IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MACHINE LEARNING APPROACHES FOR DETECTING INDUSTRIAL IOT NETWORK ATTACKS

[1]Varsha Prafull Patil, [2]Dr. Sharada N. Ohatkar

[1]Research scholar, [2]Professor

[1] Department of Electronics and Telecommunication Engineering,

[1]MKSSS's Cummins college of Engineering for Women, Pune, India

Abstract: The adoption of the Industrial Internet of Things (IIoT) and the deployment of 5G technology have led to significant advancements in industrial processes and automation. However, this proliferation of interconnected devices within the context of 5G-enabled IIoT networks has also exposed industrial systems to a heightened risk of cyberattacks. Securing IIoT networks in the era of 5G has thus become a paramount concern for ensuring the integrity, availability, and safety of critical industrial systems. This paper presents a comprehensive exploration of machine learning approaches for detecting and mitigating network attacks in the context of the 5G-enabled IIoT.

IIoT deals with large scale networks and known as use of internet of things for industrial applications. In manufacturing the networked devices, interconnected sensors and computer systems used are together refereed as Industrial IoT. Maintaining data privacy and securing the IIoT network is also a great challenge. Using various Artificial Intelligence algorithms, it is possible to provide the security to the IIoT network and data against the possible attacks on the IIoT network. This paper provides the systematic survey of possible attacks on industrial IoT network, applied ML technique for detection of attacks along with the suitable platform used.

Index Terms **- Industrial IoT, Cloud Computing, ML taxonomy, IoT edge, Machine learning and IoT applications**

Introduction

An invisible network between multiple objects scattered in various directions is called IoT or Internet of Things. Industrial Internet of Things is a subset of IoT, which focuses on the peculiar requirements of industrial applications such as manufacturing, utilities, oil and gas. Generalized IIoT network structure: Especially, IIoT refers to all interconnected sensors, devices, and different gadgets that combine with industrial packages, which includes manufacturing and electricity management, to create a complete service network that enables the application of better-stage automation. This connectivity permits the gathering, exchange and evaluation of facts because it enables performance enhancements for the duration of the manufacturing chain. Also permits the sector of manufacturing to make massive innovation leaps, received a good-sized extroversion, and increase sports that had been formerly not possible. It if focused that the entire transformation of the deliver chain right into an integrated and absolutely automated manner based totally on IIoT assumes a

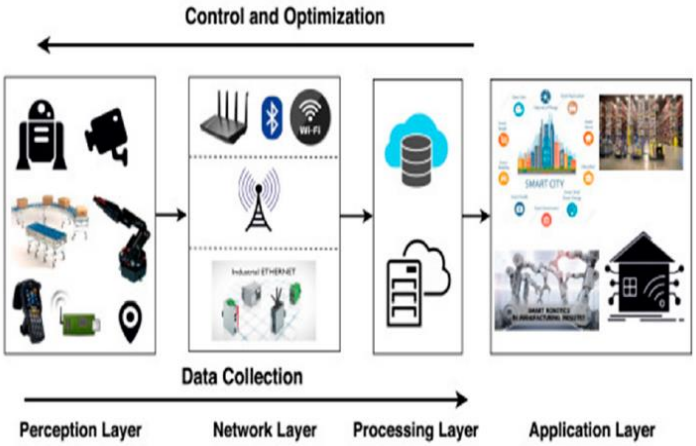

Figure 1. 1: Generalized IIoT architecture.

A more careful method to threats associated with commercial IoT structures is presented in [8], wherein the authors provide an in-depth listing of possible assaults at the layer of five useful tiers of Industrial IoT, the first three as a part of operational technologies (OT), while the other two are part of information technologies (IT). The functional level mainly includes structures, to carry out IIoT physical tactics, which includes embedded gadgets, actuators, sensors, cars and transmitters. Attacks at this stage requires incredible knowledge of IIoT design system and get admission to active device specifications, engineering plans, and specified statistics approximately their operational and installation capability. The second stage includes specialized system that communicates with and controls the first-degree devices, which includes dispensed manipulate structures (DCS), programmable common-sense controllers (percent), and gateways. Attackers at this level goal to prevent valid conversation between the two tiers and control the glide of verbal exchange. The third useful degree is SCADA with all associated manipulation of industrial automation and system telemetry which include records device acquisition, master stations and interfaces of human-device that talk thru IP. Many SCADA-degree attacks depend on techniques to craft IP packets with fake attributes which includes the source cope with to masks the sender packet identity and make recipient think it is from a valid community user. The 4th useful degree consists of enterprise making plans offerings including workplace packages, intranet, internet and services of mail. Targeted attacks at this level exploit unrecognized vulnerabilities in services and inject harmful program where utility expects legitimate statistics from the person to advantage administrative get admission to. SCADA level with all associated automation business manipulate and remote measuring structures together with information acquisition gadgets, grasp stages and interfaces of human-gadget that speak using IP. Multiple SCADA-level attacks depend upon strategies to craft IP packets with fake attributes consisting of the source cope with to mask the identification of packet sender and make recipient suppose it's from valid network user. The 4th functional degree includes planning for business. Here the services included as workplace programs, intranet, mail services and web. Attacks centered at this stage make the most acknowledged or unknown vulnerabilities in these offerings and inject malicious program where the software expects valid statistics from consumer for administrative advantage of getting right of entry. The 5th functional degree includes excessive-level offerings consisting of analytics, facts mining methods controlled with the aid of organization programs, and cloud computing offerings. assaults at this level consist of a hard and fast of malicious actions inclusive of interception and spoofing, however additionally extra advanced kinds including antagonistic assaults [1].

