ISSN No:-2456-2165

# Utilizing Machine Learning Algorithms to Improve Device Authentication in IoT

Chipo Manzini, Fungayi D. Mukoko
Department of Information Sciences and Technology,
Harare Institute of Technology
Harare, Zimbabwe

**Abstract**:- **The increase in the use of IoT (Internet Of Things) gadgets in several sectors such as smart homes, agriculture, cities, and health has resulted in a growth in security challenges, notably authentication. Authentication is a security mechanism that creates the difference between legitimate and illegitimate users, and it also encompasses the identification and verification of the users. Device authentication is a challenge in IoT setting since IoT devices are resource constraint in their make and often uses passwords set by the manufacturer. Normally end users of the IoT devices do not change the passwords which make the smart environment prone to hackers. This study is aimed at using machine learning algorithms to detect and verify IoT devices in a smart home network. Only legitimate users are to have access to use the network resources. The results show that the approach has a 96% accuracy of classifying devices based on supervised machine learning algorithms and an illegitimate device can be blocked.**

***Keywords***:- *Internet of Things (IoT), Device Authentication, Machine Learning, Security.*

## I. INTRODUCTION

IoT is the linking of physical objects that have sensors implanted in them. Actuators, programs and other technologies can also be embedded with the aim of connecting and communicating with other devices over the Internet. [1]. IoT has been applied in smart homes, smart cities, healthcare systems and agriculture to limit human interaction. The goal behind the IoT is to enable devices to report automatically in real time, improve efficiency and present important information to the surface faster than systems that rely on manual intervention.

With this huge development in IoT, there are security concerns, related with it, which can influence severely the IoT structures. Lack of proper security makes it simpler for attackers to access the network and can bargain trustworthiness, privacy, nonrepudiation and accessibility of both devices and information [2] and as a result in some cases it might cause disasters [3].

Quite a number of devices are being linked to each other, this leads to the issue of authentication. As the devices need to communicate with each other, it is important for the devices to prove that they are what they say they are. The devices need to trust that the other device that wants to communicate is a legitimate device, and if there happen to be an illegitimate device, communication must not be granted. IoT has numerous limitless devices, which are heterogeneous in their make, and contrast in sizes, storage capacity, computational power and lifespan of battery [2].

Different authentication methods have been proposed in literature which include but not limited to identity- based, context-based, token- based, one- way and two- way authentication. To guard against commands from unauthorized users, reliable IoT authentication methods are required so that connected IoT devices can be trusted.

This paper proposes to use machine learning algorithms to identify and verify IoT device in a smart home setting. Previous researches have been done that classified devices, but this work goes further to allow only legitimate users to access the resources. Device network packets captured, helps determine whether a device is legitimate or illegitimate.

The paper continues as follows; related work is covered in section II. Section III discusses the methodology used in identifying and verifying devices in a smart home environment. Section IV gives results and discussion of the work, section V concludes the paper.

## II. RELATED WORK

The growing of the IoT is unquestionable. According to the researches done device authentication is well thought out as a common security challenge in the IoT environment, therefore, effective authentication techniques must be applied to establish reliable communication[4], [5]. This section presents some device authentication techniques proposed for the IoT environment, showing the technical method used.

In[4], a context-aware technique for mobile users in smart home setting was proposed. To access services in a smart home, an authentication model for mobile users was designed. Traditional credentials and context-aware information are used in the model. The user's profile, location, calendar, and access habits and logs were all used as context sources. The objective of their work was to improve static authentication techniques by reducing the verification process whenever users request to access the service they need.

In [5], a proximity-based technique for IoT device authentication, called Move2Auth was proposed. While the IoT device delivers packets over and over, the user must hold the smart phone in front of the IoT device and do one of two hand gestures that the smart phone chooses at random. The two gestures; moving smart phone in the direction of the IoT device and away from the IoT device, as well as turning the phone, allowing the device to be detected and authenticated. It incorporates device authentication and key generation for verification. The technique cannot be applied on devices that are further apart.

Based on application usage patterns, a behavior profiling model was proposed to authenticate users in IoT networks [6]. The researchers suggested the use of behavioral authentication because many authentication methods have some weaknesses and there are complications when implementing them for authentication that is continuous and implicit.

