ISSN: 2395-7852

International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

Volume 11, Issue 3, May-June 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

IMPACT FACTOR: 7.583

| www.ijarasem.com | ijarasem@gmail.com | +91-9940572462 |

International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 7.583 | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 11, Issue 3, May-June 2024 |

Machine Learning Model for Detection & Identification of Attack in the IOT

Vaishnavi Rajesh Banchode [1], Dr. Pawan R. Bhaladhare [2] Dr. Mohammad Muqeem [3]

M.Tech (CTIS) Student, Dept. of Computer Science and Engineering, Sandip University, Mahiravani, Trambakeshwar, Nashik, Maharashtra, India

Professor, Dept. of Computer Science and Engineering, Sandip University, Mahiravani, Trambakeshwar, Nashik, Maharashtra, India

ABSTRACT: Attack and anomaly detection in the Internet of Things (IoT) infrastructure is a rising concern in the domain of IoT. With the increased use of IoT infrastructure in every domain, threats and attacks in these infrastructures are also growing commensurately. Denial of Service, Data Type Probing, Malicious Control, Malicious Operation, Scan, Spying and Wrong Setup are such attacks and anomalies which can cause an IoT system failure. In this paper, performances of several machine learning models have been compared to predict attacks and anomalies on the IoT systems accurately. the paper investigates the integration of anomaly detection and supervised learning approaches to improve the accuracy and efficiency of attack identification. Anomaly detection helps identify deviations from the expected behavior, while supervised learning utilizes labelled datasets to categorize specific attack types. This hybrid approach enables the model to continuously learn and adapt to emerging threats in real-time. The dataset NSL KDD was used in the recommended framework and the performances in terms of training, predicting time, specificity, and accuracy were compared.

KEYWORDS: IOT, Dos, Machine learning, Attack

I. INTRODUCTION

In order to detect accurately in an IoT environment, a brand-new dataset is employed for efficient feature selection. The data set documents a wide variety of cyberattacks, including those launched via botnets, common online behaviors, and the Internet of Things. This dataset, featuring efficient information features for tracking accurate traffic, was developed using a practical test platform. Other features were retrieved in a similar fashion and added to the set of extracted features to improve the effectiveness of machine learning models and forecasting models. However, the extracted features, like assault flow, groups, and subdivisions, are labelled for better performance outcomes. The (IoT) is a rapidly developing knowledge that links millions of devices in a matter of seconds. Using this technology streamlines and simplifies regular tasks. The (IoT) is one example of a technology that was once exclusively used in households and small businesses but is now widely adopted because of the benefits it provides in terms of efficiency and reliability. Yet, Internet of Things technology is rapidly replacing traditional methods. With the advancement of IoT technology, more than 27 million connected devices are expected by 2021. Due to their rapid development and widespread usage, (IoT) devices are increasingly the target of cyberattacks. Several studies have found that botnet attacks are more common in IoT environments than other types of attacks. Because most IoT devices lack the memory and processing capacity required for good security systems, they still contain many security problems. In this study, using a powerful feature assortment strategy, a lightweight, highly efficient recognition system is created. A botnet is a group of bots controlled by a single individual, the botmaster, via a command-and-control protocol to attack a specific network. Infected computers, known as bots, are utilized to perform malicious tasks under the botmaster's virtual control without leaving any traces of infection. Machine learning algorithms will be crucial in IoT security systems because they are effective at modelling systems that cannot be given by mathematical equations. Machine learning is a subfield of computer science that studies how machines might acquire knowledge through observation and experience.

II.LITERATURE REVIEW

Mahmudul Hasan, Md. Milon Islam, Md Ishrak Islam Zarif, M.M.A. Hashem "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches" In this paper, performances of several machine learning models have been compared to predict attacks and anomalies on the IoT systems accurately. Attack and anomaly detection in the Internet of Things (IoT) infrastructure is a rising concern in the domain of IoT. With the increased use of IoT infrastructure in every domain, threats and attacks in these infrastructures are also growing commensurately. Denial of Service, Data Type Probing, Malicious Control, Malicious Operation, Scan, Spying and Wrong Setup are such attacks and anomalies which can cause an IoT system failure.

