Survey on Machine Learning Based Intrusion Detection System

Ashutosh Kumar Choudhary
Assistant Professor Department of IT GMRIT
ashutoshkumar.c@gmrit.edu.in
Sirivuru Jithendra Varma
Student Department of IT GMRIT
19341a12a8@gmrit.edu.in
Varanasi Vikas
Student Department of IT GMRIT
19341a12c2@gmrit.edu.in
Dunna Jayanth
Student Department of IT GMRIT
20345a1205@gmrit.edu.in
Tegala Rakesh
Student Department of IT GMRIT
19341a12b5@gmrit.edu.in

***Abstract*—**The expeditious growth in the internet and communication fields and wireless sensor networks have resulted in a huge growth in the size of the network and increased the importance of both network and the data. As a result, many new types of attacks are being created which could cause huge potential damage to the user. These attacks have given many new challenges to detect the intruders. Intrusion detection system plays an important role in detecting these types of intrusions and malicious activates which leads to decrease the performance of the devices on the network. An intrusion detection system (IDS) is one of the tools which protects the network by inspecting the network traffic, and protecting the network's confidentiality, integrity, and availability by analyzing the network. The development of IOT has increased the usage of (WSN's) and the devices on these networks are linked to cloud technology if the attacker can compromise one device whole cloud can be compromised. In this scenario "Intrusion detection system" will be able to detect various attacks like denial of service (Dos) and DDOS etc.

***Keywords*— *Intrusion Detection System, Wireless Sensor Network, Internet of Things (IoT*)**

DOI Number: 10.48047/nq.2022.20.19.NQ99059 **NeuroQuantology 2022; 20(19): 644-651**

I.INTRODUCTION

Network usage is rapidly increasing and another side cyberattacks are also increasing. But we cannot use same technique to secure information in all these industries. To protect the data from unauthorized people we can use the IDS. Generally, an intrusion detection system can be a device or a software that prevents a network or a system for pernicious activity or policy violations by using an IDS an alert can be sent to the administrator if any intruder is detected. IDS consists of two types mainly NIDS for a whole network and HIDS for individual devices in the network. To detect malicious activity by IDS technique now a days, machine learning algorithms like Naive Bayes, Random Forest and KNN algorithm has been used**.**

II.LITERATURE SURVEY

Pankaj et al. [1] discussed about the intrusion detection system which is anomaly-based intrusion detection system this IDS uses the decision tree algorithm and a tool-based verification which is carried out with the help of (AVISPA) tool, a tool based on linux operating system. Decision tree is an algorithm which is used for supervised machine learning processes for the techniques like classification of the structured data. In this algorithm structured data is taken as the input for the working of the algorithm. Decision tree algorithm is used for the processing of the data which is taken as the input. The raw data has been taken into two parts for training and testing, after the training of the model testing data is given to the model and three types of result is generated which is used further for analyzing the data whether it is safe or unsafe. AVISPA tool which has four different types of backbends named SATMC, TA4SP, OFMC, CL-ATSE is used for the verification of the data these backbends are used for different purposes for the verification CL-ATSE protocol analyzer is used for searching the attacks and states whether either the packets are safe or unsafe.

Treepop wisanwanichthan et al. [2] describes about Wi-Fi communication in a closed loop managing machine. Concerning the layout of IDS and modelling of the loop control system is used for the transmission among the nodes of the plant machine, controller machine and network system can be challenge to cyberattacks. The IDS includes diverse dispensed detection systems that are kept in positions with nodes of network device and actuators of plant system. An IDS analyses the amount or the types of the attacks. These records can additionally be analyzed to discover community device configuration problems. The metrics can then be used for future danger checks. The main drawback of the usage of an IDS is its incapability to respond or prevent assaults upon detection. However, community-based ids make use of network sensors strategically placed in the course of the network, permitting the system to stumble on reconnaissance attacks.

Ahmad W. Al-dabbagh et al. [3] discussed about the working of (DLHA) which is a double layered hybrid approach designed specially to deal with the previously mentioned troubles. Author studied common features of various assault classes by growing principal component analysis (PCA) variables which will maximize variance from each assault type, and determined that r2l and u2r attacks have comparable actions to regular customers. Percent is a bivariate evaluation that is used to measure the linear dating among two different variables, and ranks are given by the importance. Naive Bayes (NB) is selected as a classifier for organization, (SVM) is selected as a classifier for institution ,author specially centred on detection price (DR). DR is important because it implies what number of assaults the model can discover among the all number of real attacks.

