STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 9, Issue 1, 2022
DOI Journal: 10.37591/JoAIRA

http://computers.stmjournals.com/index.php?journal=JoAIRA&page=index

Review JoAIRA

Analysis of Various Machine Learning Techniques for Smart City

Rohan Chauhan*

Abstract

A botnet represents a network of multiple bots developed to launch spiteful incidents on a target network that is managed through a command-and-control protocol by a solo object known as botmaster. Bots are compromised computers that are controlled from far by the botmaster with no indications of being hacked and used to carry out spiteful actions. The size of botnets differs from small botnets containing a few hundred bots to big botnets containing 50,000 hosts. The smart city using machine learning techniques are reviewed in this study. The various contents which are published are on the reputed journals.

Keywords: Smart city, machine learning, bots, Internet of Things, IoT sensors, machine learning algorithms

INTRODUCTION

The internet has transmuted from an appealing new innovation to a vital piece of framework. Since the common people have become dependent on the data, communication and worthwhile resources available online, there is an increasing aspiration for everything linked to the world wide web.

IoT based Smart City

A significant boom in the marketplace has incorporated sensors in appliances, equipment, and vehicles for providing additional data and amenities from regular life to the virtual reality. Providing sensors in a variety of contexts is believed to make our lives more efficient, safer, and convenient [1]. This idea is outlined as Internet of Things (IoT). The extension of Internet of Things (Extension) and radio technology have contributed a lot to the eminence of the concept of smart city, which is a natural fruit of development and growing populace besides the optimization of resource consumption. Figure 1 presents a wide-ranging view of different components required in a smart city.

Smart cities are dynamically rested upon IoT sensors and actuators to advance amenities and reduce operating costs by automation. Cities can be made smart by applying any combination of different smart gadgets. Not all components are required for a city to be considered as smart. The number of smart gadgets relies upon cost and accessible innovation. Nevertheless, IoT measures in smart cities and their compatibility can create an opportunity to cyber security issues capable of crippling crucial facilities [2].

*Author for Correspondence
Rohan Chauhan
E-mail: rohan.c199880@gmail.com

Student, Department of Computer Science & Engineering, Gautam Buddha University, Greater Noida, Uttar Pradesh, India

Received Date: March 31, 2022
Accepted Date: April 09, 2022
Published Date: April 15, 2022

Citation: Rohan Chauhan. Analysis of Various Machine Learning Techniques for Smart City. Journal of Artificial Intelligence Research & Advances. 2022; 9(1): 1–16p.

Cybersecurity in Smart Cities

With the advancement of smart city technology becoming more accessible and appropriate to the real lives of residents, concerns also emerge about security and privacy. At the place of smart terminals like customer-operated machines for

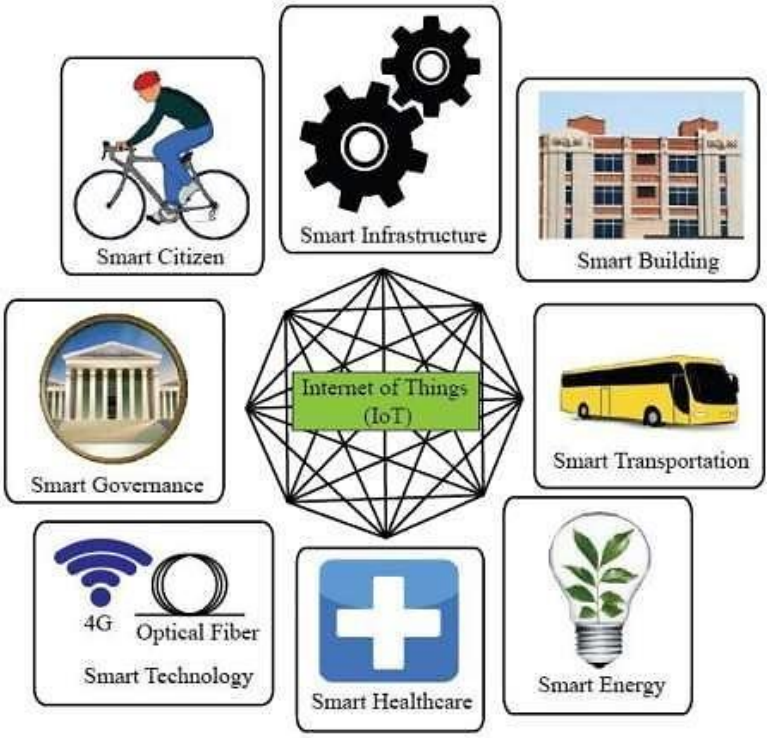

Figure 1. A Broad Overview of Smart City Components.

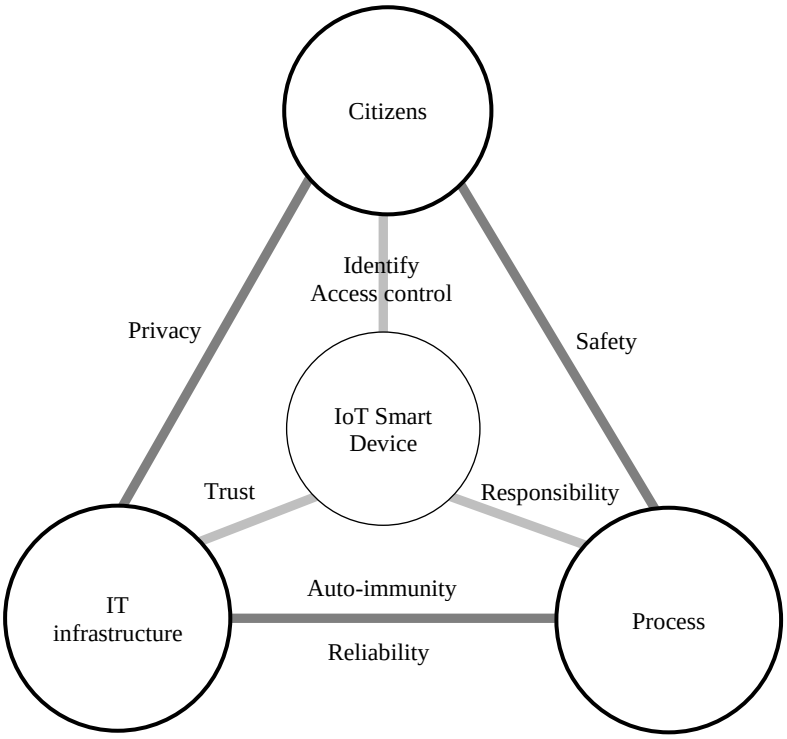

Figure 2. Pillars of Cybersecurity for Smart City.

selling or payments, there is the availability of open data. These data accesses come up with the challenges of integrity and confidentiality, since most of the data relates to identity and financial information. This concept helps in the identification of the new attack being surfaced; however, many researches are still going on to develop security measures [3]. In order to infer cyber security strategies, a model adapted for a smart city is shown in Figure 2.

