InfoTech Spectrum: Iraqi Journal of Data Science (IJDS)
ISSN: 3007-5467, Vol 3(No.1), 29-44 (2026)
DOI: https://doi.org/10.51173/ijds.v3i1.46
https://ijds.mtu.edu.iq

HUB-BASED INTERNET OF THINGS (IoT) SECURITY USING K-MEANS CLUSTERING

Hussein Ahmed Khalaf [1*]

[1] *Department of Computer Sciences, University of Monastir, Monastir, 1001, Tunisia*

** Corresponding author E-mail: hussei2nahme22d@gmail.com (Hussein Ahmed Khalaf)*

RESEARCH ARTICLE

ARTICLE INFORMATION

SUBMISSION HISTORY:
Received: 11 August 2025
Revised: 21 October 2025
Accepted: 1 November 2025
Published: 30 January 2026

KEYWORDS:
IoT;
IDATE;
HVAC;
Smart Homes;
K-Means;

ABSTRACT

The uncontrollable increase in the number of connected devices and smart homes has posed a significant security threat, particularly to systems such as smart locks, HVAC systems, and networking protocols like Zigbee and Z-Wave. The issue discussed in this paper is the issue of poor monitoring and threat detection in smart homes. Our proposal and implementation involve a prototype security module that can be used to monitor a smart home network, detecting any suspicious activity and notifying users accordingly. To measure the effect on system resources of the prototype and to observe its capability to detect malicious behavior, the prototype was implemented and tested on a smart hub platform. The experimental findings suggest that the module has the lowest performance overhead whilst having a high detection rate. Such results suggest that small-scale security surveillance can be utilized to enhance the robustness of smart home systems against cyberattacks and increase the security of individuals and the community.

1. INTRODUCTION

The Internet of Things (IoT) is a term that is gaining increasing popularity. It is an evolving discipline that is never static and therefore new applications of it ought to emerge shortly [1]. The most common understanding of the IoT is a cluster of things that are well-established or embedded using electric energy, actuators, sensors, software, and connected devices, using the Internet to connect and replace data throughout the house and all other devices [2]. The number of smart homes being constructed and the number of connected devices in use have both steadily increased in recent years [2]. With the introduction of networking technologies like Z-Wave and Zigbee, as well as smart locks and HVAC systems in recent years, the range of alternatives has expanded significantly. According to the literatures, there will be 25 billion IoT devices in use by 2020 [3]. A large number of these devices will be found in everyday homes, bringing technology closer to people than ever before.

While smart home technology has achieved many successes, there have also been numerous safety failures. Strict security precautions are necessary in this profession, as evidenced by the numerous compromised gadgets that exploited several vulnerabilities in Samsung SmartThings and its companion apps, among other things, to disable functionalities and trigger false fire alarms [4]. Passwords for eight of the sixteen smart home devices that tested in a black-box fashion were recovered [5]. There have also been more reports of legitimate users and companies getting compromised. Breaking into baby monitors and aquarium thermometers are two examples of such assaults [6]. Only a small number of them address the smart home scenario; the majority [7] focus on Wireless Sensor Network (WSN) or general IoT. To the best of our knowledge, none of these strategies takes into consideration the smart hub, which is an essential part of many contemporary smart homes. By increasing inhabitants' knowledge of potential threats, researching hub-based threat detection could enhance smart home security overall, Fig. 1.

Lighting, window and door controls, climate control systems, and many other elements are

DOI: https://doi.org/10.51173/ijds.v3i1.46

common in smart homes[8]. Modernizing control and monitoring systems to accommodate smart home applications is one of the primary objectives of IoT designers and manufacturers. This enables individuals to automate their homes and better manage their lives. Inlet switches, weather sensors, smart gates, and smoke detectors are a few examples [9]. However, there are substantial challenges to overcome due to the novelty of this goal and the immaturity of the relevant technology [10]. The concept of smart homes, which aims to regulate and monitor a wide range of household processes centrally, is based on electronic networking technologies, Fig. 2.

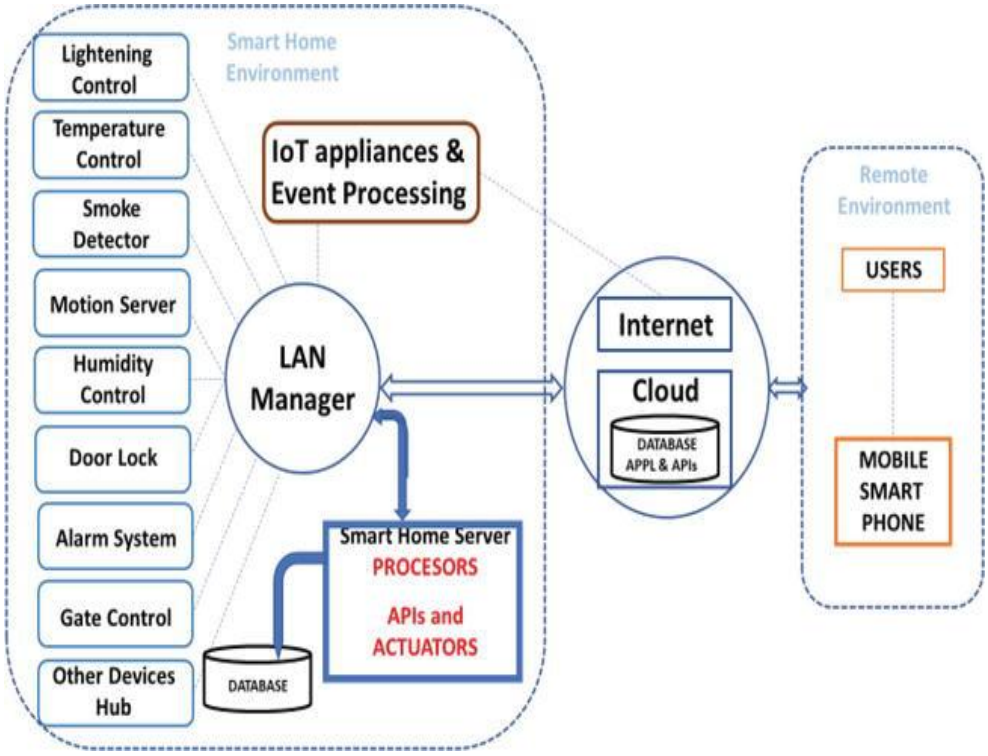

Figure 1. Smart home system using IOT

1.1. IOT and Smart Home Security

According to [11], a smart home can be remotely controlled, monitored, and maintained in response to user requests through the use of devices and sensors linked to the IoT. The first generation of smart home devices became available to householders at the start of the new millennium [12]. According to any product or technology related to smart homes, consumers can utilize a smartphone app or other networked device to monitor and control connected home appliances remotely. Therefore, security measures are already built into IoT-based smart home systems to counteract potential dangers [13]. In this context, the numerous wireless network solutions in the field of home networking, such as wireless Ethernet, ultra-wideband (UWB), and Bluetooth, among others, become relevant [14].

Ensuring the security of smart home IoT systems requires more than technical safeguards such as encryption or intrusion detection. A comprehensive governance framework provides the necessary structure to protect users, reduce risks, and establish accountability. In this context, laws, standards, guidelines, policies, and procedures serve as complementary layers of defense. Their integration helps manufacturers, service providers, and end-users align their practices with both technical and societal expectations [15], [16].

