CHAPTER 1

INTRODUCTION

1.1 Introduction To Wireless Sensor Networks (WSNs)

A Wireless Sensor Network is a wireless network which consists of several independent devices called sensors, distributed over a specified area to monitor the physical or environmental parameters. Over the past few years, Wireless Sensor Networks (WSNs) are increasingly used in several real-world applications mainly because of the transition from wired to wireless technologies. It is a network that is formed by a large number of sensor nodes over an area where each node is equipped with a sensor whose purpose is to detect physical phenomena such as light, heat, pressure, etc. Wireless sensor networks have been abundantly utilized for a variety of applications such as target tracking, environment applications, military surveillance, industrial and agricultural production, medical healthcare applications and intelligent home furnishing. In most of the cases, the sensor nodes form a multi-hop network in which the Base Station (BS) is like a central controlling point for all of the Cluster Heads or Sensor Nodes. Mostly, a sensor node has performance limitations in terms of its computational capability, power supply and its remote location. The major responsibility or function of a sensor node in a WSN is to collect the data which is given as an output by its sensing unit, sample the data observations at different time intervals and then transmit the data to the base station (sink node) or to the Cluster Head.

The major feature of the WSNs which allows their abundant usage and applications in the real world is the dependent on wireless technology as compared to the wired technology. Today's world is becoming more and more mobile every day. Due to such factors, WSNs give way to the easier deployment of sensor nodes over an experimental area and better flexibility of devices in terms of their location of use and the parameters that they are sensing.

1.2 Structure Of Sensor Nodes

A sensor node is simply a micro-electro-mechanical system (MEMS) or device which has the capability of sensing physical or environmental conditions, the capability of data transmission as well as data reception through radio communication and also the capability to perform computations and manipulations on the data. The sensor nodes can be dropped over an experimental area to perform a variety of operations as described in a section below.

A sensor node is a combined package of hardware and software typically consisting of several parts including:

- A Radio Transmitter.
- A Radio Receiver.
- A Sensing Unit which senses the conditions and parameters of the area.
- A microcontroller.
- A Power Source for energy requirements, usually a non-rechargeable battery.

The cost of the sensor nodes also varies and mainly depends on factors like the computation power, the amount of power backup it has, energy efficiency and also on the radio communication range of the sensor. These factors govern the lifetime of the sensor nodes, and subsequently govern the lifetime of the WSN also.

1.3 Communication In WSNs

It is the communication in the wireless sensor networks that is the main portion of concern. Communication in a WSN can be single-hop or it can be multi-hop also. In a single-hop communication, each sensor node is at maximum one-hop distance from the point of collection or aggregation of data (the sensor nodes are adjacent to each other or all nodes are neighbours of ClusterHead), whereas in a multi-hop environment, a sensor node can be present at a farther distance from the point of data collection and the data is then transmitted through multiple intermediate sensor nodes. A large number of sensor nodes are deployed randomly throughout the area to be monitored. These dropped sensor nodes then form a network among each other depending upon their radio communication range. Sensor nodes monitor the external parameters like temperature, atmospheric pressure, light, etc. using their Sensing Unit. The collected data is sampled and is then transmitted along to other sensor nodes by hopping (one-hop or multi-hop). This data is then finally received by the user from the Base Station or Sink Node through the internet or satellites. The user has full control over the WSN, can configures and manages the working of Base Station, publishes monitoring objectives that need to be achieved and also controls the collection of monitored data.

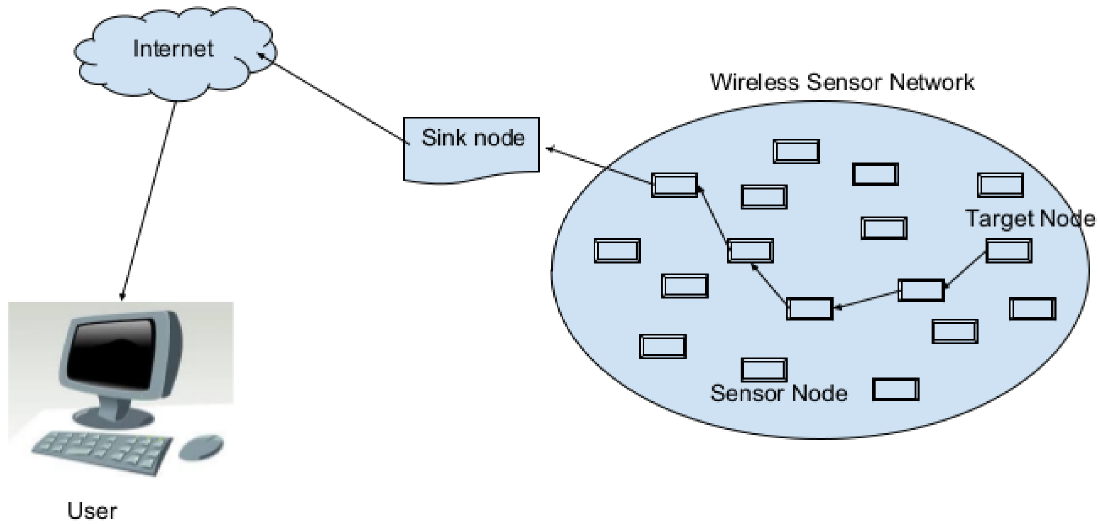


Figure 1.1: Communication In Wireless Sensor Networks

1.4 Issues And Challenges In The Field Of WSNs

1.4.1 Energy Efficiency

Energy efficiency is one of the major challenges for WSNs designs. Due to the fact that sensor nodes are deployed in remote locations and use a battery for their energy resource and since these batteries can't be changed overtime because they are basically made for one time use, so it concludes that they have a limited lifetime. As a result, most of the research is taking place to develop energy efficient algorithms and to design sensor hardware that makes wise use of the power supply.

1.4.2 Security

Data Security is another major challenge in WSNs. As the sensor nodes are dropped usually in a remote area and since they are wireless and have ad-hoc arrangement so ensuring security becomes even more challenging. Many a times, attackers attack the WSNs by taking control over a sensor node and then inserting fraudulent information within the networks. Development of efficient security mechanisms for WSNs is a hard task because of the remoteness of the sensor nodes due to which proper security control cannot be exercised over them.

1.4.3 Cost Of The Hardware

One of the main challenges in the design of WSNs is to produce hardware that is relatively cheaper and can be deployed easily. Currently, the sensor nodes are a bit more expensive in the market. There is also a trade-off between the cost and the expanded utility of the sensor nodes. The more is the hardware cost of the sensor nodes, the lesser will be their utility in day-to-day general applications.

