An Intrusion Detection System in IoT Environment Using KNN and SVM Classifiers

Abdulmalik M Alfarshouti
Computer and Information Technology, University of Tabuk, Tabuk, Saudi Arabia.
E-mail: 412010278@stu.ut.edu.sa

Saad M Almutairi
Computer and Information Technology, University of Tabuk, Tabuk, Saudi Arabia.
E-mail: S.almutairi@ut.edu.sa

Received September 18, 2021; Accepted December 16, 2021
ISSN: 1735-188X
DOI: 10.14704/WEB/V19I1/WEB19231

Abstract

IoT applications are now used in most applications in this world to facilitate data collection and remote and automatic management of all modern devices. Due to the large spread of these devices in multiple regions, they become easily vulnerable to penetration by many types of attacks. This research will focus on network layer denial of service (DOS) attacks to detect. This type of attack was chosen because of its danger to the availability of services, such as e-commerce services, financial and government services, as well as educational organizations. Failure to provide these services frequently leads to huge financial losses in addition to loss of confidence in these organizations. Machine learning techniques will be used in the proposed research to detect these attacks in a fast and efficient manner.

Keywords

IoT, Attacks, DOS, MITM, Instructions.

Introduction

1. Background

Our world is now largely dependent on information, circulation and availability. The concept of the Internet of Things appeared in 1999 by Kevin Ashton. This concept depends on the devices and equipment all over the world being connected to each other via the Internet (Lee, 2015) as shown in Figure 1.1.

Fig. 1.1 Internet of Things (IoT) (Source (D. Giusto, 2010))

As shown in the previous figure, IoT is a new concept for creating a network of devices and equipment that can communicate with each other and integrate to provide countless smart applications (D. Giusto, 2010). This concept is considered one of the most important fields of future technology to employ all computer science to serve people.

The connection methods for IoT devices vary, for example, you can use RFID, ZigBee, WSN, DSL (Digital Subscriber Line) and WLAN. IoT applications are many and very useful, as they are now used in the field of intelligent transportation, medical and agricultural applications and environmental monitoring (R. Khan, 2012). To build an IoT, there must be several elements that define the purpose of IoT and device communication methods (Marques. G, 2017). Figure 1.2 illustrates these elements. For example, for addressing devices and identify them on the IoT network, the IPv4, IPv6 and 6LoWPAN are used. The Internet of Things depends on the concept of sensing to collect and transmit data, as it contains sensors to send and receive data and actuators to receive and execute commands. Also, the communication process is an important factor in building the Internet of things, through which communication between the components of the network takes place.

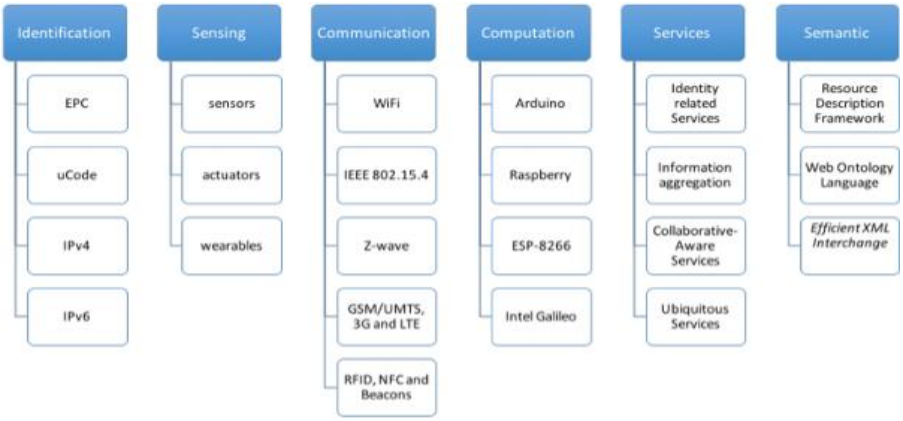

Fig. 1.2 IoT Element(Source [4])

There are multiple applications for IoT, for example smart homes. Smart home systems are integrated with the Internet of Things. Some home systems will be reviewed to demonstrate that their components can easily be exposed to attacks, and mechanisms are needed to counter and detect these attacks. An overview of smart homes is presented in (C. Wilson, 2015) as shown in Fig. 1.3. For example, smart homes and their applications are dealt with through: 1. The functional view; 2. An instrumental view; and 3. A socio-technical view. Smart home applications save energy and monitor the status of home appliances. Also in (F. Adib, 2015), a smart home application was introduced to monitor the breathing process and heartbeat of patients inside the home.

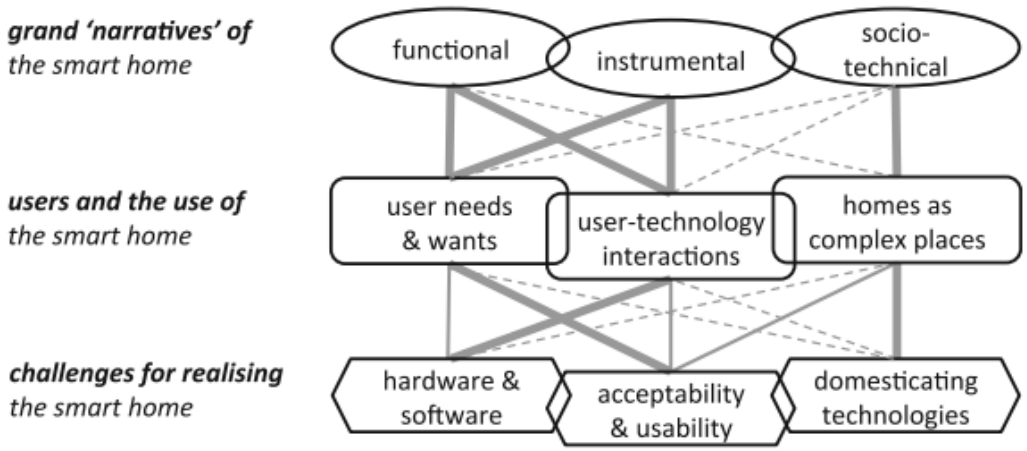

Figure 1.3 Organizing framework for smart homes (Source [5])