As people are users of the infrastructure according to its role (business, government authority, academe, citizen), they have various access steps to the smart city cyber system. The process node deals with the events of residents incorporating appropriate security principals to maintain a smart IT set-up in place. Moreover, these frameworks for smart city amenities depend on private cloud services which provide efficacy and optimal procedure. At the heart, all IoT smart devices operate independently in the best interests of their proprietors to prolong their lives be it ID tags, medical wearables, sensors, actuators, or other gadgets. For every edge, privacy concerns seek to shield the responsive data of people used on IoT and IT cloud systems. The security edge ensures multiple automatic functions from the city's cyber system to run safely for people [4].

The reliability edge links IT framework to the potential for failures and security risks. From IoT smart device, it is required to manage ID and access control besides trust that the device is secure and approved to link to IT set-up with accountable developments to improve and maintain security. In this context, the complexity of the city to interconnect the various gadgets is the result of the system's feeble reliability among the procedures that impacts on the privacy of citizens and the tools and procedures to recognize the users in the cyber environment. Many cities dynamically establish connectivity between citizens and IT set-up to maintain security that does not balance reliability, security over processes [5]. This explicates why citizens are challenging governments to secure their privacy rights.

IoT Attacks

In heterogeneous IoT infrastructures, IoT attacks take different forms. Physical attacks (such as side channel attacks and sleep denial attacks), network attacks (such as routing attacks, sybil attacks, and man in the middle attacks), and software attacks are all examples of these types of attacks (like virus, trojan and malware induction). IoT attacks grew linearly with the number of susceptible gadgets, as 70% of IoT gadgets are such gadgets. As a result, many events have happened affecting society and the economy. Hence, after a large number of events including IoT products, the security of IoT environment has turned into a main challenge [6]. Generally speaking, the majority of attacks are botnet-based attacks in the IoT set-up. Another time, on IoT devices, several security weaknesses still occur since the great majority of cases do not have enough memory and computing assets for strong security measures.

The impact of IoT botnets can be enormous, and almost all IoT malware try to produce botnets. Likewise, in October 2016, the Mirai botnet targeted a domain name server (DNS) provider company (Dyn) by exploiting a variety of susceptible IoT gadgets such as CCTVs, routers and DVRs. These mechanisms were used to send requests from ten million IP addresses. The attack created more than 1 Tbps of traffic and brought down key social media sites, for example Twitter, The Guardian, Netflix, and CNN. As a result, the attack triggered disturbance to those amenities and disturbed the internet as a whole [7].

Botnets in IoT Networks

Many breeds of malware have been reported to attack IoT gadgets and create IoT botnets. Mirai, Bashlight, Wirex, Brickerbot Reaper, and Hajime were among the botnets discovered on IoT networks. BASHLILE and Mirai are two specific IoT malware, and they have contaminated multiple IoT devices that makes them available with weaknesses and provided verification IDs. BASHLILE, also known as Gafgyt, was released in 2015. Its inheritor, namely the Mirai, was released in 2016. Contaminated gadgets next scan to discover potentially susceptible gadgets and notify them to a main database [8]. According to reports, Mirai in 2016 infected nearly 2.5 million IoT devices.

IoT Botnet Life Cycle Phases

Several studies have approved that IoT botnets perform their function in at least three main steps as shown in Figure 3:

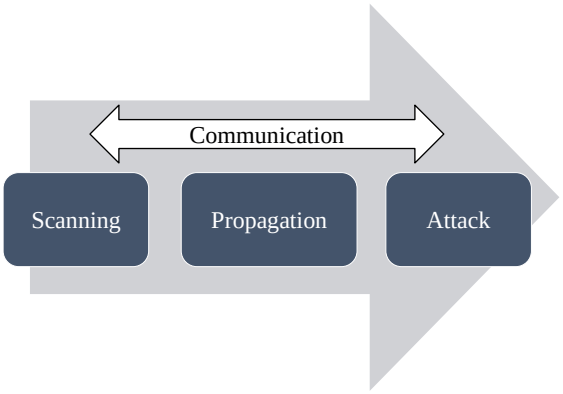

Figure 3. Phases of an IoT botnet's life cycle.

All the steps of IoT botnet's life cycle are described below:

Step 1. ***Scanning:*** To detect a weak device, a bot (or malicious code) enforces scanning and scouting. Botmaster performs scanning to identify weak IoT devices. After it discovers one, it starts infecting it either by brute force or by taking advantage of the weakness. After infecting the weak device, the infected device turns into a bot and begins to communicate with the botmaster. The Mirai virus family, for example, identifies IoT devices that emit fingerprint packets to check for pseudo-random IPv4 addresses reachable through Telnet on port 23 or port 2323. By using brute force, abusing susceptible IDs or ill-using known weaknesses of IoT devices, the bot infects new targets [9].

Step 2. ***Propagation****:* Depending on the design of the susceptible device, the appropriate version of the bot is installed and run. Often, to prevent pointing any other potentially malware-infested devices and gain full control, the bot slays the process associated with the service concerned to remove any other former malware and lock the port. Malicious code deploys new bots and proliferates expanding IoT botnets without delay. In this phase, the bots are still waiting for orders from the botmaster.

Step 3. ***Attack****:* This step is concerned with the running of spiteful events such as DDoS, crypto mining, and spam. To trigger attack, the attacker sends commands using a command-and-control server to all dispersed bots. As a result, the bots launch an attack post obtaining the same command [10].

IoT Botnet Architecture

The low cost IoT devices are the reason behind their popularity and growth. Nevertheless, these tools lack efficient security and standards, which makes them susceptible to assault and control by intruders. One of the most critical attack environments is IoT devices being infected and inducted into an IoT botnet. After compromising and damaging an IoT device, the attacker can take control of it and use it to trigger different attacks [11]. A botnet represents a network of multiple bots developed to launch spiteful incidents on a target network that is managed through a command-and-control protocol by a solo object known as botmaster. Bots are compromised computers that are controlled from far by the botmaster with no indications of being hacked and used to carry out spiteful actions. The size of botnets differs from small botnets containing a few hundred bots to big botnets containing 50,000 hosts. Hackers distribute botnet malware and work covertly without any visible indication of their incidence and may be operative and functional for years. The botnet command and control designs are illustrated in Figure 4.

