International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 16th Dec 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=Issue 12

DOI: 10.48047/IJIEMR/V11/ISSUE 12/28

Title **DETECTION OF CYBER ATTACK IN NETWORK USING MACHINE LEARNING TECHNIQUES**

Volume 11, ISSUE 12, Pages: 208-216

Paper Authors
Mohammad Reshma, Dr.V.Uma Rani

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

DETECTION OF CYBER ATTACK IN NETWORK USING MACHINE LEARNING TECHNIQUES

1 Mohammad Reshma, M.tech in software engineering (SE) SIT JNTUH

2 Dr.V.Uma Rani, Professor of CSE

ABSTRACT: 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. Machine learning techniques have been applied for major challenges in cyber security issues like intrusion detection, malware classification and detection, spam detection and phishing detection. Although machine learning cannot automate a complete cyber security system, it helps to identify cyber security threats more efficiently than other software-oriented methodologies, and thus reduces the burden on security analysts. Hence, efficient adaptive methods like various techniques of machine learning can result in higher detection rates, lower false alarm rates and reasonable computation and communication costs. 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.

Keywords: *cyber security, detection, machine learning.*

1. INTRODUCTION

Email and personal computers have both advanced significantly since their inception. Despite this, new inventions have a positive impact on people, organisations, and governments while also providing difficulties for those trying to stop them. Among the concerns are issues with information security, stockpiling and recovery security, and data availability. Contingent upon how these issues create, computerized mistreatment persuaded by dread is a critical issue in the cutting edge world. Numerous individuals and organisations have suffered from the harmful effects of digital anxiety, which has led to a number of issues. Different parties, including criminal organisations, knowledgeable individuals, and online activists, could be sufficiently alarmed to compromise openness and national security. IDS was created in part to protect against these kinds of online attacks. Support vector machine (SVM) computations that are 97% exact and 69% accurate may now be used to determine port sweep attempts using the new CICIDS2017 dataset. SVM (93.29) is inferior to approaches like Random Forest (99.93).

In any case, albeit such advancements enormously benefit people, gatherings, and state run administrations, they likewise give difficulties to the

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

individuals who go against them. Basic information security, insurances for data capacity and recovery, and data openness are a couple of them. Depending on how these concerns develop, digital oppression based on fear is a significant issue in the modern world. Digital dread has affected a large number of people and organisations and has resulted in various problems. Different parties, including criminal organisations, knowledgeable individuals, and online activists, could be sufficiently alarmed to compromise openness and national security. IDS was created in part to protect against these kinds of online attacks. The KDD99 dataset's data were analysed using PCA and Bayesian inference [9]. PCA, SVM, and KDD99 were likewise utilized by Chithik and Rabbani for IDS [10]. Aljawarneh et al. guarantee that their IDS model's results depended on the NSL-KDD dataset. Examinations of the organization show that the KDD99 dataset is regularly used for IDS [6-10]. It incorporates 41 features and was made in 1999. KDD99 is thusly obsolete and doesn't contain data on present day new attack sorts, like maltreatments that keep going for a long time. We used the brand-new CICIDS2017 dataset [12] as a result for our study.

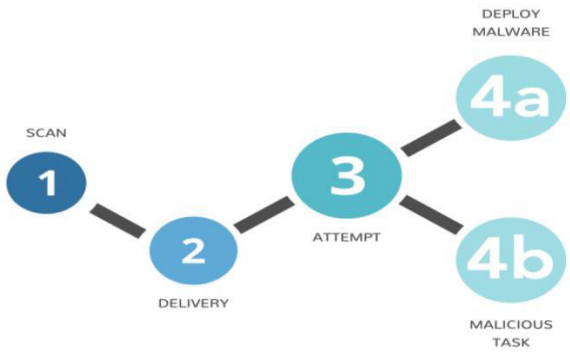

Fig.1: Early stage detection of cyber attacks.

