IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ENHANCING NETWORK SECURITY: INTRUSION DETECTION AND PREVENTION

[1] U.Satya Narayana, [2] N.Siva Nagamani

Abstract: This paper is evaluating performance of three supervised machine learning algorithms such as SVM (Support Vector Machine), Random Forest and Decision Tree. Machine learning algorithms will be used to detect whether request data contains normal or attack (anomaly) signatures. Now-a-days all services are available on internet and malicious users can attack client or server machines through this internet and to avoid such attack request IDS (Network Intrusion Detection System) will be used, IDS will monitor request data and then check if it contains normal or attack signatures, if contains attack signatures then request will be dropped. IDS will be trained with all possible attacks signatures with machine learning algorithms and then generate train model, whenever new request signatures arrived then this model applied on new request to determine whether it contains normal or attack signatures. In this paper we are evaluating performance of three machine learning algorithms such as SVM , Random Forest and Decision Tree .To avoid all attacks IDS systems has developed which process each incoming request to detect such attacks and if request is coming from genuine users then only it will forward to server for processing, if request contains attack signatures then IDS will drop that request and log such request data into dataset for future detection purpose.To detect such attacks IDS will be prior train with all possible attacks signatures coming from malicious user's request and then generate a training model. Upon receiving new request IDS will apply that request on that train model to predict it class whether request belongs to normal class or attack class. To train such models and prediction various data mining classification or prediction algorithms will be used.

Index Terms **-IDS,Random Forest,Decision Tree,SVM,anomaly,machine learning.**

Introduction

With the wide spreading usages of internet and increases in access to online contents, cybercrime is also happening at an increasing rate [1-2]. 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 detects attacks from a variety of systems and network sources by collecting information and then analyzes the information for possible security breaches [3]. 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 [4-5]. 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 [6]. 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 [7]. For the last few decades, though huge commercial investments and substantial research were done, intrusion detection technology is still immature and hence not effective [7]. 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 [8]. And before going to any wide scale deployment of anomaly based intrusion detection system, key issues remain to be solved [8]. But the literature today is limited when it comes to compare on how intrusion detection performs when using supervised machine learning techniques [9]. To protect target systems and networks against malicious activities anomaly-based network IDS is a valuable technology. Despite the variety of anomaly-based network intrusion detection techniques described in the literature in recent years [8], anomaly detection functionalities enabled security tools are just beginning to appear, and some important problems remain to be solved. Several anomaly based techniques have been proposed including Linear Regression, Support Vector Machines (SVM), Genetic Algorithm, Gaussian mixture model, knearest neighbor algorithm, Naive Bayes classifier, Decision Tree [3,5]. Among them the most widely used learning algorithm is SVM as it has already established itself on different types of problem [10]. One major issue on anomaly based detection is though all these proposed techniques can detect novel attacks but they all suffer a high false alarm rate in general. The cause behind is the complexity of generating profiles of practical normal behavior by learning from the training data sets [11]. The major challenges in evaluating performance of network IDS is the unavailability of a comprehensive network based data set [13]. Most of the proposed anomaly based techniques found in the literature were evaluated using KDD CUP 99 dataset [14]. In this paper we used SVM, Decision Tree and Random Forest Tree –three machine learning techniques, on NSLKDD [15] which is a popular benchmark dataset for network intrusion. The motivation behind doing an intrusion detection and prevention project is to enhance the security of computer systems and networks. Intrusion detection systems (IDS) and intrusion prevention systems (IPS) are designed to detect and prevent unauthorized access, misuse, and modification of computer systems and networks. In today's digital age, where businesses and individuals rely heavily on computer systems and networks, cyber attacks have become increasingly common and sophisticated. Intruders can steal sensitive information, disrupt services, and cause significant financial and reputational damage. Therefore, detecting and preventing such attacks is crucial to ensuring the integrity,availability, and confidentiality of computer systems and networks.An IDS system analyzes network traffic and system logs to detect anomalous behavior and potential security breaches. An IPS system is an extension of IDS that can automatically respond to detected threats by blocking suspicious traffic or reconfiguring the network.Together, IDS and IPS provide a multi-layered defense mechanism to protect against various types of cyber attacks.