The major component in botnet is the interaction of botmaster with its related bots. Communication with bots is required to provide commands to the bots to perform spiteful incidents. Botmaster constantly remains invisible using low bandwidth and deliver hidden services in the botnet 3 network. Botmaster always establishes communication over command-and-control server with bots. The major objective of the bots is to stay unnoticed till they are needed to complete the tasks assigned to them.

The unseen nature of bots makes the identification more complex as they do not disturb usual operation on the host and stay noiseless until they get a command from the botmaster to perform the activities assigned to them [12]. The propagation and infection phases, secondary injection, connection, malicious command and control, update and maintenance are all parts of a botnet's service period.

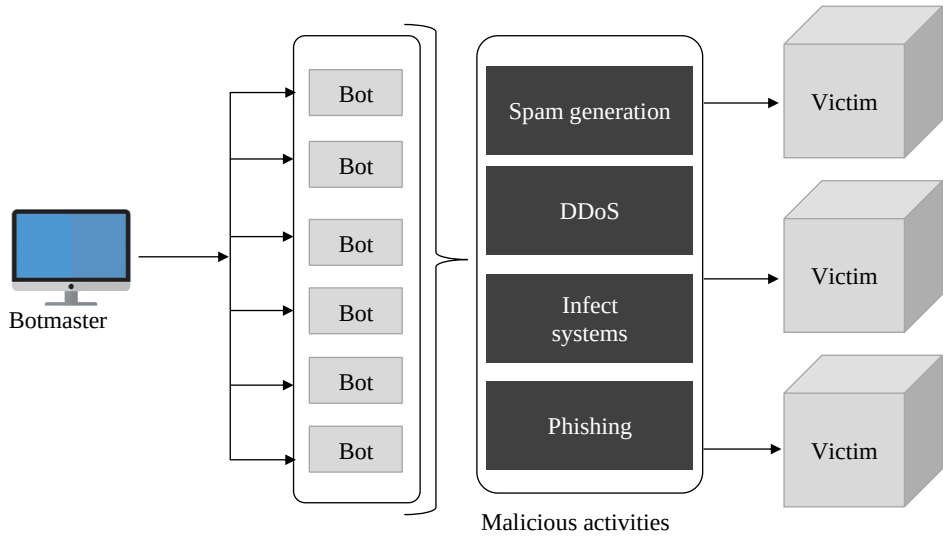

Figure 4. Botnet Architecture.

Basic Components of IoT Botnets

It is very important to understand the way internet botnets work in order to discover fresh and efficient processes to spot these bots and manage them in order to limit their damage. Inferring the functions of these bots more broadly will help to counter them and maintain cyberspace to be cleaned, and therefore assist guarantee the security of the world wide web. Post the publication of source code for Mirai, several cyber offenders cloned and modified the code and released duplicates of IoT malware intended at creating an Internet of Things (IoT) botnet and competing to control the maximum number of IoT gadgets [13]. Consequently, almost all IoT malware have identical operating steps. Figure 5 shows how the IoT botnet works.

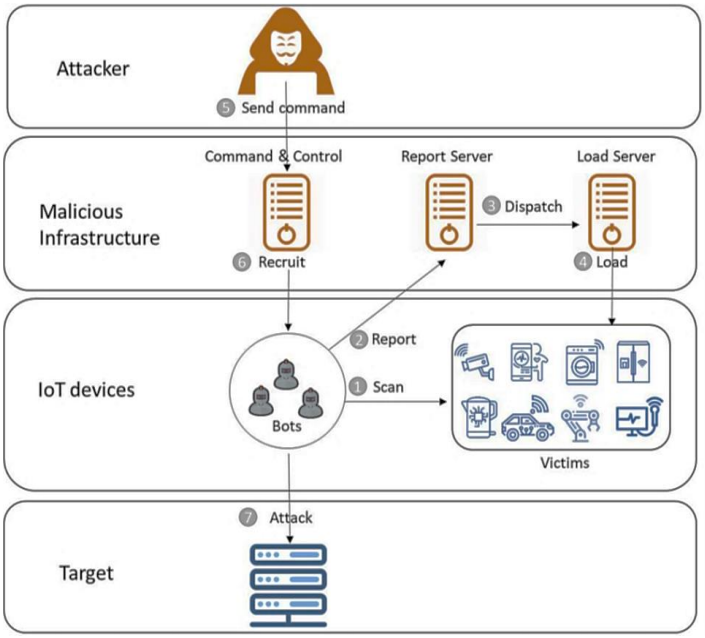

Figure 5. A flow diagram of an IoT botnet.

The figure makes it clear that botnets operate in many steps. They are summarized and generalized in the following seven steps:

1. The bots search the IP address space for Telnet or SSH-enabled devices and attempt to log in with a hard-coded IoT credential dictionary.
2. When a bot is successful, it sends the victim's IP address and relevant credentials to a report server.
3. This data is sent to the loader server by the report server.
4. At the same time, the loader server causes the device to be infected in a fashion that is reliant on the victim's architecture.
5. To begin the attack, the attacker sends a command to the command-and-control server stating the target and the needed details.
6. In advance of an attack, the command-and-control server warns the bots.
7. The bots launch an attack on the designated target.

It is worth mentioning that malware for the IoT is evolving and being advanced for various reasons. These reasons involve various purposes for IoT botnets and taking advantage of new weaknesses. Hence, preservation is done for such malware. Reconfiguring botnets or adding more devices to them expands their scale and makes them more resilient [14]. These modifications may make IoT botnets more urbane and upsurge the challenges in IoT botnet discovery.

Types of IoT Botnets

The architectures shared through traditional botnets have similarity with the IoT botnets. Centralized botnets, decentralized (peer-to-peer) botnets, and hybrid botnets are the three types of botnets:

i. *Centralized botnets:* The botmaster is in charge of administering and tracking all bots from a single central server, which helps to reduce latency. This implies that all bots receive instructions and send a report to a C&C (central server). The botmaster of this architecture can utilize one or more central servers. The protocols employed in server are HTTP and IRC. The total failure may be occurred due to involvement of only single point in botmaster server. Mirai family is the well-known centralized IoT botnets [15].
ii. *Decentralized botnets*: These are recognized as P2P (peer-to-peer) botnets. Each bot functions as both a client and a server and is linked to at least one other bot. In case of interconnection of all the bots, the commands will be sent to every bot. The coordination among bots is a complex task in this architecture. However, due to diverse communication among peers, the attack cannot be detected easily. This kind of IoT botnet makes the deployment of a P2P protocol in communication. Hajime is one among these IoT botnets.
iii. *Hybrid botnets*: A hybrid botnet consists of two kinds of bots, some of them work as servers and clients and others are executed as clients. The earlier two kinds of architectures are integrated in this. It has greater message latency.

