|  |
| --- |
| Advanced Anomaly Detection  Using a Machine Learning Algorithm  Focusing on IoT Datasets |

Abstract

The use and deployment of IoT devices have been enormously increasing, they surely bring convivence and comfort and save our time as the work done with IoT devices are fast. IoT devices have been a blessing to us however blessing can turn over as a curse by bringing unexpected risks, information leak and takeover of devices due to virus infected IoT devices. IoT devices are vulnerable for many risks like the impact of threats, devices cannot monitor the threats, cannot understand network well and have limited performance.

Therefore, it is high time we introduce new security measures for IoT devices and its dataset. Anomaly in the datasets is the uncommon and unusual deviation from the normal pattern of the data that is the dataset. Detecting Anomaly in the IoT devices dataset is a challenging task as there are various sensors deployed and the whole system is dynamic.

This paper mainly focusses on the anomaly detection of IoT devices using the AWS platform and the one that we use is AWS Sage Maker under which Random Cut Forest is the main algorithm that is implemented in the research. We will be analysis a random dataset available in the data structure and detecting the anomaly accordingly. The performance and accuracy of the algorithm will also be calculated.

Table of Contents

[Abstract 2](#_Toc21898551)

[Introduction 4](#_Toc21898552)

[Objective and Aims 6](#_Toc21898553)

[Preliminary Literature Review 7](#_Toc21898554)

[Research Problem 17](#_Toc21898555)

[Research Methodology 18](#_Toc21898556)

[Research Significance 33](#_Toc21898557)

[Conclusion 34](#_Toc21898558)

[Research Planning 35](#_Toc21898559)

[References 36](#_Toc21898560)

[Appendix 41](#_Toc21898561)

[References 51](#_Toc21898562)

**List of figures**

[**Figure 1 Graph showing anomaly detection 21**](#_Toc21899560)

[**Figure 2 Graph showing threshold 21**](#_Toc21899561)

[**Figure 3 Graph showing anomaly in RCF 29**](#_Toc21899562)

**List of tables**

[**Table 1Research Planning 36**](#_Toc21899610)

[**Table 2 Difference in requirements of 5G and 6G 47**](#_Toc21899611)

# Introduction

This is the new era of modern technology and the Internet of Things (IoT) is the most emerging topic these days. This defines the next generation of Internet where various things are connected through various devices that sense the information via the network and communication protocols, and exchange the information communicating over information media, to apply intelligent ID, setting up, tracking, organization, and extra other tasks. There has been a wide range of research and investigation tracking the wireless sensors networks and how the devices can sense making our life easier and more autonomous. IoT has a huge impact on our lifestyle, it affects our behavior as well as our reaction since it is an integral part of our modern lifestyle. The upcoming future generation will not be even to don any of their daily activities without IoT. A gigantic network of inter-connected devices is IOT, these devices assemble and share data, the way that they are used and the whole environment where the operation of these devices takes place. This whole process is worked out utilizing sensors which are embedded in all the physical device from smartwatch to smartphones. These sensors release data continuously for which they are manufactured and their working state of the environment and the essential thing is in what way do they share this huge amount of data and how can it be fruitful for us. This huge amount of data and interconnected networks has many benefits but also it has grown in its complexity and size. As there is a range of interconnected networks, there will be more and more technical problems rising with them. Managing, analyzing, monitoring, reviewing and timely detection of the network itself is a daunting task.

We are not only living in a world full of data but we are living in a world where the data is getting bigger every time. While the devices are sending all the data to the cloud and there is the huge number of devices, it is difficult to manage all the devices, therefore, continuous monitoring is needed that is design to visualize the status of IoT devices and to make sure the devices are running inactive and normal condition. Thus, there can be some states at which IoT devices have some errors. The gathered data should necessarily be analyzed in order the get actionable perceptions. One kind of job which can be commendably undertaken alongside data analysis is anomaly detection. The main objective is to find uncommon behavior which differs considerably from what has been expected before which has always been in a certain specific pattern.

There are different techniques and algorithm that is used to detect the anomaly and their performance is also studied in the literature review section. However, in this research paper, we have selected the AWS platform to detect the anomaly in IoT devices. Amazon Sage Maker is the service that we are applying here and under which many algorithms are associated like K-means clustering, PCA, Neural Topic Modeling, Linear learner, XGBoost, Binary Classification, and Random Cut Forest. But we have selected Random Cut Forest algorithm is implemented. The performance of Random Cut Forest will be calculated and based on the performance we will be choosing the best algorithm to detect the anomaly in any IoT device data sets.

# Objective and Aims

The preliminary objective of this thesis is as follows:

* Find the anomaly detection in various IoT devices dataset using AWS
* Investigate the best AWS platforms to find out which algorithm is best suitable for the research
* Analysis and observation of the performance and accuracy of the algorithm

# Preliminary Literature Review

At present, many researchers are concern about anomaly detection on various IoT devices. (Yuhuai Peng 2016) has proposed a novel about the anomaly detection scheme which is multi-dimensional and multi-source that is vastly based on the classification of computing mode. Firstly, a classification edge computing model is placed both on the side of the sensor end and at the end of the base station to realize the load balance as well as the low latency of the processing of data. Afterward, based on the fuzzy theory, an algorithm is designed to detect the anomaly of the data. This is a data anomaly detection algorithm that is a single source that can broadly examine the anomaly detection results of several consecutive instants. Lastly, at the base station end an anomaly detection algorithm that is designed to sense the data of time and space.

(MarwaKeshk 2018) An anomaly detection framework for preserving the private data, which is called PPAD-CPS, that is developed mainly focusing on keeping the information confidential as well as observing and finding out malicious information in power systems and their network trafﬁc. Two chief modules are incorporated in the framework to detect the anomaly. In the initial phase, data is filtered and transformed into a new format, the data that is filtered and transformed is collected from a pre-processing module then it will accomplish the objective of privacy preservation. Furthermore, an anomaly detection section is recommended using a Gaussian Mixture Model (GMM) and Kalman Filter (KF) for exactly approximating the later probabilities of legitimate and anomalous events. The performance of the PPAD-CPS framework is measured by the help of two public datasets, which are the Power System and UNSW-NB15 dataset. The analysis and examination of the results of the experiment make it clear that the framework is more efficient than any other modern techniques that have been proposed aimed at achieving high privacy levels. Moreover, in terms of detection rate, the framework is marked way better than the other seven peer anomaly detection techniques. Similarly, based on false-positive rate and computational time, this framework outperforms other techniques so, PPAD-CPS is a very good technique for preserving the data.

Internet of Things has affected our life so vastly that we do not even realize the challenges that come in the security and data associated with it. People are adding up more and more automation and intelligence to the machines, therefore the figure of modes of operation is increasing. Based on IoT we develop systems from home automation to smart cities, the data generated is gigantic, with the massive number of data there comes the importance of security monitoring at cloud level. These data should be handled properly at the initial stage, if not it is prone to various kinds of attacks and vulnerabilities. (Fu, Yang et al. 2011)has presented a paper exposing anomaly detection performance analysis of different IOT data sets with the help of cloud services. This system does not require the assistance of any human to detect the anomaly from a certain pattern.

While IoT devices are mounting rapidly in homes, different attacks which are targeting these policies need a smart technique that can help in detection solution to protect this heterogeneous environment. (Parth Bhatt 2018) present an approach that was completely based on machine learning for anomaly detection to protect the network from various attacks. The system is a decision module, that has the goal of classifying related attacks on IoT network. A single-board computer is used in this approach which will evaluate the attacks in the network protocol systematically. Also, it helps to find out malicious attacks on various commercial off-the-shelf IoT devices. This approach will mainly be able to authenticate the efficiency and feasibility of the system in a realistic scenario. Among the various approaches that were used before for the same cases, this approach was more feasible and easier to be implemented in comparison to others.

IoT has varied and scattered behavior that makes predictable intrusion detection methodologies tough to implement. To solve this problem, (Fu, Yang et al. 2011)proposed this paper based on the anomaly mining an intrusion detection scheme. This paper is mainly divided into two different parts; on the first part, an algorithm to detect the anomaly mining is developed which will detect anomaly in the data of perception layer, similarly, in the second part based on the anomalies that are detected a distributed intrusion detection scheme is proposed. In this distributed intrusion detection technique, the anomalies that can be triggered by attackers are easily detected but subsequently, some anomalies are not triggered by malicious intrusion, the intrusion semantic is considered to differentiate intrusion behaviors after anomalies. Conclusively, the result of this analysis mainly shows that this approach is precise and extensible in various perspectives and can be modified according to the need in the future.

In the large-scale IoT based environment where a lot of sensors are used, the main idea to observe the faulty behavior in the system is using the online anomaly detection techniques. There are a lot of researches that have been carried out on online anomaly detection although there are various thrilling developments of research that have been made, the extremely vigorous distribution of the sensors in the IoT environment makes it difficult to detect the anomaly in the online platform. (JunPing Wang 2014) presents this paper mainly targeting large scale service of IoT where an online anomaly learning and forecasting mechanism is implemented. In the first phase, the model that the system used is the reversible-jump MCMC learning which can be used online to learn anomaly-that is free of dynamics network and service data. Following, Network Utility Maximization (NUM) theory is implemented which is a fundamental consideration of IoT-based service topology. As a result of the experiment, the method demonstrates the exactitude in predicting crescendos network and service structures from synthetic data.

