South Asian Journal of Engineering and Technology Research Article

Open Access Full Text Article 123(2022) 138-145 DOI: 10.26524/sajet.2022.12.48

Detection of Cyber Attacks in Network Using Machine Learning Techniques

Mr. N. SaiKiran[1], P. D. S. Pradeep Naidu[2], K. Harshini[3], M. Venkateswarlu[4], V. Surya Narayana Reddy[5]

[1]Associate Professor, Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India

[2,3,4,5]Final Year (B. Tech), Dept. of Information Technology, QIS College of Engineering and Technology, Andhra Pradesh, India

Corresponding author.
Correspondence: N. SaiKiran
E-mail: sikaran@gmail.com

Article info
Received 26 th April 2022 Received in revised form 28 May 2022 Accepted 9 July 2022

Keywords
CNN,LSTM,image captioning, deep learning.

Abstract
The development of effective techniques has been an urgent demand in the field of the cyber security community. Machine learning for cyber security has become an issue of great importance recently due to the effectiveness of machine learning in cyber security issues. This is typically accomplished by automatically collecting information from a variety of systems and network sources, and then analysing the information for possible security problems. Our main goal is that the task of finding attacks is fundamentally different from these other applications, making it significantly harder for the intrusion detection community to employ machine learning effectively.

1. INTRODUCTION

Cyber-crime is proliferating everywhere exploiting every kind of vulnerability to the computing environment. Ethical Hackers pay more attention towards assessing vulnerabilities and recommending mitigation methodologies. The development of effective techniques has been an urgent demand in the field of the cyber security community. Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. Machine learning for cyber security has become an issue of great importance recently due to the effectiveness of machine learning in cyber security issues. Contrasted with the past, improvements in PC and correspondence innovations have given broad and propelled changes. The use of innovations gives incredible advantages to people, organizations, and governments, be that as it may, messes come up against them. For instance, the protection of significant data, security of put away information stages, accessibility of information, and so forth. Contingent upon these issues, digital fear-based oppression is one of the most significant issues in this day and age. Digital fear, which made a great deal of issues people and establishments, has arrived at a level that could undermine open and national security by different gatherings, for example, criminal association, proficient people, and digital activists. Along these lines, Intrusion Detection Systems (IDS) have been created to

maintain a strategic distance from digital assaults. Right now, learning the bolster support vector machine (SVM) calculations were utilized to recognize port sweep endeavors dependent on the new CICIDS2017 dataset with 97.80%, 69.79% precision rates were accomplished individually. Rather than SVM we can introduce some other algorithms like the random forest, CNN, ANN where these algorithms can acquire accuracies like SVM – 93.29, CNN – 63.52, Random Forest – 99.93, ANN – 99.11.

LITERATURE SURVEY

The rate of attacks against networked systems has increased melodramatically, and the strategies used by the attackers are continuing to evolve. For example, the privacy of important information, security of stored data platforms, availability of knowledge, etc. Depending on these problems, cyber terrorism is one of the most important issues in today's world. Cyber terror, which caused a lot of problems to individuals and institutions, has reached a level that could threaten public and country security by various groups such as criminal organizations, professional persons, and cyber activists. Intrusion detection is one of the solutions to these attacks. A free and effective approach for designing Intrusion Detection Systems (IDS) is Machine Learning. In this study, deep learning and support vector machine (SVM) algorithms were used to detect port scan attempts based on the new CICIDS2017 dataset Introduction Network Intrusion Detection System (IDS) is a software-based application or a hardware device that is used to identify malicious behavior in the network [1,2]. Based on the detection technique, intrusion detection is classified into anomaly-based and signature-based. IDS developers employ various techniques for intrusion detection. Information security is the process of protecting 12 information from unauthorized access, usage, disclosure, destruction, modification, or damage. The terms "Information security", "computer security" and " information insurance" are often used interchangeably.