Machine Learning Based BoT-IoT Attack Traffic Identification Process

The NIDS (Network Intrusion Detection Systems) are constructed using ML (Machine Learning) schemes for securing the connected IoT devices to deal with complex botnet attacks. This type of equipment is installed at strategic locations throughout an IoT network. The ML-based methods assist in detecting the known attacks as well as their variances. Machine learning based system of detecting the anomaly are developed from new anomaly detection model for discovering the unusual traffic on the basis of learning algorithm which exhibited an attempt of intrusions in the network. Unlike the traditional laptops and smartphones, IoT traffic behaves in diverse manner due to the communication of devices with the endpoints within a small range instead of large web servers. The ML technique is implemented to monitor this type of behavior of IoT traffic. The Figure 6 represents the entire flow of the processes.

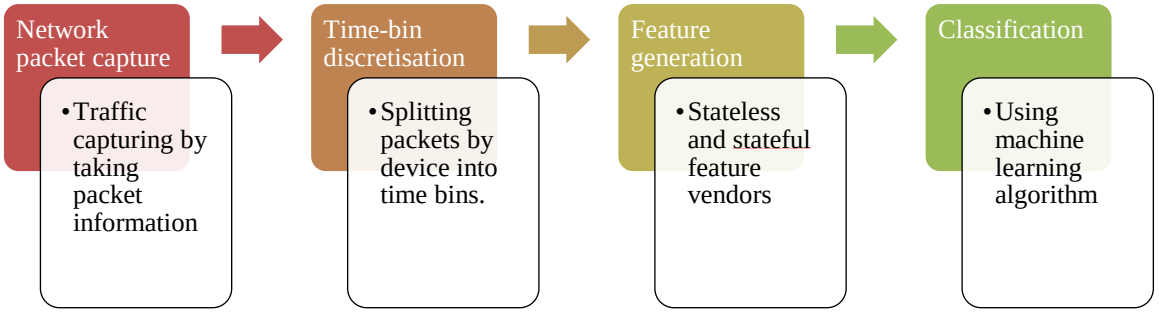

Figure 6. Process flow for the machine learning based detection framework.

The process of detecting anomaly undergoes diverse stages or phases in which the traffic is captured, the packets are grouped using device and time, attributes are extracted and finally the Binary Classification stage is carried out.

Traffic Capture Process

The major intend of this stage is to record the source IP address, source port, destination IP address, destination port, packet size, and timestamp of all sent IP packets from IoT device which is a part of some smart home application. It is complex task to gather the DDoS traffic as security risks and complexity are involved in it [16]. Three most common variations of DDoS attack such as TCP SYN flood, a UDP flood and an HTTP GET flood are stimulated in this stage for capturing the newly arrived variants in the malware properties.

Grouping

The packets are grouped from IoT devices on the basis of source IP address. After that, the partition of these packets is done into non-overlapping timestamps recorded in the initial phase.

Feature Extraction Process

This phase is executed to generate two kinds of attributes for each packet on the basis of the IoT device behavior. Stateless attributes are lightweight whose extraction is done from flow independent qualities of each sent packet. This means the IP source did not divide the incoming traffic stream to create them. Moreover, the stateful features are utilized to capture the aggregated flow of information in the network traffic concerning short time spans. Packet size and Inter-packet interval are of this kind of attributes whereas bandwidth and IP address cardinality and novelty are known as stateful features.

Classification

Finally, various classifiers such as KNN (K-nearest neighbor), RF (random forest), SVM (support vector machine) and DNN (deep neural network) are utilized to perform the binary classification so that the normal traffic can be differentiated from the DDoS traffic flow. Additionally, the DL (deep learning) classification algorithms also provide efficiency due to their utilization on the additional data generated from the real-world deployments [17].

Machine Learning Algorithms for IoT Bot Detection

Various MLAs (machine learning algorithms) are implemented for developing a mathematical model on sample data which adds the potential to computers for deciding without any explicit programming. Specially, the DL which is an enhanced ML approach has a unique ability to extract the attributes automatically from large-scale, high-speed network traffic generated through the interconnected heterogeneous IoT devices.

Artificial Neural Network (ANN)

It is a basic structure of the NN technique. This approach is based on biological neural networks (NNs) (neural networks). However, this model is utilized to learn the vector-valued functions.

Neurons are elements that act in tandem and are connected to one another. A subgroup employs to handle the elements having connection through the links is known as the layer. Each connecting link focuses on possessing the related weights and requires for the adjustment to attain the optimal weight values through the backward calculation stages. Thus, the activation function is presented as the critical role model for attaining these optimal weight vectors. In the given equations, popular activations such as sigmoid, ReLU, and tanh functions are presented.

$$f(x) = \frac{1}{1+e^{-x}}$$

$$f(x) = \frac{\sin h\,(x)}{\cos h\,(x)} = \frac{e^{x}-e^{x-x}}{e^{x}+e^{-x}}$$

$$f(x) = \max(0, x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

KNN

KNN (K Nearest Neighbor) algorithms are planned on the basis of Euclidean distance and majority of vote of its K neighbors are exploited with the entity of dissimilar classes to classify the object [18]. The accuracy of KNN algorithm is based on the selected number of neighbors which denotes the value of K. The value of K is obtained in the form of selected odd number utilized to perform the binary classification so that the possibility of two classes that attain the labels in same count, can be avoided. In case the value of K=1, the entity is simply assigned to its single nearest class. Only optimal value must be selected for the K. In case of least value of K, it could be under-fitting and its larger value leads to over-fitting of the model. The Euclidean distance amid two points (X, Y) in Euclidean n-space is expressed as:

$$d(X, Y) = \sqrt{\sum_{a=1}^{n} (Ya - Xa)^2}$$

Gaussian Naïve Bayes Machine Learning Algorithm

The NB algorithm (Nave Bayes) is a supervised machine learning (ML) algorithm based on the Bayes theorem. Bayes' theorem makes the implementation of method of maximum likelihood of case happening in accordance with the previous learning. In simple language, Bayesian probability can be defined as:

P(X/Y)=(P(Y/X))/(P(Y))

At which X and Y denote the occurrence and P(Y) is not equal to zero, P(X/Y) shows the likelihood of X case happening given that Y is true, P(Y/X): Likelihood of Y case happening given that X is true, P(X) and P(Y) are likelihood of observing case X and Y, respectively [19].

J48 Algorithm

J48, also known as C4.5 is an enhanced model of the ID3 algorithm whose construction is done using Ross Quinlan. It is a tree-based method that consists of the root node and the leave nodes. The extraction of these nodes is performed from the root. To build a tree, the information gain and entropy values for all the attribute values are computed. This model is capable of handling the uncompleted data points, continuous and discrete values and avoiding the overfitting problem due to utilization of the pruning methods.