Several studies have been conducted for anomaly detection and the recent study conducted by (Ibrahim Alrashdi 2019) on Anomaly Detection of IoT Cyberattacks Smart City Using Machine Leeming describes the IoT cybersecurity threats in a smart city where they proposed an Anomaly Detection IoT (AD-IoT) system. This system is based on the Random Forest machine learning algorithm which is an automated system. The system was able to detect attacks that occur in the fog node of IoT devices.

(Brady, Magoni et al. 2018) proposed a system that detects the anomaly in the wide range of datasets and the huge number of devices. The main aim of the paper was to identify anomalies in the IoT devices by investigating using machine learning as a tool. Here the system was able to identify the anomalies that occur in IoT network using various state-of-the-art techniques in real-time. The result offers the practitioner to select the best techniques for their usage scenario. The paper mainly focuses on the comprehensive analysis for anomaly detection using the ML technique. The system uses both non-time series and time series data to evaluate that analysis was comprehensive. Two of the time series and one of the non-time series data were analyzed. Linear Discriminant Analysis and Decision tree algorithm was used for non-series data which provided the most consistent result. Similarly, in the case of the time-series data, it was spotted that the neural networks with memory gates (LSTM and GRU) were the best at the time of analyzing data with underlying trends (for instance in case of the data set of Yahoo).

(DerisStiawan 2016) there are a lot of challenges in detecting anomaly in IoT environments mainly due to the huge number of devices that are interconnected, and their data is the cloud. Moreover, one of the objectives of anonymity in the network is to prevent the traffic of data from the devices. The main aim of the proposed system is to address the problems on how to preserve the device failures monitoring the network and how to find out the comprehensive feature for anomaly detection of IoT communication before any problems arise. This system mainly focuses on anomaly detection in IoT communication based on their profiles comparing the state when it is normal and when the data traffic is anomalous.

(Nomm and Bahsi 2018)has proposed a system to detect the anomaly of IoT botnets which mainly emphasis on feature selection. There is a significant number of increments of vulnerable devices with the rapid growth of IoT which has become a part of the botnet. This malicious traffic in the IoT should be treated well by finding out its countermeasures as early as possible. There are various attacks in IoT network and various studies have been carried out to detect the anomaly-based attack however there is a smaller number of researches where there can be a reduction in the models of learning which are induced for networks of IoT. This paper has mainly focused on the part that it is feasible to stimulate high accurate unsupervised learning models through reduced feature sizes of the set that will enable to lessen the required computational resources.

There can be many forms of defining the anomalies in the system, (Guha, Mishra et al. 2016) anomaly detection in not well understood even in the simpler context of static batch processing. In the paper presented robust random cut data structure is investigated which can be used as the synopsis of the input streams. Here a clear report showing how the sketch can be updated dynamically is presented. Many mathematical formulas are used and implemented to find out the anomalies that are present in the wide range of data streams. Generally, there are many contradictions about randomization, randomization is highly valuable in supervising learning and this is a powerful tool one can use (Breiman 2001). A combination of tree predictors is termed as random Forest and each of the trees depends on the random vector that is sampled independently no matter how the trees are distributed.

In a paper written on random forest, (Amit and Geman 1997)it defines a vast range of geometric characters and uses the random selection to search these to get the finest split at individual node. This paper is very significant in my opinion as it talks about the approach to select the features that are informative and using the inductive learning algorithm builds a tree classifier at the same moment. An estimation is offered to the full posterior by each tree where the characters depends totally on the branch that it is transferred to. Due to the large number and unique features of the queries, the standard decision tree cannot be used in this as the decision tree is based on the fixed-length feature vector and this is not feasible. This approach is used for the classification of handwritten digits and synthetic linear deformations of around 3 hundred Latex symbols.

Similarly, (Tan, Ting et al. 2011) published a paper introducing Streaming Half-Space-Trees(HS-Trees) which is used for data streaming and is a fast one-class anomaly detector. This method is used only when the data are normal in the training phase and in the cases when the data are rare it works much well. An ensemble of random HS-Trees is modeled, and the structure of the tree does not need any data to be constructed. This is the main reason why this method is highly efficient in performance as it does not need any model structuring in the case of adapting.to evolving data streams. According to the analysis done in the paper Streaming HS-Trees possess constant remunerated time difficulty and a constant memory requirement. The result of the experiment made also showed the detection performance of Streaming HS-Trees has nothing to do with the setting of the parameter.

Reviewing a paper (HongzhiWang 2019)“ Progress in outliner Detection Techniques: A survey “in which researcher provide thorough review of development of outlier detection system from early 2000 to present. This paper provides the fundamental concept of already existed outlier detection and grouping them based on techniques such as distance-clustering, density-based, ensemble-based and learning-based methods. Further, their advantages, drawbacks, and challenges with their solution of each category are described. Thus, this will certainly be broadened new researchers' knowledge in-depth with the clear and distinguishable path for the upcoming future of outlier detection methods. This paper proposed that despite progress in outlier detection there are still a lot of issues to be addressed in terms of outlier detection-based approaches.

The paper that was reviewed “Checking is Believing: Event-Aware Program Anomaly Detection in Cyber-Physical Systems “ (Long Cheng 2018)which states that cyber-physical system are vulnerable to attacks such as control-oriented attacks as code injection or code-reuse attacks and data-oriented attacks as non-control data attacks. For this researcher proposed Orpheus which is a new security methodology for defending against these attacks by enforcing runtime execution semantics checking. Also, a new program eFSA which stands as event-aware finite-state automation takes advantage of the event-driven nature of cyber-physical program and constitutes event dependent state transition. In this paper, researchers conducted a case study to successfully detect different runtime attacks.

Later another paper that was reviewed was (Paulo M. Mafra 2010)" Octopus-IIDS: An Anomaly Based Intelligent Intrusion Detection System". This paper states that IDS (Intrusion Detection System) used to identify unwanted attempts to gain access to the internal network through the internet. According to the researchers of this paper, conventional IDS are not capable of distinguishing legitimate access and attacks. Thus, this paper proposed two main artificial intelligence techniques namely Kohonen neural network (KNN) and support vector Machine (SVM) to detect anomalies. This Paper present IDS model named Octopus IIDS based on the performance of the traffic of network within analysis and categorization of messages and make use of two above mentioned AI techniques.

Another paper reviewed was Applying Intelligent Agents for Anomaly Detection of Network Traffic in Internet of Things Networks (Kotenko 2018)which states that in order to detect anomaly in IoT system algorithm called Pseudo-gradient anomaly detection and fuzzy logical interference can be used with various variant such as high-performance computers, embedded devices and system on chip which works on real-time. Also, this research relates to the control of smart agents in the security monitoring system of IoT networks.

Another paper reviewed was Context-aware intrusion detection for building automation systems. According to this paper(Pan, Hariri et al. 2019), some IoT devices which are used in Smart home or our daily life are more vulnerable to attack. This paper uses a three-layer intrusion detection system in a supervised approach to detect attacks such as Man in the middle attack, Reconnaissance attack, replay attack and DDOS (Distributed denial of attack). This system consists of three main functions such as classification of profile and type of IoT devices, identification of malicious packets into system and classification of type of attack. This system is evaluated on the application of smart home IoT devices.

According to paper (Fuller, R. et al. 2016)IoT Applications in an Adaptive Intelligent System with Responsive Anomaly Detection in which researches created the complex system with use and reuse of adaptive agents in intelligent system for automation, control and communication. This system reacts with push notification when anomaly such as halting and starting of processes, notifying humans and communication with 3rd party systems detection happens.

Another paper reviewed was Anomaly Detection for Smart Home Based on User Behavior. This paper (Yamauchi, Ohsita et al. 2019)propose a method to detect anomaly in home appliances that uses internet such as refrigerator and conditioners based on human behavior. Researcher model method by learning sequences of each event of IoT devices and detects attacks by comparing those predefined sequences with current operation sequences. For this method user collected data from four users.

Another paper reviewed was Anomaly detection and privacy preservation in the Cloud-Centric Internet of Things. According to this paper(Butun, Kantarci et al. 2015), data collected by IoT devices are prone to different attacks and those data are stored and transferred to cloud which increases this risk even more. To address this issue researcher introduces some features and applications of IoT and cloud and discuss security and confidentiality dangers to personal info and then eventually provide solution from anomaly detection perspective. This paper also delivers a clear path for forthcoming studies on anomaly detection for IoT and cloud on the sensor system.

Another paper reviewed was Machine Learning Techniques for Anomaly Detection: An Overview. In this paper (Omar, Ngadi et al. 2013) Detection of intrusion has gained comprehensive consideration and has become a productive arena for several researchers and is still the subject of extensive concern for researchers. The intrusion detection community faces difficult problems even after many years of research. During the process of recognizing unknown patterns of attack, decreasing the large number of false alerts remains an unresolved issue. Recently, however, several study findings have shown that this issue has potential alternatives. Detection of anomalies is a main problem of detection of intrusion in which disturbances of normal behaviour indicate the existence of expected or unintended caused assaults, faults, defects and others. This article provides an overview of the directions of studies for implementing monitored and unsupervised techniques to manage anomaly detection issue. The quoted references will address the main theoretical problems, guiding the scientist in interesting directions for studies.

(Estakhri 2002)A quick memory access (read or write) technique and circuit of the information is revealed to and from a memory array. Architecture wise, there are at least two address latches and two-page registers in the memory array control circuit. The first address latch includes a first information address and a second information address latch. First decode the first information address and send the respective information from the memory array to the memory array for access (read or write). When the first data address data is accessed, a second data address will start the decoding process. When the data of the first data address has been accessed, the second data address is ready for the memory array. Thus, there can be continuous fetching from or writing to the memory array. In the preferred embodiment, there are two-page registers. In a read operation, the data read from the first data address is transferred to a first page register. When the data of the second data address is being accessed, the data in the first page register is transferred to a second page register. When the operation to read the data from the second data address is completed, the data can be placed in the first register. The information in the register of the second page can be transmitted quickly to a latch and to a bus. In this way, room is always accessible for reading the information. Similarly, information is transmitted from the data bus to the register of the second page and then to the register of the first page. The information is entered into the memory array in the first register.