Md. Rumman Rafi, Mohammad Abdus Salam "Machine Learning Based Attack Detection in Internet of Things Network" In the IoT network, micro services behave differently at different times which causes deviations in normal behavior in IoT services thus creating an anomaly. Further study is needed to interpret these problems in a more in-depth way. In this research label encoding technique have been used to convert the data into a feature vector. Most of the features in this dataset contain nominal categorical value and many unique values. In this process, the dataset was collected and observed meticulously to find out the types of data. Besides, data preprocessing was implemented on the dataset. Data preprocessing consists of cleaning of data, visualization of data, feature engineering and vectorization steps. These steps converted the data into feature vectors. In the case of real-time data, there may raise different problems. A more empirical study is needed on this problem focusing on real-time data. In the IoT network, microservices behave differently at different times which causes deviations in normal behavior in IoT services thus creating an anomaly. Further study is needed to interpret these problems in a more in-depth way.

Maryam Anwer, Shariq Mahmood Khan, Muhammad Umer Farooq, Waseemullah "Attack Detection in IoT using Machine Learning" In this part, the attack detection problems are studied by statistical classification of measurements using the implementation of ML. The spam filter in cyber security separates spam from different communications services. Spam is apparently the leading ML method applied in information security. In this process, the data were collected and observed deeply to analyze the types of data. In the data preprocessing step, the data were cleaned, visualized, and feature engineering was applied along with implemented vectorizations. The proposed model for this research work is an ordinary huge organization or a smart city going through an increasing variety of IoT-associated cyber threats, such as heavy obligation DDoS attacks, achieved with an enormous botnet, e.g. Mirai, which exploit default or weak passwords.

Kalaiselvi, Hariharasudhan, Divakar, Mytheeswaran, Niveditha "Detection of malicious attacks in IOT network traffic using machine learning techniques" In order to keep tabs on and prevent undesirable business flows in the (IoT) network, it is crucial that abnormality and malicious business be identified in the IoT network. In order to prevent malicious financial transfers across the IoT network, many researchers have presented numerous machine literacy (ML) style models. Yet, because to the unfortunate point selection, a number of ML models are predisposed to misclassify significantly malicious business flows. However, additional research is needed to fully understand the critical issue of how to select efficient features for precise malicious business discovery in an IoT network. A new frame model is presented as a solution to the issue. In this paper, a framework is recommended for the detection of malicious network traffic.

III.METHODOLOGY OF PROPOSED SURVEY

This section presents the background and examines current literature that would clear up the picture for the reader about the design of the experiments conducted in this paper. Firstly, we discuss IDS including the use of ML used in attack detection and the related work which would help with selecting the algorithms to be used as well as identifying any datasets that could be utilized for testing the models. Each algorithm is explored with further research into the suitability of the algorithm for use in an IDS. The IoT is also described including the attacks that are used in the dataset that has been selected. An IDS is a tool that allows a network to be monitored for potentially harmful traffic. An IDS can be implemented using two distinct types: signature-based detection and anomaly-based detection. A signature-based IDS uses a database of existing attack signatures and compares the incoming traffic with the database, meaning that an attack can be detected only if the signature is already available in the database. An anomaly-based IDS monitors network traffic and attempts to identify any traffic that is abnormal in regards to the normal network traffic. The signature-based detection approach has a major flaw as a signature-based IDS will always be susceptible to a zero-day attack or an attacker that modifies the attack to hide from the signature database. Anomaly-based IDS are much better suited to use ML as the IDS can be trained to detect the difference between normal

traffic and attack traffic. ML is a subset of AI that involves giving an algorithm or in this case a model a dataset which will be used to identify patterns that can be used to make predictions with future data. There has been limited research devoted to IDS using ML on IoT networks.

