Survey of Traffic Classification Solution in IoT Networks

Rami J. Alzahrani
Department of ComputerScience King Abdulaziz University
Jeddah, KSA

Ahmed Alzahrani
Department of ComputerScience King Abdulaziz University
Jeddah, KSA

ABSTRACT
The Internet of Things (IoT) is creating a new evolution in the present and future Internet. The idea of IoT is to establish transmission capacities using a ubiquitous, distributed and diverse gadgets network. The rapid growth of the IoT makes the incorporation and connection of several devices a predominant procedure. The increasing numbers of IoT devices and diverse IoT traffic patterns has created the need for traffic classification methods to provide solutions for IoT applications'issues. Although it has been presented in many papers and surveys, network traffic classification is still undeveloped well in IoT because of the variations in traffic classifications in IoT and NonIoT gadgets. This paper discusses the arising patterns of IoT network traffic classifications and putting them in practical use. It also presents an overview of traditional traffic classification methods, as well as a discussion with a categorization.This paper evaluated the performance metrics such as accuracy, recall, precision and F1 score for these Machine Learning algorithms: Decision Tree (DT), K-Nearest Neighbors (K-NN), Naïve Bayes (NB) and Gradient Boosting (GRB) classifiers. The analysis of normal and attack traffic is done by using WEKA software tools and by utilizing the BoT-IoT dataset [1].

General Terms
In this paper, the traffic classification in IoT is considered as a general term. Throughout this research, present and past studies of different method and technique of traffic classification in IoT is considered to improve the security solutions of IoT.

Keywords
IoT, Security, Networks, Traffic Classification, IoT security

1. INTRODUCTION
The recent advance in information technology has created a new era named Internet of Things (IoT). This new technology allows objects (things) to be connected to the Internet such as the smart TVs, printers, cameras, smartphones, smartwatches etc. This trend provides new services and applications for many users and enhances the lifestyle. The report written by Cisco [2] predicted that almost five hundred billion devices will contain sensors and will be associated with the Internet by 2030. The report expressed that the Internet of Things (IoT) is the network that connects these gadgets for correspondence and exchanging data. The data produced by these smart devices is accumulated, evaluated, then distributed for more processes by IoT services. The data formats of the IoT network are various and they have various protocols for various applications applying various technologies. The technologies presented by IoT are improved continuously because they become an integral part of people's needs and everyday lives. Due to the challenges of lost or constrained identity facing IoT environment, maintaining security and privacy for data in IoT is crucial. The IoT has appeared as a phenomenon that develops into an essential environment in which the Internet are connected people, things, processes and data to each other. By 2022, as shown in the figure 1 below Machine to Machine (M2M) [3] networks will witness a growth by almost more than half of the world's 28.5 billion connected M2M will get from tracking applications, mechanization devices and security control. Brilliant transmission will be supplied by applications for Internet access, amusement and independent driving that will represent the most increasing part in manufacturing. Because of the increasing number of connected gadgets, it is predicted that worldwide M2M IP correspondence will develop[2]. This expansion will generate an amount to traffic larger than the number of connections, and this is caused by an increased usage of not only picture and audio but also video applications from M2M associations. Taking this into consideration, future communications will be combined with IoT connections and devices. By 2022, future communication will be based on smart phones which will constitute about 45% of global IP traffic [2].

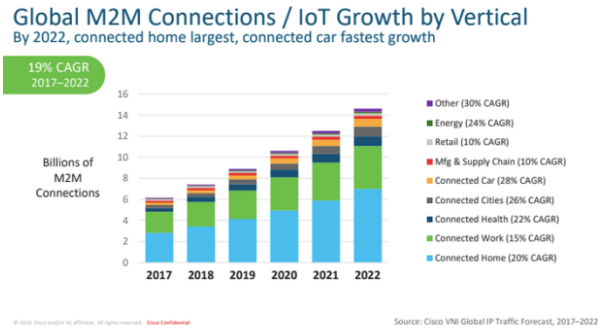

Fig 1: Global M2M Connections IoT Growth

This tendency shows the smartphones' influence on the way people use the Internet to access data. Therefore, new network needs and requirements are generated by the effect of IoT trends. The IoT has developed traffic evolution implications, heterogeneous network transformations and extensive consciousness for network security improvements[4]. Classification of network traffic is a fundamental requirement for different network applications like accounting, monitoring, security etc. It is also essential to predict future needs by analyzing the network traffic for continuous supplying of network resources. Moreover, a lot of network services for example status reports, firewalls, QoS (Quality of Service)systems, and intrusion detection systems make use of an effective categorization of network traffic.

2. PROBLEM DEFINITION Normal
The increasing number of IoT gadgets and various features of IoT traffic designs have drawn consideration to traffic categorization techniques to handle emerging problems in IoT

domain implementation. Although network traffic categorization has been examinedin many previous studies and existing papers, it is still not well developed in IoT and this is because of the variations in traffic features in IoT and Non-IoT gadgets. This review gives attention to the rising patterns of network traffic arrangement in IoT and the implementation of traffic categorization in the domains of IoT. The IoT traffic is unique in relation to different sorts of network traffic which follows steady design and expected network behaviors[5]. Based on this fact, this paperpresents an overall survey of the current IoT problems and solutions that have been tended to by network traffic categorization.

3. STUDY OBJECTIVES

This paper discusses the existing movement of IoT network traffic categorization and the use of traffic categorization in its domains. The table1 below gives a comparison of different traffic classification methods. Using tools to monitor network is a common behaviour because their usefulness encourages the data required to achieve network management. Its primary applications are collecting data from the network and analysing data in order to control the network. Using network traffic controlling tools also makes the examination of network execution obtainable.

