DETECTION OF CYBER ATTACKS IN NETWORK USING MACHINE LEARNING

J VASUDHA[1] ,B. S. MURTHY[2]

1. **Assistant Professor MCA DEPT,** Dantuluri Narayana Raju College, **Bhimavaram, Andhrapradesh**

Email id:- suryanarayanamurthy.b@gmail.com

2. **PG Student of MCA**, Dantuluri Narayana Raju College, **Bhimavaram, Andharapradesh**

Email id:- vasudhajuttiga@gmail.com

ABSTRACT

Improvements in PCs and correspondence have brought about broad changes in comparison to the past. The use of new technologies gives individuals, companies and governments unbelievable benefits, even when they are messing against them. For e.g., security of major data, safety of data stadiums, accessibility of information etc. Digital fear-based oppression is, depending on these questions, one of the biggest problems of our day. Digital fear, which has brought many problems to citizens and organizations, has come to an extent that could threaten the openness and protection of the nation through numerous events, such as criminal groups, professional individuals and digital activists. In this sense, IDS systems were built to maintain a strategic distance from digital attacks. Intrusion detection systems (IDS) At the moment, learning the computation of the SVM (Bolster Support Vector Machine) was used to identify port sweep efforts that depend on a new data set of CICIDS 2017, which achieved 69.79% individually. We should implement other algorithms, including CNN, ANN and Random Forest, rather than SVM, so that accuracy such as SVM – 93.29, CNN - 63.52, Random Forest – 99.93, ANN – 99.11 can be acquired.

1 INTRODUCTION

Improved PCs and correspondence developments have in comparison to the past, contributed to extensive and propelling improvements. The use of novel technologies gives individuals, companies and governments incredible advantages, whether or not they are destroyed. For e.g., the safety of essential data, security of revealed data phases, information accessibility, etc. Digital fear-based oppression, depending on these questions, is nowadays one of the most critical issues. Digital terror, which has posed many problems for people and organizations, has reached a degree that could threaten accessible and security for country, through numerous meetings, such as criminal alliances, professional people and digital activists. In this respect, IDS has been built to maintain a strategic distance from digital assaults. Right now, the measurements of the SVM have been used to classify port sweeping activities based on the latest CICIDS 2017 data set with 97.80%, 69.79% of precise data is carried out individually. Rather than using the SVM, it is possible to join other forest algorithms, such as RF, CNN and ASN, which can be accurate, such as SVM – 93.29, CNN – 63.52, RF – 99.93, ANN - 99.11. Instead of SVM.

1.1 MOTIVATION

The use of new technologies gives individuals, companies and governments unbelievable benefits, even when they are messing against them. For e.g., security of major data, safety of data stadiums, accessibility of information etc. Digital fear-based oppression is, depending on these questions, one of the biggest problems of our day. Digital fear, which has brought many problems to citizens and organizations, has come to an extent that could threaten the openness and protection of the nation through numerous events, such as criminal groups, professional individuals and digital activists. In this sense, IDS systems were built to maintain a strategic distance from digital attacks. Intrusion detection systems (IDS)

2. LITERATURE SURVEY AND RELATED WORK:

2.1 "Harbor scanning and defense against them," Christopher, SANS Institute, 2001. 2001. - 2001.

Port scanning is one of the most common techniques used in network resource finding for attackers. All systems connected via modem to the LAN or the Internet run on common or unknown ports services. The intruder will see the following information on the target devices by way of port scanning, which services are executed by users, support for anonymous logins and authentication of certain network services. The scan of the port is accomplished by sending a message to each port one by one. The reaction form indicates if the port is used and additional vulnerabilities can be tested. Port scanners are important for network security technicians because the target system may detect possible security vulnerabilities. Since port scans can be performed against your programs, port scans can be detected and open services information limited by the right tools. Each publicly available computer has and can operate open ports. The goal is to limit the visibility and refusal of approved users of open ports to closed ports.

2.2. J. A. Hoagland, J. A. Staniford. - Staniford. M. McAlerney, Computer Security Journal, Vol. "Practical automated port detection," 10, 1-2, 118-24, 2002. 2002. 2002

Port scanning is a common method that is very important. Software attackers often use it to classify hosts or networks they are opposed to. This makes it preliminary to classify port scans for more serious attacks useful for system administrators and other network advocates. Network defenders also use their own networks to take into account and find vulnerabilities. Accordingly, attackers must determine whether or not network advocates scan the network regularly. But Defenders don't normally want to mask their ports' scanning, even if attackers. We'll certainly speak in the rest of this article about those attackers who search the network and supporters who attempt to check. Online mailing lists and newsgroups are ongoing the legal/ethical debate of port scanning. It is necessary to scan remote network port itself, without the owners' consent, as a legal and ethical activity.

