

Unauthorized Access Prevention Model for Secure Data Transmission in Internet of Things-Based Network

¹EMMAH Victor Thomas, ²DEEDAM Fortune Baribesia, ³OKWU Hachikaru Ngozi

^{1,2,3}Department of Computer Science, Rivers State University, Nigeria P.M.B 5080

Abstract

The Internet of Things (IoT) has revolutionized connectivity by enabling devices to interact seamlessly in various domains such as healthcare, smart cities, and industrial automation. However, this connectivity also introduces vulnerabilities, with unauthorized access posing significant threats to data security. This paper presents an unauthorized access prevention model for securing data transmitted in the internet of things-based network, with a focus on providing an Intrusion Detection System (IDS) that detects botnet attacks. The approach employs a Multi-Layer Perceptron (MLP) for fault detection, specifically designed to address the challenges inherent in IoT environments. The MLP architecture consists of three dense layers with Rectified Linear Unit (ReLU) activation functions, enabling the capture of intricate data relationships. Through sequential modelling, the MLP accurately identifies complex patterns indicative of botnet attacks. The training process of the MLP yields promising results, with significant accuracy and minimal loss values. The model achieves an accuracy of 99.99% for both training and validation (testing) data. Furthermore, the classification report highlights the model's exceptional precision and recall for classifying normal and IDS classes, demonstrating its reliability for detecting botnet attacks. To bolster security measures, the system incorporates encryption and access control mechanisms. Data encryption, utilizing the Fernet symmetric encryption algorithm, ensures the confidentiality of transmitted data. Access control mechanisms differentiate between authorized and unauthorized devices, enhancing overall security.

Keywords: *Internet of things, data transmission, Random forest, encryption, Access control*

1. Introduction:

Internet of things (IoTs) is a network of physical devices that are connected to the internet, all collecting and sharing data. Connecting up all these different objects and adding sensors to them adds a level of digital intelligence to devices that would be otherwise dumb, enabling them to communicate real-time data without involving a human being (Biru, 2018). IOTs are making the fabric of the world around us smarter and more responsive, merging the digital and physical universes. The growth in the number of IoT devices is definitely beneficial with a major transformation in the ways for carrying out everyday activities. For example, smart lighting could help in reducing your electric bill and energy consumption. In addition, the benefits of connected healthcare devices have been helping people in obtaining a better impression of their

health. However, the benefits introduce prominent risks with the number of growing devices. The growth in the number of connected devices in the IoT ecosystem can present issues for security in IoT by offering more entry points for cybercriminals and hackers. The concerns of security and the issues of privacy in IoT present considerable implications for different business and public organizations. The interconnectivity of networks in IoT introduces the accessibility from anonymous and untrusted online sources. Businesses have to work on enhanced security, especially for consumer-grade IoT-enabled solutions, to encourage customer trust in IoT. Furthermore, the importance of security and privacy in IoT is also clearly evident in the gradually increasing awareness of consumers regarding the privacy of their data, (Yang & Wetherall, 2006). Data in the IoT are routed in the

form of network packets, and are split into bits. Each packet travels through the Internet via series of checkpoints. These packets are responsible for delivering high speed Internet and they can travel through air or fiber containing the destination address and the request. The Internet relies on the hardware infrastructure. Whenever you transmit data via the Internet that information cannot just go somewhere; instead, it has to be broken down, directed, and reassembled through server, switches, and routers that make their part of the Internet working as it should. The security issues in IoT basically include; inadequate password Protection, limited compliance from IoT manufacturers, Hardware issues, Lack of security in data transfer and storage, Hard-coded, weak, or guessable usernames and passwords, device update management, lack of secure interfaces, privacy concerns in IoT, abundance of data, eavesdropping, unwanted public exposure. Given the non-standard manufacturing of IoT devices and troves of data flowing through the IoT devices, we are constantly exposed to cyber-attacks. Vulnerabilities, cyber-attacks, data theft, and other risks arising from the usage of IoT devices make the need for IoT security even more, hence the need to secure data transmission in an IoT-based network.