4. OVERVIEW OF NETWORK TRAFFIC CLASSIFICATION

Nowadays, improving network security along with the Quality of Service (QoS) has become of great importance. Different services to users with a categorization of the network traffic are needed, and to provide these services by integrating Internets of Things (IoT) with Machine Learning (ML) methods are used. Therefore, it is necessary to differentiate between malicious traffic and normal traffic. Subsequent to identifying malicious traffic, it must be blocked and send the normal traffic to the accurate nodes for serving the client's needs. Here, present Machine Learning (ML) algorithms for categorizing the network traffic, for monitoring the crowding in the network. Recently, the use of IoT is growing rapidly and researchers in the field are combining their efforts to find innovative solutions for IoT effective services [6]. Various kinds of traffic design are existing in IoT, for example, Machine to Machine (M2M) transmission, Smart gadget transmission, ubiquitous sensor networks, Machine Type Communication (MTC) and Machine Oriented communication[7]. A different category is used to separate gadgets which are essential for IoT networks. The devices are separated according to the sensing mechanism into three categories: Periodic Sensing, and Event Driven Sensing Query Based Sensing [8]. The first category is Periodic Sensing gadget which able to sense the following: humidity, temperature and lights occasionally and send this data to the control unite which is located in the centre. The second category is Event Driven Sensing gadgets which is gadgets make traffic when a particular event produced. The traffic here is more random and inconsistent. The third category is Query Based Sensing gadget which is gadgets produce traffic as a consequence of the query. They create random and inconsistent traffic. Devices in IoT are gotten to and controlled remotely, consequently traffic categorization is an integral part in IoT to organize and separate the traffic for several reasons, for instance, Quality of Service (QoS,) services to users, recognition of unusual traffic and to assign a legitimate channel to the normal traffic. Although a great deal of front methods has been suggested by the experts for requesting the traffic, IoTstill needs more impressive approaches to manage distinguish traffic.

4.1 Network Traffic Classification Methods

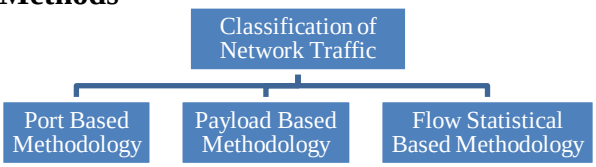

Fig 2: A Network Traffic Classification Method

As shown the figure 2 the classification of network traffic divided into three methods.

4.1.1 Port Based Methodology

Traffic classification utilizing port number is the most ordinary strategy that connect port numbers to application and differentiate the traffic. Port based classifiers use TCP and UDP headers of the packets to collect the information that related to the port number. Following to deciding the port number, the connection is made amongst the assigned TCP and UDP and isolated port numbers for the classification of the traffic. This planning technique is viewed as the fastest and easiest demanding technique for traffic classification [9]. As shown by [10], only just the percentage number between 30% to 70% of the current Internet traffic can be requested using Port Based planning methods. The problem is that this technique has limitations. As approximately ofthe applications probably won't have their ports selected with Internet Assigned Number Authority (IANA), for instance, Peer to Peer (P2P) applications such as Napster and Kazaa application. Sometimes, the ports are assigned increasingly based on the need. In addition to that, IoT interacting devices can connect with one another by interchange more data among, so this strategy isn't appropriate for categorization.

4.1.2 Payload Based Methodology

To get rid of the limitations and consistency of port based categorization, a payload based model of categorization was suggested dependent on the examination of the packets' headers as well as by utilizing their substance too. In this technique, an assessment of a packet is complete called Deep Packet Inspection (DPI). To inspect packets, correspondence flow is utilized [11]. This method identifies recognised designs interior the packets and provide best percentage of discovery.

4.1.3 Statistical Based Methodology

This method be contingent on the deference to information that can be obtain from the packet header such as inter arrival times, bytes sent, TCP window size [11]. They rely upon packet header high level information which made them a better option than to be in charge of the inaccessible payloads or dynamic ports. Machine Learning (ML) strategies are used in these strategies to accomplish categorization. ML procedures utilized information established for instructing the classifier. There are two kinds of ML algorithms which are: Supervised learning and Unsupervised learning [12]. In the first type which is supervised learning technique, the algorithm requires labelled dataset for learning, in this technique there is a necessity for supervision to train the specified Model. The reply received after training data in the model is utilized to evaluate its accuracy. This procedure is utilized to take care of categorization and minimize difficulties. In the second type which is unsupervised learning technique is completely the conflicting of the supervised learning technique, in which no necessity of labelled information the algorithm attempts to take out structures to discover patterns all alone. No need for supervision in this

type of technique. The problems of clustering and association can be addressed by utilizing unsupervised learning technique [12]. IoT traffic categorization utilizing Machine Learning (ML) for both mechanisms Port Based and Payload Based methodologies weaknesses can be under control by utilizing ML based methodologies utilized. Implementing ML procedures [12] for Network categorization utilizing next steps for classification: First, data collection is done in the network traffic capturing phase. Second, extracting features of algorithm as follows: Packet Received Rate, Node Status, Transmitted Bytes, Duration of Flow, Received Bytes, Percentage of both Lost Byte Rate and Lost Packet Rate. Finally, building the model was based on the choice of algorithm and it requires training the data set that has been created in the initial step. The table1 below shows the comparison between different traffic classification methods: The table2 below shows the pros and cons of traffic classification methods:

Table 1. Traffic Classification Methods

Category	Classification Methodology	Attribute Used	Processing Time
Port Based	Protocol port	Protocol ports	Low
Payload Based	Deep packet inspection	Use a unique pattern in the payload	High
	Stochastic packet inference	Statistical attributes inherent in packet header and payload	Moderate
Statistical Based	Packet based	Packet and payload size, interpacket arrival time	Moderate
	Flow based	Duration, transmission rate, multiple flow features	Low

Table 2. Pros and Cons of Traffic Classification Methods

Category	Pros	Cons
Port Based	-Simple to implementation -Very efficient even in large networks -The first packet can be used for classification	-Some applications don't have their ports registered -Applications may use ports other than its familiar ports (nonprivileged users) -Dynamic allocated port number (Real Video streamer) -IP layer encryption may complicate TCP and UDP header
Payload Based	-Which known as Deep Packet Inspection (DPI) -Classifier matches 'Signatures' in a payload to a known application -Regular expression for HTTP signature -Regular expression for FTP signature -It keeps away from confidence on fixed port number	-Important difficulty and processing load on the classifier -Difficult to implement on proprietary protocols -Privacy policy breaching -Cannot identify new applications -Application payload encryption
Statistical Based	-Take advantage of application range and inherent traffic footprints (flow parameters) to describe traffic and then derive categorization benchmarks over data mining techniques to recognize individual applications -Statistical categorization is viewed as lightweight and highly scalable from an operational perspective especially when constant or close to real time traffic identification is required -Beat the disadvantages of port based and payload based	-Difficulty generating signatures -Lack of accuracy

4.2 Taxonomy of Traffic Categorization in the IoT Network

This paper presents a classification of current implementations in IoT traffic categorization and designs. Regarding their attacking practices, DDoS attacks have presented a various variation over recent years and still being under examination using several options. In terms of attacking methods, there is no major difference between traditional DDoS and specific DDoS. These attacking methods use similar techniques to take advantage of weaknesses existeither in traditional systems or in an IoT gadgets. But IoT explicit DDoS attacks are more complicated and diverse because of diversity included in IoT gadgets. As a result of considerable differences amongst IoT devices' traffic characteristics and the traffic produced by other gadgets, different angles and solutions of IoT traffic categorization have been recently suggested. This partpresents several viewpoints of IoT that have been tended by utilizing traffic categorization strategies. This paper present reviews of some existing papers in IoT which have put in for exact or public IoT information sources for example testbed, gadget and dataset on their technique. This part primarily concentrates on introducing the ability to apply traffic characterization to show the probabilities and extendibility of utilizing traffic categorization in the IoT space and don't paying attention to explaining their methodologies in details. The figure 3 below shows the traffic categorization in the IoT space.

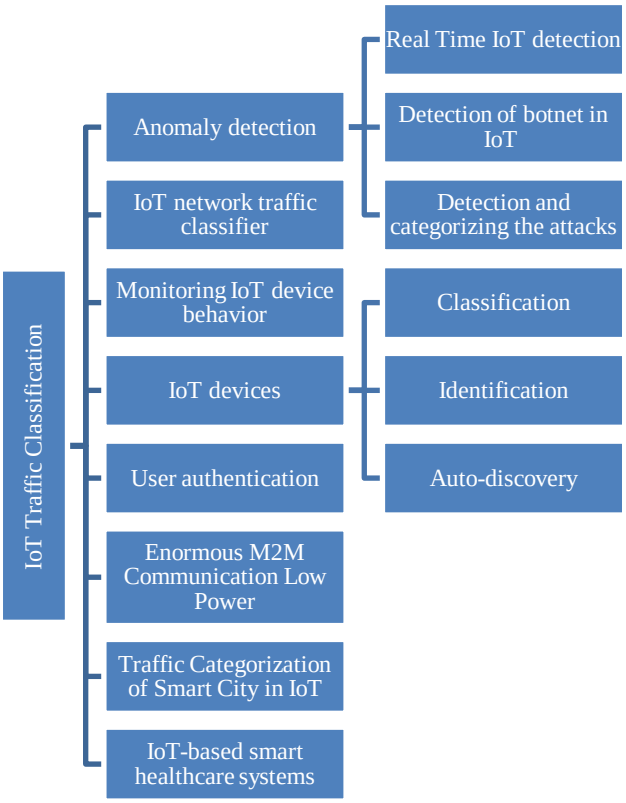

Fig 3: Traffic Categorization in the IoT Space

4.2.1 *Anomaly Detection in IoT*

By utilizing Machine Learning (ML) statistical analysis the traffic anomalies can be detect. This statistical analysis is based on comparing the behavior's signature of flowing information with the ordinary structure of it which are caught and demonstrated previously. A lot of studies have been made dependent on various targets, for example, complexity, speed and accuracy.To categorize network traffic, thepaper by[13]compressedmeta-learning techniques versus individual classifiers. Voting, Stacking, Bagging, and Boosting are among the meta-learning strategies investigated and evaluated. The paper then provided a novel experimental examination of several meta-learning strategies, also known as ensemble learners, and compressed them to their basic classifiers when they were used.