By doing an intrusion detection and prevention project, individuals and organizations can learn how to deploy and configure these systems, understand their strengths and weaknesses, and develop skills in detecting and responding to security threats. It can also help them gain knowledge in developing effective security policies and procedures to prevent future attacks.Another motivation behind doing an intrusion detection and prevention project is to comply with regulatory requirements and industry standards. Many industries, such as finance, healthcare, and government, are required to comply with regulations and standards that mandate the use of intrusion detection and prevention systems.

For example, the Payment Card Industry Data Security Standard (PCI DSS) requires merchants and service providers that accept payment cards to implement intrusion detection and prevention measures to protect against attacks on their systems. Similarly, the Health Insurance Portability and Accountability Act (HIPAA) requires healthcare organizations to implement intrusion detection and prevention systems to protect the confidentiality, integrity, and availability of electronic protected health information .In addition to complying with regulations and standards, implementing intrusion detection and prevention systems can also help organizations meet customer expectations and maintain a competitive advantage. Customers expect their personal and financial data to be protected, and a breach can lead to loss of trust and reputation damage. Therefore, investing in security measures such as IDS and IPS can help organizations demonstrate their commitment to security and protect their brand.Overall, the motivation behind doing an intrusion detection and prevention project can be multifaceted and can include a desire to enhance security, comply with regulations and standards, and meet customer expectations.

I. LITERATURE REVIEW

Incremental anomaly-based intrusion detection system using limited labeled data:

Cybercrime is increasing, and firewalls and Intrusion Detection Systems (IDS) can be used as defense mechanisms. An online classification method using an incremental naive Bayesian classifier and active learning is proposed for IDS applications. The method overcomes streaming data challenges and reduces labeling costs while improving accuracy and Kappa compared to the incremental naïve Bayesian approach.

Proposed method uses online classification and active learning for IDS. Overcomes streaming data challenges and reduces labeling costs. Improved accuracy and Kappa compared to the incremental naive Bayesian approach. Promising solution for IDS applications.

Modeling and implementation approach to Evaluate the intrusion detection system:

IDSs detect attacks in real-time or after they occur, with two objectives: reducing attack impact and evaluating IDS effectiveness. They collect network traffic information to enhance system safety. IDS evaluation is critical, noting the difference between evaluating the entire system and its components. This paper proposes an approach for IDS evaluation based on measuring component performance. A hardware platform based on embedded systems was proposed to safely implement IDS SNORT components, tested with a traffic and attack generator based on Linux KALI (Backtrack) and Metasploit 3 Framework. Results show IDS performance is closely related to component characteristics.

The model describes an approach for evaluating the performance of Intrusion Detection Systems (IDS) by measuring the performance of its components. A hardware platform based on embedded systems was proposed for implementing SNORT components safely, and testing was conducted using a traffic and attack generator. The results showed that IDS performance is closely related to the characteristics of its components. 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.

Intruders on the Internet pose a major threat. Researchers have proposed techniques like firewalls and encryption to protect computers, but intruders still manage to penetrate them. Intrusion Detection Systems (IDS) monitor computer resources and report any anomalies or strange patterns. This paper explains the evolution of IDS and its importance to various industries. It examines IDS categories, classifications, and how they can reduce network risk.

Anomaly-based network intrusion detection: Techniques, systems and challenges:

The internet and computer networks face increasing security threats, and flexible and adaptive security approaches are needed. Anomaly- based network intrusion detection techniques can help protect against malicious activities. Despite recent advancements in such methods, security tools with anomaly detection are just starting to appear, and challenges remain. This paper reviews anomaly-based intrusion detection techniques, presents available platforms, systems under development, and research projects. It also outlines challenges for the widespread deployment of anomaly- based intrusion detectors, with an emphasis on assessment issues.This paper discusses the use of anomaly-based network intrusion detection techniques to protect against increasing security threats to computer networks. The paper reviews existing techniques, presents available platforms and systems, and outlines challenges for widespread deployment of these detection systems.

