https://doi.org/10.33472/AFJBS.6.5.2024. 2758-2773

AfricanJournalofBiological Sciences

A Novel Hybrid Intruder Detection System for IoT Enabled Smart Home using Edge Computing

Priyanka Nanda*,Kamal Malik, Kapil Mehta

Research Scholar,Professor, Associate Professor

Department of Computer Science & Engineering

CT University, Punjab, India

Article History
Volume 6, Issue 5, 2024
Received: 01 May 2024
Accepted: 09 May 2024
doi:*10.33472/AFJBS.6.5.2024. 2758-2773*

Abstract

An information system security (IDS) is designed to stop any illegal access. Data availability, confidentiality, or integrity might be threatened by any kind of access. Monitoring network traffic and/or resource use and sounding an alarm in the event that malicious behaviour is detected is how an intrusion detection system (IDS) does its job.

Depending on the method used by the system to identify intrusions, the IDSs can be divided into two primary families: those that learn about the typical behaviour of the system and report any unusual events, or those that cross-check recorded events through a database of well-known infiltration methods.

Signage-based and anomaly-based are the names given to these approaches, respectively.

Keywords: Novel Hybrid IDS, Edge Computing, IoT, Home Security Systems, Machine Learning Approaches.

I.Introduction

1.1 Signature-based IDS

An attack database containing known attack "signatures" is utilised by a kind of intrusion detection systems known as signature-based IDSs (SIDSs). The present activities' signatures are retrieved, and they are compared to those in the database using matching techniques and/or protocol conventionality tests. An alert is set off if there is a match. They can function both offline, where system activity logs are examined, and online, where they can proximately monitor hosts and provide real-time alarms. Misuse Detection and Knowledge-based Detection are other names for this kind of IDS in the literature.

1.2 Anomaly-basedIDS

The shortcomings of signature-based intrusion detection systems led to the development of anomaly-based intrusion detection systems, or AIDS. AIDS usually involve a training phase where a speculativebehaviour model of the system is built. Following deployment, the IDS monitors computer hosts and compares their behaviour to what is projected. If the IDS detects a significant deviation from the hosts' behaviour as predicted by the model, it may raise an alarm. By not comparing the host's current behaviourwith attack signatures kept in a database, this technique could make it possible for an anomaly-based intrusion detection system to identify zero-day attacks. Since interacting with a target would probably cause the intrusion detection system to issue an alarm, Additionally, an attacker finds it more challenging to comprehend anomaly-based intrusion detection systems a target host's regular behaviour without doing transactions with it.

1.3 MachineLearning-basedAIDS

Training with ground-truth labels is not necessary in machine learning. We refer to the two categories of learning strategies as supervised and unsupervised learning. The supervised learning approach involves providing the data to the machine learning algorithm together with their true labels (abnormality / non anomaly). Training approaches such as Artificial Neural Networks (ANNs), Decision Trees, Support Vector Machines (SVMs), and others employ this method.

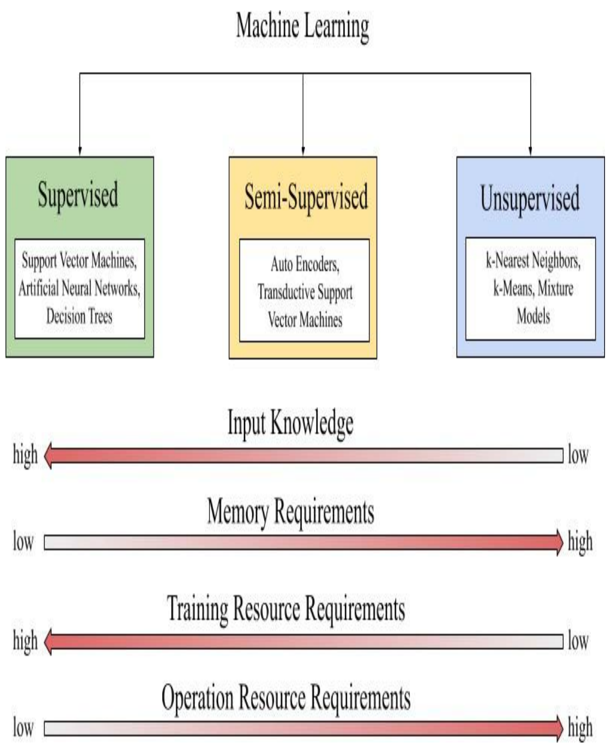

Machine Learning
Supervised
Support Vector Machines, Artificial Neural Networks, Decision Trees
Semi-Supervised
Auto Encoders, Transductive Support Vector Machines
Unsupervised
k-Nearest Neighbors, k-Means, Mixture Models
Input Knowledge
high
low
Memory Requirements
low
high
Training Resource Requirements
high
low
Operation Resource Requirements
low
high

Figure 1 – A broad categorization of machine learning methods used in IDS and their primary prerequisites

1.4 IDS for IoT

IoT-specific and IoT-agnostic IDSs are two different types of IDSs that target the Internet of Things, as was mentioned in Section 1. An IoT-specific intrusion detection system (IDS) targets devices that make use of certain communication technologies, such 6LoWPAN, BLE, LoRaWAN, etc. On the same network as the device, this sort of IDS should be installed. Their predictions are often executed by means of messages received from Internet of Things (IoT) devices that provide control information relevant to that technology, like protocol compliance checks. But, IoT-agnostic IDSs are independent of any specific IoT technology. They employ the available data on the technology—such as TCP/IP traffic—that the devices are currently using. An edge environment is a good fit for this kind of IDS since it can handle with trafficgeneratedbyheterogeneousdevicesleveraging different communication technologies.

1.5 The Edge-Enabled Approach

A number of factors led to the proposal of edge computing to improve the dependability and characteristics of conventional Internet of Things applications [60, 61]. Computing, storage, and administrative duties can be delegated to the edge nodes by the IoT application. Real-time network management, improved data management, and a reduction in latency are a few of the anticipated quality improvements. Here, security apps like an IDS might also be "migrated" to the edge (see Figure 1's red box for an example). With increased computing power at its disposal to employ more sophisticated algorithms and greater storage capacity to store systems logs for later analysis or to execute memory-intensive operations, an intrusion detection system (IDS) could gain from this transformation.

For real-time Internet of Things applications, an edge node may also be able to provide lower latency than the cloud. Furthermore, an IDS that is placed at the edge need to be IoT-agnostic, which means it shouldn't rely on any particular IoT communication platforms. Without the need to install a separate IoT-specific IDS for each subnetwork of devices, such an IDS may handle numerous heterogeneous devices utilizing various communication methods in an integrated way.