2. Literature Review

Several studies have been carried out to secure the IoT network and prevent unauthorised access. Hwang and Kim (2019) developed a hybrid encryption model combining ECC and AES to balance security and performance. Their approach showed significant improvement in securing real-time IoT data while maintaining low computational overhead.

Ahmed *et al.* (2021) proposed an ensemble-based intrusion detection system utilizing random forests and neural networks to identify unauthorized activities with high accuracy. Meanwhile, Kumar *et al.* (2023) demonstrated the effectiveness of federated learning in detecting attacks without compromising device privacy.

Li *et al.* (2018) proposed a decentralized authentication framework leveraging blockchain technology. This approach ensures tamper-proof and transparent device registration, eliminating the risk of a single point of failure.

Shams *et al.* (2020) introduced a biometric-based authentication protocol specifically for healthcare IoT, ensuring that only legitimate users with verified biometrics can access sensitive data.

Kumar *et al.* (2019) developed a hash-chain-based authentication protocol that reduces computational complexity while maintaining security, making it ideal for resource-constrained devices.

Hwang and Kim (2019) proposed a hybrid encryption scheme combining Elliptic Curve Cryptography (ECC) with Advanced Encryption Standard (AES). Their approach effectively balances security and performance, particularly for low-power IoT devices.

Zhang *et al.* (2021) explored homomorphic encryption for secure data aggregation in IoT. This method allows computations on encrypted data, preserving privacy without decrypting sensitive information.

Bhowmik *et al.* (2022) investigated lattice-based encryption techniques to secure IoT communications against emerging quantum computing threats.

Jun and Chi (2014) proposed an IDS for IoT systems based on Complex Event Processing (CEP) technology which is an emerging and efficient technology to filter and process real-time events. It is a good solution for large volumes of messages with low latency. Santos *et al.* (2019) detected the dangerous sinkhole attack on the routing services in IoT. They proposed Intrusion detection of SiNkhole attacks on 6LoWPAN for Internet of Things. It combines watchdog, reputation and trust strategies for detection of attackers. First, as a hierarchical structure, the nodes (grouped or separated) are classified as leaders. Then, the nodes can change role over the time based on network requirements. Each node monitors a number of transmissions performed by a superior node. If an attack is detected, an alert message is broadcasted and a cooperative isolation of the malicious node is performed. Their simulation results show sinkhole detection rate of 92% on 50 fixed nodes scenario and of 75% for 50 mobile nodes.

Prabavathy *et al.* (2018) proposed a novel fog computing based intrusion detection technique using Online Sequential Extreme Learning Machine (OS-ELM). The distributed security mechanism (guaranteed by the fog computing idea) respects interoperability, flexibility, scalability and heterogeneity aspects of IoT systems. The proposed system is composed of two major parts which include Attack detection at fog nodes which uses OSELM algorithm to detect intrusions in fog nodes and Summarization at cloud server, where detected intrusions are sent from the fog node to the cloud server in order to

have a general idea about the global security state of the IoT system.

3. Methodology

The proposed system architecture shown in figure 1 comprises of a hybrid algorithm namely, Random forest and AES Algorithm, which is intended in classifying the attack into different modules thereby providing solution to the data leakage and privacy problems caused by intruders in the traditional system called DDoS attacks and

spoofing or impersonation attack etc. The Random forest Algorithm is used to detect and classify IOT attacks faster. It records the signature of each attack present in the IoT devices; the data clustered and then predicts the type of the attacks since some of the attacks possess similar characteristics. The intelligent algorithm is able to extract the intruder profile, save and record the data and is stored in different nodes in a decentralized manner. The AES algorithm will be used to secure the data from unauthorized users.

