classifying network traffic. To capture the local aspects of traffic data, we also use numerous convolutional layers. We call the BAT model BAT-MC since it uses several convolutional layers to process data samples. Network traffic is classified using the softmax classifier. The suggested end-to-end model can automatically learn the hierarchy's essential features without the need for feature engineering expertise. It can effectively explain the behavior of network traffic and enhance anomaly detection capabilities. We evaluate our model using a publicly available benchmark dataset, and the experimental findings show that it outperforms alternative comparison techniques.

1.INTRODUCTION

With the development and improvement of Internet technology, the Internet is providing various convenient services for people. However, they are also facing various security threats. Network viruses, eavesdropping and malicious attacks are on the rise, causing network security to become the focus of attention of the society and government departments. Fortunately, these problems can be well solved via intrusion detection. Intrusion detection plays an important part in ensuring network information security. However, with the explosive growth of Internet business, traffic types in the network are increasing day by day, and network behavior characteristics are becoming increasingly complex, which brings great challenges to intrusion detection. How to identify various malicious network traffics, especially unexpected malicious network traffics, is a key problem that cannot be avoided.

In fact, network traffic can be divided into two categories (normal traffics and malicious traffics). Furthermore, network traffic can also be divided into five categories: Normal, DoS (Denial of Service attacks), R2L (Root to Local attacks), U2R (User to Root attack) and Probe (Probing attacks). Hence, intrusion detection can be considered as a classification problem. By improving the performance of classifiers in effectively identifying malicious traffics, intrusion detection accuracy can be largely improved.

2. LITERATURE SURVEY

A SURVEY: INTRUSION DETECTION TECHNIQUES FOR INTERNET OF THINGS
AUTHORS: Sarika Choudhary and Nishtha Kesswani (1991)
The latest buzzword in internet technology now a days is the Internet of Things. The Internet of Things (IOT) is an ever-growing network which will transform real-world objects into smart or intelligent virtual objects. IOT is a heterogeneous network in which devices with different protocols can connect

with each other in order to exchange information. These days, human life depends upon the smart things and their activities. Therefore, implementing protected communications in the IOT network is a challenge. Since the IOT network is secured with authentication and encryption, but not secured against cyber-attacks, an Intrusion Detection System is needed. This research article focuses on IOT introduction, architecture, technologies, attacks and IDS. The main objective of this article is to provide a general idea of the Internet of Things, various intrusion detection techniques, and security attacks associated with IOT.

NETWORK INTRUSION DETECTION
AUTHORS: B. Mukherjee, L.T. Heberlein and K.N. Levitt (1994)

Intrusion detection is a new, retrofit approach for providing a sense of security in existing computers and data networks, while allowing them to operate in their current "open" mode. The goal of intrusion detection is to identify unauthorized use, misuse, and abuse of computer systems by both system insiders and external penetrators. The intrusion detection problem is becoming a challenging task due to the proliferation of heterogeneous computer networks since the increased connectivity of computer systems gives greater access to outsiders and makes it easier for intruders to avoid identification. Intrusion detection systems (IDSs) are based on the beliefs that an intruder's behavior will be noticeably different from that of a legitimate user and that many unauthorized actions are detectable. Typically, IDSs employ statistical anomaly and rule based misuse models in order to detect intrusions. A number of prototype IDSs have been developed at several institutions, and some of them have also been deployed on an experimental basis in operational systems. In the present paper, several host-based and network-based IDSs are surveyed, and the characteristics of the corresponding systems are identified. The hostbased systems employ the host operating system's audit trails as the main source of input to detect intrusive activity, while most of the network-based IDSs build their detection mechanism on monitored network traffic, and some employ host audit trails as well. An outline of a statistical anomaly detection algorithm employed in a typical IDS is also included.

SURVEY ON SDN BASED NETWORK INTRUSION DETECTION USING MACHINE LEARNING APPROACH
AUTHORS: N. Sultana, N. Chilamkurti, W. Peng, and R. Alhadad (2019)

Software Defined Networking Technology (SDN) provides a prospect to effectively detect and monitor network security problems ascribing to the emergence of the programmable features. Recently, Machine Learning (ML) approaches have been implemented in the SDN-based Network Intrusion Detection Systems (NIDS) to protect computer networks and to overcome network security issues. A stream of advanced machine learning approaches – the deep learning technology (DL) commences to emerge in the SDN context. In this survey, we reviewed various recent works on machine learning (ML) methods that leverage SDN to implement NIDS. More specifically, we evaluated the techniques of deep learning in developing SDN-based NIDS. In the meantime, in this survey, we covered tools that can be used to develop NIDS models in SDN environment. This survey is concluded with a discussion of ongoing challenges in implementing NIDS using ML/DL and future works.