4.2.1.1 *Detection and Categorizing the Attacks*

Since the appearance of DDoS/DoS attacks, detecting them has been an essential priority in network security. The work in [14] suggested an architecture to reveal abnormal activities of IoT devices by merging spread equivalent computation for preparing of a big data capacity connected with Machine Learning (ML) algorithms. The suggestedmethod utilizes the idea of equivalent computing to achieve a training and examining phase of the created technique from ML algorithms on numerous computation nodes for exploiting the speed. To accomplish this idea, a VM (Virtual Machine) was utilized that tool Hadoop Distributed File System (HDFS) and the Yet Another Resource Negotiator (YARN) resource manager tools and Spark of Hadoop.In terms of time, the model training stage and testing stage performed roughly 7 and 6.3 times faster, respectively.The study in[15] offered a heterogeneous IDS (Intrusion Detection System) by applying deep learning methods to differentiate between normal actions and unnormal ones such as flooding, impersonation and injection. The work suggested a deep auto encoded Dense Neural Network (DNN) model structure which ensures significant level of precision and reducing the latency. This technique suggested the auto encoder which offers a compacted and less boisterous shape of the input area utilizing an unsupervised pretraining technique on the information. It likewise utilizes a supervised categorizer in the last DNN intrusion detection situation. The accuracy was 99.8%for flooding, impersonation and injection.

4.2.1.2 *Detection of Botnet in IoT*

The greatest Distributed Denial of Service (DDoS) attack was executed through the IoT malicious which is Mirai in 2016. This type of attack was extensively known and it caused inconvenience to some controlling service facilities of the Internet and that done by focusing on the Domain Name Service (DNS) servers and this type of malicious affect for example Amazon, Twitter, and Netflix [16]. The research paper [17] suggested an algorithm in a wide-ranging network to recognize the attack like Mirai in IoT malware bots. This suggested technique investigates the network traffic's signature produced via the IoT gadgets which the Mirai malware undermined. The study classifies the IoT gadgets into two categories: gadgets defenceless to attacks and gadgets that are not, then expansion role is implemented for devices that are exposed to attacks and those which are not. Moreover, the suggested methodology indicates a sampling density and allocates the incidence to IoT gadgets for controlling and monitoring motivation to recognize any potential corruption.The highest values foraverage detectiondelays between 0 and10,000 (per packetelapsed)

4.2.1.3 *Real Time IoT Detection*

The authors in this study [18] presented model that is called IoT-guard which can predict and recognize malicious and unmalicious gadgets behaviours, and this guard is a self-adjustable categorization structure that utilizes a semi managed learning method. IoT guard utilizes the network traffic created by gadgets in addition to the information taken out from network logarithms which are created by Access Points (AP) and entries. The suggested work joins all the data generated by the device from various kinds of traffic to recognize patterns of the network. Due to the unreachability of session marks for greater part of the network logarithms, so the execution of an unsupervised learning technique is done by utilizing this methodology by which it gains from the different designs discovered in network traffic.Without specialist hardware, predict traffic class in 250 milliseconds.

4.2.2 *IoT Devices Autodiscover (Network Monitoring)*

In IoT gadgets, achieving security and avoiding malicious attacks can be guaranteed by utilizing effective disclosure tools to monitor IoT gadgets on the networks. The method is otherwise called gadget acknowledgement by utilizing network traffic examination [19]. It is considered an empirical implementation that can prohibit malicious attacks in the network and detect any undesired task by monitoring their generated traffic. This can be explained by taking the traffic generated by a brilliant plug and it is expected that it will change things powers on/off as an alternative of viewing videos from YouTube. consequently, in this situation any abnormalities will be easily monitoring and detected. An Autodiscover IoT gadget protocol in real time utilizing two phase design is presented to locate IoT gadgets in a network and observe the border of Intranet [5]. The first stage is monitoring suspicious IP parts by a stateless TCP SYN filtering technique for the disclosure of active hosts TCP stack to accomplish a quicker information handing out rate. The

second stage is distinguishing every IoT gadget on different protocols utilizing the Per Instance Algorithm Configuration (PIAC) algorithm which ensures the precision and coverage of IoT gadget observation. The suggested technique succeeded to accomplish an individual port [20] implemented a ML based technique to classify IoT gadget in the network. The goal was to recognize if traffic stream is associated to any exact identified IoT gadget. The suggested technique can differentiate IoT gadgets from nonIoT ones. Similarly, it is also able to recognize the model of the device such as IP camera or smart TV in a network utilizing a technique that is based on network behavior. A group of categorizations dependent on ML algorithms was implemented in a multistage procedure through traffic movements produced via a specific gadget possibly categorized through IP address. In the same way, [21] proposed an approach to recognize unidentified IoT gadgets and differentiate between IoT gadgets and non-IoT gadgets. In this method, the generated traffic is used to model the behavior of a device. The technique uses a deep LSTM (Long Short Term Memory) auto encoder network to demonstrate the TCP structures of every gadget and group the structures utilizing Bayesian Hyper Parameter Tuning model. The suggested approach appeared to have a powerful precision in corresponding gadgets to their categorized equivalents. The performance was 82% averageF1 scoreand 70% accuracy formthe correct the class ofunseen gadgets. Two techniques called packet inspection, and statistical classification were employed in another work in [22] to recognize IoT gadgets and the created traffic runs by the gadgets. The assessment is implemented over the utilize of random forest learning for gadgets determination and traffic categorization. This study as well explained that network traffic categorization with mechanized learning algorithms can be used to make the device identification feasible. The performance was 98.7% overall accuracyprecision. IOT SENTINEL [23], [24] suggested and applied a framework to recognize the IoT gadgets' types which is a group of gadget technique and software edition. This framework similarly has the capacity to apply a set of strategies for limiting the association of defenseless gadgets. Utilizing a ML based categorization model, IOT SENTINEL will be able to identify device types by their signature in the network traffic. It controls and manage the activity of gadgets in network correspondence throughout the installation procedure. The implemented approach in [5] analyses both the network traffic to recognize IoT gadgets and the raw network traffic to withdraw two-way flows. The examination work prepares and examine different ML algorithms to achieve network traffic categorization. Similarly, the work in [25] built up a ML framework that categorizes IoT gadgets. A characterization procedure was established upon numerous traffic structures such as signaling patterns, action cycles, transmission protocols, and cipher suites algorithms. Additionally, a framework utilizes a categorization established upon ML in a multistage situation that exceptionally identifies IoT gadgets with over 99% precision. The suggested technique assessed compromises between speed, costs and precision of the categorize in a constant way. In a similar way, to recognize new and unknown gadgets in the network, the research in [26] implemented a mechanized categorization technique that utilize the structures of the created traffic by IoT gadgets. This technique defines the specifications of various devices by making use of the rich data transported through the traffic streams of IoT networks. To assess the efficiency of the implemented technique, an actual IoT dataset was used. Discoveries from the evaluation stated that the implemented technique is effective of transporting an acceptable level of performance in recognizing gadgets.