Layer		Components	Possible Attacks
IT	V	Business Applications, Cloud Computing, Data Analytics, Internet and Mobile Devices	DoS, Side channel attacks, Cloud malware, Injection, Authentication Attacks , Man-in – the-Middle, Mobile device attacks.
	IV	Data centers, Office Application, Intranet, Mail and Web Services	Phishing, SQL Injections, Malwares, DNS poisoning, Remote code Execution Brute Force Attacks, Web Application Attacks.
DeMilitarized zone			
OT	III	SCADA Control, HMI, Control Room and Operator Stations	IP Spoofing, Data sniffing, data manipulation, Malwares
	II	Distributed Control systems, PLC's , and Gateways	Replay attack, Man-in-the -Middle attack, Sniffing, Wireless device attacks, Brute force Password guessing
	I	Sensors, Motors, Actuators, Transmitters, Embedded Devices.	Reverse Engineering, Malware, injecting crafted packet or input, Eavesdropping, Brute force search attacks.

Table 1.1: Layer wise possible attacks.

II LITERATURE SURVEY:

Recently the popularity of Cloud computing use has become increasing. The personalized data centers are used to provide inexpensive infrastructure solution to the business plans. Variety of Internet services are provided by Cloud computing. Cloud computing with various online resources is used to assist users/organizations in reducing infrastructure costs. In [4] the five ML algorithms, Support Vector Machine, Naive Bayes, Decision Tree, K nearest neighbor and Random Forests, were used. Comparison of evaluation parameters mentioning the F1 score, accuracy, precision and recall were taken. It is observed that the accuracy of the SVM algorithm is high and that of Naive Bayes is low.

End users uses widely distributed, as Platform as a service (PaaS), Software as a service (SaaS) and Infrastructure as a service (IaaS). Hence there is no need for user's ownership of the cloud computing infrastructure for knowledge, control and for deployment management of their applications. For detection of

the anomalous activity and reporting to the admin authority IDS is used, which may be hardware or software. Prior information given to the system administrator regarding the attack. Definition of IDS is given by James Anderson in 1980.With the help of collection methods examination of access logs & server case logs carried out by him. Intrusion Detection Expert System (IDES) an anomaly-based IDS based on statistics created by Dorothy E. Denning in 1986. Artificial Neural Networks (ANN) is applied to IDES boosting by Teresa F. Lunt. A rule-based anomaly detector with mathematical analysis is created at the University of California by researchers Wisdom & Sense. Anomaly and Signature based Distributed Intrusion Detection System in 1991 (DIDS) is given by Davis. In order to detect unknown attacks Anomaly based Intrusion Detection System (AIDS) is used for matching the incoming network traffic with the existing patterns extracted Signature-based detection is used. This is named as or Knowledge-based detection or Misuse based detection. In [5] the principle of abnormal network traffic could be malicious is used. Using this it is possible to identify the network activity as an exception and leads to further investigation to calculate the nature of the traffic. Machine Learning (ML) algorithms work effectively and are best suited for anomaly detection.