2. Attacks Targeting IoT Devices

As a result of the IoT architecture, as it consists of Heterogeneous components and communicates in various ways such as wireless communications, it is constantly exposed to threats (Mendez Mena.D, 2018). Also, the resource limitations of IoT devices such as the sensors make them unable to protect themselves. The attacks that can be exposed to IoT devices vary greatly (Tabassum, A., & Lebda, W. 2019), Figure 1.4 shows a detail of these attacks. In this research the MITM and DOS attacks will be choosing as they have detrimental effect on IoT applications. In MITM attacks, unencrypted com- munication paths are attacked, and an illegitimate connection is established between the devices, resulting in the attacker gaining access to the data. With this attack, the attacker could eavesdrop, change and delete data (Anthi, 2019). Therefore, MITM attacks are invisible because they are carried out in silence and have no apparent effect. On the other hand, DOS attacks prevent the service from reaching the legitimate users (Rathore H, 2017). For example, the health care system is attacked, which leads to the deprivation of patients and doctors of access to information in a timely manner.

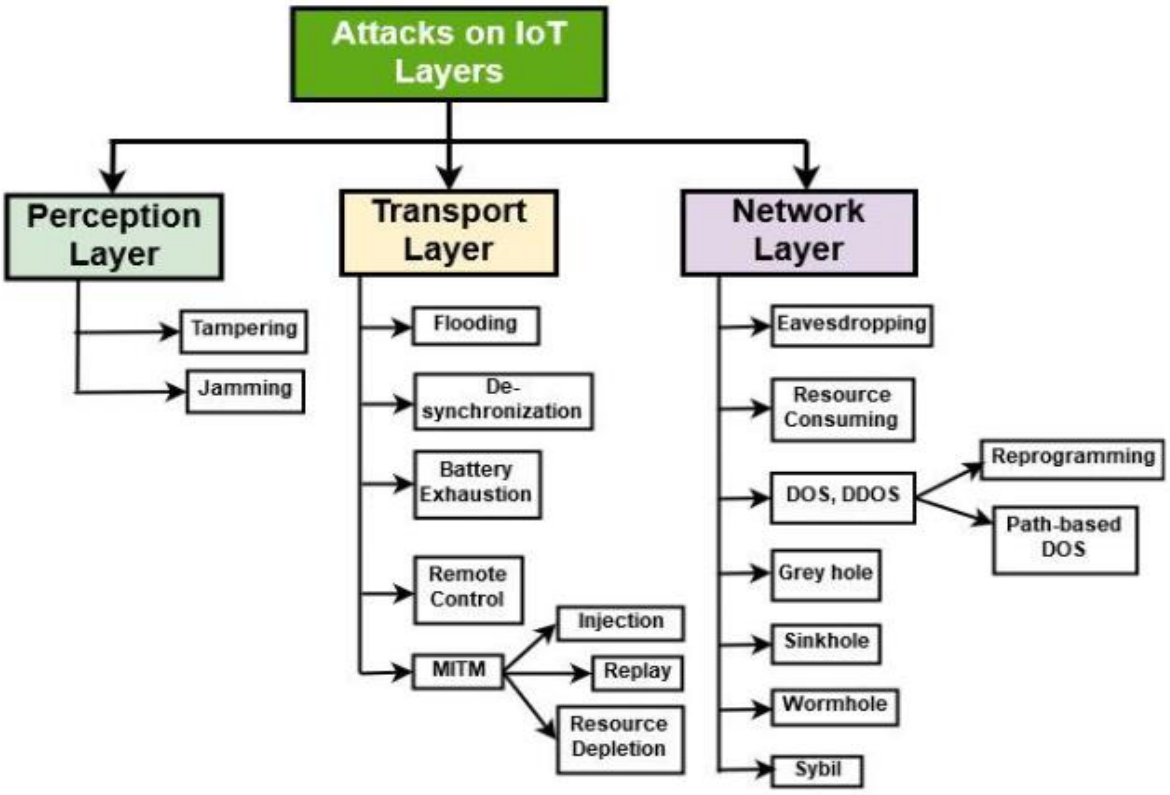

Figure 1.4 IoT Attacks (Source [8])

3. Problem Definition

The IoT elements are constantly communicating with each other by using various types of communication media such as wireless communication. Also, most of the IoT elements are constantly connected to the Internet, which makes it easier to attack it by penetrating the wireless network or the Internet (Mendez Mena. D, 2018). The most famous of these attacks are Denial of Service (DOS) over the network layer. These attack are very harmful and dangerous for people and organizations, as IOT devices can be controlled, such as company, home security or smart cars (Mendez Mena. D, 2018). Therefore, these attacks must be constantly detected to protect IoT devices.

4. Objective

The main objectives of the presented research can be summarized as follows:

- Get a complete picture of the IoT concept, its elements, and how to secure them.
- Study the possible attacks on IoT elements.
- Securing IoT against these attacks by proposed a model based on machine learning techniques to detect these attacks intelligently.

5. Research Questions

In this research the following research questions that will be answered.

How to deal with security issues while using the IoT elements?

What are the most used machine learning techniques to secure IoT Network and devices?

What is the most appropriate machine learning technique to secure IoT Network and devices?

The answers to these questions are aimed at studying securing IoT devices using machine learning techniques.