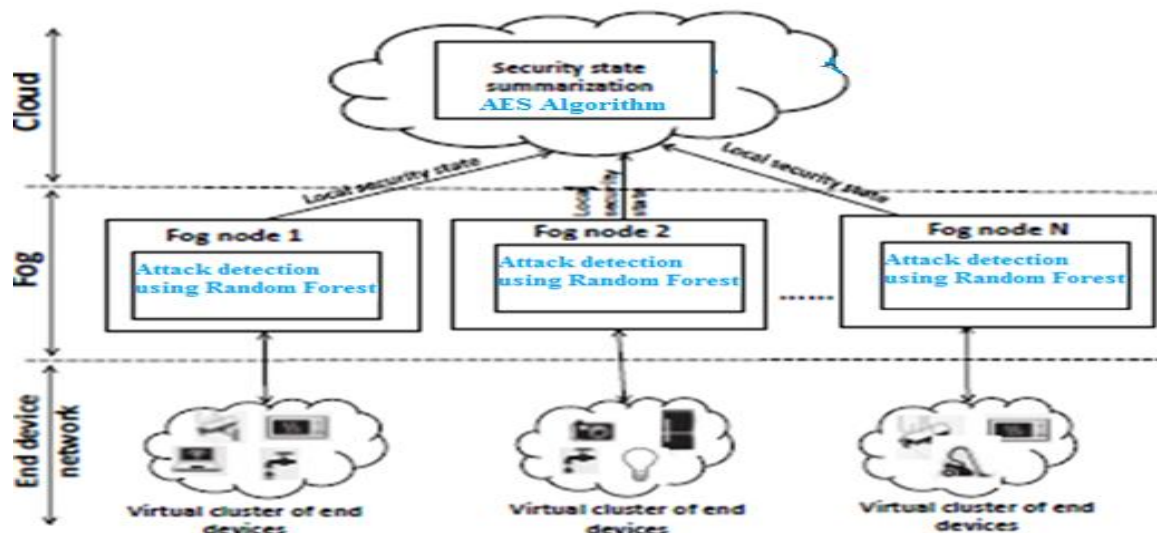


Figure 1: Architecture of the Proposed System

Figure 1 describes the activities in the proposed system where the end device network indicates the users sending data into the cloud for data storage using the, the fog nodes, indicates the storage point at which data is stored which an unauthorised users attacks the data stored in the cloud in form of malware where the summary of the results is kept for proper sanctions. Random forest algorithm was used to detect and classify the IOT attack from unauthorised users, thereafter, AES algorithm is used to verify the packets and secure the data.

4. Experimental Setup

The proposed system based its addition in the security state summarization. The random forest does the detection of the attacks and classifies the attacks in form of data and then stores the result (different attacks e.g., Probe, U2L, R2L, DOS, etc). After the classification is completed, the AES

algorithm is applied to the classified data to verify the security hash function that is attached to each threat, mined the threat and encrypt the data.

Dataset

NSL-KDD benchmark dataset for intrusion detection was used to during the system experiments. The dataset contains separate training and test records. The training set contains 125,973 records and the test set contains 22,544 records with 41 features. The dataset can be modeled for binary classification with 2-class and multi-class classification with 4-class. The major attacks found in IoT environment is matched with attacks in NSL-KDD dataset. Also the dataset used shows some attacks and its type based on the IOT environment. Table 1 illustrates result obtained, when random forest algorithm is applied to the dataset

Table 1: Dataset Sample with Random Forest Applied

Dataset	Total	Normal	DoS	Probe	R2L	U2R
KDD Train+	125,973	67,343	45,827	11,456	995	49
KDD Test+	25,192	13,449	9234	2289	209	11
KDD Test-21	22,542	12,709	7749	1867	175	42
UNSW-NB15 Train+	175,341	56,000	12,264	11,450	985	48
UNSW-NB15 Test+	82,332	37,000	4089	2012	201	14

The Random forest model saved once, so to use as reference for making prediction about new incoming packet. After the training phase is executed, the implementation phase, use the real time a packet arrived, it get analysed, verified, extracts its features and at last, feed forward, make a prediction about packet i.e. it is attacked or clean by referencing model.

Also the system built provide a pop-up notification in form of SMS and Email alert through which, the user get informed about IP address of the intruder. Due to this feature, user get aware about the source of attack and also be aware of not receiving any data from such source.

The AES algorithm works in the framework of public-key cryptosystems and is based on the

5. System Implementation And Results

In the implementation, Exploratory Data Analysis was carried out in extracting valuable insights from the IDS dataset by utilising visualisations; then, a Multi-Layer Perceptron (MLP) is applied on IoTs for for fault detection in the IoT devices. The final phase has to do with access control and data encryption.

