

A Review of Data Gathering Algorithms for Real-Time Processing in Internet of Things Environment

Atheer A. Kadhim¹, Norfaradilla Wahid²
Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia
Johor, Malaysia

Abstract—Today, Wireless Sensor Network (WSN) has become an enabler technology for the Internet of Things (IoT) applications. The emergence of various applications has then enabled the need for robust and efficient data collection and transfer algorithms. This paper presents a comprehensive review for the existing data gathering algorithms and the technologies adopted for that applications. After reviewing the algorithms and the challenges related to them, which extend the physical reach of the monitoring capability; they possess several constraints such as limited energy availability, low memory size, and low processing speed, which are the principal obstacles to designing efficient management protocols for WSN-IoT integration.

Keywords—Internet of Things (IoT); Wireless Sensor Network (WSN) and Data Gathering; Virtual Machine (VM); Virtualization Cloud (VC); Data Reduction (DR); Access point (AP); Mobile Ad hoc Network (MANET)

I. INTRODUCTION

The Internet of Things (IoT) is one of the emerging technologies in the area of information technology [1]. It is widely called IoT which means that many things or objects are interconnected to each other through the Internet [2]. Internet technology is known for a long time and has been used for connecting computers using Internet protocol (TCP/IP) so that millions of networks are interconnected around the world [2]. Those networks are used for different kinds of purpose such as private, public, academic, business, and government networks. Technically, these networks might be connected to each other using fiber optics or wirelessly [3].

The availability of Internet networks has triggered an interest in connecting all the objects using Internet networks [4]. As such, several researchers have paid attention over the last ten years to enable connectivity for worldwide networks. This, in turn, has enhanced the vision for global networking for all the objects. IoT is virtual shall provide unlimited opportunities and connections to occur [5]. Nowadays, IoT research and development has become one of the hot topics in many countries around the world. Although IoT has provided a lot of opportunities, many challenges have been arisen such as security considering the huge number of devices connected to each other for achieving a certain type of function.

The next generation of IoT is required to provide new services to meet the demand of the fourth industrial revolution. Therefore, it must be able to deal effectively in the data transfer process without any human involvement in the interconnected objects such as computing devices and digital machines [3].

Although there are tremendous works have been done by several international standardization bodies, industry players, researchers, developer and other parties, there are several issues need to be addressed to reach the peak of IoT capabilities [6].

However, in this work, several studies are reviewed, discussed and critically analyzed to providing a solid literature review for future research. Additionally, a wide range of case studies from the past up to date is presented for a better understanding of the theory related to each proposed algorithm and the applicable technology in each study. Articles, journals, books, and previous works had been listed in the following tasks.

In the manner, Dias et al., [6] review numerous techniques that are used for Prediction-Based Data Reduction in Wireless Sensor Networks. Meanwhile, the work of Maraiya et al., [7], reviews the most common Data Aggregation methods in WSN. In the work of Cheng et al., [8], reviewed the state of the art of approximate data collection algorithms in IoT and WSN. Fig. 1 shows the number of reviewed and discussed articles in this work based on the years, note, * represents the number of review articles.

A total of 41 research articles and four review articles are covered in this work. The review emphasizes on the fundamental of IoT based on WSN which are Sensor Technology, Characteristic Features, Overview of IoT Sensor, IoT Hardware Prototype and Saving Energy Techniques. This review approach allows us to improve the scope and shape the direction of IoT based on WSN.

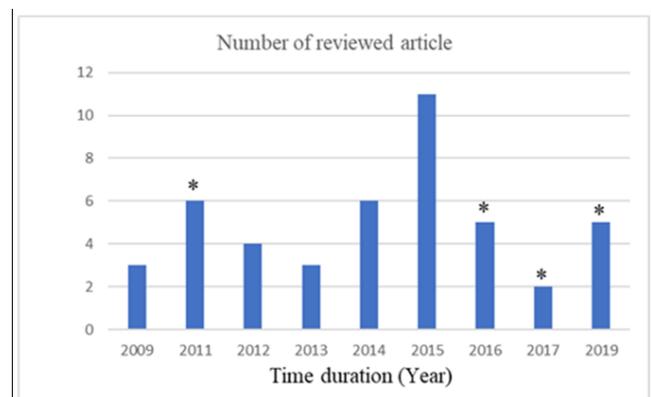


Fig. 1. An Overview of the Reviewed Articles.

This research is partly sponsored by Universiti Tun Hussein Onn Malaysia through Tier 1 Grant H108.

This paper segmented into four parts starting with the section of introduction which describes the IoT, WSN, and IoT based on WSN. Furthermore, an overview of the IoT based on WSN has been discussed under Section II. Then, Section III reviews the IoT sensor node energy-saving methods in WSN is given. Whereas, the analysis and discussion of this work have been given in Section IV. Finally, Section V presents the conclusion of this paper.

II. OVERVIEW OF IoT BASED ON WSN

The vision for the Internet in the future is to be a global network consisting of many objects connected together using a specific IP address based on the relevant standard. Accordingly, all the devices including computers, sensors, RFID tags or mobile phones will have the capability to access the network and then communicating with other devices for the purpose of performing a certain task [9]. The WSN technology has increased the importance of IoT by combining the technologies of WSN and the internet resulting in IoT infrastructure. WSN is widely utilized for different kinds of applications. It is used mainly for gathering data from the field or environment through sensors. The key element for the IoT paradigm is RFID and WSN. RFID is used for identification purposes and tracking. Meanwhile, WSN is a very good option for proving sensing actuation function to IoT [3]. WSN has been adopted in many applications as an effective solution in several applications and research.

Unfortunately, the adoption of WSN in different types of applications has made a lot of challenges to specify the WSN requirement to be used for IoT. Generally, the common WSN platform can be applied with reasonable results in a wide range of IoT monitoring applications [6]. Furthermore, it is required to deploy WSN not only in monitoring applications but also in many applications such as security, biomedical research and tracking [10]. So, IoT is expected to play a significant role in very vital applications for emergency services. Based on the type of network, IoT networks can be classified into many types according to its intended application such as environmental data collection, military applications, security monitoring, health applications, home applications, and so on.

Although the generic form of WSN is applicable for monitoring applications, it still requires tough requirements such as employing a huge number of nodes with very low cost [9, 10]. The designer needs to consider that the nodes have to stand alone for a long time before the service time. Other factors also need to consider such as simplicity positioning the nodes with cheap maintenance cost. These requirements made generic WSN platforms less preferred [7]. This is would limit the applicability while it is suitable for many applications including but not limited to agricultural, medical and military applications as depicted in Fig. 2.