Related Work

Due to the design of the IoT that makes it Due to the design of the IoT that makes it most often connected to the internet, it is vulnerable to Network Intrusion from hackers. As a result, many tools and frameworks have been designed to help detect and protect against these Intru- sion. Machine learning has now become one of the popular techniques used in these tools and frameworks, because of its self-learning ability. In this chapter, a discuss of previous work that works to discover these Intrusion and how to implement this detect and protection from these Intrusion is explained. In (A. Nagisetty, 2019) a framework is introduced that is able to detect malicious activities in IoT networks using the Keras Deep Learning Library. Four different deep learning methods were used and compared, namely, Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNN), Deep Neural Networks (DNN) and Auto encoder. This research uses UNSW- NB15 and NSL-KDD99 dataset to train the deep learning models. Figure 2.1 illustrates the proposed framework. In (Hao, Z., 2020) an intrusion detection system in IoT networks using artificial neural networks (ANN) was introduced. This system addresses the problem of False positive rate and false negative rate by means of a sequential classifier as shown in Figure 2.2. To implement this idea, five consecutive ANN networks were designed, in which the first network classifies traffic data into normal data or intrusion on the network. In the second network, normal data is passed to detect that there is intrusion that were not detected in the first network. This process is carried out until the last network that gives the final classification results.

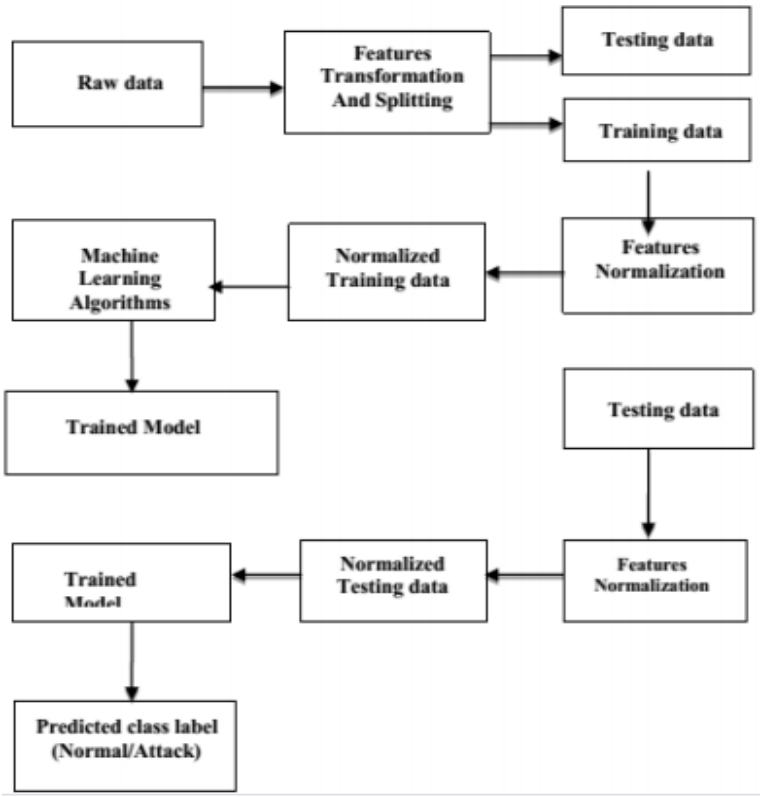

Figure 2.1 Framework for Detection of Malicious Activities using deep Learning Model (A. Nagisetty, 2019)

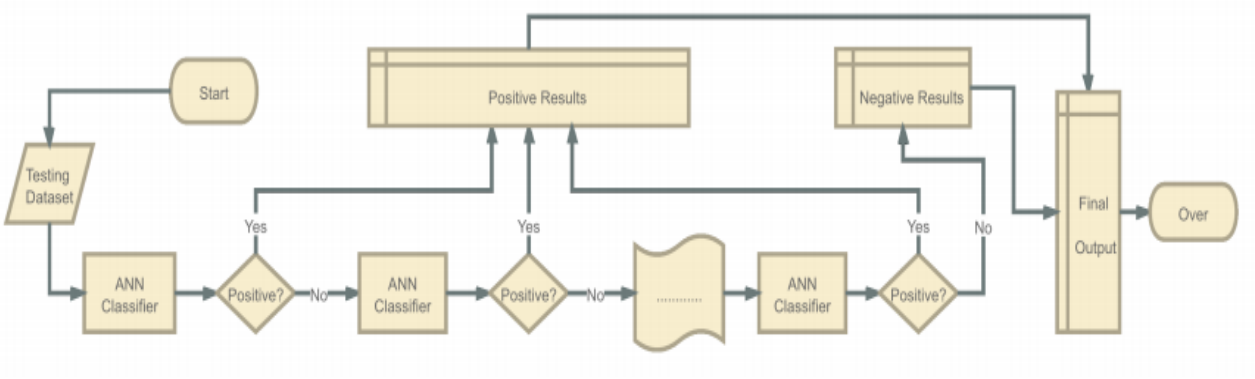

Figure 2.2 Sequential Detection System (Hao, Z., 2020)

In (S. S. Swarna Sugi, 2020), an intrusion detection system was introduced in the Internet of Things, using two methods, Long Short- Term Memory (LSTM) and K-Nearest Neighbor (KNN) as illustrated in figure 2.3. These two methods are applied to detect intrusion then a comparison for the results is proposed. By comparing the two techniques, LSTM has a high speed in detecting intrusion, and also has higher detection accuracy than KNN.

In (S. Latif, 2020), a new Intrusion Detection System for Internet of Things based on deep learning (DRANN) was developed using UNSW-NB15 dataset. DRANN consists of one input layer, 5 hidden layer and 1 output layer. The input layer contains 41 neurons as the number of input features in the dataset. The accuracy that achieved by this system is 99.54 %.

