International Journal for Modern Trends in Science and Technology, 9(04): 46-51, 2023

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0904008
Available online at: http://www.ijmtst.com/vol9issue04.html

An Efficent Cyber Security Intrusion Detection and Analysis using Deep Learning

Chandra Sekhar Koppireddy | Annavarapu Sujana Subha Vyshnavi | Akula Ramkumar | Jonnada Satya Sri | Pottimurthi Sai Sampath Kumar | Gollamudi Ashita

Department of Computer Science and Engineering, Pragati Engineering College (A), Surampalem (East Godavari) A.P,India.

To Cite this Article
Chandra Sekhar Koppireddy, Annavarapu Sujana Subha Vyshnavi, Akula Ramkumar, Jonnada Satya Sri, Pottimurthi Sai Sampath Kumar and Gollamudi Ashita. An Efficent Cyber Security Intrusion Detection and Analysis using Deep Learning. International Journal for Modern Trends in Science and Technology 2023, 9(04), pp. 46-51. https://doi.org/10.46501/IJMTST0903008

Article Info
Received: 02 March 2023; Accepted: 25 March 2023; Published: 30 March 2023.

ABSTRACT

Machine learning techniques are being widely used to develop an intrusion detection system (IDS) for detecting and classifying cyber-attacks at the network-level and host-level in a timely and automatic manner. However, many challenges arise since malicious attacks are continually changing and are occurring in very large volumes requiring a scalable solution. There are different malware datasets available publicly for further research by cyber security community.

In this paper, deep neural network (DNN), a type of deep learning model is explored to develop a flexible and effective IDS to detect and classify unforeseen and unpredictable cyber-attacks. The continuous change in network behaviour and rapid evolution of attacks makes it necessary to evaluate various datasets which are generated over the years through static and dynamic approaches. This type of study facilitates to identify the best algorithm which can effectively work in detecting future cyber-attacks. A comprehensive evaluation of experiments of DNNs and other classical machine learning classifiers are shown on various publicly available benchmark malware datasets. The optimal network parameters and network topologies for DNNs is chosen through following hyper parameter selection methods with KDDCup 99 dataset. All experiments of DNNs are run till 1,000 epochs with learning rate varying in the range [0.01-0.5]. The DNN model which performed well on KDDCup 99 is applied on other datasets such as NSL-KDD, UNSW-NB15, Kyoto, WSN-DS and CICIDS 2017 to conduct the benchmark. Our DNN model learns the abstract and high dimensional feature representation of the IDS data by passing them into many hidden layers. Through a rigorous experimental testing it is confirmed that DNNs perform well in comparison to the classical machine learning classifiers. Finally, we propose a highly scalable and hybrid DNNs framework called Scale-Hybrid-IDS-AlertNet (SHIA) which can be used in real time to effectively monitor the network traffic and host-level events to proactively alert possible cyber-attacks

1. INTRODUCTION

With the wide spreading usages of internet and increases in access to online contents, cybercrime is also happening at an increasing rate. Intrusion detection is the first step to prevent security attack. Hence the security solutions such as Firewall, Intrusion Detection System (IDS), Unified Threat Modeling (UTM) and Intrusion Prevention System (IPS) are getting much attention in studies. IDS detect attacks from a variety of systems and network sources by collecting information and then analyze the information for possible security breaches. The network based IDS analyzes the data packets that travel over a network and this analysis are carried out in two ways. Till today anomaly based

detection is far behind than the detection that works based on signature and hence anomaly based detection still remains a major area for research. The challenges with anomaly based intrusion detection are that it needs to deal with novel attack for which there is no prior knowledge to identify the anomaly. Hence the system somehow needs to have the intelligence to segregate which traffic is harmless and which one is malicious or anomalous and for that machine learning techniques are being explored by the researchers over the last few years. IDS however is not an answer to all security related problems. For example, IDS cannot compensate weak identification and authentication mechanisms or if there is a weakness in the network protocols. Studying the field of intrusion detection first started in 1980 and the first such model was published in 1987. For the last few decades, though huge commercial investments and substantial research were done, intrusion detection technology is still immature and hence not effective. While network IDS that works based on signature have seen commercial success and widespread adoption by the technology based organization throughout the globe, anomaly based network IDS have not gained success in the same scale. Due to that reason in the field of IDS, currently anomaly based detection is a major focus area of research and development. And before going to any wide scale deployment of anomaly based intrusion detection system, key issues remain to be solved . But the literature today is limited when it comes to compare on how intrusion detection performs when using supervised machine learning techniques.