- Laws and Rules: There are national and international legal frameworks that are needed to enforce compliance and clarify accountability. The need to implement tougher legislative measures in protecting citizens against cyberattacks and the misuse of information has been raised by many governments, industry organizations, and consumer advocates.

- Standards: The use of standardization (e.g., ISO/IEC, ETSI EN 303 645, NIST frameworks) to implement security requirements ensures that these requirements are applied consistently across both devices and ecosystems. This enhances fragmentation, improves interoperability and reduces vulnerabilities that can be used by attackers [17].
- Guidelines: Guidelines should be provided in a practical way to enable the users to apply security controls effectively. The abstract standards are put into practice in everyday smart home settings through step-by-step recommendations (including password management, firmware update practice, and safe network settings).
- Policies: Security policies transform standards and guidelines into requirements that are enforceable to households, organizations and service providers. These policies ensure the uninterrupted protection of personal information, system accessibility, and equipment integrity under the broader set of legislations.
- Procedures: Lastly, operationalization policies enact security policies by giving clear and repeatable procedures. They may involve guidelines on the secure devices onboarding, making regular updates, or incident response. This clarity in the procedures becomes essential to the resilience and the long-term maintenance of the system [18].

These governance factors can be embedded in the design, deployment and use of smart home IoT to ensure that the stakeholders go beyond ad hoc solutions and implement a sustainable model of security-by-design and security-in-practice.

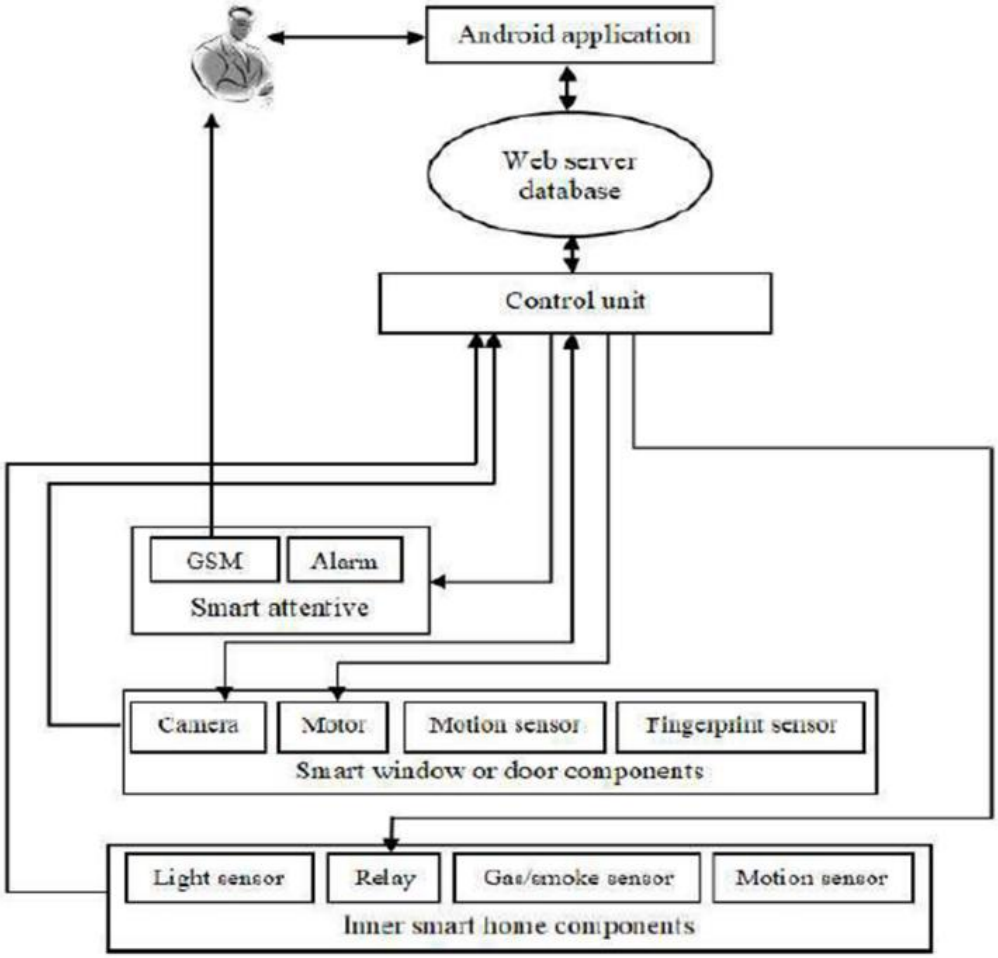

Figure 2. Smart Home Security System

1.2. Control And Monitor Components

A standard smart home security system is a set of hardware and software components which combine to provide a dependable system of control, monitoring, and intrusion detection. All the components have a specific part in the system functionality and user data protection. The key

components of the system are described in the following subsections, and they are accompanied by reference to the pertinent literature and implementations, Fig. 2.

- Android Application: An application is the primary interface of end users. Customers are able to view and control devices remotely, e.g. smart locks, lighting, and cameras, via the app. Push notifications, user authentication, and secure communication protocols are also supported using mobile applications [19].
- Web Server and Database: The web server serves as the core node for processing the commands and providing communication between the client and the devices. Sensitive credentials (e.g., usernames and passwords) are frequently stored by databases connected to the server and should be encrypted and guarded against unauthorised access. The security research indicates that ineffectively configured web servers are an important attack point in smart home systems.
- Control Unit: The control unit acts as the brain of the smart home system and coordinates the communications between the devices and processes sensor information. Popular platforms are based on microcontrollers like Arduino-based platforms, Raspberry Pi devices, or ESP modules. These platforms facilitate wireless communication (e.g., Wi-Fi, RF, Zigbee, etc.) as well as allow the development of secure automation systems, which are modular and cost-effective[20].
- Motion Sensor: Motion sensors (e.g., PIR sensors) provide intrusion detection by monitoring activity near entry points such as doors and windows. When motion is detected, the system can trigger alarms or initiate video recording. Integration of low-cost sensors with IoT platforms is widely discussed as a foundational component of smart surveillance and intrusion prevention [21].

1.3. IOT Network Security

The smart home community still has a major problem with network security. Several security enhancement solutions have been developed and disseminated across the Internet community in an effort to counteract network attacks. Surprisingly often, these assaults are either modified to avoid detection or are entirely novel. The most prevalent and persistent issue is an attacker trying to get into systems connected to the Internet, which is beyond the comprehension of the typical user. Because of this, many tiers of network security have been implemented in the IoT. Customers buying IoT gadgets for smart homes do so on the faith that the manufacturer has implemented adequate security measures. Since this is the case, a primary method of network security is to detect and eliminate common security threats [22]. The proliferation of electronic gadgets beyond traditional computers like desktops and laptops has compounded previously pressing security problems. These computers run lightweight operating systems, such as macOS, Linux, or Windows, which have limited storage space and fewer security features. Usually, these gadgets may link up with one another through other gadgets or networks, employing wireless-like communication protocols as BLE, Bluetooth, NFC, ZigBee, Wi-Fi, LoRaWan, Thread, etc. [23]. To complement the current IP network, the Internet Engineering Task Force (IETF) has exerted considerable effort in the development of necessary lightweight communication protocols for use in regulated settings [24]. These include the Constrained Application Protocol (CoAP: RFC 7252) and the IPv6 over Low Power Wireless Personal Area Networks (6LoWPAN: RFC 6282).

