International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Real Time Security Enhancement for IOT Enabled Intelligent Network

Richa Singhai[1], Rama Sushil[2]

Submitted:12/02/2023 **Revised:**15/04/2023 **Accepted:**09/05/2023

Abstract: The Internet of Things (IoT) will continue to have an increasing impact on our economic, commercial, and social lives. Because they frequently have limited resources, IoT network nodes become appealing targets for hackers. In order to address the security and privacy challenges that IoT networks face, a lot of work has been put forth, primarily using conventional encryption approaches. IoT networks have a number of security issues, but current solutions are limited by the characteristics of IoT nodes. It can be difficult to have secure and private conversations due to the Internet of Things' (IoT) extensive use and implementation. Different security concerns in IoT networks and devices can be addressed using machine learning (ML) techniques. This proposal offers a comprehensive examination of the security requirements for IoT networks, potential attack vectors, and current security measures. In order to address new security issues in cyber-physical systems (CPS), a number of research directions have been investigated. One of these directions is machine learning (ML), which has been hailed as the most cutting-edge and promising strategy. It addresses the most recent developments in machine learning techniques to address IoT device security challenges, describes IoT system designs, and analyzes various IoT system assaults. The use of machine learning techniques to offer decentralized privacy and protection was recently proposed. This proposal also discusses potential research challenges brought on by IoT devices' potential use of security measures in the future.

Index Terms: *Machine Learning (ML), IoT Applications, Privacy and protection, Internet of Things (IoT) System,*

1. Introduction

Securing them gets harder as more connected gadgets hit the market. Protecting these new, expansive networks that overburden network security personnel is the responsibility of organizations and companies around the world. To uncover threats, these cybersecurity experts must filter through large lakes of host and network-based information. These rules include limitations in their detecting abilities, a high rate of false positives, and reliance on the knowledge and competence of human authors.

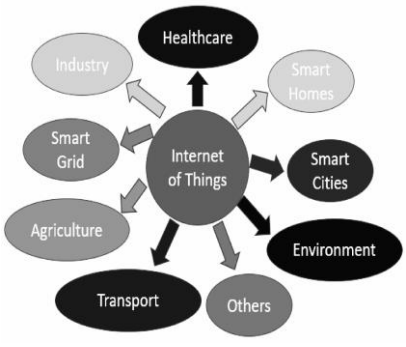

Fig 1: Biometrics of IoT

The Internet of Things (IoT) is understood to be a distributed, interconnected, and wirelessly or wired-connected network of embedded systems [1]. A physical thing or network of things having constrained computation, storage, and communication capabilities is another way to characterize it. Data can also be sent and received thanks to electronics (including sensors and actuators), software, and network connections. When it refers to "things," we mean commonplace items like IP cameras, smart light bulbs, smart adapters, smart meters, smart refrigerators, smart ovens, air conditioners, temperature sensors, and other smart household appliances. point. equipment like radio frequency identification (RFID), accelerometers, parking sensors, heart rate monitors, and other sensors found in cars are in figure 1.

ML can offer completely new and more dependable detection techniques [2-5]. Without constant human supervision, ML models can recognize attacks and take the necessary measures. In addition, the network is constantly receiving hundreds of millions of network connections and records, making it possible to use it to address the big data issue at the root of cybersecurity worries. In order to do this, ML for cybersecurity has been the subject of extensive research. Many researchers train and test their algorithms on publicly accessible datasets in order to examine different machine learning techniques [6]. Robots and smart devices can deduce pertinent information from device- or human-generated data with the use of machine learning (ML). This is also known as the capability of intelligent devices to automate or adjust circumstances and actions depending on knowledge. It is regarded as a crucial element of IoT systems. Applications including classification, regression, and density

[1]*Dept of CSE, DIT University, Dehradun, India*
[2]*Dept of CSE, DIT University, Dehradun, India*
E-mail address: btirtjs@gmail.com, rama.sushil@dituniversity.edu.in

estimation have all been used with ML techniques. Because many of its components have significance only on particular networks, cybersecurity data is distinctive [7-11].