A deep learning approach for network intrusion detection system:

A Network Intrusion Detection System (NIDS) helps system administrators to detect network security breaches in their organizations. However, many challenges arise while developing a flexible and efficient NIDS for unforeseen and unpredictable attacks. We propose a deep learning based approach for developing such an efficient and flexible NIDS. We use Self- taught Learning (STL), a deep learning based technique, on NSL-KDD - a benchmark dataset for network intrusion. We present the performance of our approach and compare it with a few previous work. Compared metrics include accuracy, precision, recall, and f-measure values.

Proposing a deep learning approach for developing an efficient and flexible Network Intrusion Detection System (NIDS), using Self-taught Learning (STL) on the NSL-KDD benchmark dataset. Performance is compared to previous work based on accuracy, precision, recall, and f- measure values.

II. PROPOSED METHODOLOGY

Machine Learning algorithms are totally subject to data since it is the most vital perspective that makes model training possible. On the other hand, if won't be able to make sense out of that data, before feeding it to ML algorithms, a machine will be useless. In straightforward words, we generally need to take care of the right data for example the data in the right scale, group, and containing important features, for the problem we need a machine to solve.

This makes data preparation the most important step in the ML process. Data preparation defined as the procedure that makes our dataset more appropriate to work with in the ML process.

DATASET:

The NSL-KDD dataset is a benchmark dataset commonly used for intrusion detection research. It is an updated version of the original KDD Cup 99 dataset, which was widely used in the intrusion detection research community.

The NSL-KDD dataset was created to address some of the limitations of the KDD Cup 99 dataset, including the use of redundant and irrelevant features and the unrealistic traffic patterns. The NSL-KDD dataset is a cleaned and preprocessed version of the KDD Cup 99 dataset, and it contains a more representative sample of network traffic.

The NSL-KDD dataset consists of two sets: the training set and the testing set. The training set contains 125,973 instances, and the testing set contains 22,544 instances. The instances in the dataset are classified into one of four categories: normal, DoS, probe, or R2L (remote-to-local) and U2R (user-to-root) attacks.

The dataset contains 41 features, including basic features such as the duration of the connection, the protocol used, and the number of bytes sent and received, as well as more advanced features such as the number of failed login attempts and the number of root accesses. The features have been preprocessed and normalized to have a similar range of values.

The NSL-KDD dataset has been used extensively in research on intrusion detection systems, and it has been shown to be a useful benchmark dataset for evaluating the performance of different machine learning algorithms. However, it is important to note that the dataset may not fully represent real-world network traffic, and care should be taken when interpreting the results of experiments using the dataset.

ALGORITHMS:

Algorithms used are Support Vector Machine(SVM), Decision Tree, Random Forest Tree.

SUPPORT VECTOR MACHINE (SVM):

It is a popular method for solving complex classification problems where the number of features is high. The hyperplane is a line that separates the two classesSVM finds the optimal hyperplane by maximizing the margin, which is the distance between the hyperplane and the closest data points of each class.

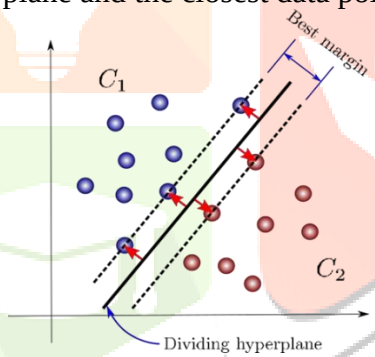


Fig. 1 Support Vector Machine

DECISION TREE:

The decision tree is built from a set of training data that contains both normal and malicious traffic.Each node in the decision tree is associated with a feature of the network traffic, such as the source IP address,the destination IP address,or the protocol used.Tree that can classify the network traffic as normal or malicious.

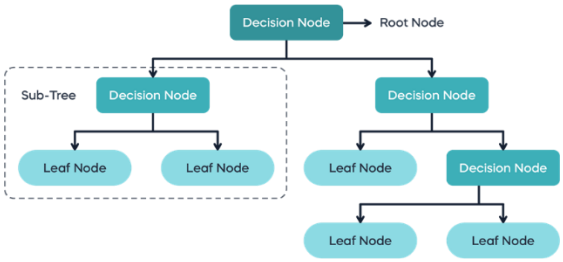


Fig. 2 Decision Tree

RANDOM FOREST TREE:

It is an extension of the decision tree algorithm.Multiple decision trees are used to build a more accurate and robust model.