Additionally, the IPv6 Routing Protocol for Low-Power and Lossy Networks (RPL: RFC 6550) [21]. The network has residential routers, which, like their commercial counterparts, were probably purchased from different manufacturers and include premium management interfaces. These providers may agree to allow an external amp module to manage network behavior and abandon the optimization of user interfaces (the prototype controls open-source platforms, for example, OpenWRT). The cloud-based control model (Fig. 3) is able to provide the best reaction on the aspect of using the device, thereby reducing the difficulty associated with growth for manufacturers, allowing them to focus on realistic optimization.

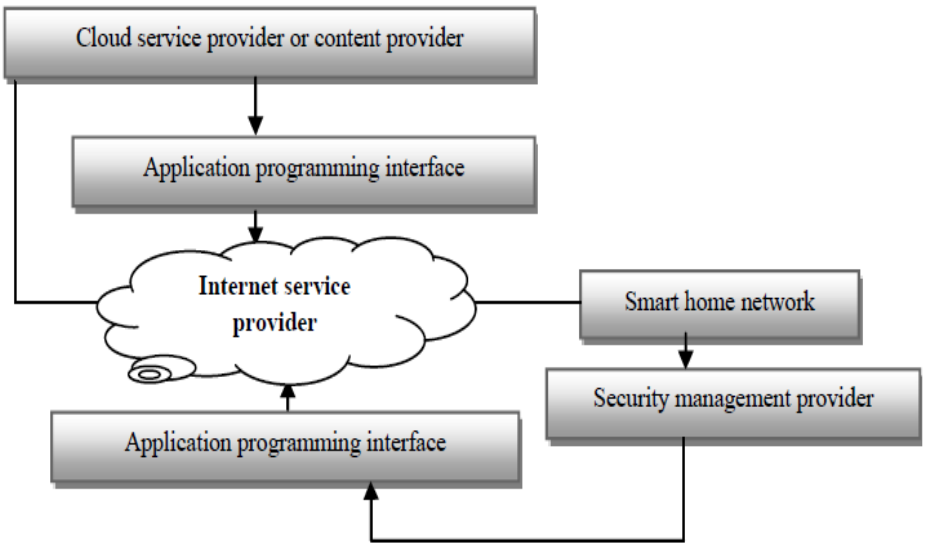

Figure 3. Network Level Security

1.4. Impact Levels by Threat

Most IoT devices used in smart homes lack robust, built-in security mechanisms, largely due to hardware limitations such as low processing power, minimal storage capacity, and restricted battery life [22]. The absence of sufficient security software exposes these devices to a wide range of threats that compromise the confidentiality, integrity, and availability (CIA triad) of smart home systems.

Smart-home attacks can generally be classified into two categories: passive and active. Passive attacks involve manipulators trying to acquire information without any change in system functionality. An instance is the network traffic analysis and packet sniffing, which enable attackers to track devices' communication. Tools like Wireshark may capture information exchanged between smart home parts and allow attackers to study sensitive information without affecting the system functionality, which is especially hard to detect.

Active attacks, on the contrary, directly interfere with device functionality or data. This is usually denial-of-service (DoS), code injection and tampering of messages. As an example, hackers can modify the energy consumption messages to overbill houses by making them pay more than they consumed [22].

IoT device security is also a problem that is worsened by the limitations of protocols and packet size. The IEEE 802.15.4 standard has a maximum packet size of 127 bytes in the physical layer. A 25-byte overhead was still taken into consideration, thus leaving 102 bytes at the MAC layer. With the use of encryption such as the AES-CCM-128, the payload is only 81 bytes, whereas using AES-CCM-32, the payload is only 93 bytes. These constraints indicate why strong encryption is challenging to deploy in IoT resource-constrained devices. Designers therefore have to tradeoff between security and performance of the device, as well as its energy efficiency [25].

Altogether, the lack of functionality in IoT devices, as well as the lack of information on built-in security, contributes to privacy loss and insecure communications as an urgent issue in smart homes. The solutions to these problems include not only effective encryption algorithms but also very compact security systems built with constrained IoT in mind.

1.5. Problem Description

Smart home security is a relatively unexplored and complex field. Here, we comment on some of the most burning points concerning smart home security,

1.5.1. Heterogeneity

There are a lot of different kinds of gadgets that make up smart homes. To begin with, there is

a wide variety of wireless standards used by various gadgets, from Wi-Fi and Bluetooth to Zigbee and Z-Wave. Moreover, devices use a wide variety of data protocols, such as constrained application protocol (CoAP) and message queueing telemetry transfer (MQTT) [26]. A comprehensive security system would have to be familiar with and flexible enough to work with any protocol. Overall, it is difficult to find a single solution that meets everyone's requirements due to heterogeneity, especially considering the yearly influx of new items.

1.5.2. Educating the user

Everyone in the middle class is who smart homes are aimed towards. This demographic typically lacks the knowledge necessary to protect themselves from cyber threats and keep their devices secure. Consumers aren't willing to pay a premium for a more secure product because they don't understand the importance of security. As a result, the industry has little incentive to improve smart home security. The conclusion that raising end-user awareness can improve smart home device security is compelling.

1.5.3. Defence position

To protect the network from both external and internal threats, the positioning of defense systems in smart homes is a challenge. If a gadget only communicates within its own private network, the network provider will never know if it is sending harmful data. If a ZWave node employs the multi-hop functionality, in which it can act as an intermediate router, to forge or discard messages, the router will be unaware of the problem. Due to its lack of resources, a certain node cannot perform advanced detection [27]. These cases show that there is no clear starting point from which to build a bulletproof defense that can prevent all possible threats. Our defense system is centred around a central processing unit. Data flows within the network may be understood in great detail thanks to the smart hub's central location in the local network and proximity to the end nodes. The smart hub may also report on the status of connected gadgets. The objective of this study is to design, implement, and evaluate a hub-based IoT security monitoring module that can detect and alert users to suspicious network activity with minimal impact on system performance. Specifically, this research seeks to:

- Develop a lightweight intrusion detection prototype suitable for deployment on a consumer-grade smart hub.
- Evaluate detection accuracy against common IoT attack scenarios (e.g., spoofing, replay, DoS).
- Assess system feasibility, including CPU/memory overhead and latency introduced by the monitoring mechanism.