2. Related Work

The authors' Collaborative Intrusion Detection System (CIDS), published in Arkady et al. (Joint Intrusion Detection to Secure IoT and Cloud Networks Using Deep Blockchain Frameworks), will be implemented in 2021. Enterprise adoption of numerous cloud infrastructures is increasing and Challenge is motivating our proposal [12-15]. It can be challenging to recognize assaults that affect many infrastructures. To guarantee the confidentiality and privacy of data delivered to the author's CIDS from the cloud or IoT infrastructure, the author suggests employing blockchain technology. For its own part, CIDS deployed BiLSTM RNNs to spot threats in the data.Model training and testing were conducted using the Bot-IoT dataset and its precursor UNSW-BN15 [16]. As opposed to SVM, RF, NB, and MLO models, which do not include training or test data are in figure 2.

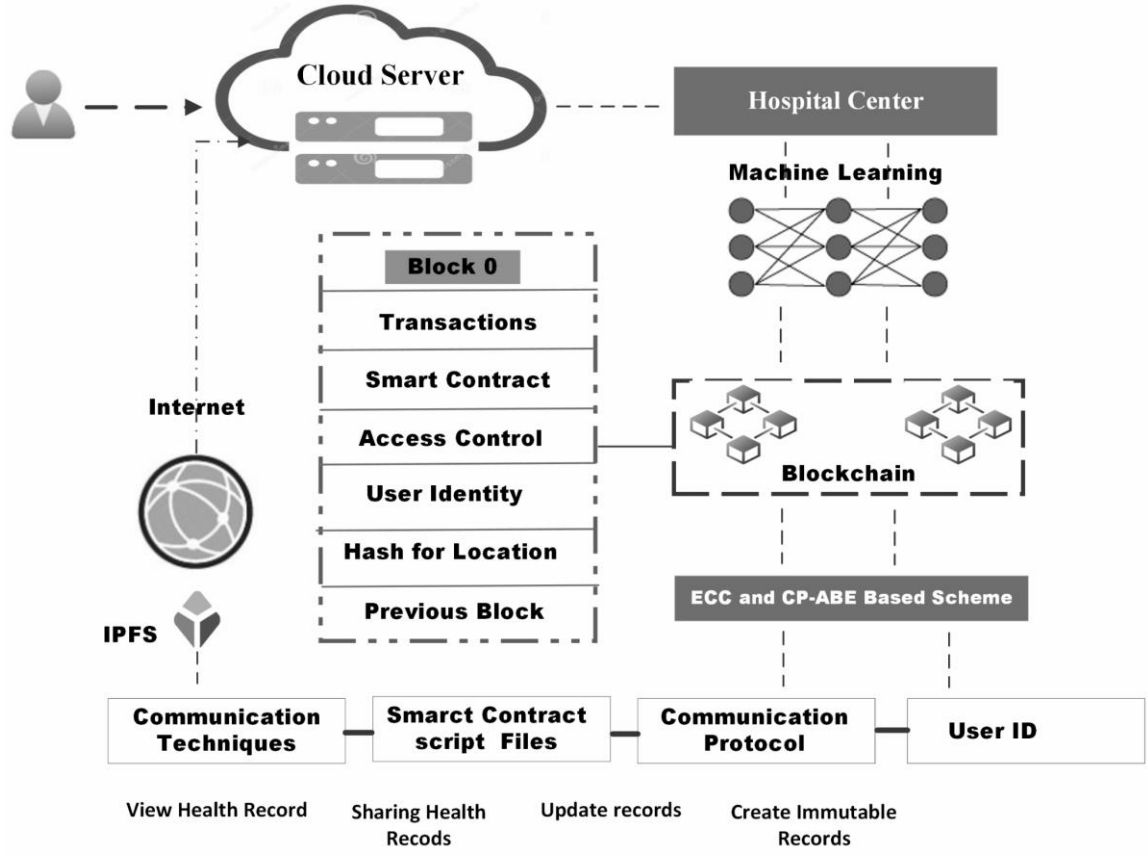

Fig 2: Blockchain based IoT Security

Blockchain is utilized in the Internet of Things to recognise a variety of machine learning-based attacks. Chema et al. created his decentralized blockchain-based IDS [16]. Make use of a lot of distributed networks. In his experiments, the researcher employed MATLAB, Python, and Ethereum. The accuracy and false positive rate of BiLSTM RNN, according to the authors, surpasses those of all other models. The ML model applied in this experiment is SVM. In our binary testing, SVM did quite well.