(Yildirim, Arslan et al. 2015) There are a few factors affecting data move throughput in start to finish data moves, for example, arrange highlights (e.g., organize transmission capacity, round-trip-time, foundation traffic) ; end-framework highlights (e.g., NIC limit, number of CPU centers and their clock rate, number of circle drives and I/O rate) ; and data set highlights (e.g., normal record size, dataset size, and so on.). Number of PC plates and informational collection highlights (e.g., normal record size, dataset size, document size dispersion). Enormous information move streamlining crosswise over between cloud and intra-cloud systems is a troublesome assignment that necessities joint thought of every one of these parameters. This activity of advancement turns out to be much progressively troublesome when moving informational collections comprising of heterogeneous record sizes (for example mixed huge records and minor documents). In this paper the creator gives models and rules to setting the best qualities for these parameters and present two unique calculations for exchange enhancement utilizing the created models. The examinations performed on fast systems administration and cloud proving grounds demonstrate that in many cases, our calculations beat the most widely recognized data move instruments, for example, Globus Online and UDT.

In this paper (Rahm and Do 2000) the author has classify data quality issues that data cleaning addresses and provide an overview of the primary solutions to solutions. Especially when incorporating heterogeneous information sources, data cleaning is needed and should be resolved in conjunction with schematic data transformations. Data cleaning is an important component of the so-called ETL process in data warehouses. The author also addresses present information cleaning tool assistance.

A survey paper (Fatima, Nazir et al. 2017)introduces issues related to information cleaning and the methods presently used for pre-processing. Determining which pre-processing method is best in which scenario to enhance Data Warehouse efficiency is the primary objective of this article. Many data cleaning methods were evaluated using certain assessment characteristics and tested on various data sets. For conclusive outcomes, data quality instruments such as YALE, ALTERYX, and WEKA were used to prepare information in data warehouse and guarantee that only cleaned information populates the warehouse, thereby improving the warehouse's usability. In many future operations such as cleaning, standardization, correction, matching, and conversion, this paper results can be helpful. This study can contribute to data auditing and data pattern detection.

We have reviewed a paper(Karami and Guerrero-Zapata 2015) that present a new fuzzy anomaly detection scheme in this article that operates in two stages. The author has proposed a hybridization of the Particle Swarm Optimization (PSO) and K-means algorithm in the first phase–the training phase–with two immediate cost purposes as well as well-separated clusters and native optimization to determine the optimum amount of clusters. When the ideal positioning of centroids and items clusters is established, the second stage begins. In this stage–the detection phase–a blurred strategy to identify anomalies in fresh surveillance information by combining two distance-based techniques is used as classification and outlier. Experimental findings show that the suggested algorithm can reach the ideal number of clusters, fully distributed clusters, boost the elevated detection rate and at the same moment reduce the false positive rate associated to some other famous clustering algorithms.

# Research Problem

There are quite a big number of networks in IoT that are connected and the data from the devices are tremendous. There is a high chance of failure and it is difficult to predict when the path will breakdown and causes the error. The traditional approaches cannot be used as the system is massive. This will not only hinder the privacy of personal data but also affect all the aspects of the environment like industrial aspects, health aspects, etc. For instance, if there is an anomaly in the data of an aircraft that delivers 2.5 beta data per flight and there is hundreds of flight per day; it will cause a problem which can be harmful.

Devices send data every second and are connected to mountains of other devices thus, a data platform for processing data is required that will deal alongside billions of such incoming events every day. Managing this very extent of data is itself a daunting task plus to store the data it is tougher and not actionable. To get illegal insights, the data which is collected has to be analyzed.

Based on the literature review that I have done, a kind of task that can be challenged with analysis of data is called anomaly detection. The problem of finding patterns in the dataset that do not imitate normal behavior is anomaly detection. Anomaly detection can help to overcome the limitation of statistical approaches.

Many researchers have done various researches on Anomaly detection in IoT devices, some of them were focusing on intrusion detection some had machine learning and deep learning approaches. Some of the anomaly detection is cognitive anomaly detection in IoT devices and some of them have also reviewed the performance investigation of anomaly detection of distinctive IoT datasets by the use of cloud microservices. But only a few of them have proposed a paper on anomaly detection of IoT devices using the AWS platform.

# Research Methodology

As we are familiar with the fact that IOT has become really prominent and with the proliferation of IOT devices we have to look up for measures to connect them, collect and store device data. AWS IOT provides comprehensive and very profound functionality as it deals with the data in cloud which makes it even easier to build virtually any use cases underneath the eclectic range of IOT devices. Artificial Intelligence is the modern technology that is used in the AWS platform therefore AWS IOT can work in progress even without the internet connectivity. There are various reasons to select AWS IOT; AWS provides broad and deep IOT services, multilayered security is provided by the AWS IOT, artificial intelligence is added to IOT for more enhancement in technology, this is built on a scalable, secure and proven cloud infrastructure. There are range of companies that use AWS IOT today, some of them are Vizio, LG, Robot, NASA JPL, PHILIPS etc. As discussed earlier among the various use cases there are certain use cases that are built in AWS IOT such as Industrial use cases like predictive quality, asset condition monitoring and predictive maintenance; Connected Homes and Commercial.

Industrial IOT (IIOT) is a new form of technology which acts as a bridge between the ancient equipment that are used in the industries from a long ago, its infrastructure and modern technology like machine learning, cloud computing and big data. IIOT is mainly applied in the areas where continuous monitoring and predictive quality and maintenance is in the top priority. Consumers remotely monitor the operations of these equipment which are used in the industries via the new technology from anywhere. The Industrial IOT use cases comprises predictive quality, asset condition monitoring and predictive maintenance.

In the first use case which is predictive quality, to enhance the quality of factory production, it abstracts actionable ideas from industrial information sources like production machinery, environmental circumstances and human observations. Industrial companies can use AWS IoT to construct predictive models of quality which assist them create healthier products. Products which are of superior quality increase customer happiness and decrease product recalls

Another use case is asset condition monitoring, in this use case asset performance is determined taking into consideration the condition of equipment and machineries. Using the AWS IOT one can store all the data for example temperature, pressure, vibration, and error message that will point out whether the machineries are performing ideally. We will be able to boost the use of assets and make full use of the investment with enhanced visibility.

There is another use case in the industrial IOT which is called predictive maintenance where it captures the state of industrial machinery to detect future breakdowns before impacting manufacturing resulting in increased equipment lifespan, employee security, and supply chain optimization. We can monitor and infer equipment status, safety and efficiency continually with AWS IoT to detect problems.

Another awesome technology developed by AWS IOT is Connected Homes. Connected Homes puts our home in schedule. A closely linked home brings together equipment and services for an integrated, independent experience that enhances the life of a consumer. Connected home involvements incorporate everything from lights that are voice-controlled, gesture control TV, home-cleaning robots, machine-learning security cameras, Alexa controller, smart blinds and trouble-shooting Wi-Fi routers. Due to lower expenses and increased connectivity choices, these smart home devices help in decreasing the costs and mounting the vast choices for connectivity.

New features are initialized and abilities to smart devices such as interconnectivity, analytical maintenance, security, offline communication, consumer insight analytics, and machine learning, IoT enables the connected home. Home automation, home security and monitoring, and home networking, each of these abilities plays a diverse role for connecting home use cases.

There are different instances of use cases related to connected households, one of which is home automation. Home automation involves a broad variety of linked IoT systems such as a washing machine, oven, fridge, dryer, TV, coffee maker, or light bulb. AWS in IoT benefits enhance the customer experience by allowing a device to quickly, reliably and easily connect and execute an intervention. A client who purchases an intelligent coffee maker, for instance, will want a simple setup using a smartphone app. Additionally, phones can also profit from a seamless customer experience using voice services such as Alexa.

The other is home security and tracking where security related devices such as linked door locks, doorbells, secured cameras, and precise surveillance systems such as water over flow and leak detectors, pressure and energy management systems, and connected thermostats are included in the home security and surveillance segment. Devices constructed with AWS IoT can routine machine learning to identify threats automatically, perform action, and send homeowners alerts. AWS IoT allows low latency devices to operate and calculate information locally without internet connectivity.

In linked households, home network management is also a situation of use. Network operators are looking for fresh ways to assist clients find, troubleshoot and solve home network problems, including Wi-Fi and cable TV connectivity, rapidly and easily. Set-top boxes enabled by AWS IoT can automatically log and proactively send network diagnostics to the customer service center or enable clients to monitor and troubleshoot their network health via a mobile app.

Another very good technology developed is in the commercial field. AWS IoT clients are building traffic surveillance, public safety and health monitoring business applications.

This research methodology that we will be implementing in the Thesis is quantitative. Here we will be using various mathematical interpretations of the data. Similarly, various statistical interpretation will also take place and we will be using computational technique to detect the anomaly in IoT devices using an algorithm. The findings and results will be calculated on the numerical form and the findings will be particularly only for this research and the other research done based on this research will only be hypothetical.

Generally, anomaly is an observation that diverges from otherwise well-structured data. Real-time anomaly detection in IoT devices is very important because faster we get to know about them the quicker, we will be able to find out how to rectify the problem. Below is a figure that shows the anomaly for the temperature to time graph.