Researchers in [7] proposed the use of hardware serialization to authenticate IoT devices. The technique included the pre-registration and authentication steps. Every IoT device, according to the researchers, has a device identification chip. The goal of this chip is to offer an unclonable device identification for authentication.

According to [8], gadgets may be recognized just using machine learning technologies. To collect message packets from IoT devices in specific IP address ranges, a low-level scanner called NMAP (network mapper) was employed. The packets are collected by the AP (access point) and transferred to the server for processing. To classify IoT devices, the server use a pre-trained model. Researchers in [8] proposed using Artificial Intelligence (AI)/machine learning to solve authentication and permission problems in end devices. The suggested approach is built on a fog computing paradigm within the context of a smart home, but it does not rely on end device computational capacity, storage, or power.

Researchers in [9] presented a method for detecting devices that are categorized as white-listed devices and specific device instances connecting to a network automatically. They also created a security system model that permits the implementation of rules to limit IoT device communications based on device privileges.

At the physical layer, researchers in [10] implemented an authentication technique based on ECC (elliptic curve cryptography) between an IoT device's RFID (radio frequency and identification) tag and reader, and a server. The model that was used was set into four phases which are; pre-preparation, setup phase where the server computes an elliptic curve equation in excess of binary field, Polynomial Addition of Elliptic Curve point phase and authentication phase. When compared to other open key cryptography systems such as RSA, the use of ECC makes the convention more useful, secure, and requires less resources and calculations.

A ticket-based authentication mechanism was developed in [11] between a low-powered sensor node and a mobile device that belonged to distinct networks. Although the two directing networks have a trust agreement, the client nodes, such as sensor nodes and mobile nodes, do not have. The nodes and their respective authentication servers have pre- shared private keys. The authentication servers both authenticate the nodes. Both nodes are certain that they are talking with their legitimate counterparts because both authentication servers are involved during authentication. The authenticating objects affirm that they share the same session key based on their common trust[12].

In [13], the authors proposed a common and lightweight physical layer authentication system for the Internet of Things devices in smart cities. The tags were used for verification purposes. The tag embedding method goal is to insert a legitimate IoT device's authentication tag signal. The tag verification method determines whether or not the signal received at the receiver is from a genuine IoT device[13].

Researchers in [14] proposed a mutual authentication mechanism utilizing Elliptic Curve Cryptography (ECC) between server the and the IoT device using RFID tags. Elliptical curve on binary field is calculated and polynomial addition of elliptical points is done on binary field. The pseudo random number generator retains an internal state which comprise of a key and seed. The tag and server must exchange messages during the authentication step. The server produces the random number, and the tag chooses it.

To successfully authenticate users accurately and at the same time reducing the use of much resources, [15] proposed an authentication technique for IoT systems that is light-weight called Li-GAT (Lightweight Gait Authentication Technique) that make use of various information collected from smart devices. Small IoT devices use Li-GAT to preserve energy and compute power, which are both important parts of the IoT ecosystem. This method required users to keep their devices in their hands for authentication to take place.

The study by S. Jung and S. Jung [16] proposed the push Open Authorization (OAuth) that changes the OAuth protocol and issues the OAuth token when the OAuth authorization server registers to the OAuth client first. According to the researchers, using a personal OAuth authorization server to authenticate is more reliable than using a third-party authorization server since users may limit access to the data created by IoT devices directly. Users may directly authorize the OAuth client that can access the information on their IoT devices[16].

Researchers in [17] proposed an authentication scheme that relied on PUF to reduce the risk of authentication key exposure and the capacity of authentication server. The suggested method only stores and updates a CRP when it comes into touch with the device that needs to be authorized. By encrypting authentication messages with secret key generated by the CRP, it also provides for more secure device authentication.

A multi-key based mutual authentication technique is proposed in[18].Secure vault; that is a collection of keys with equal size is a common secret key between the IoT server and the IoT device. The secure vault's initial contents are exchanged between the server and the IoT device, and the secure vault's information must be update after every successful communication.