EXISTING SYSTEM:

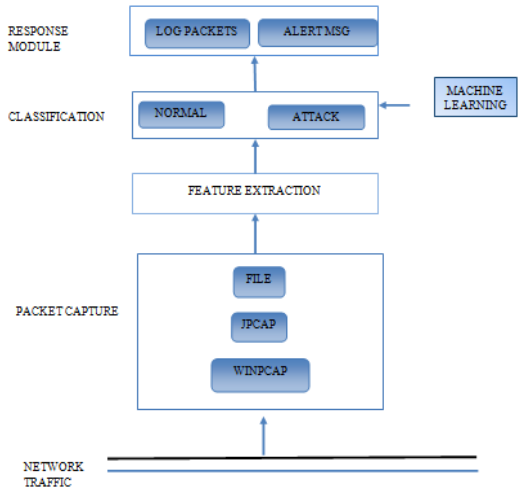
Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. Machine learning for cyber security has become an issue of great importance recently due to the effectiveness of machine learning in cyber security issues.

An IDS will not register these intrusions until they are deeper into the network, which leaves your systems vulnerable until the intrusion is discovered. This is a huge concern as encryption is becoming more prevalent to keep our data secure. One significant issue with an IDS is that they regularly alert you to false positives. In many cases false positives are more frequent than actual threats. If they don't take care to monitor the false positives, real attacks can slip through or be ignored.

When an IDS detects suspicious activity, the violation is typically reported to a security information and event management (SIEM) system where real threats are ultimately determined amid benign traffic abnormalities or other false alarms. However, the longer it takes to distinguish a threat, the more damage can be done. An IDS is immensely helpful for monitoring the network, but their usefulness all depends on what you do with the information that they give you. Because detection tools don't block or resolve potential issues, they are ineffective at adding a layer of security unless you have the right personnel and policy to administer them and act on any threats. An IDS cannot see into encrypted packets, so intruders can use them to slip into the network.

An IDS will not register these intrusions until they are deeper into the network, which leaves your systems vulnerable until the intrusion is discovered. This is a huge concern as encryption is becoming more prevalent to keep our data secure. One significant issue with an IDS is that they regularly alert you to false positives. In many cases false positives are more frequent than actual threats. An IDS can be tuned to reduce the number of false positives; however, your engineers will still have to spend time responding to them. If they don't take care to monitor the false positives, real attacks can slip through or be ignored

SYSTEM ARCHITECTURE:



Network Traffic: Network traffic refers to the amount of data moving across a network at a given point of time. Network data is mostly encapsulated in network packets, which provide the load in the network. Network traffic is the main component for network traffic measurement, network traffic control and simulation. The Proper analysis of network traffic provides the organization with the following benefits: Identifying network bottlenecks - There could be users or applications that consume high amounts of bandwidth, thus constituting a major part of the network traffic. Different solutions can be implemented to tackle these. Network security - Unusual amount of traffic in a network is a possible sign of an attack. Network traffic reports provide valuable insights into preventing such attacks. Network engineering - Knowing the usage levels of the network allows future requirements to be analysed.

Packet capture :Packet Capture is a networking term for intercepting a data packet that is crossing a specific point in a data network. Once a packet is captured in real-time, it is stored for a period of time so that it can be analysed, and then either be downloaded, archived or discarded. Packets are captured and examined to help diagnose and solve network problems such as: Identifying security threats Troubleshooting undesirable network behaviours Identifying network congestion Identifying data/packet loss Forensic network analysis.

Classification :Classification is another extensively used supervisory machine learning task. In cyber security, spam detection is successfully implemented by ML based classifiers which involves discriminating a given email message as spam or not. The spam filter models are able to separate spam messages from non-spam messages. Machine learning techniques for classification include Logistic

Regression, K-Nearest Neighbours, Support Vector Machine, Naïve Bayes, Decision Tree, Random Forest Classification

IMPLEMENTATION DETAILS:

The system is implemented by using ANACONDA software , Anaconda is the world's most popular data science platform and the foundation of modern machine learning

Securing confidential information Algorithms:

Artificial Neural Network (ANN).

The plan thought of an ANN is to mirror the manner in which human cerebrum work. An ANN contains an info layer, a few secret layers and a yield layer. The units in neighbouring layers are completely associated. An ANN contains a colossal number of units and can hypothetically estimated subjective capacities; subsequently, it has solid fitting capacity, particularly for nonlinear capacities. Because of the perplexing model design, preparing ANNs is tedious

Support Vector Machine (SVM).

The system in SVMs is to discover a maximum edge partition hyperplane in the n-measurement highlight space. SVMs can accomplish satisfying outcomes even with limited scope preparing sets in light of the fact that the partition hyperplane is resolved simply by few help vectors. In any case, SVMs are delicate to commotion close the hyperplane.