3. EXISTING SYSTEM

SYST Machine learning methods have been widely used in intrusion detection to identify malicious traffic. However, these methods belong to shallow learning and often emphasize feature engineering and selection. They have difficulty in features selection and cannot effectively solve the massive intrusion data classification problem, which leads to low recognition accuracy and high false alarm rate. The traffic anomaly detection methods based on machine learning have achieved a lot of success. It proposes a new method of feature selection and classification based on support vector machine (SVM). Experimental results on NSL-KDD cup 99 of intrusion detection data set showed that the classification accuracy of this method with all training features reached 99%. Combination k-mean clustering on the basis of KNN classifier. The experimental results on NSL-KDD dataset show that this method greatly improves the performance of KNN classifier. It proposes a new framework to combine the misuse and the anomaly detection in which they apply the random forests algorithm. Experimental results show that the overall detection rate of the hybrid system is 94.7% and the overall false positive rate is 2%. In the performance of NSL-KDD dataset is evaluated via Artificial Neural Network (ANN). The detection rate obtained is 81.2% and 79.9% for intrusion detection and attack type classification task respectively for NSL-KDD dataset.

3.1.1DISADVANTAGES OF EXISTING SYSTEM

- Low Accuracy
- False Positive Rate

3.2 PROPOSED SYSTEM

The accuracy of the BAT-MC network can reach 84.25%, which is about 4.12% and 2.96% higher than the existing CNN and RNN model, respectively.

The following are some of the key contributions and findings of our work:

- It proposes an end-to-end deep learning model BAT-MC that is composed of BLSTM and attention mechanism. BAT-MC can well solve the problem of intrusion detection and provide a new research method for intrusion detection.
- They introduce the attention mechanism into the BLSTM model to highlight the key input. Attention mechanism conducts feature learning on sequential data composed of data package vectors. The obtained feature information is reasonable and accurate.
- Compare the performance of BAT-MC with traditional deep learning methods, the BAT-MC model can extract information from each packet. By making full use of the structure information of network traffic, the BAT-MC model can capture features more comprehensively.

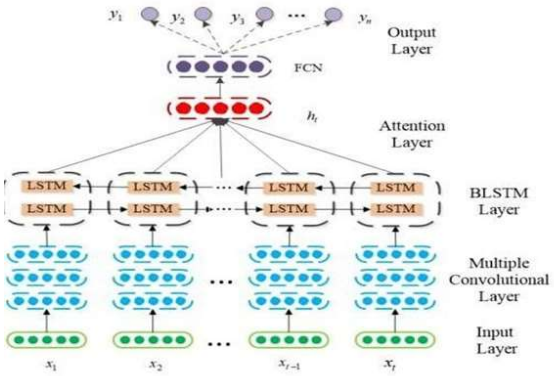
- Evaluate our proposed network with a real NSL-KDD dataset. The experimental results show that the performance of BAT-MC is better than the traditional methods.

3.2.1 ADVANTAGES OF PROPOSED SYSTEM

- Better Performance
- High Accuracy

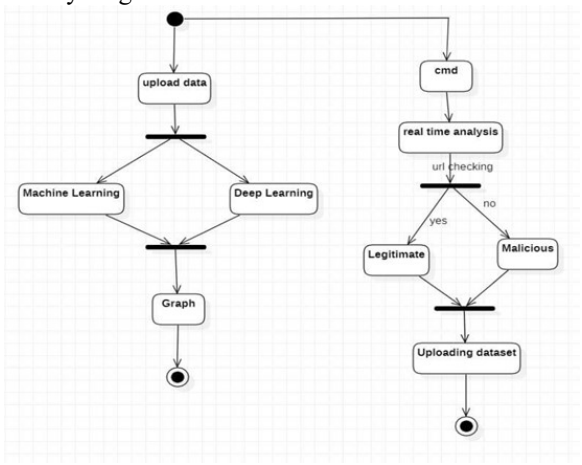
SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture of deep learning methods on net work intrusion detection using NSL-KDD dataset.



