International Journal of Advanced Technology in Engineering and Science
Vol. No. 10, Issue No. 01, January 2022
www.ijates.com
ijates
ISSN 2348 - 7550

A Study Of intrusion Detection System Using Machine Learning Algorithms

P.Santhiya[1],M.Arun[2], D.Kalaiabirami[3]

[1,2,3]Assistant Professor, Department of CSE,

Builders Engineering College,Kangayam, TiruppurDt, Tamilnadu, India.

ABSTRACT

The massive number of people connecting to a network is increasing now a days, the network security is also becoming more important. With the development of technology, the range of human beings use internet for storing or gaining access to essential statistics has increased. Hence, the need to secure the data integrity, and confidentiality also increased. Cyber Security practices is much more important to protect our data from the cyber attackers who steel the information from us, and use it for harm. Among these detecting of Cyber-attacks is more complex. Intrusion Detection System (IDS) is a cyber-security technique, monitors the state of software and hardware running in the network. An IDS is used to analyse, protect the system and predict the behaviour of the system. Intrusion Detecting System still faces the challenges in improving the detection accuracy. To solve this problem, machine learning algorithms are used to detect the intrusion. This paper discussed various ML algorithms which can predict the intrusion in the network.

Keywords: Intrusion Detection, cyber security, Firewall, machine learning algorithmssignature-based detection, anomaly based detection, Malware Detection, Support Vector Machines.

I. INTRODUCTION

In computer world, the rapid growth of the Internet, which increases cyber-attacks not only in numbers but also in diversity: ransomware are on the rise like never before, and zero day attack exploits become so critical that they are gaining media coverage. A cyber-attack (or an intrusion) is defined as all unauthorized activities that compromise one, two, or all of these three fundamental components of an information system i.e. Confidentiality, integrity, and availability (also known as the CIA triad)[1]. Antiviruses and firewalls are no longer sufficient to ensure the protection of a particular network, which should be based on multiple layers of security.Intrusion detection system (IDS) is a security technique attempting to detect various attacks. They are the set of techniques that are used to detect suspicious activity both on host and network level.So, Intrusion Detection System is used to detect all types of malicious network traffic andcomputer usage that can't be detected by a conventional firewall. In order to detect unwanted activity and events such as illegal and malicious

traffic, traffic that violates security policy, and traffic that violates acceptable use policies Intrusion Detection System is used.

An Intrusion Detection System(IDS) is a piece of installed software or a physical appliance that monitors network traffic. Many IDS tools is used to store a detected event in a log to be reviewed at a later date or will combine events with other data to make decisions regarding policies or damage control. An IPS is a type of IDS that can prevent or stop unwanted traffic.

The Intrusion Detection System classified into two major categories: (i) signature-based detection (or "misuse detection") and (ii) anomaly based detection. The signature-based detection very effective and reliable but only known attacks can be detected that have already exist in a database. But in anomaly detection, unknown attacks can be detected by building a model of normal behaviour of the system and then looks for deviations in the monitored data but it often generates an overwhelming number of false alarms.

IDS detect intrusions in different places. 1. Network intrusion detection (NIDS) is a strategically placed (single or multiple locations) system to monitor all the network traffic. 2. Host intrusion detection (HIDS) runs on all devices in the network which is connected to the internet/intranet of the organization. They can detect malicious traffic which originates from within (for example, when malware is trying to spread to other systems from a host in the organization).

Based on their action IDS can be classified into two types: one is Active, other is Passive. (i) Active - it is also known as an intrusion detection and prevention system which generates alerts and logs entries along with commands to change the configuration to protect the network. (ii) Passive, it just detects malicious activity and generates an alert or logs, but it doesn't take any action.

Some of the Intrusion Detection functions are listed below: 1.)Monitoring and analysing both user and system activities.2.)Analysing system configurations and vulnerabilities. 3.)Assessing system and file integrity.4.)Ability to recognize patterns typical of attacks.5.)Analysis of abnormal activity patterns.6.)Tracking user policy violations.

