International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Data Reduction Techniques in Wireless Sensor Networks with Internet of Things

[1]Dr. S. Venkatesh, [2]Shiv Prasad Kori, [3]P. William, [4]Dr. M. L. Meena, [5]A. Deepak, [6]Dr. Dler Salih Hasan, [7]Dr. Anurag Shrivastava

Submitted: 12/10/2023 **Revised**: 28/11/2023 **Accepted**: 08/12/2023

Abstract: Data is sensed and routed by nodes, which are sometimes referred to as sensors and sinks, in a wireless sensor network using a number of protocols that alter based on the intended function of the network. These protocols vary from network to network. the use of wireless sensor networks (WSN) in a variety of different sectors, including the military, healthcare facilities, medical equipment, environmental monitoring, and other areas. In this piece, we focused our attention largely on the energy consumption of sensor nodes and the redundant storage of data. Both the number of items that are connected to the Internet of items (IoT) and the amount of data that these connected devices send will continue to quickly increase. WSN-based sensor nodes (SNs) generate some Internet of Things data and transmit it to gateways (GWs), which leads the sensor nodes to soon run out of both energy and storage space as a result of the data transfer. The majority of the approaches that have been proposed are only capable of decreasing data at a certain level of an Internet of Things architecture, such as at gateways. The Two-Tier Data Reduction (TTDR) method is strongly urged to be used by both the sensor nodes and the gateway, which, individually, stand for the network's top and bottom tiers, respectively. This results in a gradual reduction in the total number of data sets. In the end, the effectiveness of the TTDR is evaluated by utilising the OM Net++ simulator in conjunction with real sensory data. The findings that were acquired demonstrate how successful the strategy that was recommended was at both transferring data and utilising energy.

Keywords: *Wireless Sensor Networks, Internet of Things, Sensor Node, Data Compression, MDL*

1. Introduction

WSN is a relatively new technology that has the potential to be used in a variety of settings, such as environmental monitoring, surveillance, medical systems, robotic exploration, and the armed forces. A wireless sensor network (WSN) is comprised of individual nodes, each of which has a limited amount of computational power, storage capacity, and communication bandwidth. As a result, the individual nodes' access to resources is effectively restricted. After deployment, it is the responsibility of the sensor nodes to self-organize an appropriate network architecture and often establish connections with other nodes in the network via a sequence of hops. After then, the sensors already on board will begin to collect crucial data. WSNs are constructed from a number of dispersed nodes that collaborate with one another to create a number of different multi-hop wireless networks. Each mode will typically need batteries, but it may also include a central processing unit, a sensor, or low-power radios depending on its specific configuration. There is potential for cost and size differences associated with sensor nodes. A sensor node is often a piece of hardware that combines sensing, which means it can collect data from its surrounding environment, processing, which means it can analyse and store data locally, and communication, which means it can communicate data. Sensing allows the device to acquire data from the environment. Processing allows the device to analyse and store data locally. Academics have started to take notice of Wireless Sensor Networks (WSNs) as a direct result of the rapid development of wireless technology and embedded electronics. A typical wireless sensor network (WSN) is built from minuscule building blocks known as nodes [1].

These nodes are equipped with a central processing unit (CPU), a limited amount of processing power, and many sophisticated sensors. These sensors are used by nodes in order to monitor a variety of environmental factors,

[1]Assistant Professor, Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, kattankulathur, Chennai
venkates9@srmist.edu.in
[2]HOD & Department of Electronics and Telecommunication, Government Polytechnic College, Raisen, Madhya Pradesh
shivprasadkori@gmail.com
[3]Department of Information Technology, Sanjivani College of Engineering, Savitribai Phule Pune University, Pune
william160891@gmail.com
[4]Assistant Professor, Department of Electronics Engineering, Rajasthan Technical University, Kota
mlmeena@rtu.ac.in
[5]Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu
**deepakarun@saveetha.com*
[6]Computer Science and Information Tech., College of Science / University of Salahaddin-Erbil, Iraq
dler.hasan@su.edu.krd
[7]Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences,
Chennai, Tamilnadu
Anuragshri76@gmail.com

including temperature, pressure, vibration, and humidity. A typical wireless sensor network (WSN) node has the following components: a transceiver, a processor unit, and a power unit. These devices carry out key activities such as allowing nodes to communicate with one another and relaying data acquired by their sensors. Communication between the system's nodes is essential for the operation of centralised systems. Because of the need for this system, the concept of the internet of things, sometimes known as the IoT, came into being. The Internet of Things makes it feasible to have fast access to data on the surrounding environment. As a result, dramatically increased levels of productivity and efficiency are achieved across a wide range of processes [2].

This article presents a comprehensive analysis of wireless sensor networks (WSNs). The objectives of this study are to analyse the technology and characteristics of WSNs, investigate the applications of WSNs, and give information on the challenges and opportunities associated with WSNs. The definition of WSNs may be found at the beginning of Section 2, along with an explanation of their structure. The third section offers an account of the history of WSNs, while the fourth section discusses how these networks function. The advantages and disadvantages of wireless sensor networks.

It is common practise to define a WSN as a network of nodes that collaborate in order to jointly observe and manage their local environments. These nodes are linked together via the usage of wireless media. The nodes are able to communicate with one another by using this connection in their interactions. The architecture of a typical WSN is comprised of the following three fundamental components: sensor nodes, gateways, and observers (users). Gateways and sensor nodes are the components that make up the sensor field. Gateways and observers are connected to one another by means of specialist networks or, more often, the internet (please refer to Figure 1 for more explanation) [2].

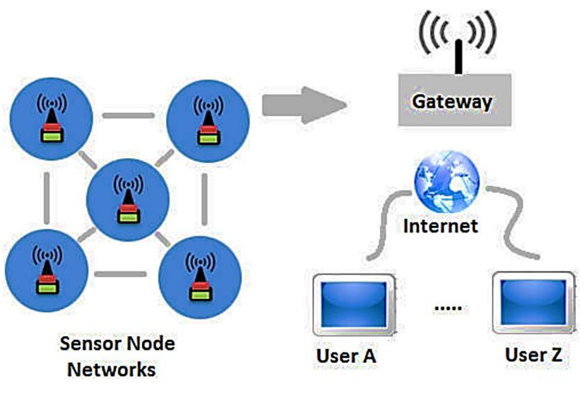

Fig.1: Wireless Sensor Network (Wsn)

If one were to believe the formula, Sensing plus processing power plus radio equals plenty of potential. A sensor device is necessary in order to monitor environmental elements such as vibration, pressure, and humidity. After the activities of monitoring and sensing have been completed, the central processing unit will carry out the necessary computations. Last but not least, the Radio Unit is responsible for sending computed environmental data to the other nodes in the network through wireless communication channels. At long last, these data are sent to the Gateway [3].