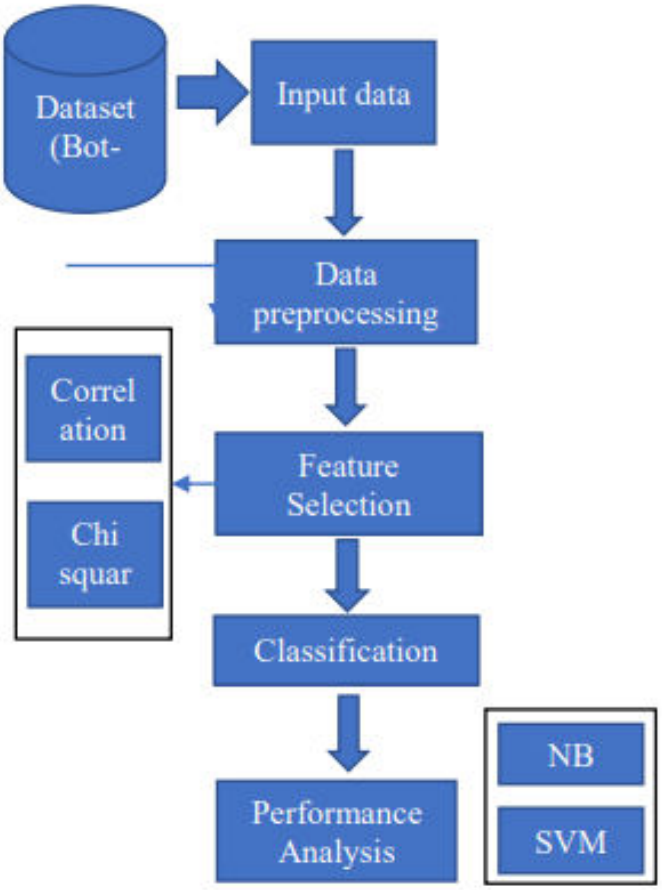


Figure 1. System Architecture

The primary purpose of employing machine learning in IoT security is to enhance the overall security posture of IoT systems. IoT devices are often resource-constrained and may lack sophisticated security mechanisms. Machine learning algorithms can help in detecting anomalous behaviors or patterns indicative of attacks, thereby bolstering the security of these systems.

Machine learning algorithms excel at detecting anomalies or deviations from normal behavior in large datasets. In IoT environments, anomalies could signify potential security threats such as malware infections, unauthorized access attempts, or denial-of-service attacks. By continuously analyzing data from IoT devices, machine learning models can identify suspicious activities in real-time.

Machine learning algorithms can learn patterns from historical data and use this knowledge to identify known attack patterns or signatures. This capability is particularly useful for recognizing common types of attacks, such as Distributed Denial of Service (DDoS) attacks, Man-in-the-Middle (MitM) attacks, or data exfiltration attempts targeting IoT devices. One of the key advantages of machine learning-based security solutions is their ability to adapt to evolving threats. As attackers develop new techniques and strategies, machine learning models can be retrained using updated datasets to recognize emerging threats. This adaptability is crucial in the dynamic landscape of IoT security, where new vulnerabilities and attack vectors constantly emerge. Traditional rule-based intrusion detection systems may generate a significant number of false positives, leading to alert fatigue and decreased efficiency in incident response. Machine learning algorithms can help mitigate this issue by dynamically adjusting to the specific characteristics of IoT environments, resulting in more accurate detection and fewer false alarms. Machine learning techniques can be designed to analyze data while preserving the privacy of sensitive information. By leveraging techniques such as federated learning or differential privacy, IoT systems can benefit from the

insights provided by machine learning models without compromising the privacy of user data.

Machine Learning is the subfield of Artificial Intelligence. It is one of the best methods that can provide a result without being explicitly programmed. Machine learning has four big different techniques. Supervised Learning, Unsupervised Learning, Reinforcement Learning, and Semi-Supervised Learning. K-Nearest Neighbor (KNN) is the subfield of the supervised learning method. This algorithm works for regression and classification problems. The rule of this algorithm is similar things should have similar characteristics. When the algorithm classifies all characteristic things, it should be close to proximity. This algorithm estimates the distances between input value and classified value. K refers to when we classify the new data, algorithm looks to classified data for examine the closeness between these data. KNN is extremely easy to implement in its most basic form, and yet performs quite complex classification tasks. It is a lazy learning algorithm since it does not have a specialized training phase. Rather, it uses all data for training while classifying a new data point or instance. KNN is a nonparametric learning algorithm, which means that it does not assume anything about the underlying data. This is an extremely useful feature since most of the real-world data does not really follow any theoretical assumption e.g., linear separability, uniform distribution, etc. The intuition behind the KNN algorithm is one of the simplest of all the supervised machine learning algorithms. It simply calculates the distance of a new data point to all other training data points. The distance can be of any type e.g., Euclidean or Manhattan, etc. It then selects the K-nearest data points, where K can be any integer. Finally, it assigns the data point to the class to which most of the K data points belong. Figure 1 shows KNN algorithm.