2. LITERATURE REVIEW

IoT is a concept that is getting more popular in the industry and in research. IoT is a term that discusses a connected network of interrelated objects such as sensors, actuators, embedded systems, and software applications that are linked together, communicate, and share data through the Internet to facilitate the automation, monitoring, and control of various environments, such as a smart home. It is a rather dynamically developing sphere, where hardware and communication technologies keep being improved and their areas of usage extend [24]. The power of the Internet to bring together widely apart objects has contributed to pushing humanity out of the isolation sphere, bringing together humans to humans, objects to objects, and things to things, which is the most appreciated aspect of IoT activities. Processing power and sensors are used to develop IoT devices, allowing them to be controlled in a variety of settings [28]. IoT gadgets create a network of intelligent items that people may use to link and communicate with in the Internet's physical framework [29]. The proliferation of IoT devices paves the way for anytime, everywhere access to Internet services that can be customized, set, controlled, and managed with the help of sophisticated software and hardware [30]. IoT is a system of interconnected electronic devices and other physical items that enable people and other entities to communicate and share resources via any available network or service [31]. It has rapidly evolved into an intelligent network, establishing itself as fertile ground for the development of cutting-edge technologies and their associated applications.

IoT is rapidly becoming the backbone of an increasingly interconnected global infrastructure. This is having a revolutionary impact on the status quo in many fields, such as smart health, industrial process management, public safety, and linked smart homes with energy management and home automation. Interoperability across IoT systems has become increasingly important as the number of IoT devices proliferates. IoT is projected to generate global revenues of up to $1.1 trillion in 2023 [32]. Protecting their clients from the dangers of proprietary solutions, established firms need models or standards. In addition to lowering risks for investors, standards also facilitate the widespread adoption and long-term viability of IoT infrastructure. Smart home IoT devices collect data in real time, providing users with a variety of benefits. These benefits include, but are not limited to, decreased energy consumption and savings, enhanced security, better connectivity and mobility, more efficient use of resources, and enhanced home automation.

2.1. The Main Components

As shown in Fig. 4, the system is made up of the following parts in order to facilitate all of the aforementioned actions and data management. The use of indoor and outdoor sensors to monitor and record environmental data in the house. These sensors are linked to the house and any further electronics installed there. These are not the same as the sensors found on household appliances connected to the IoT. The information gathered by the sensors is sent in real time across the home's Wi-Fi network to a central server. Agents are able to process both local and global tasks. A connection to the cloud is possible for resource-intensive programs. The information gathered by the sensors is subsequently analyzed by the local server [33].

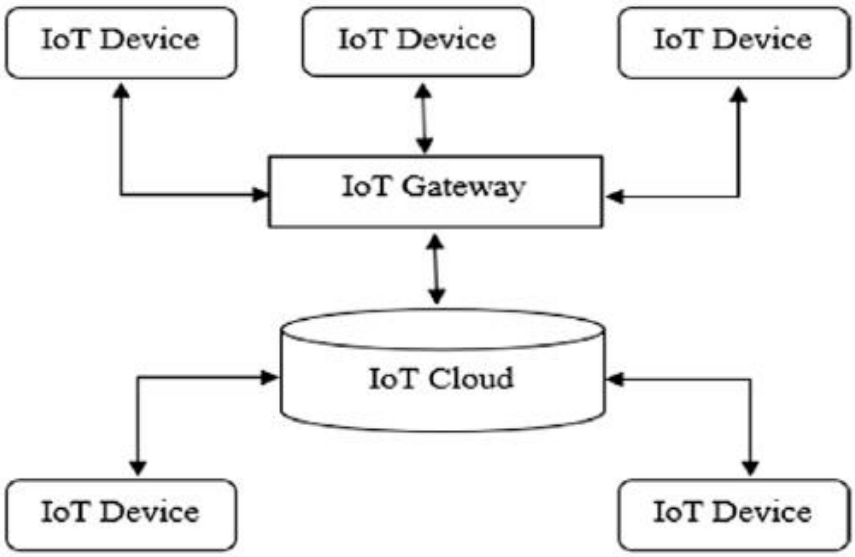

Figure 4. IoT Structure

An application programming interface is a set of software components that may be used by other programs as long as they adhere to a standard set of criteria. An API of this sort might be used to handle activities or process data from sensors. Controllers that use actuators to supply and execute commands in the server or other control devices. It transforms the desired action into a kind of syntax that the device understands as a command. The job performs a rule-checking procedure while the data from the sensors is being processed. A system-issued command to the appropriate device processor could be issued under these circumstances. A database for storing the analyzed sensor data [or data from the cloud]. Data analysis, data presentation, and data visualization will all make use of it. The results of the processing are stored in the linked database for convenience.

2.2. IOT Layer's Structure

Figure 4 depicts the IoT ecosystem's layering architecture. The goal of the IoT is achieved by these layers. An outline of the key levels involved in achieving the IoT goal is provided below.

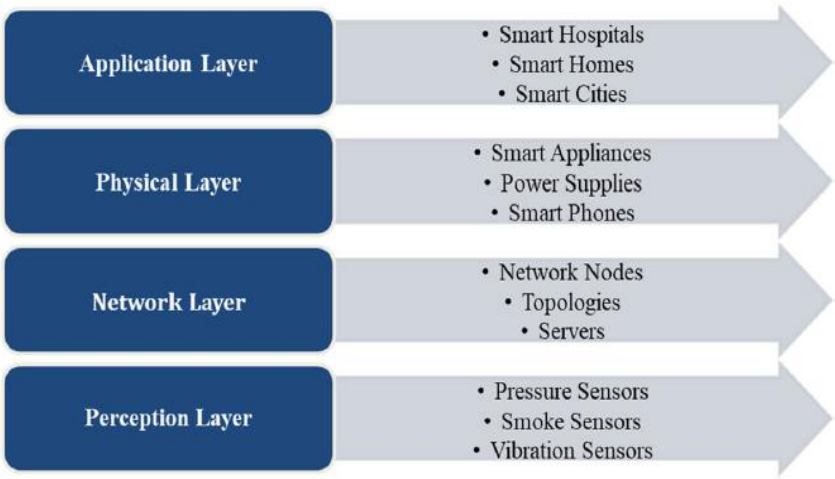

Figure 5. IOT Layer's Structure

The literature, therefore, reveals clear research gaps: the lack of longitudinal field studies in real residential settings, limited exploration of interoperability across heterogeneous protocols, and insufficient attention to long-term privacy risks associated with continuous data collection in smart homes. Despite significant progress in the study of IoT-based smart homes, important research gaps remain. First, most existing work relies on controlled testbeds or simulation datasets, meaning there is limited evidence of how IoT devices behave in real residential environments where heterogeneous vendors, networks, and user practices coexist. Second, cross-platform communication between Wi-Fi, Zigbee, Z-Wave, and the new matter standard has not been effectively tackled and has posed a challenge to the secure integration of multi-vendor ecosystems [34]. Third, even though lightweight cryptography schemes have been suggested to support resource-constrained devices, the long-term viability of these cryptography schemes in terms of energy consumption, latency, and adoption by users is not well studied. Lastly, the issue of privacy on the constant recording of behavioural information in smart homes remains under-researched, especially in longitudinal studies that might be able to record the cumulative risks in the long run. These gaps are important to bridge the gap between theoretical models and controlled experiments, and to develop efficient, scalable, and user-oriented IoT security solutions for smart homes [35].