LITERATURE REVIEW

Bot-IoT Attacks Traffic Identification using Machine Learning Technique

Shafiq *et al*. suggested a novel hybrid approach for recognizing the anomaly traffic in IoT (Internet of Things). At first, BoT-IoT dataset was implemented. This dataset contained 44 effective attributes whose selection was done from various features for the ML (machine learning) algorithm [20]. Thereafter, five ML methods were adopted in order to recognize the malicious and anomaly traffic. This approach focused on selecting extensive parameters of ML. The effectiveness of machine learning model was proved while recognizing the IoT anomaly and traffic. The deployment of a bijective soft set approach and its algorithm was also done. Moreover, the suggested approach was exploited on the basis of bijective soft set approach. The results obtained in experiments exhibited the

efficiency of the suggested approach while selecting the finest ML algorithm among various algorithms. Moreover, the NB (Naïve Bayes) algorithm performed well for detecting the anomaly and intrusion in IoT network. An innovative system was introduced by Shafiq *et al*. for recognizing the traffic in Bot-IoT [21]. Initially, a technique named CorrAUC was implemented for selecting the features. This technique was planned on the basis of wrapper method for filtering the attributes in accurate manner and selecting the suitable features for the selected ML (machine learning) algorithm with regard to AUC (area under the curve) metric. Subsequently, the TOPSIS was incorporated with the Shannon entropy on the basis of a bijective soft set so that the selected features were computed to identify the malicious traffic in the IoT network. The Bot-IoT data set was applied for quantifying the introduced system. The experimental results revealed that the introduced system performed successfully and yielded accuracy of around 96% on average. The ML (machine learning) technique was developed by Alshamkhany *et al.* for detecting the Botnet attacks [22]. There were four ML techniques known as NB (Naïve Bayes), SVM (Support Vector Machine), KNN (K-Nearest Neighbor) and DT (Decision Tree) to achieve this. The Bot-IoT dataset and UNSW (University of New South Wales) dataset that contained 82000 records had been utilized. The results confirmed the supremacy of DT among all the algorithms. This algorithm offered the accuracy of 99.89%, precision up to 100% and recall of 100% to detect the botnet attack in IoT (Internet of Things). Nguyen *et al.* proposed a new high-level PSI-rooted subgraph-based feature model to detect IoT botnets [23]. Next, a finite amount of attributes were extracted which contained precise behavioral descriptions as small space was required for them and they were useful for mitigating the processing time. Five machine learning models were used to exploit the PSI-rooted subgraph-based attributes: 1) RF (Random Forest), 2) DT (Decision Tree), 3) KNN (K-Nearest Neighbor), 4) SVM (Support Vector Machine), and 5) Bagging. The outcomes indicated that the detection rate obtained from every model was calculated 97% and the projected model consumed least time in comparison with other algorithms. A graph-based ML (machine learning) system was investigated by Alharbi *et al*. to detect the botnet attack in IoT (Internet of Things) [24]. The graph attributes were taken in account to construct a generalized model so that the botnets were detected on the basis of chosen attributes [24].

Table 1. Summary of Literature Review.

Author	Year	Technique used	Dataset	Outcomes
Muhammad Shafiq *et al.* [20]	2020	A novel hybrid approach	Bot-IoT dataset	The NB (Naïve Bayes) algorithm performed well for detecting the anomaly and intrusion in IoT network.
Muhammad Shafiq *et al.* [21]	2021	CorrAUC	Bot-IoT data set	The experimental results revealed that the introduced system performed successfully and yielded accuracy around 96% on average.
Mustafa Alshamkhany *et al.* [22]	2020	NB (Naïve Bayes), SVM (Support Vector Machine), KNN (K-Nearest Neighbor) and DT (Decision Tree)	UNSW-NB15 dataset and Bot-IoT	The outcomes indicated that the detection rate obtained from every model was calculated 97% and the projected model consumed least time in comparison with other algorithms.
Huy-Trung Nguyen *et al.* [23]	2020	PSI-rooted subgraph-based feature model	GitHub	The outcomes indicated that the detection rate obtained from every model was calculated 97% and the projected model consumed least time.
Afnan Alharbi *et al.* [24]	2021	Graph-based ML (machine learning) system	CTU-13 and IoT-23	The attributes of investigated system led to alleviate the training time as well as model complexity and higher bots' detection rate in contrast to other methods.

Five filter-based feature evaluation measures generated through a variety of theories namely consistency, correlation, and information were employed to select the various feature sets that were further discovered. The investigated system was quantified on two heterogeneous botnet datasets such as CTU-13 and IoT-23 and its comparison was done with other ML techniques. The experimental outcomes validated that the attributes of investigated system led to alleviate the training time as well as model complexity and higher bots detection rate in contrast to other methods. Furthermore, diverse kinds of botnet families were detected, and this system was robust against zero-day attacks (Table 1).

Bot-IoT Attacks Traffic Identification using Deep Learning Technique

Popoola *et al*. emphasized on mitigating the feature dimensionality of large-scale IoT (Internet of Things) network traffic data with the execution of the encoding stage of LAE (long short-term memory autoencoder) system [25]. The network traffic samples were classified in exact manner by analyzing the long-term inter-related changes in the low-dimensional feature set generated from the LAE. For this purpose, deep BLSTM (bidirectional long short-term memory) was deployed. The BoT-IoT data set was executed in the experimentation for computing the presented system. The LAE was able to reduce the memory space used for large-scale network traffic data storage by 91.89%, according to the findings. The presented approach performed better than traditional techniques and enhanced the accuracy by 27.03%. The deep BLSTM model was robust against the model underfitting and overfitting and attained superior generalization potential in binary and multiclass classification scenarios. Sriram *et al.* discussed that smart cities were based on the internet-enabled devices utilized in various applications such as medical sector and traffic control, etc. for improving its efficiency [26]. In order to provide increased cyber security solutions to IoT devices and smart city applications, a DL (deep learning) based botnet detection system that runs on network traffic flows was created. This system focused on gathering the network traffic flows, transforming them into connection records and utilizing a DL (deep learning) model for detecting attacks launched via the compromised IoT devices. Experiments on benchmark data sets were used to determine which model was the best. The results exhibited the superiority of the designed system over the traditional methods. A new technique known as N-BaIoT (network-based anomaly detection for the IoT) was formulated by Meidan *et al*. for extracting the behavior snapshots of the network [27]. Deep autoencoders were used to detect unusual network traffic coming from compromised IoT devices. There were nine financial IoT devices employed in the lab with two botnets such as Mirai and BASHLITE for computing the formulated technique. The outcomes depicted that the formulated method was capable of detecting the attacks in exact and quick rate due to their activation from the compromised IoT devices which were part of a botnet. An approach was recommended by McDermott *et al.* with the objective of detecting the botnet activity in IoT (Internet of Things) networks [28]. The DL (Deep Learning) application was adopted for building a framework named BLSTM-RNN (Bidirectional Long Short-Term Memory based Recurrent Neural Network) so that the attack was detected. The text was identified, and the attack packets were transformed into tokenized integer format using WE (Word Embedding). A comparative analysis was carried out on the recommended framework against the existing model to detect four attack vectors utilized through the Mirai botnet. The results indicated that the recommended framework provided 99% accuracy to detect Mirai botnet attack, 98% for UDP and 98% for DNS. A DL (deep learning) based technique was established by Liu *et al*. to detect the botnet in IoT (Internet of Things) [29]. The damped incremental statistics was adopted for extracting the basic traffic attributes of IoT devices and the Z-Score technique was utilized for normalizing the attributes. After that, a CNN (convolutional neural network) algorithm was developed for learning the dataset and implementing the trained CNN for detecting the traffic. The experimental results demonstrated that the established technique had potential for distinguishing the benign traffic and various kinds of attack traffic in efficient way and this technique yielded 99.57% accuracy (Table 2).