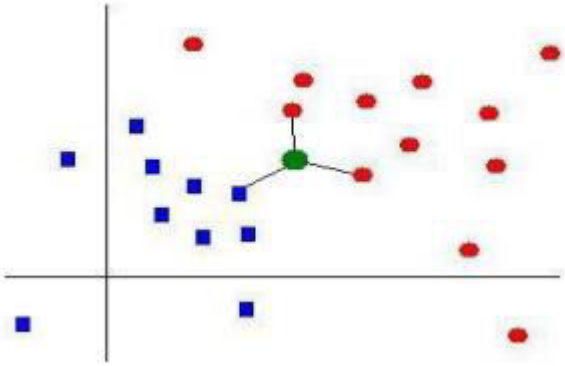


Figure 2. K-Nearest Neighbor Algorithm

Logistic Regression is a supervised machine learning algorithm from the field of Statistics. It is the method for binary classification problems. Logistic Regression is named for the function used at the core of the method known as logistic function. This function is also called sigmoid function. Logistic Regression is known as logit regression, maximum-entropy classification, or the log-linear classifier. In this model the probabilities describing the possible outcomes of a single trial are modeled using logistic function or sigmoid function. A logistic function f(x) or logistic curve is a common "S" shape (sigmoid curve) in Figure 2. Where f(x) is the prediction probability and x are the input data, f(x)=1/1+e^-(x)

Random Forest is a supervised machine learning method that operates by constructing multiple decision trees. A decision tree is a tree-shaped diagram used to determine the output. Figure 3 shows that each branch gives the output of individual possible outcomes. The base estimator of this model is the decision tree. Each estimator is trained on a different bootstrap sample having the same size as the training set. The Random Forest model introduces further randomization on features in the training of individual trees. Random Forest algorithm is used for both classification and regression problems. Classification aggregates predictions by the majority votes. And regression aggregates predictions by averaging.

The evaluation of the results of the implementation of the algorithms has been calculated according to the confusion matrix. Size of the matrix is depending on the number of the class in the testing dataset. There are a couple of criteria terms produced by the confusion matrix. Those criteria's names are respectively precision (Pr), recall (Rc), and accuracy.

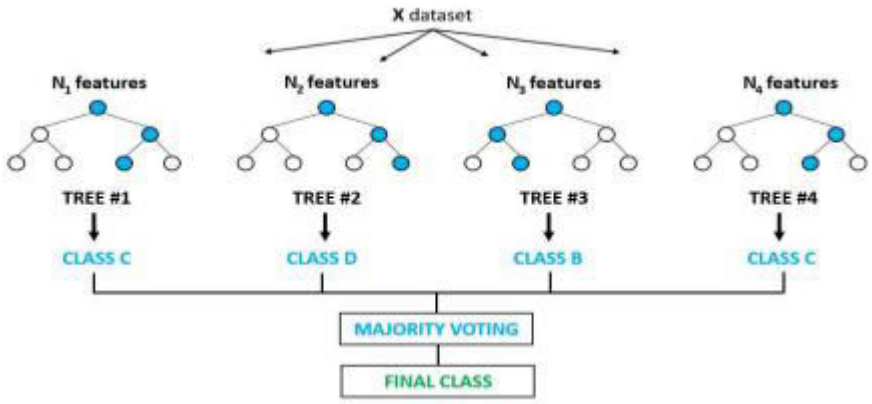


Figure 3. Random Forest Model

Support Vector Machine (SVM) is a supervised learning algorithm that uses a hyperplane to separate the training data to classify future predictions. The hyperplanes divide a dataset into two classes and they are decision boundaries that help classify the data points. A hyperplane can be represented as a line or a plane in a multi-dimensional space and is used to separate the data based on the class they belong to. Other more recent research compared SVM and ANN's ability to classify attack data as previously mentioned, SVM relies on placing a hyperplane to separate data which can be expressed as follows:
ax + b = 0
where a is the vector of the same dimensions as the input feature vector x and b is the bias. In this case, ax can be written as a 1x 1 + a 2x 2 + ... + a nx n where n is the number of dimensions of the feature vector x. When making predictions, the following expression is used:
y = sign (ax − b)
where sign is a function that returns either +1 or −1 depending if the input is a positive number or a negative number respectively. This value is used to determine the prediction of what class the feature vector belongs to. xi is the feature vector and i and yi is the label that can either be +1 or −1 and can be written as the follows: axi − b ≥ +1 i f yi = +1. axi − b ≤ −1 i f yi = −1 SVMs use kernels and kernel is basically a set of mathematical functions.