Anomaly

Anomaly

Y-temperature

X-timestamp

Figure 1 Graph showing anomaly detection

Here, two data points manifest from the out layers of data. These data that deviates from the ordinary data and looks unusual are anomaly data. Detecting those data is the most important concern for most of the industries and big technology companies these days. One of the most common techniques is using the thresholds. We can have predefined limits that can set to modify something that is abnormal or if the data exceeds the threshold limit.

Threshold

Threshold

Figure 2 Graph showing threshold

However, there can be some weaknesses with the threshold like it is difficult to know what should the threshold limit be set to, threshold does not define the prediction of whether the error will occur or not, it does not adapt to new states or dynamic of the data.

Machine learning is a subset of artificial intelligence (AI) which provides machines the capability to acquire automatically and recover from experience exclusive of being openly programmed. Simply, machine learning is the process in which humans can train the machine on the basis of past data and perform the task as humans but in a faster and efficient manner. It is a part of artificial intelligence in which systems can learn from the data, patterns of data and take the decision with very less involvement with the human. Machine learning is starting to redefine the way we live and we have to understand how it works and why this matters. Machine learning enables to automatically shape the model that is necessary based on the data that is given to it. This data will be given in the training phase and is the training data. Machine learning can be used for detecting the anomalies in the data. As the difficulty and number of various attacks are mounting, machine learning techniques acknowledges construction and upholding anomaly detection system (ADS) through less human interference, this looks like the only practical method through which we will be able to realise the next generation of anomaly detection in IOT devices.

There are various machine learning techniques for detecting the anomaly such as supervised, unsupervised and reinforcement anomaly detection. Supervised methods which are also called classification methods needs a categorized training set of data that will include both normal and anomalous data sample in order to build a predictive model. As supervised means to oversee and direct a certain activity and make sure it is done correctly so in this method the machine learns under guidance. Basically, here the machine learns as we feed the data and we also have to tell the machine what the input is and how the output should look like. Supervised method provides better rate of detection of anomaly in comparison to the unsupervised and semi-supervised data theoretically since in supervised method there is access to more information. However, in a practical approach the method does not appears to be accurate due to some technical issues. There is lack of accuracy in this method practically. Moreover, there are many false alarm rates in the training sets which is why it is a daunting task to obtain the accuracy. This technique usually solves problems of regression and classification. Generally, predicting a label or class is classification and regression is predicting a continuous quantity. There are various supervised algorithms, among which are Supervised Neural Networks, k-Nearest Neighbors, Decision Tree, Bayesian Networks and Support Vector Machines.

Bayesian Network(BN) is defined by (Heckerman 1995) as “A model which encodes probabilistic relationships between variables interest is Bayesian Network (BN). It is commonly used in conjunction with numerical schemes for intrusion detection. It has various advantages like the capability of programming interdependencies among variables and of predicting events, and the capability to combine together prior information and data.”

K-Nearest Neighbour(k-NN) is a supervised algorithm which is modest and conventional nonparametric techniques for the classification of the samples. In this method the distance between numerous points on the input vectors is approximately calculated afterwards it assigns the points which are unlabelled to the class of its K-nearest neighbours.

Decision Tree(DT) is defined by (Quinlan 2014) as “powerful and common tools for classification and prediction. There are three different components in the decision tree that are: nodes, arcs and leaves. A decision tree is used at a data point through initiating at the root of the tree and moving this till a leaf node is reached. The classification of the data point is provided by the leaf. ID3 and C4.5 which is established by Quinlan are the most implemented Decision Tree.”

In NNs, the behaviour of different users and daemons are predicted. NNs use the rule-based approach to address various problems that are encountered when they are appropriately designed and implemented. The main benefit of this algorithm is their tolerance to imprecise data and indeterminate information, then their capability to determine solutions from data exclusive of having preceding knowledge of the regularities in the data.

Support vector machine is same as other machine learning algorithm in which a randomly selected training set is classified in advance. (Vapnik and Vapnik 1998) proposed the support vector machine in which it firstly plots the contribution vector into a higher-dimensional feature space then after that it gives the output in high dimensional feature space separating the hyper-plane.

Unsupervised anomaly detection technique is a machine learning technique in which data provided to it is not labelled and the machine has to learn without any supervision provided by human. Here the machine has to find out data set, hidden patterns and trends in the data. This kind of technique is used to solve the association and clustering problems. Association patterns involves discovering patterns and data, cooccurrences etc. A very good example of association is bread and jam, the people who buy bread have a high probability of buying jam as well. In association it is usually things that occur together or similar things that go together. Apart from association problems unsupervised techniques also deals with clustering and anomaly mainly. Its main aim is to discover underlying patterns as only input vector is given to this technique. The approach for unsupervised is to understand patterns and discover output, unlabelled input is given to it then it explores the patterns and trends of the data until it reaches the expected output. Here there is no feedback mechanism as the machine is not aware about the output during the training phase. There are various unsupervised technique that are used to solve various problems like clustering technique, K-means, Fuzzy C- Means (FCM) etc.

In the clustering technique the data observed is grouped into clusters which will be in accordance to the similarities or distance measure. This clustering techniques is based mainly on two approaches. The first approach is the one in which the unlabelled data is used to do the anomaly detection in which the data can be both normal and attack traffic. Another approach is the one in which the model is trained with only normal data creating the profile of normal activity.

K- Means algorithm is an ancient algorithm in which data is divided into k clusters and the data in this clusters are similar for sure but the data in different clusters have very low correspondences.

Here in our thesis unsupervised machine learning algorithm is used. Hence the input data is only given, and it is not labelled. The output can be found out by examining the hidden patterns and trends in the data and the output will also be unlabelled.

Data cleaning which is also called data cleansing is the process that deals with detecting and correcting(or removing) corrupt, inconsistencies and error data from data set or database which will improve the quality of data mainly by identifying irrelevant parts of the data and updating the coarse data.

Data cleaning is done using various data cleaning tool. Data cleaning will fix common data quality issues using a variety of parameters, allow us to replace nulls, modify case and remove unwanted characters across any fields we select be its numeric and string. Generally, data cleaning has various phases. The first phase is data analysis; in this phase deep analysis of data is done that makes it clear which kinds of errors and irregularities are to be removed. Not only manual analysis, an analysis program is used to detect the problems in quality of data and its properties. On the basis of the degree of the errors and dirtiness of data, various data transformation and cleaning steps are to be implemented. The transformed data after the transformation process should be verified and evaluated. A large number of iterations of the steps used earlier i.e. analysis, design and verification is needed. Verification is followed by transformation in which the execution of the transformation steps is executed. The next step is backflow of cleaned data in which the dirty data is removed, and the cleaned data is changed in the original sources so that we can escape redoing the cleaning work for forthcoming data extractions. There are various data cleaning tool which are available in the market. The ones that are used frequently are data analysis and reengineering tools, specialized cleaning tools and ETL tools.

Data analysis and reengineering tools can be shared into two parts that are data profiling and data mining tool according to (Rahm and Do 2000). A healthy and good feature of the data is delivered by the data profiling tool either by illuminating critical parts of collected data. Numeric feature of the data is also analysed with their values under particular range also it will determine dependencies and repetition of data. Additionally, they perform different process like summarization and aggregations which will provide related information. According to all this information that are gathered, decision is made. Migration Architect (Evoke Software) is the most popular data profiling tool used commercially. It determines the following actual metadata for each attribute: information type, duration, cardinality, separate and percentage values, smallest and supreme values, lost values, and uniqueness. Migration Architecture also provides for the development of the data migration target scheme. Alteryx is one of the most popular data analysis tools used commercially. It provides a general platform in order to make decision deliberately plus it also take into consideration and highlights the irregularities among the data. Data mining instruments such as WIZRULE (WizSoft) and DATAMININGSUITE (Information Discovery), infer attribute relationships and their values, and compute a confidence level showing the number of qualifying rows. Reengineering instruments for information, such as INTEGRITY (Vality), use patterns and regulations that have been found to define and achieve cleaning transformations, i.e. reengineer heritage data.

Specialized cleaning tool is another data cleaning tool that deals with one specific domain, mainly the name and address data, or concentrated on identical elimination. Special domain cleaning and duplicate elimination are the specialized cleaning tool which are included in specialized cleaning tool.

There are a range of exposed source and utensils for ETL processes which are available in the market. Big companies like Microsoft and Amazon has launched these tools; “EXTRACT”,” DATA TRANSFORMATION SERVICE” provided by Microsoft and “WAREHOUSE ADMINISTRATION” are some examples of these tools. There is a built-in DBMS which will manage the internal as well as the external sources, the sources that will map the components, the codes of sources and schemas and many more. In the first place, OLTP (Online Transaction Processing) system is used to extract all the data’s, after this with the help of various tools data is cleaned by using mapping rule. A Graphical User Interface (GUI) is provided which will show all the data that makes it easy for the users to interface. In ETL process not only cleaning of data is carried out but also it integrates API’s of cleaning tools in it which will provide better and efficient results.

A big number of commercial tools provide extensive assistance for the ETL method for cleaning the data mainly for data warehouses. AWS also uses the extract, transform and load (ETL) algorithm. AWS has its own AWS data cleaning tool which is called AWS Glue. AWS Glue enables us for analysis of data by preparing and loading the data which is fully managed by extracting, transforming and loading service. The extract transforms and load (ETL) is a fully manageable service and ETL job can be easily run with some few clicks in the AWS Management Console. AWS Glue can be simply point to the data which is stored in AWS and this will determine and store the related metadata in AWS Glue Data Catalog. The data can be searched, queried and is available as soon as the data is catalogued. A lot of benefits are there when we use AWS Glue such as less hassle, cost effective, more power. There are different steps of working AWS. In the first step, data catalog is built, AWS Glue will generate ETL code in Scala or Python for the extraction of data, change and loading the data into the target. This data can be edited, debugged and tested. At last, the ETL jobs will be scheduled and run.