Lawal et al. created a framework for assaults based on fog computing, cloud-based monitoring [17-19]. The computational load of IoT devices is carried by their design, which makes use of edge computing capabilities offered by the "nebula." They experimented with the Extreme Gradient Boosting (XGBoost) algorithm for the system's binary and multiclass classification [20–22]. The author's 10 updated algorithms and the entire Bot-IoT 7 dataset. With 98.1% and 93.4% F1 results, respectively, excellent results for binary and multiple classification were achieved.

3. Proposed Methodology

The model needs to have training data and rule-making capabilities in order to go beyond the range of available datasets. Understanding the characteristics present in the dataset and their implications is crucial for this goal. All features in the BoT IoT dataset will be defined, used, and analyzed in this proposal. Additionally, some dataset characteristics that could be problematic given their use in the current investigation are emphasized. Figure 5 illustrates ML based malware prediction.

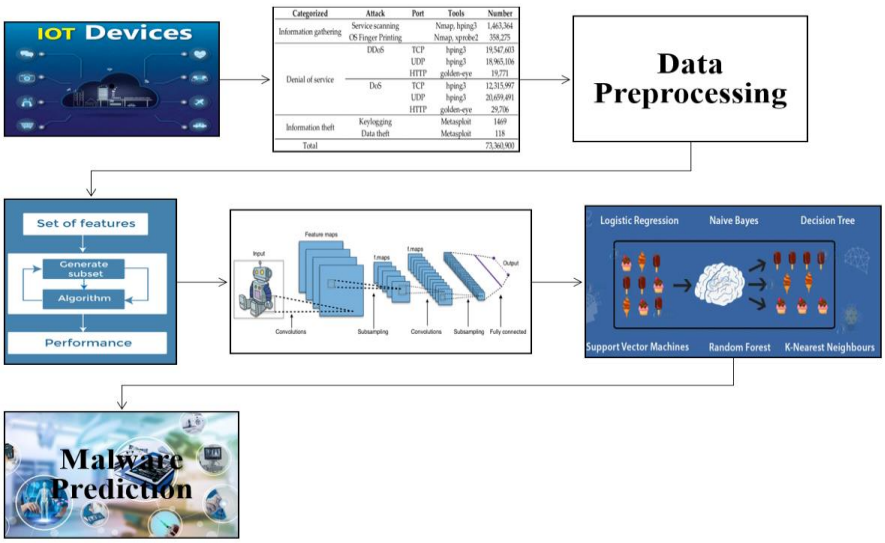

Fig 5: Malware prediction based on machine learning

As a result of the use of machine learning, malware can be identified in IoT devices by the following steps:

- The NSL-KDD dataset should be imported first, and then it should be divided into categories (training and testing).
- By using the EDA process, analyze the features from the dataset that are relevant to the model, such as protocol type, service, flag, and attack distributions.
- A CNN model with different layers and activation functions can be created
- The use of machine learning to identify malware in IoT devices

A. Dataset and Data Preprocessing

The total number of records in the BoT IoT data collection is 15,000 (80% training records, 20% testing records). The epoch here represents the number of times the loop has finished. The complete collection of data cannot be given to a neural network simultaneously. After that, a stack is created using the practice data set. The BoT IoT dataset, which includes all pertinent information from the BoT IoT dataset, is one of the most frequently used datasets for assessing network intrusion detection systems as shown in figure 6.