1.4.4 Real World Protocols

Protocols developed for WSNs should be complete in all the respects so that they can be employed for all the real world problems faced by the WSNs. These real world protocols give way to the increased usage of WSNs for real-world applications .

1.4.5 Analytical And Practical Results

Very few analytical and experimental results exist for WSNs till date. All new protocols and applications have to be first tested & analysed practically, the results then have to be compared with the existing protocols. Only after this comparison, a confidence is gained over those new applications.

1.5 Applications Of WSNs

1.5.1 Medical Applications

In advanced hospitals and medical areas, the integrated monitoring of a patient can be done by using WSNs. Diagnosis of diseases and internal body problems can be detected efficiently and accurately. They also aid the doctors and surgeons in operating the patient's body. They also help in keeping a check on drug level present in a patient's body. An example of this is 'Artificial Retina' which helps the visually impaired patients in detecting the presence of light and the objects. They can also locate and count.

1.5.2 Area Monitoring

Area monitoring is the one of the most common and widespread applications of WSNs. In area monitoring, the sensor nodes are dropped over an experimental region where some condition or parameter is to be monitored. An example is to monitor the temperature in a particular area.

1.5.3 Military Applications

WSNs have an advantage that they can be deployed rapidly and easily. And the fact that they organize themselves on their own makes them very useful in military operations for sensing and monitoring friendly or hostile moves. WSNs can be used for Battlefield Surveillance to keep a check on all the aspects in order to supply the military with more equipment, forces or ammunitions, if needed. All types of attacks like nuclear or biological, chemical, can be detected efficiently through the use of different types of sensor nodes. e.g. 'Sniper Detection System' uses acoustic sensors to detect the incoming fire and the location of the opponent by processing the detected audio from the microphone.

1.5.4 Disaster Relief Operations

Wireless sensor networks can be effectively used to prevent or minimize the after effects of natural disasters like floods, earthquakes, forest fires, tsunamis, cyclones, etc. The sensor nodes can be dropped over an area that seems to be vulnerable to a particular calamity in order to monitor the physical parameters related to that calamity. The aggregated data can then be used to study or detect the outbreak of the calamity which will prove helpful to reduce the damage caused by the calamity. Wireless nodes can also be used to monitor the changes in water levels in real time by deploying them in rivers.

1.5.5 Environmental Applications

WSNs take advantage of the ad-hoc wireless links that they use compared to the wired installations, which makes them more flexible for collecting and aggregating readings in different areas of a particular city. They can be used to track the motion and lifestyle of animals, birds and record them which makes them useful in studying the flora and fauna of a particular area. They can be used even for monitoring the earth and soil properties in an area which can help the farmers have better productivity.

They can also be used for the detection of forest fires, flood, earthquakes, volcanos and chemical/biological outbreak etc. e.g. several cities like Stockholm, London and Brisbane make use of WSNs to monitor the concentration of poisonous gases to keep a check on the air pollution.

1.6 Clustering In WSNs

Clustering is a reliable and efficient mechanism in large multi-hop wireless sensor networks for achieving scalability (so that more number of sensor nodes can be added to the WSN without affecting the performance), improved energy efficiency (to increase the network lifetime) and achieving better and increased network performance. Clustering improves the data collection mechanism in WSNs. Clustering is an important task for performing the analysis and exploration of data efficiently. It finds its place in a wide range of applications from computational biology to computer vision to information retrieval. There are many variants of Clustering and it is employed in different fields of science.

1.6.1 Basic Mechanism Of Clustering

The basic mechanism of clustering is that the sensor nodes are grouped into disjoint sets or categories, where each set is supervised by a designated sensor node called the Cluster-Head (CH), selected from among the sensor nodes. The sensor nodes belonging to a cluster transmit their collected observations (which are likely to be highly correlated) to their CH rather than sending to the sink node. The CH removes the redundancies present in the aggregated data and transmits the compressed data to the base station.

The formation of clusters in a WSN ultimately leads to a two-level hierarchy consisting of the CH nodes at the higher level of the hierarchy and the cluster-member nodes at the lower level of the hierarchy. The sensor nodes keep on sensing the parameters and send the sampled data periodically the corresponding CH nodes. The CH nodes receive the sampled data from the cluster member nodes and then aggregate the data. The aggregated data is then compressed by supressing local redundancies and is then transmitted to the Base Station (BS) either directly or through the intermediate communication with other CH nodes.

However, because the CH nodes send their data to higher distances than the common nodes, they naturally spend energy at higher rates. Periodical re-clustering and re-election of CH balances the energy consumption among the sensor nodes, thus rotating the role of CH among all the nodes in a cluster.

The BS is the node at which all the aggregated and compressed data is received from the Cluster Heads and is processed so that the user can access this data through internet or satellite communications. In most of the WSNs, the location of the BS is assumed to be fixed and at a far distance from the sensor nodes. The CH nodes actually perform the role of gateways between the sensor nodes (where data is created) and the BS (where data is stored and processed for access). In some way, the CH act as sink nodes for the cluster nodes, and the BS acts as the sink for the CHs. The structure or hierarchy formed between the sensor nodes, the sink (CH), and the BS can be replicated as many times as it is needed, creating multiple layers of the hierarchical WSN giving way to the multi-level cluster hierarchy.

1.6.2 Objectives Of Clustering

The main objectives for which Clustering is performed in WSNs is to achieve overall system scalability, lifetime, and energy efficiency. The hierarchical structure formed by Clustering helps the sensor nodes to lower their energy consumption. Clustering allows the CHs to perform data aggregation and compression in order to decrease the total amount of data transmitted to the BS. This helps saving the network bandwidth and allows the addition of more sensor nodes into the WSN without actually affecting or degrading the existing performance of the system.

On the other hand, in a single-tier network which does not employs clustering, the base station gets flooded and overloaded without a huge bulk of data from all the sensor nodes present in the network. Such overload can become the cause of latency in communication and can also result to congestion of data packets. Usually, the sensor nodes have a limited communication range and are not capable of transmitting data to far away locations. This leads to the loss of scalability of WSNs and inhibits their application for monitoring a large area which involves huge number of sensor nodes in the network. Hierarchical structure of WSNs that use clustering proves to be useful for the applications that require high scalability to hundreds or thousands of sensor nodes.

In addition to improving network scalability and decreasing the energy consumption through data aggregation, clustering has numerous other advantages and corresponding objectives :

- It leads to localization of the route setup within a cluster and thus reduces the number of entries in the routing table stored at each individual sensor node.
- It can also leads to less bandwidth utilisation because it limits the scope of intercluster communications to CHs and avoids the redundant exchange of messages among sensor nodes in a cluster.
- There is a little responsibility on sensor nodes. They have to focus only on their connection with the CH. They are now independent of the changes taking place at inter-cluster level of hierarchy.