Hong sheng yin et al. [4] discussed about better k-dependency bayesian community (KDBN) which is primarily based upon the IDCM and use's the common intrusion detection framework (CIDF) model as a reference. It contains four components: event generator, event aanalyser, event database, and response unit. The improved KDBN technology is the core unit of IDCM .This model can describe the relationship among the system.The system is developed primarily to detect the U2R,R2L attacks. The system can be able to detect R2L,U2R attacks at very good accuracy levels and stability at these attacks is also very good but the system cannot be used for the detecting other type of attacks.

Geetha Priya thamilarasu et al [5] discussed about how the cybersecurity threats are being done in health care industry. In three-five years the threats and attacks on internet linked medical devices can probably cause huge potential damage or a bodily harm and existence-threatening harm to the patients. For an example if any attacker hacked the network and changed the data related to amount of insulin to be given to a patient, if the attacker increased dosage, the patient may die. Related to the cardiac diseases the devices used for prevention from cardiac arrest such as a pacemaker can be hacked also which can keep the patient life in the risk. Author also stated about diverse attacks on scientific gadgets along with sniffing, information alteration, wrong data injection, and denial of provide or servicer assaults that can reduce or takeaway the affected person protection, safety and availability on important systems.The author designed and evolved a proper and good intrusion detector prototype which can be used for internet of medical things.The author tested different polynomial orders for the accuracy and performance and come to an end and stated that the 0.33 order polynomial changed into most relevant for the model without experiencing better computational sources. In the earlier studies, the author developed an initial framework using mobile sellers for WBAN and furnished a balanced evaluation of the device with other mobile agent based on totally intrusion detection.The author significantly enlarge the research to cope with both tool and network degree anomaly detection throughout the complete series of internet of medical matters.

Chih-Che Sun et al [6] discussed about IDS which can be used for the smart meters. As the usage of the smart meters is growing rapidly, the intruders are inventing various techniques to attack these meters, as the meters consist of hardware components and communication systems. The IDS consists of two stages to detect the intruder. In the first stage SVM algorithm is used to classify the abnormal behavior of mechanism in the meters. After detecting the abnormal behavior, the second stage comes to activation immediately. After this the IDS is also able to find the type of attack the victim is suffering from the attacker.

Valentina Casola et al. [7] discussed about in what ways the cloud services can be used for the healthcare system and for the usage and storage management in clinical statistical management. Each worker in the healthcare system will get right of entry to the host a cloud platform, which can be made to save, procedure, and balanced facts among patients, health system employees, and other relevant stakeholders (which includes facilities for ailment manage and if there is a pandemic detected preventive measures are taken). This type of platform also can host offerings for coping with the information of all the registered customers, patient record, and patient related health data and reviews. These type of cloud platforms also can support the healthcare company's administrative procedures, which include producing and updating the bills and reviews of the billing and dispensing price range. As like any technologies, cloud deployments inside the health system enterprises are also susceptible to attacks which can be done by both outside attackers (people not included in the network) and people working in the system or vendors employees associated with the cloud provider issuer (this are insider attacks). Researchers have attempted to clear up such demanding situations, for instance, by using cryptographic answers together with privacy-keeping cloud solutions.

Orly Stan et al. [8] discussed about the intrusion detection which, consists of two modules: (1) a far-off terminal (rt) a model in which authentication is done and detects the legitimate connected components (2) a chain-primarily based anomaly detection module that detects intrusions within the working of the system. The rt authentication module managed to authenticate RTS with +0. 99 precision and +0. 98 bears in mind; and stumble on illegitimate thing (or a legitimate thing that impersonates other additives) with +0.98 precision and +0.99. The collection primarily based anomaly detection module controlled to flawlessly discover each regular and bizarre behavior. Moreover, the sequence- based anomaly detection module controlled to accurately (i.e., zero false positives) model the normal conduct of an actual gadget short period of time (~22 seconds). The author planned to improve the proposed ids with the aid of growing an extra module aimed toward examining the payload of information and standing words.

Usman Shuaibu et al. [9] discussed about the considerable growth in the usage of computer networks, making it more challenging to ensure network availability, integrity, and security. Threats or malicious behavior can be discovered by an intrusion detection system (ids). The ids provide computer network security at the network level. The intrusion might be recognized as odd by a network. Intruders can damage community security by exploiting network weaknesses, programming faults like buffer overflows, and other community vulnerabilities like rules susceptible to protection. The invaders are either less privileged device users who want to have more access to power or hackers who are common internet users who want to steal or harm sensitive data from the victim's device. The earliest intrusion detection systems were primarily signature-based, which means that they relied on pre-defined and configured lists of known attack signatures to identify malicious activity. The effectiveness of single, hybrid, and ensemble ML techniques in conjunction with intrusion detection system is evaluated in this study.