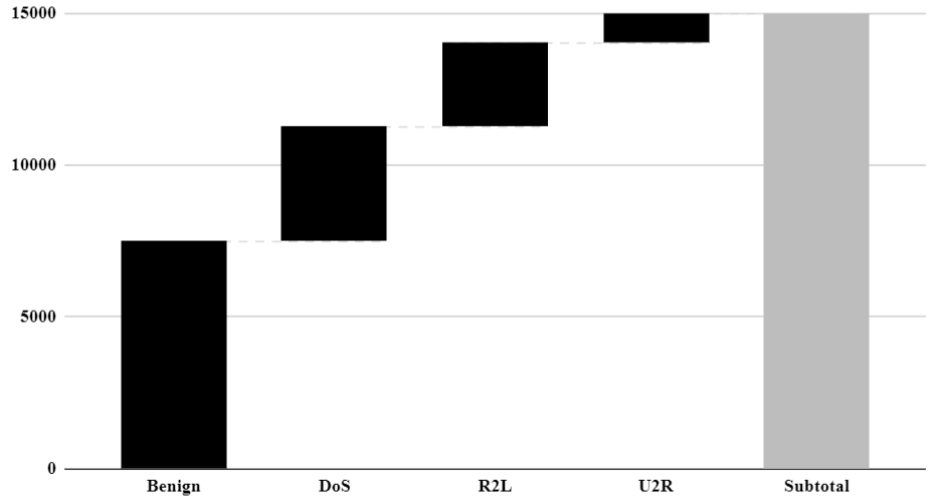

Fig 6: Attack categories for the NSL-KDD dataset

According to the statistics of the main BoT IoT dataset, the number of datasets chosen at each degree of difficulty is inversely correlated. Attacks such as user-to-root, remote-to-local, denial-of-service, and others are included in this data collection (DoS). In this dataset, feature types may be categorized into four different groups. Ten continuous, 23 discrete, six binaries, four categorical, and six binary categories. Built-in, content-based, host-based, and time-based IDS traffic logging fall into these four categories.

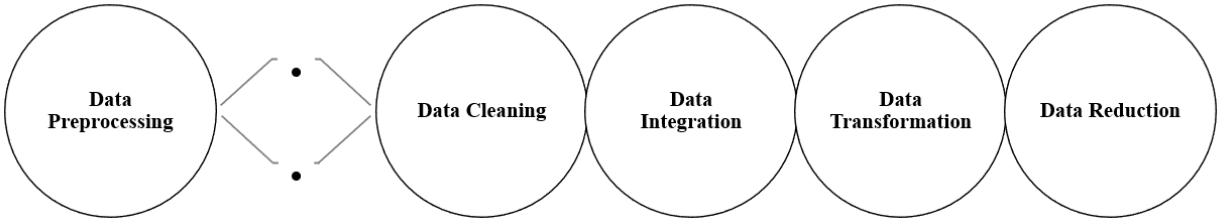

Fig 7. Data Preprocessing

The data had to be processed after gathering the information from the dataset depicted in figure 7. The ideal model feature is applied in this circumstance. The most critical attributes are chosen and included in the model via feature selection. Choosing optional features is made easier as a result. While reducing overfitting, it can increase performance.

B. Feature Extraction

- ***Dependent Features***

The three dependent properties of both IoT records are category, subcategory, and attack characteristics. Every sentence and sub sentence aside from raw sentences contains all three elements. There are five potential possibilities for the "Category" string feature: DDoS and DoS

Typically, theft and spying. These figures show the kind of attack for a specific event. To categorize cases with more precise attack kinds, combine these subcategory values with category values. It's crucial to keep in mind that the values for the DoS category and DDoS category subcategory are identical.

- ***Independent Features***

There are 42 distinct features in the entire Bot-IoT dataset. The experts at Argus Security have compiled a list of the top 25 features. They can either be directly generated from the set's attributes or they can be included in the whole set. These properties are only present in the 5% and 10% subsets and must be calculated using a script. Features are deemed invalid if they cannot produce a generalizable model during analysis.

The EDA methodology is used for data analysis. Users can examine the dataset and make inferences regarding potential patterns and outliers. EDA is a technique for examining how data affects a model. Attack Distribution, Distribution, Protocol Types, Services, Flags, and Distribution all require EDA. Disabled, Unknown, and Enabled are the three main classifications for both IoT functionalities. Based on the set or subset of IoT features used in the study, as well as the features themselves, studies are classed and separated into subsections.

C. Convolutional Neural Network (CNN)

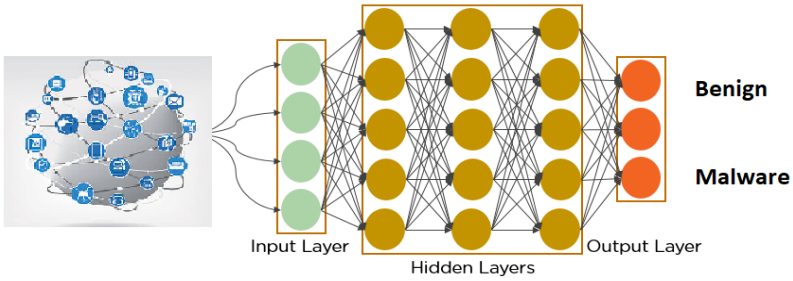

Fig 8. CNN

As seen in Figure 8, CNN classifies the data automatically and offers a better categorization in this instance. The features that the CNN pulls from the input data set are categorized by a second neural network. The feature extraction network uses a number of input data sources. The neural network categorizes the received feature signal using this data. The output layer of the network also includes three average pooling layers, a convolutional layer, and a fully connected softmax. In this instance, they convolve the convolution kernel with the input layer using the Conv1D layer of CNN to create an output tensor with a single spatial dimension.Weighted averages of the inputs are computed by the neurons in the layers of the neural network, and the results are propagated through nonlinear functions.