ActivityDiagram

Activity diagrams are graphical representations of workflows of step wise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by- step workflows of components in a system. An activity diagram shows the overall flow of control..



5.SYSTEM IMPLEMENTATION

•Uploading the Dataset and Data Preprocessing.
•Create evaluation metrics using BAT-MC.
•Comparing the experimental accuracy results.
•Performance analysis of BAT-MC and visualizing performance analysis.
•Home
•Our Database
•Dataset Download
•About

Here we upload the preprocessed NSL-KDD dataset, which is a open source dataset.

There are three symbolic data types in NSL-KDD data features: protocol type, flag and service. We use one-hot encoder mapping these features into binary vectors.

One-Hot Processing: NSL-KDD dataset is processed by one-hot method to transform symbolic features into numerical features. For example, the second feature of the NSL-KDD data sample is protocol type. The protocol type has three values: tcp, udp, and icmp. One-hot method is processed into a binary code that can be recognized by a computer, where tcp is [1, 0, 0], udp is [0, 1, 0], and icmp is [0, 0, 1].

Normalization Processing: The value of the original data may be too large, resulting in problems such as ''large numbers to eat decimals'', data processing overflows, and inconsistent weights so on. We use standard scaler to normalize the continuous data into the range [0, 1]. Normalization processing eliminates the influence of the measurement unit on the model training, and makes the training result more dependent on the characteristics of the data itself.

5.1.2 Create evaluation metrics using BAT-MC

The BAT-MC model can extract information from each packet. By making full use of the structure information of network traffic, the BAT-MC model can capture features more comprehensively. The evaluate of our proposed network with a real NSL-KDD dataset. The experimental results show that the performance of BAT-MC is better than the traditional methods.

5.1.3 Comparing the experimental accuracy results

Experimental results show that the overall detection rate of the hybrid system is 94.7% and the overall false positive rate is 2%. In, the performance of NSL-KDD dataset is evaluated via Artificial Neural Network (ANN). The detection rate obtained is 81.2% and 79.9% for intrusion detection and attack type classification task respectively for NSL-KDD dataset. In, an intrusion detection method based on decision tree (DT) is proposed.

5.1.4 Performance analysis of BAT-MC and visualizing performance analysis

The BAT model combines BLSTM (Bidirectional Long Short-term memory) and attention mechanism. After the LSTM () analysis is performed, the blstm_accuracy () is performed where it shows the loss of 1.13 and the accuracy of 92.00 percent. The accuracy is performed based on BLSTM and the attention mechanism.

5.1.5 Home

For checking the website URL is malicious or not, running the Django project from visual studio code by the commands. Python manage.py migrate (for migrating to the server).Python manage.py run server (for running the server). The linkhttp://127.0.0.1:8000//is copied to the any browser. The home page where the URL are scanned which is done by clicking 'Real Time Analysis'.Then we will get the URL as legitimate or malicious.

5.1.6 Our Database

Our Database contains the detailed information of the URL which we have given in the web application developed. It contains the information of the URL like its address, organization, state, country, domain, rank, etc.

5.1.7 Dataset Download

Here all the verified URL's are stored in the dataset which we can download and use later for the future purpose.We can download in the form of Excel sheet.

5.1.8 About

An intrusion detection system (IDS) is a device, or software application that monitors a network or systems for malicious activity or policy violations. Any intrusion activity or violation is typically reported either to an administrator or collected centrally using a security information and event management (SIEM) system.

6. SYSTEM TESTING

Unittesting

Unit testing involves the design of test cases that validate that the internal program logic is

functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

6.1.2Integrationtesting

Integration tests are designed to test integrated software components to determine if they

actually, run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

6.1.3Functionaltest

Functional tests provide systematic demonstrations that functions tested are available as

specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centred on the following items:

7.RESULTS

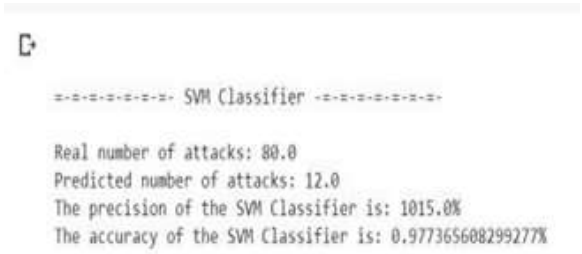
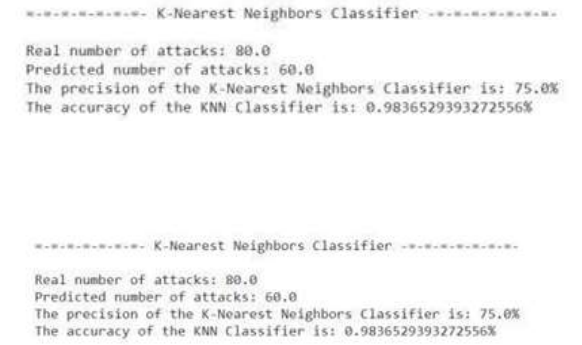


Fig 7.1 Result of SVM classifier

Fig 7.2 Result of Decision Tree



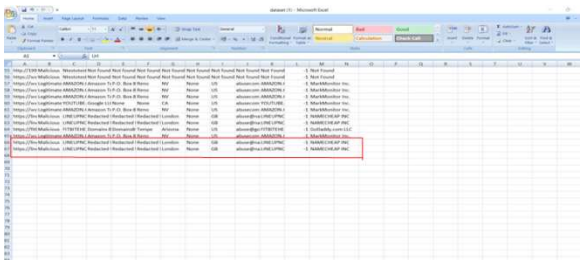


Fig : The excel sheet shows the data updated

8. CONCLUSION

The structured information in network traffic is not fully utilized by the deep learning techniques used in network traffic classification research at the moment. Using the two-phase learning of BLSTM and focusing on the time series characteristics for intrusion detection using the NSL-KDD dataset, it suggests a unique model, BAT-MC, based on the application methods of deep learning in the field of natural language processing. The forward LSTM and backward LSTM are connected by the BLSTM layer, which is used to extract features from each packet's traffic bytes. A packet vector can be generated by each data packet. A network flow vector is created by arranging these packet vectors. The network flow vector made up of packet vectors is subjected to feature learning using the attention layer. Without the use of feature engineering technologies, deep neural networks automatically complete the aforementioned feature learning process. The issue of manual design features is successfully avoided by this model. The KDD Test+ and KDDTest-21 datasets are used to evaluate the BAT-MC method's performance. According to experimental results on the NSL-KDD dataset, the accuracy of the BAT-MC model is fairly high. By comparing with some standard classifier, these comparisons show that BAT-MC models results are very promising when compared to other current deep learning-based methods. Hence, it believes that the proposed method is a powerful tool for the intrusion detection problem.

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. [Abney, 1991]
- 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.
- S. Garg and S. Batra, ‘‘A novel ensembled technique for anomaly detection,’’ Int. J. Commun. Syst., vol. 30, no. 11, p. e3248, Jul. 2017.
- F. Kuang, W. Xu, and S. Zhang, ‘‘A novel hybrid KPCA and SVM with GA model for intrusion

detection,'' Appl. Soft Comput.., vol. 18, pp. 178–184, May 2014.

• W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, ''Malware traffic classification using convolutional neural network for representation learning,'' in Proc. Int. Conf. Inf. Netw. (ICOIN), 2017, pp. 712–717.

• P. Torres, C. Catania, S. Garcia, and C. G. Garino, ''An analysis of Recurrent Neural Networks for Botnet detection behavior,'' in Proc. IEEE Biennial Congr. Argentina (ARGENCON), Jun. 2016, pp. 1–6.

• R. C. Staudemeyer and C. W. Omlin, ''ACM press the south African institute for computer scientists and information technologists conference - east London, south Africa (2013.10.07- 2013.10.09) proceedings of the south African institute for computerscientists and information technologists co,'' in Proc. South African Inst. Comput. Scientists Inf. Technol. Conf., 2013, pp. 252–261.

• S. Cornegruta, R. Bakewell, S. Withey, and G. Montana, ''Modelling radiological language with bidirectional long short-term memory networks,'' in Proc. 7th Int. Workshop Health Text Mining Inf. Anal., 2016, pp. 1–11.

• O. Firat, K. Cho, and Y. Bengio, ''Multi-way, multilingual neural machine translation with a shared attention mechanism,'' in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., 2016, pp. 1–10.