2.LITERATURE SURVEY

A macro-social exploratory analysis of the rate of interstate cyber-victimization-A Macro-Social Exploratory Analysis of the Rate of Interstate Cyber-Victimization Hyojong Song, M. Lynch, John K. Cochran Law 2016

This study examines whether macro-level opportunity indicators affect cyber-theft victimization. Based on the arguments from criminal opportunity theory, exposure to risk is measured by state-level patterns of internet access (where users access the internet). Other structural characteristics of states were measured to determine if variation in social structure impacted cyber-victimization across states. The current study found that structural conditions such as unemployment and non-urban population are associated with where users access the internet. Also, this study found that the proportion of users who access the internet only at home was positively associated with state-level counts of cyber-theft victimization. The theoretical implications of these findings are discussed.

Incremental anomaly-based intrusion detection system using limited labeled data

With the proliferation of the internet and increased global access to online media, cybercrime is also occurring at an increasing rate. Currently, both personal users and companies are vulnerable to cybercrime. A number of tools including firewalls and Intrusion Detection Systems (IDS) can be used as defense mechanisms. A firewall acts as a checkpoint which allows packets to pass through according to predetermined conditions. In extreme cases, it may even disconnect all network traffic. An IDS, on the other hand, automates the monitoring process in computer networks. The streaming nature of data in computer networks poses a significant challenge in building IDS. 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.

Modeling and implementation approach to evaluate the intrusion detection system

Intrusions detection systems (IDSs) are systems that try to detect attacks as they occur or when they were over. Research in this area had two objectives: first, reducing the impact of attacks; and secondly the evaluation of the system IDS. Indeed, in one hand the IDSs collect network traffic information from some sources present in the network or the computer system and then use these data to enhance the systems safety. In the other hand, the evaluation of IDS is a critical task. In fact, its important to note the difference between evaluating the effectiveness of an entire system and evaluating the characteristics of the system components. In this paper, we present an approach for IDS evaluating based on measuring the performance of its components. First of all, in order to implement the IDS SNORT components safely we have proposed a hardware platform based on embedded systems. Then we have tested it by using a generator of traffics and attacks based on Linux KALI (Backtrack) and

Metasploite 3 Framework. The obtained results show that the IDS performance is closely related to the characteristics of these components.

Toward credible evaluation of anomaly-based intrusion-detection methods

Since the first introduction of anomaly-based intrusion detection to the research community in 1987, the field has grown tremendously. A variety of methods and techniques introducing new capabilities in detecting novel attacks were developed. Most of these techniques report a high detection rate of 98% at the low false alarm rate of 1%. In spite of the anomaly-based approach's appeal, the industry generally favors signature-based detection for mainstream implementation of intrusion-detection systems. While a variety of anomaly-detection techniques have been proposed, adequate comparison of these methods' strengths and limitations that can lead to potential commercial application is difficult. Since the validity of experimental research in academic computer science, in general, is questionable, it is plausible to assume that research in anomaly detection shares the above problem. The concerns about the validity of these methods may partially explain why anomaly-based intrusion-detection methods are not adopted by industry. To investigate this issue, we review the current state of the experimental practice in the area of anomaly-based intrusion detection and survey 276 studies in this area published during the period of 2000-2008. We summarize our observations and identify the common pitfalls among surveyed works.

Importance of intrusion detection system (IDS)

Intruders computers, who are spread across the Internet have become a major threat in our world, The researchers proposed a number of techniques such as (firewall, encryption) to prevent such penetration and protect the infrastructure of computers, but with this, the intruders managed to penetrate the computers. IDS has taken much of the attention of researchers, IDS monitors the resources computer and sends reports on the activities of any anomaly or strange patterns The aim of this paper is to explain the stages of the evolution of the idea of IDS and its importance to researchers and research centres, security, military and to examine the importance of intrusion detection systems and categories , classifications, and where can put IDS to reduce the risk to the network.