2. Review of Literature

Data is sensed and routed by nodes, which are sometimes referred to as sensors and sinks, in a wireless sensor network using a number of protocols that alter based on the intended function of the network. These protocols vary from network to network. The use of wireless sensor networks (WSN) in a variety of domains, such as the military, hospitals, medical equipment, environmental monitoring, and so on. A few of the shortcomings of the WSN are its short lifespan, poor battery life, low bandwidth, high energy consumption, redundant data storage, and inefficient routing. The redundant data storage and the energy usage of sensor nodes were the primary focuses of our attention throughout the whole of this piece. The process of data reduction is one of the preprocessing methods for data mining that may potentially save money and enhance the efficiency of storage. The purpose of data reduction, often known as DR, is to get rid of unnecessary data during the transmission process. In order to accomplish this aim in a manner that is consistent with the WSN environment, a great number of data reduction solutions have been created. This paper provides an introduction to contemporary data reduction-based algorithms and tactics that aid in extending the lifespan of networks and reducing the amount of energy that is used [3].

As wireless technology and digital electronics continue to progress, more and more parts of daily life are beginning to make use of a variety of different kinds of small devices. These devices are able to perceive their surroundings, compute, and interact with one another. In most cases, they are composed of a small number of intelligent sensors, low-power radios, and embedded CPUs (Central Processing Units). These devices are put to use in the formation of a wireless sensor network (WSN), which is essential for the provision of sensing services and the monitoring of one's surrounding environment. Concurrently with the growth of wireless sensor networks (WSNs), the idea of an internet of things (IoT) is being established. The IoT may be defined as a link between identifiable entities that are located inside an internet connection and participate in sensing and monitoring activities. This paper devotes considerable attention to the topic of WSNs. In addition to this, it conducts an analysis of the characteristics and capabilities of WSNs. In addition to this, it provides a review of the applications of WSN and the Internet of Things [4].

The Internet is now going through a transition in which it is moving seamlessly from an Internet of People to an Internet of Things (IoT). It is estimated that consumers will have access to fifty billion things over the internet by the year 2020. Because of this mobility, it is more difficult to manage the interoperability of the many components that make up the Internet, such as RFIDs (Radio Frequency Identification), portable mobile devices, and wireless sensors. In the past, this industry has been responsible for the development of a number of different protocols, such as IPv6, 6LoWPAN (IPv6 over Low power Wireless Personal Area Networks), and M2M (Machine to Machine communications). In this research, we focus on the challenges that are involved with integrating wireless sensor networks into the Internet of Things (IoT), and we elucidate those challenges. We utilise wireless sensors to regulate electrical appliances that are located inside of a smart building so that we can illustrate how a test bed for the real world is produced. The difficulties that have been experienced are stressed, and suitable solutions are provided [5].

Recent research has shown that Internet of Things (IoT)-based wireless technologies have advanced rapidly in a variety of business sectors. The Internet of Things (IoT) is a network that allows communication to take place between physical objects, machines, sensors, and other pieces of technology without the need for human intervention. The Internet of Things (IoT), which is often referred to as a wireless sensor network (WSN), is a collection of real-time applications that is quickly increasing. The abbreviation for "wireless sensor network" is "WSN." Wireless sensor networks (WSNs) and the Internet of Things (IoT) both have a multitude of uses, with some of those applications being very important to our day-to-day lives and others having a lesser impact. The nodes that make up a WSN are often very small devices that run on batteries. Therefore, methods of data aggregation that are both energy efficient and help prolong the life of the network are quite significant. Several different approaches and algorithms have been proposed in order to achieve energy-efficient data aggregation in IoT-WSN systems. The concerns of wireless networking for the collection of data and the conservation of energy are the primary focus of this review of the relevant literature [6].

Both the number of items that are connected to the Internet of items (IoT) and the amount of data that these connected devices send will continue to quickly increase. WSN-based sensor nodes (SNs) generate some Internet of Things data and transmit it to gateways (GWs), which leads the sensor nodes to soon run out of both energy and storage space as a result of the data transfer. The cheap cost of SNs is a factor that limits both their energy capacity and their storage capacity. It is recommended that the quantity of data stored at the source nodes be kept to a minimum in order to alleviate these concerns. This will reduce the amount of storage space required as well as the amount of energy that is used. The majority of the approaches that have been proposed are only capable of decreasing data at a certain level of an Internet of Things architecture, such as at gateways [7]. The Two-Tier Data Reduction (TTDR) method is strongly urged to be used by both the sensor nodes and the gateway, which, individually, stand for the network's top and bottom tiers, respectively. We provide a data compression method that is both straightforward and efficient for use by sensor nodes that are part of the Internet of Things and have limited resources at their disposal. The approaches take use of Run-Length Encoding (RLE) following Delta Encoding in order to optimise the temporal correlation that is present in sensor data. This is accomplished by encoding the data in run lengths. In order to structure the data sets in accordance with the Minimum Description Length (MDL) concept, we make use of hierarchical clustering at the gateway layer. These data sets were made available by the sensor nodes. If the MDL concept can be used to compress any of the possible combinations of the data sets that are being input, then a cluster will be produced as a result. This leads to a gradual decrease in the number of data sets, and the clustering of sets is stopped if it is difficult to locate a match between sets that can be compressed in the same way. This results in a gradual reduction in the total number of data sets. At the end of the day, the OM Net++ simulator and real sensory data are what are utilised to evaluate how effective the TTDR was. The findings that were acquired demonstrate how successful the strategy that was recommended was at both transferring data and utilising energy [8].