- A CH has full control over the sensor nodes in its cluster and can schedule activities in the cluster.

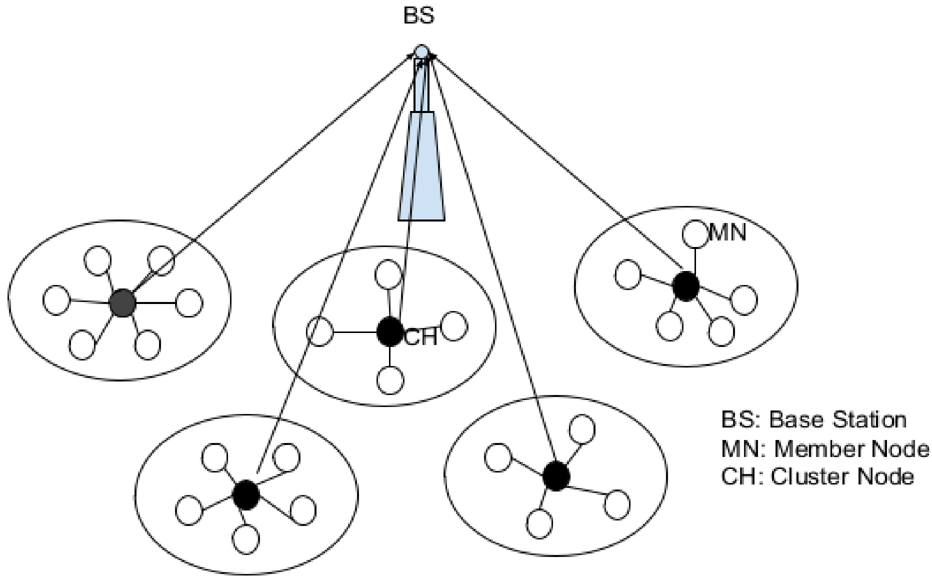


Figure 1.2 : Clustering In Wireless Sensor Networks

1.7 Motivation

Wireless Sensor Networks are usually small in size and operate on low power. Energy resources are often limited for wireless sensor networks. There are various clustering algorithms in existence that have been proposed to categorize sensor nodes in a wireless sensor network into clusters. Most of these protocols or clustering algorithms mainly focus on the energy consumption of the sensor nodes. Due to this constrained and strict focus, these algorithms apply methods to elect Cluster Heads only on the basis of maximum residual energy considerations and the number of neighbours in the related set which is called connection degree (D).

However, one of the most important properties that the sensor nodes present in one cluster should have is that their observed data should be highly correlated with each other. If a clustering algorithm does not take into consideration the existence of data correlation among the sensor nodes, then the clustered structure that is generated by the clustering algorithm does not exploit or take advantage of these correlations and leads to poorly efficient aggregation and compression by the CH nodes.

Moreover, when the initial cluster formation takes place then the calculation of average entropy and the connection degree calculation takes place which results in flooding of messages over the network. As a result of messages overhead, the energy consumption of the sensor nodes increases as a result the network lifetime decreases because there is no external source of energy is available in case of wireless sensor networks. Also the nodes present at the boundary of the clusters should make a decision to go on with which cluster and transfer its data to which Cluster Head.

Motivated by these two factors, the proposed algorithm tries to integrate data correlation and the modifications in initial cluster formation so as to reduce messages overhead which ultimately results in high energy efficiency and increased scalability.

1.8 Problem Statement

The main problem in the wireless sensor networks is to design and implement an efficient data trasmission for Wireless Sensor Networks (WSNs) which incorporates or exploits the presence of data correlation and the modifications in the initial cluster formation so as to reduce the total energy consumption of the network as well as increases the lifetime and scalability of the WSN. The proposed algorithm uses average entropy to measure the data correlation among sensor nodes. Apart from average entropy a special type of clustering scheme called static clustering (grid based) is also used at the beginning of the cluster formation to conserve the extra energy used initially for cluster formation. Apart for average entropy, which is used for correlation calculation, we uses the connection degree and the distance measures also.

So the main aim of the project is to categorise the sensor nodes into clusters such that the messages overhead and the complexity of the information exchange are reduced and the total entropy of the system is minimized and the sensor nodes present in the cluster are highly correlated.

CHAPTER 2

LITERATURE REVIEW

2.1 Clustering

Clustering has tested to be an economical methodology that increases the network life time by dropping the energy utilization and provides the required quantifiability. To achieve high quantifiability and magnified energy potency and to reinforce the network life time the researchers have extremely adopted the scheme of forming clusters i.e. grouping the sensor nodes in large scale wireless sensor network environments.

In this scheme, the member nodes periodically transmit there information to the heads of the clusters they belong and it becomes the responsibility of the cluster head to combine this information and transmit it to the base station. This transmission can either be direct or via alternative cluster heads. This scheme eventually creates two level structures wherever higher level constitutes of the cluster head nodes and therefore the member nodes become a vicinity of lower level hierarchy thereby decreasing the number of relayed packets. A cluster head node has an additional load because it should settle for messages from its cluster members, combine them, broadcast the collective message to the next hop towards the sink and relay the collective messages originated by alternative cluster head nodes. Re-clustering the network is usually necessary so as to attain the load balancing.

Ideal implementation of this is usually energy economical if the cluster heads are befittingly positioned so the position of cluster head becomes a key criteria in scheme for achieving energy potency. In clusering, the cluster head nodes are elected among the deployed sensors within the network wherever this network is undiversified in nature. Communication locality and distance from base station are major considerations that require to think about whereas implementing clustering in wireless sensing network. Another key side of clustering is that the communication between the cluster head and the base station, if this is often not direct than multihop routing is required that generates the importance of inter-cluster head connectivity. And conjointly the cluster head shouldn't be exhausted unnecessarily which can otherwise result in loss of energy of cluster head nodes.

2.2 Review Of Various Clustering Algorithms

2.2.1 Tree Based Clustering Algorithm Using PCA

This algorithm proposes a collaborative data reduction method to remove the redundancy existing in the data coming from multiple sensors as well as from a single sensor. The algorithm uses a tree-based data propagation model to characterize the collaboration structure among multiple sensors. This algorithm uses Principal Component Analysis (PCA) and partial correlation so that it can detect only linear correlation among the sensor nodes and can remove only linear data redundancy.