This is actually a grey field in most jurisdictions. But our experience in monitoring unwanted remote scanning is that virtually all of them come from endangered, hostile host systems. Therefore, it is fair to regard a port at least as aggressive and to warn the remote network administrators it came from. However, this paper focuses on the technical issues of identifying port scans which rely on how important and how you want to respond. In the case of an intrusion detector through the network, we are also concerned (NIDS).

However, we aim to take into account certain of the most obvious ways that an attacker might use to avoid detections, taking an approach that is practical with busy networks. In the remaining section, we first explain port scanning and provide examples of how attackers try to be stealthy. A number of earlier works on ports identification are discussed in the following section. Next, we present our algorithms to clarify our approach with some preliminary information. Lastly, we consider possible extensions of this work along with other applications which may be considered.

We believe that readers are familiar with Internet protocols, basic ideas for network intrusion detection or digital analysis, and the basic theory of probability, theory of information and the linear algebra. There are two general objectives for an intruder while performing a port scan: primary and secondary. The main objective is to collect access and status information for these IP addresses and port combinations (either TCP or UDP). The second goal is to provide alerts to flood-intrusion detection systems to distract or deter network advocates. This paper mainly collects information that collects the port scans, as it is simple to recognize flood port cans (thus, ICMP scans are not discussed directly in this paper, but ideas can obviously be applied to this matter). But it is critical for us to be maliciously overflowing with knowledge in our algorithm design.

We will use the term scan base print for the Port/IP combination set to be specified by the attacker. It is useful to differentiate the footprint of the scan from the document in which the attacker is looking at the footprint. It's referred to the time sequence. The sample is irrespective of the script's aspects, including the rapidity, the randomness, etc. The footprint represents the information gathering requirements for an intruder and creates a search script that complies with these requirements and maybe other non-information gathering requirements (such as not being detected by a NIDS). The most popular type of a port scan footprint is currently a horizontal scan. This means that an intruder is interested in discovering some hosts that reveal the service and have an advantage for a specific service. It then scans the port of interest for all IP addresses in a number of interests. Especially in TCP port 53 sequence this is also done today (DNS)

2.2 M. And C. And that. Take a trip. M.A. Rabbani (2016, p. 5 Hybrid vector support analysis and design component analysis of identities), 'International IEEE Communications and electronics systems conference'

A universal critical problem has emerged that affects individuals, firms or governments compared with the previous networked systems security. Networked networks have been targeted melodramatically, and the techniques of attackers are still evolving. For instance, data security, knowledge availability etc. are important information. important information. Based on these issues, cyber terror is one of the most important subjects in the world today. Cyber-terror has reached a degree in which numerous entities, including criminal organizations, professional organizations, and Internet activists, could place the public and security of the country at risk causing citizens and institutions great problems.

One of the remedies is the detection of intruders. The free and productive approach to IDS development is machine learning. The aim of this survey was to identify profound learning and support for the port-based application or hardware Vector Machine (SVM), which is used in a network to detect malignant behavior, through the new CICIDS2017 intrusion detection system (IDS). Intrusion detection is the technique of detection anomaly-based and signature-based. IDS developers are using intrusion detection techniques. Information protection measures are designed to protect information from unauthorized access, usage, touch, deterioration or harm. The words "information security," "data security" and "information insurance" are interdependent as well These topics serve similar purposes to include access to intelligence, confidentiality and integrity. Studies suggest that the first step in the attack is discovery.

Awareness is made at this point to obtain system information. Finding an open server ports list gives very useful information to an intruder. That's why several resources, such as antivirus and IDS, are available for recognizing open ports. One of these approaches is machine learning. Machine learning technology (ML) can anticipate and identify threats before major security incidents. The grouping of binary instances in two classes is called the classification. On the other hand, multi-class classification refers to classifying instances into three or more classes. In this study, the confidentiality of information in both classifications shall be covered against unauthorized access, use, disclosure, killing, modification or harm. The words "information security," "data security" and "information insurance" are interdependent as well These topics serve similar purposes to include access to intelligence, confidentiality and integrity. Studies suggest that the first step in the attack is discovery. Awareness is made at this point to obtain system information.