Since the nodes that make up an IoT sensor network each operate on their own limited batteries, these networks often include methods for conserving energy. Although processing data takes a far less amount of energy than transmitting data, both activities are highly costly and need a significant amount of energy in IoT sensor nodes. There are several strategies and concepts available for conserving energy, the majority of which focus on reducing the amount of data transfer. This indicates that large power savings in IoT sensor networks may be realised by restricting the amount of data that is delivered. Based on the findings of this research, a compression-based data reduction (CBDR) approach was suggested. This method functions at the level of Internet of Things sensor nodes. The compression method used by the CBDR is a two-step process [9]. During the first stage, the lossy SAX Quantization stage narrows the dynamic range of sensor data. During the second stage, the lossless LZW stage compresses the output of the loss quantization stage. On each of these levels, there is no risk of losing anything. It is beneficial to quantize the data readings from the sensor nodes to the size of the SAX alphabet in order to obtain the optimum compression sizes; nevertheless, from the point of view of LZW, this leads in a poorer compression ratio. It was also suggested that the CBDR strategy be improved by including Dynamic Transmission (DT-CBDR), which would cut down on both the total volume of data that would be supplied to the gateway as well as the amount of processing that would be required. Actual sensory data that was gathered at Intel Lab is integrated with the OM Net++ simulator so that the usefulness of the strategy that was presented may be shown. The results of the simulation tests demonstrate that the CBDR technique is superior to the other strategies that are discussed in the relevant literature [10].

3. Data Reduction Techniques

❖ **Adaptive Filter:**

A piece of technological equipment known as a filter is used to remove information from a signal that is deemed to be unnecessary. Both stationary and non-stationary filters are considered to be types of filters. Since stationary filters do not allow for the alteration of their component, monitoring temporal fluctuation and adjusting the coefficient requires non-stationary filters. Stationary filters also do not allow for the modification of their component. The adaptive filters work according to the same principles. It employs techniques such as LMS (Least Mean Square), Kalman Filter, LMS-VSS (LMS Variable Step-Size), LMS-SSA, and Extended Kalman Filter in order to provide correct forecasts as quickly as possible and enhance its accuracy by adjusting its coefficients. Additionally, it does this in order to make its predictions more accurate. When used for filtering in a WSN, LMS-SSA (also known as LMS Step Size Adaption) yields more accurate results while requiring fewer computations than other methods. In addition to this, prior knowledge of correlation functions or matrix inversions is not required in order to use it in any way. It is suitable for use in settings with relatively tiny sensor nodes, such as those seen in WSN contexts. Through manipulation of the filter settings, prediction in WSN is able to lessen the amount of data sent by sensor nodes. The data will be supplied together with certain prediction values, and these values will be compared to the counter values of the other data. If there is a disparity, the sink will send out an alarm message, and both ends will go into standalone mode until the disparity is rectified. During this time, the sink will continue to monitor the situation. However, the error must be lower than e max (where e max is the highest prediction error that the filter can accept). When a certain prediction error happens, the LMS will transition from normal to standalone mode. Additionally, the adaptive filter may be used in conjunction with the Dual Prediction framework, which requires both the sensor and the sink to possess prediction models. It forecasts a reduction in the amount of communication that takes place between nodes and sinks, as well as in the number of local readings taken by sensor nodes.

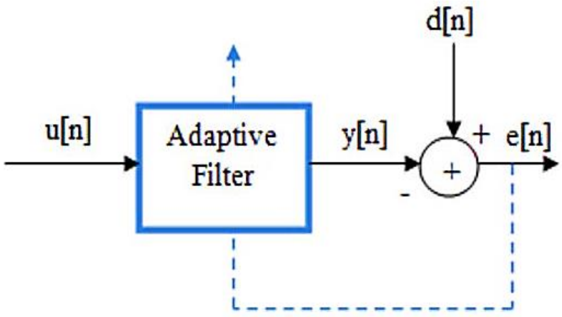

Fig 2: General Adaptive Filter

Predictions may be produced in both time and space by using a pre-defined model, the parameters of which are dependent on prior knowledge or historical data. This allows for the prediction of future events. This allows for the creation of forecasts. The LMS was based on a Field Programmable Gate Array, which was done so in order to reduce the amount of communication that took place between the sensor nodes and the base station.

❖ **Tree Based Methods:**

In addition to that, it makes advantage of in-network data reduction on occasion. There are a few different tree-based data reduction methods that may be used in WSN. When it comes to mobility, we may use energy consumption models that are dependent on distance, such as Khepera, Robomote, LocalPos, and FIRA. To calculate the amount of energy required to travel a distance of d, an equation known as EM(d) = kd is used. There is a correlation between the speed of the node and the value of the parameter k. A graph called a spanning tree is one that does not include any cycles and uses each of the nodes as a vertex. SBTT and E-span are two further ways of spanning trees that are cognizant of energy use. On the other hand, distance-based approaches do nothing more than save energy and are completely worthless when it comes to data compression.

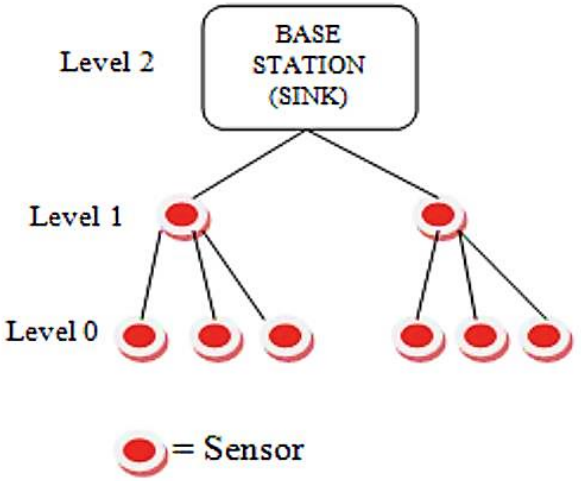

Fig 3: Tree Based Aggregation

Because of this, the MNDGT (Minimum Node Data Gathering Tree) approach, which is based on association rule mining, provides a higher level of energy efficiency for WSN in-network data reduction. Association rule mining, which participates actively in the process of resource management, contributes to an improvement in the Quality of Service (QoS) in wireless sensor networks. The sensor association rules, sometimes known as SAR for short, are what are responsible for capturing the temporal link between the sensors. On top of the in-network DR (data reduction) system sits the data collecting tree, which will be used to transfer data to the washbasin after it is finished being created. Discovering patterns that occur repeatedly may be accomplished effectively with SAR. In addition to that, it made use of a variety of methods for data collecting, such as LEACH, PEGASIS, Chain-based tree-level Schema, and so on; however, these strategies do not take into account the possibility of repeated sensor activity.