2.2.2 Correlation And Random Update Based On Data Change Rate For Wireless Sensor Networks

This algorithm (EECRU) proposes the energy efficient clustering way in which the whole procedure is completed in two stages :

1. Initial Cluster Construction
2. The update of clusters.

In the initial cluster construction, a distributed data correlation based clustering algorithm is used to construct the clusters. The average entropy is computed to calculate the correlation. It uses the fact that nodes in the same clusters are highly correlated to each other so the data redundancy can be diminished or removed. The average entropy describes the average data rate from every node to their centre node. It measures the capability of becoming a CH[1].

In the update of cluster stage, the data change rate and the rotation of cluster-head takes place. But in this method the messages overhead and the complexity of the information exchange during the cluster construction and the update process are the major drawbacks.

ICC procedure consists of 3 phases:

1. Initialization phase
2. Cluster Head Selection phase (CHS)
3. Cluster Formation phase (CF).

1. Initialization Phase: Each node broadcasts its sensor data so that other nodes within the radio range are able to calculate their own average entropy. The related set and unrelated set are initialized.

2. Cluster Head Selection Phase (CHS): In the CHS phase, each node exchanges initialization message with all its neighbors. The local optimal nodes are selected as clusterheads and are added to a global cluster head set C.

3. Cluster Formation phase(CF): CH nodes broadcast notifications to their neighbors. Other nodes must choose a cluster to join according to certain rules.

The methods of measures used in this algorithm are:

- **Average Entropy (Ah_i):** The average entropy descibes average data rate from every node in the neighbour set to their center node.

  For any node n_i belonging to any given network P = {n_i|i = 1,2,3.....,N}, n_j belonging to it's neighbour set, the average entropy Ah_i can be calculated as:

  $$Ah_i = \sum_j \left(\frac{Y_j / X_i}{|NS_i|} \right)$$

  The correlation between the node and its neighbour decreases with the increase in the average entropy. Hence the average entropy should be small, as it implies small data volume and hence energy consumption is minimized.

- **Connection Degree (D_i):** Connection degree gives the measure of the relative location of the neighbouring or correlated nodes. More is the value of connection degree more likely the node is in centre of some correlated node.

- **Data Change Rate:** The data rate can be defined as the rate of change in the values of data i.e. variation in the data. The sensor nodes in inactive gives the stable data indicating that there is less uncertainty of the events of the sensor nodes whereas the data of the active nodes are very unstable and fluctuates more.

  The data change rate is defined as:

  $$K = \frac{1}{\left(1 + e^{-(a \Delta H + b)}\right)}$$

- **Random Update:** This algorithm uses the random update method to reduce the extra energy consumption. Here the update frequency of the clusters are adjusted depending upon the data change rate of the sensor nodes.

  If data change rate is low then the update frequency is low, more the data change rate more is the update frequency.

As described above major drawback is due to the message overhead and complexity of the information exchange. Our proposed algorithm works on the reduction of this message overhead which be further discussed in detail in Chapter 3.

2.2.3 Entropy-Based Correlation Clustering For Wireless Sensor Networks In Multi-Correlated Regional Environments

This research proposes a new simple method of clustering sensor nodes into correlation groups in multiple-correlation areas. There are mainly 3 steps involved:

1. Evaluating Joint Entropy for the sensed data.
2. Defining the correlation region based on entropy theory.
3. Proposing a correlation clustering scheme which has less computation to be done.

The Correlation Coefficient describes the correlation level of a pair of data, independent of the independent entropy values, and thus, can be used to compare the correlation level of given two pairs of data. The correlation coefficient is given by[3]:

$$\rho(X,Y)=2\frac{I(X,Y)}{H(X)+H(Y)}=2-2\frac{H(X,Y)}{H(X)+H(Y)}$$

Where 'I' represents Mutual Information which can be calculated as[3]:

$$I(X,Y)=-\sum_{x\in X}\sum_{y\in Y}P(x,y)\log_2\frac{P(x,y)}{P(x)P(y)}$$

DEFINING THE CORRELATION REGION BASED ON ENTROPY THEORY

The correlated nodes can be collected into a correlated group. Also because of the nodes may share information with each other, their joint entropy evaluates to a value less than the total of all nodes' entropy. This step comprises of estimating the joint entropy of a group of nodes from entropies of a single nodes and correlation coefficients of all pairs in the group. Firstly, we need to determine the upper and lower bounds of the joint entropy in order to find the correlation region.

PROPOSING A CORRELATION CLUSTERING SCHEME WHICH HAS LESS COMPUTATION TO BE DONE

The sensor field can be divided into correlation regions with a specified base entropy and correlation level using the definition of correlation region. Based on this, the required algorithm can be proposed and the results can be evaluated accordingly.

In this research paper, a correlation clustering scheme is proposed which has less computations and is based on the entropy concept. The correlation region is defined according to the entropy and entropy correlation coefficient, which makes the correlation clustering algorithm simple with less computation as compared to the algorithms proposed prior to this approach.

2.2.4 E-BACH: Entropy-Based Clustering Hierarchy For Wireless Sensor Networks

According to this method, network optimization is done using the data quality. The concept of the entropy based on the data is used to develop the hierarchical heterogeneous network. Potential information gain is used to sample the data of each individual node. The proposed approach here, uses the entropy-baesd clustering and this approach is data-centric[2]. The properties of the data that is sensed by the nodes is used as base for the wireless sensor network partitioning and also for forming the cluser head and deciding the schedule for the nodes. According to some entropy measures hierarchy of the clusters is developed based on their content and the nodes providing the similar data are grouped. Basically this approach develop the hierarchy of the nodes based on data entropy, which lowers the energy consumption in transmitting the data as the nodes are in high correlation with other nodes.

2.2.5 LEACH (Low Energy Adaptive Clustering Hierarchy)

LEACH is protocol for WSNs for cluster-based architecture, this is typically a renowned algorithm, which works by choosing the CHs in rounds[9]. LEACH is a mainstream energy economical adaptative formula that structures nodes groups, supported the signal quality and utilizes these native cluster heads as routers to the sink. Since data exchange to the bottom station devours a lot of energy, all the member nodes within a cluster act with the transmission by rotating the cluster heads. This prompts balanced energy utilization of all nodes and hence forward a lot of extended life of the system. A predefined value, P (the desired share of cluster heads within the network), is set before starting this algorithm. LEACH works in many rounds where every round has two stages, the setup stage and the steady stage. In the setup stage, every node chooses whether or not to finish up a cluster head or not. Every node picks a random number p between zero and one, which is it's likelihood to elect itself as a cluster head. If the probability p is a smaller amount than a threshold T (n) for node n, node n will become a cluster head for the present round r. This T (n) is calculated by victimization the Equation as follows:

$$T(n)=\begin{cases} \dfrac{p}{1-p*(r \bmod \dfrac{1}{p})} & if \in G \\ 0 & otherwise \end{cases}$$

During the steady phase, the sensor nodes will begin sensing and transmitting the information to the cluster heads. The cluster heads combine the information from the sensor nodes in their cluster and sends information to the base station. When specific time is spent on the steady phase, the network goes into another round of choosing the cluster heads. The length of the steady phase is longer than the span of the setup part with a particular finish goal to minimize the overhead. LEACH provides optimized behavior for communication in WSNs taking under consideration self-organization techniques. Quality is additionally supported by LEACH, though new nodes should be synchronous to current round. Node failures might prompt less cluster heads to be chosen than sought in light-weight of the very fact that the predefined P could be a proportion of the mixture variety of sensor nodes.