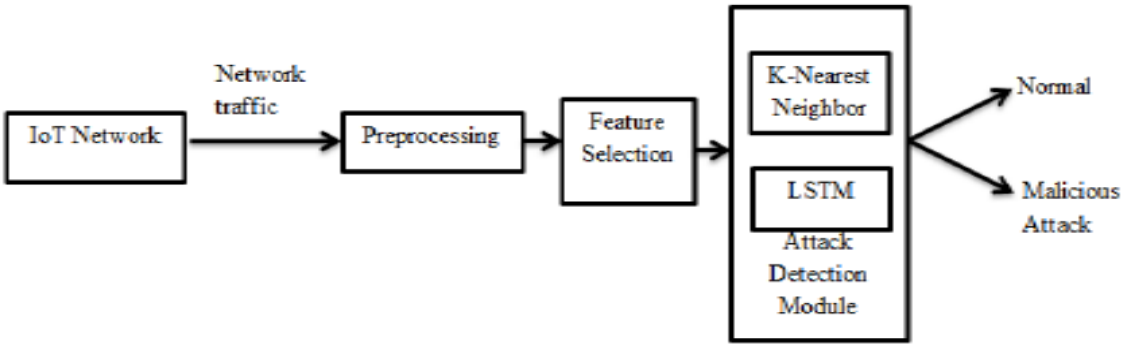

Figure 2.3. Intrusion Detection System (S.S. Swarna Sugi, 2020)

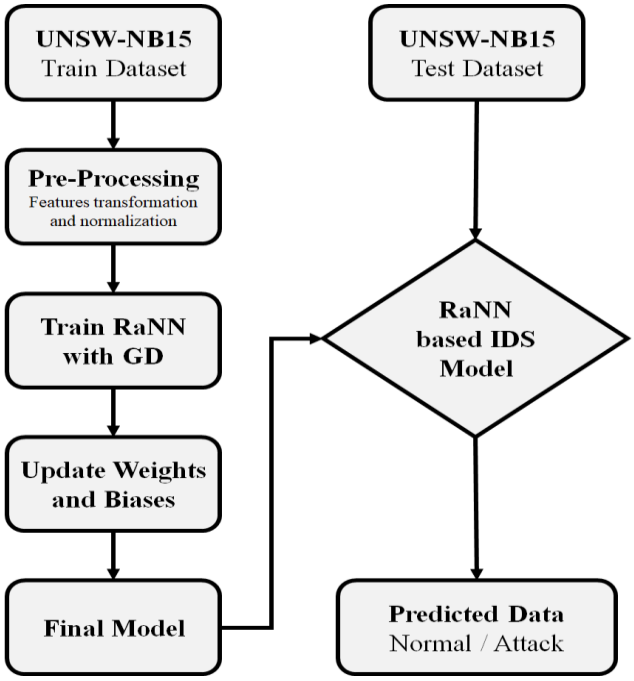

Figure 2.4 Intrusion Detection Scheme. (S. Latif, 2020)

In (E. D. Alalade, 2020), a system has been established to detect intrusion in Internet of things networks, using hybrid learning techniques Artificial Immune System and Extreme Learning Machine (AIS-ELM). This system focuses on smart homes to protect them from intruder attacks. The system architecture is explained in figure 2.5.

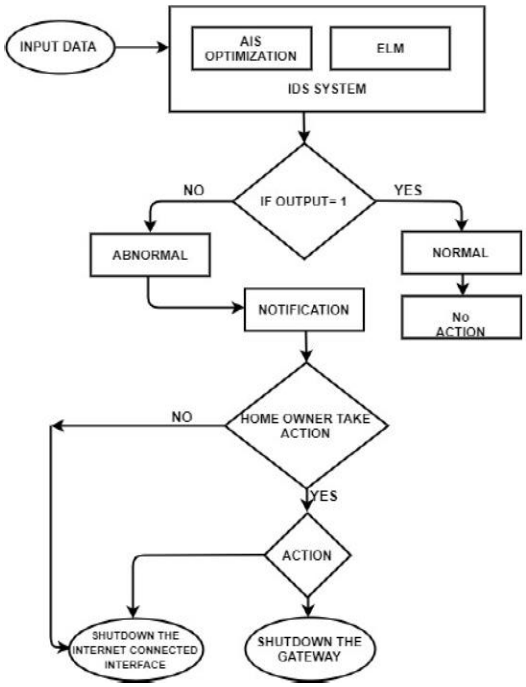

Figure 2.5 Intrusion Detection Scheme. (E.D. Alalade, 2020)

By using Blockchain technology (Cheema, M.A., 2020), a machine learning method (SVM)) is applied to detect intrusion in IoT networks. The training process in this research is illustrated in figure 2.6, Bot-IoT data set is used. Blockchain technology is used to share the attackers' information (IP addresses) IoT divisions.

A supervised learning models is used in (M. G. Desai, 2020) to build a classifier to detect intrusion in IoT Network using KDD dataset. Decision Tree (DT), Random Forest classifier (RFC), and Support Vector Machines (SVM) are used in this research as supervised learning models to build the classifier. A comparison between these three methods is explained, SVM is the most accurate.

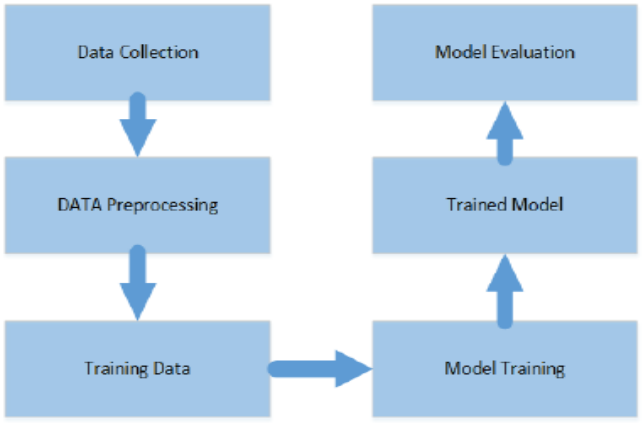

Figure 2.6 Intrusion Detection Scheme. (Cheema, M.A., 2020)

In (Eltanbouly, S., 2020), several machine learning algorithms that are applied to intrusion detection on network have been reviewed. Various datasets are used in this research to train the various machine learning algorithms. Researchers in (Illavarason, P., 2019) presented a study to evaluate the performance of feature extraction and machine learning classification techniques in the intrusion detection system.