D. Classifier

The way that various algorithms process data varies. For instance, there are classification algorithms that can handle massive volumes of input data.Classification is a crucial component of supervised learning. The classification methods employed in this proposal include

Logistic Regression, Naive Bayes, Decision Trees, Random Forests, Support Vector Machines, and KNNs.

- ***Logistic Regression***

The probability of a target variable can be predicted using logistic regression, a supervised learning classification technique. It is one of the few ML strategies utilized to tackle different categorization issues, such as the prediction of malware. Although applying and analyzing logistic regression is significantly simpler, coaching can be highly beneficial. It divides the data points into two groups using logistic regression. The output class to which a given output (1 or 0) belongs is determined using the categorical categorization presented in Figure 9.

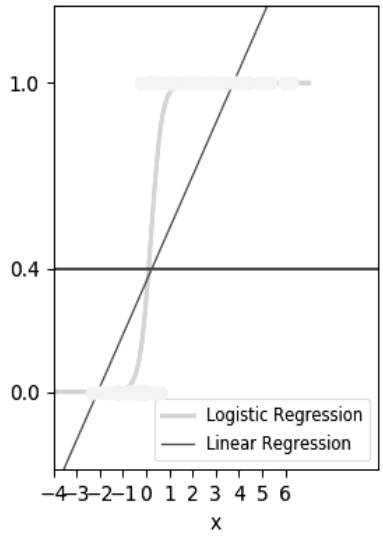

Fig 9. Regression

- ***Naive Bayes Algorithm***

It is used to resolve classification issues, and can be thought of as a supervised learning method owing to Bayes' theorem. This is a collection of algorithms that all operate according to the same fundamental principles, not just one. In other words, every pair of classified features is distinct. One of the best and simplest classification techniques for building rapid machine learning models that can predict outcomes is the naive bayes classifier. Naive Bayes is one of the most effective machine learning approaches for classification. The Bayes theorem is expanded to take into account the independence of each attribute. It is employed for many different things, including spam detection and text classification.

- ***Decision Tree Algorithm***

It is a method which is used in machine learning for both categorization and prediction. Given a set of inputs, a decision tree can be used to map various outputs that are results or outcomes of decisions. The end result of several hierarchical decision-making procedures is this decision tree. It uses two techniques to construct this tree: introduction and pruning. While induction grows the tree, pruning removes its different complexities.

- ***K-Nearest Neighbors Algorithm***

It is one of the most fundamental and significant categorization methods in machine learning. When nonparametric techniques are needed in real-world situations, these kNNs are applied. These algorithms don't assume anything about distribution.

- ***Support Vector Machine Algorithm***

It is a type of supervised learning, examining data for regression and classification. Although SVM can be used for regression, its primary application is classification and an n-dimensional spatial plot should be made. A coordinate value corresponds to each feature value. The ideal hyperplane is then identified to divide the two classes.

- ***Random Forest Algorithm***

It is a sort of ensemble learning technique, is used to conduct classification, regression, and other tasks that may be accomplished using decision trees. During training, these decision trees can be created using either classification or regression as the class outcome. It can stop overfitting the training set with the aid of these random forests.

E. Attack Prediction

The proposed model's accuracy and DR ratings demonstrate how well it can forecast attack trends across all classes. In addition to DR and ACC, a crucial indicator called FPR tracks the frequency of shared records labeled as attacks. A high FPR may not be trustworthy even if the DR and ACC match. Since precision and recall might not be sufficient on their own, the F1 score offers a more precise evaluation of performance. Matrix of confusion

displaying the values that the model thinks belong to a specific class. This enables conception of the model's presentation. The columns indicate actual grades and the rows reflect expected grades, hence they are N x N in size.