2.2.6 PEGASIS (Power-Efficient Gathering In Sensor Information Systems)

PEGASIS that is associate improvement over the LEACH. It is chain primarily based protocol, in which nodes got to speak with their nearest neighbors and alternate in speaking with BS[11]. Each node within the system utilizes signal quality to search out the closest neighbor. The chain in PEGASIS contains of nodes nearest to every different that form a way to the BS. The accumulated kind of the information are sent to the BS by any node within the chain and therefore the nodes within the chain can alternate sending to the BS. This diminishes the ability needed to transmit information per round on the grounds that the ability depleting is unfold systematically over all nodes.Anyhow, the assumptions in PEGASIS might not usually be practical.

1. PEGASIS expect that each sensor node will speak with the BS directly. In practiical cases, sensor nodes use multi-hop routing to send data to the BS.
2. It considers that each node keep an entire information regarding the location of each different node within the system ; but the strategy by that the node location are find is not precise or accurate.

3. It considers that all sensor nodes have a similar level of energy and all nodes seemingly to die at a similar time. In spite of the actual fact that in several situations sensors are fastened or immobile as assumed in PEGASIS, a couple of sensors could be permitted to move and in this way influence the protocol functions.

2.2.7 TEEN

A hierarchical clustering based protocol produced for responsive systems within which nodes respond instantly to sudden and extreme changes in surroundings famous as TEEN. Cluster formation and information transfer are done as in the LEACH protocol. Threshold values aboard completely different traits - Hard Threshold (HT) and Soft Threshold (ST). These values as well the environmental measures are perceived by the nodes endlessly. At the point once the node finds that the detected attribute has achieved HT, the node switches on its transmitter and sends the perceived data[13].

The perceived value is put away in associate inner variable SV within the node. In the present cluster period, the node will next transmit data simply once the current estimation of the perceived attribute is higher than HT and therefore the present estimation of the perceived attribute varies from SV by a sum equivalent to or higher than the ST. The utilization of HT and ST will decrease the amount of transmissions within the network and consequently it diminishes the general energy dissipation within the network. This approach is not suited for time crucial information sensing applications.

2.2.7 EEHC

EEHC is the randomized algorithm for WSNs. This methodology is divided into two stages in particular single-level clustering and multi-level clustering. In the single-level clustering, each device node declares itself as a CH with likelihood p to the neighboring node within its communication range. These CHs are named as the volunteer CHs. All nodes that are within k hops range of a CH get this announcement either by direct correspondence or by forwarding. Forced CHs are nodes that are neither CH nor belong to a cluster. On the off likelihood that the announcement does not accomplish a node within a pre-set interval t that is computed in lightweight of the length for a packet to succeed in a node that is k hops away, the node can be converted into a forced CH expecting that it's not within k hops of all volunteer CHs. The second section, known as multi-level clustering builds h levels of cluster hierarchy. The algorithmic rule guarantees h-jump network amongst CHs and therefore the base station. The CHs nearest to the bottom station have disadvantage since they are going to concern as relays for different Chs.

2.2.9 HEED(Hybrid, Energy Efficient And Distributed)

Another energy-proficient node cluster algorithmic program is the Hybrid, Energy Efficient and Distributed (HEED) clustering approach for ad hoc sensor networks. HEED is distributed cluster protocol that was proposed with four essential objectives as follows:

- Increasing network lifespan by distributing energy utilization.
- Ending the clustering procedure within a steady number of iterations.
- Reducing control overhead.
- Generating well - distributed cluster heads and compact clusters.

HEED periodically chooses cluster heads based on a hybrid of two cluster parameters: The essential parameter is the leftover energy of each detector node and therefore the secondary parameter is intra-cluster correspondence value as a component of neighbor vicinity or cluster density. The first parameter is used to probabilistically select associate initial set of cluster heads whereas the secondary parameter is employed for breaking ties[15].

The grouping procedure at each sensor node needs a couple of rounds. Every round is sufficiently long to urge messages from any neighbor within the cluster range. As in LEACH, an underlying rate of cluster heads within the system, C_{prob} , is predefined. The parameter C_{prob} is simply used to constrain the underlying cluster head announcements and has no direct impact on the ultimate cluster structure. In HEED, every sensor node sets the probabilty CH_{prob} of changing into a cluster head as follows:

$$CH_{prob} = \frac{C_{prob} * E_{residual}}{E_{max}}$$

Where $E_{residual}$ is that the calculable current residual energy during this sensor node and E_{max} is the maximum (corresponding to a fully charged battery), that is usually identical for homogeneous sensor nodes. The CH_{prob} value should be larger than a minimum threshold p min. A cluster head is either a tentative cluster head, if its CH_{prob} < 1, or a final cluster head, if its CH_{prob} has reached 1[16].

Amid every round of HEED, every sensor node that never got notification from a cluster head chooses itself to become a cluster head with probability CH_{prob} . The recently selected cluster heads are added to this arrangement of cluster heads. If a sensor node is chosen to finish up as cluster head, it telecasts announcement message as a tentative cluster head or a final cluster head. A device node being attentive to the cluster head list chooses the cluster head with the foremost negligible expense from this arrangement of cluster heads. Each node then doubles its CH_{prob} and goes to succeeding step. If a node finishes the HEED execution while selecting itself to become a cluster head or joining a cluster, it declares itself as a final cluster-head. A tentative cluster head node will become a regular node at a later iteration if it hears from a lower cost cluster head. Note that a node is chosen as a cluster head back to back clustering intervals if it's higher remaining energy with lower cost. Since a WSN is assumed to be a stationary network, where node do not die suddenly, the neighbor set of every node doesn't change frequently.