With the help of target CPPS (including, SCADA, MES, ERP, etc.), the network traffic traces (PCAP files) are captured at the start. Followed by extraction of features of two way traffic flow which includes parameters such as packet length, packet inter arrival time number of received/sent packets, and flow duration. For data mapping, the PCA algorithm is applied by feature selection module, which allows elimination of less/redundant feature without losing data. An alert about the malicious network traffic pattern is generated by detection engine with the help of mapped data provided to it. For timely retraining the detection engine's Machine Learning models, the feature selection module's saved dataset is used. Analysis of the detection engine output has been carried out by the alert manager. This then generates the alert if the occurrence of intrusion goes up beyond the predefined security level. Here for attacks detection in Industry 4.0 CPPSs machine learning is applied. From the semiconductor based large production factory the traces of network traffic are taken known as PCAP files. 11 types of semi-supervised, supervised and unsupervised machine learning algorithms are applied for detecting the anomaly in traffic network. As per the supervised algorithms simulation results, the better performance than unsupervised and semi-supervised. For detecting DDoS attacks, DT,K-NN and RF algorithms are used with Precision, Recall & accuracy of 0.999 and FPR as 0.001.K-means & EM unsupervised algorithms showed a best performance with Recall > 0.9, FPR > 0.09 Precision > 0.9 and Accuracy of 0.95 and the slightly decreasing performance is observed after application of PCA algorithm (with variance retain about 95 %).Here for unsupervised & supervised learning data labelling is not required, which need human intervention & efforts and a difficult task.

In [6] for classifying the Industrial IoT malware attacks artificial neural network (ANN), long short-term memory (LSTM) and gated recurrent unit (GRU) can be used with 99 % accuracy. Accuracy of GRU and LSTM is found as 98 % and that of ANN is 99 %. While the F1 score, and Cohen's Kappa are same for all with the value 0.98 each. The IAUC- ROC value is 0.83 for GRU, 0.84 for LSTM and 0.85 for ANN. GRU is giving maximum precision value as 0.99 and it is 0.98 for LSTM and ANN. Recall value is 0.98 for LSTM and GRU while the recall value of ANN is 0.99.

In [7] the attacks with low accuracy & low precision are termed as challenging attacks and are identified by the model. Due to these types of attacks the performance of the model get reduced. Hence it is necessary to develop a two-level model for identification of challenging attacks in industrial IoT. Hence, the proposed work seeks to develop a two-level IDS that identifies attacks in the IIoT and focuses on challenging attacks. Tremendous traffic in the Industrial IoT network is generated by smart meters and sensors. The data will have a wide range and magnitude. The normalization and data preprocessing are carried out before the first-level detection because of the broad range and magnitude of the data.

Level-1 Detection: In First level detection the challenging attacks are categorized as misclassified or benign attacks then these were given to Level-2 detection.

Level-2 Detection: In level 2 misclassified challenging IIoT attacks are detected. For identification NSA, two models and DNN (trained using enhanced dragonfly algorithm) are used. Number of benefits are given by Industrial Internet of Things (IIoT). Here the study of various applications, security issues in the IIoT with a Deep Learning-based two-level Intrusion Detection System are mentioned. The proposed model first separate challenging attacks in the first-level detection and carry out advanced detection in the second level.

For evaluating the model IoT and network datasets and benchmark is used. Demonstration shows the better performance than the existing system models. Better performance can be observed on new attacks or zero-day. attacks using two-level IDS. Also, this can be extended for particular environments like Machine-to-Machine (M2M) and Internet of Medical Things (IoMT) for data and device protection.

In [8] the random neural network based lightweight prediction model is proposed for evaluating the F1 score, recall, precision and accuracy as performance parameters for attack detection in Industrial IoT network. These parameters are then compared with conventional support vector machine, decision tree and artificial neural network. The results show an accuracy of 99.20% can be achieved with the rate of learning as 0.01 with a 34.51 milliseconds prediction time. Other parameters of performance like recall, precision, F1 score are 99.13%, 99.11 % and 99.20 %. Using this model, it is possible to improve the accuracy of attack detection by 5.65 % average as compared to conventional machine learning methods used in the IoT security. RaNN proposed here is used to detect the mentioned attacks with a maximum 99 % accuracy with 34.51 milliseconds prediction time as compared to other techniques. With 0.01 learning rate better results of RaNN are observed. Proposed RaNN based prediction model gives greater accuracy than the available DT, ANN and SVM machine learning algorithms. Recall, F1 score, and precision parameter values are also found improved using the RaNN based approach. Here the deployment of hardware using Intel Neural Computing Stick-based architecture and Rasberry Pi 4 for detection of attacks is also mentioned for detecting the attacks at the edge for Industrial IoT. Model proposed here is tested for a DS2OS open-source dataset.