4. Results and Discussion

The suggested model clearly outperforms the opposition in every category, particularly in terms of detection rate. A variety of algorithms process data differently. One of the key components of supervised learning is classification. There are algorithms for handling massive quantities of input data, such as logistic regression, naive bayes, decision trees, random forests, support vector machines, and KNN. The model indicates whether the data represents an invasion or falls under the category of common binary classifiers. Figures 10 and 11 show the proposed model's results on binary classification of the BoT IoT dataset based on Accuracy, and Recall.

It measures how well a classifier can predict class labels and predict predictor attributes in fresh data. The predictor accuracy of a classifier measures the accuracy of its prediction. In order to determine which features enhance the prediction accuracy of the IoT dataset, Figure 10 outlines an accuracy analysis of the machine learning algorithms. When compared to other algorithms, the test algorithm's Random Forest method improves predictability accurately on IoT datasets.

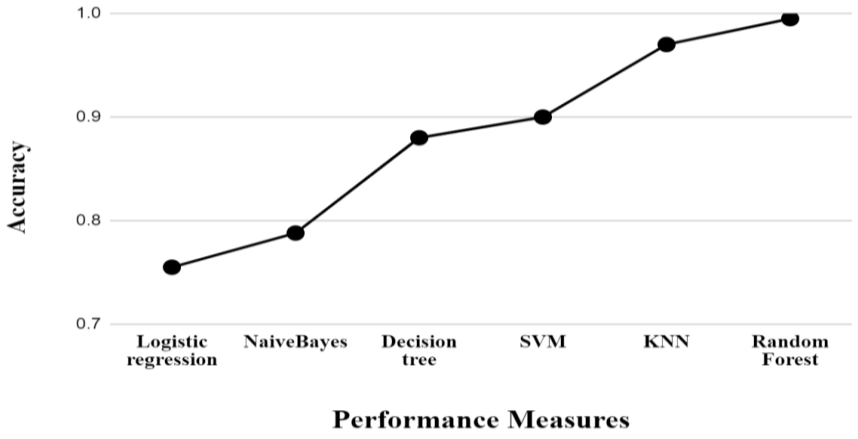

Fig 10. IoT Attack Prediction Accuracy Based on Classifiers

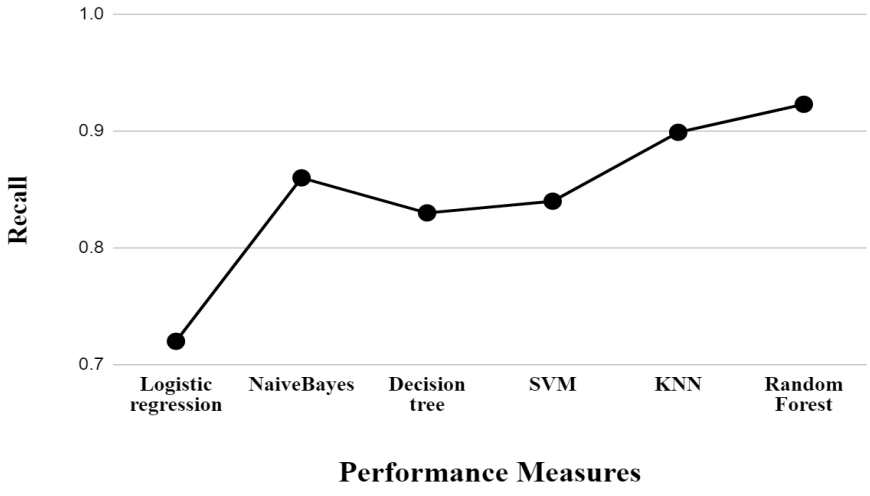

Fig 11. IoT Attack Prediction Recall Based on Classifiers

The recall, a metric, determines how many pertinent IoT detections are chosen. The analysis of recall measurements is shown in Figure 11. The outcomes demonstrate that the algorithm outperforms the experimental algorithm both qualitatively and numerically.

5. Conclusion

The IoT industry's expansion is reliant on IoT privacy and security. The dynamic nature of IoT networks poses a challenge to traditional security and privacy solutions. Using ML, IoT devices may react to configuration changes. By analyzing ambient statistical data, these

learning strategies can promote self-organizing behavior and improve system performance. These approaches to distributed learning don't need centralized controller-to-device communication. Using ML models to detect network breaches and assaults has already had a significant influence in the cybersecurity industry. Users must be familiar with a variety of computer operations linked to networks, operating systems, and cyber security data.