Exploratory Data Analysis (EDA) is an essential initial step that allows for a comprehensive understanding of the data's features and establishes the basis for future modelling efforts. Figure 2 shows the countplot of the proto column, while figure 3 shows the correlation matrix of numerical features. The correlation matrix shows the relationship between features of the dataset. Finally, random forest classifier was used to perform a ranking on the dataset features. The ranking of the features can be seen in Table 2 and Figure 4.

algebraic properties of modular exponentiation, together with the discrete logarithm problem, which is considered to be computationally intractable. The algorithm uses a key pair consisting of a public key and a private key. The private key is used to generate a digital signature for a message, and such a signature can be verified by using the signer's corresponding public key. The digital signature provides message authentication (the receiver can verify the origin of the message), integrity (the receiver can verify that the message has not been modified since it was signed) and non-repudiation (the sender cannot falsely claim that they have not signed the message).

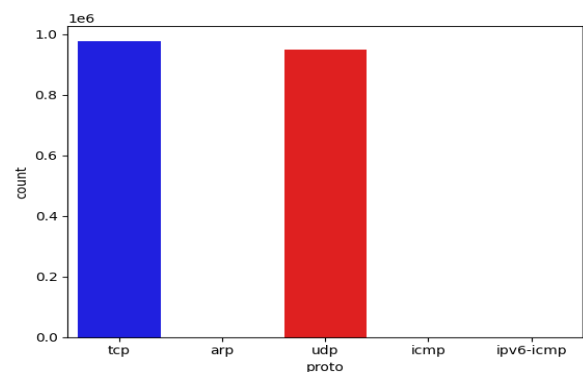


Figure 2: Countplot of the proto column

The countplot indicates that TCP and UDP are the most common protocols in the dataset, as they have the maximum number of bars, demonstrating their prevalence compared to other protocols. This knowledge is useful for comprehending the structure of network traffic or examining particular communication patterns in the dataset.

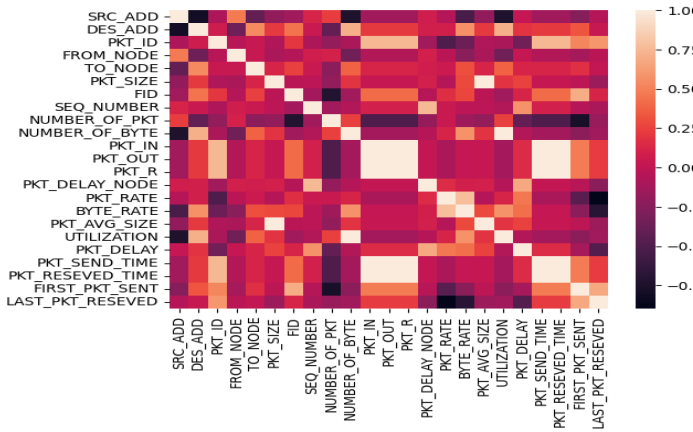


Figure 3: Correlation Matrix of the dataset features

Table 2: Feature Ranking

Features	Important_Features
time	0.235698
N_IN_Conn_P_DstIP	0.204543
ltime	0.144103
pkSeqID	0.115731
TnP_PDstIP	0.064887
seq	0.03169
TnBPDstIP	0.030107
AR_P_Proto_P_SrcIP	0.028172
stddev	0.02744

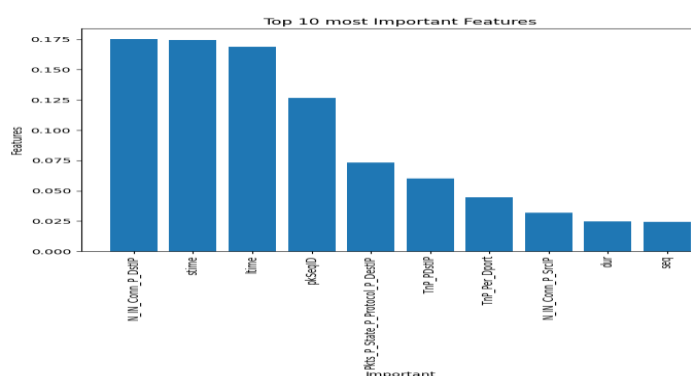


Figure 4: Histogram of ten most important features

The histogram depicts the first ten most important features. The feature with the highest bar signifies the most important feature while the feature with the lost bar represents the least important feature.