WSN is characterized as a two-direction system where the data on long-distance can be transferred between the sink node and sensors through a jumping mechanism. As such, the sensed data such as temperature, moistness, light and so on can be sent to the destination point and afterward pass on to the preparing gear [11]. The detecting hubs convey in multi-jump. Each sensor is a handset having reception apparatus, a small-scale controller and an interfacing circuit for the sensors as a

correspondence, activation and detecting unit individually alongside a wellspring of intensity which could be both battery or any vitality reaping innovation However, in this work we emphases on the fundamental parts of IoT based on WSN as shown in Fig. 3.

A. Sensor Technology

As a keyway to obtain information, sensor generation and communicator generation, personal computer technology constitutes the three pillars of data technology. Sensors are the main components of the Internet of Things awareness layer, which can assist the Internet of Things data and access the external physical world in a timely manner [9]. The sensor network technology formed by the sensor network technology's sensor and communication network has laid the foundation for the development of the Internet of Things. The sensor warehouse is mainly used for the environment and small parts tracking to meet the requirements of the product to the surrounding environment and safety monitoring. Inductive devices, including in particular sensors that select sensing gadgets (including light sensors, temperature sensors) are widely used in warehouse control of multipurpose systems [11].

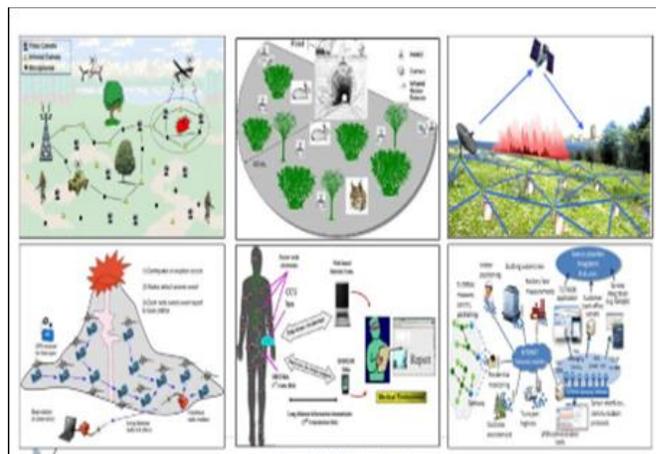


Fig. 2. WSNs Applications [10].

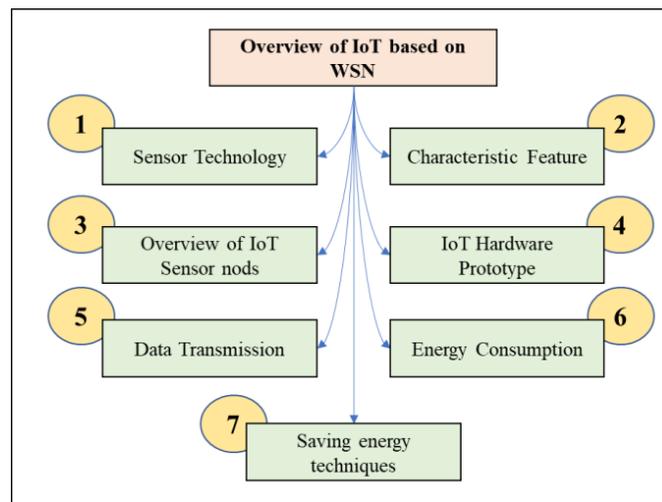


Fig. 3. Overview IoT based on WSN.

A sensor is a tool that converts the physical measurement sign. The sensor's selection must conform to the requirements of the surrounding environment or the object. It can be collected by the body, and the organic chemistry effect is associated with the measurement sensor as the development of the system. In well-known circumstances, there are no special requirements for the environment of daily warehouse items (including temperature, humidity, etc.), so special sensors should be installed near the well-known warehouses. This type of sensor may be a combination of various sensing environments [10]. In the measurement to meet the needs of general merchandise. For unique items or unique business needs, often more feature definition areas are needed, and integrated sensors are set up, including gravity, strain, fuel density, and sensor noise. The steeper price included in the number of sensors, not always the same as the use of smooth, can be placed when the product is replaced or the items necessary to alternate, move the sensor, and open and close the component provider as required to increase the sensor usage rate [8].

B. Characteristic of WSNs

WSN can be generally defined as a set of nodes that are connected to each system so that all the nodes perform together with the sensing and control of the environment. This process enables the communication between humans or computers and the environment under control [12]. WSNs these days, as a rule, incorporate sensor hubs, actuator hubs, passages, and customers. An expansive number of sensor hubs conveyed haphazardly within or close to the observing region (sensor field), structure arranges through self-association. The gathered information is then checked by sensor hubs and pass forward to other sensor hubs via the jump mechanism [13-15]. For the purpose of reliable transfer of data, the hubs need to take care of the availability of jump to the portal hub which then transfers the data to the administration hub via the web or satellite [7]. The client designs and deals with the WSN with the administration hub, distribute checking missions and gathering of the observed information. As related advancements develop, the expense of WSN hardware has dropped drastically, and their applications are slowly growing from the military regions to mechanical and business fields. In the interim, guidelines for WSN innovation have been all around grown, for example, ZigBee.

C. Overview of IoT Sensor Nodes

Although WSN provides good service for sensing areas with difficult to access, it experienced a lot of challenges which limit its deployment. One of those challenges is to provide a sustainable power supply for the wireless sensor nodes [6], [9]. Up to our knowledge, there are a large number of published papers have proposed different kinds of solutions to overcome this drawback. Recently, it has been proposed to use energy harvesting technology and wireless charging technology to provide sustainable energy resources [16].

One of the basic elements in IoT based WSN is the IoT sensor hub which composes of four sections (i) power module, (ii) a sensor, (iii) a microcontroller, and (iv) a remote handset. The function of the power module is to provide power for the framework whereas the sensor is used for capturing the status

of the sensed raw data. It is in charge about data transfer and exchange to other sensors. In fact, it converts the sensed data such light into an electrical signal and then passing it to the microcontroller [8]. The microcontroller is employed to receive the information sent by the sensor and perform the required operations on it. The RF module is the last element which is located at that point where the information is exchanged so physical acknowledgment can be accomplished in this element. It is highly required to consider all these elements need to be with small size and low power consumption.