Table 2. Study of Used Techniques.

Author	Year	Technique used	Dataset	Outcomes
Segun I. Popoola *et al*. [25]	2021	LAE (long short-term memory autoencoder) and BLSTM (bidirectional long short-term memory)	Bot-IoT dataset	The presented approach performed better than traditional techniques and enhanced the accuracy by 27.03%.
S. Sriram *et al*. [26]	2020	DL (deep learning) based botnet detection system	N-BaIoT	The results exhibited the superiority of the designed system over the traditional methods
Yair Meidan *et al*. [27]	2018	N-BaIoT based on deep autoencoders	Bot-IoT dataset	The outcomes depicted that the formulated method was capable of detecting the attacks in exact and quick rate
Christopher D. McDermott *et al*. [28]	2018	BLSTM-RNN (Bidirectional Long Short-Term Memory based Recurrent Neural Network)	GitHub	The recommended framework provided 99% accuracy to detect Mirai botnet attack, 98% for UDP and 98% for DNS.
Junyi Liu *et al.* [29]	2019	CNN (convolutional neural network)	UCI Machine Learning	The established technique had potential for distinguishing the benign traffic and various kinds of attack traffic in efficient way and this technique yielded 99.57% accuracy

Bot-IoT Attacks Traffic Identification using Optimized Machine Learning

Injadat *et al*. suggested an optimized ML (machine learning)-based system named BO-GP-DT in which BO-GP (Bayesian optimization Gaussian Process) algorithm was integrated with DT (decision tree) classifier for detecting the attacks on IoT devices effectively and efficiently [30]. The Bot-IoT-2018 dataset was executed for quantifying the performance of the suggested system. The experimental results depicted that the suggested system performed well concerning superior accuracy, precision, recall and F-score and proved to be efficient and robust to detect the botnet attacks in IoT environments. A hybrid framework named BD-PSO-V was introduced by Asadi *et al.* in which botnet detection was done with integration of PSO and voting system for dealing with the anomalies [31]. The significant and efficient attributes were selected using PSO (particle swarm optimization) to detect the botnets. The SVM (support vector machine) and C4.5 DT (decision tree) were included in the voting system for recognizing the botnets and classifying the samples. The decision was made in this system on the basis of majority of votes. The introduced framework was evaluated on ISOT and Bot-IoT datasets. The simulation validated that the introduced framework provided enhanced the accuracy by 0.42% on initial dataset and 0.17% on the Bot-IoT dataset in comparison with the other techniques. A SSLPSO-SVM (Smart Self-Adaptive Learning Based Particle Swarm Optimization Support Vector Machine) technique was developed by Moodi *et al*. for recognizing the android botnet accurately [32]. A new method called SSS (Smart Selection Strategy) was adopted in this technique. The results indicated that the developed technique performed more efficiently with regard to sensitivity, specificity, precision and accuracy. An ESVNN (enhanced support vector neural network) was projected by Jagadeesan *et al*. to detect the botnet attack [33]. The accuracy was improved by selecting the efficient attributes of traffic flows. These features were utilized for input. The AF (Artificial Flora) algorithm was put forward to improve the performance of this algorithm. The simulation results confirmed that the projected algorithm yielded superior accuracy and F-measure in contrast to other models. The recall of projected algorithm was calculated as 0.8636, accuracy was 0.8684 and F-score was 0.8669. A Fuzzy Logic based feature engineering technique was constructed by Joshi *et al.* to detect botnet attack. The fundamental purpose of this technique was to discover the fuzzy components in the dataset and create the fuzzy sets [34]. The ANN (Artificial Neural Network) employed the generated attributes to classify the botnet attack. The ANN algorithm was trained and computed on CTU-13 dataset. The constructed technique and implemented classification algorithm provided the accuracy of around 99.94% (Table 3).

Table 3. Review Methodology of Comparison Table.

Author	Year	Technique used	Dataset	Outcomes
Mohammad Noor Injadat *et al.* [30]	2020	A hybrid technique named BO-GP-DT	Bot-IoT-2018 dataset	The suggested system performed well concerning superior accuracy, precision, recall and F-score and proved efficient and robust to detect the botnet attacks.
Mehdi Asadi *et al.* [31]	2020	BD-PSO-V	ISOT dataset and the Bot-IoT dataset	The introduced framework provided enhanced the accuracy by 0.42% on initial dataset and 0.17% on the Bot-IoT dataset.
Mahdi Moodi *et al.* [32]	2021	SSLPSO-SVM (Smart Self-Adaptive Learning Based Particle Swarm Optimization Support Vector Machine)	Android Botnet dataset	The developed technique performed more Efficiently with regard to sensitivity, specificity, precision and accuracy.
S. Jagadeesan *et al.* [33]	2021	ESVNN (enhanced support vector neural network)	Bot-IoT dataset	The recall of projected algorithm was calculated 0.8636, accuracy was 0.8684 and F-score was 0.8669.
Chirag Joshi *et al.* [34]	2021	Fuzzy Logic based feature engineering technique and ANN	CTU-13 dataset	The constructed technique and implemented classification algorithm provided the accuracy Around 99.94%.

REVIEW METHODOLOGY

The state-of-the-art review layout is shown in this section, which is a step-by-step procedure for the literature described in the preceding sections. The goal of this study is to classify the present literature on smart cities and machine learning, as well as to assess current trends.