a. Implementation of MLP for Detecting IDS Botnet attacks

In the implementation of a Multi-Layer Perceptron (MLP) for fault detection in the IoT devices, a sequential model is employed with three dense layers. The first layer consists of 50 units, utilizes the rectified linear unit (ReLU) activation function, and takes the input shape derived from the flattened training data. The second hidden layer also comprises 50 units with ReLU activation. The final layer, with 5 units and a softmax activation function, serves as the output layer, representing the five distinct fault classes. The model is compiled using the Adam optimizer, categorical crossentropy loss function, and accuracy as the evaluation metric. This architecture is designed to capture complex relationships within the input data and detect botnet attacks accurately on IoT devices. The training process of the MLP can be seen in Table 3. Figure 5 shows the accuracy of the MLP model for both training and testing, and Figure 6 shows the loss values of the MLP model for both training and testing. Figure 7 shows the classification report of the MLP model, and Figure 8 shows the confusion matrix for the MLP model.

Table 3: MLP training Steps For Botnet IDS Detection on IoT Devices

```

Epoch 1/10
24/24 [=====] - 2s 24ms/step - loss: 0.3425 - accuracy: 0.870
2 - val_loss: 0.1195 - val_accuracy: 0.9895
Epoch 2/10
24/24 [=====] - 0s 9ms/step - loss: 0.0576 - accuracy: 0.9987
- val_loss: 0.0298 - val_accuracy: 1.0000
Epoch 3/10
24/24 [=====] - 0s 8ms/step - loss: 0.0162 - accuracy: 1.0000
- val_loss: 0.0112 - val_accuracy: 1.0000
Epoch 4/10
24/24 [=====] - 0s 7ms/step - loss: 0.0072 - accuracy: 1.0000
- val_loss: 0.0062 - val_accuracy: 1.0000
Epoch 5/10
24/24 [=====] - 0s 10ms/step - loss: 0.0043 - accuracy: 1.000
0 - val_loss: 0.0040 - val_accuracy: 1.0000
Epoch 6/10
24/24 [=====] - 0s 11ms/step - loss: 0.0029 - accuracy: 1.000
0 - val_loss: 0.0028 - val_accuracy: 1.0000
Epoch 7/10
24/24 [=====] - 0s 9ms/step - loss: 0.0021 - accuracy: 1.0000
- val_loss: 0.0021 - val_accuracy: 1.0000
Epoch 8/10
24/24 [=====] - 0s 8ms/step - loss: 0.0015 - accuracy: 1.0000
- val_loss: 0.0017 - val_accuracy: 1.0000
Epoch 9/10
24/24 [=====] - 0s 8ms/step - loss: 0.0012 - accuracy: 1.0000
- val_loss: 0.0013 - val_accuracy: 1.0000
Epoch 10/10
24/24 [=====] - 0s 7ms/step - loss: 9.7540e-04 - accuracy: 1.
0000 - val_loss: 0.0011 - val_accuracy: 1.0000
CPU times: total: 4.3 s
Wall time: 4.18 s

```

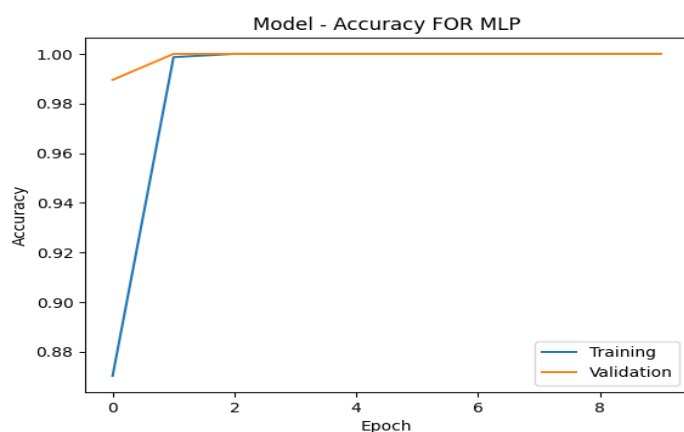


Figure 5: Accuracy of the MLP model for both Training and Testing

The accuracy, demonstrates how well the model performed during training. This shows that model achieved an accuracy 99.99% for the training data and 99.99% for the validation or testing data. The blue line represents the model training accuracy, whereas the orange line represents the validation test accuracy which evaluates the model performance by using a testing data.