Here HEED doesn't got to do neighbor revelation regularly. HEED increases the lifespan of all the nodes within the network, thus supporting steadiness of the neighbor set. Nodes also automatically update their neighbor sets in multi-hop networks by automatically sending and receiving messages. The HEED clustering enhances system lifespan over LEACH clump since LEACH haphazardly chooses cluster heads (and thus cluster sizes), which can create speedier end of a number of nodes. The ultimate cluster heads selected in HEED are very much circulated over the network and also the energy consumption is minimized.

2.2.10 UCS (Unequal Clustering Size)

It is unequal clustering model, called Unequal clustering size (UCS) to regulate energy utilization. The sensor field is separated into two concentric circles known as layers and every layer has some variety of clusters of same size[17]. The size and shapes of the clusters of two layers are distinctive. The protocol assumes that the BS is placed within the center of the system and CHs area are determined "priori" that are situated symmetrically in concentric circles round the BS. To minimize the energy utilization within the cluster, each CH ought to be set at the middle of the cluster.

CHs are deterministically placed within the system. The coverage of the clusters are often shifted by differing the first of the primary layer round the BS, that the variety of nodes in an exceedingly specific cluster are additionally modified. Every CH transmits data to BS by picking the closest CH toward BS.

The UCS has two preferences contrasted with LEACH. The UCS will continue uniform energy utilization among CHs. This can be accomplished by differing the number of nodes in every cluster as for the traditional communication load. Also, protocol makes two layered network model and two-hop inter-cluster communication technique, this outcome in an exceedingly shorter average transmission distance contrasted and LEACH, in this way there is successfullly the whole energy utilization.

2.2.11 PEACH (Power-Efficient And Adjustive Cluster Hierarchy)

PEACH convention is proposed for WSNs to increase system period by lessening the energy utilization. The nodes within the system will perceive the supply and destination of the data packets by overhearing attributes of wireless correspondence. In PEACH, the clusters are framed without extra transmission overhead, for instance, notice, declaration, connection and scheduling messages. PEACH is probabilistic directive algorithm and provides a flexible multi-level cluster[18]. PEACH is extremely practiced and ascendible below totally different circumstances than this cluster protocols. PEACH could be applicable to each aware and unaware WSNs with relation to location. In specific applications, the location data knowledge of the node isn't notable. In such applications, location unaware PEACH convention may be utilized. The situation aware PEACH works once the confinement mechanism, for instance, a GPS-like instrumentality is accessible on sensor nodes[22].

2.2.12 TTCRP (Two Tier Cluster Primarily Based Routing Protocol)

TTCRP arranges the nodes as clusters at two levels. At the first level sensor nodes be a part of pre allotted quality made CHs. These CHs structure the second level of clusters to convey information to the BS. The CHs area unit outfitted with double channels during which distinctive channels area unit used is for correspondence at each levels. The CHs get info from their cluster node at one channel and utilize second channel to send it to the BS through different CHs. The planned set up executes an influence management algorithm to allow the disengaged sensor nodes and additionally cluster heads to more and more amendment their transmission power for interfacing sensor nodes with inaccessible clusters and subsequently provides system strength.

2.2.13 DAIC(Distance Aware Intelligent Clustering)

Distance Aware Intelligent Clustering (DAIC) is a progressive routing convention projected to reduce the energy utilization and expand the system life. The theme isolates the system into two levels: primary and secondary. The CHs of the first level are chosen by considering the gap between the CH and BS. The protocol decides the number of CHs powerfully in view of the number of alive nodes within the system, that keeps away from the determination of superfluously large range of CHs. The non-CH nodes transmit the data to the essential CHs and also the CH nodes at the secondary level transmit the information to the BS. For uniform distribution of energy load, DAIC utilizes rotation of CHs as a vicinity of every spherical of communication and chooses CHs on the premise of residual energy.

CHAPTER 3

PROPOSED APPROACH

3.1 Assumptions

1. All sensor nodes and cluster heads within the network are stationary.
2. The physical location of the nodes are known to themselves and to the base station.
3. The transmission range of all the nodes are known and all nodes have same transmission range.
4. All the sensor nodes and the cluster heads have unique identities.
5. The base station is the centrally controlling authority and a knows the location all the nodes, virtually has no computation and resource contraints.
6. The base station communicates with the cluster head and each cluster head manages all the sensor nodes in it's own group.
7. All the grids are provided with the grid ids and the base station and the nodes are aware of these grid ids.
8. Initially when the nodes are deployed they know which grid they belong.

3.2 Paramters Used

- **Entropy**

  Entropy is a measure of the uncertainty of a random variable. The amount of disorder which is present in a vector of data values is most suitably determined by calculating the value of Entropy of that vector. Let the sampled data in sensor node X and any of its neighbour Y can be denoted as *Xi={x1,x2,x3, ...,xm}* and *Yi={y1,y2,y3,.....,ym}*. The probability mass function is denoted by p(x). Thus, p(x) and p(y) correspond to the probabilities of two different random variables.

  The entropy H(X) of a discrete random variable X is defined by:

  $$H(X) = -\sum_i p(x)_i \log p(x_i)$$

- **Pearson's Correlation Coefficient**

  The Pearson correlation coefficient is used to measure the strength of a linear association between two variables, where the value ρ = 1 means a perfect positive correlation and the value ρ = -1 means a perfect negative correlation. The Pearson’s Correlation Coefficient is calculated by the following formula:

$$p_{ij} = \frac{Cov(X_i, Y_j)}{\sigma_{x_i} \sigma_{y_j}}$$

Pearson’s Correlation Coefficient can categorise the type of correlation by considering as one variable increases, what happens to the other variable:

Positive correlation: The other variable has a tendency to also increase

Negative correlation: The other variable has a tendency to decrease

No correlation: The other variable does not tend to either increase or decrease.

The Pearson’s Correlation Coefficient will always take on a value between 1 and – 1:

If the correlation coefficient is +1, the variables have a perfect positive correlation. This means that if one variable moves a given amount in one direction, the second moves proportionally in the same direction. A positive correlation coefficient less than one indicates a less than perfect positive correlation, with the strength of the correlation growing as the number approaches one.

If the correlation coefficient is zero, no relationship exists between the variables. If one variable moves, no predictions can be made about the movement of the other variable. They are uncorrelated.

If the correlation coefficient is –1, the variables are perfectly negatively correlated (or inversely correlated) and move in opposition to each other. If one variable increases, the other variable decreases proportionally. A negative correlation coefficient greater than –1 indicates a less than perfect negative correlation, with the strength of the correlation growing as the number approaches –1.