II. Literature Review

[1] According to Mohammed Ali Al-Garadi and Amr Mohamed (2020), the Internet of Things (IoT) connects billions of intelligent objects that can converse with one another with little help from humans. By 2020, there will be a projected 50 billion devices in this industry, making it one of the fastest growing in computer history. While there are many real-world smart applications that can enhance quality of life, Internet of Things technologies are essential to many of these applications. However, new security issues have emerged as a result of the interconnectedness of IoT systems and the interdisciplinary elements involved in their implementation. Putting security measures in place for Internet of Things devices and their inherent vulnerabilities is ineffective. These methods include encryption, authentication, access control, network security, and application security.

[2] In 2020, Roberto Magan-Carrio talked on the necessity for new security event monitoring and detection systems to handle the new problems this scenario presents. One such problem is the real-time requirement that enables prompt security event detection and, as a result, prompt attack response. In this regard, intrusion detection systems are commonly employed, despite the fact that their applicability in actual settings is restricted by the fact that their assessment frequently depends on the usage of predetermined network statistics. An easy-to-use, real-time tool for tracking and identifying security occurrences is presented in this work.

[3] According to AbdelouahidDerhab (2020), connected things generate massive amounts of data traffic that power big data analytics, which may be used to spot unusual traffic and uncover patterns that weren't previously seen.

[4] Mateusz Krzysztoń (2022) talked on how machine learning algorithms are used in IoT networks to identify harmful behavior and classify data. As part of the suggested approach, many watchdogs that are dispersed throughout the network under consideration and are primarily in charge of processing local traffic collaborate to make decisions.

[5] Using the OMNET++ INET emulator, Marcelo Paulon J. V. (2021) described data collecting on a Bluetooth mesh network. The wandering sink node that collects the data might be a smartphone or other portable device that is being carried by a cyclist, a pedestrian, an animal, or a drone.

[6] Anca Jurcut (2022) and Eric Gyamfi (*) talked about how IoT devices gather important data that helps businesses or individual consumers make important decisions that affect their daily lives. The majority of these Internet of Things devices have poor CPU speeds, little memory, and minimal energy storage. Because these devices lack the capability to run current general-purpose security software, they are therefore susceptible to cyber-attacks. IoT networks become inherently risky as a result.

[7] Marcelo Paulon J. V. (2022) talked about utilizing the OMNET++ INET simulator to assess irregular data gathering on a Bluetooth mesh network. The wandering sink node that collects the data might be a smartphone or other portable device that is being carried by a cyclist, a pedestrian, an animal, or a drone. In difficult-to-reach places without internet connectivity, the sink node might establish a mesh network connection to gather sensor data. We were able to suggest and assess methods for mobility-aware, adaptive routing of sensor data towards the sink node after integrating Bluetooth Mesh relay extensions, Low Power, and Friend features in OMNET++.

[8] According to Thiago de Souza Lamenza (2022), testing network algorithms on actual hardware in physical settings is a crucial step in closing the knowledge gap between theory and practice in the field. It's also an intriguing method to learn more about emerging technologies like Bluetooth Mesh. Using ESP-32 microcontrollers, we developed a Bluetooth Mesh data gathering approach and deployed it in both indoor and outdoor environments.

[9] The fundamental characteristics of Bluetooth Low Energy (BLE) that enhanced its potential and opened the door for the creation of new networking paradigms were discussed by Giacomini E. and D'Alterio F.. The ability for a single node to participate in more than a piconet and to assume both the role of master and slave makes it possible the formation of multi-hop networks that can be used in several emerging application scenarios. Additionally, the inherent low power consumption at the cost of contained throughput makes this technology appealing for Internet of Things (IoT), where power memory and capacity constrained devices exchange messages at relatively low data rates.

[10] Bing Lin1 , Hang Su2(2020) talked about how to define typical application scenarios, propose new measurement metrics, and design an experimental system. After analyzing the experimental results, we have determined that under reasonable improvements, the impact of broadcast storms on mesh message control success rate should not be overstated.

[11] Mansura Habiba(2023) talked about the physical devices to connect directly to the internet and upload data continuously. Insecure access makes IoT platforms vulnerable to different network intrusion attacks. As a result, the Intrusion Detection System (IDS) is a core component of a modern IoT platform.

[12] Ibbad and Hafeez (2020) talked on how to use conventional endpoint and network security solutions to safeguard IoT deployments. We introduce IoT-Keeper, a lightweight system that protects IoT connectivity, as a solution to this issue. We offer an anomaly detection method that IoT-Keeper employs to analyze traffic at edge gateways.

[13] According to Tushar M. Jois (2021), Internet-of-things (IoT) devices are the fundamental components of an increasing number of automated smart settings, such as offices, schools, residences, and neighbourhoods. IoT devices have a lot of market area, are growing quickly, and are prone to privacy issues and unidentified vulnerabilities.

[14] Convolutional neural networks have been applied to a wide range of tasks, the most common being object identification, picture recognition, and speech recognition, according to Arohan Ajit (2020). All of these industries have worked to improve these networks' accuracy and efficiency.

[15] IbomoiyeDomorMienye (2020) talked about the two-stage approach that is suggested as an efficient way to forecast cardiac disease. To find the optimal representation of the training data, an enhanced sparse autoencoder (SAE), an unsupervised neural

network, must be trained in the first stage. Using the learnt records, an artificial neural network (ANN) is used in the second step to forecast the health state.

[16] In 2020, MANASSÉS RIBEIRO talked about how the classifier is trained using samples of a class that is regarded as normal, allowing extraordinary patterns to be recognized as anomalies. In fact, since one or more clusters might characterize distinct features of normalcy, how the normal class is represented in the feature space matters for real-world situations.

[17] Sajjadul Islam Nader(2022) talked about Modern civilization is gradually improving dayby day.One of themostpromising developmentin modern technology is homeautomation system. The rapid advancementin technology and automation systemmadehumanlifemuch easier.In this project, an advancedsmarthome with animprovedsecurity system isintroducedusingthe internetof things (IoT). The usercan control andmonitor the system using web-basedinterfaces like IFTTT or byusing smartphone applications like Blynk. The main purpose of this project is to makea low-cost smart home automation and security system to ensure a better life for usandourfamilies.

[18] Hikmat Yar (2021) talked about the Smart home applications are ubiquitous and havegained popularity due to the overwhelming use of Internet of Things (IoT)-basedtechnology.Therevolutionintechnologieshasmadehomesmoreconvenient,efficient, and even more secure.