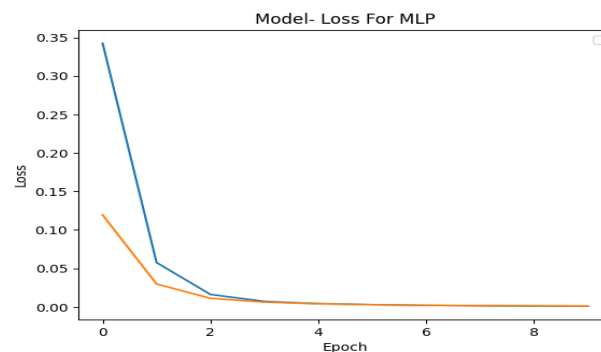


Figure 6: Loss values for both the training and testing of the MLP model.

The line graph above shows a representation of the losses acquired by the model during training and testing. The blue line indicates the loss acquired by the model during training, and the orange line indicate the loss acquired by the model during testing. The loss values are acquired at each training steps, starting from step 1 to step 10. The loss values indicate the losses the model had during training. This shows that the model achieved a loss value of about 0.02% for the training data and the validation or testing data.

Classification_Report For MLP				
	precision	recall	f1-score	support
Normal	1.00	1.00	1.00	91
IDS	1.00	1.00	1.00	100
accuracy			1.00	191
macro avg	1.00	1.00	1.00	191
weighted avg	1.00	1.00	1.00	191

Figure 7: Classification Report of the MLP model

The classification report indicates outstanding performance of the Multilayer Perceptron (MLP) model with flawless precision, recall, and F1-score for the "Normal" and "IDS" classes. This shows that the model correctly classified all occurrences of both classes in the dataset. The 100% accuracy rate solidifies the model's ability in differentiating between the two classes. The MLP model shows outstanding classification performance with great precision, recall, and accuracy, making it very reliable for the task.

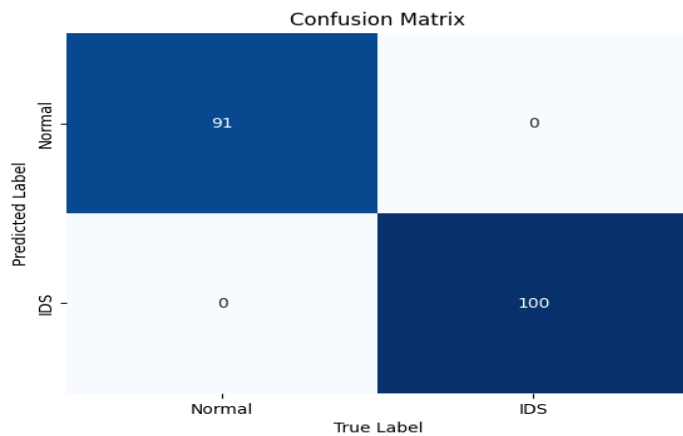


Figure 8: Confusion Matrix of the MLP model

The confusion matrix shows the number of correct and incorrect prediction of the model on the test data. The result of the confusion matrix shows that MLP model makes correct predictions for all the classes with a misclassification of 0%.

b. Encryption and Access Control

The system operates on the principle of providing each IoT device with a unique encryption key. This ensures that data transmitted between devices and the server remains encrypted, safeguarding it from potential threats such as eavesdropping or data interception. Additionally, the system incorporates access control mechanisms to differentiate between authorized and unauthorized devices. This differentiation is crucial in restricting data transmission privileges to only trusted devices, thereby mitigating the risk of unauthorized access or data breaches.