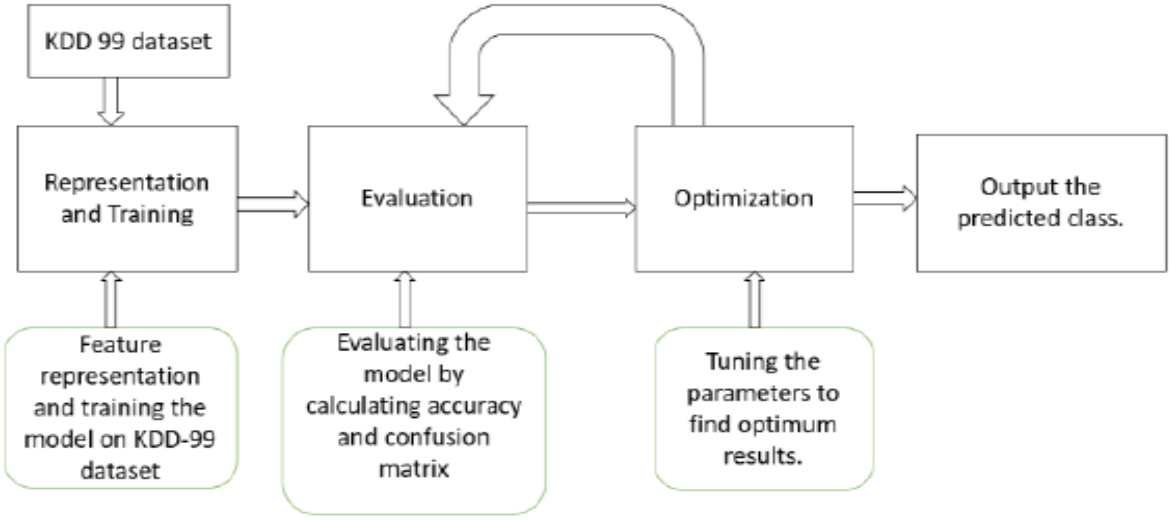

Figure 2.7 Block diagram of anomaly detection (Vikram, A., 2020)

Researchers at (Vikram, A., 2020) have presented a machine learning-based method for detecting intrusion on the IoT network. KDD dataset was used and as a result of its duplication of values and the multiplicity of data types, a preprocessing was performed on it in the first step before training the machine learning algorithm. Figure 2.7 explain the anomaly detection system.

Research in (Hodo, E., 2016) aims to search for a better classifier capable of detects anomaly traffic from N_BaIoT dataset. Several classifiers based on machine learning techniques have been experimented with. Decision Trees, Extra Trees Classifiers, Random Forests, and Support Vector Machines represent the machine learning techniques used in this research. The data varied between Thermostat – Baby Monitor – Security Camera.

The researchers review in (Elrawy, M., 2018) the architecture of IoT with a focus on its weaknesses that lead to security breaches. The four types of security problems present in the different layers of the IoT model are shown in Figure 2.8. The four types from bottom layer to up layer are authentication-related problem, Confidentiality-related risks, the data integrity between services and applications and privacy.

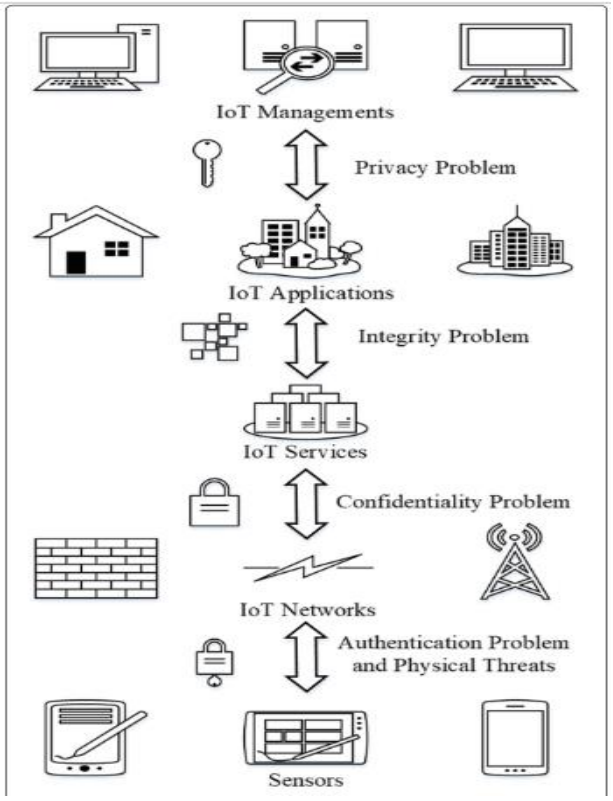

Figure 2.8 The security challenges in the different IoT layers. (Elrawy, M., 2018)

Methodology

In this chapter, the methodology of the introduced research for building the proposed system to detect IoT Network intrusion as shown in Figure 3.1 the KNN and SVM classifiers are used to detect the network intrusion. After that, the results for each classifier will be compared and the best one will be chosen to be relied upon in designing a system capable of detecting various attacks on the IoT network.

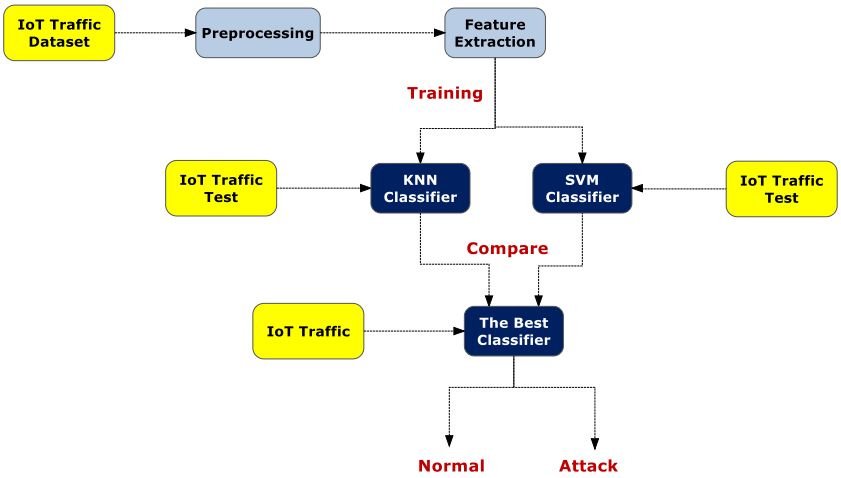

Fig. 3.1 The proposed System to detect IoT Network Intrusion

Dataset

In this research, the Bot-IoT Dataset was used to design a model to detect botnet attacks (DOS-DDOS) using machine learning algorithms (KNN - SVM). Botnets are IoT devices connected to the Internet infected with malware, through which attackers can infiltrate and take control of these devices. The most famous of these attacks is DOS-DDOS. Bot-IoT Dataset generated by designing a realistic network environment in the Cyber Range Lab of UNSW Canberra (N. Koroniotis, 2019). A lightweight protocol (MQTT protocol) is used which makes this data applicable to various IoT solutions. This data is available in CSV files of 72 million records with 42 features (27 Integer, 13 Float, and 2 String types). Only 5 % of this data has been choose to train and test the proposed model.