[19] In his talk, Muaadh A. Alsouf (2021) talked about how the Internet of Things (IoT) has come to be seen as a way to make people's lives better by offering a variety of smart and connected devices and applications in a number of areas, including smart farming, smart homes, smart transportation, smart health, smart grid, smart cities, and smart environment. Cyberattacks might, however, affect IoT devices. Researchers have appropriately accepted the application of deep learning algorithms as a means of protecting the Internet of Things environment.

[20] Debatosh Debnath (2021) and Faisal Alghayadh talked about how smart home systems regulate appliances, lighting, temperature, and security camera systems. These internet-connected gadgets and sensors are easily vulnerable to assault because of their connectivity.

III. Problem Formulation

Followingarethevariousinferencesdrawn fromtheliteraturesurvey

- The researchers have not particularly addressed the detection of real-time identity-based parameters of invader people.
- Issues with data latency and smart object reaction times exist in cloud computing.
- Not much research has been done to measure identification criteria so that users and security professionals may make informed decisions.
- The smart object in cloud computing has an issue with power consumption and energy inefficiency.
- The cloud computing smart object's storage capacity and bandwidth costs are an issue.
- Using previously developed techniques, the homeowner receives a push notice as soon as the door is opened from the sensors installed on it. However, if the user is permitted to enter, the alarm system is temporarily turned off to allow the user to enter.
- Not much evidence has been provided to support the regular monitoring of home security and associated features by the monitoring authorities, which jeopardizes home security.

IV. Proposed Methodology

4.1 DeviceClassification

Devices (such as motion sensors, security cameras, smart plugs and lights, etc.) can be categorized by classifiers using network packets. or to collect fingerprints from devices so that an intrusion detection system (IDS) can trigger alerts based just on passive network readings in the event that an unauthorized device joins the network. IDSs intended for edge deployment are limited to using network traffic to detect intrusions. Statistical techniques are used to justify each feature's usefulness. The researchers employed 15-minute intervals and a segment of those intervals, dubbed the "activity period," to examine traffic patterns and gather data. The time interval between the first and last packets being received by the device is represented by this time component. Depending on the device, the duration of this part of the activity may vary. Device traffic sessions are utilized by the system's IoT gateways to extract characteristics. These attributes are sent to central edge nodes, where they are gathered and utilized for classifying and training machine learning models. Once these classifiers are returned to the gateways, they are utilized to finalize the device's identity. It is not necessary for the system to be aware of the traffic signatures of the connected devices in advance because it can recognize new devices automatically using the extracted fingerprint. In particular, the gateways classifier identifies that traffic as relatively uncommon when a device is connected to the network or when an existing device varies its normal traffic—for instance, due to a firmware upgrade. Inthiscasethegatewaysends the captured features to the edge node.When another devicebelongingtothesameunknownclass(i.e.with thesametrafficsignature)connectstoanothergateway, thisonealsosendsfeaturestotheedgenode.Nowthe edgenodeisabletoclusterizeandtoidentifythenewdevicecategory.Ifthereisnotaseconddeviceconnecting to another gateway this strategy does not work.The whole system can be controlled as a Software Defined Network (SDN) function.The categorization is done flow-wise, using a predetermined time window frame, rather than packet-wise, which might be resource-intensive. The system's objective is to assign a device to one of the four distinct classes—security cameras, smart plugs, smart lights, or motion sensors—that are taken into consideration. Additionally, the authors have employed t-Distributed Stochastic Neighbour Embedding (tSNE) to decrease the dataset's dimensionality.Thepresentedapproachusesonlybasicfeaturesofthe TCP/IPstack,nonethelessachievesgoodaccuracyand recallscores,whichtranslatesinaclassifiernotrequiringintensivecalculationsduringbothtrainingandpredictionphases.Howeveralsothenumberofconsidered devicecategoriesisveryrestricted,indirectlyimproving theaccuracyoftheclassifier.

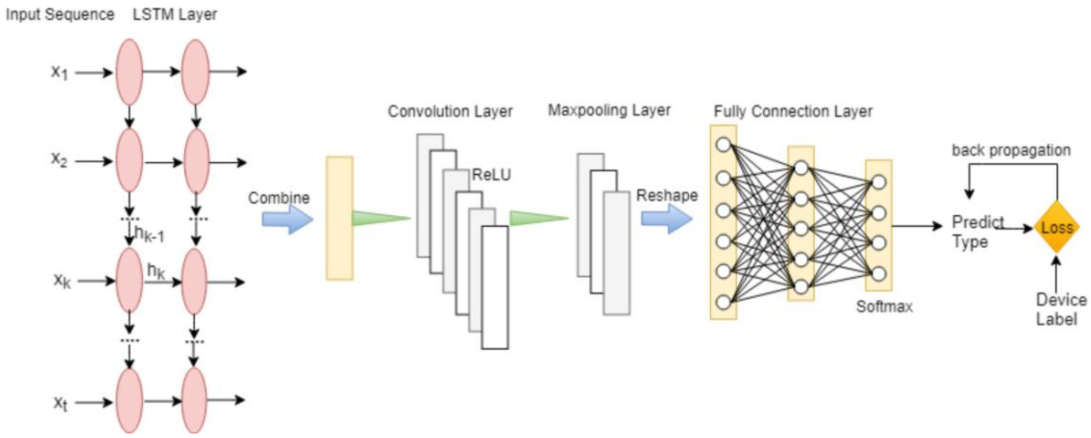
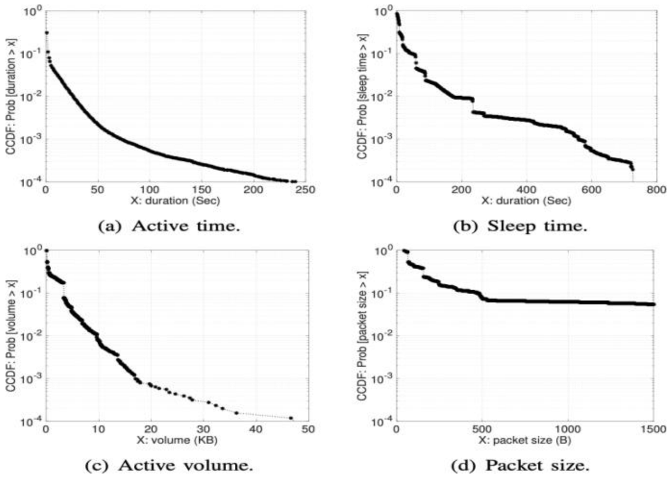


Fig.2–Neural network architecture in reference [74]. An encoding of the network traffic flows is produced by the LSTM cells. To forecast the device label, the encoding is fed to a convolutional ANN and then a fully-connected ANN. Standard backpropagation/gradient descent methods are used to train the network.