4.2.3 Verify and Monitor the Behavior of IoT Device

IoT gadgets that are distantly associated through mediums for example the Internet are vulnerable to attackers either interior or exterior a network. Moreover, attackers are able to exploit TCP and UDP ports which are unprotected in the gadgets inside a house or organization. They utilized these structures to indicate or expand their attacks to enter a network. These type of security worries have driven internal and worldwide governments to establish criterions rules for the Internet population to give a safe IoT gadgets and services facilities. An Internet Engineering Task Force (IETF) have proposed Manufacturer Usage Description (MUD) [27] describes an authorized frame to get the behavior of the gadget's runtime and also a capacity to assess it clearly. This structure calls the producer of IoT gadgets to give the categorization and ethical description of the gadgets. Consequently, the ethical and categorization of the gadget from the time of connection in a network can be supposed. These conditions differ for each IoT gadgets from various producers. The good understanding of IoT gadget requirement will benefit to decrease the dangers of attacks by offering the network manager a capacity to arrange the streaming and working gadgets with a determined group of ACL (Access Control List). These studies, [28], [29] implemented and established a framework to create Manufacturer Usage Description side view for IoT gadget dependent on the features and attitude of traffic indication of gadget. There are two components implemented: first, it catches and monitor all TCP ad UDP streams on both way: to the device and from the device. Second, it creates MUD side view from the taken stream categorization. The framework then legitimately complies with IoT gadget MUD side view utilizing the MUDdy tool. To confirm MUD side view, it utilizes the idea of Metagraphs to assess MUD side view for interior uniformity and its implementation with the specified guidelines. MUD is a developing idea and a few implementations presented on how producers can design behavioral side view for users' IoT gadgets or how normal forms must utilize these side view to users' network security.

4.2.4 Monitor and manage IoT Network Traffic Categorization

The work in [11] implemented a supervised Deep Learning (DL) technique depending upon movement measurements of networks. It utilizes the measurements from packet headers throughout the stream existence and creates a period sequences of attribute directions. Afterwards, a technique will be able to recognize the connected service and application to a stream. In this work, the categorize depends upon the collaboration of both deep learning techniques CNN and RNN which known as Convolutional Neural Network and Recurrent Neural Network. They found that 5 to 15 packets were enough to achieveexcellent detectionresults.

4.2.5 IoT Networks Clients' Verification Technique

The work in [30] utilizes the categorization of network traffic throughout app entry of end client gadgets. This technique categorizes each individual incident and makes a categorization technique dependent upon learning the users' past admission designs of the actions. Therefore, to recognize any anomalies in the designs, the created traffic throughout the applications on the end client gadget is supervised by the ML categorization technique through the both access requests and session. The accuracy was 95%.

4.2.6 *An Enormous M2M Communication Low Power*

Nowadays, there are many researches works which try to decrease the consumption of energy and power saving of sensor nodes by utilizing enhancement of communication collection and different path routing. Nevertheless, utilizing network traffic categorization, the study in [31] implemented a traffic conscious of AB (Ambient Backscatter) transmission technique to decrease the expendedpower in the network to encourage the collecting and communication chances for AB transmission. The technique uses a BNP (Bayesian Non-Parametric) ML algorithm for traffic applications' classification. Similarly, a stochastic geometric examination was utilized to enhance the traffic design option standard. To ensure the precision of traffic designs' classification, the implemented technique reflect on three structures of network traffic, for example, packet length and time and difference in packet length. By utilizing the procedures mentioned previously. The implemented technique acquires an enhanced method of correspondence that limits power utilization in the cellular network including primary and secondary users.

4.2.7 *Traffic Categorization of Smart City in IoT*

The massive measure of created information is a major primary structures of smart city networks. The ability to analyze this massive data plays an important part for many reasons for instance quality of service and security. Artificial Intelligent (AI) methods in this area are considered a helpful implementation for examining this enormous information. In addition, network traffic classification is of great importance to ensure network security and prevent attacks. Moreover, traffic categorization enhances effectiveness in the performance and ensure the QoS of IoT gadgets. Taking this into consideration, the study [32] offers a deep learning technique by implementing the concept of container network to contract with the categorization of massive information. Container network is chosen because it can take advantage of DL procedure in relations of preparing information on a largescale just as a better coarseness attribute extraction [33]. It implemented end to end categorization technique to expand the execution of the CNN[34] as it is informed [32] that the end-to-end categorization technique can rearrange the difficulty of attribute displacement in the traditional traffic categorization algorithm. The implemented technique utilizes an image processing learning technique which changes every information stream to two phases of extent matrix after division. Then the matrix will be transmitted to the categorization procedure.In terms of distinguishing between malware and benign traffic, the accuracy was 100 %.