Deris stiawan et al.[10] discussed about how to optimize the intrusion detection systems through the traffic which is coming to the network and by analysing the traffic further and extracting the information from the traffic. The dataset used is (ITD-UTM) dataset which is developed by a university in Malaysia these datasets are considered because datasets are very much close to real time traffic which have attacks on the systems. The dataset is pre-processed using the data division because normalization gives the results in long decimal values which will cause overflow problems. For the developing of the system six feature ranked techniques are used those are IG, GR, SO, RF , CS. These techniques uses four feature classifiers which describes about the accuracy and used for the comparison.

III. Comparative study:

Title of the paper	Tools used & Methodology Followed	Merits	Demerits
Machine Learning Based Novel Approach for Intrusion Detection and Prevention System: A Tool Based Verification (2018) [1]	1.Decision tree algorithm. 2.AVISPA tool (Linux operating system).	AVISPA tool is used for verification and structured data is taken as input so good results can be obtained by implementing the decision tree algorithm.	The paper does not discuss about the previous existing systems for the working of the model and decision tree. No comparisons of different models are done.
An Intrusion Detection System for Cyber Attacks in Wireless Networked Control Systems (2018) [2]	1.IDS is used for the closed control loop system using different types of distributed detectors along the network.	An IDS can be used to detect different type of intrusions and attacks. With the help of IDS. This information can be gathered and can be further used to change the security control or to introduce new laws that can protect the system.	The major downside of using an IDS in network is its disability to counter or stop these attacks after the detection.

A Double-Layered Hybrid Approach for Network Intrusion Detection System Using Combined Naive Bayes and SVM” IEEE 2021[3]	1.Principial component analysis and one hot encoding is done for the data transformation. 2.Navie bayes and Support vector machine.	The experimental results suggest that DLHA out performs several existing states of the threats. Ids techniques and is significantly better than any single machine learning classifier by large margins.	The proposed system consists of an algorithm called (DLHA)which is a double layered hybrid approach which is very much useful in the new type of datasets.
Intrusion Detection Classification Model on an Improved k-Dependence Bayesian Network [4]	1.KDDCup99 dataset for verification. 2.k-Dependndce Bayesian network and improved KDBN for the building of the model.	The virtual augmentation order to achieve high efficiency, high accuracy and high reliability of network.	It perfectly solves the problem of low detection accuracy and low stability for small categories (U2R and R2L) mentioned by many sources but not preferable for other attacks.
An Intrusion Detection System for Internet of Medical Things [5]	1. Special agent protocol algorithms. 2.PCA technique for feature extraction and polynomial regression for intrusion detection algorithm assessment is done through feedback reliability ratio and training time.	This methodology provides very good intrusion detection system including the hardware sensors, also it is very much useful for internet of things devices in the service of health care and by using this model we can also further develop it for different devices also.	The main disadvantage of tis methodology is as it is implementing for the sensors it is better to use state graphs rather than regression because regression deals with data samples only.
Intrusion Detection for Cybersecurity of Smart meters [6]	1.Support Vector Machine 2.Pattern reorganization algorithm for intrusion detection. 3.For hardware	This methodology provides a very good detection system on smart meters using various hardware	All types of attacks related to the smart grid are considered but these attacks may be updated day by day.

	pattern reorganization RAM connections are analyzed.	observations also.	
Health care related data in the cloud: challenges and opportunities [7]	Methodology is mainly focused on the dependencies of cloud in the health care domain and data management in the cloud Based data.	This paper provides very much data related to cloud data management and tells the challenges for the achieving the security in the cloud data management.	There is no proper data related to overcome the challenges received by the cloud data management.
Intrusion Detection System for the MIL-STD-1553 Communication Bus [8]	1.A tailor-made IDS. 2.RT authentication method is used for authentication.	As this intrusion detection system mainly focuses on the MIL-Communication bus this can be used at the time of cyberwarfare and military purposes also.	This IDS which is used in this model is unable to cover attack methods that utilize only data and status words.
Intrusion Detection System using Machine Learning Techniques: A review [9]	As this is a review paper there are many different types of techniques included for the intrusion detection system.	This paper focuses on various machine learning techniques which are related to both network intrusion and host Intrusion also.	Research papers should also be taken the intrusion detection system related to cloud but there is no proper papers mentioned in the above paper.
An Approach for Optimizing Ensemble Intrusion Detection Systems [10]	1.Gain ratio, Chi-Squared, Relief-F(RF), One-R, Symmetric uncertainty for feature selection methods. 2Bayesian Network, J48, Naive Bayes, SOM for Ensemble. 3.Hold-Up, F-Measure, static	This paper gives a very good feature selection techniques for the intrusion detection system features to be considered and covers various aspects of the features using different feature selection methods.	This methodology may be accurate on the dataset which is used in the working of the paper but it may be not implemented accurately for other datasets. Some other feature selection techniques should also be

	for validation.		considered to select more accurate features.