(a) Active time.

(b) Sleep time.

(c) Active volume.

(d) Packet size.

Fig.3–Distribution of several characteristics in [75]. Over the course of three weeks, the writers gathered the unprocessed IoT data. The graphs display the (a) packet size, (b) sleep time, (c) quantity of data shared via the connection, and (d) duration of the connection using the Complementary Cumulative Distribution Function (CCDF).

The classification isdoneusinginformationstreamsgeneratedbydevices and then using a LSTM-CNN model leveraging time-dependenciesofnetworktraffic.First,foreachcaptured packetfeaturessuchastimestamp,length,andvarious addresses are saved.Then features are extracted di- vidingthetrafficintotimewindowsoflength-seconds. They extracted features differentiating betweenincomingoroutgoingpacketsanduser(TCP, UDP, MQTT, HTTP) or control packets (ICMP, ARP, DNS). Various statistics of the packets are extracted. Finally,theprocesseddataisgiventoaLSTMnetwork learning an encoding of the data.The LSTM is then attachedtoaCNNnetworkwhichlearnsthefinalclassification. Thenetworkistrainedwithstandardbackpropagationalgorithms. Theachievedresultsarequitegoodin accuracy, even if there is room for improvement. Instead of using a dataset that was already available, the authors recorded campus device network activity for three weeks. Figure 4 shows the feature distribution that they reported for the dataset samples. The authors employed clusterization algorithms for every feature in order to enhance the performance of the classifiers. It is unclear, although, whose algorithm and how they have used it. The acquired results show high accuracy; nevertheless, the classifier is not suitable for classifying unknown devices since it has to be taught using network traffic from each target device.Several classification algorithms are then trained, including logistic regression, SVM, decision tree and random forest.To improvetheperformancealsoaPCAdecompositionof the features is applied.The final classification accuracyisgood,consideringtherelativelyhighnumberof consideredclasses.However,noimplementationofflow extraction or feature extraction has been carried out, whichcouldbechallenginginareal-worldsystembased on the complexity of the used features.

4.2 Open issues for Edge-Enabled Architecture

The edge network creates surfaces for attacks that bad actors can exploit. Figure 1 shows the schematic depiction of an edge-enabled Internet of Things application. Defending against malevolent The primary goal of conventional IoT-oriented intrusion detection systems (IDS), which are implemented at the gateway or device level, is to identify IoT devices. Even so, an intentional attack against the edge network has the potential to render an edge node hostile. These kinds of things can happen as a result of physical node tampering or remote exploits. The ubiquity of the edge makes it easier for a hacker to interfere with a device if nodes are positioned in public areas. An attacker in control might alter every bit of data that travels through an edge node.

One method they can do this is by pretending to be trustworthy IoT gateways or devices and creating packet streams in the edge, or they can filter out packets and decide which ones to accept and send. While it has been covered in the literature [63], controlling malevolent edge nodes is frequently not a fully automated process.

In an edge situation, already-existing intrusion detection systems (IDSs) might be implemented and utilized, but additional difficulties emerge that compromise the intrusion detection system's dependability.

They include:

- Encryption of traffic. You can install the IDS on several external edge nodes or IoT gateways. Assuming that secure protocols are used for communication between IoT devices and the cloud, traffic seen by edge nodes that are deployed with encryption is supported. This might also occur if the IDS is installed on IoT gateways and the IoT devices are TCP/IP stack capable, meaning they can speak to the cloud directly and the gateway is simply used for routing. Because of packet encryption, an intrusion detection system can only operate using non-encrypted fields—like timestamps and TCP/IP headers—instead of knowing the content that is stored inside.
- Changes in Resources. IDSs are able to use a variety of detection strategies, but the number of computational resources needed for each methodology might vary greatly. The computing resources of edge nodes, however, are also highly variable; these resources might be anything from a Raspberry Pi to a generic PC with customized hardware. It is possible that the edge node running the system will be unable to provide the resources necessary for the IDS to function, which might cause communication delays and prevent the system from operating as a whole.

  However, an edge node with greater resources also costs more, and that extra money is squandered if the IDS doesn't use it. Edge IDSs must to be flexible enough to adapt to the available resources, employing a range of algorithms with varying requirements and choosing them according to the platform on which they are currently running.

- IoT/Edge distributed IDS architecture: The network edge's IDS execution should be slightly spread due to the resource unpredictability between the edge and the Internet of Things. A single intrusion detection system (IDS) may consist of several subsystems that work together to enhance detection performance or ensure proper system operation. However, the collaboration of several subsystems increases the intrusion detection system's complexity by introducing distributed systems difficulties.
- *Aggregated traffic.*An observer on the edge, such as an IDS, may not be aware of the source end-device of a packet if the protocol stacks of IoT devices and the edge differ.
- ThisiscausedbyIoTgatewayswhichreceivepack- etsfromend-devicesusingtheirspecificIoTcom- municationtechnologyandcraftnewpacketsusingtheprotocolsofthenetworkedge,suchasTCP/IP.

V. Results and Discussions

5.1 Machine Learning Techniques Applied to IDS

This section covers some of the popular machine learning approaches used with intrusion detection systems (IDS) and how an edge IDS may benefit from them. We outline each technique's underlying principle in brief before illuminating the anticipated benefits and drawbacks of an IDS built around it. The specifications that determine whether a method is appropriate for an edge-oriented intrusion detection system (IDS) are emphasized, including processing power, storage capacity, and real-time response.