Some of the Limitation of traditional IDS are several real attacks are far less than the number of false alarms raised which leads to unnoticed of real threats.Constant software updates are required for signature-based IDS to keep up with the new threats. IDS monitor the whole network, so are vulnerable to the same attacks the network's hosts are. Protocol-based attacks can cause the IDS to fail. Network IDS can only detect network anomalies which limit the variety of attacks it can discover.

II. MACHINE LEARNING IN IDS

In the Intrusion Detection System, Machine Learning is one of the technique used to detect attacks which is concerned with the design and development of algorithms and methods that allow computer systems to autonomously acquire and integrate knowledge to continuously to improve their efficiently and effectively. It also gives high accuracy and good detection where their calculations are driven by science and insights. These

calculations used to find their examples, relationships, and peculiarities in the given datasets.The data set contains a collection of data instances each of which can be described using a set of attributes (features) and the associated labels. The attributes can be of different types such as categorical or continuous. The labels associated with data instances are usually in the form of binary values i.e. normal and anomalous.By applying machine learning techniques for intrusion detection, it automatically build the model based on the training data set.

III. LITERATURE SURVEY

"Statistical analysis of CIDDS-001 dataset for Network Intrusion Detection Systems using Distance-based Machine Learning", [1] discussed the statistical analysis by evaluating the labelled flow based CIDDS-001dataset which used for Anomaly based Network Intrusion Detection Systems. Here, two ML algorithms used for evaluation, they are k-nearest neighbour classification and k-means clustering which used to measure the complexity in terms of prominent metrics. As a result, the both k-nearest neighbour classification and k-means clustering perform well over CIDDS-001 dataset by using prominent metrics. "Automated multi-level malware detection system based on reconstructed semantic view of executables using machine learning techniques at VMM", In[2] AMMDS extensively reduces the manual effortand accurately identify the malware from thesemantically reconstructed and forensicallyextracted executables as compared to other existingVMI and MFA based out-of-VM approaches.Finally, the AMMDS used to measure the malwaredetection rate with an accuracy of 100% by evaluating a largenumber of real-world Windows malware.

"Advanced Machine Language Approach to Detect DDoS Attack Using DBSCANClustering Technology with Entropy", [4]proposedan advanced machine learning approach on cloudcomputing environment to detectthe distributed denial of services attacks with entropy usingclustering technology. For implementing DDoShybrid detection schemes more additional regressive testing needed to perform at bothvulnerable side of the cloud computingenvironment (the network and host level) can be done. "A Novel algorithm for Network Anomaly Detection using Adaptive MachineLearning", [5] proposed a method by using the labelled dataset for training but canadapt/learn itself and can detect new attacks. Normally, a modelis built with the existing data and the system istrained which helps to detectintrusions. The main issue in this is the network traffic changes overtime, for such cases the system should be trainedautomatically or retrained. Still theperformance of the algorithm to beimproved by combining this algorithm with featureweights. "Anomaly detection in wide area network mesh using two machine learning anomaly detection algorithms", [6] explains the two new techniques for distinguishingsystem execution in view of split-examplecharacterization: AdaBoost and Simple feedforwardneural system. Both the methods are used to simulated datasets to check their sensitivity withrespect to duration and amplitude of anomaly. Thedecision tree method is used to be very fast(4seconds evaluation per one hour of data tested)and detected all the simulated anomalies.

"A Novel Malware Analysis Framework for Malware Detection and Classification using Machine Learning Approach", [7]in this approach only 220 samples of files are taken for analysis which may be biased, because not all the features may have incorporated using these number of samples.Based on similarity in their behaviour

novel intelligent malware analysis framework has been developed for dynamic andstatic analysis of malware samples. By using machine-learning models the malicious files can be detected and classify. As a result, J48 Decision tree shows the best performance in terms of accuracy and precision.