After the cleaning of data, data should be pipelined this is because data is growing exponentially in a rapid basis. As there are a lot of data that are to be managed and this vigorous amount of data is difficult to be managed similarly it is also time consuming. Data pipeline makes it easy for users to integrate the data which is spread across various sources of data. It also transforms, analyse and store the data across the same location. AWS Data Pipeline is a web service which makes the movement and transformation of data automatic. AWS Data Pipeline is a web service which helps to reliably process and transfer data between various AWS calculate and storing services, in addition to this on-premises data sources, at specified intervals. The parameters of data transformation should be defined, and AWS Data Pipeline will look up and implement the logic that have been set up.

Let’s consider a simple example for AWS Data Pipeline. Using AWS Data Pipeline web server’s logs can be archived to Amazon Simple Storage Service (Amazon S3) buckets on daily basis then run the Amazon EMR cluster over these logs to generate report on the weekly basis.

AWS Data Pipeline has three different components that works together in order to manage the data. The first component is pipeline definition. The pipeline definition will specify the logic of the business in the database management system. These have different information like data nodes, activities, schedule, precondition, compute resources and actions. Data nodes are the names, location and format of the data sources like Amazon S3, Dynamo DB or it can be on-premise data store as well. Activities transforms the data like editing the data, moving from one source to other source, queries in the data. There is schedule for these activities. Also, there are preconditions that should be completed before the activities is schedule. There are various compute resources like AWS EC2 instances or EMR clusters. Finally, there are actions, these are the updates about the pipeline, for example send a notification, reply message, trigger an alarm etc. Another component is pipeline itself in which a defined activity is performed scheduling the pipeline and it routes different tasks generating EC2 instances in AWS. The pipeline definition of the first stage is uploaded, and the pipeline is then activated. The pipeline definition can also be edited anytime during the run time and it can be activated again so that it can be effective. Data source can be modified by deactivating the pipeline and after the modification is done the pipeline can be activated again. The pipeline can be deleted when the task is done. The last component is the task runner polls. In this component, polling for the task is carried out and then those tasks are performed accordingly. According to the resources that are created in the pipeline definition task runner is installed and automatically run. Task Runner application can be customised in accordance to the need or else a default AWS Data Pipeline is offered by AWS.

The AWS Data Pipeline firstly schedules the task then the task runner polls for these tasks and receive certain task to run. Afterward this task runner that will run the task continuously reports for these tasks progress as it works on the particular task. Similarly, task runner also reports about the task when it finishes the task. And if the whole process is completed and the task is successful finally the task ends however if the task is not succussed again it is tried, and the iteration process continues until the whole task is completely ended.

AWS Pipeline has input data stores which can possibly be AWS S3 or Dynamo DB tables or on-premise data stores. These data from these input data source can be passed into the Data Pipeline to carry out compute resources. In the pipeline these data are processed, analysed, transformed according to the wish then the result is pushed into the output data stores. These output data stores also have options like these can be Amazon S3 buckets, Dynamo DB table or a red shift.

AWS Data Pipeline provides a range of benefits. Here, it provides a drag and drop console mechanism through which we can easily and quickly create a data pipeline. Also, it is built on distributed, reliable infrastructure so if there is any fault while creating the data pipeline, it will automatically redraw the data pipeline. It also helps in scheduling, dependency tracking and error handling. With the help of AWS Data Pipeline work can be dispatched through single machine or many machines in a serial or a parallel manner. AWS Data Pipeline is not so expensive to use as the bill occurs monthly which is low in rates. AWS Data Pipeline provides a full control over the computational resources creating data pipeline such as EC2 instances or an EMR cluster.

There are many ways through which AWS Data Pipeline can be accessed. Pipeline can be created, access and manage the pipeline with various interfaces like AWS Management Console, AWS Command Line Interface (AWS CLI), AWS SDKs and Query API.

In the development of detecting anomalies in IoT devices, there are various ways of doing so. And as the threshold the one mentioned above it is daunting task to be familiar with the limit of threshold in certain cases. This is the reason why we choose Amazon Web Service as the technique of detecting anomaly. We have selected the AWS platform to detect the anomaly that arises in IoT devices. AWS enables a wide range of platforms and provides detail of how one can look forward to the sudden unusual behavior in the data of IoT devices.

The AWS service we use is Amazon Sage Maker, Amazon Sage Maker is the end to end machine learning platform and a smartly achieved service which makes it easy for developers and data scientist to quickly build machine and execute the pipelines into production. It supports all framework and has built-in high-performance machine learning algorithms that can be used out of the box to solve various problems.

There were many algorithms under Amazon Sage Maker. Among which, the Random cut forest is the one that we use in this research. “Random Cut Forest is the newest built-in algorithm that are made available for Amazon Sage Maker, a fully-managed platform which enables developers and data scientists to rapidly and effortlessly shape, train and organize machine learning models at any scale.” [www.aws.amazon.com]. For simple understanding, let us see the figure below.

Y-temperature

X-timestamp

Anomaly

Figure 3 Graph showing anomaly in RCF

Here let’s assume that we have various data points in a two-dimensional space. The out-layer data points in orange are the anomalies in the system. The algorithm associate’s anomaly score with each point in the data set based on how far it is from the circle. There are different score values for the data set, the low score point indicates that it is inside the circle whereas the high score values indicate that the dataset is out layered. The definition of low and high depends on the application itself. The RCF is used to detect the data points that deviate from the well-structured data that is in patterned and this is an unsupervised machine learning algorithm. This algorithm helps to detect anomalies in crucial cases like monitoring the health of emergency room patients, detecting the fraud moves and overvalued priced products marked by the retailer. This algorithm delivers high performance although it deals with massive data sets on large-scale.

First of all, this algorithm operates by acquiring a random sample of the training mode data. A sampling technique called reservoir sampling is used in the algorithm as the data can be very large that it cannot fit on a single machine. Sub-samples are then well distributed to each component tree in the algorithm. To randomly subdivide the bounding boxes, a binary tree is built to arrange each sample until each leaf embodies a bounding box containing a single data point. The anomaly score assigned to an input data point is inversely proportional to its forest-wide average depth.

The primary concept behind the RCF algorithm is to build a trees forest where a partition of a sample of training data is used to obtain each tree. A random sample of input information, for instance, is first identified. Then the random sample is divided by the number of trees in the forest. Such a partition is provided to each tree and organizes that sub-set of points into a k-d tree. The anomaly score allocated by the tree to a data point is described as the anticipated tree complexity shift resulting in adding that point to the tree; which is in proportion inversely proportional to the resulting point depth in the tree.

The random cutting forest assigns an anomaly score by calculating each component tree's median score and scaling the sample size outcome. On the basis of reference (Guha, Mishra et al. 2016) the RCF algorithm is described.

A random sample of the training data is the first phase in the RCF algorithm. Suppose we want a sample of K size from N complete information points in specific. If the training data is small enough, the entire dataset can be used, and we could randomly draw K elements from this set. The training data, however, is often too big to fit all at once, and this strategy is not viable. In its place, a technique which is called reservoir sampling is used.

An extensively acknowledged basis for estimating large data sets which outstrip accessible memory of the computer is random sampling. The total size of data is infinite when it comes in stream and typically only one pass through the data is feasible. A method of keeping a secured size random sample from the data that is streaming is reservoir sampling. In the past there were reservoir schemes introduced which are random sampling without replacement, in the sample no repetition is allowed. In this paper a new method to replace reservoir sampling is introduced. Firstly, we show that proposed method certainly maintains a random sample by replacement at any given time. Then we present a advanced version that knowingly speeds up the overall sampling procedure.

An algorithm which is used for drawing the random samples from a set of data i.e. where the elements can be observed individually at a time or in group is called Reservoir Sampling. In this sampling, even if the value of N is not known earlier the sampling method still works. When there is only one sample that is requested, for example when K=1, the algorithm is as follows:

* The input dataset or data stream is
* The random sample initially is
* For particular observed sample
  + Here, we pick a uniform random number
  + If
  + Then we Set
* Finally, we return X

Reservoir Sample selects a random sample such that

And for all  . The algorithm is more complicated when K>1. In addition to this, a difference must be drawn between random and non-replacement sampling. Based on the algorithms outlined in RCF (Park, Ostrouchov et al. 2004) conducts an enhanced reservoir sampling without replacing the training data.

The next step in RCF is using the random data sample to build a random cut forest. First, the sample is divided into several partitions of equal size equivalent to the number of trees in the forest. Separately partition is then sent to a single tree. The tree systematizes its partitioning recursively into a binary tree through dividing the data domain hooked on boundary boxes.

# Research Significance

While there are tons of idea that is related with the anomaly detection and this is not a new topic plus it has been widely studied, however, very less attention has been given to select the technique used for anomaly detection using the AWS platform. Amazon practices various decentralized or dispersed IT foundation to make several IT resources obtainable which is on-demand. Amazon Web Service is a cloud computing platform that is launched by Amazon.

This will help big business and enterprise as the policy of AWS is pay as you go and also there are various platforms that we can use to identify the anomaly in IOT devices. This will give full security for all the IOT devices data that is used in our daily life. Among the platforms provided AWS Sage Maker is the one that we use under which Random Cut Forest algorithm is the one that we have implement.