3.EXISTINGSYSTEM

A novel supervised machine learning system is developed to classify network traffic whether it is malicious or benign. To find the best model considering detection success rate, combination of supervised learning algorithm and feature selection method have been used. Through this study, it is found that Artificial Neural Network (ANN) based machine learning with wrapper feature selection outperform support vector machine (SVM) technique while classifying network traffic.

DISADVANTAGES OF EXISTING SYSTEM:

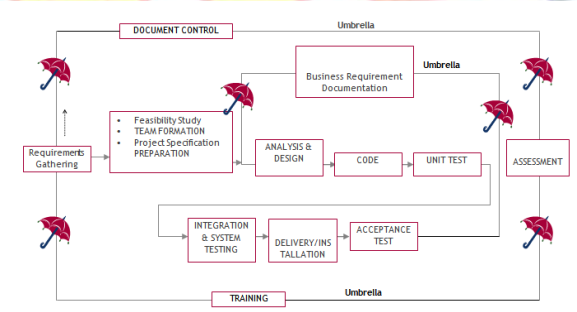
1. Less Accuracy.

3.2 PROPOSED SYSTEM

To evaluate the performance, NSL-KDD dataset is used to classify network traffic using SVM and ANN supervised machine learning techniques. Comparative study shows that the proposed model is efficient than other existing models with respect to intrusion detection success rate.

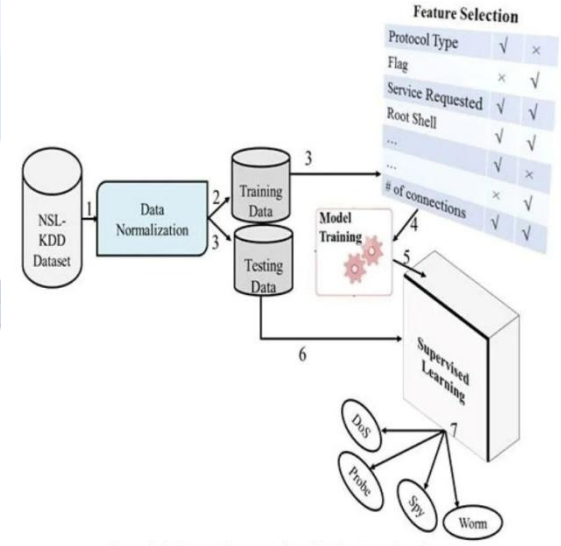
SDLC (Umbrella Model):

SDLC is nothing but Software Development Life Cycle. It is a standard which is used by software industry to develop good software.



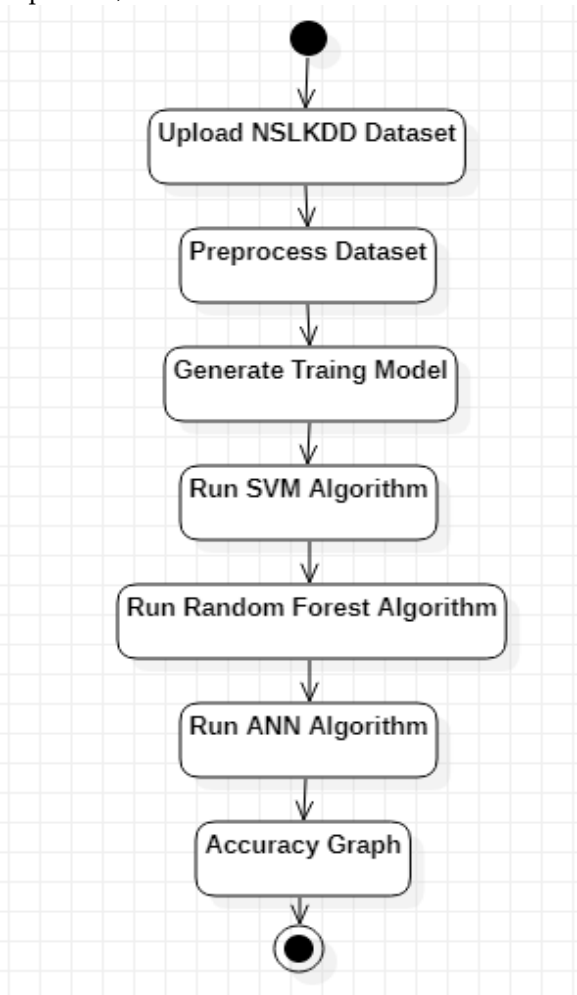
SYSTEM ARCHITECTURE

BelowdiagramdepictsthewholesystemarchitectureofThe mosttrendingarticlesevery yearusingNLP technique.'