K-Nearest Neighbour (KNN).

The centre thought of KNN depends on the complex theory. On the off chance that the majority of an example's neighbours have a place with a similar class, the example has a high likelihood of having a place with the class. In this manner, the grouping result issimply identified with the top-k closest neighbours. The boundary k enormously impacts the presentation of KNN models. The more modest k is, the more intricate the model is and the higher the danger of overfitting. On the other hand ,the bigger k is, the easier the model is and the more fragile the fitting capacity.

PROPOSED SYSTEM:

Machine Learning algorithms can be used to train and detect if there has been a cyber attack. As soon as the attack is detected, an email notification can be sent to the security engineers or users. Any classification algorithm can be used to categorize if it is a DoS/DDoS attack or not. One example of a classification algorithm is Support Vector Machine (SVM) which is a supervised learning method that analyses data and recognizes patterns. Since we cannot control when, where or how an attack may come our way, and absolute prevention against these cannot be guaranteed yet, our best shot for now is early detection which will help mitigate the risk of irreparable damage such incidents can cause. Organizations can use existing solutions or build their own to detect cyber attacks at a very early stage to minimize the impact. Any system that requires minimal human intervention would be ideal.

Problem Modelling Network admins The following steps are the functions of Network admin: Intercept network traffic. Read and store the data packets information. Check for alerts regarding the cyber-attacks and network.

SYSTEM ANALYSIS:

a) System: A system is an orderly group of interdependent components linked together according to a plan to achieve a specific objective. Its main characteristics are organization, interaction, interdependence, integration and a central objective.

b) System Analysis: System analysis and design are the application of the system approach to problem solving generally using computers. To reconstruct a system the analyst must consider its elements output and inputs, processors, controls, feedback and environment.

IMPLEMENTATION:

The system is implemented by using ANACONDA software , Anaconda is the world's most popular data science platform and the foundation of modern machine learning.

Securing confidential information Algorithms:

Artificial Neural Network (ANN).

The plan thought of an ANN is to mirror the manner in which human cerebrum work. An ANN contains an info layer, a few secret layers and a yield layer. The units in neighboring layers are completely associated. An ANN contains a colossal number of units and can hypothetically estimated subjective capacities; subsequently, it has solid fitting capacity, particularly for nonlinear capacities. Because of the perplexing model design, preparing ANNs is tedious.

Support Vector Machine (SVM).

The system in SVMs is to discover a maximum edge partition hyperplane in the n-measurement highlight space. SVMs can accomplish satisfying outcomes even with limited scope preparing sets in light of the fact that the partition hyperplane is resolved simply by few help vectors. In any case, SVMs are delicate to commotion close the hyperplane

K-Nearest Neighbor (KNN).

The center thought of KNN depends on the complex theory. On the off chance that the majority of an example's neighbors have a place with a similar class, the example has a high likelihood of having a place with the class. In this manner, the grouping result issimply identified with the top-k closest neighbors. The boundary k enormously impacts the presentation of KNN models. The more modest k is, the more intricate the model is and the higher the danger of overfitting. On the other hand ,the bigger k is, the easier the model is and the more fragile the fitting capacity.

SOFTWARE REQUIREMENTS

Processor : Intel(R)Core (TM)I5

RAM : 2.00GB

System Type : 64Bit Operating System.

HARDWARE REQUIREMENTS:

Python 3.8.3

Visual studio code

Python Django Web framework

Beautiful Soup.

FUNDAMENTALS OF MACHINE LEARNING:

Support vector machine (SVM) is another widely used supervised machine learning model. SVM works to find hyperplane with most suitable dataset distribution by classifying the data into two classes on both sides of the hyperplane. Both sides of the hyperplane donate a separate class. The class of every data point depends on the side of the hyperplane it lands. Support vector machine has a high consumption of space and time to handle larger and noisier datasets . The computational complexity of SVM is O(n2) where n represents the number of instances . A matrix that is used to evaluate the performance of machine learning classifier is called a confusion matrix .