There are only a few researchers that have presented the paper about the anomaly detection in IOT devices and the system I have proposed is new and convenient in every form. Random Cut Forest (RCF) algorithm is a new form of anomaly detecting algorithm that is used for anomaly detection even in the large-scale dataset. Here, we have researched and presented a report about RFC and the performance matrix that the algorithm delivers. Similarly, how this algorithm deviates from other algorithms for anomaly detection and the performance comparison of other algorithms and this algorithm.

# Conclusion

This paper mainly provides an anomaly detection technique for IoT devices using Random Cut Forest in the dataset. Random Cut Forest to detect the anomaly in IoT devices is a new approach in the online anomaly detection technique. Random Cut Forest operates under the AWS Sage Maker. This is only the proposal of the research; shortly we will be explaining the methodology of RCF to detect the anomaly in IoT devices in detail. Also, this research can be a lot more progressive and a lot of improvement can be made in anomaly detection contemplating numerous IoT datasets. The potential performance of RCF is measured and compared with other algorithms. Here, we will be calculating the efficiency of the algorithm that we have used which is RCF. Also, there will be a comparison in the efficiency and performance of others algorithm and the deployed algorithm; finding out which one is the best to be implemented. In the future, the system can be potentially improved, the efficiency of the algorithm to detect the anomaly can also be done by taking into consideration additional parameters in the current algorithm.

# References

Amit, Y. and D. Geman (1997). "Shape quantization and recognition with randomized trees." Neural computation **9**(7): 1545-1588.

Brady, S., et al. (2018). Analysis of Machine Learning Techniques for Anomaly Detection in the Internet of Things, IEEE**:** 1.

Breiman, L. (2001). "Random forests." Machine learning **45**(1): 5-32.

Butun, I., et al. (2015). Anomaly detection and privacy preservation in cloud-centric Internet of Things, Institute of Electrical and Electronics Engineers Inc.

DerisStiawan, M. Y. I., Reza Firsandaya Malik, SitiNurmaini, RahmatBudiarto (2016). "Anomaly Detection and Monitoring in Internet of Things Communication." 8th International Conference on Information Technology and Electrical Engineering (ICITEE), IEEE.

Estakhri, P. (2002). Data pipelining method and apparatus for memory control circuit, Google Patents.

Fatima, A., et al. (2017). "Data cleaning in data warehouse: a survey of data pre-processing techniques and tools." IJ Information Technology and Computer Science **3**: 50-61.

Fu, R., et al. (2011). An intrusion detection scheme based on anomaly mining in internet of things.

Fuller, T., et al. (2016). IoT applications in an adaptive intelligent system with responsive anomaly detection, IEEE**:** 754.

Guha, S., et al. (2016). Robust random cut forest based anomaly detection on streams. International conference on machine learning.

Heckerman, D. (1995). "A tutorial on learning with bayesian networks. Microsoft Research."

HongzhiWang, M. J. B. a. M. H. (2019). "Progress in Outlier Detection Techniques: A Survey." IEEE Access, Access, IEEE: 107964.

Ibrahim Alrashdi, A. A., EsamAloufi, RaedAlharthi,MohamedZohdy (2019). "AD-IoT: Anomaly Detection of IoT Cyberattacks Smart City Using Machine Leaming 1n." IEEE Access, Access, IEEE.

JunPing Wang, S. (2014). "An Online Anomaly Learning and Forecasting Model for Large-Scale Service of Internet of Thing." IEEE Access, Access, IEEE.

Karami, A. and M. Guerrero-Zapata (2015). "A fuzzy anomaly detection system based on hybrid PSO-Kmeans algorithm in content-centric networks." Neurocomputing **149**: 1253-1269.

Kotenko, I. S., Igor Ageev, Sergey (2018). Applying Intelligent Agents for Anomaly Detection of Network Traffic in Internet of Things Networks, IEEE**:** 123.

Long Cheng, M., Ke Tian, Danfeng (Daphne) Yao, Senior Member, Lui Sha, Fellow, and Raheem A. Beyah (2018). "Checking is Believing: Event-Aware Program Anomaly Detection in Cyber-Physical Systems." IEEE Access, Access, IEEE.

MarwaKeshk, E. S., NourMoustafa, Jiankun Hu and Ibrahim Khalil (2018). "An Integrated Framework for Privacy-Preserving based Anomaly Detection for Cyber-Physical Systems."

Nomm, S. and H. Bahsi (2018). Unsupervised Anomaly Based Botnet Detection in IoT Networks, IEEE**:** 1048.

Omar, S., et al. (2013). "Machine learning techniques for anomaly detection: an overview." International Journal of Computer Applications **79**(2).

Pan, Z., et al. (2019). "Context aware intrusion detection for building automation systems." Computers & Security **85**: 181-201.

Park, B.-H., et al. (2004). Reservoir-based random sampling with replacement from data stream. Proceedings of the 2004 SIAM International Conference on Data Mining, SIAM.

Parth Bhatt, A. M. (2018). "HADS: Hybrid Anomaly Detection System for IoT Environments." IEEE Access, Access, IEEE.

Paulo M. Mafra, V. M., Joni da Silva Fraga (2010). "Octopus-IIDS: An Anomaly Based Intelligent Intrusion Detection System." IEEE Access, Access, IEEE.

Quinlan, J. R. (2014). C4. 5: programs for machine learning, Elsevier.

Rahm, E. and H. H. Do (2000). "Data cleaning: Problems and current approaches." IEEE Data Eng. Bull. **23**(4): 3-13.

Tan, S. C., et al. (2011). Fast anomaly detection for streaming data. Twenty-Second International Joint Conference on Artificial Intelligence.

Vapnik, V. and V. Vapnik (1998). "Statistical learning theory Wiley." New York: 156-160.

Yamauchi, M., et al. (2019). Anomaly Detection for Smart Home Based on User Behavior, IEEE**:** 1.

Yildirim, E., et al. (2015). "Application-level optimization of big data transfers through pipelining, parallelism and concurrency." IEEE Transactions on Cloud Computing **4**(1): 63-75.

Yuhuai Peng, A. T., Jingjing Wu,Yuanguo Bi (2016). "Hierarchical Edge Computing: A Novel Multi-source Multi-dimensional Data Anomaly Detection Scheme for Industrial Internet of Things." IEEE Access, Access, IEEE.

# Appendix

**The Impacts of the 6G in the Future Connected Societies: A Comprehensive Survey**

Erika Maharjan

erika.maharjan @live.vu.edu.au, Victoria University

Ashkan Yousefi

Ashkan.Yousefi@vu.edu.au, Victoria University

Omid Ameri Sianaki

Omid.AmeriSianaki@vu.edu.au, Victoria University

**Abstract**

Technological revolution has started with the emergence of 5G, with 5G technology that has not even been implemented in many parts of the world, various areas of the globe still using 4G and even 3G. Technologists from various countries are already working to figure out 6G technology. 6G technology will have more advance features than 5G in terms of speed and latency, performance and use of artificial intelligence. This paper provides the comprehensive survey of 6G and a brief discussion about how it will be connected with the society. The paper provides an overview of primary drivers of 6G technology, set of service classes, discussion on some enabling technologies and finally open research problems with 6G technology. Also, the paper is a general overview of the future of 6G oriented IOT taking into consideration how 5G is deployed in IOT. A part of IOT which is IONT is described in brief and the paper also provides an overview of 5G oriented IONT. The paper basically gives the overall overview of IONT and the fields where it can be used in 5G technology. Mainly, it focuses on the challenges that arises in the areas where IONT is deployed and how the theories and general patterns of IONT affects the areas.

1. **Introduction**

There is an advancement of new mobile generation every decade since the 1980s, there are improvement in various features to each new generation than the previous in various ways. On the basis of observed trends, a new technology is being developed every 10 years plus their incremental versions originated in between. With the implementation of 5G which is not even developed properly in many parts of the globe, 6G technology has become the trending talk for technologist these days. New mobile generation originates from the imperfection and flaws that are present in the previous generation. While this demand for wireless capacity will continue to grow, the advent of the Internet of All (IoE) network, linking millions of people and billions of devices, is bringing about a revolutionary paradigm shift from yesteryear's rate-centric mobile broadband (eMBB) services to ultra-reliable, low-latency (URLLC). Even though, the 5G has been stated as the generation where IOE will be enabled completely but there are still some flaws that 6G needs to fulfil.

As 5G was marked to support IOE it is still debatable that it will be able to support the future coming societies which will eventually be smart and autonomous. In [1], there is a brief discussion for what is beyond 5G. In order to fully support IOE services like extended reality (XR) and connected autonomous system the wireless communication system should be able to deliver high reliability, data rates and low latency. To overcome all of these deficiencies of 5G in IOE, a new technology which is 6G is in the phase of development. The motive of this paper is to present a forward-looking idea of 6G system and how it can be enabled with the new technology that will be growing in the near future like artificial intelligence, implants, XR, computing, sensing and 3D mapping and so on. There will be improvements in terms of performance, data rates, latency, device densities and IOT connectivity, energy efficiency, spectral efficiency, more artificial and machine learning application. Optical networking in 6G will be more advance in terms of comparison with other generations. There is also an expectation of new dimensions like quantum communication and satellite integration in 6G. In, a future generation flexibility connected application is presented through a flexibility analysis model which is mainly for high throughput satellites.