D. IoT Hardware / Prototype

The rapid advancement of computing hardware technology has resulted in developing small scale devices at a reasonable cost. Consequently, IoT has got attention for various applications [3].

For IoT, the microcontrollers are integrated with processors, wireless chips and other components to form the Prototyping Development Kit through embedded software packages with reprogramming capability. Other studies, as in this future research, employ Arduino hardware since it is an open-source device for controlling a large number of sensors rather than personal computers [17]. Briefly, Arduino has the capability to be registered based on the microcontroller board and the programming composer of the board. Arduino hardware is programmed using C or C++ programming language. Based on the handling sight and the programming condition, Arduino load represent logical wiring similar to physical connections, [18]. The microcontroller has a very important microchip called AVR provided by Atmel organization. This chip operates at 16 MHz with an 8-bit, but it has limited and has limited accessible memory (32 Kbytes of capacity) and 2 Kbytes of irregular access memory. Due to this Atmel chip, Arduino became popular for work especially for many DIY applications [16].

With respect to the product advancement, programming of the shield was begun earlier with sensors programming and further included clock module. DHT22 and RTC modules are used to provide information with aid of the libraries of Arduino IDE [18]. The data are read in real-time through the analog pins of sensors and then send a copy of data to IDE's serial port for testing purposes. Arduino gives an SD card library, which was utilized to make a capacity SD Card Data-log. The capacity spares sensor information and the opportunity to the SD card on the W5100 arrange shield. To give clients remotely screen their nursery through site page, the webserver should have been built up. Webserver libraries were made for Arduino however they didn't meet the necessities of the framework. Better web content help was expected to the venture, so new webserver was intended to fit exactly the framework prerequisites. Live to outline was intended to show information for the client on the site [18]. Without JavaScript support on the webserver, web association would have been required. The new webserver empowers the framework to be utilized disconnected in a neighborhood without web association.

E. Data Transmission in IoT based WSN

Al-Fagih, [19] has reviewed the data transmission in IoT in detail. The data transmission in IoT-WSN is directly related to the type of application for data transfer from the sensor toward

the access point. Generally, data transmission can be performed through one of the following types: continuous data transmission where the data are sent by the sensors in specific times. The second type is event-dependent where the transmission is enabled when such an event occurs. The last is query-based data transmission where data are sent once the access point transmits query. Despite these three types are different in operation, the continuous type can be jointly applied with an event or query-based data transmission forming a hybrid model. It's observed that the introduced framework facilitates applications that are more relevant to the query-based model.

In the literature, several protocols are proposed for data delivery in WSN such as, [5, 6, 19]. These protocols are arranged in three different architectures: a hierarchical structure, a data-centric structure, or location-based structure as depicted in Fig. 4. In the hierarchical structure, nodes are arranged in a clustering paradigm so that the head of the cluster collected together for the purpose of minimizing the transferred data and thus reducing the cost of energy.

The protocols related to the data-centric paradigm are query-based and as such only, the desirable data will be transmitted. In this process, the duplication of transmission for the same data is avoided. In the third type, the data is sent selectively to the targeted location. This is a good alternative to the transmission of data to all the locations. As a result, the bandwidth and power are saved significantly.

It is vital to take in mind that IoT is strongly dependent on WSN. In WSN, remote communication has existed between the sensors and the network hub. The major function of IoT is to construct an overall system among all the conceivable articles. Besides, WSN is a genuinely enhanced innovation that guides the client to accomplish the importance of IoT [8]. The primary thought of WSN is to associate the detecting layer and system layer in the IoT.

Fig. 5 shows one scenario of WSN where the operation is event related. Once the sensor node identifies an event, it gathers the data and send it to the next nearest node and so on until reaches to the destination node. This is the simple structure of independent WSN. In this scenario, data are delivered only to a single gateway. In addition to that, the data can be transmitted in access point scheme or hybrid scheme as shown in Fig. 6, respectively.

The independent WSN has been improved and enhanced. The improved version of the independent WSN is called hybrid WSN. In contract to independent WSN, hybrid WSN has multiple gateways for the purpose of data transmission. In this case, WSN will maximize its performance.

In the third scheme, Access Point-WSN is not similar to the other schemes. Basically, this scheme adopts WLAN structure [20]. As shown in Fig. 7, there are many of hubs in WSN, each one of them compose with the other to make the connections. The system is interfacing WSN and Web through one portal. When the entryway is separated, there is an alternative method to interface the two systems. Nonetheless, it is divides the nodes into two parts to support and fix the lacking that may happen during data transmission among the nodes. It is more

grounded from the system of one portal in first scenario. In contrast to the past strategy, pathway arrange the fills as a self-arranged in WSN.

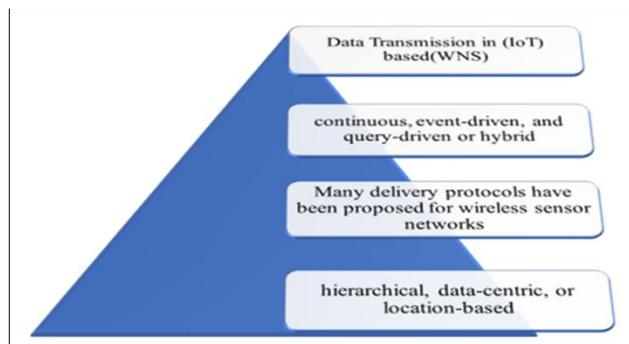


Fig. 4. Data Transmission in (IoT) based (WNS).

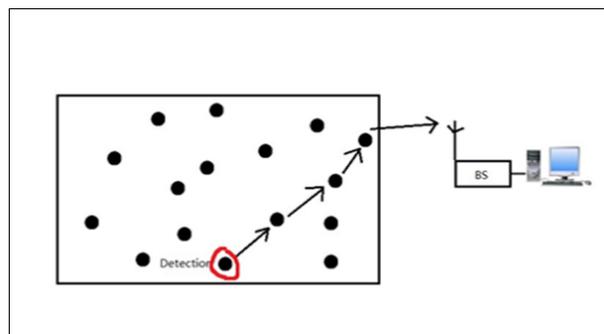


Fig. 5. Independent WSN Scheme.