ActivityDiagram

Activity diagram is another important diagram in UML to describe dynamic aspects of the system. It is basically a flow chart to represent the flow form one activity to another activity. The activity can be described as an operation of the system. So the control flow is drawn from one operation to another. This flow can be sequential, branched or concurrent.



5.SYSTEM IMPLEMENTATION

MODULES

1. Upload NSLKDD Dataset
2. Preprocess Dataset
3. Generate Training Model
4. Run SVM Algorithm
5. Run ANN Algorithm
6. Run Random Forest Algorithm
7. Accuracy Graph

MODULE DESCRIPTION

Upload NSLKDD Dataset: Upload NSL KDD Dataset Module Is used to upload dataset.

Pre-process Dataset: Pre-process Dataset Module is used to assign numeric values to each attack names as algorithms will not understand string names.

Generate Training Model: Generate Training Model Module is used to generate model for training purpose. We can see dataset arrange in such a format so algorithms can build training and test set for prediction and accuracy result

Run SVM Algorithm: Run SVM Algorithm to get its prediction accuracy. We can see SVM prediction accuracy is 52%.

Run Random Forest Algorithm: Run Random Forest Algorithm to get its accuracy. We can see random forest also got same accuracy.

Run ANN Algorithm: Run ANN Algorithm to get its accuracy. We can see ANN accuracy is better than other two algorithms. ANN algorithm accuracy may be vary different times as it hidden layer will be chosen randomly from dataset.

Accuracy Graph: Accuracy Graph module to get Accuracy graph. In graph x-axis represents algorithm name and y-axis represents accuracy and ANN is the propose technique.

6. SYSTEM TESTING

Implementation and Testing:

Implementation is one of the most important tasks in project is the phase in which one has to be cautions because all the efforts undertaken during the project will be very interactive. Implementation is the most crucial stage in achieving successful system and giving the users confidence that the new system is workable and effective. Each program is tested individually at the time of development using the sample data and has verified that these programs link together in the way specified in the program specification. The computer system and its environment are tested to the satisfaction of the user.

Implementation

The implementation phase is less creative than system design. It is primarily concerned with user training, and file conversion. The system may be requiring extensive user training. The initial parameters of the system should be modifies as a result of a programming. A simple operating procedure is provided so that the user can understand the different functions clearly and quickly. The different reports can be obtained either on the inkjet or dot matrix printer, which is available at the disposal of the user. The proposed system is very easy to implement. In general implementation is used to mean the process of converting a new or revised system design into an operational one.

Testing

Testing is the process where the test data is prepared and is used for testing the modules individually and later the validation given for the fields. Then the system testing takes place which makes sure that all components of the

system property functions as a unit. The test data should be chosen such that it passed through all possible condition. Actually testing is the state of implementation which aimed at ensuring that the system works accurately and efficiently before the actual operation commence. The following is the description of the testing strategies, which were carried out during the testing period.

7.RESULTS

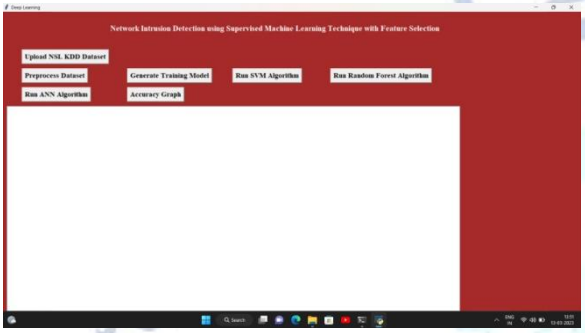


Fig.7.1 Homepageofour project

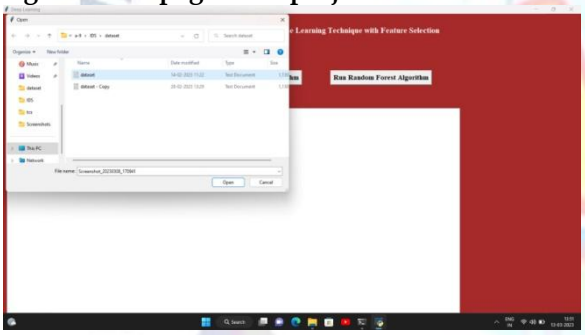


Fig. 7.2 Selecting "Upload NSL-KDD Dataset"