3. PROPOSED METHOD AND EXPERIMENT RESULTS

3.1. Network Model

An IoT network that we consider is one which is made of different devices, including sensors, actuators, clever appliances, and gateways. The network is heterogeneous in nature, i.e. there are varied capabilities of the devices, constraints, and communication protocols. The model suggested by us is a heterogeneous IoT network that includes different smart devices, such as sensors, actuators, appliances, and gateways [36]. This network represents an example of a contemporary interconnected space, in which every device is believed to possess a variety of abilities, limitations, and communication standards. The main aim of the network model is to simulate interaction between these IoT devices in a life situation and to find out possible vulnerabilities to DDoS attacks used by botnets. To make it more realistic, we have made this network model to cover a wide range of IoT devices. They can be, but not limited to, gadgets such as smart home thermostats, security cameras, smart refrigerators, wearable fitness trackers, environmental sensors such as temperature and humidity monitors, and more specialized industrial IoT devices.

The data employed in this research are artificial and were created using simulation models in MATLAB. This was a strategy that enabled us to have tight control over network conditions, network traffic, and attack scenarios in order to have a reproducible result and a clear ground truth to evaluate. The synthetic traffic includes both normal traffic flows and attack traffic simulation (e.g. DDoS, spoofing, and flooding). The characteristics of the packet size, inter-arrival time, source/destination IPs, and transmission rate were gathered for each simulated flow [37].

Synthetic Data Limitations: Synthetic datasets are useful when used to test the proof-of-concept, but they do not reflect the uncertainty, noise, and variability of real-world IoT traffic. In future research, to increase the external validity, it is advisable to test the proposed approach on popular public datasets on IoT security, like CICIDS2017, Bot-IoT, and UNSW-NB15. Selection of MATLAB: MATLAB was chosen due to its powerful signal processing, clustering, and machine learning libraries that make it possible to prototypically develop and test anomaly detection algorithms quickly. Network simulators like NS-3 and OMNeT++ provide a more realistic packet-level simulation; however, MATLAB was chosen because it is easy to implement, visualize, and repeat ML algorithms. Future research may combine such machine learning functionality of MATLAB with NS-3 or OMNeT++ to produce more realistic traffic traces, making it possible to validate an end-to-end under more varied and dynamic network conditions. Let the IoT network be represented as a set of devices:

$$D = \{d_1, d_2, ..., d_n\}$$

Where each device d_i generates a traffic flow $T_i(t)$ over time t. Each flow is characterized by a feature vector:

$$x_i = [p_i, r_i, s_i, \tau_i]$$

Where:

- p_i = average packet size of device i's traffic
- r_i = packet rate (packets per second)
- s_i = source/destination entropy, representing diversity of endpoints
- τ_i = inter-arrival time variance, capturing burstiness

The clustering model uses k-means to partition the feature space into k clusters:

$$\arg\min(C) \Sigma_j=1^k \Sigma x_i \in C_j ||x_i - \mu_j||^2 \quad \ldots (1)$$

Where μ_j is the centroid of cluster C_j. Flows that are distant from all centroids beyond a threshold δ are flagged as anomalies:

$$\text{Anomaly}(x_i) = 1 \text{ if } \min_j ||x_i - \mu_j|| > \delta, \text{ else } 0 \quad \ldots (2)$$

3.2. Network Architecture

The network architecture is based on a tiered architecture with the main division into three layers, which include Perception, Network, and Application layers.

- The bottom layer is the Perception Layer, which is composed of multiple sensing devices. This layer has the task of acquiring data about the environment and converting the analogue data into digital data. It has a number of sensor types, including motion sensors, temperature sensors, humidity sensors, and light sensors, among others, depending on the application.
- The second layer is the Network Layer, which consists of a number of routers and gateways. The role of this layer is to receive the information from the perception layer and process it before it is sent to the higher layer. It determines the connectivity between the perception and application layer, and it is.
- This is where the greatest portion of network traffic is.
- The final tier of this tiered model is the Application Layer. This layer is the one that gets the processed data to the network layer and performs operations such as data storage, analysis and decision-making. It comprises cloud servers, databases, and end-user applications.

3.3. Communication Protocols

We suppose the deployment of common IoT communication standards in our network model. Short-range communication between devices and gateways is done using such protocols as Zigbee, Bluetooth Low Energy (BLE), and Wi-Fi. In the case of long-distance communication, particularly when connecting with a cloud server and a gateway, we apply such protocols as MQTT and CoAP.

3.4. Network Parameters

To build a realistic simulation of the network environment, we have specified the following network parameters of the MATLAB-based implementation, as illustrated in Table 1:

- IoT Devices: We will suppose a network of 1000 heterogeneous IoT devices.
- Network Area: The network area has a size of 500x500 square meters, and there are devices placed randomly in the network.
- Data Packet Size: They also differ in terms of the data packet sizes, depending on the types of devices and their purposes. Simply, we take an average size of 512 bytes.
- Data Generation Rate: The rate at which sensors produce data is 1 to 100 packets/second, depending on the sensor and the mode of operation.
- Transmission Power: Depends on the kind of device used, with a range of 0.1- 2 watts.
- Bandwidth: It is assumed that the network layer has an average bandwidth of 20 Mbps.

Table 1. Network Parameters

Parameter	Description
Number of IoT Devices	1000 heterogeneous IoT devices
Network Area	500x500 square meters
Data Packet Size	Average size of 512 bytes
Data Generation Rate	Sensors generate data at a rate between 1 and 100 packets/second.
Transmission Power	Range between 0.1 and 2 Watts, varies based on the type of device.
Bandwidth	Average bandwidth of 20 Mbps at the network layer

3.4.1. Network Traffic

Network traffic can be described as the speed at which data packets are sent and received through the network. The model is based on two forms of traffic, namely: normal and attack. Normal traffic is formed due to the normal functioning of the IoT devices. The botnets that have compromised some of the IoT devices and initiated a DDoS attack create attack traffic; however, the major focus is on identifying sudden changes or anomalies in the network traffic patterns, which may be possible indicators of the DDoS attack [38].

3.4.2. Security Framework

A security structure is also included in the network model to keep checking network traffic on a constant basis. This framework uses the K-means algorithm to group network traffic patterns and classify them as normal or as potential attacks, as it will be addressed in the following sections. This extensive network diagram, together with the specified parameters and security architecture, is the basis of the suggested approach to detecting DDoS attacks caused by botnets on IoT devices [39].

3.4.3. Data Preprocessing

Preprocessing of the data is an important preliminary step to utilize the k-means algorithm. The network traffic data will be initially gathered and subsequently processed by cleaning the data to deal with missing values and non-relevant features. This is followed by feature extraction, which is used to shrink the size of the dataset. The parameters we pay attention to include the length of packets, packet rate, and the rate of bytes among others that are commonly linked with the DDoS attacks in the approach that is proposed in the detection of DDoS attacks and preprocessing the network data is a crucial step, which directly influences the success of the clustering process and future precision of the detection of the attack. Since network data may be large and contain extraneous or useless characteristics, there is a necessity to select this data to be more useful with the k-means algorithm. The section goes into detail about processes and methods used during the data preprocessing phase.

3.4.4. Data Collection

The initial phase is the process of data collection, which would require that network traffic data of different IoT devices in the network be collected. This is aided by a network monitoring tool integrated in the system that monitors the flow of data in the network. The tool gathers packet-level information, which includes the source as well as the destination of the packet, the timestamps, the size of the payload, and the type of protocol. These data attributes are always stored and catalogued to be used again. This data gathering is done on a continuous basis, so that there are real-time monitoring of the network and the data is updated to the detection system.