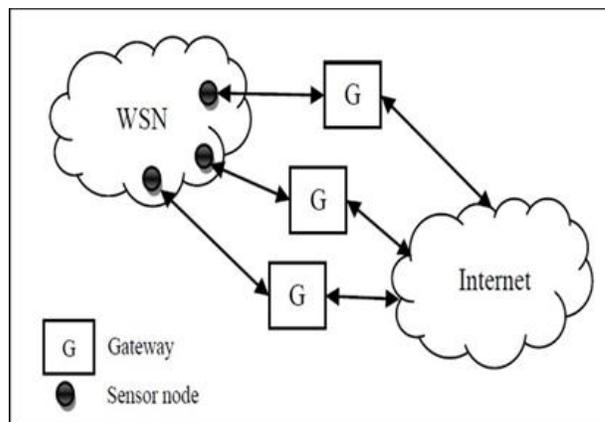


Fig. 6. Scheme of Hybrid WSN Data Transmission.

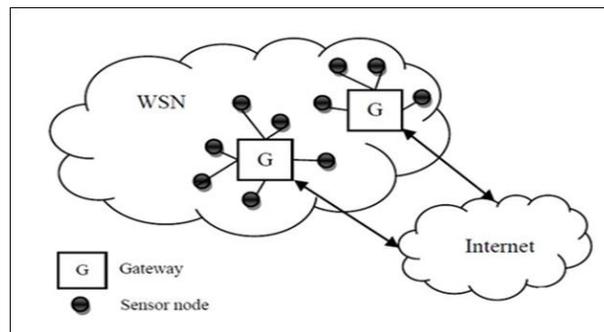


Fig. 7. Scheme of hybrid WSN data transmission

In WSN, the applicable protocols related to data routing are based on the data transmission model specifically in managing network active status and power consumption. In this regard, it is found that hierarchical scheme protocols are the most proper option for environmental monitoring applications since data are sent unremittingly to the AP. This is expected since applications can produce a large number of duplicated data which gathered together to the receiver. This, in turn, would shrink data overwhelming on the route and thus optimize the energy consumption. However, the typical WSN is too limited to meet the needs of the Internet of Things in terms of heterogeneity and transmission/processing load. Therefore, they adopted an extended concept of a sensor network that contains MANET nodes.

F. IoT Energy Consumption

energy waste has been described in [21]: In IoT based WSNs, sensors consume energy while sensing, handling, transferring and transferring or getting data to complete the tasks. Data gathering are normally collected by sensor dedicated for that job. Obviously, as the produced data are minimized, the saved energy would be increased. The inherent redundancy of a wireless sensor network generates a huge similar report, and the network is responsible for routing to the receiver. It has been proven practically, that the communication system is responsible for the significant amount of power wastage. In terms of communication, from the perspective of applications, useless countries have also wasted a lot of energy [9].

G. Energy Saving Techniques in WSN

There are various solutions for tackling the issue of energy consumption. Some of those solutions are to minimize the sent data, to reduce the overhead which accompanies the data as well as applying a smart and efficient routing algorithm. Additionally, increasing the time for idle status with no transmission which implies a limited number of transmissions. The other solution is to focus on topology control as described by [9]:

- Minimize of data transmission: the transfer of data duplication or unnecessary data can be avoided by applying these certain types of solutions. Many related works have been discussed in the related works section in this chapter.
- Decreasing the overhead: The data transmission will be maximized by minimizing the overhead attached to the data. Several existing techniques have been discussed in the related works section in this paper.
- Optimizing the routing: For better performance, the routes need to be “available” during all the time but in such a way to keep minimum power consumption. In this regard, some routing protocols are utilizing the node movement or “send to all” characteristics of WSN. In the other type, the GPS coordinate of the nodes is employed to create a path to the destination. In a different way, a hierarchy structure of nodes is used for a proper path with less overhead. Additionally, many paths can be employed for reaching the destination by balancing the data on different paths resulting in more.

Finally, a statistics-centric protocol transmits the most effective information to the involved nodes to avoid costly transmissions.

- Duty cycling: The duty cycle refers to the proportion of time that a node is active during its life cycle. Nodes should coordinate and adapt to specific application needs during sleep or activity. The nodes with High granularity technology are the active nodes in the WSN while the nodes with low granularity imply shut down the RF transmission of the active nodes since there is no data to send. They are highly relevant to media access agreements.

III. REVIEW OF THE IOT SENSOR NODE ENERGY SAVING METHODS IN WSN/IOT

Chaudhary et al., [20] praised the improved technologies of energy-saving technologies for data gathering in WSN. In that article, the battery has been used wisely and effectively and thus low power consumption is identified. As such, people can maximize the benefit of using the same resource but with a large number of messages. The production of cluster overhead and node selection methods are based on certain parameters in order to use the correct decisions in this case, using global weight calculations such as nodes, cluster head data collection, and data aggregation using data cube clustering.

Rohankar et al., [21] had made reviewed the latest developments in WSN including data gathering methods. The review paper had categorized each consideration technology based on the underlying topology. The second classification is based on the saved energy scheme. Those technologies are assessed qualitatively by comparing one to each other. At the end of the review, the limitations of the technologies considered were discussed.

In the previous work of Quwaider and Jararweh, [22], they proposed a cloud-based and efficient data collection system in a cloud-based WBAN. The main objective was to provide a wide range of WBAN monitoring data to end-users or service providers in a robust way. In this case, the data gathering in WBAN has been simulated using a prototype composed of a virtual machine (VM) and a virtualization cloud (VC). Using this prototype system, they provide a scalable storage and processing infrastructure for large WBAN systems.

In Xu and Song [23], the authors studied instantaneous periodic query scheduling in multi-hop data gathering (WSNs). Given a set of heterogeneous data gathering queries in the WSN, each query desires that data from the source node be assembled into the control center. Firstly, they proposed a series of almost urgent requirements for different queries that can be scheduled by the WSN. Then, three effective data gathering algorithms were developed to meet instantaneous requirements under the limitation of resources. The issue has been addressed through three vital tasks: (1) data gathering, building of routing tree, (2) schedule of path activity, and 3)) scheduling on Packet-level.

In Luo et al., [24], the work investigated the speed of raw data gathering from all nodes to the receiver. In this case, the TDMA technique has been employed on the same frequency

channel. Additionally, a centralized and distributed fast data gathering algorithm is employed to find the optimal solution in polynomial time when no interfering links occur. The study also proposed the RCTS algorithm for identifying the best solution. It is found that RCTS is time efficient and it is a good candidate for eliminating the major parts of the interference captured in indoor and outdoor environments.