Decision Tree is a supervised learning algorithm that is useful to present a visual representation of the model. A DT uses a hierarchical model that resembles a flow chart which has several connected nodes. These nodes represent tests on the attribute in the dataset with a branch that leads to either another node or a decision on the data being classified. This research showed that the model that used DT was the best model with high accuracy and a low false positive rate. As previously mentioned, DT creates a hierarchical model using the training data to create nodes that act as tests for making predictions. When making DT, the root node needs to be selected as well as selecting the nodes that make up the DT. In this regard, there are many ways to do this with entropy being used in this case. Entropy is used to measure the probability of a data point being incorrectly classified when randomly chosen and is expressed as follows:

$$E = \sum_{i=1}^{c} -p_i \log_2 (p_i)$$

where pi is the probability of the data being classified to a given class of i and c is the number of classes. The attribute with the lowest entropy would be used for the root node.

IV.CONCLUSION AND FUTURE WORK

In this technology age, it shows us the electronic devices which relate to the internet, it covers human life. The data science world is developing new techniques for the security of those devices. When we look for the collection of data in networks, IoT devices are closing to the computer systems. This study aims for improving the security level of IoT networks. This study reached our goals, we detected most of the attacks in our dataset. Data science will grow up with our next research and articles. Data is very important, so we always try to secure all types of data. In the future, we will research to apply Deep learning and Reinforcement learning models for bigger datasets to get better accuracy result in detecting attacks. To identify that traffic accurately, NSL KDD dataset is critically evaluated by making use of ML techniques. This dataset is used for the comparison of the given framework by employing functions such as selection and classification. Overall, the RF algorithm

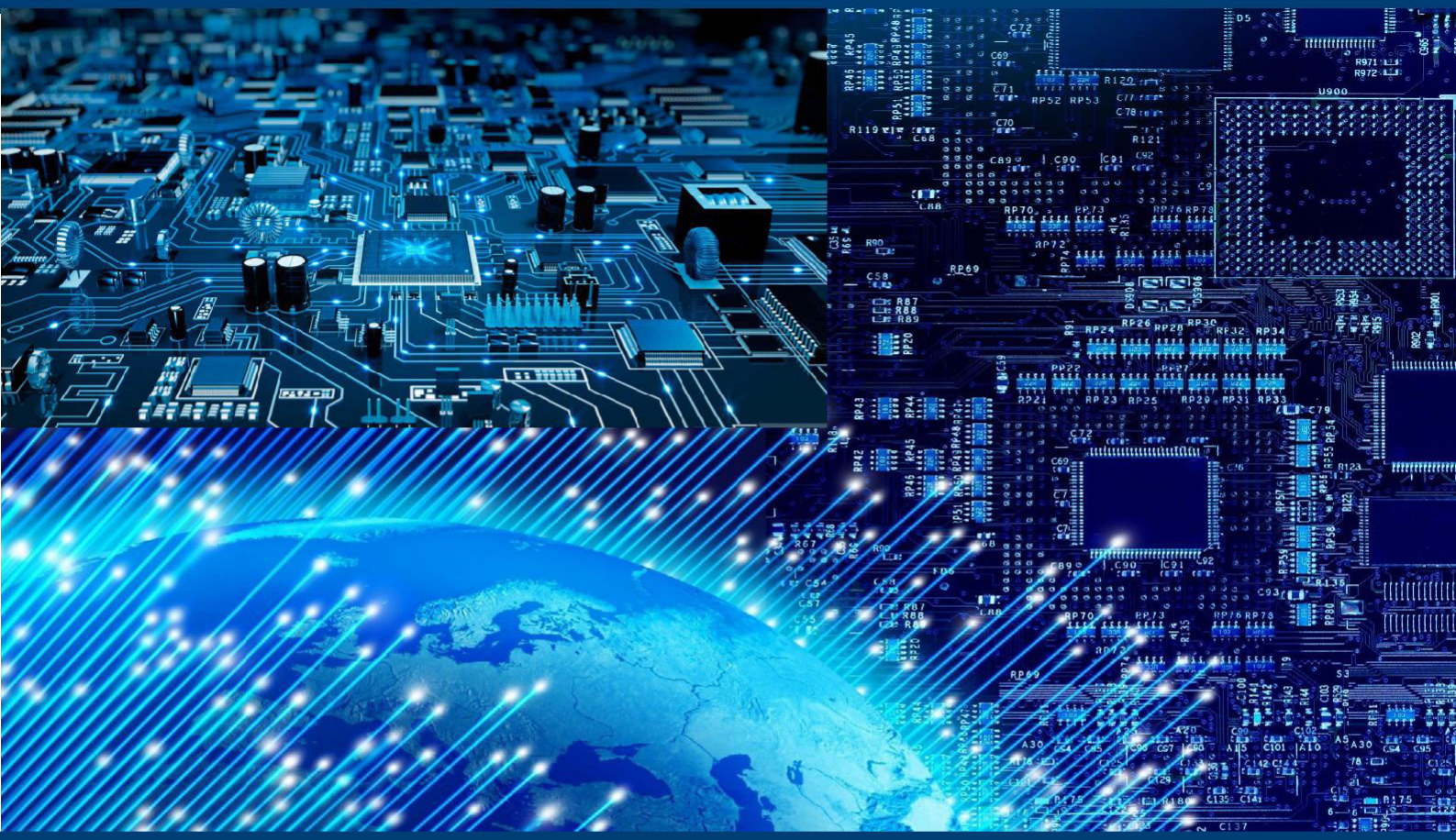
provided the best accuracy of 85.34% on the fog layer in comparison with the other two learning algorithms. In the future, it is planned to analyze different IoT devices, explore further technologies and, testing with different data of IoT devices infected by malware and cyber-attacks.