Implementation Tools

Experiments in this research were carried out using Matlab 2019 on a computer containing an Intel i7 processor with 8 gigabyte of RAM. Windows 10 is the operating system of the used computer.

Feature Selection and Dimensionality Reduction

In this research, a dataset has been relied on, which contains many features, some of which can be dispensed with. Only the features relevant to attack detection will be selected, as this will contribute to raising the efficiency of the machine learning model to predict the presence of attacks. In addition to, the selecting appropriate features has the feasibility to reduce any overfitting that might occur using all the features. Twenty percent of the data is used for testing and the remaining for training. In the first, the data will be read with the deletion of all unselected columns in data analysis. In this research to apply the machine learning techniques to the Bot-IoT dataset, the features described in table 1. are chosen.

Table 3.1 Dataset Features and Description

Feature	Description
stime	Record start time
sport	Port that data is being sent from
dport	Port that data is being received from
pkts	Total number of packets transferred
bytes	Total number of bytes transferred
ltime	Record last time
seq	Sequence number
dur	Record total duration
mean	Average duration of aggregated records
sum	Total duration of aggregated records
min	Minimum duration of aggregated records
max	Maximum duration of aggregated records
spkts	Source to destination packet count
dpkts	Destination to source packet count
sbytes	Source to destination byte count
dbytes	Destination to source byte count
rate	Total packets per second in transaction
srate	Source to destination packets per second
drate	Destination to source packets per second

Machine Learning Algorithms

In this research, the KNN and SVM algorithms are used to analysis dataset that contain the DOS-DDOS attack. KNN and SVM algorithms are types of supervised learning. SVM is less computationally demanding for most machine learning algorithms, which increases the speed of detection of attacks. As for KNN, it is one of the best performance machine learning algorithms in recognizing very complex patterns, which makes it suitable for dealing with IoT architecture data. Where training data is used with their labels for learning and then tests are performed on a portion of this data to verify its validity and accuracy.

The K Nearest Neighbor algorithm classifies data based on a similarity measure. The way it works depends on the classification of points in relation to the classification of neighboring points. It can be used, for example, to classify email data as normal or spam, based on prior training data containing both types. Figure 3.2 shows how the models based on the KNN algorithm work, where any new data is classified based on its closest data set.

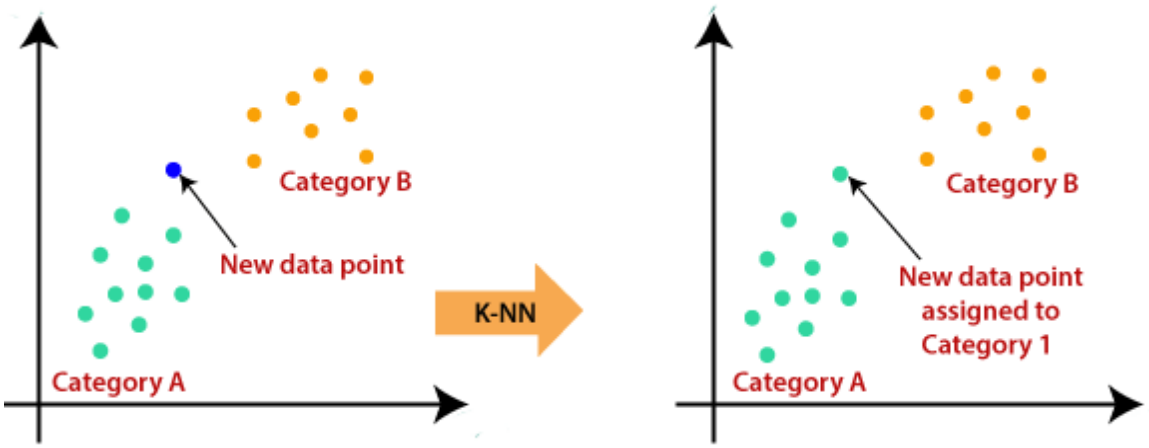

Figure 3.2 KNN Classification

The SVM algorithm is a supervised machine learning algorithm, commonly used for classification and regression. In 2-dimensional space, this algorithm separates data types by creating a line between them. This algorithm relies on the features in the data to separate the data into two types, each with similar data. Figure 3.3 shows this line and how to separate the data by the best line that passes between them. The SVM algorithm can also be used to classify more than two types of data in a post-modification procedure as found in the fitcecoc function in Matlab.

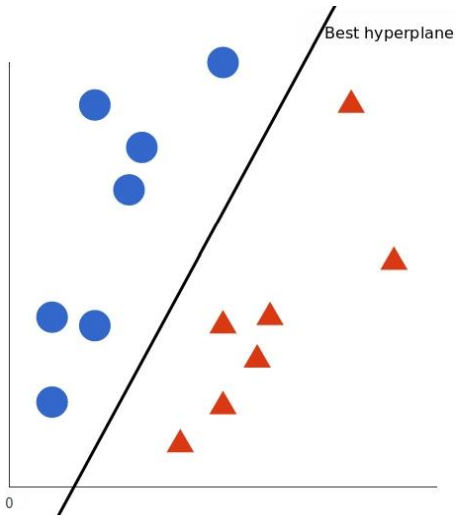

Figure 3.3 SVM Classification

Results

Performance Metrics

To evaluate the performance of the KNN and SVM models that introduced in this paper, will rely on Accuracy, Precision, Recall & F1 Score metrics. To calculate these metrics, the following values True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) will be calculated. These values can be viewed on confusion matrix as illustrated in figure 4.1 and also can be defined as:

TP: is the success of the to detect the occurrence of the condition when it actually exists.
TN: is the failure of the model to detect the occurrence of the condition when it does not actually exist.
FP: is the success of the to detect the occurrence of a condition when it does not really exist
FN: is the failure of the model to detect the occurrence of the condition when it actually exists.