3.3 Implementation Details

Our proposed algorithm mainly focus on the reduction of the message overhead. In most of the clustering algorithms there is the message flooding for the cluster formation and cluster head selection like EECRU (Energy-efficient clustering method using random update). In EECRU, firstly the clusters are formed. While forming the clusters each node broadcast it's data to other nodes and calculate the entropy based on the data. On the basis of this average entropy related and unrelated sets are formed. Each related set contains the nodes with lower entropy that is nodes having higher correlation are grouped together. The local optimal nodes are selected as the cluster heads.

But in our algorithm works on reducing these message overhead. Our proposed algorithm can be explained in details in following 3 steps:

- Grid based clustering
- Cluster Head Selection
- Dynamic Clustering

3.3.1 Grid Based Clustering

Among all clustering Grid-based clustering and routing schemes, in which clusters are equally-sized square grids in a two-dimension plane, with less routing management overhead, and all nodes in one grid are equivalent from the routing perspective. With the assistance of GPS or localization techniques, the square grid also provides easier coordination among all sensor nodes in the network. Therefore, it allows for a theoretical analysis while still being useful enough to incorporate all the important elements of a network. Extensive research work has been done in grid-based clustering. In the early work of GAF(Geographic Adaptive Fidelity) , the grid size s is chosen such that any two nodes in horizontally or vertically adjacent grids are within the transmission range, r, of each other, which is referred to as Manhattan walk as shown in figure 3.1.

By investigating the worst-case scenario, the grid size should be $s \leq r/\sqrt{5}$. Recently, the work of also uses this clustering structure. For the one-dimension case , s should be less than r/2. More recent work of and used a smaller grid size, $s' \leq r/\sqrt{8}$, allowing nodes in diagonal grids to be in the same transmission range as well, as shown in figure 3.2. With the same transmission range r, there are fewer grids in figure 3.1, but it may take more hops to reach the sink. Thus the tradeoff between these two gridding approaches is still an open question.The figures 3.3 and 3.4 shows the sensors nodes deployed in the area of interest.

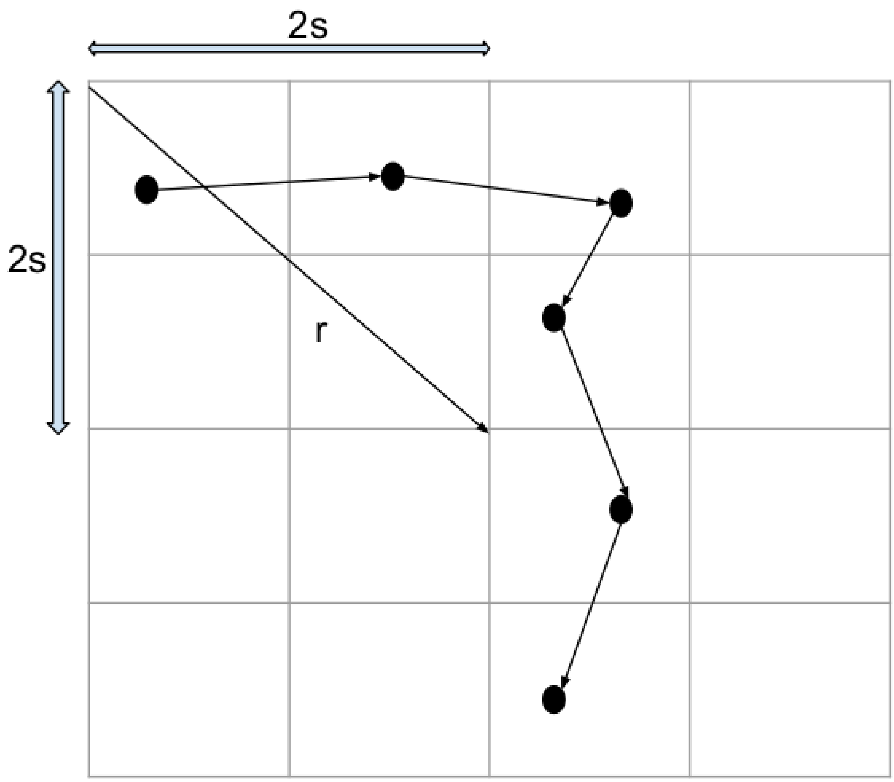


Figure 3.1: Grid Formation Using Manhattan Walk (s ≤r/√5)

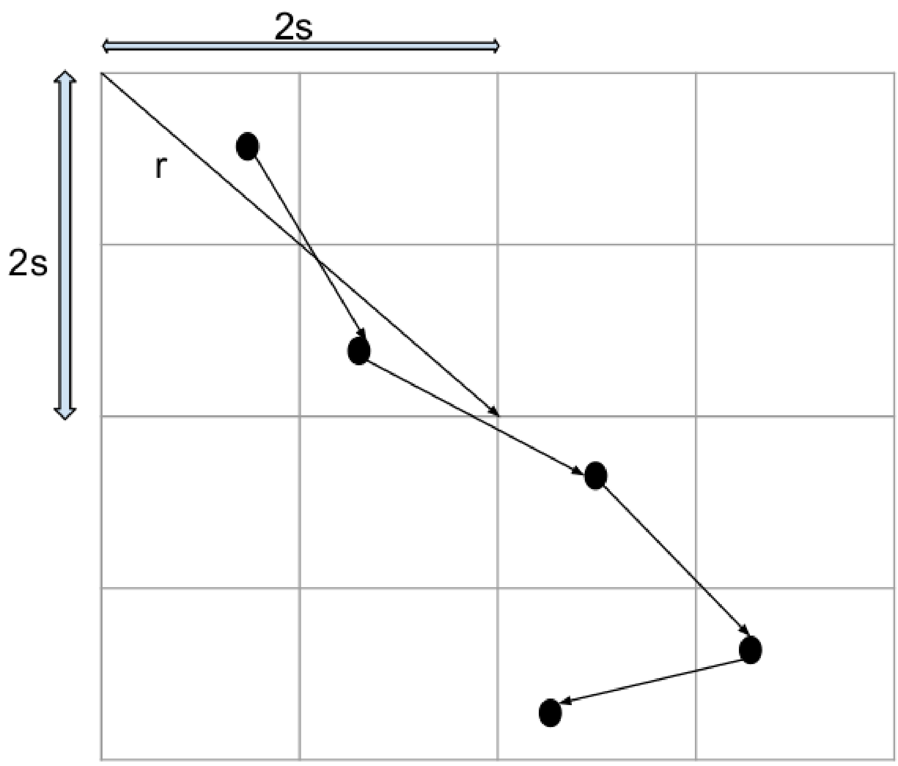


Figure 3.2: Grid Formation Using Diagonal-First (s ≤ r/√8)

Figure 3.3 shows the sensor nodes randomly deployed while figure 3.4 shows how the clusters which are made based on the grid boundary.