5.2Support Vector Machine(SVM)

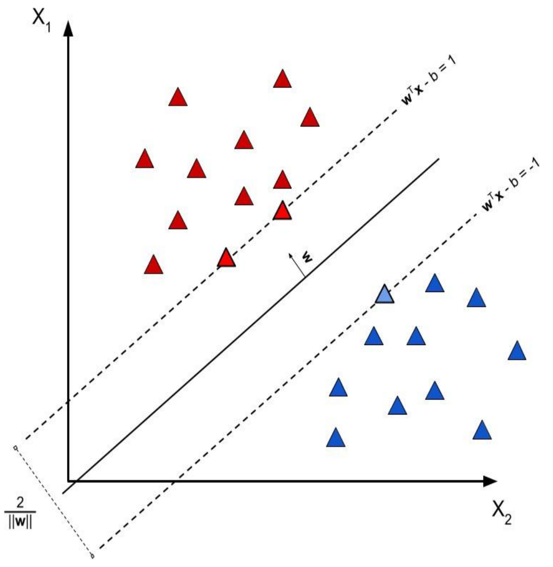
A well-liked machine learning approach for classification and regression is called Support Vector Machines (SVMs). The hyperplane w. x – b = 0 divides the Rn space with instances having the same (labels) falling in the same area. In a classification job, training data of size N is provided as input in the form $\{(x, y)\}N$, $x\in \mathbb{R}n$, $y\in$ R. Due to this division, the forecast for a fresh instance is provided by:

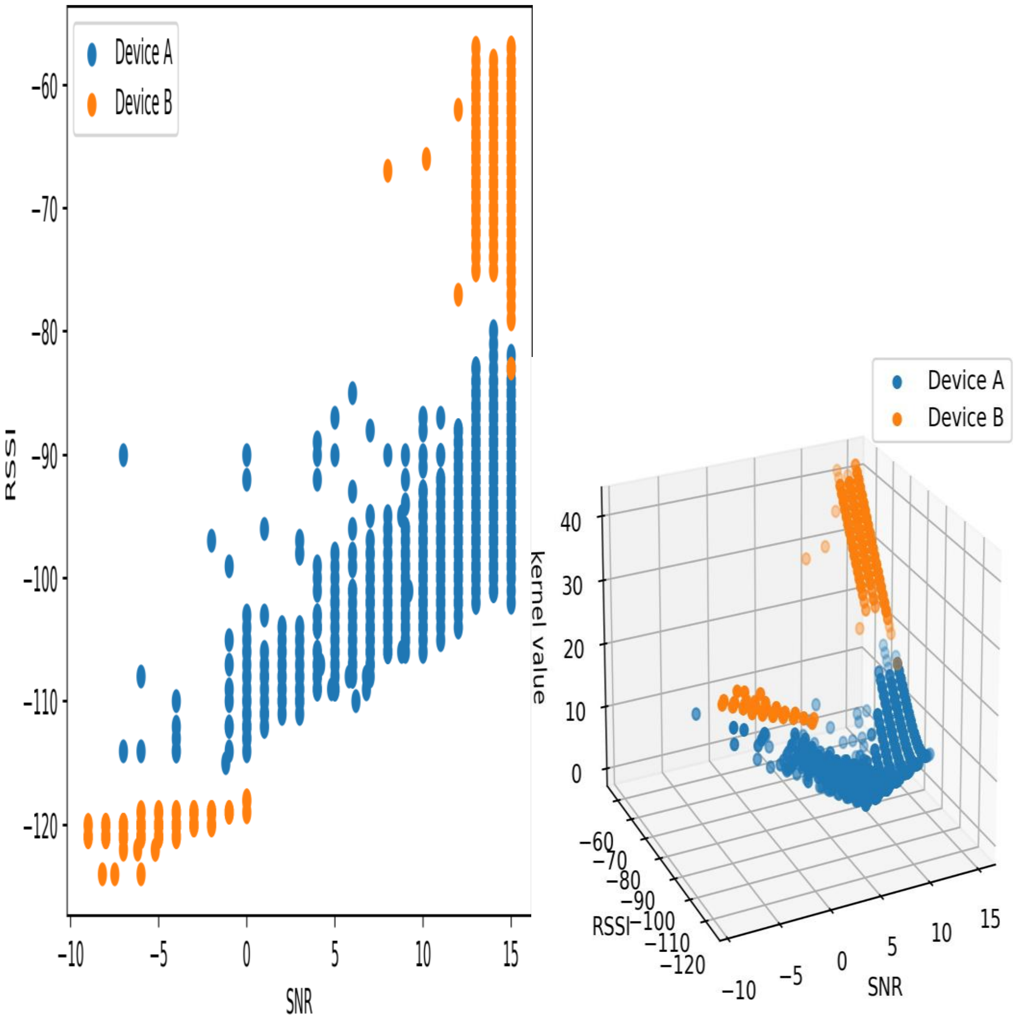
$$y'=\text{sign}(w\cdot x'-b)$$

Optimizing the distance between the space areas it divides is the primary characteristic of the hyperplane that the Support Vector Machine (SVM) constructs. Assuming a normalized dataset, let's investigate the case of $y\in \{-1, +1\}$. This allows for the definition of hyperplanes

w.x – b = 1 and w. x – b = -1. The margin, defined geometrically as 2/||W||, is the distance between these hyperplanes.

Maximizing the margin, or minimizing, is what an SVM aims to do. Adding the limits is necessary since we do not want any points to fall into the margin.


X_1
X_2
$\mathbf{w}^T\mathbf{x} - b = 1$
$\mathbf{w}^T\mathbf{x} - b = -1$
$\mathbf{w}$
$\frac{2}{\|\mathbf{w}\|}$



The problem can be formulated as:

(a) Exampleofmeasurements of SNR and RSSIApplicationofakernel function

$\{ \min||w||+\lambda\sum^{N}\xi_i$,theRBFisoftenusedasadistancemeasure.The

highaccuracyofthesystem,showthathighIDSperformancecouldbeobtainedalsowithsimpleapproaches. Pervez et al.[84] have proposed a feature-filtering algorithmbasedonSVMtoenhancetheperformanceon NSL-KDDdatasetpredictions.Thealgorithmconsiders a set of input features and trains a first SVM classifier. After that, the algorithm iterates by changing the feature space by using a specific policy to remove one of the features. After that, a new SVM is trained in the new feature space. If the new classifier's accuracy is higher than the old one, the algorithm keeps iterating; if not, it goes backwards. This technique's poor generalization makes it unsuitable for identifying new network intrusions, despite its good accuracy, recall, and low false negative ratio. Chandrasekhar et al. [85] suggested a method that uses ANNs, SVMs, and fuzzy C-means (FCM) in a confederated manner. The dataset is first partitioned into clusters using FCM, one for each kind of assault that the system can detect, and one for nominal network traffic.

The effectiveness of subsequent applied classifiers is enhanced by this initial division, which lowers the variance of the data points relative to the variance of the entire dataset. One ANN is applied each cluster after clustering in order to identify the assault patterns. Thefinalclassificationiscarriedoutvia abinarySVM.The "majority voting" strategy's evaluation in public is based on the possibility that a class with a large number of samples may be more likely to "vote" for the classification, which might affect prediction performance if the class distribution is excessively skewed. Having labelled samples is not an easy problem in an IDS environment, as we have shown, and having balanced labelled samples may be much more difficult. One way to overcome this would be to use one of the various KNN versions, such Large Margin Nearest Neighbour or Neighbourhood Components Analysis [88], or to add a distance weight to the voting process.