- **Cluster-Head (CH) Based Reduction:**

CHR is essential for lowering the amount of energy that is used by WSNs, increasing the life of the network, and improving its scalability. Clustering is vital for high-density networks because it makes the management of a group of cluster heads (CH) from each cluster more easier than the management of individual nodes. Taking the positions of the nodes as their point of departure, a number of algorithms, notably CACC, partition the region being monitored into cells (often hexagonal cells). Each cluster is composed of around 6-7 cells. In a sensor network operating in a heterogeneous environment, the VAP-E protocol is implemented whenever a greater amount of transmission power from each node varies and relies on VAP. PEZCA, VoGC, BARC, KOCA, CFL, EECS, EEUC, FoVs, and PDCH are some further examples of algorithms that save energy. However, the LEACH distributed algorithm (Low Energy Adaptive Clustering Hierarchy) and its numerous offspring are the ones that are used the most often in wireless sensor networks owing to the efficiency with which they lower the amount of power that is consumed. The generation of CHs is only the responsibility of a select few sensor nodes in the GROUP, and these nodes HAS. Harmony Search Algorithms are used by musicians in order to limit the intra-cluster distance of the network and continually improve the pitches. All of these are examples of cluster algorithms that are energy-dependent.

There are three separate stages, as follows: NACHO: During initialization, the base station will calculate the key parameters before sending a message to each and every node. This will cause each node to transmit a list of the other nodes that are in its immediate vicinity. The primary components that are utilised to construct the cluster are referred to as NRE (Node Residual Energy), centrality, CH-frequency, and concentration. The data transmission phase starts when CH receives data and passes the resulting package to the sink. This is when the phase really begins. At this time, it may be said that the phase is already in progress. This makes it feasible to maintain track of the energy level, and when a CH passes away, it causes all of the nodes in the network to detach from their CH by sending out a re-cluster message. This makes it possible to keep track of the energy level. NACHO is hence able to generate better clusters, enhance energy, and improve packet reduction, in contrast to all other methods.

- **Data Stream Based Reduction:**

When information is sent to the internet in a manner that is not sequentially organised, a method known as "data streams" is used to handle the information. When real-time applications running on WSN encounter delays, it is sometimes necessary to reduce the number of data streams being sent. Therefore, reduction must take place during transmission in order for there to be less of a delay in the delivery of data in wireless sensor networks. The use of sampling techniques for data streams is strongly recommended in order to cut down on the amount of data that must be sent.

WSN is used by Data Stream Reduction (DSR), also known as stream processing, as a distributed database while it is calculating functions and utilising resources from the sensor database or buffer. In order for us to get delay metrics, sample size impacts, and other parameters, among others, from the appropriate approaches, each of these tactics has to be implemented in real-time applications. An example of a real-time application that was developed utilising data reduction is shown in Figure 5. Sensor nodes are responsible for the collection of data streams from wireless settings, the classification of these streams via the use of stream organisers, and the determination of a sample size or method for creating samples through the use of stream processing.

➢ ***Sampling:***

Using this method, the best possible data quality is achieved while taking into account the requirements of the network in terms of energy, routing, bandwidth, and so on. Using this method, a histogram of the sensor data is first generated. Subsequently, samples are generated using a random method, and then they are based on the histogram. After that, the data samples will be arranged in a hierarchy that corresponds to their position in the actual data stream. After reduction, research indicates that the complexity of communication is O(md), where d is the greatest hop in a WSN; O(n) is the number of samples; and O(md) is also the complexity of the overall system. Nonetheless, the method relies on the following premises: a random topology, an EF-Tree, stream creation, a sufficient sample size, and NS-2 simulation. This technique fulfils all of the real-time application requirements, meets the RTS deadline, consumes less energy, keeps the data quality intact, and enables judgements on data reduction.

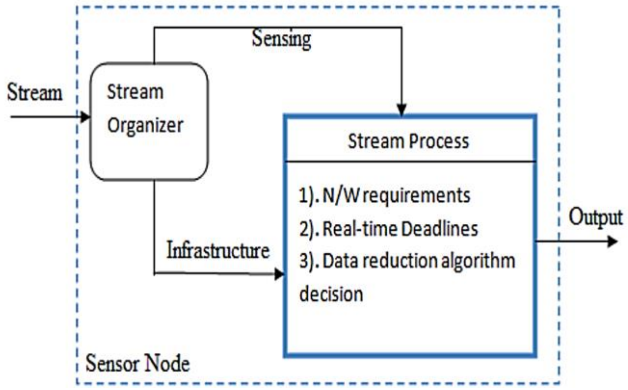

Fig 4: Real Time Application

❖ **Hybrid Data Reduction (HDR):**

Data driven reduction, event driven reduction, and time driven reduction are the three fundamental layers that make up the many methods to data reduction. Data driven reduction is used when particular data management is required, event driven reduction is used when certain major changes in sensed parameters are present, and time driven reduction is utilised when sensors frequently detect and send the data. However, real-time applications, which are essentially event- and data-driven, still make use of time-driven reduction approaches. These methodologies are used in applications such as fire systems and warning systems. For this reason, the process for collecting the data makes use of reduction techniques known as HDR (Hybrid Data Reduction), which take into consideration all three components.