SVM and Random Forest are just a few examples of machine learning algorithms that will be employed in this research to identify cyberattacks. These forecasts may be made using two algorithms, including SVM and RF. With the help of this study, we can determine which algorithm is most accurate in determining if a cyberattack has taken place. Adopting machine learning techniques, cyberattacks in networks using this tactic are discovered.

2. EXISTING SYSTEM

Almansob and Lomte utilized Principal Component Analysis(PCA) on the KDD99 dataset[9]. PCA, SVM, and KDD99 were additionally utilized by Chithik and Rabbani for IDS [10]. Aljawarneh et al. guarantee that their IDS model's results depended on the NSL-KDD dataset. Assessments of the arrangement show that the KDD99 dataset is regularly used for IDS [6-10]. The 1999 film KDD99 features 41 memorable moments. As a result, KDD99 is out of current and is missing information on modern new assault kinds, such as abuses that last more than one day. We used the brand-new CICIDS2017 dataset [12] as a result for our study.

Drawbacks:

1. Putting strict restrictions into place
2. Resource-restrictive;
3. Difficult for non-technical people to use
4. Needs Continuous Patching
5. Constantly targeted for attacks

3. PROPOSED SYSTEM

The calculation's fundamental stages are portrayed in the segments that follow. Each dataset must be

standardized. Utilizing the dataset you have arranged, form the testing and preparing datasets. Make IDS models utilizing the RF and SVM calculations. Assess each model's general presentation.

Steps:-

1. Normalization of every dataset.
2. Convert that dataset into the testing and training.
3. Form IDS models with the help of using RF and SVMalgorithms.
4. Evaluate every model's performances
5. After Evaluation, it use better performed model to detect attacks based on inputs by using logistic regression.

Advantages:-

1. Defense against malicious actors attacking your network.
2. Prevents people from getting access to the network in an unauthorised manner.
3. Maintaining the confidentiality of sensitive data

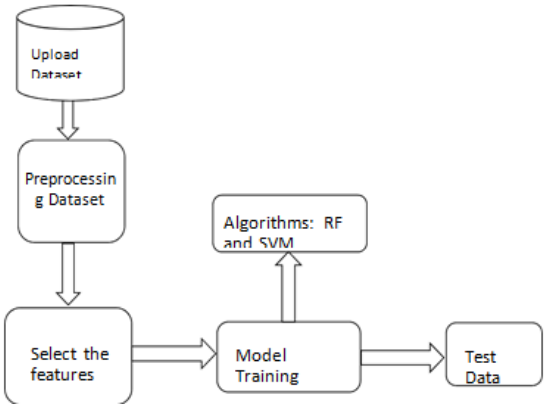

Fig : System architecture

4. LITERATURE REVIEW

4.1 Port scanning techniques and the defense against them.

Before starting an attack, cybercriminals frequently employ port scanning to find vulnerabilities in computer systems. You'll frequently encounter services listening on both well-known and unknown ports while utilising a modem or local area network. An attacker can use port scanning to find out what services are active on a target system, who is the owner of those services, and if anonymous logins are permitted. Sending a message to each port individually allows for port scanning. You can decide if a port is being used and whether it could be additionally tested for security openings and weaknesses by taking a gander at the kind of answer you get. Network security experts employ port scanners because they can find security weaknesses on the targeted system. By employing the proper tools and doing port scans on your systems, you may find and restrict the amount of information that is

provided by open services. Every system that is exposed to the public has ports that are available and open. The goal is to prevent unauthorised users from utilising locked ports and to limit access to open ports to just those who are authorised.

4.2 Practical automated detection of stealthy portscans:

Port scanning is a routine yet important task. It is a word that computer attackers commonly use to describe targets or networks where they believe hostile behaviour is occurring. It is therefore very helpful for system administrators and other network security workers to be able to recognise portscans as potential precursory assaults. Network defenders make considerable use of it in order to better comprehend and discover vulnerabilities in their own networks. As a result, attackers closely monitor whether a network's defences consistently do port scans. In contrast, defenders are less inclined than attackers to seek to conceal their port monitoring. It is essential to keep in mind that this article will refer to attackers scanning the network and defenses attempting to stop the scan. To be more clear, this is done. On mailing lists and newsgroups on the Internet, arguments over the ethics and law of port scanning regularly emerge.