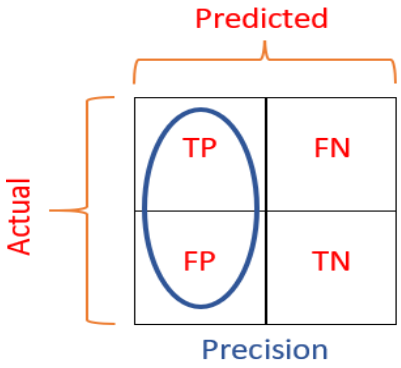

Figure 4.1 TP, TN, FP, FN from Confusion Matrix

Can now calculate the metrics that will be used in evaluating machine learning models to classify IoT attack data as follows:

Accuracy = Number of items correctly identified as either truly positive or truly negative out of the total number of items =TP+TN/(TP+FP+FN+TN)

For Multiclass (K class) classification the accuracy can be computed as:

$$\frac{\sum_{i=1}^{k} \frac{tp_i+tn_i}{tp_i+tn_i+fp_i+tn_i}}{k}$$

Precision = Number of items correctly identified as positive out of the total items identified as positive= TP / (TP + FP)

For Multiclass (K class) classification the precision can be computed as:

$$\frac{\sum_{i=1}^{k} tp_i}{\sum_{i=1}^{k}(tp_i + fp_i)}$$

Recall = Number of items correctly identified as positive out of the total actual positives= TP / (TP + FN)

For Multiclass (K class) classification the recall can be computed as:

$$\frac{\sum_{i=1}^{k} tp_i}{\sum_{i=1}^{k}(tp_i + fn_i)}$$

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

KNN and SVM Comparison Results

Now in this section the results of data analysis and detection of (DOS, DDOS) attacks using KNN and SVM for multiclass techniques will be displayed and compared. In the first the results for KNN techniques will be displayed as in figure 4.2:

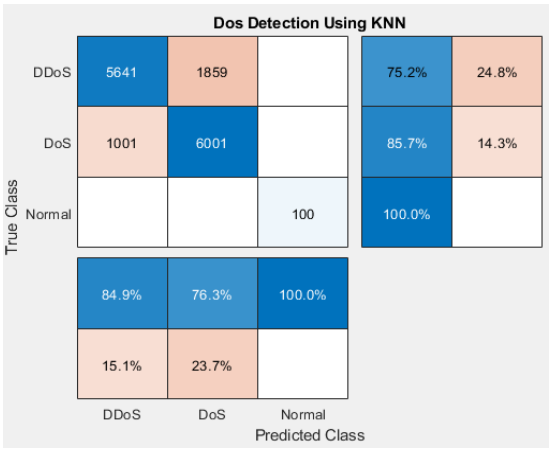

Figure 4.2 Confusion Matrix for the KNN model

Average Accuracy= (80.5 +80.5 +100) / 3 = 87%

The KNN model was able to detect DOS attacks with 80.5 % accuracy, DDOS attacks with 80.5% accuracy and finally detect normal data with 100% accuracy. The mean accuracy of the model was calculated for both attacks and normal data, and the result was 87%. Also precision for DOS 85.7%, for DDOS is 75.2 %, finally for Normal data is 100%. The overall precision for the KNN model has the following value: Precision =0.87.

As for Recall, the value for DOS 76.3%, for DDOS is 84.9%, finally for Normal data is 100%. The overall precision for the KNN model has the following value: Recall = 0.87

F1 Score = 2*(Recall * Precision) / (Recall + Precision) = 2(0.87*0.87)/ (0.87+0.87) = 0.87

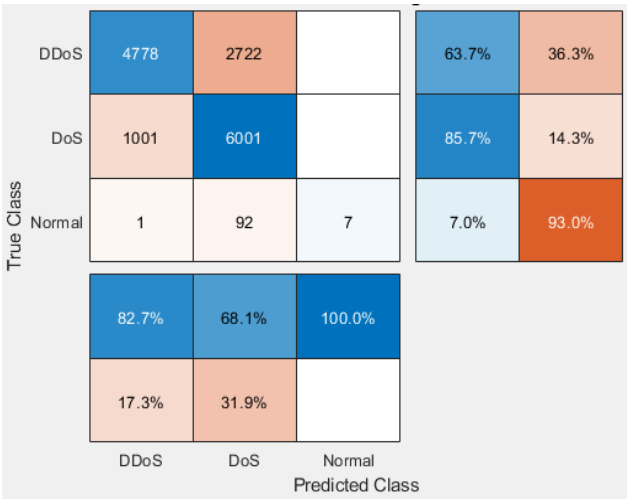

Figure 4.3 Confusion Matrix for the SVM model

Average Accuracy= (63.7 + 85.7 + 7) /3= 52%

The SVM model was able to detect DOS attacks with 85.7 % accuracy, DDOS attacks with 63.7 % accuracy and finally detect normal data with 7% accuracy. The mean accuracy of the model was calculated for both attacks and normal data, and the result was 52%. Also precision for DOS 85.7%, for DDOS is 63.7%, finally for Normal data is 7%. the overall precision for the SVM model has the following value: Precision =0.52

As for Recall, the value for DOS 68.1%, for DDOS is 82.7%, finally for Normal data is 100%. The overall precision for the KNN model has the following value: Recall = 0.836