3 EXISTING SYSTEM

The Almansob and Lomte KDD99 dataset was used in Blamless Bayes and Principal Component Analysis (PCA). Similarly, Chithik and Rabbani have also used PCA, SVM and KDD99 for IDS. In Aljavarskaya et al. The papers, their evaluations and examinations have been transmitted based on the NSL-KDD data set for their IDS model Composite inspectorates indicate that the KDD99 dataset is used continuously for IDS. KDD99 is therefore older and provides little knowledge about cutting-edge new forms of attack, for example, multi-day misuse etc. In our investigation, we used an innovative and new dataset from CICIDS2017.

4 PROPOSED WORK AND ALGORITHM

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

5. METHODOLOGIES

MODULES

The algorithm

- **ANN**
- **CNN**
- **Woodland Random**

Applications

This technique was used to detect cyber-attack using machine learning technologies on the networK

DATA PREPROCESSING

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So for this, we use data preprocessing task.

DATA EDI

Electronic Data Interchange (EDI) is a computer-to-computer exchange of business documents in a standard electronic format between two or more trading partners. It enables companies to exchange information electronically in a structured format, eliminating the need for manual data entry and reducing the cost and time associated with paper-based transactions.

Model building

Building a model in machine learning is creating a mathematical representation by generalizing and learning from training data. Then, the built machine learning model is applied to new data to make predictions and obtain results.

ML Deploy

Deploying a machine learning model, known as model deployment, simply means to integrate a machine learning model and integrate it into an existing production environment where it can take in an input and return an output. The purpose of deploying your model is so that you can make the predictions from a trained ML model available to others, whether that be users, management, or other systems. Model deployment is closely related to ML systems architecture, which refers to the arrangement and interactions of software components within a system to achieve a predefined goal

Network Intrusion Detection System

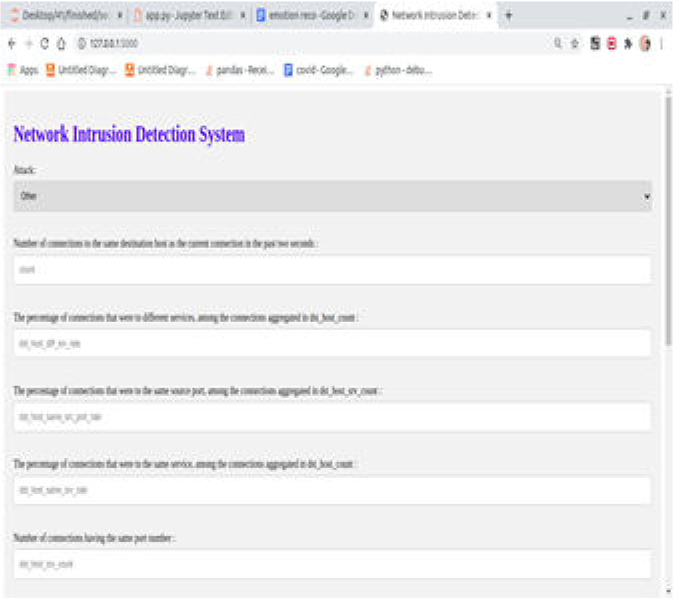
An Intrusion Detection System (IDS) is a network security technology originally built for detecting vulnerability exploits against a target application or computer. The IDS is also a listen-only device. The IDS monitors traffic and reports results to an administrator.

Predict Attack

Attack graphs depict ways in which a hacker can exploit any given vulnerability in a network. Researchers can use these graphs to identify nearly all possible weaknesses in a given system.

Predictive analytics in cybersecurity is a rapidly evolving field, and it's centered around leveraging statistical techniques and machine learning algorithms to predict potential threats and security incidents before they happen. Prediction as a classification problem, Networking sectors have to predict the type of Network attack. from given dataset using machine learning techniques. The analysis of dataset by supervised machine. learning technique (SMLT) to capture several information's.

6 RESULTS AND DISCUSSION



Network Intrusion Detection System

Attack:

Other

Number of connections to the same destination host as the current connection in the past two seconds :

count

The percentage of connections that were to different services, among the connections aggregated in dst_host_count :

dst_host_diff_srv_rate

The percentage of connections that were to the same source port, among the connections aggregated in dst_host_srv_count :