4.2.8 *Smart Healthcare Implementation in IoT Domain*

IoT offerings plenty of applications and implementations to serve practically all parts of life. The healthcare system is the greatest significant fields in the current situations. It utilizes IoT applications to offer a good service for residents. The research work in [35] implemented a technique to proficiently accomplish the created traffic from Wireless Body Area Network. It takes advantage of the idea of SDN (Software Defined Networking) through it categorizes the traffic created by a wireless body area network to three collections, for example, emergency information, sensor health traffic and environmental information. In this implemented design, the PrivacyPreserving Data Aggregation (PDA) gets information from sensor nodes and categorize different traffic kinds. Therefore, it sends the categorized information to the essential server for additional examination. In the implementation, an SDN controller distantly set up and modifies stream tables. PDA utilizing the communication services, for instance, Wi-Fi and 3G or 4G networks based on the service provider and network rules.For health packets to reach the server, the average delay and jitter were 0.00016 s and 0.000239 s, respectively. The delay's standard deviation was 0.000379 seconds.

5. PERFORMANCE EVALUATION

In this research the BoT-IoT dataset which was developed at the University of New South Wales Canberra at Australia is used. The dataset contains a legitimate traffic and simulated IoT traffic alongside different types of attacks. The reason behind choosing this dataset was because it represents a realistic IoT ecosystem environment. The dataset contains DDoS, DoS, UDP, TCP, Theft, HTTP OS Fingerprint and Service Scan, Keylogging, Reconnaissance and Data exfiltration attacks. All these data were pre-processed to identify network-level patterns for different kinds of traffic that devices create and use these patterns to detect any intrusion behaviours in the IoT Infrastructure[1]. The experiments in this study are using a hardware specification of Intel® Core™ i7-8650U CPU @ 1.90GHz processor, 16 GB RAM with the operating system Windows 10, 64bit. The experiments are done using WEKA (Waikato Environment for Knowledge Analysis) version 3.9.4 which is a data mining tool used for data pre-processing, classification, regression, clustering, association rules and visualization. WEKA was written in Java code and it is an open-source tool developed at the University of Waikato in New Zealand. All the algorithms used in this research are supported in WEKA such as Decision Tree (DT), K-Nearest Neighbors (K-NN), Naïve Bayes (NB) and Gradient Boosting (GRB) classifiers using k-fold cross-validation k = 10. The table 3 below shows binary categorization precision, recall, and F1 score.

Table 3. Binary categorization precision, recall, and F1

ML		Rec	Pre	F1
DT	Normal	87%	87%	87%
	Attack	100%	100%	100%
	Average	93.5%	93.5%	93.5%
K-NN	Normal	32%	77%	45%
	Attack	100%	100%	100%
	Average	66%	88.5%	72.5%
NB	Normal	9%	2%	4%
	Attack	100%	100%	100%
	Average	54.5%	51%	52%
GRB	Normal	11%	73%	19%
	Attack	100%	100%	100%
	Average	55.5%	86.5%	59%

The table 4 below showsthesummary of binary and multiclass results.

Table 4. Summary of binary and multiclass results

	Binary Classification			
	DT	K-NN	NB	GRB
Accuracy (%)	99.96	99.97	99.85	99.99
Recall (%)	93.5	66	54.5	55.5
Precision (%)	93.5	88.5	51	86.5
F1 Score (%)	93.5	72.5	52	59.5
	Multiclass Classification			
	DT	K-NN	NB	GRB
Accuracy (%)	99.96	93.00	78.17	99.88
Recall (%)	95	63	42	93
Precision (%)	97	69	26	95
F1 Score (%)	96	65	22	93

The figure 4 below shows the average performance of classifiers in multiclass classification for all normal and attack instances. The DT classifier recorded better results in terms of the average accuracy and average precision in normal instances and attacks categorization with 99.96% and 97%, respectively. In contrast, the NB recorded the lowest average accuracy and average precision with 78.17% and 26%, respectively. I summary, the results show that the better accuracy found in DT and GRB algorithms. This indicates its effectiveness in distinguishing normal and attack instances.