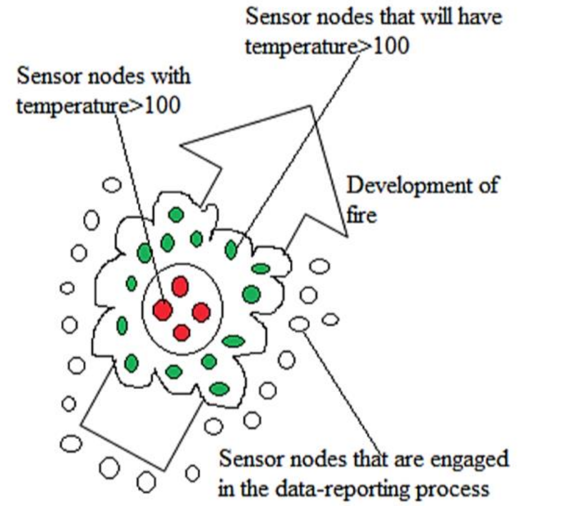

Fig 5: Hybrid Wsn

In order to reduce the amount of energy used and the amount of data that is sent, the hybrid data gathering protocol (HDG) permits switching between data driven and event driven data reporting (DDDR). In the future, we are going to collect information about incidents by employing a data reporting mechanism that is based on WSN applications such as fire and alarm systems. Cluster-based routing and other similar technologies are now available for use in this context. The whole network is broken up into clusters, the data is compressed, and messages are transmitted to the sink; nevertheless, the key challenge is selecting the appropriate cluster or CH (Cluster Head), which will result in increased energy consumption and a decreased lifespan for the network.

❖ **Data Prediction Based Reduction:**

The process of estimating a value by making use of a number of different methods and then picking the suitable data is referred to as data prediction. There are a wide variety of distinct sorts of algorithms that may help with the reduction of data via prediction. The approach known as Least Mean Square (LMS) is often used as an example due to the fact that it is straightforward and does not include any complicated steps. The use of it in conjunction with linear filters, such as adaptive filters, is also shown in the previous section.

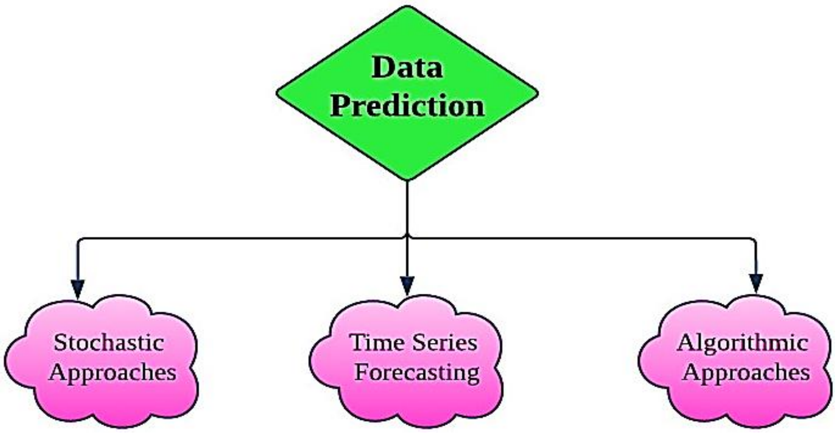

Fig 6: Data Prediction for Energy Management

4. Research Methodology

In this part, we will talk about the method that is often suggested for the data reduction technology that is used in IoT sensors. The sensor node tier and the GW tier are the two essential steps that are included in this method to data reduction that is based on compression and MDL. The recommended effort to increase energy efficiency has as its primary objective the reduction of the amount of data that is sent. Encoding schemes such as Delta encoding and RLE encoding are applied for the purpose of data compression at the first layer of the system. According to the MDL principle, GW will cluster the data sets that it gets from sensor nodes that are part of the second tier. As a consequence of this, each and every combination of incoming data sets that may be compressed by making use of the MDL principle will be consolidated into a single cluster. The techniques that are offered will be broken down and examined in further detail in the sections that follow. Due to the fact that Internet of Things applications need a great number of sensor nodes, the cost of the sensor node, namely in terms of energy, has to be as low as is practically possible. Moving forward, we are going to proceed on the basis that the sensor node i obtains a new reading of the data each and every time that a time interval s elapses. After the completion of each period, the sensor node i will generate a data set DS, which is an example of time-series data, and then it will transmit this information to the GW. This will be done so that analysis can take place. In the event that the time interval s is inadequate, the sensor node will repeatedly record the same data (or data that is very similar to the original data). The quantity of data that is sent across the network will be cut down by making advantage of the temporal correlation of the data readings that occur inside each sensor node. The conventional data compression methods, such as RLE encoding and Delta encoding, can be used to effectually alter our circumstance if we so like.

According to TTDR, the ecosystem of the Internet of Things is made up of a wireless sensor network (WSN) that is comprised of a range of sensor nodes, each of which transmits data to the Gateway (GW), which is a specific data sink. The datasets that were analysed as part of this investigation were given by the Intel Laboratory. the time-stamped measurements of light, temperature, relative humidity, and voltage that are included as part of the meteorological data that is given by the databases maintained by Intel. These data are a great illustration of a time series that may be used for Internet of Things applications. Significant temperature-based simulation experiments were done out making use of the OM Net++ simulation. When analysing a compression method, the compression ratio is the only factor that is often taken into consideration. When compressing data in IoT sensor nodes, a number of extra considerations need to be taken into account. In this section, we will go through a range of measurements that give a full evaluation of the performance of compression algorithms when applied to IoT sensor nodes. These nodes collect data from other IoT devices. The effectiveness of TTDR is evaluated based on the variables presented in Table 1.

Table 1: Parameters Settings

Parameter	Value
WSI size	49 sensors
periods of reading	data detected at 20, 50, and 100
K for Cardinality	8-bit (3-value)
E_{etec}	50 nj/bit
F_{AMP}	A 100 Pj/bit/m2

- **SAX Quantization**

In order for the LZW technique to work on the received data readings that are supplied by IoT sensor nodes (which represent an ideal paradigm of a timeseries data), some type of time series preparation is necessary. In order to proceed with the analysis, the time series that are reflective of the data readings from the different IoT sensor nodes need to be converted into a format that can be used. When managing time series, you may want to take into consideration using both a symbolic representation and a normalisation as two different ways. The extensive use of symbolic representation may be attributed to a number of different variables. One of these characteristics is the use of techniques derived from a variety of different fields, such as text processing, information retrieval, or bioinformatics. Additional considerations include the effectiveness, readability, and straightforwardness of the time series representation. Symbolic Aggregate Approximation, more often referred to by its acronym SAX, is widely regarded as one of the most powerful approaches to symbolically representing data. The piecewise aggregate approximation (PAA) transformation and the translation of the numerical input to a collection of symbols are the two components that come together to generate the SAX algebraic expression. Piecewise aggregate approximation is what the letters "PAA" stand for in the acronym. In this particular investigation, just the second aspect of the SAX scale is relevant. It is possible to apply the SAX method in order to translate the normalised data readings obtained from Internet of Things sensor nodes into a symbolic representation. In order to accomplish this conversion, the SAX quantization makes use of breakpoints of the number N 1, which split the area that is covered by the Gaussian distribution into regions that are proportional to one another. A breakpoint is a sorted series of values that begins with the value B = 1... and ends with the value 1. The area under a Gaussian curve with parameters N (0, 1) is equal to 1/a from point I to point i+1, where 0 and a, respectively, stand for I and I. From here to i+1, the scenario looks like this. If you search for the breakpoints in a statistical table, you should be able to discover them. Table 2 presents a lookup table of the breakpoints for the values in the range of three to ten, which can be found in the previous table.