References

[1] "35.100 Open systems interconnection (OSI)," ISO, 04-Apr-2020. [Online]. Available: https://www.iso.org/ics/35.100/x/. [Accessed: 11-Oct-2021].

[2] A. Alhowaide, I. Alsmadi, and J. Tang, "Ensemble Detection Model for IoT IDS," Internet of Things, p. 100435, 2021.

[3] A. Alhowaide, I. Alsmadi, and J. Tang, "PCA, Random-Forest and Pearson Correlation for Dimensionality Reduction in IoT IDS," 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2020.

[4] A. Churcher, R. Ullah, J. Ahmad, S. ur Rehman, F. Masood, M. Gogate, F. Alqahtani, B. Nour, and W. J. Buchanan, "An Experimental Analysis of Attack Classification Using Machine Learning in IoT Networks," Sensors, vol. 21, no. 2, p. 446, 2021.

[5] Demirpolat, A. K. Sarica, and P. Angin, "ProtÉdge: A Few-Shot Ensemble Learning Approach to Software-Defined Networking-Assisted Edge Security," Transactions on Emerging Telecommunications Technologies, vol. 32, no. 6, 2020.

[6] Djenna, D. E. Saidouni, and W. Abada, "A Pragmatic Cybersecurity Strategies for Combating IoT-Cyberattacks," 2020 International Symposium on Networks, Computers and Communications (ISNCC), 2020. 97

[7] "About the Metasploit Meterpreter," Offensive Security. [Online]. Available: https://www.offensive-security.com/metasploit-unleashed/about-meterpreter/. [Accessed: 16-Oct-2021].

[8] "About Wireshark," Wireshark · Go Deep. [Online]. Available: https://www.wireshark.org/. [Accessed: 11-Oct-2021].

[9] Ahlashkari, "CICFlowmeter," GitHub. [Online]. Available: https://github.com/ahlashkari/CICFlowMeter. [Accessed: 03-Oct-2021].

[10] "Argus Manual Pages." [Online]. Available: https://openargus.org/oldsite/man/man8/argus.8.pdf . [Accessed: 05-Oct-2021].

[11] B. Susilo and R. F. Sari, "Intrusion Detection in IoT Networks Using Deep Learning Algorithm," Information, vol. 11, no. 5, p. 279, 2020.

[12] "Bare Metal hypervisor," VMware. [Online]. Available: https://www.vmware.com/topics/glossary/content/bare-metal-hypervisor. [Accessed: 16-Oct-2021].

[13] C. Kemp, C. Calvert, and T. M. Khoshgoftaar, "Detection Methods of Slow Read DOS Using Full Packet Capture Data," 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), 2020.

[14] B. B. Zarpelo, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in internet of things," Journal of Network and Computer Applications, vol. 84, pp. 25 – 37, 2017.

[15] M. binti Mohamad Noor and W. H. Hassan, "Current research on internet of things (iot) security: A survey," Computer Networks, vol. 148, pp. 283 – 294, 2019.

[16] K. T. Nguyen, M. Laurent, and N. Oualha, "Survey on secure communication protocols for the internet of things," Ad Hoc Networks,vol. 32, pp. 17 – 31, 2015. Internet of Things security and privacy:design methods and optimization.

[17] K. Sha, W. Wei, T. A. Yang, Z. Wang, and W. Shi, "On security challenges and open issues in internet of things," Future Generation Computer Systems, vol. 83, pp. 326 – 337, 2018.

[18] M. Tao, K. Ota, M. Dong, and Z. Qian, "Accessauth: Capacity-aware security access authentication in federated-iot-enabled v2g networks,"Journal of Parallel and Distributed Computing, vol. 118, pp. 107 –117, 2018.

[19] I. Farris, T. Taleb, Y. Khettab, and J. Song, "A survey on emerging sdn and nfv security mechanisms for iot systems," IEEE Communications Surveys Tutorials, vol. 21, pp. 812–837, Firstquarter 2019.

[20] M. at. el, "Machine Learning for Internet of Things Data Analysis: A Survey," Journal of Digital Communications and Networks, Elsevier,vol. 1, pp. 1–56, Feburary 2018.

[21] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in internet-of-things," IEEE Internet of Things Journal,vol. 4, pp. 1250–1258, Oct 2017.

[22] A. Chowdhury and S. A. Raut, "A survey study on internet of things resource management," Journal of Network and Computer Applications,vol. 120, pp. 42 – 60, 2018.