Table 2.1: Comparison of ML approaches used with the Application & Platform of implementation

Sr. No.	Machine Learning approach used	Application	Platform
1	Support Vector Machine, Naive Bayes, Decision Tree, K nearest neighbor and Random Forests	Cloud Computing	Google Colaboratory under Python 3
2	DT,K-NN and RF	Industry 4.0	Python and Google colab
3	artificial neural network (ANN), long short-term memory (LSTM) and gated recurrent unit (GRU)	Industrial IoT	Python 3.8 using Keras, Tensorflow, and Scikit-Learn libraries.
4	Deep Learning-based two-level Intrusion Detection System	Industrial IoT	Python and Google colab
5	DT, ANN and SVM machine learning algorithms	Industrial IoT	Anaconda Navigator and Python language

III TAXONOMY OF MACHINE LEARNING [9]:

For securing wireless communication in IIoT-based systems the Machine learning (ML) algorithms have been used. Also helps to resolve different types of cybersecurity challenges. One of the efficient intrusion detection methods is use of machine learning approaches. As these approaches are applied to various datasets and used to analyze data in real time to in order to get best results. Supervised learning, deep learning, decision trees and adaptive learning approaches are used by many researchers nowadays.

Machine learning algorithms are categorized as supervised and unsupervised. Supervised Learning deals with the learning from labelled data and applying labels based on patterns. Naïve Bayes is a simple and effective learning algorithm that predicts end result based on the similarity of each class's attributes. It simplifies probability calculations and is often used with categorical data. Decision trees are a powerful supervised machine learning method that can perform regression and classification problems. They are more efficient and faster than remaining classification algorithms. Logistic regression is simple model that predicts a class probability of being 1 given an observation. It is best suitable for binary classification and can also be enlarged to multiclass classification with some extra efforts. The k-Nearest Neighbors approach is a machine learning method used to resolve classification issues. Overall, supervised and unsupervised learning algorithms play crucial roles in the area of machine learning.

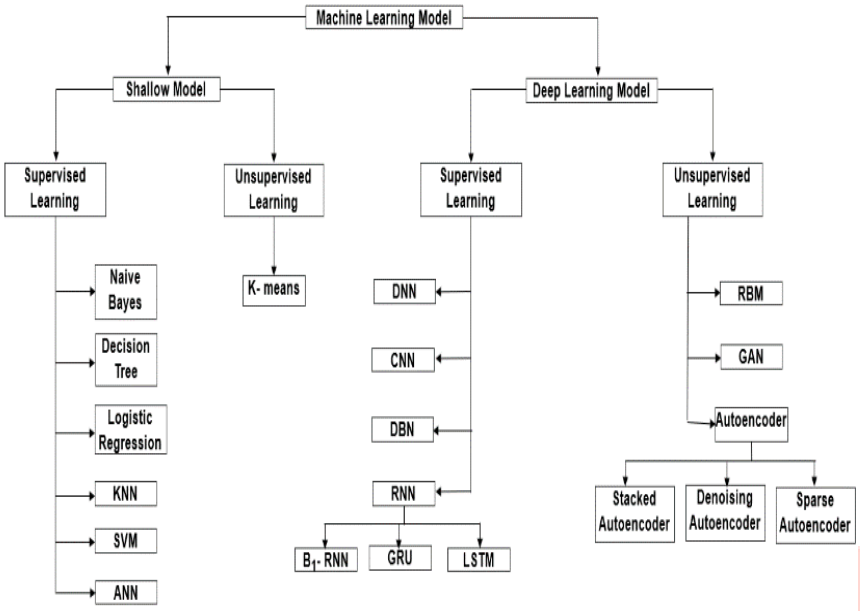

Table 3.1: Taxonomy of Machine Learning

KNN is a slow and parameter free learning approach that deals with zero assumptions on basic data distribution. The lack of assumptions about the primary data distribution is known as “nonparametric.” Identification of the number of nearest neighbors is done by k in kNN, the very important decision factor is the quantity of neighbors. It calculates the distance between the input and its test data before generating the appropriate decision.

The SVM (Support Vector Machine) is a widely used framework in machine learning and embedded systems, allowing for the classification of nonlinear and linear objects. It is particularly favorable in big data environments for multidisciplinary environment. Artificial Neural Networks (ANN) can be created using various types of neurons, with sizes ranging from tens of thousands to fewer than 10. Unsupervised Learning is used to analyze data structure, extract insights, and increase efficiency. An unsupervised approach called K-Means Clustering is used in dividing data into k-clusters, with known number of clusters.