Table 2: A Lookup Table For Breakpoints For A Variety Of Values

	a	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
β1	β1	-0.43	-0.56	-0.87	-0.88	-1.18	-1.28	-1.33	-1.23
	β2	0.54	0	-0.43	-0.54	-0.54	-0.78	-0.88	-0.23
	β3		0.76	0.34	0	-0.34	-0.33	-0.44	-0.54
	β4			0.96	0.74	0.21	0	-0.23	-0.23
	β5				0.89	0.58	0.38	0.21	0
	β6					1.17	0.79	0.87	0.24
	β7						1.45	0.98	0.58
	β8							1.23	0.89
	β9								1.36

5. Analysis and Interpretation

The first experiment acts as a recorder for the system's activities. During this experiment, each mote reads its sensory input once per second and creates WSN communication. We are able to measure the delay and monitor how it varies in relation to the amount of traffic. In principle, the routing protocol gives precedence to routing control messages over data packets when determining which to process first. As a consequence of this, this priority may have an effect on the latency experienced across the network.

❖ **Obtained Results in the Sensor Node Tier**

In the first tier (sensor node), the recommended compression approach, which is based on Delta encoding followed by RLE encoding, is explored and compared with ATP and PFF methods. These are the studies that we have decided to look at since they make use of the same data set, have an idea that is sufficiently comparable to ours, and we are able to acquire the findings of these studies directly from the researchers who conducted them. The performance of anything is evaluated by determining what percentage of the original data is still accessible after aggregation and compression, what percentage of the original information is passed on to the subsequent layer, and how much energy is used. Figure 9 illustrates the percentage of unused data for each Internet of Things sensor node, both with and without the use of data aggregation and compression using a variety of methods (including TTDR, ATP, and PFF). According to the findings, applying compression techniques to TTDR will result in a maximum remaining data reduction of 20.53 percent, applying aggregation to the ATP protocol will result in a maximum remaining data reduction of 31.68 percent, and applying neither will result in a maximum remaining data reduction of 100 percent, as is the case with PFF.

Table 3: Percentage of Readings After Applying Compression

Readings / period	ATP (0.03)	ATP (0.05)	ATP (0.07)	TTDR	PFF
20	35	23	21	20	21
50	33	17	15	18	19
100	25	8	10	16	20

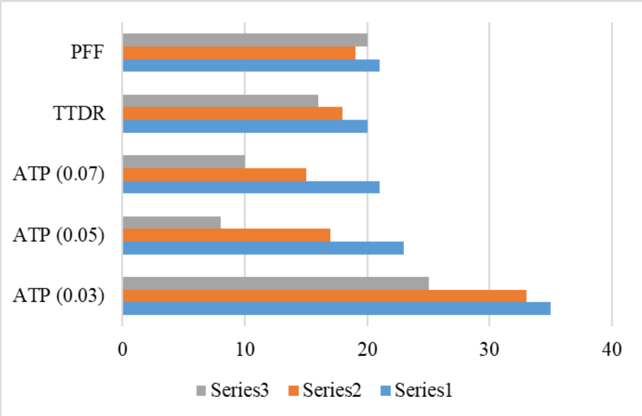

Fig 7: Percentage of Readings After Applying Compression

Figure 7 illustrates the proportion of data sets that were sent to the subsequent tier by IoT sensor nodes making use of a variety of methodologies (namely, TTDR, ATP, and PFF). The comparison reveals that utilising TTDR may reduce the number of data sets that are sent by a maximum of 62%, while using ATP can reduce the number of data sets by a maximum of 83%. While in PFF, the proportion of data sets that are transferred is always one hundred percent. When compared to ATP and PFF, the data shown in Figure 10 demonstrate that TTDR generates better outcomes. In order to accomplish this goal, the number of data sets that are moved on to the subsequent layer during each session was minimised. When directly applying a compression method to IoT sensor nodes, the primary objective is to minimise the amount of energy that is used by the system.

Table 4: Percentage Of Send By Iot Sensor Node

Readings / period	ATP (0.03)	ATP (0.05)	ATP (0.07)	TTDR	PFF
20	83	85	87	60	100
50	85	87	89	55	100
100	89	90	91	50	100

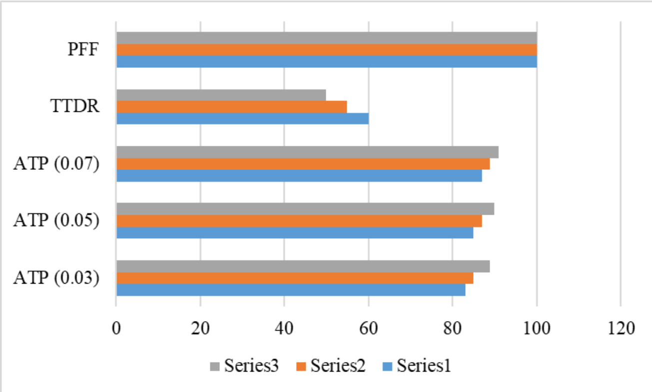

Fig 8: Percentage of send by IOT sensor node

Figure 8 depicts the "First Order Radio Model" developed by Heinzelman. This model is the basis for TTDR's energy consumption analysis. It was decided to go with this model since it was compatible with the topology that was chosen for the study. We are making the assumption that data transmission is what causes the Internet of Things sensor node to use such a large amount of energy. The C_{Ratio} may be determined by applying the following equation to the data.

$$C_{Ratio} = \left(1 - \frac{Compressed\_DS}{Original\_DS}\right) \times 100\%$$

Where the uncompressed DS indicates the number of bits that were present in the original dataset and the original DS indicates the size of the dataset before compression. The compressed DS gives a representation of the number of bits that remain after compression. 64 bits are used by TTDR so that uncompressed data values may be expressed. The quantity of data readings encoded during each period determines the range for the TTDR compression ratio, which may be anywhere from (80.45% to 83.93%). It is imperative that you bear in mind that the first reading will always be the uncompressed version of the material.