Combines the results of the individual decision trees to make a final prediction.

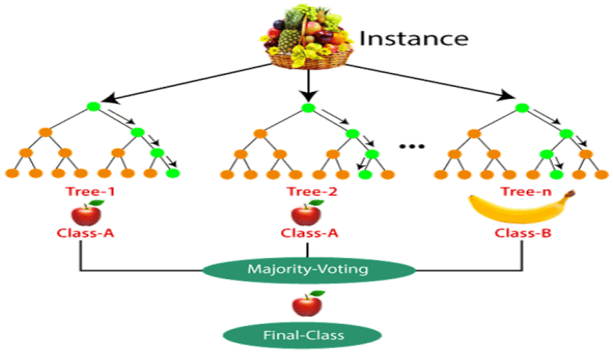


Fig. 3 Random Forest

Proposed System

The execution of the process will be explained clearly with the help of the continuous screenshots.The whole processes includes upload of the datasets and training the machine learning algorithm.The entire process occurs in simple steps in the case of frontend. We use NSL KDD data set to train the machine learning modelFirst we open CMD in the source program file and run the program file by giving command Python IDS.py . After giving that command the program starts executing and display a pie chart. we can see the overview of our data set how many data sets are present and which dataset uses which protocol and differentiating them in two categories such as normal and attack.

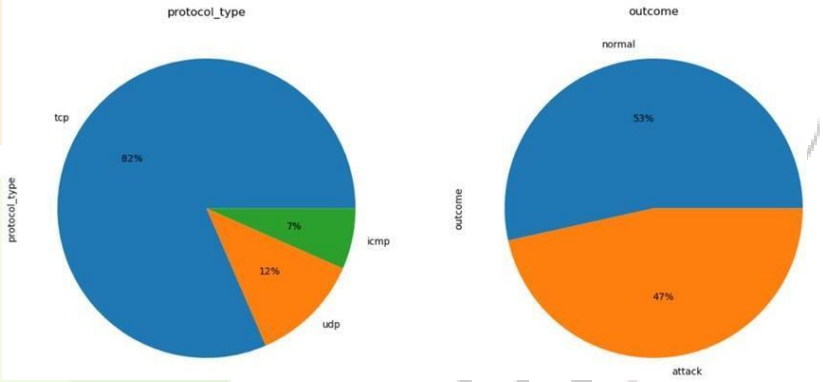


Fig 4 Data in Pie Chart

In this stage the dataset is preprocessed. the data set is verified and null values are removed and features which are unwanted to these algorithms are also removed.After closing this window we can see another pop up window which shows the importance of the features which has given priority.

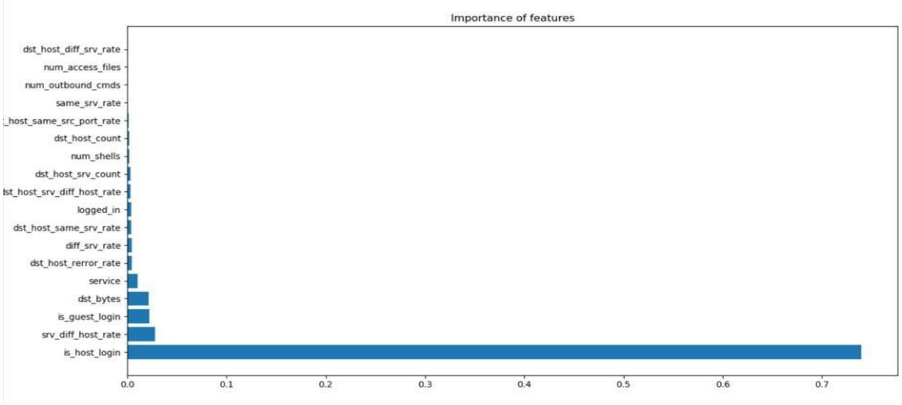
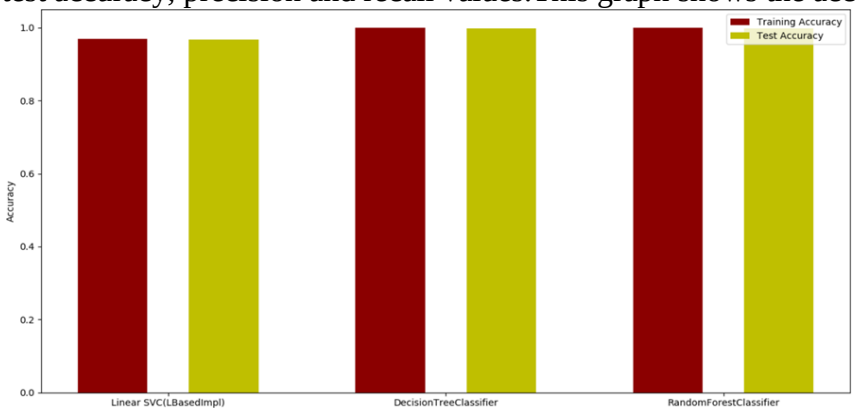