Guo et al. [25] proposed a new method called Event-based Data Aggregation (EDA) which uses using the fuzzy logic cloud member model to gather the data of events in the WSN. In this method, the base station has the capability to restore the whole event data once it receives the data packets of the event. EDA method provides a degree of balance between delay factor, savings of flow and the accuracy of the restored events. The performance evaluation of this method has been conducted by Guo using both analysis and simulations.

Jacques et al., [26], proposed a new filtering technique for specifying the redundancy of data received in constant time slots. Further to that, the data gathering method has been proposed based on grouping data which share most of the same features as such the data integration would be maintained.

In the work of Pfletschinger et al., [27], a network coding scheme in WSN has been proposed and its effectiveness has been evaluated through the following factors: reliability enhancement, power efficiency and resilience to network protocol failures. The main challenge in that work was to identify the number of eavesdropping so that space diversity can be employed but with low power consumption.

In the prior work of Raza, [17], the purpose of this work is to explore the complex interactions between application features and adaptive mechanisms across the network stack by using specific real-world deployments. Moreover, the paper proposed a generic framework that integrates adaptation into near-optimal energy efficiency for heterogeneous applications.

In Bahi et al., [28], the author introduced an energy-saving technique for data gathering in structure with constant on time slots transmission. This study has discussed the issue of locating the pairs of nodes that generates redundant data. In addition, this study provides a frequency filtering method to solve this problem. In Enam et al., [29], the work was to build up a vitality productive information gathering condition for a substantial scale, haphazardly sent group based remote sensor organizes by utilizing a virtual lattice-based instrument to limit the bunches and balance out the group bulks in the system. This was one of the requirements for implementing the proposed differential information total plan for the spatially related information in a bunch.

In the work of Laiymani and Makhoul [30], they displayed a productive versatile testing approach dependent on the reliance of restrictive difference on estimations shifts after some time. At that point, in extra they proposed a various dimensions action display that utilizes conduct capacities demonstrated by altered Bezier bends to characterize application classes and take into consideration examining versatile rate. The proposed strategy was effectively tried in a genuine sensor informational collection as the researcher said.

In Enam [31], the author built up a novel and a versatile technique for information conglomeration that abuses the spatial connection between the sensor hubs. The primary element of the proposed accumulation technique is that notwithstanding lessening the expense of excess information move in the system, it additionally ideally uses the accessible space in a bundle at each group head.

Trade-off between them depends to a large extent on the certain application. In this regard, one of the techniques for data gathering is called Prefix Frequency Filtering (PFF) where power consumption and data accuracy are targeted in that study. The main target of PFF is to identify the data groups produced by neighboring nodes with shared features resulting in canceling the redundant data and thus avoiding energy dissipation. Although this method is simple it requires tedious computational time. In prior work of Harb, Makhoul, and Laiymani [32], PFF has been improved by integrating K-means of clustering algorithm as such it is called KPFF. The KPFF was able to minimize the time needed for detecting similar pairs and therefore the data latency is minimized as well.

Li et al., [33] had analyzed the complexity for many factors such as data message complexity and energy cost complexity. In this work, the lower bound of the complexity of the optimal method has been employed but for the other factors, an efficient distributed algorithm has been used. This in result provided gradually matching with the upper bound of complexity.

In Carlos-Mancilla et al., [34], this work proposes and builds up a proficient information collection technique for remote sensor systems (WSN). In the proposed information collection strategy, each bunch of head (CH) hub contains a nearby sending history to choose whether to advance or dispose of the most as of late gotten parcel. At the point when the new parcel touches base at the CH hub, the limit is determined dependent on the data of the sending history; at that point, an arbitrary number is created and contrasted with the edge an incentive with deciding if the information bundle ought to be disposed of. Truth be told, the CH hub advances the new bundle with a likelihood of $1-p$ and disposes of it with the likelihood p that decides the parameter p dependent on the sending history.

The energy-saving with optimism use for a long time has become a recent challenge facing the researchers around the world. It is required to minimize the power usage in WSN but has to ensure the reliable and robust functional performance of WSN. It has to meet the minimum requirement of normal operation without failure due to power supply. Whereas, the authors proposed a reactive data acquisition scheme called SWIFNET in [35]. It is based on the synergetic effect of a combination of data reduction methods and energy-saving data compression schemes. In particular, it combines compressive sensing, data prediction, and adaptive sampling strategies.

The Internet of Things represents advances in miniaturization, wireless connectivity, and increased data storage, driven by various sensors. Sensors detect and measure any changes in location, temperature, light, etc.; in addition, they need to convert billions of objects into data-generating "things" to report their status and often interact with their

environment. Application and service development methods and frameworks are needed to support the implementation of solutions that cover data collection, transmission and data processing, analysis, reporting, and advanced querying. In the previous work of Lengyel et al. [3], this article introduced the Sensor HUB framework, which utilizes the most advanced open source technologies and provides a unified toolchain for IoT related applications and service development. Sensor HUB is both a method and an environment that supports the development of applications and services related to the Internet of Things. In addition, it supports data monetization methods that provide a way to define data views and analyze data on different data sources. The framework uses the platform-as-a-service (PaaS) model and has been applied in the areas of vehicles, health, production lines, and smart cities.

Data collection and propagation in the Internet of Things (IoT) Wireless Sensor Network (WSN) requires a stable multi-hop network path from source to sink. However, due to limited energy, the battery consumption of the interrupting node can cause the path to be disconnected and result in the end-to-end data transmission failure of the WSN-based IoT. Therefore, in addition to its own energy-saving, each sensor involved in multi-hop transmission activity also needs a feasible strategy for selecting a relay node through utilization. Its remaining energy and multi-hop IoT network connection. In Luo et al. [5], the author first analyzed the energy consumption model and data relay model in wireless sensor networks, and then proposed the concept of "equivalent node" to select relay nodes to achieve data transmission and Energy-saving optimization. A probabilistic propagation algorithm called ENS PD is designed to select the best energy strategy and extend the lifetime of the entire network. Extensive simulations and actual test results show that our models and algorithms can minimize power consumption compared to other methods while ensuring the quality of communications in WSN-based IoT.