6. Result and Discussion

During the process of evaluating the system, we looked at how well it performs consistently and whether or not it gives a level of performance that is adequate. Two separate tests were carried out in order to provide an accurate picture of how well the system worked.

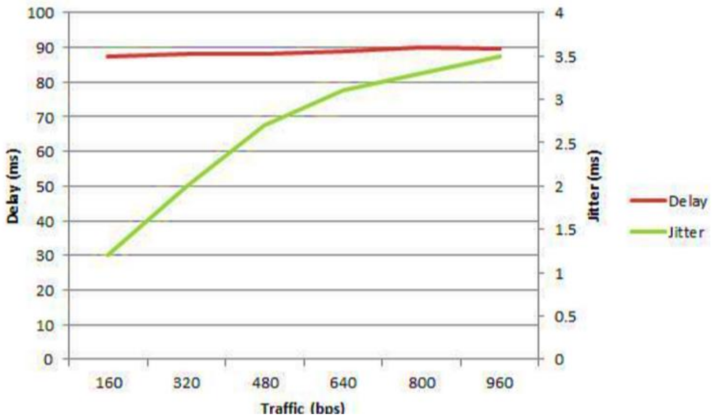

Fig 9: Delay and Jitter Variation with Increasing Traffic

The primary purpose of the first experiment is to act as a recorder for the system's activities. During this experiment, each mote reads its sensory input once per second and creates WSN communication. We are able to measure the delay and monitor how it varies in relation to the amount of traffic. In principle, the routing protocol gives precedence to routing control messages over data packets when determining which to process first. As a consequence of this, this priority may have an effect on the latency experienced across the network. The conclusion that the quantity of motes in the network does not have a significant influence on the average latency can be drawn from the data shown in Figure 9. Jitter, on the other hand, is strongly influenced by the number of motes that are present in the network. Jitter is more sensitive to variations in traffic than delay is, and this is due to the fact that there are moments when the network load is higher than other times. As a direct consequence of this, the latency continues to be the same while the jitter becomes worse.In the second experiment, the amount of time needed to complete the Gateway Packet Transformation operation, which was a contributor to the overall communication delay, was measured and computed. To put it another way, how much more time would the packet translation procedure contribute to the total delay? The results of the research demonstrate the average delay and jitter that were computed for the time frame starting with packet sniffing by the gateway's packet transformation process and finishing with the transformation and sending to the receiver. These results are shown in the phrase "average delay" and "jitter," respectively. This information was sent in around 200 packets, which were later rearranged by the process once they were intercepted[7-9].

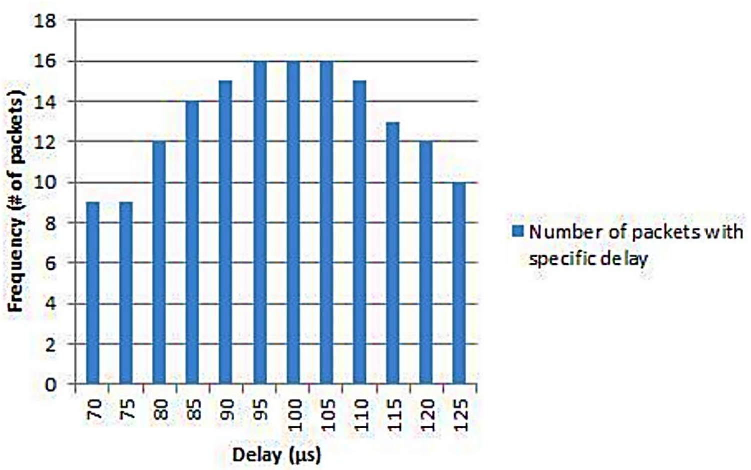

Fig 10: Delay Frequency Histogram

The time it takes for the transition to take place is measured in milliarcseconds. The jitter lasts for around 30 microseconds, although the average delay lasts for 100 microseconds. This has a direct impact on how long it takes for the processing to complete, which might take anything from a few microseconds up to a maximum of 150 microseconds. In addition, the histogram that can be seen in figure 10 illustrates how the distribution of delay frequency shifts based on the load that is being put on the system. From this image, it is possible to draw the conclusion that the delay follows a regular distribution. Additionally, the Gateway Packet Transformation technique[10-12] is shown to not contribute a significant amount to the overall delay as a result of this experiment's findings.

❖ **Obtained Results in the GW Tier**

The GW acts as a link between a typical WSN and a base station or end user that is connected to a local network or the Internet. This connection may be point-to-point or multipoint. In addition to transmitting the aggregate data readings to the primary server, it also broadcasts instructions and updates to the Wireless Sensor Network. It is necessary to initiate communication between the sensor nodes and the GW by delivering compressed data sets from the sensor nodes. Following then, the compression will be removed from the new data sets that are being received. In each and every sensor data collection, we have made the decision to condense the data readings into an 8-value (3-bit) format in accordance with the cardinality (k) of the MDL principle. Our exhaustive simulation experiments, on the other hand, have shown that each reading of uncompressed data takes 64 bits to represent it. It is important that you keep in mind that the hypothesis for each cluster is always presented in its original, unedited form.

Because each data reading is only represented by three bits, we can see that the amount of data that is sent has been significantly reduced. At the second level, the GW is responsible for receiving all of the data sets that have been detected by the Internet of Things sensor nodes that are presently a part of it. As a consequence of this, the energy that is consumed at the GW tier is a combination of the energy that is needed to transfer data to the cloud and the energy that is used to receive data sets. This energy is computed as follows:

$$EN_{GW}(DS, d) = E_{TX}(DS, d) + E_{RX}(DS, d)$$

In this part, the performance and simulation results are evaluated with the use of graphs and discussion pertaining to the CBDR technique that was recommended in section 3. There are two parts to the goal: First, test the performance of the CBDR using real sensory data as well as various other performance indicators. Every sensor node will have access to the intended CBDR, but making use of it will depend on how the dataset from the Intel Berkeley Research Lab is used. This observed meteorological data, which consists of the temperature, humidity, and light, is gathered once every 31 seconds. Throughout the course of the experimental simulations, performance measurements served as the primary method for determining whether or not the CBDR approach was successful. These performance measurements include lifetime, the quantity of data that remains after compression, the fraction of data supplied to the GW, the compression ratio, and the amount of energy that is utilised. The second stage is to evaluate the suggested CBDR in comparison to similar competing methods.