"Wired LAN and Wireless LAN Attack Detection Using Signature Based and Machine Learning Tools", [8]author describe about internal attack which occurs in both WirelessLAN and Wired LAN. Now-a-days there are various Signature based tools,used for detecting these types ofattacks but these are not sufficient due to high falsealarm rate. Subsequently, identify these kinds ofassaults with three routes: through Wireshark, withsignature based apparatuses (Snort and Kismet) andwith machine learning instruments (WEKA). PINGscan or PING flood, NMAP scan (portsweep) andARP spoofing attacks seen in wired LAN attack. In wireless LAN attacks, Deauthentication attack, Disassociationattack and Access point (AP) spoofing attack are carried out.Signature based tools detect these types of theattacks based on the stored signature and timingthreshold. Nevertheless, machine-learning toolstake several different features to detect these typesof attacks with more accuracy and low falsepositive rate."Survey on SDN based network intrusion detection system using machine learning approaches", [10]In this model, Machine Learning (ML) approaches have beenimplemented in the SDN-based Network IntrusionDetection Systems (NIDS) to protect computernetworks to overcome network security issues.SDN Technologyprovides a prospect to detect and monitornetwork security problems ascribing to theemergence of the programmable features. It evaluateddifferent late chips away at machine learning (ML)techniques that use SDN to execute NIDS.Meanwhile, it secured models that can be utilizedto create NIDS models in SDN condition.

"Performance comparison of intrusion detection systems and application of machine learning to Snort system", [14] it investigates themalicious traffic on computer networks by using two opensource intrusion detection systems (IDSs) namelySnort and Suricata.For higherdetection accuracy snort is selected for furtherexperiments. Snortadaptive plug-in and SupportVector Machine (SVM) was selected for empirical study with different learning algorithms. Hybridversion of SVM and Fuzzy logic produced a betterdetection accuracy. Finally, the best result wasachieved using an optimized SVM with fireflyalgorithm with FPR (false positive rate) as 8.6% andFNR (false negative rate) as 2.2%, which is a goodresult."Cyber Security Intrusion Detection for Agriculture 4.0: Machine Learning-Based Solutions, Datasets, and Future Directions", [16] this paper discussed about Agriculture 4.0 cyber security by using intrusion detection system. Evaluate intrusion detection systems along with the emerging technologies like Cloud computing, Fog/Edge computing, Network virtualization, Autonomous tractors, Drones, Internet of things, Industrial agriculture, and Smart Grids. With help of machine learning algorithm a comprehensive classification of intrusion detection systems is used in each emerging technology. Performance evaluation of intrusion detection systems for Agriculture 4.0have been implemented by using public datasets. "Network Intrusion Detection Using Linear and Ensemble ML Modeling"[17]To monitor traffic in the network a network intrusion detection system is created in the network to breach the security by invading foreign entities. Experimental analysis have been performed on the NSL-KDD dataset instead of the KDD dataset because it does not have redundant data so the output produced from classifiers will

not be biased. The four major types of attacks are: denial of service (DoS), probe attack, user to root attack (U2R), remote to local attack (R2L). In this research, an intense study on linear and ensemble models such as logistic regression, stochastic gradient descent (SGD), naïve bayes, light GBM (LGBM), and XGBoost. Lastly, a stacked model is developed which is trained by classifiers, and helps to detect intrusion in networks. By using these methods, maximum accuracy (98.6%) from stacked model have been detected.

"Poligraph: Intrusion-Tolerant and Distributed Fake News Detection System"[18]This paper mainly aims to address architectural, system, technical, and social challenges of building a practical, long-term fake news detection platform. A case study for fake news detection by authors' institute, showing that machine learning-based reviews are less accurate but timely, while human reviews, in particular, experts reviews, are more accurate but time-consuming which leads to combine both approaches. At the core of Poligraph, by combining machine learning techniques and human expert determination the two-layer consensus using Byzantine fault-tolerant (BFT) and asynchronous threshold common coin protocols is constructed. By implementing of Poligraph the performance can be evaluated on Amazon EC2 using a variety of news from online publications and social media. Poligraph achieves throughput of more than 5,000 transactions per second and latency as low as 0.05 second. "A System to automate the development of anomaly-basednetwork intrusion detection model" [21]describes about the anomaly based intrusion detection technique to develop an automation system to both train and test supervised machine learning models , which is developed to classify real time network traffic as to whether it is malicious or not. In this, both detection success rate and the false positives rate are Artificial Neural Networks(ANN) followed by Support Vector Machines(SVM). To evaluate the performance of the system, NSL-KDD dataset is used to train and test the SVM and ANN models and finally classify real time network traffic using these models. This system can be used to carry out model building automatically on the new datasets and also for classifying the behaviour of the provided dataset without having to code.