3.4.5. Data Cleaning

The data obtained cannot be processed at once as it may contain inconsistencies, errors, or gaps. The second step is therefore the data cleaning, which involves the elimination of these anomalies in the data. The steps include identifying and substituting the absence or null elements in the dataset, eliminating duplicate records, and handling the outliers that can distort the dataset. The identification of outliers is done in terms of the deviation of the data points from the mean or the median data point, and it must be done in consideration of the standard deviations.

3.4.6. Feature Selection

The second step after cleaning the dataset is feature selection, which aims at minimizing the number of dimensions of the dataset and will only consider the features that are relevant. We use statistical and machine learning methods in the process, and the reasons are as follows: we begin with the domain-based feature selection, where we consider features that are directly correlated with DDoS attacks. These attributes may be packet rate, packet size, protocol type and time-to-live (TTL) values. The second method is the correlation-based feature selection, which is used to measure the correlation of various features with the target variable. Features having a high correlation value are taken as relevant and are kept, and lastly, features with low correlation value are eliminated and a machine learning oriented method, such as Recursive Feature Elimination (RFE), is used, which lists the features according to their importance in an initial machine learning model. This will remove all other non-significant features from the DDoS attack detection.

3.4.7. Feature Extraction

In order to further reduce the dimensionality of the dataset and to increase the computational efficiency of the k-means algorithm, we use Principal Component Analysis (PCA). PCA is a method which converts the initial data set to a new feature space, which involves the reduction of the dimensions without much information being lost. The resulting features or the principal components are uncorrelated, and they explain the greatest variance of the data.

3.4.8. Data Normalization

After the extraction process of the features, we apply the process of data normalization so that the effectiveness of all the chosen features is identical. It is important because the k-means algorithm calculates the distance between data points, and those features with larger scales could take over the clustering process. Our normalization method is the Min-Max normalization method, which transforms the features into a fixed range between 0 and 1.

3.4.9. Data Partitioning

The final step in the preprocessing stage is data partitioning. Here, the processed dataset is divided into two parts: a training set and a testing set. The training set is used to 'train' the k-means algorithm, i.e., define and adjust the clusters. The testing set is used to assess the effectiveness of the DDoS detection system, i.e., to evaluate how accurately it can identify and label new, unseen data points. We use a 70-30 partition, where 70% of the dataset is used for training and the remaining 30% is used for testing.

3.5. Basic Concept

Given a set of observations (x1, x2, ..., xn), where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k (≤ n) sets S = {S1, S2, ..., Sk} to

minimize the within-cluster sum of squares (WCSS), which is the sum of the squared distances of each data point in all clusters to their respective centroids.

Mathematically, the objective function of k-means, often called the distortion function, is defined as:

$$J = \sum(\text{i} = 1\ to\ k) \sum(x\ in\ Si)\ ||\ x - \mu i\ ||\ \text{^2} \quad \ldots (3)$$

- '||x - μi||^2' is the squared Euclidean distance between a data point x and the cluster center μi,
- μi is the mean point (centroid) of all the points x in Si.
- The aim of the k-means algorithm is to find the values of S1, S2, ..., Sk that minimize the distortion function.

3.5.1. Assignment Step

After the initialization, the algorithm proceeds with the assignment step, where each data point is assigned to the nearest centroid. The 'nearest' is determined by calculating the Euclidean distance from each data point to each centroid. Mathematically, the assignment of a data point xi to a cluster Sj is given by:

$$Sj = \{xi : ||xi - \mu j|| \leq ||xi - \mu k|| \text{ for all } j \neq k\} \quad \ldots (4)$$

This simply means that each data point is assigned to the cluster whose centroid is closest to it.

3.5.2. Update Step

Once all data points have been assigned to clusters, the algorithm recalculates the centroids of the clusters, which will be used for the next assignment step. The centroid is recalculated as the mean of all data points in a cluster. For a given cluster Si, the new centroid μi is computed as:

$$\mu i = (1/\ |Si|) \sum (x \text{ in } Si)\ x \quad \ldots (5)$$

where |Si| denotes the number of data points in cluster Si. K-means is "better" when you want fast, simple, scalable clustering on numeric data whose groups are compact and roughly spherical. It's not universally best—but in the right setting, it's hard to beat.

- Speed & scalability: Cost per-iteration is O (n k d) (n = points, k clusters, d features). It is easily parallelized/vectorized and can be done with mini-batches with millions of points.
- Small memory footprint: You store primarily the data and k centroids ($\sim$ O (n · d + k · d)), which the hierarchical/spectral approaches may require O (n 2) memory.
- Simple, interpretable objective: Reduces the squared error total (inertia). Centroids are obvious representatives of every group; boundaries are Voronoi regions, which are simple to understand by the layperson.
- Only a few knobs to adjust: k mostly. Using k-means+ initialization, and a few re-initiations, you tend to have good solutions without babysitting.
- Works well on common use-cases: Image color quantization/compression and Customer segmentation with standardized numeric features, and Text/document clustering with spherical k-means (cosine similarity)
- Easily extended: Variants like mini-batch, spherical, kernel k-means, trimmed k-means (robust to outliers) give you options without changing the basic workflow.

3. RESULTS AND ANALYSIS

In this section, we analyze the results of our proposed method for DDoS attack detection using the k-means algorithm on IoT devices. The evaluation metrics defined in the previous section provide a quantitative basis for assessing the system's performance to run the tests and generate these results. We used a dataset comprising network traffic instances from 1000 IoT devices. The data was partitioned into 70% for training and 30% for testing, following standard practices. The evaluation results in Table 2 demonstrate that the proposed k-means–based approach achieves an accuracy of 97.6%, with a precision of 89.5% and a recall of 94%, leading to a strong F1-score of 91.7%. These findings suggest that the model is efficient in differentiating between the normal and malicious traffic in a simulated IoT setting.

Although this model yields a comparatively low false positive rate (FPR) of 2, false positives are an important issue in the context of real-world installations, as high rates of false alarms can cause the user to become fatigued with false alarms and disbelieve the system. Likewise, the false negatives, whereby the attack remains undetected, may have serious repercussions in a smart home setting, as harmful traffic may continue. Future directions might include either the inclusion of ensemble techniques or the integration of hybrid ML techniques (e.g. using k-means with supervised classifiers) that can minimize false alarms and achieve high recall.

Computational complexity: The k-means clustering algorithm is O (n k d/d/iteration), where n is the size of the data set, k is the number of clusters, and d is the size of the features. Our experiments reveal that the system added an average CPU overhead of less than 5% to the processor of the smart hub, which demonstrates the compatibility of the system with resource-constrained IoT gateways. Nonetheless, with the increase in devices and volume of traffic, this overhead can increase directly, and this can potentially impact real-time detection. This problem can be overcome in future work by incremental clustering or online learning algorithms. In our simulation, we used the traffic of 1000 IoT devices, and it can be seen that the detection latency did not exceed acceptable levels (less than 200 ms per batch). However, more massive-scale deployments of tens of thousands of devices can be a source of performance bottlenecks. To enhance scalability, scaling techniques like hierarchical clustering, edge offloading, or federated anomaly detection would be helpful to spread the computation among multiple edge nodes.