1. **Paper Presentation**

As 5G promised to be a communication layer of IOE, we argue that the vision of 5G is still missing as there are implicit limitation of the system that are frequently exposing. Regardless to the fact that 5G is still-forthcoming wireless system in various part of the world, there seem to be a need to encourage the next generation 6G wireless system which can genuinely incorporate significant extensive application that will range from autonomous method to trendy existence and haptics. The paper has presented a comprehensive forthcoming principle of 6G. 6G is going to be better, faster and more powerful than ever before. Here, we discussed about the main operator of 6G system with regards to the implementation and technological trends that are growing day by day. Here we proposed an innovative ambience and environment that represents the target 6G functioning and its requirements. We then identify the enabling techniques for the 6G services implemented and outline a basic study agenda using these technologies. This paper is a call for all the researchers to research about 6G by presenting an open research platform that gives the basic idea of 6G.

Firstly, we discussed about the driving application behind 6G and their requirements. There are four novel and main application domains that helps to define the driving application of 6G. The first one is of the multisensory XR (Extended Reality) application in which 6G can be applied across VR, AR and MR spectrums. 5G has already gave powerful VR, AR system but there is still room for improvement as it is not been able to provide a complete immersive XR experience seizing every part of sensory inputs because of their incapability to provide extremely low latencies for data rate intensive to XR application. 6G must support all the requirements of XR services for instance quality of-physical-experience (QOPE)[2].The second one is connected robotics and autonomous systems (CRAS) which incorporates smart vehicles, drone delivery systems, smart cities, intelligent robotics, communication between cars and so on. Perhaps CRAS is an important use case requiring stringent requirements across the latency spectrum of speed-reliability which is a balance that is not yet available in 5G. [3]The third driving application is wireless brain-computer interactions (BCI), BCI are much more sensitive then XR as it involves physical insights. For the few years BCI have been narrowed only to healthcare scenarios where humans were able to control their prosthetic parts, EEG or use the computing devices via some brain implants. Though, there is going to be a new revolution coming soon relating to BCI and introduce new use-case scenarios that require 6G connectivity. Mind to mind communication, brain to internet communication and transferring memories are some of the examples of BCI where 6G will be operated however, BCI service need ultra-low latency, very high rates, high computing and high reliability and this will also need QOPE.[4] The last driving application will be one of the greatest innovative IOE technologies which is Blockchains and distributed ledger technologies (DLT). Blockchain is todays big talking points and experts see it as a technology which will influence all of our life and this will surely need low-latency, scalability, reliable connectivity and so on.

Driving Trends and performance metrics are also discussed which usually provides the goalmouth for 6G. 6G requires higher bit rates in comparison to 5G for driving application such as XR and BCI. 6G also requires more spectrum and higher reliability. Also, 6G system must deliver high spectral and energy efficiency (SEE) that is required to measure in bps/Hz/Joules. The evolution of this started with 2G which worked on bps and then 3G which supported bps/Hz which was followed by 4G (bps/Hz/ and 5G dealing with bps/Hz/Joules. 6G deals with volumetric rather than spatial bandwidth. 6G architecture will be driven to evolve if smart surfaces and environments will be used in the wireless communications system. Big and small data sets should be used across the 6G devices infrastructure to improve network functionality and deliver fresh services. This will give rise to new methods of machine learning and data analytics which will go across traditional big data. 6G should be able to adapt with networks that are self-organising which are simply known as self-organising network (SON) to self-sustaining network (SSN) which are mainly used applications like CRAS and DLT. All the previous five generation of cellular system had one common function which is wireless communication. 6G should be able to advance various functions which incorporates communications, computing, control, localization and sensing which are essential for the applications like CRAS, XR and DLT. Recently, there is a shift of using wearables from smart phones that were centralised to 4G and 5G. Application such as XR and BCI are catalyst to this trending technological shift. A majority of 6G use cases are driven by these devices. These driving trends will basically disrupt the novel targets to perform better according to the requirements in the next generation.



With the development of wireless system there will be evolution of new 6G service classes such as mobile broadband reliable low latency communication (MBRLLC) that is used in 6G in place of eMBB and URLLC which is used in 5G as these will no longer support applications like XR, CRAS or BCI. Energy efficiency, impact on reliability and rate are the factors which are main for MBRLLC, the reason for this is 6G devices are constantly decreasing in size and mounting in functionality. Another service class of 6G is massive URLLC, 6G has to scale up previous URLLC to a new massive URLLC (mURLLC) which will basically combine 5G URLLC with the previous mMTC. Alternatively, it is essential to have a principled and scalable structure that accounts for delay, reliability, packet size, architecture network, topology (across access, edge and core) and uncertain decision-making[5]. A new class of 6G service is introduced which is called human-centric service, which basically needs QOPE targets instead of raw rate-reliability-latency metrics. Another new class of 6G service is multi-purpose 3CLS and energy services.[6]

In order to allow the above-mentioned services and to assure their performance, a new disruptive technology must be combined into 6G. Technologies like frequency higher than 6 GHz from small cells to tiny cells should be enabled. 6G will operate in terahertz (THz) which will exploit frequency that is over sub-6 GHz.[7]Transceivers with integrated frequency bands can be enabled in 6G which will be able to liberate various frequencies across the microwave/mm wave/THz spectra that provides seamless connectivity in local as well as wide area levels. There has to communication with large intelligent surfaces. Edge AI is another brilliant technology that should be enabled. Integrated terrestrial, Airborne and Satellite Networks will have to enable. Energy transfer and harvesting is another very important technology that should be assisted. Beyond 6G is another handful technology that will mature along the same time of 6G and, hence, potentially play a role towards the end of the 6G standardization and research process. Transferring energy and harvesting could be enabled in 6G through which 6G will be the first generation of cellular system to provide energy along with 3CLS.

With all the trends that will be enabled in 6G and enabling technologies, research agendas and open problems in 6G are discussed. 6G systems needs to have high performance of 3D fundamentals taking into consideration rate-reliability-latency adjustments. There can be various open problems when exploitation of mm waves and THz. 6G should must be able to operate in 3D space communication. This can be an open research topic as this whole concept is totally new. The research can be carried out by planning, propagating and modelling the 3D environment. In [8] it is already shown that the 2D plan is subsequently different than the 3D planning. Communication with LIS is another good research but open research problem here will be optimized deployment of passive reflectors and meta surfaces to AI-powered operation of reconfigurable LIS. Many research problems for 6G can be derived from joint design of Artificial Intelligence such as analytics of very small data, using machine learning (ML), AI based SSN, ML algorithm to apply reliably. An important 6G research area is the design of QOPE metrics that combines physical considerations from human physiology or from a control system. In order to make CRAS work in 6G system, a communication and control co-design is required which is not available in 5G as the 5G operates in traditional radio-centric focus. As the specifics of wireless network is limited to cellular communication which makes it a very trendy topic for research. This idea of joining the communication and control has to be extended to make the 3CLS functions. RF and non- RF links design is an open research and the way how holographic radio is realised is another widely open area for research.

Table 2 Difference in requirements of 5G and 6G

|  |  |  |
| --- | --- | --- |
|  | 5G | 6G |
| Required rate | 1Gbps | 1Tbps |
| End-to-End Delay Requirements | 5ms | <1ms |
| Processing Delay | 100ns | 10ns |
| Frequency Bands | * Sub-6 GHz. * MmWave for fixed access. | * Sub-6 GHz. * MmWave for mobile access. * Exploration of THz bands (above 140 GHz). * Non-RF (e.g., optical, VLC, etc) |
| End-to-End Reliability Requirements | Five 9s | Seven 9s |
| Spectral and Energy Efficiency Gains with Respect to Today’s Networks | 10x in bps/Hz/m2 | 1000x in bps/Hz/m3 (volumetric) |
| Radio-Only Delay Requirements | 100 ns | 10 ns |

6G cellular system will be highly be enabled in IOT. According to Forbes data, until 2020 there will be 156k smart vehicles so we can imagine how IOT can be implemented in huge number and this will be enabled with 6G and the services. IOT has given rise to various wearables devices that we are using these days. For an advance use of the wearables IOT has a section called IONT. Internet of nano things (IONT) is connection between the Nano technology and the internet with high computational data fields. This paper mainly deals with the crucial properties for the IONT like the security issue, reliability source, the confidentiality and ability to work in the evolving big- 5G oriented patterns. The paper basically gave the overall overview of IONT and the fields where it can used in 5G technology. Mainly, it focuses on the challenges that arises in the areas where IONT is deployed and how the theories and general patterns of IONT affects the areas. Generally, the writer of this paper has discussed issues relating to techniques and intelligence of IONT, security issues and the measures that should be taken for easy access. Overview of anticipated and usual attacks furthermore to classifying the kinds of attackers. Possible communication technologies have been overviewed and categorized. Precise tools and methods of evaluation have also been stated.

This paper will be structured in mainly 12 sections. The first section provides the market op-opportunity that relies on the six V's in the big data project. The second section gives the architecture mutual to the IONT. Section 3 discusses important design factors in IONT. Applied intelligence techniques were stated and categorised in Section 4.

Section 5 discusses security needs in the IoNT theories and patterns which is IONT paradigm. In sixth and seventh section respectively, the physical layer debates similarly, the frequently used communication techniques are provided. The eight section outlines critical re-strictions affected by Nano-communications in the IoNT paradigm. Similarly, in Section 9, helpful assessment processes and benchmarking tools are overviewed. Section 10 provides an overview of classified attackers and safety assaults. In section 11 provides instructions for future studies and reviews some accessible research problems. At last, with section 12 the author concludes the survey article.