Development of Review Protocol

This analysis identifies relevant research publications from reliable electronic databases and the field's most prestigious conferences. The number of papers that were considered was then reduced using inclusion and exclusion criteria. The final research studies were then picked based on a range of factors. The data presented here is the result of extensive research.

Sources of Information

Various electronic database sources were researched for this review study; some of the most common electronic databases used in this search are indicated in Figure 7.

Figure 7. Popular Electronic Databases.

Inclusion and Exclusion Criteria

This literature review spans the years 1990 through 2021 and contains both quantitative and qualitative research studies. As this subject is investigated for a range of smart city techniques, the keywords "smart city", "machine learning algorithms for smart city", "smart city algorithms" pointed to a significant number of results. The search was done using the search term "machine learning approach [in, for] [Smart city]" in the abstract and title. There are research studies from various conferences, journals, and book chapters included in this collection. Only the relevant articles have been chosen from the retrieved results. The relevance of the methodologies, publications, and conferences are the selection factors for this work.

Extraction Outcomes

The relevant work of fake news algorithms is recovered from the large collection of data provided by search engines using the inclusion criterion, which is based primarily on approaches. Figure 8(a and b) show the interpretation of the selected kind of articles and year of work. The accompanying graph illustrates those journals (51%) account for the majority of the work in this study, with conferences accounting for 40% and book chapters accounting for 9%. In addition, the graph below shows a year-by-year analysis of smart city-related effort.

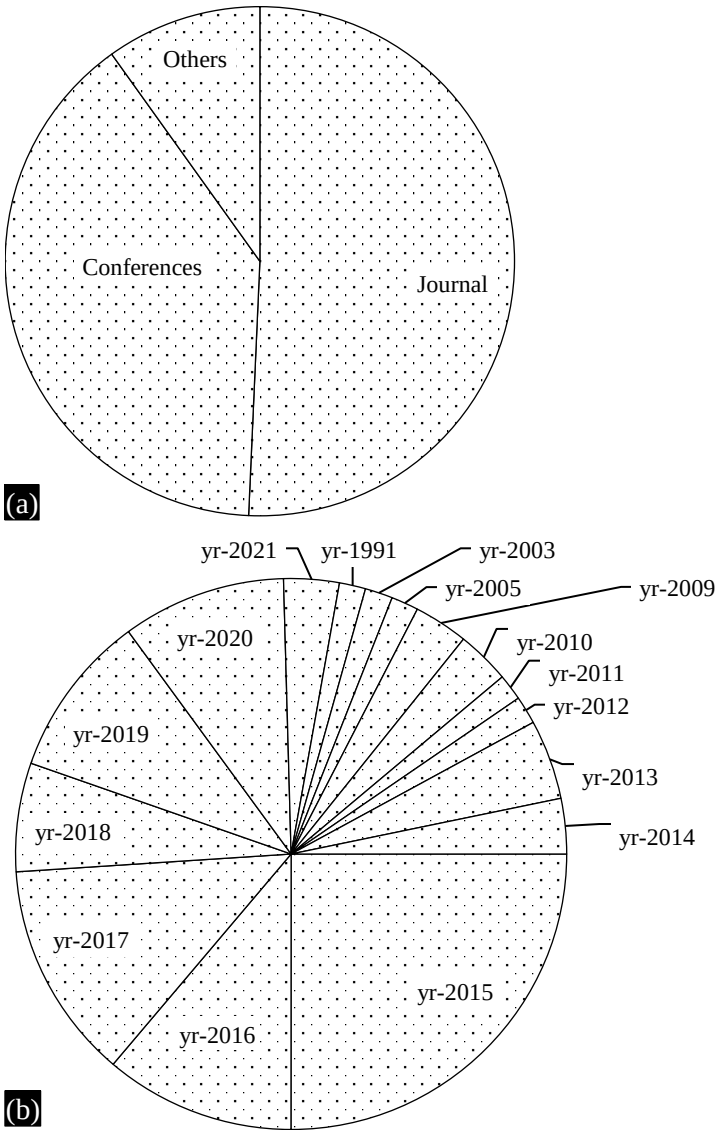

Figure 8. (a) Included research works-type of paper, (b) year-wise included research.

CONCLUSION

The internet has transmuted from an appealing new innovation to a vital piece of framework. Since the common people have become dependent on the data, communication and worthwhile resources available online, there is an increasing aspiration for everything linked to the world wide web. The reliability edge links IT framework to the potential for failures and security risks. From IoT smart device, it is required to manage ID and access control besides trust that the device is secure and approved to link to IT set-up with accountable developments to improve and maintain security. The various types of schemes for the smart city are reviewed from different sources like journals, conferences and others. In is concluded that the machine learning schemes are the most advanced schemes for the smart city and machine learning.