It would be useful to assess the efficacy of various unsupervised algorithms in upcoming research. We also used many machine learning techniques, each in isolation. In the future, we hope to enhance the detection performance by combining many machine learning methods into a multi-layered model.

REFERENCES

1. Mahmudul Hasan∗, Md. Milon Islam, Md Ishrak Islam Zarif, M.M.A. Hashem "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches" M. Hasan, Md. M. Islam and Md. I.I. Zarif et al. / Internet of Things 7 (2019) 100059

2. Md. Rumman Rafi, Mohammad Abdus Salam "Machine Learning Based Attack Detection in Internet of Things Network" International Journal of Computer Science and Information Security (IJCSIS), Vol. 19, No. 8, August 2021
3. Maryam Anwer, Shariq Mahmood Khan, Muhammad Umer Farooq, Waseemullah "Attack Detection in IoT using Machine Learning" Engineering, Technology & Applied Science Research Vol. 11, No. 3, 2021
4. Kalaiselvi S, Hariharasudhan R, Divakar A, Mytheeswaran V, Niveditha TP "detection of malicious attacks in IOT network traffic using machine learning techniques" Detection of Malicious Attacks in IOT Network Traffic Using Section A-Research paper Machine Learning Techniques 2023.
5. M.-O. Pahl, F.-X. Aubet, All eyes on you: distributed multi-dimensional IoT microservice anomaly detection, in: Proceedings of the 2018 Fourteenth International Conference on Network and Service Management (CNSM)(CNSM 2018), 2018. Rome, Italy
6. C.C. Aggarwal, J. Han, J. Wang, P.S. Yu, A framework for clustering evolving data streams, in: Proceedings of the Twenty-ninth International Conference on Very Large Data Bases-Volume 29, VLDB Endowment, 2003, pp. 81–92
7. O. Brun, Y. Yin, E. Gelenbe, Y.M. Kadioglu, J. Augusto-Gonzalez, M. Ramos, Deep learning with dense random neural networks for detecting attacks against IoT-connected home environments, in: Proceedings of the 2018 ISCIS Security Workshop, Imperial College London. Recent Cybersecurity Research in Europe. Lecture Notes CCIS, in: 821, 2018.
8. M. Roopak, G. Y. Tian and J. Chambers, "An Intrusion Detection System against DDoS Attacks in IoT Networks," 2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 2020.
9. N. K. Sahu and I. Mukherjee, "Machine Learning based anomaly detection for IoT Network," International Conference on Trends in Electronics and Informatics (ICOEI 2020), 2020
10. R. Doshi, N. Apthorpe and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," 2018 IEEE Symposium on Security and Privacy Workshops, 2018.



International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| Mobile No: +91-9940572462 | Whatsapp: +91-9940572462 | ijarasem@gmail.com |