One benefit of KNN—and instance-based approaches in general—for an IDS is that it does not require a training phase, making it appropriate for systems whose circumstances change dynamically over time. Assume that after we apply anomaly detection to a system, its operational parameters alter for a justifiable reason (like seasonality, for example). Rebuilding a model for anomaly detection can be necessary. When using techniques like SVMs, ANNs, or anything similar, we have to retrain the entire model, which causes the IDS to experience downtime and leaves the system susceptible. On the other hand, all it takes to create a new KNN instance is feeding the algorithm with labeled samples, and the system will instantly function. Predictions made without a model may be slower than those made using other ML approaches, which might affect the IDS's real-time feature.

5.3DecisionTree

A decision tree is a machine learning approach that quickly generates predictions by using input data to create a classification tree. A data feature is represented by each node in the tree, a feature value by each branch, and a potential classification result by each terminal leaf. Typically, the Decision Tree model is constructed top-down, choosing features one after the other in accordance with a predetermined strategy like Information Gain (IG). Information Gain quantifies the amount of knowledge that may be obtained about one random variable by assessing another random variable. It is defined as the difference in entropy between itself and the attribute value with respect to prior information in the context of decision trees.

$_i$givenasknown:

T $\mathrm{IG}(T,X_i)=\mathrm{H}(T)-\mathrm{H}(T|X_i)$

the conditional entropy is represented by H i, while the Shannon's entropy is represented by H. A "rank-ing" of features is provided by Information Gain, which causes the tree to prioritize characteristics with strong judicial potential. The most popular decision tree building algorithms, ID3, C4.5, and CART [95], all employ information gain to generate new tree nodes. Regardless of the technique used to create it, a decision tree's prediction process involves picking branches based on if-then-else operations on the input characteristics, working your way down the tree from the root node. This reduces the need for specialized hardware needed for other machine learning approaches, making the prediction stage of a Decision Tree computationally efficient.

Overfitting of training data is a major problem when utilizing a Decision Tree. Although overfitting is a problem with all machine learning approaches, Decision Trees are more prone to it since they simply do independent feature comparisons without any relationship. Methods like post-pruning might be employed to prevent overfitting. By deleting part of the internal nodes from a tree, a pruning technique aims to increase model generalization by reducing the number of training set-specific tests. Selecting a threshold for the amount of pruning that should be done to a decision tree is not an easy process. Using the test set to create the model biases the model itself, therefore alternating between pruning and evaluation may not be the best course of action.

The difficulty of modeling intricate interactions between attributes is another problem. Every feature is handled separately during the prediction step, which may result in certain specific inputs being incorrectly classified. Data processing using PCA might be the first step towards solving this issue. The inability to find patterns that connect various data pieces is another significant problem with using decision trees in an IDS. Decision trees do, in fact, handle each input separately from the others and do not preserve the internal state of the earlier predictions. a resolution to the decision tree's Small Disjoint issue.

When nodes closest to the leafs distinguish between a limited number of instances, the training data becomes overfit, resulting in the Small Disjoint issue. To demonstrate the coverage of those rules that address the Small Disjoint issue, a genetic algorithm is employed. The system is composed of two modules: one that generates the rules and the other that optimizes them.

VI. Conclusions

Intrusion Detection Systems for the Internet of Things are presented in this paper from both an architecture and methodology standpoint, enabling them to detect abnormalities and cyberattacks. From an architecture standpoint, more attack surfaces are available for malevolent actors to take advantage of, even though conventional IoT IDS are implemented at the device or gateway level. We spoke about the new problems and the solutions that have been proposed to deal with them. There's talk about new IDSs made especially for the edge. Next, we concentrate on how Machine Learning techniques are being adopted and used by IDSs nowadays. We first discussed the theory underlying each strategy before providing examples of its anticipated benefits and drawbacks. The requirements for each technique's processing power, storage capacity, and real-time responsiveness are emphasized. These factors determine whether a method is appropriate for an edge-oriented IDS or not.

References

[1] M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, X. Du, I. Ali, and M. Guizani. “A Survey of Ma- chine and Deep Learning Methods for Internet of Things (IoT) Security”. In: *IEEE Communi- cationsSurveysTutorials*22.3 (2020).

[2] RobertoMagán-Carrión,JoséCamacho,Gabriel Maciá-Fernández,andAngelRuíz-Zafra.“Mul- tivariateStatisticalNetworkMonitoringSensor: An effective tool for real-time monitor- ing and anomaly detection in complex net- works and systems”. In: *International Journal of Distributed Sensor Networks* 16.5 (May 2020).

[3] AneBlázquez-García,AngelConde,UsueMori, and Jose A. Lozano. “A review on out- lier/anomaly detection in time series data”. In: *ArXive-prints*(2020).

[4] Mateusz Krzysztoń and Michał Marks. “Sim- ulation of watchdog placement for cooperative anomaly detection in Bluetooth Mesh Intru- sionDetectionSystem”.In:*SimulationModelling Practice and Theory* 101 (May 2020).

[5] HamidTahaei,FirdausAfifi,AdelehAsemi,FaizZaki,andNorBadrulAnuar.“Theriseoftraf- ficclassificationinIoTnetworks:Asurvey”.In: *Journal of Network and Computer Applications* 154 (Mar. 2020).

[6] A. Ajit, K. Acharya, and A. Samanta. “A Re- viewof ConvolutionalNeural Networks”.In: *2020 InternationalConferenceonEmergingTrends in Information Technology and Engineering (ic- ETITE)*. 2020, pp. 1–5.

[7] IbomoiyeDomorMienye, Yanxia Sun, and ZenghuiWang.“Improvedsparseautoencoder basedartificialneuralnetworkapproachforpre- diction of heart disease”. In: *Informatics in Medicine Unlocked* 18 (2020), p. 100307. doi: 10.1016/j.imu.2020.100307.

[8] M.Ribeiro,M.Gutoski,A.E.Lazzaretti,and H.S.Lopes.“One-ClassClassificationinImages andVideosUsingaConvolutionalAutoencoder With Compact Embedding”. In: *IEEE Access* 8 (2020), pp. 86520–86535.

[9] S.Krishnaveni,PalaniVigneshwar,S.Kishore, Jothi, and S. Sivamohan. “Anomaly-Based IntrusionDetectionSystemUsingSupportVec- tor Machine”. In: *Advances in Intelligent Sys- temsandComputing*. Springer Singapore, 2020, pp. 723–731. doi: 10.1007/978-981-15-0199-9_62.