"Protocol Based Deep Intrusion Detectionfor DoS and DDoS attacks usingUNSW-NB15 and Bot-IoT data-sets", [22] this paper introduce the Protocol Based Deep Intrusion Detection (PB-DID) architecture, where data-set of packets are created from IoT traffic by comparing features from the UNSWNB15 andBot-IoT data-sets based on flow and Transmission Control Protocol (TCP). Here, we differentiated non-anomalous,DoS, and DDoS traffic uniquely by taking care of the problems like imbalanced and over-fitting by using deep learning (DL) technique classification accuracy of 96.3% is achieved."Cyber Intrusion Detection Using MachineLearning Classification Techniques" [23]to detect intrusions in cyber, powerful machinelearning classification algorithms, namely Bayesian Network, Naive Bayes classifier,Decision Tree, Random Decision Forest, Random Tree, Decision Table, andArtificial Neural Network, to provide intelligent servicesin the domain of cyber-security. By using these algorithms, the effectiveness of various experimentson cyber-security datasets having several categories of cyber-attacks can be detected and also evaluate the effectiveness of the performance metrics, precision, recall, f1-score,and accuracy.

IV. CONCLUSION AND FUTURE ENHANCEMENT

This survey paper specifies that how machine learning algorithm useful for intrusion detection is shown with a lot of efficiency. Before diving into complex algorithms and statistical models, take a moment to think carefully about the problem you are trying to solve, and the data available to you. By using advanced machine algorithms, it may be to generate a more complete and descriptive set of input to solve many problems in various field with their accuracy measures. In future case, we can implement this methodology to solve our problems and also detect and classified with more accuracy by using machine algorithms.

REFERENCES

[1] AbhishekVermaa, VirenderRangaa,"Statistical analysis of CIDDS-001 dataset forNetwork Intrusion Detection Systems usingDistance-based Machine Learning", ProcediaComputer Science 125, (2018), 709–716

[2] Ajay Kumara M.A., Jaidhar C.D, "Automatedmulti-level malware detection system based onreconstructed semantic view of executablesusing machine learning techniques at VMM",Future Generation Computer Systems, Vol.79,1(2018), 431-446.

[3] Ana Sima, Kurt Stockinger, KatrinAffolter,Martin Braschler, Peter Monte, Lukas Kaiser, "AHybrid Approach for Alarm Verification usingStream Processing, Machine Learning and TextAnalytics", ACM, (2018),https://doi.org/10.21256/zhaw-3487

[4] AntenehGirma, Mosses Garuba, and RajiniGoel, "Advanced Machine Language Approachto Detect DDoS Attack Using DBSCANClustering Technology with Entropy",Information Technology - New Generations.Advances in Intelligent Systems andComputing, Vol. 558, (2018), DOIhttps://doi.org/10.1007/978-3-319-54978-1_17

[5] Ashok Kumar. D, S. R. Venugopalan, "A Novelalgorithm for Network Anomaly Detectionusing Adaptive Machine learning", AdvancedComputing and Intelligent Engineering.Advances in Intelligent Systems andComputing, Vol. 564, (2018), DOIhttps://doi.org/10.1007/978-981-10-6875-1_7

[6] James Zhang, IlijaVukotic, Robert Gardner,"Anomaly detection in wide area network meshusing two machine learning anomaly detectionalgorithms", Networking and InternetArchitecture, (2018)

[7] KamalakantaSethi, Shankar Kumar Chaudhary,Bata KrishanTripathy, PadmalochanBera, "ANovel Malware Analysis Framework forMalware Detection and Classification usingMachine Learning Approach", ACM, (2018),doi>10.1145/3154273.3154326.