IV.Comparative analysis of Algorithms

Algorithm (Methodology)	Accuracy %	Error %
DLHA (R2L)	96.6%	3.33%
KDBN	95.99%	4.11%
DLHA(U2R)	100%	0
SMART GRID	95.45%	4.65%
NN(MLP)	98.22%	1.78%
SVM(RBF)	98.71%	1.29%
Bayesian	92.12%	7.88%
Naïve Bayes	60%	40%
J48	85%	15%

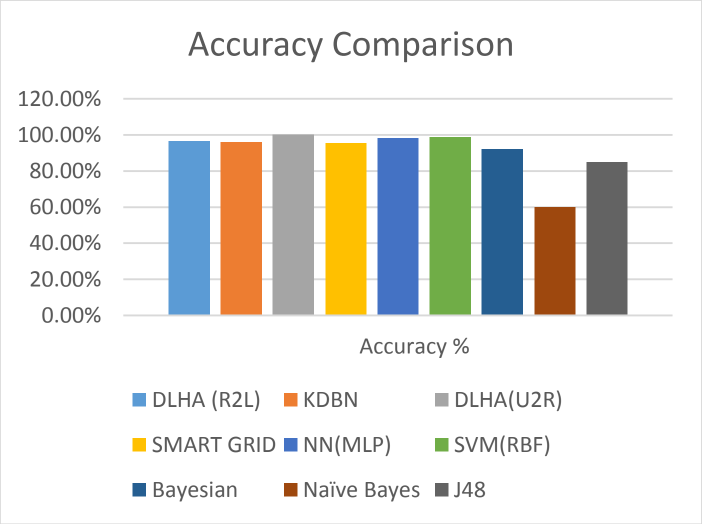

Fig1: Representation of Accuracy Rat

V. Conclusion:

Intrusion detection system is very important tool for the protection of the system or a complete network. An IDS can be prepared by using different techniques like history based or machine learning based. A machine learning based IDS can be prepared by using various machine learning algorithms like KNN, KDBN, etc. Algorithms reviewed in this paper have different accuracy levels. DLHA has high accuracy but it can be only preferable for the U2R attacks. SVM algorithm has high accuracy comparing to the other algorithms which consists of all the attacks so SVM can be used for developing a machine learning based IDS which can be able to detect the intruder properly and help the user from intruders or unauthorized people.

REFERNCES:

1.Pankaj Ramchandra Chandre, Parikshit Narendra Mahalle, Gitanjali Rahui Shinde“Machine Learning Based Novel Approach for Intrusion Detection and Prevention System: A Tool Based Verification” 2018 IEEE

2. Treepop wisanwanichthan, Mason Thammawichai “A Double-Layered Hybrid Approach for Network Intrusion Detection System Using Combined Naive Bayes and SVM” IEEE 2021

3. Ahmad W. Al-Dabbagh and Tongwen Chen “An Intrusion Detection System for Cyber Attacks in Wireless Networked Control Systems” IEEE 2018

4. Hongsheng yin, mengyang xue, yuteng xiao, kaijian xia, and guofang yu1"Intrusion Detection Classification Model on an Improved k-Dependence Bayesian Network", 10.1109/ACCESS.2019.2949890 2019 IEEE

5. Geethapriya thamilarasu, adedayo odesile, and andrew hoang “An Intrusion Detection System for Internet of Medical Things” 10.1109/access.2020.3026260 IEEE 2020

6. Chih-Che Sun, Adam Hahn, D. Jonathan Sebastian, Chen-Ching Liu “Intrusion Detection for Cybersecurity of Smart Meters”, 10.1109/tsg.2020.3010230, IEEE 2020.

7 Valentina Casola, Aniello Castiglione, Kim-Kwang Raymond Choo, Christian Esposito and P. S. Chatterjee, " Healthcare Related Data in the Cloud: Challenges and Opportunities," 2016 IEEE.

8. Orly Stan, Adi Cohen, Yuval Elovici, Asaf Shabtai “Intrusion Detection System for the MIL-STD-1553 Communication Bus”, 10.1109/TAES.2019.2961824, IEEE 2019

9. Usman Shuaibu Musa Megha Chhabra Aniso Ali Mandeep Kaur “Intrusion Detection System using Machine Learning Techniques: A Review” IEEE Xplore Part Number: CFP20V90-ART; ISBN: 978-1-7281-5461-9 2020

10. Deris stiawan, Ahmad heryanto, Ali bardadi, Dian palupi rini, Imam much Ibnu subroto, Kurniabudi, mohd yazid bin idris, Abdul hanan Abdullah, Bedine kerim , and rahmat budiarto “An Approach for Optimizing Ensemble Intrusion Detection Systems”, IEEE 2021