- Data Encryption:** One of the fundamental components of the system is data encryption. Employing the Fernet symmetric encryption algorithm, the system encrypts data before transmission, rendering it indecipherable to unauthorized parties. This encryption mechanism ensures that even if data is intercepted during transmission, it remains protected and unintelligible. By encrypting data at the source and decrypting it at the destination using the corresponding encryption key, the system establishes a secure communication channel, safeguarding sensitive information from potential threats.
- Access Control:** Another critical aspect of the system is access control, which determines which devices are permitted to send data. Through a predefined set of rules or criteria, the system identifies and

distinguishes between authorized and unauthorized devices. Devices that meet the specified criteria are granted permission to participate in data exchange activities, while those that do not meet the criteria are restricted from sending data. This access control mechanism ensures that only trusted devices with legitimate credentials can contribute to the data flow within the network, enhancing overall security and minimizing the risk of unauthorized access.

Simulation Results: The simulation results provide insights into the system's performance and effectiveness in enforcing security measures. Devices that are authorized to send data are clearly identified, and the data they transmit is encrypted using the corresponding encryption key. This encryption ensures the confidentiality and integrity of the transmitted data, thereby reducing the likelihood of data compromise. Conversely, devices that are not authorized to send data are also identified, highlighting the system's ability to enforce access control policies and restrict unauthorized access. By simulating various scenarios, the system demonstrates its capability to maintain data security and privacy in different operational contexts. A sample of simulated result can be seen in Figure 9, while the chart showing other simulations for different devices is represented in figure 10.

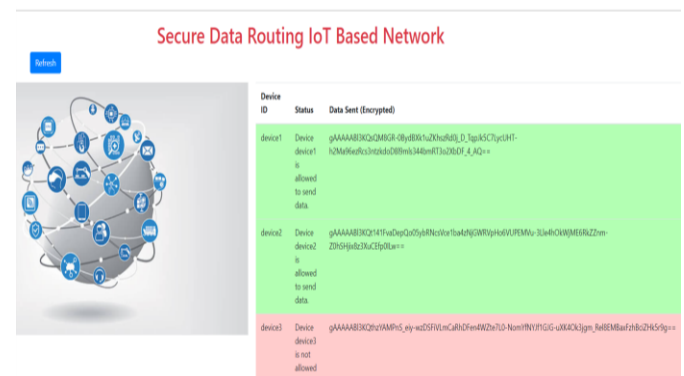


Figure 9: Simulated Result Sample

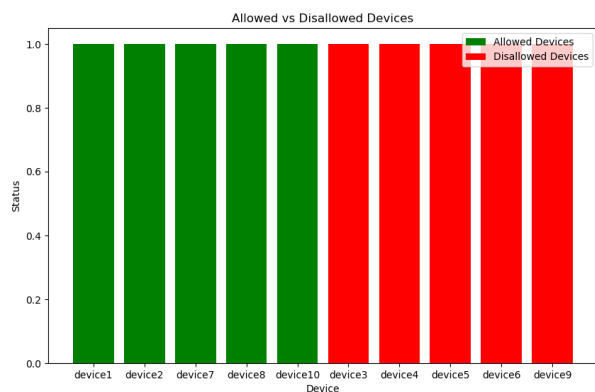


Figure 10: Allowed vs Disallowed Devices

The simulation result shows the devices that allowed and devices that are not allowed. It also shows the encrypted data sent. Devices that are allowed are shown in green while devices that are not allowed are shown in red.

6. Discussion

From the results of the Exploratory Data Analysis (EDA) of IoT data, the insights about the data are clearly shown through visualizations. The countplot of the 'proto' column illuminates the prevalence of TCP and UDP protocols, offering valuable insights into network traffic patterns. Meanwhile, the correlation matrix provides a deeper understanding of feature relationships within the dataset. Additionally, the feature ranking and histogram of the ten most important features offer insights into feature relevance and distribution, respectively, facilitating informed decision-making in subsequent modeling efforts. Transitioning to the implementation of the Multi-Layer Perceptron (MLP) for detecting IDS botnet attacks, the discussion provides a comprehensive overview of the model architecture and training process. The sequential model with three dense layers is meticulously described, along with the choice of activation functions, optimizer, loss function, and evaluation metric. The training provides a step-by-step account of the model's training progression, culminating in high accuracy and minimal loss.