Researchers describe a PUF- based Authentication Scheme (PAS) with session keying, which allow smart devices and the IoT gateway to communicate securely. On the grounds of a CRP, a registration and authentication technique was given, as well as session keying. Authentication requests are delivered by the PAS using the device's footprint. The gateway issues a challenge to the PUF-based device seeking authentication and generation of a session key for further communication. A secure command execution protocol was also given for executing commands and requested parameter choices initiated by requesting devices in the Internet of Things, such as cell phones and wearable devices[19].

## III. METHODOLOGY

The main objective is to identify and verify an IoT device in a network. Real IoT devices like smart bulbs, smart sensors and smartphones that control the devices are connected to a raspberry Pi to form a network as shown in fig. 1.

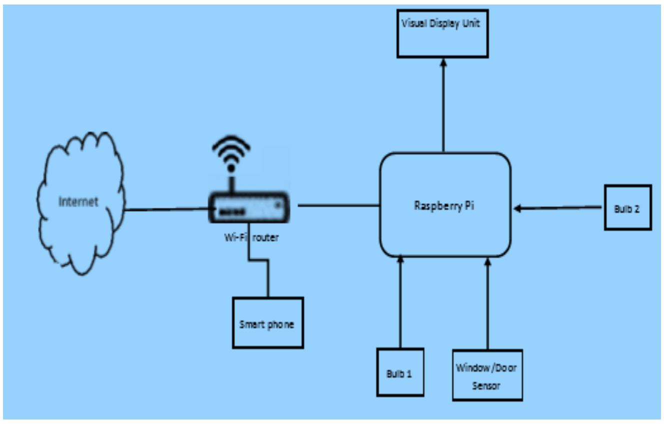
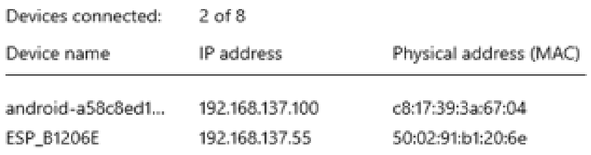


Fig. 1 Devices connected to a raspberry Pi

The smart bulbs, and smartphone are currently connected to the raspberry Pi. The Pi is acting as a wireless access point for the IoT devices. Fig. 2 shows the mac addresses of the devices currently connected.



Devices connected: 2 of 8

| Device name | IP address | Physical address (MAC) |
|---|---|---|
| android-a58c8ed1... | 192.168.137.100 | c8:17:39:3a:67:04 |
| ESP_B1206E | 192.168.137.55 | 50:02:91:b1:20:6e |

Fig. 2 Mac addresses of connected devices

Network packets of real IoT devices are captured using packet sniffer software. The network packet is the flow between the transmitter IP, receiver IP, transmitter port and receiver port. Each flow of data is used to extract features that are used to train the model. The approach is based on supervised learning where labels are used. Feature selection is done in order to identify the devices. The protocols used were TLS, UDP, TCP, ARP, etc. the communication flow from each device are captured based on sender IP, receiver IP, sender port and receiver port.

Data cleaning is done to remove unwanted features or to get rid of that which is not important to the device classification.

A dataset is created where the rows represent traffic flow and the columns represent feature vector. If a new device is added, training is done again to add new features to the model.

## IV. RESULTS AND DISCUSSION

Logistic Regression ()LR, Linear Discriminant Analysis (LDA), Decision Tree Classifier (DTC), Random Forest Classifier (RF) and Gaussian NB were the algorithms that were used in training the model. Decision Tree classifier had a score of approximately 96.32%, which made it an algorithm that classifies the IoT devices with minimum errors. Gaussian NB had a percentage of 74.38%, which made the algorithm less accurate as compared to other algorithms.

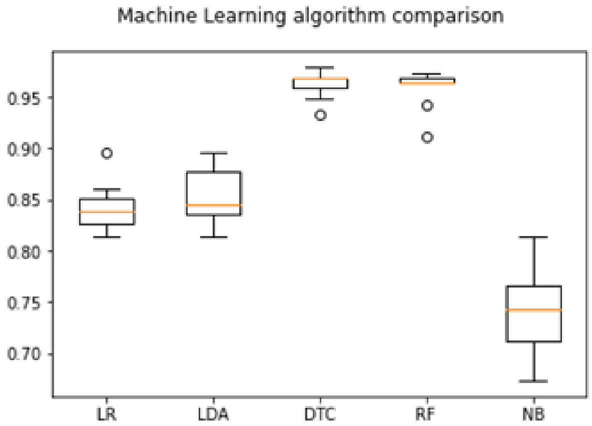


Fig. 3 Comparison of machine learning algorithms

Confusion matrices are performance indicators used when working with classification problems. They give remarkable summary on how well a model is performing. A 2 × 2 confusion matrix per binary classifier is used.