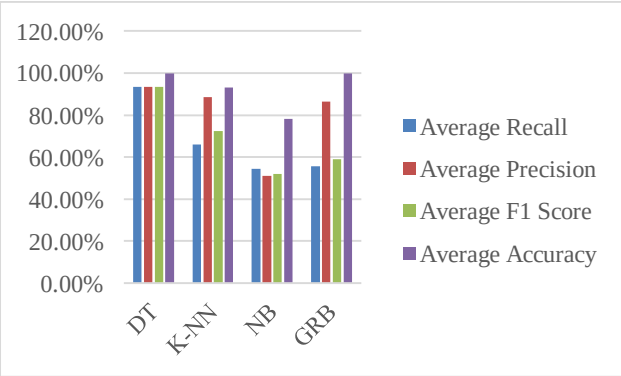

Fig 4: Average performance of classifiers in multiclass

6. DISCUSSION AND RECOMMENDATION

Because of its capacities to examine and finish up the network traffic, IoT traffic characterization has been as of late focused by the scholarly local area. Nonetheless, in light of the fact that this zone of exploration is as yet in the beginning phase, it will experience numerous difficulties and open many examination problems. The idea of traffic arrangement is indistinguishable in IoT and Non-IoT spaces by which network traffic is examined for various objectives. The classification in traffic IoT faces various difficulties, for example, infrastructure, data source, standardization, and quality of service. These difficulties should be completely tended to in IoT to be conveyed as a solid arrangement. The greater part of the IoT gadgets is wireless from which various traffic designs are created. Security violation should be accurately examined through considering substantial and modern datasets. Various conditions and frameworks demand explicit investigations and examinations. These are viewed as seriously exciting issues because of the diverse nature of IoT. After the presence of the IoT notion, different kind of gadgets are associated with one another. IoT gadgets expose an enormous scope of varieties in sorts containing wearable gadgets, static, implantable and portable. It is understandable that the nature of IoT traffic is unique in relation to the regular network traffic. Consequently, traffic arrangement in the IoT area is facing difficulties when compared with non-IoT space. This is on the grounds that, in IoT space, the assortment of gadget sand its diverseness is exceedingly higher than in the non-IoT area. In spite of the fact that the primary worries in traffic arrangement are security and Quality of Service (QoS) issues. This paper introduced a complete review of the present and past studies in IoT traffic characterization with regards to application and design in IoT. In the survey presented, it is obviously communicated that the fundamental attention of the papers in IoT is the traffic characterization towards security worries. The performance evaluation of this research shows that DT and GRB recorded a better accuracy result. These good performances will help improve the security of the IoT networks. It is value referencing that IoT traffic characterization studies are as yet considered in a beginning level and needed to be examined in actual practical assessments and datasets. Building up a dataset from actual world use cases and gadgets is profoundly needed to realize and understand IoT traffic characterization studies.

7. ACKNOWLEDGMENT

I am glad that I completed this work successfully. This work would be impossible without the help of Dr. Ahmed Alzahrani. I would like to thank him for his expert advice and usual support.

8. REFERENCES

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

[2] “Cisco Visual Networking Index: Forecast and Trends, 2017–2022,” 2019.

[3] O. Garcia-Morchon, S. Kumar, and M. Sethi, “Internet of Things (IoT) Security: State of the Art and Challenges,” Apr. 2019. doi: 10.17487/RFC8576.

[4] A. Sabella, R. Irons-Mclean, and M. Yannuzzi, Orchestrating and Automating Security for the Internet of Things: Delivering Advanced Security Capabilities from Edge to Cloud for IoT. 2018.

[5] M. R. Shahid, G. Blanc, Z. Zhang, and H. Debar, “IoT Devices Recognition Through Network Traffic Analysis,” in Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018, Jan. 2019, pp. 5187–5192, doi: 10.1109/BigData.2018.8622243.

[6] S. P. Khedkar and R. Aroulcanessane, “SDN enabled cloud, IoT and DCNs: A comprehensive Survey,” Sep. 2019, doi: 10.1109/ICCUBEA47591.2019.9129091.

[7] S. P. Khedkar and R. AroulCanessane, "Machine Learning Model for classification of IoT Network Traffic," Nov. 2020, pp. 166–170, doi: 10.1109/i-smac49090.2020.9243468.

[8] F. Tang, Z. M. Fadlullah, B. Mao, and N. Kato, "An Intelligent Traffic Load Prediction-Based Adaptive Channel Assignment Algorithm in SDN-IoT: A Deep Learning Approach," IEEE Internet of Things Journal, vol. 5, no. 6, pp. 5141–5154, Dec. 2018, doi: 10.1109/JIOT.2018.2838574.

[9] "Discriminators for use in flow-based classification," 2013.https://qmro.qmul.ac.uk/xmlui/handle/123456789/5050 (accessed Jan. 01, 2021).

[10] Y. Qi, L. Xu, B. Yang, Y. Xue, and J. Li, "Packet classification algorithms: From theory to practice," in Proceedings - IEEE INFOCOM, 2009, pp. 648–656, doi: 10.1109/INFCOM.2009.5061972.

[11] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Network Traffic Classifier with Convolutional and Recurrent Neural Networks for Internet of Things," IEEE Access, vol. 5, pp. 18042–18050, Sep. 2017, doi: 10.1109/ACCESS.2017.2747560.

[12] M. Shafiq, X. Yu, A. A. Laghari, L. Yao, N. K. Karn, and F. Abdessamia, "Network Traffic Classification techniques and comparative analysis using Machine Learning algorithms," in 2016 2nd IEEE International Conference on Computer and Communications, ICCC 2016 - Proceedings, May 2017, pp. 2451–2455, doi: 10.1109/CompComm.2016.7925139.

[13] I. P. Possebon, A. S. Silva, L. Z. Granville, A. Schaeffer-Filho, and A. Marnerides, "Improved Network Traffic Classification Using Ensemble Learning," in Proceedings - IEEE Symposium on Computers and Communications, Jun. 2019, vol. 2019-June, doi: 10.1109/ISCC47284.2019.8969637.

[14] I. Kotenko, I. Saenko, A. Kushnerevich, and A. Branitskiy, "Attack Detection in IoT Critical Infrastructures: A Machine Learning and Big Data Processing Approach," in Proceedings - 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing, PDP 2019, Mar. 2019, pp. 340–347, doi: 10.1109/EMPDP.2019.8671571.

[15] S. Rezvy, Y. Luo, M. Petridis, A. Lasebae, and T. Zebin, "An efficient deep learning model for intrusion classification and prediction in 5G and IoT networks," Apr. 2019, doi: 10.1109/CISS.2019.8693059.

[16] "Hacked cameras, DVRs powered today's massive internet outage. Krebson Security - Google Search." .