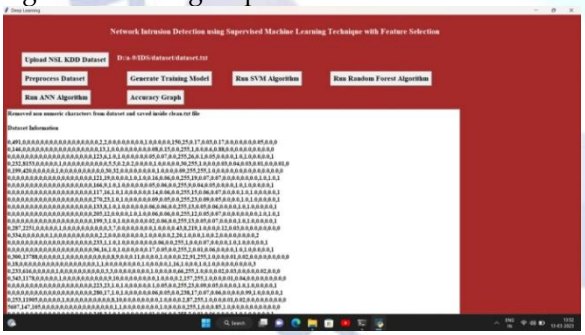


Fig. "Preprocess Dataset"

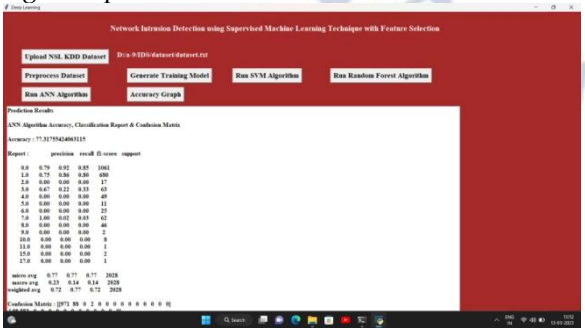


Fig. "ANN Algorithm Accuracy"

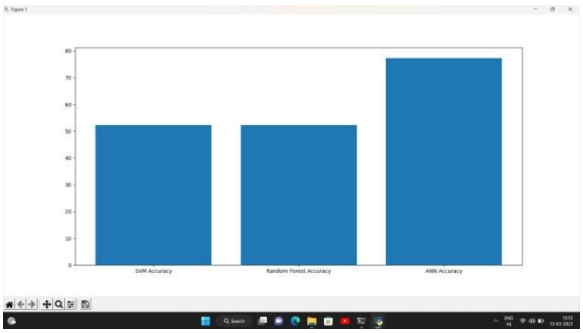


Fig. "Accuracy Graph"

8.CONCLUSION & FUTURE WORK

In this paper, we have presented different machine learning models using different machine learning algorithms and different feature selection methods to find a best model. The analysis of the result shows that the model built using ANN and wrapper feature selection outperformed all other models in classifying network traffic correctly with detection rate of 94.02%. We believe that these findings will contribute to research further in the domain of building a detection system that can detect known attacks as well as novel attacks. The intrusion detection system exist today can only detect known attacks. Detecting new attacks or zero day attack still remains a research topic due to the high false positive rate of the existing systems.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of interstate cyber-victimization," American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, 2016.

[2] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in Web Research (ICWR), 2017 3th International Conference on, 2017, pp. 178–184.

[3] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in International Conference on Networked Systems, 2015, pp. 513–517.

[4] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516–524, 2010.

[5] A. S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, 2011.

[6] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," arXiv preprint arXiv:1312.2177, 2013.

[7] N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," International Journal of Computing and Business Research (IJCBR) ISSN (Online), pp. 2229–6166, 2013.

[8] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," computers & security, vol. 28, no. 1–2, pp. 18–28, 2009.

[9] [9] Manjula Devarakonda Venkata1, Sumalatha Lingamgunta & K Murali, Health Care Automation in Compliance to Industry 4.0 Standards: A Case Study of Liver Disease Prediction ,Journal of Scientific & Industrial Research , Vol. 82, February 2023, pp. 263-268, DOI: 10.56042/jsir.v82i2.70215

[10] Manjula Devarakonda Venkata1, Sumalatha Lingamgunta & K Murali, Health Care Automation in Compliance to Industry 4.0 Standards: A Case Study of Liver Disease Prediction ,Journal of Scientific & Industrial Research , Vol. 82, February 2023, pp. 263-268, DOI: 10.56042/jsir.v82i2.70215

[11] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," Procedia Computer Science, vol. 89, pp. 117–123, 2016.

[12] J. Zheng, F. Shen, H. Fan, and J. Zhao, "An online incremental learning support vector machine for large-scale data," Neural Computing and Applications, vol. 22, no. 5, pp. 1023–1035, 2013.