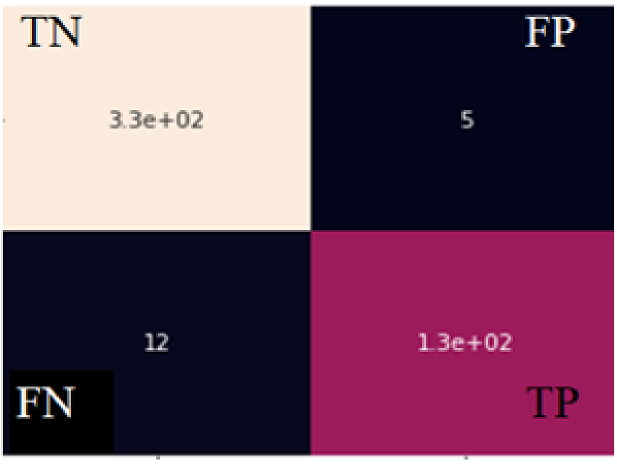


Fig. 4 Confusion matrix

Confusion matrix shows the accuracy of the model. Accuracy is given as follows:

$$
\begin{aligned}
\text{Accuracy} &= (TP + TN) \,/\, (TP + FP + TN + FN) \\
&= (316 + 150) / (150 + 5 + 316 + 12) \\
&= 0.964\ldots.. \\
&= 0.96 \qquad (1)
\end{aligned}
$$

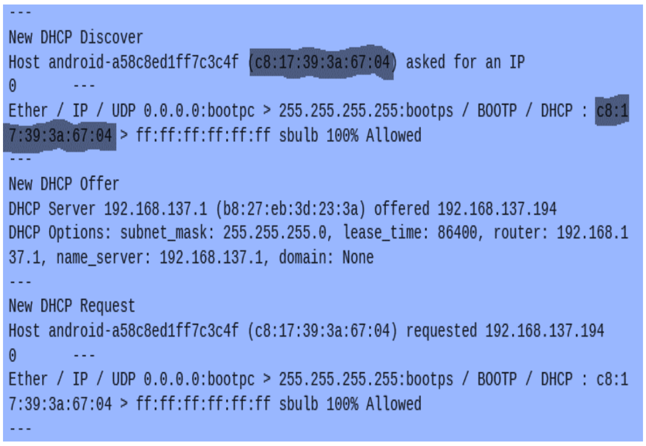
Where; TP is True Positive, TN is True Negative, FN is False Negative and FP is False Positive.

From the confusion matrix, precision, which is the estimate of how many predictions are correct, is also calculated.

$$
\begin{aligned}
\text{Precision} &= TP \,/\, (TP + FP) \\
&= 316 / (150 + 5) \\
&= 1.58 \qquad (2)
\end{aligned}
$$

Legitimate devices are granted access to the network resources. From the diagram given below, the mac address highlighted shows a device that was given access to the network.