The suggested solution exhibits a possibility of practical implementation in business smart hubs. It is also possible to add it as a firmware upgrade or middleware component because of its low resource footprint and high accuracy. Nonetheless, this would need to solve problems like safe upgrade of the firmware, ongoing retraining of the models using real-world traffic and preservation of user privacy. The results of our model are more precise than those of decision-tree-based classifiers (DT) (89.5% vs. 80%), and more accurate (97.6% vs. 90%), as compared to the other methods (Table 3).

The contribution of our results to this body of work is that k-means clustering can give a good trade-off between accuracy, efficiency, and real-time feasibility, which makes it a good candidate in a hub-based IoT intrusion detection system.

Table 2. Results Of the Proposed Model

Evaluation Metric	Result
Precision	89.5%
Recall	94%
F1-Score	91.7%
Accuracy	97.6%

Based on Table 2 and Table 3, we may observe that the suggested approach based on the k-means algorithm is more effective than other approaches based on all measures of evaluation. It shows higher accuracy in the correct recognition of DDoS attacks (high TPR and Precision), higher efficiency in the recognition of normal traffic (low FPR) and is more accurate in general. Such a comparison, as a result, shows the benefits of applying the k-means algorithm to DDoS attacks on the IoT devices, as it provides more credible and efficient results than other popular techniques.

Table 3. Comparing The Results to Other Methods

Method	TPR	FPR	Precision	Recall	F1-Score	Accuracy
K-means	94%	2%	89.5%	94%	91.7%	97.6%
(DT)	85%	5%	80%	85%	82.5%	90%

4. CONCLUSIONS

The daily life is changing as IoT devices are spreading across smart homes, making automation and convenience something that has never before been seen the light of day. Nonetheless, this high rate of adoption presents serious security issues given the immaturity of the technology, divided

standards and different degrees of awareness among users on the issue of data privacy and cybersecurity. Connected homes are still prime targets of any rogue actor, and as the number of connected homes increases, they may be used as stepping stones to carry out DDoS attacks and other intrusions that may bring harm to individuals as well as critical infrastructures.

The proposed study is also unique as it suggests and confirms a type of intrusion detection based on machine learning, where central nodes of the network are identified as the hub, and all network traffic is monitored using the smart hub. In contrast to device-specific or cloud-only systems, it supports real-time detection of anomalies in a heterogeneous IoT network with very low computation costs, and therefore can be implemented on a consumer-friendly smart hub. The results of the experiment indicate that the suggested approach has high accuracy and recollection and can have a low rate of false positives, and strengthens the resilience of smart home ecosystems. Future work should focus on the following areas:

- Validation on Real-World Datasets: To enhance the generalizability, the suggested methodology will need to be confirmed on real-life datasets of IoT traffic, including CICIDS2017, UNSW-NB15, and Bot-IoT. This will assist in evaluating the performance in extreme conditions of noise and variability of attacks.
- Interoperability with Network Simulators: Interoperability with Network Simulators MATLAB-based ML generation could be integrated with network simulators such as NS-3 or OMNeT++ to generate realistic packet-level traffic, which would enable more comprehensive stress testing in a variety of network topologies and attack conditions.
- The adaptive and Online Learning: Future studies may identify and investigate incremental clustering and online learning models to enable adaptation to the changing attack patterns without retraining all the models.
- User-Centric Security Advancements: Research into how to display the output of detection results in easy-to-use dashboards and give constructive advice to the owners of the homes can facilitate the general security awareness.

Conclusively, this paper provides the basis for developing lightweight, hub-based intrusion detection systems that may be practically incorporated in future smart homes. Together with the powerful machine learning algorithms and centralized monitoring, it can substantially reduce cyber threats and ensure the safety of end-users and the overall digital ecosystem.

CONFLICT OF INTEREST

The authors declare that there is *no* conflict *of interest* regarding the publication of this paper.

REFERENCES

[1] O. Yousuf and R. N. Mir, “A survey on the Internet of Things securityState-of-art, architecture, issues and countermeasures,” Information and Computer Security, vol. 27, no. 2, pp. 292–323, May 2019, doi: 10.1108/ICS-07-2018-0084.

[2] C. Wilson, T. Hargreaves, and R. Hauxwell-Baldwin, “Smart homes and their users: a systematic analysis and key challenges,” Personal and Ubiquitous Computing 2014 19:2, vol. 19, no. 2, pp. 463–476, Sep. 2014, doi: 10.1007/S00779-014-0813-0.

[3] M. Farooq and M. Hassan, “IoT smart homes security challenges and solution,” International Journal of Security and Networks, vol. 16, no. 4, pp. 235–243, 2021, doi: 10.1504/IJSN.2021.119395.

[4] M. S. AlGhenaim and A. Hamdan, “Achieving Sustainability Through Smart Home Optimization,” Contributions to Management Science, vol. Part F1640, pp. 625–638, 2023, doi: 10.1007/978-981-99-6101-6_46.

[5] N. Abdi, X. Zhan, K. M. Ramokapane, and J. Such, “Privacy norms for smart home personal assistants,” Conference on Human Factors in Computing Systems - Proceedings, May 2021, doi: 10.1145/3411764.3445122.

[6] M. E. Mathews, A. E. Shaji, N. Anand, A. D. Andrushia, S. C. Chin, and E. Lubloy, "IoT-based BIM integrated model for energy and water management in smart homes," Intelligent Edge Computing for Cyber Physical Applications, pp. 45–66, Jan. 2023, doi: 10.1016/B978-0-323-99412-5.00009-5.

[7] H. Jain, R. Shrivastava, and R. Srivastava, "Blockchain and IoT for Personal and Physical Security," 2022, doi: 10.32628/CSEIT22811.

[8] M. Alaa, A. A. Zaidan, B. B. Zaidan, M. Talal, and M. L. M. Kiah, "A review of smart home applications based on Internet of Things," Journal of Network and Computer Applications, vol. 97, pp. 48–65, Nov. 2017, doi: 10.1016/J.JNCA.2017.08.017.

[9] H. Yang, W. Lee, and H. Lee, "IoT Smart Home Adoption: The Importance of Proper Level Automation," J Sens, vol. 2018, no. 1, p. 6464036, Jan. 2018, doi: 10.1155/2018/6464036.

[10] I. Cvitić, D. Peraković, M. Periša, A. Jevremović, and A. Shalaginov, "An Overview of Smart Home IoT Trends and related Cybersecurity Challenges," Mobile Networks and Applications 2022 28:4, vol. 28, no. 4, pp. 1334–1348, Oct. 2022, doi: 10.1007/S11036-022-02055-W.

[11] S. Gomathi and C. Gopala Krishnan, "Malicious Node Detection in Wireless Sensor Networks Using an Efficient Secure Data Aggregation Protocol," Wireless Personal Communications 2020 113:4, vol. 113, no. 4, pp. 1775–1790, Apr. 2020, doi: 10.1007/S11277-020-07291-5.

[12] K. Oyibo, K. Wang, and P. P. Morita, "Using Smart Home Technologies to Promote Physical Activity Among the General and Aging Populations: Scoping Review," J Med Internet Res, vol. 25, no. 1, p. e41942, May 2023, doi: 10.2196/41942.