[17] A. Kumar and T. J. Lim, "Early Detection OfMirai-Like IoT Bots In Large-Scale Networks Through Sub-Sampled Packet Traffic Analysis," Lecture Notes in Networks and Systems, vol. 70, pp. 847–867, Jan. 2019.

[18] I. Hafeez, A. Y. Ding, M. Antikainen, and S. Tarkoma, "Real-time IoT device activity detection in edge networks," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Aug. 2018, vol. 11058 LNCS, pp. 221–236, doi: 10.1007/978-3-030-02744-5_17.

[19] J. Shen, Yi. Li, B. Li, H. Chen, and J. Li, "IoT Eye An Efficient System for Dynamic IoT Devices Auto-discovery on Organization Level," in Proceedings - 4th IEEE International Conference on Cyber Security and Cloud Computing, CSCloud 2017 and 3rd IEEE International Conference of Scalable and Smart Cloud, SSC 2017, Jul. 2017, pp. 294–299, doi: 10.1109/CSCloud.2017.66.

[20] Y. Meidanet al., "ProfilIoT: A machine learning approach for IoT device identification based on network traffic analysis," in Proceedings of the ACM Symposium on Applied Computing, Apr. 2017, vol. Part F128005, pp. 506–509, doi: 10.1145/3019612.3019878.

[21] J. Ortiz, C. Crawford, and F. Le, "DeviceMien: Network device behavior modeling for identifying unknown IoT devices," in IoTDI 2019 - Proceedings of the 2019 Internet of Things Design and Implementation, Apr. 2019, pp. 106–117, doi: 10.1145/3302505.3310073.

[22] M. R. P. Santos, R. M. C. Andrade, D. G. Gomes, and A. C. Callado, "An efficient approach for device identification and traffic classification in IoT ecosystems," in Proceedings - IEEE Symposium on Computers and Communications, Nov. 2018, vol. 2018-June, pp. 304–309, doi: 10.1109/ISCC.2018.8538630.

[23] M. Miettinen, S. Marchal, I. Hafeez, N. Asokan, A. R. Sadeghi, and S. Tarkoma, "IoT SENTINEL: Automated Device-Type Identification for Security Enforcement in IoT," in Proceedings - International Conference on Distributed Computing Systems, Jul. 2017, pp. 2177–2184, doi: 10.1109/ICDCS.2017.283.

[24] M. Miettinen et al., "IoT Sentinel Demo: Automated Device-Type Identification for Security Enforcement in IoT," in Proceedings - International Conference on Distributed Computing Systems, Jul. 2017, pp. 2511–2514, doi: 10.1109/ICDCS.2017.284.

[25] A. Sivanathanet al., "Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics," IEEE Transactions on Mobile Computing, vol. 18, no. 8, pp. 1745–1759, Aug. 2019, doi: 10.1109/TMC.2018.2866249.

[26] L. Bai, L. Yao, S. S. Kanhere, X. Wang, and Z. Yang, "Automatic Device Classification from Network Traffic Streams of Internet of Things," in Proceedings - Conference on Local Computer Networks, LCN, Feb. 2019, vol. 2018-October, pp. 597–605, doi: 10.1109/LCN.2018.8638232.

[27] "Manufacturer Usage Description Specification," 2019. https://tools.ietf.org/html/rfc8520 (accessed Dec. 30, 2020).

[28] A. Hamza, D. Ranathunga, H. H. Gharakheili, M. Roughan, and V. Sivaraman, "Clear as MUD: Generating, validating and applying IoT behavioral profiles," in IoT S and P 2018 - Proceedings of the 2018 Workshop on IoT Security and Privacy, Part of SIGCOMM 2018, Aug. 2018, vol. 18, pp. 8–14, doi: 10.1145/3229565.3229566.

[29] A. Hamza, D. Ranathunga, H. H. Gharakheili, T. A. Benson, M. Roughan, and V. Sivaraman, "Verifying and Monitoring IoTs Network Behavior using MUD Profiles," arXiv, Feb. 2019.

[30] Y. Ashibani and Q. H. Mahmoud, "A User Authentication Model for IoT Networks Based on App Traffic Patterns," in 2018 IEEE 9th AnnualInformation Technology, Electronics and Mobile Communication Conference, IEMCON 2018, Jan. 2019, pp. 632–638, doi: 10.1109/IEMCON.2018.8614892.

[31] S. H. Kim and D. I. Kim, "Traffic-Aware Backscatter Communications in Wireless-Powered Heterogeneous Networks," IEEE Transactions on Mobile Computing, vol. 19, no. 7, pp. 1731–1744, Jul. 2020, doi: 10.1109/TMC.2019.2913386.

[32] H. Yao, P. Gao, J. Wang, P. Zhang, C. Jiang, and Z. Han, "Capsule Network Assisted IoT Traffic Classification Mechanism for Smart Cities," IEEE Internet of Things Journal, vol. 6, no. 5, pp. 7515–7525, Oct. 2019, doi: 10.1109/JIOT.2019.2901348.

[33] H. Zheng, F. Yang, X. Tian, X. Gan, X. Wang, and S. Xiao, "Data gathering with compressive sensing in wireless sensor networks: A random walk based approach," IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 1, pp. 35–44, Jan. 2015, doi: 10.1109/TPDS.2014.2308212.

[34] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," 2017. doi: 10.5555/3294996.3295142.

[35] F. Sallabi, F. Naeem, M. Awad, and K. Shuaib, "Managing IoT-Based Smart Healthcare Systems Traffic with Software Defined Networks," Nov. 2018, doi: 10.1109/ISNCC.2018.8530920.