[10] AneBlázquez-García, Angel Conde, Usue Mori, and Jose A. Lozano. “A review on out lier/anomaly detection in time series data”. In: ArXiv e-prints (2020). eprint: 2002.04236.

[11] Andrea Lacava, Emanuele Giacomini, Francesco D’Alterio, and Francesca Cuomo. “Intrusion De tection System for Bluetooth Mesh Networks: Data Gathering and Experimental Evaluations”. In: 2021 IEEE International Conference on Per vasive Computing and Communications Work shops and other Affiliated Events (PerCom Workshops). 2021, pp. 661–666.

[12] M. B. C, K. K. J, L. N, P. K. H, and S. J. “In truder Detection System- A LoRa Based Ap proach”. In: 2020 5th International Conference on Communication and Electronics Systems (IC CES). 2020, pp. 255–258.

[13] M. Eskandari, Z. H. Janjua, M. Vecchio, and F. Antonelli. “Passban IDS: An Intelligent Anomaly-Based Intrusion Detection System for IoT Edge Devices”. In: IEEE Internet of Things Journal 7.8 (2020), pp. 6882–6897.

[14] Ibbad Hafeez, Markku Antikainen, Aaron Yi Ding, and SasuTarkoma. “IoT-KEEPER: De tecting Malicious IoT Network Activity Using Online Traffic Analysis at the Edge”. In: IEEE Transactions on Network and Service Manage ment 17.1 (Mar. 2020), pp. 45–59. doi: 10.1109/ tnsm.2020.2966951.

[15] Yulong Lu and Jianfeng Lu. “A Universal Ap proximation Theorem of Deep Neural Networks for Expressing Distributions”. In: ArXiv e-prints (2020). arXiv: 2004.08867 [cs.LG].

[16] IbomoiyeDomorMienye, Yanxia Sun, and Zenghui Wang. “Improved sparse autoencoder based artificial neural network approach for pre diction of heart disease”. In: Informatics in Medicine Unlocked 18 (2020), p. 100307. doi: 10.1016/j.imu.2020.100307.

[17] M. Ribeiro, M. Gutoski, A. E. Lazzaretti, and H. S. Lopes. “One-Class Classification in Images and Videos Using a Convolutional Autoencoder With Compact Embedding”. In: IEEE Access 8 (2020), pp. 86520–86535.

[18] RaniyahWazirali. “An Improved Intrusion Detection System Based on KNN Hyperparameter Tuning and Cross-Validation”. In: Arabian Journal for Science and Engineering (Aug. 2020). doi: 10.1007/s13369-020-04907-7.

[19] Yee Jian Chew, Shih Yin Ooi, Kok-Seng Wong, and Ying Han Pang. “Decision Tree with Sen sitive Pruning in Network-based Intrusion De tection System”. In: Lecture Notes in Electrical Engineering. Springer Singapore, 2020, pp. 1–10. doi: 10.1007/978-981-15-0058-9_1.

[20] S. Krishnaveni, Palani Vigneshwar, S. Kishore, B. Jothi, and S. Sivamohan. “Anomaly-Based Intrusion Detection System Using Support Vec tor Machine”. In: Advances in Intelligent Sys tems and Computing. Springer Singapore, 2020, pp. 723–731. doi: 10.1007/978-981-15-0199 9_62

[21] Chaofei Tang, NurbolLuktarhan, and Yuxin Zhao. “An Efficient Intrusion Detection Method Based on LightGBM and Autoencoder”. In: Symmetry 12.9 (Sept. 2020), p. 1458. doi: 10.3390/ sym12091458.

[22] Stolojescu-Crisan, C.; Crisan, C.; Butunoi, B.-P. An IoT-Based Smart Home Automation System. Sensors 2021, 21, 3784.

[23] D. Kundu, M. E. Khallil, T. K. Das, A. A. Mamun, and A. Musha, “Smart home automation system using on IoT,” International Journal of Scientific Engineering and Research, vol. 11, no. 6, pp. 697–701, 2020

[24] M. V. Gladence, M. V. Anu, R. Rathna, and E. Brumancia, “Recommender system for home automation using IoT and artificial intelligence,” Journal of Ambient Intelligence and Humanized Computing, pp. 1–9, 2020.

[25] H. Basly, W. Ouarda, F. E. Sayadi, B. Ouni, and A. M. Alimi, "CNN-SVM learning approach based human activity rec ognition," in Proceedings of the Image and Signal Processing ICISO 2020, Marrakesh, Morocco, June 2020

[26]U.Pujari, P.Patil, N. Bahadure, and M.Asnodkar,"Internet of things based integrated smart home automation system," in Proceedings of the International Conference on Communication and Information Processing (ICCIP), Tokyo, India, No vember 2020.

[27] H. Basly, W. Ouarda, F. E. Sayadi, B. Ouni, and A. M. Alimi, "CNN-SVM learning approach based human activity rec ognition," in Proceedings of the Image and Signal Processing ICISO 2020, Marrakesh, Morocco, June 2020.

[28]U.Pujari, P.Patil, N. Bahadure, and M.Asnodkar,"Internet of things based integrated smart home automation system," in Proceedings of the International Conference on Communication and Information Processing (ICCIP), Tokyo, India, No vember 2020.

[29] M. N. Aman, U. Javaid, and B. Sikdar, "IoT-proctor: a secure and lightweight device patching framework for mitigating malware spread in IoT networks," IEEE Systems Journal, pp. 1–12, 2021.

[30] O. Taiwo, L. L. Gabralla, and A. E. Ezugwu, "Smart home automation: taxonomy,Composition, challenges and future direction," in Computational Science and its Applications– ICCSA 2020, O. Gervasi et al., Ed., Springer Nature Switzerland, Switzerland, 2020.

[31] A. A. Zaidan and B. B. Zaidan, "A review on intelligent process for smart home applications based on IoT: coherent taxonomy, motivation, open challenges, and recommendations," Artificial Intelligence Review, vol. 53, no.1, pp.141–165, 2020.

[32] U. Javaid and B. Sikdar, "Lightweight and secure energy trading framework for electric vehicles," in Proceedings of the 021 International Conference on Sustainable Energy and Fu ture Electric Transportation, SEFET, Hyderabad, India, Jan 2021.

[33] P. Sharma and P. kantha, "'Blynk' cloud server based mon itoring and control using 'NodeMCU'," International Re search Journal of Engineering and Technology (IRJET), vol. 7, no. 10, pp. 1362–1366, 2020.