Fig 5. Important Features

After closing this window we can see a confusion matrix of both support vector machine and decision tree by seeing these confusion matrix we can get a overview about how both models are trained.By closing these two confusion matrix we can see another window which show the importance of the features which are selected for the random forest tree in random forest tree we take few decision tree outputs and by bagging

method we finalize a single output.After closing these window we can see another popup window which show the confusion matrix of the random forest tree. Then again we get three windows which shows graphs about test accuracy, precision and recall values.This graph shows the accuracy.



And final we can see the accuracy of all the algorithms in the Command prompt.

IV. CONCLUSION

In conclusion, the use of machine learning algorithms such as SVM, decision tree, and random forest can be effective in enhancing network security intrusion detection and prevention.

SVM is a powerful algorithm that can effectively classify data into different categories based on a set of features. Decision trees are easy to interpret and provide clear insights into the decision-making process. Random forest, on the other hand, can provide high accuracy and reduce the risk of overfitting.

By combining these machine learning algorithms with other network security measures such as firewalls and encryption, it is possible to develop a robust network security system that can detect and prevent a wide range of security threats.

However, it is important to note that machine learning algorithms are not a silver bullet and should be used in conjunction with other security measures. Additionally, the effectiveness of these algorithms depends on the quality of the data used to train them. Therefore, it is important to ensure that the data used to train these algorithms is accurate, diverse, and representative of real-world network security threats.

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 andimplementation 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) 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) 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.

10) 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.

11) F. Gharibian and A. A. Ghorbani, “Comparative study of supervised machine learning techniques for intrusion detection,” in Communication Networks and Services Research, 2007. CNSR’07. Fifth Annual Conference on, 2007, pp. 350–358.

12) J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural networks, vol. 61, pp. 85–117, 2015.
13) N. Moustafa and J. Slay, “UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set),” in Military Communications and Information Systems Conference (MilCIS), 2015, 2015, pp. 1–6.
14) T. Janarthanan and S. Zargari, “Feature selection in UNSW-NB15 and KDDCUP’99 datasets,” in Industrial Electronics (ISIE), 2017 IEEE 26th International Symposium on, 2017, pp. 1881–1886.
15) L. Dhanabal and S. P. Shantharajah, “A study on NSL-KDD dataset for intrusion detection system based on classification algorithms,” International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 6, pp. 446–452, 2015.
16) Javaid, Q. Niyaz, W. Sun, and M. Alam, “A deep learning approach for network intrusion detection system,” in Proceedings of the 9th EAI International Conference on Bio- inspired Information and Communications Technologies (formerly BIONETICS), 2016, pp. 21–26.
17) M. Panda, A. Abraham, and M. R. Patra, “Discriminative multinomial naive bayes for network intrusion detection,” inInformation Assurance and Security (IAS), 2010 Sixth International Conference on, 2010, pp. 5–10.
18) Ingre and A. Yadav, “Performance analysis of NSL-KDD dataset using ANN,” in Signal Processing And Communication Engineering Systems (SPACES), 2015 International Conference on, 2015, pp. 92–96.
19) L. M. Ibrahim, D. T. Basheer, and M. S. Mahmod, “A comparison study for intrusion database (Kdd99, Nsl-Kdd) based on self organization map (SOM) artificial neural network,” Journal of Engineering Science and Technology, vol. 8, no. 1, pp. 107–119, 2013.