[8] Kaur. J, "Wired LAN and Wireless LAN AttackDetection Using Signature Based and MachineLearning Tools", Networking Communicationand Data Knowledge Engineering, (2018), 15 -24

[9] Liang Xiao, Xiaoyue Wan, Xiaozhen Lu,Yanyong Zhang, Di Wu, "IoT SecurityTechniques Based on Machine Learning",(2018), https://arxiv.org/abs/1801.06275

[10] Nasrin Sultana, Naveen Chilamkurti, WeiPeng, RabeiAlhadad, "Survey on SDN basednetwork intrusion detection system usingmachine learning approaches", Peer to PeerNetworking and Applications, (2018), 1-9.https://doi.org/10.1007/s12083-017-0630-0

[11] Pete Burnap, Richard French, Frederick Turner,Kevin Jones, "Malware classification using self-organising feature maps and machine activitydata", Computers and Security, Vol.73, (2018),399 - 410

[12] Rama Rao. KVSN, Sivakannan S, M.A.Prasad,R.AgileshSaravanan, "Technical challenges andperspectives in batch and stream big datamachine learning", International Journal of Engineering & Technology, 7 (1.3) (2018) 48-51.

[13] Santiago López-Tapia, Rafael Molina,Nicolás Pérez de la Blanca, "Using machinelearning to detect and localize concealedobjects in passive millimeter-wave images",Engineering Applications of Artificial Intelligence, Vol.67, (2018), 81 – 90.

[14] Syed Ali Raza Shah, BijuIssac, "Performancecomparison of intrusion detection systems andapplication of machine learning to Snortsystem", Future Generation Computer Systems,Vol.80, (2018), 157 – 170.

[15] ZivKatzir , Yuval Elovici, "Quantifying theResilience of Machine Learning Classifiers Usedfor Cyber Security", Expert Systems withApplications, Vol. 92, (2018), 419 – 429.

[16]"Cyber Security Intrusion Detection for Agriculture 4.0: Machine Learning-Based Solutions, Datasets, and Future Directions", Mohamed Amine Ferrag, Lei Shu, OthmaneFriha, Xing Yang, JUNE 2022.

[17] ShilpiHiteshkumar Parikh, AnushkaGaurangSandesaraChintan Bhatt"Network Intrusion Detection Using Linear and Ensemble ML Modeling",© 2020 The Authors. Transactions on Emerging Telecommunications Technologies published by John Wiley & Sons Ltd.

[18] Guohou Shan; Boxin Zhao; James R. Clavin; Haibin Zhang; SisiDuan"Poligraph: Intrusion-Tolerant and Distributed Fake News Detection System",2021.

[19] Kathryn-Ann Tait, Jan Sher Khan, FehaidAlqahtani, Awais Aziz Shah, Fadia Ali Khan, Mujeeb Ur Rehman, WadiiBoulila, Jawad Ahmad"Intrusion Detection using Machine Learning Techniques: An Experimental Comparison,2021.

[20] Zeeshan Ahmad, Adnan Shahid Khan, CheahWaiShiang, Johari Abdullah1 Farhan Ahmad "Network intrusion detection system: A systematic study of machine learning and deep learning approaches"2021.

[21] B Padmaja, K SaiSravan, E Krishna Rao Patro, G Chandra Sekhar"A System to automate the development of anomaly-based network intrusion detection model",Journal of Physics: Conference Series 2089 (2021) 012006 ,2021.

[22] Muhammad zeeshan(senior member, ieee), Qaiser riaz1 (member, ieee), Muhammad a. bilal, Muhammad k. shahzad, hajirajabeen, syedalihaider, azizur Rahim,2021 "Protocol Based Deep Intrusion Detection for DoS and DDoS attacks using UNSW-NB15 and Bot-IoT data-sets",Digital Object Identifier 10.1109/ACCESS.2017.DOI 2021.

[23] Hamed Alqahtani1, Iqbal H. Sarker, AsraKalim, Syed Md. Minhaz Hossain,Sheikh Ikhlaq, and Sohrab Hossain, "Cyber Intrusion Detection Using Machine Learning Classification Techniques" ,© Springer Nature Singapore Pte Ltd. 2020.