REFERENCES

1. Faisal Hussain, Syed Ghazanfar Abbas, Fayyaz Ubaid U, Shah Ghalib A, Abdullah Toqeer, Ahmad Ali. Towards a Universal Features Set for IoT Botnet Attacks Detection. IEEE 23rd International Multitopic Conference (INMIC). 2020; 1–6.
2. Christos Tzagkarakis, Nikolaos Petroulakis, Sotiris Ioannidis. Botnet Attack Detection at the IoT Edge Based on Sparse Representation. Global IoT Summit (GIoTS). 2019; 1–6.
3. Yair Meidan, Michael Bohadana, Yael Mathov, Yisroel Mirsky, Asaf Shabtai, Dominik Breitenbacher, Yuval Elovici. N-BaIoT: Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders. IEEE Pervas Comput. 2018; 17(3): 12–22.
4. Ashley Woodiss-Field, Johnstone Michael N. Assessing the Suitability of Traditional Botnet Detection against Contemporary Threats. Workshop on Emerging Technologies for Security in IoT (ETSecIoT). 2020; 18–21.
5. Syed Muhammad Sajjad, Muhammad Yousaf. UCAM: Usage, Communication and Access Monitoring Based Detection System for IoT Botnets. 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE). 2018; 1547–1550.
6. Nikolaos Giachoudis, Georgios-Paraskevas Damiris, Georgios Theodoridis, Georgios Spathoulas. Collaborative Agent-based Detection of DDoS IoT Botnets. 15th International Conference on Distributed Computing in Sensor Systems (DCOSS). 2019; 205–211.
7. Yan Naung Soe, Paulus Insap Santosa, Rudy Hartanto. DDoS Attack Detection Based on Simple ANN with SMOTE for IoT Environment. 4th International Conference on Informatics and Computing (ICIC). 2019; 1–5.
8. Arash Mahboubi, Seyit Camtepe, Keyvan Ansari. Stochastic Modeling of IoT Botnet Spread: A Short Survey on Mobile Malware Spread Modeling. IEEE Access. 2020; 8: 228818–228830.
9. Riaz Ullah Khan, Rajesh Kumar, Mamoun Alazab, Xiaosong Zhang. A Hybrid Technique to Detect Botnets, Based on P2P Traffic Similarity. Cybersecurity and Cyberforensics Conference (CCC). 2019; 136–142.
10. Lihua Yin, Xi Luo, Chunsheng Zhu, Liming Wang, Zhen Xu, Hui Lu. ConnSpoiler: Disrupting C&C Communication of IoT-Based Botnet through Fast Detection of Anomalous Domain Queries. IEEE Trans Industr Inform. 2020; 16(2): 1373–1384.
11. Gokhan Sagirlar, Barbara Carminati, Elena Ferrari. AutoBotCatcher: Blockchain-Based P2P Botnet Detection for the Internet of Things. IEEE 4th International Conference on Collaboration and Internet Computing (CIC). 2018; 1–8.
12. Bharath Sudharsan, Dineshkumar Sundaram, Pankesh Patel, Breslin John G, Muhammad Intizar Ali. Edge2Guard: Botnet Attacks Detecting Offline Models for Resource-Constrained IoT Devices. IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops). 2021; 680–685.
13. Dwyer Owen P, Marnerides Angelos K, Vasileios Giotsas, Troy Mursch. Profiling IoT-Based Botnet Traffic Using DNS. IEEE Global Communications Conference (GLOBECOM). 2019; 1–6.
14. Hassan Jalil Hadi; Syed Muhammad Sajjad; Khaleeq un-Nisa. BoDMitM: Botnet Detection and Mitigation System for Home Router Base on MUD. 2019 International Conference on Frontiers of Information Technology (FIT). 16-18 Dec. 2019. Islamabad, Pakistan: IEEE; 2020.

15. Ayush Kumar, Teng Joon Lim. EDIMA: Early Detection of IoT Malware Network Activity Using Machine Learning Techniques. IEEE 5th World Forum on Internet of Things (WF-IoT). 2019; 289–294.
16. Thomas Lange, Houssain Kettani. On Security Threats of Botnets to Cyber Systems. 6th International Conference on Signal Processing and Integrated Networks (SPIN). 2019; 176–183.
17. Madhuri Gurunathrao Desai, Yong Shi, Kun Suo. IoT Bonet and Network Intrusion Detection using Dimensionality Reduction and Supervised Machine Learning. 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). 2020; 0316–0322.
18. Ly Vu, Van Loi Cao, Quang Uy Nguyen, Nguyen Diep N, Dinh Thai Hoang, Eryk Dutkiewicz. Learning Latent Representation for IoT Anomaly Detection. IEEE Trans Cybern. 2020; 1–14.
19. Myint Soe Khaing, Yee Mon Thant, Thazin Tun, Chaw Su Htwe, Mie Mie Su Thwin. IoT Botnet Detection Mechanism Based on UDP Protocol. IEEE Conference on Computer Applications (ICCA). 2020; 1–7.
20. Muhammad Shafiq, Zhihong Tian, Mohsen Guizani. Selection of effective machine learning algorithm and Bot-IoT attacks traffic identification for internet of things in smart city. Future Gener Comput Syst. 2020; 107: 433–442.
21. Muhammad Shafiq, Zhihong Tian, Ali Kashif Bashir, Xiaojiang Du, Mohsen Guizani. CorrAUC: A Malicious Bot-IoT Traffic Detection Method in IoT Network Using Machine-Learning Techniques. IEEE Internet Things J. 2021; 8(5): 3242–3254.
22. Mustafa Alshamkhany, Wisam Alshamkhany, Mohamed Mansour, Mueez Khan, Salam Dhou, Fadi Aloul. Botnet Attack Detection using Machine Learning. 14th International Conference on Innovations in Information Technology (IIT). 2020; 203–208.
23. Huy-Trung Nguyen, Quoc-Dung Ngo, Van-Hoang Le. PSI-rooted subgraph: A novel feature for IoT botnet detection using classifier algorithms. ICT Express. 2020; 6(2): 128–138.
24. Afnan Alharbi, Khalid Alsubhi. Botnet Detection Approach Using Graph-Based Machine Learning. IEEE Access. 2021; 9: 99166–99180.
25. Popoola Segun I, Bamidele Adebisi, Mohammad Hammoudeh, Guan Gui, Haris Gacanin. Hybrid Deep Learning for Botnet Attack Detection in the Internet-of-Things Networks. IEEE Internet Things J. 2021; 8(6): 4944–4956.
26. Sriram S, Vinayakumar R, Mamoun Alazab, Soman KP. Network Flow based IoT Botnet Attack Detection using Deep Learning. IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). 2020; 189–194.
27. Yair Meidan, Michael Bohadana, Yael Mathov, Yisroel Mirsky, Asaf Shabtai, Dominik Breitenbacher, Yuval Elovici. N-BaIoT: Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders. IEEE Pervas Comput. 2018; 17(3): 12–22.
28. McDermott Christopher D, Farzan Majdani, Petrovski Andrei V. Botnet Detection in the Internet of Things using Deep Learning Approaches. International Joint Conference on Neural Networks (IJCNN). 2018; 1–8.
29. Junyi Liu, Shiyue Liu, Sihua Zhang. Detection of IoT Botnet Based on Deep Learning. Chinese Control Conference (CCC). 2019; 8381–8385.
30. Mohammad Noor Injadat, Abdallah Moubayed, Abdallah Shami. Detecting Botnet Attacks in IoT Environments: An Optimized Machine Learning Approach. 32nd International Conference on Microelectronics (ICM). 2020; 1–4.
31. Mehdi Asadi, Mohammad Ali Jabraeil Jamali, Vahid Majidnezhad. Detecting botnet by using particle swarm optimization algorithm based on voting system. Future Gener Comput Syst. 2020; 107: 95–111.
32. Mahdi Moodi, Mahdieh Ghazvini, Hossein Moodi. A hybrid intelligent approach to detect Android Botnet using Smart Self-Adaptive Learning-based PSO-SVM. Knowl-Based Syst. 2021; 222: 106988.

33. Jagadeesan S, Amutha B. An efficient botnet detection with the enhanced support vector neural network. Measurement. 2021; 176: 109140.
34. Chirag Joshi, Ranjeet Kumar Ranjan, Vishal Bharti. A Fuzzy Logic based feature engineering approach for Botnet detection using ANN. J King Saud Univ Comput Inf Sci. 2021.