In the previous work of Rault et al. [36], the paper proposes a novel sensor network data acquisition framework using flight sensor nodes. Since sensor nodes are usually limited by energy, efficient data communication within the network is required. In contrast to its conventional role in sensor networks, the proposed framework utilizes various entities that form networks for different utilities. The use of flight sensor nodes is often considered the traditional purpose of sensing and monitoring. Flight sensing nodes are commonly used in the form of an airborne sensor network and they cannot be used as data collection entities as proposed in this framework. Similarly, it is often desirable for a cluster head (CH) to transmit aggregated data to a neighboring CH or directly to a base station (BS). In the proposed framework, the CH transmits data directly to the flight sensor nodes, avoiding the need for energy-intensive multi-hop inter-cluster communication to communicate information to the BS. Flight sensor nodes are called sensor flights.

In Mudgule et al., [37], the author focuses on the data redundancy and energy of sensor nodes. Data simplification is one of the data pre-processing techniques for data mining, which can improve storage efficiency and reduce costs. Data Reduction (DR) is designed to remove unnecessary data when transferred. For this reason, according to WSN, many data

reduction strategies will be introduced. This survey introduces the latest data reduction-based algorithms and techniques that help increase the network's energy and longevity.

In Maraiya et al., [7], discussed the data aggregation approaches based on the routing protocols, the algorithm in the wireless sensor network. And also discuss the advantages and disadvantages of various performance measures of the data aggregation in the network.

In Pandey and Kaur [38], the authors' attention to various data aggregation algorithms in a wireless sensor network. Data aggregation technique increases the lifetime of sensor networks by decreasing the number of packets to be sent to the sink or base station. Here, they first explore the data aggregation algorithms on the basis of network topology, then they explored various tradeoffs in data aggregation algorithms and finally they highlighted security issues in data aggregation.

The work in (Hung et al., [39], proposed a centralized algorithm to determine a set of representative nodes with high energy and wide data coverage. Here, the sensor node's data coverage is considered to be a set of sensor nodes that have very close reading behavior to a particular sensor node. In order to further reduce the extra cost in the messages used to select representative nodes, a distributed algorithm was developed. In addition, when the energy of the original representative node is insufficient or cannot capture the spatial correlation within its respective data coverage, a maintenance mechanism is proposed to dynamically select the alternative representative node. Through experimental research on synthetic and actual data sets, the proposed algorithm has been proved to be able to effectively provide approximate data collection while extending the network lifetime.

A performance assessment of information reduction techniques for IoT based WSN Multimedia applications has been provided in [40]. In this article, the authors study the performance of various BS algorithms and compression techniques in computing and communication energy, time, and quality. They have chosen five different BS algorithms and two compression techniques and implemented them on the Android platform. Considering the fact that these BS algorithms operate under the WMSN environment where data is subject to packet loss and error, they also studied the packet loss rate performance of the network under various packet sizes. Experimental results show that the highest energy efficiency BS algorithm can also provide the best prospect detection quality. The results also show that data compression techniques including BS algorithms and compression techniques can provide significant energy savings in terms of transmission energy costs.

In the prior work of Dias et al., [6], the authors analyzed and classified the existing prediction-based data reduction mechanisms for wireless sensor network design. Their meaning is based on the constraints of the wireless sensor network, the characteristics of the prediction method, and the monitoring data, and a systematic procedure for selecting the prediction scheme in the wireless sensor network. Finally, this article concludes this article and discusses future challenges and open research directions for the use of predictive methods to support the development of wireless sensor networks.

A data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications have been proposed in [41]. In this work, the author proposes another structure that consolidates information expectation, pressure, and recuperation together to accomplish the exactness and effectiveness of grouped remote sensor arrange information handling. The principle reason for this system is to diminish correspondence costs while guaranteeing the precision of information handling and information expectation. In this system, information expectation is accomplished by executing the Least Mean Square (LMS) bi-forecast calculation with the best advance size with the least mean squared subsidiary (MSD), where the group head (CH) can get a decent guess from the sensor the genuine information of the hub. On this premise, the brought together primary segment investigation (PCA) innovation is utilized to pack and recoup the forecast information of the CHS and the sink individually, in order to spare the correspondence cost and wipe out the spatial repetition of the detecting information on the earth. Every one of the blunders that outcome from these procedures will, in the long run, be assessed hypothetically and these are controllable. In light of hypothetical examination, the creators structured some usage calculations. Reproductions utilizing certifiable information have demonstrated that (LMS) system gives a financially savvy answer for natural checking applications in bunch based WSNs.

In recent work of Alduais et al. [17], the authors have strategized to diminish the number of information transmissions and lessen the measure of information that prompts an all-encompassing system lifetime. The proposed strategy intends to diminish the number of transmitted messages by means of hubs supporting single and different sensors depending on the relative or relative contrasts between the transmitted present and last sensor estimations. The outcomes demonstrate that the proposed technique appears to demonstrate the best execution in decreasing the number of message transformations and parcel measures. In that article, the normal level of decrease in message transmission times is 74%, and 80% is the normal rate decrease of hub information. Bundle measure. From the outcomes, it tends to be obviously observed that diminishing the number of bundle transformations and decreasing the extent of hub information parcels lessens vitality utilization and broadens the administration life of the framework.

Raza et al. [11], depicts subsidiary based expectation (DBP), another kind of information forecast method that is less complex than the writing. Assessments utilizing genuine datasets from various WSN arrangements demonstrate that DBPs, for the most part, perform superior to contenders, with information pressure rates as high as 99% and great expectation exactness. In any case, tests led on genuine remote sensor arranges in expressway burrows have appeared considering the system stacking, DBP just triples its lifetime - a critical outcome in itself, however, it is a long way from the above information concealment rate. So as to completely understand the vitality funds acknowledged by information forecast, the information layer and the system layer must be together enhanced. In that review test explore, considering the

activity of DBP, a basic change of the MAC and wiring stack can altogether expand the lifetime by multiple times.

In Ait-Sahalia, and Xiu, [42] this work proposes an algorithm based on "Principal Component Analysis" to perform multivariate data reduction. It was considered an air quality monitoring scenario as a case study. The results showed that using the proposed technique, the outcome of the study reduced the data sent preserving its representativeness. Moreover, it's showed that energy consumption and delay were reduced proportionally to the amount of reduced data.