7. Conclusions

The primary objective of this survey article is to evaluate the data reduction strategy for the purpose of conserving energy in wireless sensor networks (WSN). In this survey report, we present an overview of recent improvements that have been made to various reduction initiatives. This study cites a number of articles, many of which contribute to both the practical application of the findings and to further research. Research groups have been concentrating their efforts on developing the most effective data reduction-based strategy for use in WSNs, despite the fact that data reduction methodologies are reliant on the topology of the WSN, the needs of the application, and the environment of the network. As part of our research, we have devised a two-pronged strategy to reduce the amount of data generated by Internet of Things sensors. The sensor nodes and the gateway tier, which are the two layers of the network architecture in which they are located, are where our proposed method operates. At the sensor node layer, we developed a data compression strategy that was both straightforward to build and well-suited for the constrained sensor nodes used in the Internet of Things. Delta encoding, followed by RLE encoding, is used in the approaches in order to make advantage of the temporal correlation that is present in the sensor data. When it came to grouping the data sets that were received from sensor nodes, we established a technique to reduce the amount of data sets based on the Minimum Description Length (MDL) idea. This was done at the gateway layer. The second layer implements a basic strategy for reducing the quantity of data sets that are sent. If there are any combinations of incoming data sets that can be streamlined by using the MDL principle, then those pairs will be merged into a single cluster.

References

[1] K. Watfa, O. Mirza, and J. Kawtharani "BARC: A battery aware reliable clustering algorithm for sensor networks", *Journal of Network and Computer Applications*, 32,6, 1183-1193

[2] M. Rajalakshmi M. E., R. Karthika M. E., "Hybrid Data Reduction Scheme for Energy Saving in Wireless Sensor Networks", *International journal of Science*, *Engineering and Technology Research (IJSETR)*, volume 2, issue 4, ISSN: 2278-7798, April 2013.

[3] AnkitTripathi, Sanjeev Gupta, BhartiChourasiya, "Survey on Data Aggregation Techniques for Wireless Sensor Networks", *International Journal of*

Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, July 2014.

[4] BiljanaStojkoska, dimitarSolev, DancoDavcev "Data Prediction in WSN using variable step size LMS Algorithm". SENSORCOMM 2011: *The Fifth International Conference on Sensor Technologies and Applications*.

[5] Heinzelman, W.R., Chandrakasan, A., and Balakrishnan, H., "Energy-Efficient Communication Protocol for Wireless Microsensor Networks". *In proceedings of the 33rd Hawaii international Conference on System Science*. Hawaii, January 2000.

[6] D.Chu, A. Deshpande, J.M.Hellerstein, W.Hong, "Approximate data collection in sensor networks using probabilistic models", in: proc. *22nd International Conference on Data Engineering (ICDE06)*, Atlanta, GA, , p. 48, April 3-8, 2006

[7] Neha Sharma, P. William, Kushagra Kulshreshtha, Gunjan Sharma, Bhadrappa Haralayya, Yogesh Chauhan, Anurag Shrivastava, "Human Resource Management Model with ICT Architecture: Solution of Management & Understanding of Psychology of Human Resources and Corporate Social Responsibility", *JRTDD*, vol. 6, no. 9s(2), pp. 219–230, Aug. 2023.

[8] William, P., Shrivastava, A., Chauhan, P.S., Raja, M., Ojha, S.B., Kumar, K. (2023). Natural Language Processing Implementation for Sentiment Analysis on Tweets. In: Marriwala, N., Tripathi, C., Jain, S., Kumar, D. (eds) Mobile Radio Communications and 5G Networks. Lecture Notes in Networks and Systems, vol 588. Springer, Singapore. https://doi.org/10.1007/978-981-19-7982-8_26

[9] K. Maheswari, P. William, Gunjan Sharma, Firas Tayseer Mohammad Ayasrah, Ahmad Y. A. Bani Ahmad, Gowtham Ramkumar, Anurag Shrivastava, "Enterprise Human Resource Management Model by Artificial Intelligence to Get Befitted in Psychology of Consumers Towards Digital Technology", *JRTDD*, vol. 6, no. 10s(2), pp. 209–220, Sep. 2023.

[10] Kumar, A., More, C., Shinde, N. K., Muralidhar, N. V., Shrivastava, A., Reddy, C. V. K., & William, P. (2023). Distributed Electromagnetic Radiation Based Renewable Energy Assessment Using Novel Ensembling Approach. *Journal of Nano-and Electronic Physics*, *15*(4).

[11] William, P., Shrivastava, A., Shunmuga Karpagam, N., Mohanaprakash, T.A., Tongkachok, K., Kumar, K. (2023). Crime Analysis Using Computer Vision Approach with Machine Learning. In: Marriwala, N., Tripathi, C., Jain, S., Kumar, D. (eds) Mobile Radio Communications and 5G Networks. Lecture Notes in Networks and Systems, vol 588. Springer, Singapore. https://doi.org/10.1007/978-981-19-7982-8_25

[12] RI, Z. Yun, Z. Xian-ha, and Z. Zibo " A clustering algorithm based on cell combination for wireless sensor networks" *In Second International Workshop on Education Technology and Computer Science*, 2,74-77.

[13] R. Wang, L. Guozhi, and C. Zheng" A clustering algorithm based on virtual area partition for heterogeneous wireless sensor networks", *In International Conference on Mechatronics and Automation*, 372-376

[14] W. Yang, and W. T. Zhu " Voting-on-grid clustering for secures localization in wireless sensor networks", *In proceedings of the IEEE International Conference on communication*, 1-5.

[15] S. Zainalie and M. Yaghmaee "CFL: A clustering algorithm for localization in wireless sensor networks", *In International Symposium on Telecommunications* 435-439.