Cyber attack detection techniques fall into two categories: signature-based and inconsistency-based. In both cases, machine learning techniques are used. The authors of improved the detection of Denial-of-Service (DoS) attacks. The Naive Bayes classifier was created based on the element vectors, which included different User Datagram Protocol (UDP) and Transmission Control Protocol (TCP) bundles and their sizes. It has also been shown that Discrete Wavelet Transform and Matching Pursuit may successfully be used to calculate highlights depending on various organisational boundaries.

Error Rate:

The error rate (ERate) is a percentage of the total number of misclassified instances to all instances of the dataset.

ERate = (FPositive + FNegative)/ (TNegative + FPositive + FNegative + TPositive)

Recall:

The recall is a percentage of correctly classified positive instances to the total number of positive instances classified in the dataset.

Recall = TPositive / (TPositive + FNegative)

Precision

The precision is a percentage of the total number of positive instances classified to the total number of positive instances.

Precision= TPositive / (TPositive + FPositive)

MODULES:

Classification

Feature extraction

Detection

Evaluation

METHODOLOGY:

Support vector machine (SVM)

It is a classification method.

In this algorithm, we plot each data item as a point in n-dimensional space (where n is number offeatures you have) with the value of each feature being the value of a particular coordinate.

RESULT:

The experiments were conducted in Machine learning libraries like numpy , pandas, scikit learn. Python language is used to develop the application with jupyter notebook IDE .Predictions can be done by four algorithms like SVM,ANN, RF, CNN this paper helps to identify which algorithm predicts the best accuracy rates which helps to predict best results to identify the cyber attacks happened or not. Fig: 2 Protocol Type Distribution.

CONCLUSION:

At the present time, assessments of help vector machine ,ANN, CNN, Random Forest and significant learning estimations reliant upon current CICIDS2017datasetwere presented moderately. Results show that the significantlearning estimation performed generally best results over SVM, ANN, RFand CNN. We will use port scope attempts just as other attack types with AI and significant learningcomputations, a pache Hadoop and shimmer advancements together ward on this dataset later on. Every one of these estimation assists us with recognizing the digital assault in network. It occurs in the manner that when we think about long back a long time there might be such countless assaults occurred so when these assaults are perceived then the highlights at which esteems these assaults are going on will be put away in some datasets. So by utilizing these datasets we will anticipate if digital assault is finished. These forecasts should be possible by four calculations like SVM,ANN, RF, CNN this paper assists with distinguishing which calculation predicts the best precision rates which assists with foreseeing best outcomes to recognize the digital assaults occurred or not.

REFERENCES:

1. K. Graves, Ceh: Official certified ethical hacker review guide: Exam 312-50. John Wiley & Sons, 2007.
2. R. Christopher, "Port scanning techniques and the defense against them," SANS Institute, 2001.
3. M. Baykara, R. Das¸, and I. Karado gan, "Bilgi g ¨uvenligisistemlerinde kullanilan arac¸ larin incelenmesi," in 1st InternationalSymposium on Digital Forensics and Security (ISDFS13), 2013, pp.231– 239.
4. RashmiT V. "Predicting the System Failures Using Machine LearningAlgorithms".International Journal of Advanced Scientific Innovation,vol. 1, no. 1, Dec. 2020, doi:10.5281/zenodo.4641686.
5. S. Robertson, E. V. Siegel, M. Miller, and S. J. Stolfo, "Surveillance detection in high bandwidth environments," in DARPA Information Survivability Conference and Exposition, 2003. Proceedings, vol. 1.IEEE, 2003, pp. 130– 138.
6. K. Ibrahimi and M. Ouaddane, "Management of intrusion detectionsystems based-kdd99: Analysis with lda and pca," in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1
7. M. Thangamani, and Jafar Ali Ibrahim. S, "Knowledge Exploration in Image Text Data using Data Hiding Scheme," Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2018, 14-16 March, 2018, Hong Kong, pp352-357 http://www.iaeng.org/publication/IMECS2018/IMECS2018_pp352-357.pdf
8. Jeyaselvi, M., M. Sathya, S. Suchitra, S. Jafar Ali Ibrahim, and N. S. Kalyan Chakravarthy. "SVM-Based Cloning and Jamming Attack Detection in IoT Sensor Networks." *Advances in Information Communication Technology and Computing, pp. 461-471. Springer, Singapore, 2022. https://link.springer.com/chapter/10.1007/978-981-19-0619-0_41*