[34] S. Garg, A. Yadav, S. Jamloki, A. Sadana, and K. arani, "IoT based home automation," Journal of Information and Opti mization Sciences, vol. 41, no. 1, pp. 261–271, 2020.

[35] P. Shivaprasad, "Understanding confusion matrix, pre cision-recall and F1 score," 2020, [Online]. Available: https:// towardsdatascience.com/understanding-confusion-matrix-pr ecision-recall-and%20f1-score-8061c9270011.

[36] O. Taiwo and A. E. Ezugwu, "Smart healthcare support for remote patient monitoring during covid-19 quarantine," Informatics in Medicine Unlocked, vol. 20, no. 100428, pp. 100428–100512, 2020.

[37] F. Hall, L. Maglaras, T. Aivaliotis, L. Xagoraris, and L. Kantzavelou, "Smart homes: security challenges and pri-vacy concerns," pp. 1–6, 2020.

[38] P. Sharma and P. kantha, "'Blynk' cloud server based monitoring and control using 'NodeMCU'," International Re search Journal of Engineering and Technology (IRJET), vol. 7, no. 10, pp. 1362–1366, 2020.

[39] P. Shivaprasad, "Understanding confusion matrix, pre cision-recall and F1 score," 2020, [Online]. Available: https:// towardsdatascience.com/understanding-confusion-matrix-pr ecision-recall-and%20f1-score-8061c9270011.

[40] W. Ahmad, A. Rasool, A. R. Javed, T. Baker, and Z. Jalil, "Cyber security in IoT-based cloud computing: a comprehensive survey," Electronics, vol. 11, no. 1, article 16, 2022.

[41] M. Shafiq, Z. Tian, A. K. Bashir, X. Du, and M. Guizani, "Cor rAUC: a malicious Bot-IoT traffic detection method in IoT network using machine learning techniques," IEEE Internet of Things Journal, vol. 8, no. 5, pp. 3242–3254, 2021.

[42]Shafiq,Z.Tian,Y.Sun,X.Du,andM.Guizani,"Selectionof effective machine learning algorithm and Bot-IoT attacks traf f ic identification for Internet of Things in smart city," Future Generation Computer Systems, vol. 107, pp. 433–442, 2020.

[43] M. Shafiq, Z. Tian, A. K. Bashir, X. Du, and M. Guizani, "IoT malicious traffic identification using wrapper-based feature selection mechanisms," Computers & Security, vol. 94, Article ID 101863, 2020.

[44] S. Weisman, What Are Denial of Service (DoS) Attacks? DoS Attacks Explained, Norton Lifelock, United States, 2020.

[45] A. Mustafa, B. Poudel, A. Bidram, and H. Modares, "Detection and mitigation of data manipulation attacks in AC micro grids," IEEE Transactions on Smart Grid, vol. 11, no. 3, pp. 2588–2603, 2020.

[46] A. R. Javed, M. Usman, S. U. Rehman, M. U. Khan, and M. S. Haghighi, "Anomaly detection in automated vehicles using multistage attention-based convolutional neural network," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 7, pp. 4291–4300, 2021.

[47] S. Shukla, S. Thakur, and J. G. Breslin, "Anomaly detection in smart grid network using FC-based blockchain model and lin ear SVM," in International Conference on Machine Learning, Optimization, and Data Science, pp. 157–171, Springer, Cham, 2021.

[48] A. Chehri, I. Fofana, and X. Yang, "Security risk modeling in smart grid critical infrastructures in the era of big data and artificial intelligence," Sustainability, vol. 13, no. 6, article 3196, 2021

[49] V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriyeh, A. Dehghantanha, and G. Srivastava, "Federated-learning based anomaly detection for IoT security attacks," IEEE Inter net of Things Journal, vol. 9, no. 4, pp. 2545–2554, 2022.

[50] A. Al Hayajneh, M. Z. A. Bhuiyan, and I. McAndrew, "Improving Internet of Things (IoT) security with software defined networking (SDN)," Computers, vol. 9, no. 1, 2020.

[51] J. Galeano-Brajones, J. Carmona-Murillo, J. F. Valenzuela Valdés, and F. Luna-Valero, "Detection and mitigation of DoS and DDoS attacks in IoT-based stateful SDN: an experi mental approach," Sensors, vol. 20, no. 3, 2020.

[52] M. D. Sánchez-Hernández, M. C. Herrera-Enríquez, and F. Expósito, "Controlling behaviors in couple relationships in the digital age: acceptability of gender violence, sexism, and myths about romantic love," Psychosocial Intervention, vol. 29, no. 2, 2020.

[53] G. Beaumier and K. Kalomeni, "Ruling through technology: politicizing blockchain services," Review of International Polit ical Economy, pp. 1–24, 2021.

[54] G. Thawre, N. Bahekar, and B. R. Chandavarkar, "Use cases of authentication protocols in the context of digital payment sys tem," in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–6, Kharagpur, India, July 2020.

[55] M. Shafiq, Z. Tian, A. K. Bashir, A. Jolfaei, and X. Yu, "Data mining and machine learning methods for sustainable smart cities traffic classification: a survey," Sustainable Cities and Society, vol. 60, p. 102177, 2020.

[56] H. Abbasi, N. Ezzati-Jivan, M. Bellaiche, C. Talhi, and M. R. Dagenais, "Machine learning-based EDoS attack detection technique using execution trace analysis," Journal of Hardware and Systems Security, vol. 3, no. 2, pp. 164–176, 2019.

[57] Syafa'Ah, L.; Minarno, A.E.; Sumadi, F.D.S.; Rahayu, D.A.P. ESP 8266 for Control and Monitoring in Smart Home Application. In Proceedings of the 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), Jember, Indonesia, 16–17 October 2019; pp. 123–128.

[58] Makhanya, S.P.; Dogo, E.M.; Nwulu, N.I.; Damisa, U. A Smart Switch Control System Using ESP8266 Wi-Fi Module Integrated with an Android Application. In Proceedings of the 2019 IEEE 7th International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, 12–14 August 2019; pp. 125–128.

[59] Stolojescu-Crisan, C.; Crisan, C.; Butunoi, B.P. IoT Based Intelligent Building Applications in the Context of COVID-19 Pan-demic. In Proceedings of the International Symposium on Electronics and Telecommunications (ISETC), Timisoara, Romania, 5–6 November 2020.

[60] Shahajan, M.; Islam, G.M.J.; Das, S.K.; Islam, S.; Islam, M.; Islam, K.K. Internet of Things (IoT) based automatic electrical energy meter billing system. J. Electron. Commun. Eng. 2019, 14, 39–50.