In prior work Mccorrie et al. [43], another strategy for specifically sifting detected information dependent on state acknowledgment has been concocted that utilizes skewed twofold exponentially weighted moving normal channels for exact state expectation. This is genuine regardless of whether a critical temperature step change happens. A test system was executed to create a flight temperature profile as the flight temperature experienced, all things considered, so the calculation could be balanced and assessed. The outcomes abridged a reenacted trip of 280 variable lengths (from roughly 58 minutes to 14 hours). The outcomes demonstrated that in the departure, cruising and landing stages, the number of transmissions was diminished by a normal of 95, 99.8 and 91% with the detecting and transmitting framework. Correlation of the transmissions experienced when examining at the same rate. The algorithm produces an average error of 0.11 ± 0.04 °C in the 927 °C range.

In de Carvalho et al. [44], the authors proposed to use a method based on multiple linear regression to improve prediction accuracy. The improvement is achieved by the multivariate correlation of readings gathered by sensor nodes in the field. The authors claimed that the solution has outperforms some current solutions adopted in the literature.

A survey for approximate sensory data collection has been presented in the recent work of Cheng et al., [45], that survey reviews the state of the art approximated by a collection algorithm. They classified the min into three categories: the model-based ones, the compressive sensing-based ones, and the query-driven ones. For each category of algorithms, the advantages and disadvantages are elaborated, some challenges and unsolved problems are pointed out.

In the recent work of Alduais et al., [1], that work displayed another pointer to assess the execution of various multivariate information decrease models in remote sensor systems (WSNs). The proposed measurement is known as the update recurrence metric (UFM), which is characterized as the recurrence of refreshing model reference parameters amid information accumulation. A strategy for evaluating the mistake limit amid the preparation stage has additionally been proposed. The prescribed blunder edge is utilized to refresh the show reference parameters when vital. Numerical examination and recreation results demonstrate that the proposed measurement confirms the adequacy of the multivariate information decrease show in vitality utilization of sensor hubs. Furthermore, the proposed versatile limit improves the execution of the model more than the non-versatile edge in decreasing the recurrence of refreshing model reference parameters, which correctly extends the lifetime of the node.

Compared to the non-adaptive thresholds of the multivariate data reduction model of MLR-B and PCA-B, the adaptive threshold increases the frequency of parameter updates by 80% and 52%, respectively.

IV. CONCLUSION

The vast usage of the Internet of Things (IoT) innovation for different applications has empowered the requirement for hearty and effective information accumulation and exchange calculations. This paper introduced a complete audit for the current information gathering calculations and the innovations received for those applications. It reviewed the proposed algorithm for tracking this issue. Although the existing algorithms for data gathering can perform well, it still needs to be further enhanced in the future to overcome all the deficiencies such as low power consumption for standalone sensors. This paper has covered a comprehensive review for those algorithms. However, this paper is a platform for developing many solutions by the researchers for an efficient algorithm for wireless sensors. That solution will be investigated when proposed in future work.

ACKNOWLEDGMENT

The authors would like to thank Universiti Tun Hussein Onn Malaysia for supporting this research through Tier 1 Grant H108.