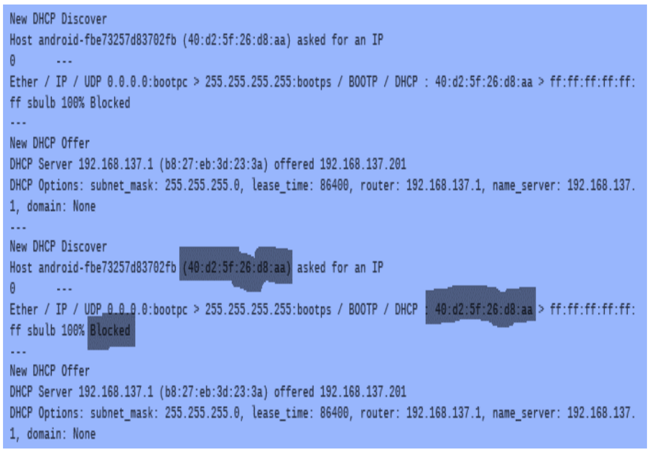
.



```
---
New DHCP Discover
Host android-a58c8ed1ff7c3c4f (c8:17:39:3a:67:04) asked for an IP
0       ---
Ether / IP / UDP 0.0.0.0:bootpc > 255.255.255.255:bootps / BOOTP / DHCP : c8:1
7:39:3a:67:04 > ff:ff:ff:ff:ff:ff sbulb 100% Allowed
---
New DHCP Offer
DHCP Server 192.168.137.1 (b8:27:eb:3d:23:3a) offered 192.168.137.194
DHCP Options: subnet_mask: 255.255.255.0, lease_time: 86400, router: 192.168.1
37.1, name_server: 192.168.137.1, domain: None
---
New DHCP Request
Host android-a58c8ed1ff7c3c4f (c8:17:39:3a:67:04) requested 192.168.137.194
0       ---
Ether / IP / UDP 0.0.0.0:bootpc > 255.255.255.255:bootps / BOOTP / DHCP : c8:1
7:39:3a:67:04 > ff:ff:ff:ff:ff:ff sbulb 100% Allowed
---
```

Fig. 5 Access granted to a legitimate device

Illegitimate devices are blocked from accessing the network. From the diagram given below, the mac address highlighted shows a device that is blocked from the network.



```
New DHCP Discover
Host android-fbe73257d83702fb (40:d2:5f:26:d8:aa) asked for an IP
0       ---
Ether / IP / UDP 0.0.0.0:bootpc > 255.255.255.255:bootps / BOOTP / DHCP : 40:d2:5f:26:d8:aa > ff:ff:ff:ff:ff:
ff sbulb 100% Blocked
---
New DHCP Offer
DHCP Server 192.168.137.1 (b8:27:eb:3d:23:3a) offered 192.168.137.201
DHCP Options: subnet_mask: 255.255.255.0, lease_time: 86400, router: 192.168.137.1, name_server: 192.168.137.
1, domain: None
---
New DHCP Discover
Host android-fbe73257d83702fb (40:d2:5f:26:d8:aa) asked for an IP
0       ---
Ether / IP / UDP 0.0.0.0:bootpc > 255.255.255.255:bootps / BOOTP / DHCP : 40:d2:5f:26:d8:aa > ff:ff:ff:ff:ff:
ff sbulb 100% Blocked
---
New DHCP Offer
DHCP Server 192.168.137.1 (b8:27:eb:3d:23:3a) offered 192.168.137.201
DHCP Options: subnet_mask: 255.255.255.0, lease_time: 86400, router: 192.168.137.1, name_server: 192.168.137.
1, domain: None
```

Fig. 6 Device blocked from accessing resources

## V. CONCLUSIONS

This paper presented a method of identifying legitimate devices in a local smart home network. Features are extracted from network traffic flow. Logistic Regression, Linear Discriminant Analysis, Decision Tree Classifier, Random Forest Classifier and Gaussian NB were used. Decision tree classifier shows the highest accuracy of 96%. Devices can be blocked from accessing resources using device features.

## REFERENCES

[1] Margarate Rousse, “What is IoT (Internet of Things) and How Does it Work?” 2020.

[2] O. Lucia, B. Isong, N. Gasela, and A. M. Abu-Mahfouz, “Device Authentication Schemes in IoT: A Review,” Proc. - 2019 Int. Multidiscip. Inf. Technol. Eng. Conf. IMITEC 2019, 2019, doi: 10.1109/IMITEC45504.2019.9015902.

[3] M. El-Hajj, A. Fadlallah, M. Chamoun, and A. Serhrouchni, “A survey of internet of things (IoT) authentication schemes,” Sensors (Switzerland), vol. 19, no. 5, pp. 1–43, 2019, doi: 10.3390/s19051141.

[4] Y. Ashibani, D. Kauling, and Q. H. Mahmoud, “A context-aware authentication service for smart homes,” 2017 14th IEEE Annu. Consum. Commun. Netw. Conf. CCNC 2017, pp. 588–589, 2017, doi: 10.1109/CCNC.2017.7983179.