Without the owners' consent, port scanning of distant networks raises the issue of whether the practise is morally and legally acceptable. Most jurisdictions are presently unclear about how to address this. Our experience has revealed that virtually all unsolicited remote port scans originate from hacked systems, making them very likely to be hostile. Therefore, a portscan should be considered potentially hostile and reported to the network administrators in charge of maintaining the security of that distant network. This article examines the technical issues of how to detect portscans, independent of how significant one believes they are or how one responds to them, rather than focusing on the ramifications of portscans. Another important problem of ours is detecting a portscan using a network intrusion detection system (NIDS). Although we make an attempt to take into account some of the more blatant techniques an attacker may employ to escape detection, we choose to stick with an approach that is simple to apply in networks with a lot of traffic.

4.3 Combined analysis of support vector machine and principle component analysis for ids:

People from many walks of life, including corporations, governments, and the general public, are increasingly highly concerned about secure networked systems. Lately, there have been much more endeavors to go after organized frameworks, and assailant strategies are changing simultaneously. Consider, for example, the availability of information, the protection and security of touchy information, and the security of information stockpiling frameworks. These issues have made digital psychological oppression perhaps of the most critical test in the cutting edge world. It's time to act since cyberterrorism is now at a dangerously high level, putting public and national security at risk from organisations like criminal gangs, professionals, and activists. One method for preventing assaults of this nature is intrusion detection. It is easy to create intrusion detection systems using machine learning (IDS). In order to identify a port scan attempt, this

study employed support vector machine (SVM) and deep learning methods. Introduction A software-based or hardware-based intrusion detection system (IDS) can be used to identify malicious activities on a network. There are two forms of intrusion detection: anomaly-based and signature-based. IDS programmers employ a range of ways to find intrusions. When it comes to information security, your goal is to prevent unauthorised access to and improper use of the data. While talking about data security, a few terms — including "PC security" and "data protection" — are utilized reciprocally. These segments are connected and endeavor toward similar goals to offer availability, secrecy, and trustworthiness of data. Research shows that the primary stage in an attack is disclosure. Observation is currently being done to see more about the framework.

4.4 Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model:

A safe network environment needs a reliable intrusion detection system. The accuracy of intrusion detection will change with different feature selection techniques. In a number of intrusion detection situations, the individual feature selection technique might be inaccurate. In order to fine-tune feature subsets in an ensemble environment, feature selection is used in this research. To decide if an element is fundamental or superfluous, the determination system is changed into a two-class issue with an odd number of component choice methods. True tasks utilize procedures like the mean abatement pollutant, arbitrary timberland classifiers, and steadiness choice calculations. We will adapt using our new technique till we obtain ensemble feature subsets from their feature subsets. There are a few procedures for evaluating the viability of outfit highlight subsets, including support vector machines, choice trees, knn, and 3033multi-facet insight. In addition to the KDD Cup 99 data, the UNSW NB 15 and CIDDS 001 intrusion detection datasets are also employed in this study. The most accurate outcome is cidds-001, with a classification accuracy of 99.40 percent. The study's results show that our method increases the precision of intrusion detection categorization. There are a number of gaps and problems in the eleven publicly accessible IDS datasets produced since 1998 that highlight the genuine need for a thorough and accurate dataset. Advanced Military Studies and Technology Agency (Lincoln Laboratory, 1998-1999) The dataset, which was developed for network security research, showed problems brought on by malicious and unnecessary traffic that was purposefully put into the network. Email, surfing, FTP, Telnet, IRC, and SNMP activity are all included in this dataset. Denial of Service (DoS), password guessing (PW), buffer overflows (BO), remote FTP (RF), Syn flooding (SF), and rootkits are just a few of the risks it contains (R). This dataset, which does not reflect actual network traffic, has anomalies including the absence of false positives. Because of this, the dataset that is currently available cannot be used to assess IDSs on contemporary networks.