REFERENCES

- [1] Alduais, N., J. Abdullah, A. Jamil, and H. Heidari, Performance evaluation of real-time multivariate data reduction models for adaptive-threshold in wireless sensor networks. *IEEE sensors letters*, 2017. 1(6): p. 1-4.
- [2] Moreno, M.V., J. Santa, M.A. Zamora, and A.F. Skarmeta. A holistic IoT-based management platform for smart environments. in 2014 IEEE International Conference on Communications (ICC). 2014. IEEE.
- [3] Lengyel, L., P. Ekler, T. Ujj, T. Balogh, and H. Charaf, SensorHUB: an IoT driver framework for supporting sensor networks and data analysis. *International Journal of Distributed Sensor Networks*, 2015. 11(7): p. 454379.
- [4] Noori, a., et al., Wireless Sensor Network Deployment Based On Machine Learning For Prolonging Network Lifetime And pdr. *Journal of Theoretical and Applied Information Technology*, 2019. 97(14).
- [5] Luo, J., D. Wu, C. Pan, and J. Zha, Optimal energy strategy for node selection and data relay in WSN-based IoT. *Mobile Networks and Applications*, 2015. 20(2): p. 169-180.
- [6] Dias, G.M., B. Bellalta, and S. Oechsner, A survey about prediction-based data reduction in wireless sensor networks. *ACM Computing Surveys (CSUR)*, 2016. 49(3): p. 1-35.
- [7] Maraiya, K., K. Kant, and N. Gupta, Wireless sensor network: a review on data aggregation. *International Journal of Scientific & Engineering Research*, 2011. 2(4): p. 1-6.
- [8] Rohankar, R., C. Katti, and S. Kumar, Comparison of energy efficient data collection techniques in wireless sensor network. *Procedia Computer Science*, 2015. 57: p. 146-151.
- [9] Minet, P., Energy efficient routing. *Ad Hoc and Sensor Wireless Networks: Architectures: Algorithms and Protocols*, 2009.
- [10] Luo, S., Y. Sun, and Y. Ji, Data collection for time-critical applications in the low-duty-cycle wireless sensor networks. *International Journal of Distributed Sensor Networks*, 2015. 11(8): p. 931913.
- [11] Raza, U., A. Camera, A.L. Murphy, T. Palpanas, and G.P. Picco, Practical data prediction for real-world wireless sensor networks. *IEEE Transactions on Knowledge and Data Engineering*, 2015. 27(8): p. 2231-2244.
- [12] Bröring, A., et al., New generation sensor web enablement. *Sensors*, 2011. 11(3): p. 2652-2699.
- [13] Shujaa, m., lagged multi-objective jumping particle swarm optimization for wireless sensor network Deployment. *Journal of Theoretical and Applied Information Technology*, 2019. 97(2).
- [14] Hammid, A.T. and M.H.B. Sulaiman, Series division method based on PSO and FA to optimize Long-Term Hydro Generation Scheduling. *Sustainable Energy Technologies and Assessments*, 2018. 29: p. 106-118.
- [15] Hammid, A.T., M.H.B. Sulaiman, and O.I. Awad, A robust firefly algorithm with backpropagation Neural networks for solving hydrogeneration prediction. *Electrical Engineering*, 2018. 100(4): p. 2617-2633.
- [16] Alduais, N.A.M., J. Abdullah, and A. Jamil, RDCM: An Efficient Real-Time Data Collection Model for IoT/WSN Edge With Multivariate Sensors. *IEEE Access*, 2019. 7: p. 89063-89082.
- [17] Alduais N. A. M., J. Abdullah, A. Jamil, and L. Audah., An Efficient Data Collection and Dissemination for IOT Based WSN. In *Proceeding of the 7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference*, IEEE, Vancouver, 2016: p. 1-16.
- [18] Yu, S. Arduino based Balancing Robot. in 2019 International Conference on Electronics, Information, and Communication (ICEIC). 2019. IEEE.
- [19] Al-Fagih, A.E., A framework for data delivery in integrated Internet of Things architectures. 2013: Queen's University (Canada).
- [20] Chaudhary, S., N. Singh, A. Pathak, and A. Vatsa, Energy Efficient Techniques for Data aggregation and collection in WSN. *International Journal of Computer Science, Engineering and Applications*, 2012. 2(4): p. 37.
- [21] Reinhardt, A., D. Christin, M. Hollick, and R. Steinmetz. On the energy efficiency of lossless data compression in wireless sensor networks. in 2009 IEEE 34th Conference on Local Computer Networks. 2009. IEEE.
- [22] Quwaidar, M. and Y. Jararweh, Cloudlet-based efficient data collection in wireless body area networks. *Simulation Modelling Practice and Theory*, 2015. 50: p. 57-71.
- [23] Beaulah, H.L., K. Thangaraj, and M. Chitra, Design of a New Energy Efficient L4 Leach Protocol based Visual Sensor Network for Forest Monitoring System. *Asian Journal of Research in Social Sciences and Humanities*, 2016. 6(7): p. 662-671.
- [24] Palpanas, T., Real-time data analytics in sensor networks. in *Managing and Mining Sensor Data*. 2013, Springer. p. 173-210.
- [25] Guo, Y., F. Hong, Z. Guo, Z. Jin, and Y. Feng. EDA: Event-oriented data aggregation in sensor networks. in 2009 IEEE 28th International Performance Computing and Communications Conference. 2009. IEEE.
- [26] Bahi, J.M., A. Makhoul, and M. Medlej. Data aggregation for periodic sensor networks using sets similarity functions. in 2011 7th International Wireless Communications and Mobile Computing Conference. 2011. IEEE.
- [27] Pfletschinger, S., M. Navarro, and C. Ibars. Energy-efficient data collection in WSN with network coding. in 2011 IEEE GLOBECOM Workshops (GC Wkshps). 2011. IEEE.
- [28] Bahi, J.M., A. Makhoul, and M. Medlej. Frequency filtering approach for data aggregation in periodic sensor networks. in 2012 IEEE Network Operations and Management Symposium. 2012. IEEE.
- [29] Enam, R.N., R. Qureshi, and S. Misbahuddin, A uniform clustering mechanism for wireless sensor networks. *International Journal of Distributed Sensor Networks*, 2014. 10(3): p. 924012.
- [30] Laiymani, D. and A. Makhoul. Adaptive data collection approach for periodic sensor networks. in 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC). 2013. IEEE.
- [31] Enam, R.N. and R. Qureshi. An adaptive data aggregation technique for dynamic cluster based wireless sensor networks. in 2014 23rd International Conference on Computer Communication and Networks (ICCCN). 2014. IEEE.
- [32] Harb, H., A. Makhoul, D. Laiymani, A. Jaber, and R. Tawil. K-means based clustering approach for data aggregation in periodic sensor networks. in 2014 IEEE 10th international conference on wireless and mobile computing, networking and communications (WiMob). 2014. IEEE.

- [33] Li, X.-Y., Y. Wang, and Y. Wang, Complexity of data collection, aggregation, and selection for wireless sensor networks. *IEEE Transactions on Computers*, 2010. 60(3): p. 386-399.
- [34] Carlos-Mancilla, M., E. López-Mellado, and M. Siller, Wireless sensor networks formation: approaches and techniques. *Journal of Sensors*, 2016. 2016.
- [35] Aderohunmu, F.A., D. Brunelli, J.D. Deng, and M.K. Purvis, A data acquisition protocol for a reactive wireless sensor network monitoring application. *Sensors*, 2015. 15(5): p. 10221-10254.
- [36] Rault, T., A. Bouabdallah, and Y. Challal, Energy efficiency in wireless sensor networks: A top-down survey. *Computer Networks*, 2014. 67: p. 104-122.
- [37] C. B. Mudgule, U.N., & P. D. Ganjewar, Data Compression in Wireless Sensor Network: A Survey. *International Journal of Innovative Research in Computer and Communication Engineering*, 2014. 2: p. 1-10.
- [38] Pandey, I.K., M. Natarajan, and S. Kaur-Ghumaan, Hydrogen generation: aromatic dithiolate-bridged metal carbonyl complexes as hydrogenase catalytic site models. *Journal of inorganic biochemistry*, 2015. 143: p. 88-110.
- [39] Hung, C.-C., W.-C. Peng, and W.-C. Lee, Energy-aware set-covering approaches for approximate data collection in wireless sensor networks. *IEEE Transactions on Knowledge and Data Engineering*, 2011. 24(11): p. 1993-2007.
- [40] Sarisaray-Boluk, P. and K. Akkaya, Performance comparison of data reduction techniques for wireless multimedia sensor network applications. *International Journal of Distributed Sensor Networks*, 2015. 11(8): p. 873495.
- [41] Wu, M., L. Tan, and N. Xiong, Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications. *Information Sciences*, 2016. 329: p. 800-818.
- [42] Ait-Sahalia, Y. and D. Xiu, Principal component analysis of high-frequency data. *Journal of the American Statistical Association*, 2019. 114(525): p. 287-303.
- [43] McCorrie, D.J., E. Gaura, K. Burnham, N. Poole, and R. Hazelden, Predictive data reduction in wireless sensor networks using selective filtering for engine monitoring, in *Wireless Sensor and Mobile Ad-Hoc Networks*. 2015, Springer. p. 129-148.
- [44] de Carvalho, C.G.N., D.G. Gomes, J.N. de Souza, and N. Agoulmine. Multiple linear regression to improve prediction accuracy in WSN data reduction. in *2011 7th Latin American Network Operations and Management Symposium*. 2011. IEEE.
- [45] Cheng, S., Z. Cai, and J. Li, Approximate sensory data collection: a survey. *Sensors*, 2017. 17(3): p. 564.