It is anticipated that the number of linked appliances will rise quickly in the coming years. Intelligent communication paradigms to ensure effective interaction between these devices, they're needed. These smart paradigms can efficiently process information of variable dimensions and complexities to fulfil the fast mounting big data project. We have to keep in mind that big data is characterized primarily by 6 dimensions frequently referred to as the 6 V's: Volume, Velocity, Variety, Veracity, Variability, Value. The increasing use of nanotechnology in the 6Vs is one of the most significant considerations controlling to global growth of the IONT market. Nevertheless, concerns about interoperability and the lack of realistic test beds can significantly degrade IONT's market success.

In section 2, there is a brief description of IONT architecture in 5G/ Big Data, the clear vision of security can be known if the architecture of IONT is understood. There are various devices that are connected to the network and communicates back, and forth which are called endpoints. Here in IONT it is called Nano nodes. There are Nano-routers that are better in terms of computational and communication capabilities in comparison to Nano-nodes. Interface devices are there which can handle both nanoscale and microscale devices. It acts as a gateway and interface to the internet. The IONT architecture can be easily customized in accordance to the customers need.

One of the main goals of 5th cellular network generation specified is to have omnipresent communication between anyone and anything at all times and anywhere. In this section, security, energy efficiency, and quality of service of IONT are considered which are very important aspects of the IONT theories and patterns. Energy harvesting is another factor that is used in IONT design. Capacity enhancement and channel bonding are the features of QOS in which the channel can be linked to an end-to-end channel which has the identical frequency band to improve data transmission rate, symmetrical streaming, and enhanced upload speeds. (Al-Turjman 2018)

In this section, IONT uses the Artificial Intelligence to perform task which mainly includes the machine learning techniques. Machine learning can be divided into three main categorises, they are supervised learning, unsupervised learning and reinforcement learning. According to security and privacy requirements there are many credentials to be dealt with like confidentiality and authentication, privacy, trust, time-criticality, availability, intrusion detection and predication, non-repudiation etc.

The fifth section is security and privacy requirements for IONT in terms of 5G/Big Data applications. Encryption based technique is used to make sure the data transmitted between the IONT hubs are confidential which is working together towards a consistent storage in the 5G-oriented Big-Data. Encryption techniques along with privacy protection mechanisms should be implemented. Time constraints data should be handled with proper care as if any mistake does takes place then the consequences will be really bad for this. Smart devices and applications of IONT should be available anytime and it should also be responsible to prevent any attacks that comes with in its path. Security components in these devices should be solid and adaptative so that it can tackle attacks intelligently. Users in IONT are highly concerned about their privacy issue and service providers should take care of this trust issue by considering mutual controlled with the third-party authority. Another main important security aspect is intrusion detection where other approaches like anomaly detection, specification detection and misuse detection are used. These approaches have to develop a bit more so that it can cope up with IONT capabilities. Users should be really careful about the IONT service providers can cause illegal actions which can be easily identified and reported. Thus, reliable IONT provider should be trusted.(Liaskos, Tsioliaridou et al. 2016)

Various security actions should be considered to secure the data. The first step is scanning and identification which is followed by authentication and authorization review. Then cluster configuration and deployment control are implemented, and then big data security planning is done. The authentication and authorization action play a vital role in security perspective in IONT. The authentication is classified as follows, encrypted password, proxy-based method, single sign-on, cognitive passwords, graphical passwords, cognitive authentication, hardware tokens, mobile phone method, biometric methods etc.

The sixth part is the IONT physical layer and 5G Big Data section where the communication mechanism of Nano-networks is briefly described. The nanonetwork uses various communication modules like acoustic, Nano-mechanical, electromagnetic and chemical. At the terahertz (THz) level the physical signalling is performed that is why various sizes antenna are having to be used so instead a special modulation technique is used. Similarly, in order to overcome the signalling difficulties, various research studies are being done.

The seventh section deals with the IONT communication technologies which are Nano sensor to Nano sensor interface, interface to internet communication technique.

The eight section in this paper describes the restriction in secured IONT. There are many restrictions in a secured IONT paradigm which can rightly be classified as primary and secondary restrictions. The restriction which has to be fulfilled in order to achieve a fully secured IONT is primary restrictions. It is the most important constraint of restriction in secure IONT.

In the ninth section, assessment and tools part is covered. 5G/ Big Data deals with humongous data which are manipulated plus it also takes into consideration its usage. ALL the devices are connected to the internet, even nanoscale data also. Hence, massive amount of data is generated. Hence, high-level evaluation methods and benchmarking tools are utmost necessary. Some of the tools listed in the paper are Hadoop, Spark, Sqoop, Mahout, Hive, Apache, hbase etc.

In this section, security attackers and attacks are promptly discussed. Here, different possible attackers and the related attacks are mentioned which are stated in the literature review of the paper. Active/passive, local/global, malicious/rational and insider/outsider, are the general attacker’s type. This classification is important to know the nature of the attacks and how the attacks can be prevented. Attacks can also be categorised usually as intentional and unintentional attacks, if the nodes tend to misbehave the messages are denied intentionally to be forwarded to another node in IONT. Activities like intentionally modifying the message, injecting fake information are intentional attacks. Similarly, unintentional attacks are the ones where if the nodes in IONT misbehaves then the message forwarding is unintentionally denied, modified.

The eleventh section generally deals with the open research issues where it has clearly stated that terahertz channel modelling can be studied to implement it in 5G/Big Data. There are challenges in terms of transmission of huge data in timely and reliable manner, security issues and channel capacity, also the utilised MAC protocols used in the system. As all the IONT devices are connected to the internet, security will be another issue that continues to be an open source research. There can also be issue of theft for the handheld gadgets which will affect the privacy. There can be various hazards physically in the sensors that will affect the overall IONT system.

Last section is the conclusion section, IONT has long run to go and it will advance the life of people in many ways. This will affect various sectors like agriculture, heath, cellular system, environmental monitoring and many other sectors.

# References

[1] K. David and H. Berndt, "6G Vision and Requirements: Is There Any Need for Beyond 5G?," *IEEE Vehicular Technology Magazine, Vehicular Technology Magazine, IEEE, IEEE Veh. Technol. Mag.,* Periodical no. 3, p. 72, 2018, doi: 10.1109/MVT.2018.2848498.

[2] A. T. Z. Kasgari, W. Saad, and M. Debbah, "Human-in-the-Loop Wireless Communications: Machine Learning and Brain-Aware Resource Management," ed, 2018.

[3] B. Chang, L. Zhang, L. Li, G. Zhao, and Z. Chen, "Optimizing Resource Allocation in URLLC for Real-Time Wireless Control Systems," *IEEE Transactions on Vehicular Technology, Vehicular Technology, IEEE Transactions on, IEEE Trans. Veh. Technol.,* Periodical no. 9, p. 8916, 2019, doi: 10.1109/TVT.2019.2930153.

[4] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proceedings of the IEEE,* vol. 104, no. 9, pp. 1711-1726, 2016.

[5] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and Low-Latency Wireless Communication: Tail, Risk, and Scale," *Proceedings of the IEEE, Proc. IEEE,* Periodical no. 10, p. 1834, 2018, doi: 10.1109/JPROC.2018.2867029.

[6] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks," *arXiv preprint arXiv:1710.02913,* 2017.

[7] Y. Xing and T. S. Rappaport, "Propagation Measurement System and Approach at 140 GHz-Moving to 6G and Above 100 GHz," ed: IEEE, 2018, p. 1.

[8] M. Mozaffari, A. Taleb Zadeh Kasgari, W. Saad, M. Bennis, and M. Debbah, "Beyond 5G With UAVs: Foundations of a 3D Wireless Cellular Network," *IEEE Transactions on Wireless Communications, Wireless Communications, IEEE Transactions on, IEEE Trans. Wireless Commun.,* Periodical no. 1, p. 357, 2019, doi: 10.1109/TWC.2018.2879940.

[9] I. F. Akyildiz, S. Nie, S.-C. Lin, and M. Chandrasekaran, "5G roadmap: 10 key enabling technologies," *Computer Networks,* Article vol. 106, pp. 17-48, 09/04/September 2016 2016, doi: 10.1016/j.comnet.2016.06.010.

[10] F. Al-Turjman, "QoS—aware data delivery framework for safety-inspired multimedia in integrated vehicular-IoT," (in English), *Computer Communications,* Article vol. 121, pp. 33-43, 05 / 01 / 2018, doi: 10.1016/j.comcom.2018.02.012.

[11] C. Liaskos, A. Tsioliaridou, S. Ioannidis, N. Kantartzis, and A. Pitsillides, "A deployable routing system for nanonetworks," ed: IEEE, 2016, p. 1.

[12] J. M. Jornet and I. F. Akyildiz, "Graphene-based plasmonic nano-antenna for terahertz band communication in nanonetworks," *IEEE Journal on selected areas in communications,* vol. 31, no. 12, pp. 685-694, 2013.

[13] C. Park and T. S. Rappaport, "Short-range wireless communications for next-generation networks: UWB, 60 GHz millimeter-wave WPAN, and ZigBee," *IEEE Wireless Communications,* vol. 14, no. 4, pp. 70-78, 2007.

[14] A. Morgado, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "A survey of 5G technologies: Regulatory, standardization and industrial perspectives," *Digital Communications and Networks,* vol. 4, no. 2, pp. 87-97, 2018.

[15] F. Al-Turjman, "5G-enabled devices and smart-spaces in social-IoT: An overview," *Future Generation Computer Systems,* Article vol. 92, pp. 732-744, 03/01/March 2019 2019, doi: 10.1016/j.future.2017.11.035.