[5] J. Zhang, Z. Wang, Z. Yang, and Q. Zhang, “Proximity based IoT device authentication,” Proc. - IEEE INFOCOM, 2017, doi: 10.1109/INFOCOM.2017.8057145.

[6] A. Hasan, “Internet of Things Device Authentication Scheme using Hardware Serialization,” pp. 109–114, 2018.

[7] A. Hasan and K. Qureshi, “Internet of Things Device Authentication Scheme Using Hardware Serialization,” ICAEM 2018 - 2018 Int. Conf. Appl. Eng. Math. Proc., pp. 109–114, 2018, doi: 10.1109/ICAEM.2018.8536286.

[8] S. Zareen, “Artificial Intelligence / Machine Learning in IoT for Authentication and Authorization of Edge Devices,” 2019 Int. Conf. Appl. Eng. Math., pp. 220–224.

[9] S. A. Hamad, W. E. Zhang, and Q. Z. Sheng, “IoT Device Identification via Network-Flow Based Fingerprinting and Learning,” 2019 18th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. IEEE Int. Conf. Big Data Sci. Eng., pp. 103–111, 2019, doi: 10.1109/TrustCom/BigDataSE.2019.00023.

[10] S. Khan and R. K. Aggarwal, “Efficient Mutual Authentication mechanism to Secure Internet of Things (IoT),” Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019, pp. 409–412, 2019, doi: 10.1109/COMITCon.2019.8862196.

[11] P. Gope, R. Amin, S. K. H. Islam, N. Kumar, and V. K. Bhalla, “Lightweight and privacy-preserving RFID authentication scheme for distributed IoT infrastructure with secure localization services for smart city environment,” Futur. Gener. Comput. Syst., 2017, doi: 10.1016/j.future.2017.06.023.

[12] A. P. Shrestha, S. M. R. Islam, and K. S. Kwak, “Ticket-Based Authentication for Securing Internet of Things,” 2020 10th Annu. Comput. Commun. Work. Conf., pp. 938–941, 2020, doi: 10.1109/ccwc47524.2020.9031254.

[13] P. Zhang, S. Member, J. Liu, and Y. Shen, “Lightweight Tag-based PHY-layer Authentication for IoT Devices in Smart Cities,” IEEE Internet Things J., vol. PP, no. c, p. 1, 2019, doi: 10.1109/JIOT.2019.2958079.

[14] A. Tewari and B. . Gupta, “A Mutual Authentication Protocol for IoT Devices using Elliptic Curve Cryptography,” 2018 8th Int. Conf. Cloud Comput. Data Sci. Eng. (Confluence)., pp. 716–720, 2018, doi: 10.1109/confluence.2018.8442962.

[15] P. Musale, D. Baek, and B. J. Choi, “Lightweight gait based authentication technique for IoT using subconscious level activities,” IEEE World Forum Internet Things, WF-IoT 2018 - Proc., vol. 2018-Janua, pp. 564–567, 2018, doi: 10.1109/WF-IoT.2018.8355210.

[16] S. W. Jung and S. Jung, “Personal OAuth authorization server and push OAuth for Internet of Things,” Int. J. Distrib. Sens. Networks, vol. 13, no. 6, pp. 0–10, 2017, doi: 10.1177/1550147717712627.

[17] B. Kim, S. Yoon, Y. Kang, and D. Choi, “PUF based IoT Device Authentication Scheme,” 2019 Int. Conf. Inf. Technol. Converg., pp. 2019–2021, 2019, doi: 10.1109/ICTC46691.2019.8939751.

[18] T. Shah and S. Venkatesan, “Authentication of IoT Device and IoT Server Using Secure Vaults,” 2018 17th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. 12th IEEE Int. Conf. Big Data Sci. Eng., pp. 819–824, 2018, doi: 10.1109/TrustCom/BigDataSE.2018.00117.

[19] M. A. Mughal, X. Luo, Z. Mahmood, and A. Ullah, “Physical Unclonable Function Based Authentication Scheme for Smart Devices in Internet of Things,” no. Ic, 2018, doi: 10.1109/SmartIoT.2018.00037.