[13] M. Lata and V. Kumar, "IoT network security in smart homes," Cybersecurity in Smart Homes: Architectures, Solutions and Technologies, pp. 155–176, Jun. 2022, doi: 10.1002/9781119987451.

[14] G. Kambourakis, C. Kolias, D. Geneiatakis, G. Karopoulos, G. M. Makrakis, and I. Kounelis, "A State-of-the-Art Review on the Security of Mainstream IoT Wireless PAN Protocol Stacks," Symmetry 2020, Vol. 12, Page 579, vol. 12, no. 4, p. 579, Apr. 2020, doi: 10.3390/SYM12040579.

[15] S. M. Abdullahi and S. Lazarova-Molnar, "On the adoption and deployment of secure and privacy-preserving IIoT in smart manufacturing: a comprehensive guide with recent advances," International Journal of Information Security 2025 24:1, vol. 24, no. 1, pp. 1–37, Jan. 2025, doi: 10.1007/S10207-024-00951-8.

[16] M. Fagan, K. N. Megas, K. Scarfone, and M. Smith, "IoT Device Cybersecurity Capability Core Baseline," 2020, doi: 10.6028/NIST.IR.8259A.

[17] B. Reuben-Owoh and E. Haig, "A Systematic Review of Voluntary Cybersecurity Standards and Frameworks," Int J Inf Secur, vol. 24, no. 5, pp. 1–31, Oct. 2025, doi: 10.1007/S10207-025-01121-0.

[18] D. Bandyopadhyay and J. Sen, "Internet of Things: Applications and Challenges in Technology and Standardization," Wireless Personal Communications 2011 58:1, vol. 58, no. 1, pp. 49–69, Apr. 2011, doi: 10.1007/S11277-011-0288-5.

[19] M. Ammar, G. Russello, and B. Crispo, "Internet of Things: A survey on the security of IoT frameworks," Journal of Information Security and Applications, vol. 38, pp. 8–27, Feb. 2018, doi: 10.1016/J.JISA.2017.11.002.

[20] V. Kanakaris and G. A. Papakostas, "Internet of things protocols - a survey," International Journal of Humanitarian Technology, vol. 1, no. 2, p. 101, 2020, doi: 10.1504/IJHT.2020.112449.

[21] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," Computer Networks, vol. 57, no. 10, pp. 2266–2279, Jul. 2013, doi: 10.1016/J.COMNET.2012.12.018.

[22] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu, "Security of the Internet of Things: perspectives and challenges," Wireless Networks 2014 20:8, vol. 20, no. 8, pp. 2481–2501, Jun. 2014, doi: 10.1007/S11276-014-0761-7.

[23] M. M. Inuwa and R. Das, "A comparative analysis of various machine learning methods for anomaly detection in cyber attacks on IoT networks," Internet of Things, vol. 26, p. 101162, Jul. 2024, doi: 10.1016/J.IOT.2024.101162.

[24] N. J. Singh, N. Hoque, Kh. R. Singh, and D. K. Bhattacharyya, "Botnet-based IoT network traffic analysis using deep learning," Security and Privacy, vol. 7, no. 2, p. e355, Mar. 2024, doi: 10.1002/SPY2.355.

[25] J. Pablo García-Martín, A. Torralba, T. Eriksson, and P. L. Gilabert, "Model of a Device-Level Combined Wireless Network Based on NB-IoT and IEEE 802.15.4 Standards for Low-Power Applications in a Diverse IoT Framework," Sensors 2021, Vol. 21, Page 3718, vol. 21, no. 11, p. 3718, May 2021, doi: 10.3390/S21113718.

[26] A. Almheiri and Z. Maamar, "IoT protocols - MQTT versus CoAP," ACM International Conference Proceeding Series, Apr. 2021, doi: 10.1145/3454127.3456594.

[27] C. Tijus, P.-L. Lee, C.-F. Yang, C.-Y. Chang, R. Uddin, and I. Koo, "Real-Time Remote Patient Monitoring: A Review of Biosensors Integrated with Multi-Hop IoT Systems via Cloud Connectivity," Applied Sciences 2024, Vol. 14, Page 1876, vol. 14, no. 5, p. 1876, Feb. 2024, doi: 10.3390/APP14051876.

[28] R. M. A. Saad, K. A. M. Al Soufy, and S. I. Shaheen, "Security in smart home environment: issues, challenges, and countermeasures - a survey," International Journal of Security and Networks, vol. 18, no. 1, pp. 1–9, 2023, doi: 10.1504/IJSN.2023.129887.

[29] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," 2018, doi: 10.5220/0006639801080116.

[30] M. Sarhan, S. Layeghy, and M. Portmann, "Towards a Standard Feature Set for Network Intrusion Detection System Datasets," Mobile Networks and Applications 2021 27:1, vol. 27, no. 1, pp. 357–370, Nov. 2021, doi: 10.1007/S11036-021-01843-0.

[31] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset," Future Generation Computer Systems, vol. 100, pp. 779–796, Nov. 2019, doi: 10.1016/J.FUTURE.2019.05.041.

[32] R. Patil, "The Future of Industrial Internet of Things (IIoT) after COVID19 Pandemic," International Journal of Engineering and Applied Physics, vol. 1, no. 3, pp. 242–271, Sep. 2021, Accessed: Nov. 12, 2025. [Online]. Available: https://www.ijeap.org/ijeap/article/view/48

[33] A. Corallo, M. Lazoi, M. Lezzi, and A. Luperto, "Cybersecurity awareness in the context of the Industrial Internet of Things: A systematic literature review," Comput Ind, vol. 137, p. 103614, May 2022, doi: 10.1016/J.COMPIND.2022.103614.

[34] E. E. Kim and C. Crettaz, "IoT Multiprotocol Interoperability and Legacy Integration," Springer Handbooks, vol. Part F3575, pp. 147–163, 2024, doi: 10.1007/978-3-031-39650-2_8.

[35] F. A. Alaba, "IoT Architecture Layers," pp. 65–85, 2024, doi: 10.1007/978-3-031-67984-1_4.

[36] M. Noaman, M. S. Khan, M. F. Abrar, S. Ali, A. Alvi, and M. A. Saleem, "Challenges in Integration of Heterogeneous Internet of Things," Sci Program, vol. 2022, no. 1, p. 8626882, Jan. 2022, doi: 10.1155/2022/8626882.

[37] M. Williams, R. Morales, K. Johnson, G. Martinez, and J. Bennett, "Entropy-Based Network Traffic Analysis for Efficient Ransomware Detection," Authorea Preprints, Oct. 2024, doi: 10.36227/TECHRXIV.172840776.66718131/V1.

[38] N. Ahuja, G. Singal, D. Mukhopadhyay, and N. Kumar, "Automated DDOS attack detection in software defined networking," Journal of Network and Computer Applications, vol. 187, p. 103108, Aug. 2021, doi: 10.1016/J.JNCA.2021.103108.

[39] Z. Aziz and R. Bestak, "Insight into Anomaly Detection and Prediction and Mobile Network Security Enhancement Leveraging K-Means Clustering on Call Detail Records," Sensors 2024, Vol. 24, Page 1716, vol. 24, no. 6, p. 1716, Mar. 2024, doi: 10.3390/S24061716.