Deep Learning (DL) is a subset of ML techniques that consists of several linked layers, with the first layer being the input and the output layer being the output layer. Signal intensity of a neuron is affected by parameters such as weight, activation function and each covered layer is made up of many neurons. Supervised and unsupervised learning methods are two categories of Deep learning techniques, including Convolutional Neural Network (CNN), Convolutional Neural Network (DBN), recurrent neural networks (RNN) and deep neural networks (DNN).

Unsupervised learning methods include generative adversarial network (GAN), restricted Boltz-mann machine (RBM) and auto encoder (AE). In summary, the SVM, ANN, and DL are essential tools for machine learning and embedded systems. These techniques can be applied to extensive data classification challenges and are highly useful in multidomain applications.

Machine learning techniques	Advantages	Disadvantages
kNN	Effective and simple classification performance in various domains	Shows less run time performance on big training set.
SVM	High accuracy and low computational cost.	Lack of transparency in results.
ANN	It can learn without the need to be programmed.	Extra processing time for a big neural network.
Naïve Bayes	It requires small space during classification and training.	It can be oversensitive to irrelevant attributes.
Decision tree	Decision trees are very fast and simple.	It has long training time.
***K*-means**	Implementation is easy.	It is too hard to anticipate *K*-value.
DNN	Has a capacity of processing unrecognized data.	Highly dependent on the availability of the data
CNN	High accuracy in image recognition problems.	Big computational cost.
RNN	RNN used in a collection of records modelling.	Gradient exploding and vanishing problems.
GAN	GAN generated images are very true.	Poor interpretability of neural networks.

IV CONCLUSION & FUTURE SCOPE

Performance analysis and comparative study of threat detection model for cloud services using machine learning techniques like Native Bayes & Support Vector Machine has been carried out by considering various performance parameters. Different ML algorithms can be applied to Industry 4.0 CPPSs for detecting DDoS attacks with the comparative performance evaluation of each approach. In future it is possible to develop a combined IDS using federated learning by combining the mentioned ML approaches. Two level Deep learning-based IDS can be used to identify the challenging attacks at the first level and identification of new ttacks and zero-day attack has been done at the second level with improved performance than the existing methods. In future for protecting the data and device, this model can be applied to Machine to Machine and Internet of Medical Things (IoMT) environments with the performance evaluation. DT & RaNN algorithms can be used for Attack Detection in the Industrial IoT. As per the ML taxonomy variety of ML approaches are available which are to be applied to the dataset in order to detect the threats in industrial IoT network.

V REFERENCES

[1] A. Abdullah, R. Hamad, M. Abdulrahman, H. Moala, and S. Elkhediri, “Cybersecurity: a review of internet of things (iot) security issues, chal-lenges and techniques”, in 2019 2nd International Conference on Com-puter Applications & Information Security (ICCAIS). IEEE, 2019, pp.1_6.
[2] K. Tsiknas, D. Taketzis, K. Demertzis, and C. Skianis, “Cyber threats to industrial iot: a survey on attacks and countermeasures”, IoT, vol. 2, no. 1, pp. 163_186, 2021.
[3] H. Vargas, C. Lozano-Garzon, G. A. Montoya, and Y. Donoso, “Detection of security attacks in industrial iot networks: A blockchain and machine learning approach”, Electronics, vol. 10, no. 21, p. 2662, 2021.
[4] R. K. Sadavarte and G. Kurundkar, “Survey and performance analysis of machine learning based security threats detection approaches in cloud computing”, 2021.
[5] F. B. Saghezchi, G. Mantas, M. A. Violas, A. M. de Oliveira Duarte, and J. Rodriguez, “Machine learning for ddos attack detection in industry 4.0” cppss, Electronics, vol. 11, no. 4, p. 602, 2022.

[6] M. Mudassir, D. Unal, M. Hammoudeh, and F. Azzedin,"Detection of botnet attacks against industrial iot systems by multilayer deep learning approaches", Wireless Communications and Mobile Computing, vol. 2022.
[7] K. Raja, K. Karthikeyan, B. Abilash, K. Dev, and G. Raja, "Deep learn-ing based attack detection in iiot using two-level intrusion detection sys-tem", 2021.
[8] SHAHID LATIF & ZHUO ZOU. "A Novel Attack Detection Scheme for the Industrial Internet of Things Using a Lightweight Random Neural Network", IEEE Access Green Internet of Things 2020.
[9] Asadullah Momand, Sana Ullah Jan and Naeem Ramzan "A Systematic and Comprehensive Survey of Recent Advances in Intrusion Detection Systems Using Machine Learning: Deep Learning, Datasets, and Attack Taxonomy", Hindawi Journal of Sensors 2023, https://doi.org/10.1155/2023/6048087