Figures 5 and 6 further illustrate the model's performance through accuracy and loss values for both training and testing datasets, respectively. Moreover, the classification report underscores the model's precision, recall, and F1-score, attesting to its robustness in differentiating between "Normal" and "IDS" classes. Finally, the confusion matrix offers a visual representation of the model's accurate predictions across all classes,

affirming its reliability in detecting botnet attacks on IoT devices.

The significance of encryption and access control mechanisms is seen during the implementation phase. These components work synergistically to secure data transmission between IoT devices and the server. Encryption, utilizing the Fernet symmetric encryption algorithm, ensures that transmitted data remains indecipherable to unauthorized parties, thereby mitigating risks such as eavesdropping or interception. Access control mechanisms further bolster security by distinguishing between authorized and unauthorized devices, thereby limiting data transmission privileges exclusively to trusted devices.

The simulation results vividly illustrate the system's efficacy in identifying permitted and prohibited devices, showcasing encrypted data transmission and robust access control policies.

7. Conclusion

This paper underscores the efficacy of robust security measures, exploratory data analysis (EDA) techniques, and machine learning (ML) models in fortifying the security of Internet of Things (IoT) devices. The implementation of encryption and access control mechanisms has been instrumental in ensuring secure data transmission and mitigating potential threats such as eavesdropping and unauthorized access. Specifically, the simulation results yielded promising outcomes, showcasing the system's ability to accurately distinguish between authorized and unauthorized devices. Notably, encrypted data transmission emerged as a pivotal feature, underscoring the system's commitment to data security.

Moreover, the utilization of EDA techniques has provided valuable insights into network traffic patterns, feature relationships, and class imbalances within the dataset.

Overall, achieved results from this paper's model implementation affirm the effectiveness of a comprehensive approach encompassing encryption, access control, EDA, and ML techniques in safeguarding IoT ecosystems. These findings contribute to advancing knowledge in IoT security and provide practical insights for designing resilient security frameworks tailored to mitigate emerging cyber threats in interconnected environments.

References

1. Biru, B. A. (2018). *Open Source Solutions for the Industrial Internet of Things* (Doctoral dissertation, Politecnico di Torino).
2. Yang, X., & Wetherall, D. (2006). Source selectable path diversity via routing deflections. *ACM SIGCOMM Computer Communication Review*, 36(4), 159-170.
3. Ahmed, M., Raza, A., & Khan, M. (2021). "Anomaly Detection in IoT Networks Using Ensemble Learning." *Computers & Security*, 110, 102466.
4. Kumar, K. N., Mohan, C. K., & Cenkeramaddi, L. R. (2023). The impact of adversarial attacks on federated learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5), 2672-2691.
5. Li, X., Lu, Q., & Yuan, Y. (2018). "Decentralized Authentication Framework Using Blockchain for IoT Devices." *IEEE Internet of Things Journal*, 5(4), 3576–3585.
6. Shams, T., Hussain, M., & Al-Fuqaha, A. (2020). "Biometric-Based Secure Access Control for Healthcare IoT." *Sensors*, 20(12), 3411.
7. Kumar, A., Patel, N., & Singh, R. (2019). "Lightweight Hash-Chain-Based Authentication for IoT." *Journal of Information Security and Applications*, 44, 123–134.
8. Hwang, S., & Kim, J. (2019). "Hybrid Encryption Protocol for IoT Data Security." *Sensors*, 19(14), 3073.
9. Zhang, X., Wang, H., & Chen, Y. (2021). "Homomorphic Encryption for Secure Data Aggregation in IoT." *IEEE Transactions on Information Forensics and Security*, 16, 4032–4045.
10. Jun, C., & Chi, C. (2014). Design of complex event-processing IDS in internet of things. In *2014 sixth international conference on measuring technology and mechatronics automation* (pp. 226-229). IEEE.
11. Santos, A. L., Cervantes, C. A., Nogueira, M., & Kantarci, B. (2019). Clustering and reliability-driven mitigation of routing attacks in massive IoT systems. *Journal of Internet Services and Applications*, 10(1), 18.
12. Prabavathy, S., Sundarakantham, K., & Shalinie, S. M. (2018). Design of cognitive fog computing for intrusion detection in Internet of Things. *Journal of Communications and Networks*, 20(3), 291-298.