4.5 Toward generating a new intrusion detection dataset and intrusion traffic characterization:

As created programmes and computer networks expand tremendously in size, it is becoming more and

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

more clear that launching assaults might inflict enormous harm. IDS and IPS systems are crucial in preventing network assaults as they grow more complex. Anomaly-based techniques in intrusion detection systems are constrained by erroneous dataset deployment, analysis, and assessment. Various intrusion detection and prevention techniques have been put to the test utilising datasets like DARPA98, KDD99, ISC2012, and ADFA13, among others. Using eleven datasets going back to 1998, we discovered that many of the publicly accessible datasets are outdated or otherwise untrustworthy. In certain instances, the data isn't sufficiently vast or varied to account for all potential threats. Others contain packet and payload data that has been anonymised, which does not accurately represent recent patterns. We had the option to create a dependable dataset utilizing this study that is transparently accessible to people in general and contains both harmless and seven regular assault network streams. The scientists break down their discoveries in this distribution to distinguish which organization traffic qualities and machine learning techniques are best at recognizing specific kinds of attacks.

5. ALGORITHMS

Support Vector Machine (SVM):

The SVM is a typical supervised machine learning model utilized in classification. Given a two-class training sample, the objective of a support vector machine is to select the optimal hyperplane with the highest margin of separation between the two classes. For better generalization, the hyperplane shouldn't be closer to data points from the other class. Choose the hyperplane that is the farthest away from the data points in each category. The points closest to the classifier's margin are the support vectors. Utilizing the WEKA interface, the investigation's exactness is evaluated. The SVM determines the ideal separating hyperplane by maximizing the distance between the two decision boundaries.

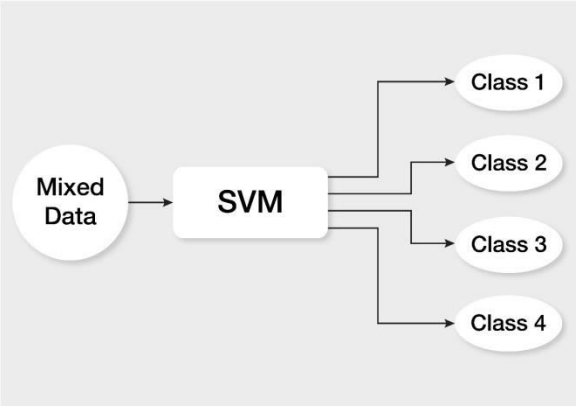

Fig.3: SVM model

Random Forest:

In classification and regression problems, supervised machine learning algorithms like random forest are frequently used. It builds decision trees using a variety of samples, using the average for classification and the majority vote for regression. The Random Forest approach can be utilized for both classification and regression tasks. Accuracy is improved by using cross validation. While the random forest classifier handles the missing values, a significant portion of the data will remain accurate.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

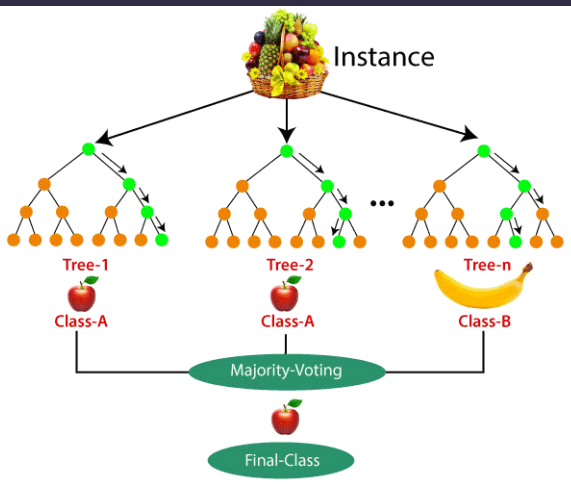

Fig.5: Random forest model

Logistic Regression:

It is a machine learning classification algorithm is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data code as 1 (yes, success etc) or 0 (no, failure etc).

It consider the best performed model to predicting the target variable. i.e., Attack has occurred or not.