dst_host_same_src_port_rate

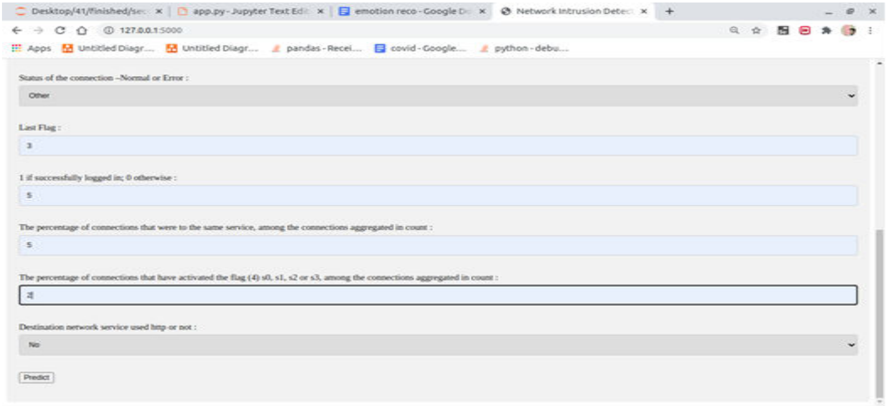
The percentage of connections that were to the same service, among the connections aggregated in dst_host_count :

dst_host_same_srv_rate

Number of connections having the same port number :

dst_host_srv_count

FIG1 : NETWORK IDS



Status of the connection –Normal or Error :

Other

Last Flag :

3

1 if successfully logged in; 0 otherwise :

5

The percentage of connections that were to the same service, among the connections aggregated in count :

5

The percentage of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in count :

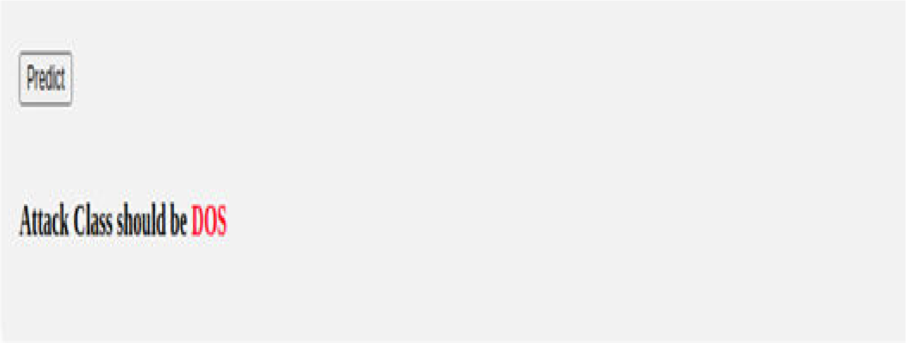
2

Destination network service used http or not :

No

Predict

FIG 2 INPUT SCREEN



Predict

Attack Class should be DOS

FIG3 PREDICT ATTACK

6.CONCLUSION AND FUTURE SCOPE

CONCLUSION

Right now, vector aid estimates, ANN, CNN, Random Forest, and deep learnings based on modern CICIDS2017 dataset have been relatively added. The findings show that the in-depth estimation of learning has obtained essentially better results than SVM, ANN, RF and CNN. With AI and in-depth learning calculations, Apache Hadoop and sparkling inventions, we will make common use of port sweep efforts as well as other assault forms that rely on that dataset. All of this allows us to identify the network cyber threat. It occurs in a way that when we consider the many attacks that occurred over a long period of time, the features of these attacks are preserved in those datasets if they are remembered. We will also predict whether cyber attack is conducted or not using these datasets. This document aims to assess the best prediction algorithms to avoid the best outcomes of cyber attacks. This article can be found in four algorithms including SVM, ANN, RF, CNN

FUTURE SCOPE:

We add some ML algorithms for improvement to improve the precision..

7 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 ¨uvenli ˘gi sistemlerinde kullanilan arac¸larin incelenmesi,” in 1st International Symposium on Digital Forensics and Security (ISDFS13), 2013, pp. 231–239.

[4] S. Staniford, J. A. Hoagland, and J. M. McAlerney, “Practical automated detection of stealthy portscans,” Journal of Computer Security, vol. 10, no. 1-2, pp. 105–136, 2002.

[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 detection systems based-kdd99: Analysis with lda and pca,” in Wireless Networks and Mobile Communications (WINCOM), 2017 International Conference on. IEEE, 2017, pp. 1–6.

[7] N. Moustafa and J. Slay, “The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems,” in Building Analysis Datasets and Gathering

Experience Returns for Security (BADGERS), 2015 4th International Workshop on. IEEE, 2015, pp. 25–31.

[8] L. Sun, T. Anthony, H. Z. Xia, J. Chen, X. Huang, and Y. Zhang, “Detection and classification of malicious patterns in network traffic using benford’s law,” in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017. IEEE, 2017, pp. 864–872.