F1 Score = 2*(Recall * Precision) / (Recall + Precision) = 2(0.836*0. 52)/ (0.836+0. 52) = 0.64

As it is clear from the previous results, the values of the KNN model are better than the SVM model

Table 4.1 Performance matrices for KNN and SVM Technique

Model	Accuracy	Precision	Recall	F1 Score
KNN	87%	0.87	0.87	0.87
SVM	52%	0.52	0.836	0.64

Conclusion and Future Work

Internet of things applications are nowadays the dominant and most prevalent applications for the services they provide covering all fields. In this research, two machine learning models were presented based on the KNN algorithm and SVM algorithm, to detect attacks against IoT devices. DOS and DDOS attacks are focused on in this research. Matlab software was used to develop the two machine learning models, the two models were tested and the performance measure for each model was calculated. It can be concluded from the results of this research that machine learning algorithms are able to detect attacks on the IoT architecture. In addition to, the results showed that the developed model based on the KNN algorithm outperformed the developed based on the SVM algorithm.

There are many types of machine learning algorithms, differing in their accuracy and relevance to the quality of the problem or the data they are analyzing. Therefore, among the future development goals of the proposed research is to choose several machine learning algorithms and compare them to choose the best. Also researching the nature of the data that contains normal data and attacks, to extract some features that can help us increase the accuracy of the selected machine learning models.

References

Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons, 58*, 431- 440.

Giusto, D., Iera, A., Morabito, G., & Atzori, L. (Eds.). (2010). *The internet of things: 20th Tyrrhenian workshop on digital communications.* Springer Science & Business Media.

Khan, R., Khan, S.U., Zaheer, R., & Khan, S. (2012). Future internet: the internet of things architecture, possible applications and key challenges. *In 10th international conference on frontiers of information technology*, 257-260.

Marques, G., Garcia, N., & Pombo, N. (2017). A survey on IoT: architectures, elements, applications, QoS, platforms and security concepts. *In Advances in mobile cloud computing and big data in the 5G era*, 115-130.

Wilson, C., Hargreaves, T., & Hauxwell-Baldwin, R. (2015). Smart homes and their users: a systematic analysis and key challenges. Personal and Ubiquitous Computing, 19(2), 463-476.

Adib, F., Mao, H., Kabelac, Z., Katabi, D., & Miller, R. C. (2015). Smart homes that monitor breathing and heart rate. *In Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 837-846.

Mendez Mena, D., Papapanagiotou, I., & Yang, B. (2018). Internet of things: Survey on security. *Information Security Journal: A Global Perspective, 27*(3), 162-182.

Tabassum, A., & Lebda, W. (2019). Security Framework for IoT Devices against Cyber-Attacks. arXiv preprint arXiv:1912.01712.

Anthi, E., Williams, L., Słowińska, M., Theodorakopoulos, G., & Burnap, P. (2019). A supervised intrusion detection system for smart home IoT devices. *IEEE Internet of Things Journal, 6*(5), 9042-9053.

Rathore, H., Mohamed, A., Al-Ali, A., Du, X., & Guizani, M. (2017). A review of security challenges, attacks and resolutions for wireless medical devices. *In 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 1495-1501.

Nagisetty, A., & Gupta, G. P. (2019, March). Framework for detection of malicious activities in iot networks using keras deep learning library. *In 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 633-637.

Hao, Z., Feng, Y., Koide, H., & Sakurai, K. (2020). A sequential detection method for intrusion detection system based on artificial neural networks. *International Journal of Networking and Computing, 10*(2), 213-226.

Sugi, S.S.S., & Ratna, S.R. (2020). Investigation of machine learning techniques in intrusion detection system for IoT network. In 3rd International Conference on Intelligent Sustainable Systems (ICISS), 1164-1167.

Latif, S., Idrees, Z., Zou, Z., & Ahmad, J. (2020). DRaNN: A deep random neural network model for intrusion detection in industrial IoT. *In International Conference on UK-China Emerging Technologies (UCET)*, 1-4.

Alalade, E.D. (2020). Intrusion Detection System in Smart Home Network Using Artificial Immune System and Extreme Learning Machine Hybrid Approach. *In IEEE 6th World Forum on Internet of Things (WF-IoT)*, 1-2.

Cheema, M.A., Qureshi, H.K., Chrysostomou, C., & Lestas, M. (2020). Utilizing Blockchain for Distributed Machine Learning based Intrusion Detection in Internet of Things. *In 16th International Conference on Distributed Computing in Sensor Systems (DCOSS),* 429-435.

Desai, M.G., Shi, Y., & Suo, K. (2020). IoT Bonet and Network Intrusion Detection using Dimensionality Reduction and Supervised Machine Learning. *In 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON),* 0316-0322.

Eltanbouly, S., Bashendy, M., AlNaimi, N., Chkirbene, Z., & Erbad, A. (2020). Machine learning techniques for network anomaly detection: A survey. *In IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT),* 156-162E.

Illavarason, P., & Sundaram, B.K. (2019). A study of intrusion detection system using machine learning classification algorithm based on different feature selection approach. *In Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC),* 295-299.

Vikram, A. (2020). Anomaly detection in Network Traffic Using Unsupervised Machine learning Approach. *In 5th International Conference on Communication and Electronics Systems (ICCES),* 476-479.

Hodo, E., Bellekens, X., Hamilton, A.W., Dubouilh, P., Iorkyase, E., Tachtatzis, C., & Atkinson, R.C. (2016). Threat analysis of IoT net- works using artificial neural network intrusion detection system. *International Symposium on Networks, Computers and Communications (ISNCC),* 1-6.

Elrawy, M.F., Awad, A.I., & Hamed, H.F. (2018). Intrusion detection systems for IoT-based smart environments: a survey. *Journal of Cloud Computing, 7*(1), 1-20.

Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset. *Future Generation Computer Systems, 100,* 779-796.
https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/bot_iot.php

An Introduction to KNN Algorithm.
https://www.section.io/engineering-education/introduction-to-knn-algorithm/

Support Vector Machines (SVM) Algorithm Explained. 2021.
https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/