6. EXPERIMENTAL RESULTS

Datasets description here given below:

duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins	logged_in	num_compromised	root_
0	0	udp	other	SF	146	0	0	0	0	0	0	0	0
1	0	tcp	private	S0	0	0	0	0	0	0	0	0	0
2	0	tcp	http	SF	232	8153	0	0	0	0	0	1	0
3	0	tcp	http	SF	199	420	0	0	0	0	0	1	0
4	0	tcp	private	REJ	0	0	0	0	0	0	0	0	0

Fig.1: Dataset

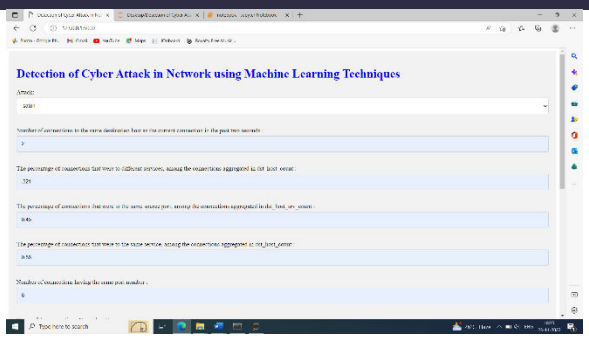

Fig.2: Main page

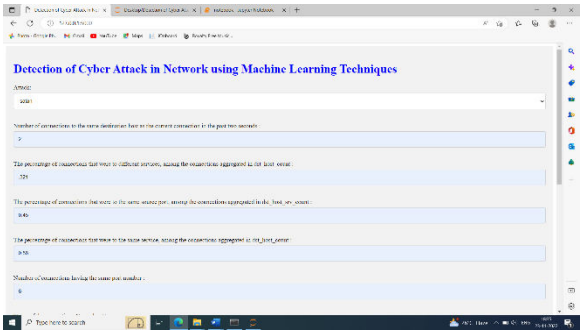

Fig.3: User input screen

Fig.4: Prediction result

7. CONCLUSION

This paper presents an automated solution to one of the biggest challenges in networks, which may leading to the biggest data loss for individual as well as company Our main goal is to detecting cyber-attacks accurately in network using Machine

Learning techniques, algorithms Random forest and support vector machine to construct a model and Recursive feature elimination process to select best feature to predict the target variable. the prediction methodology using the pre- processed dataset and use the most accurate and appropriate ML techniques.

8. FUTURE SCOPE

In enhancement better to add some machine learning methods to increase accuracy.

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 ˘gisistemlerindekullanilanarac¸larinincelenmesi," 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.

[9] S. M. Almansob and S. S. Lomte, "Addressing challenges for intrusion detection system using naive bayes and pca algorithm," in Convergence in Technology (I2CT), 2017 2nd International Conference for. IEEE, 2017, pp. 565–568.

[10] M. C. Raja and M. M. A. Rabbani, "Combined analysis of support vector machine and principle component analysis for ids," in IEEE International Conference on Communication and Electronics Systems, 2016, pp. 1–5.

[11] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," Journal of Computational Science, vol. 25, pp. 152–160, 2018.

[12] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization." in ICISSP, 2018, pp. 108–116.

[13] D. Aksu, S. Ustebay, M. A. Aydin, and T. Atmaca, "Intrusion detection with comparative analysis of supervised learning techniques and fisher score feature selection algorithm," in International Symposium on Computer and Information Sciences. Springer, 2018, pp. 141–149.

[14] N. Marir, H. Wang, G. Feng, B. Li, and M. Jia, "Distributed abnormal behavior detection approach based on deep belief network and ensemble svm using spark," IEEE Access, 2018.

[15] P. A. A. Resende and A. C. Drummond, "Adaptive anomaly-based intrusion detection system using genetic algorithm and profiling," Security and Privacy, vol. 1, no. 4, p. e36, 2018.

[16] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[17] R. Shouval, O. Bondi, H. Mishan, A. Shimoni, R. Unger, and A. Nagler, "Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in sct," Bone marrow transplantation, vol. 49, no. 3, p. 332, 2014.