We are implementing the grid based clustering algorithm for initial static clustering.

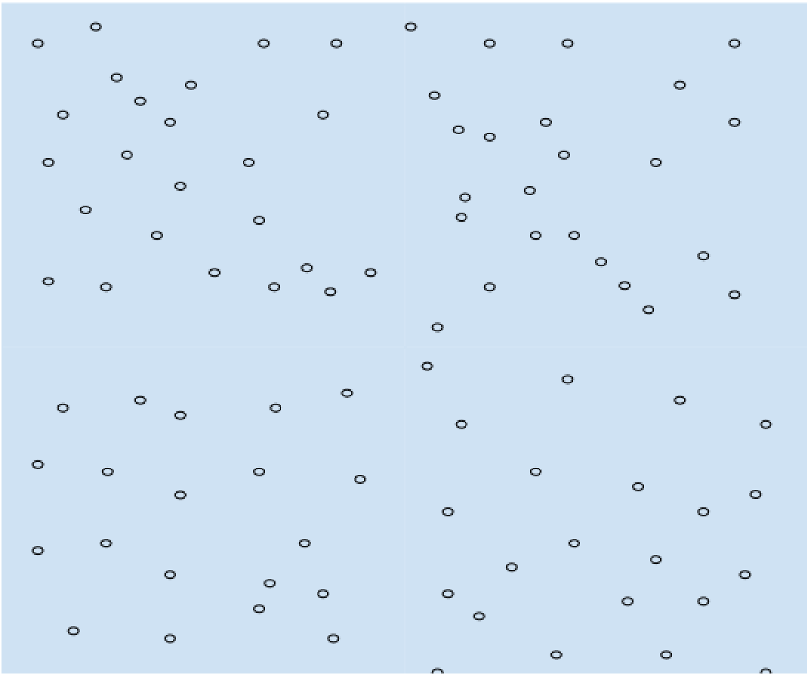


Figure 3.3: Initially Deployed Sensor Nodes In The Area Of Interest

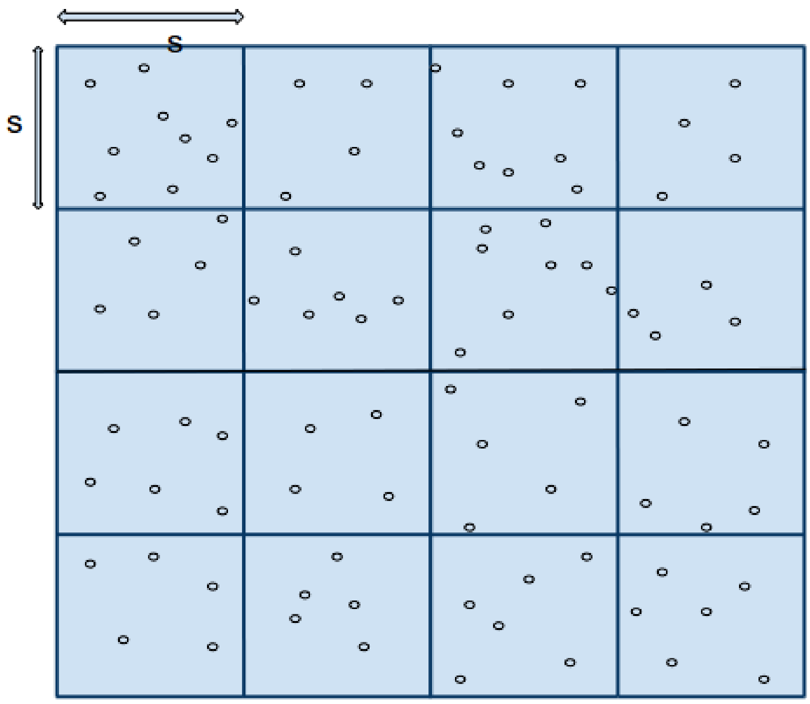


Figure 3.4: Clusters Formation Using Grid Method

3.3.2 Cluster Head Selection

Cluster head is basically that node that is in local optima than all the other nodes in the cluster i.e. the node that lie in close approximation to the center of the cluster. Here the criteria for the cluster head selection is that, the node more closer to the center node tends to become the cluster head. Since the base station aware of each node's location and coordinates select the cluster head for each cluster. The selection of the cluster head for the cluters of figure 3.4 is represented in figure 3.5

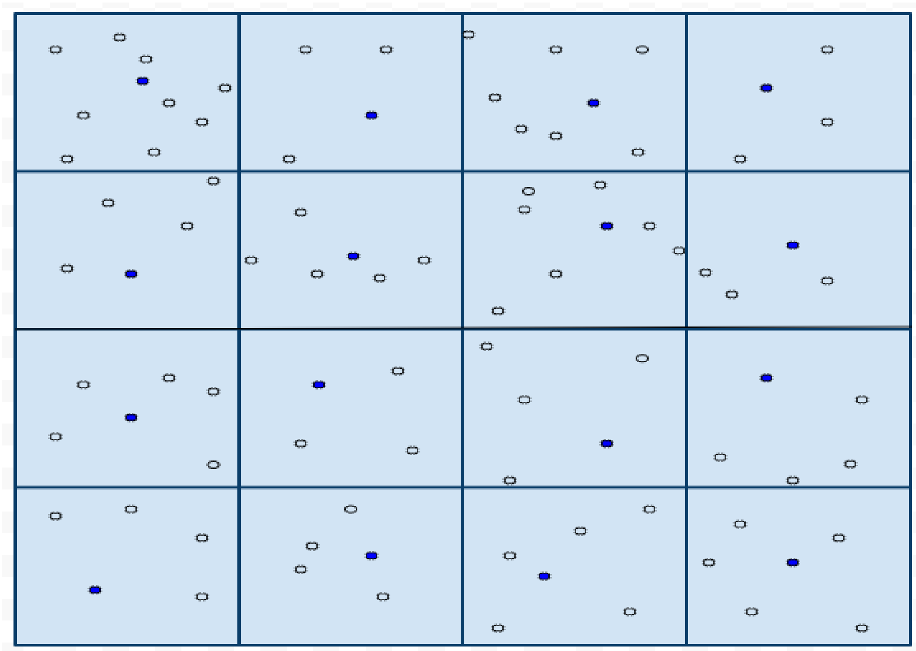


Figure 3.5 Cluster Head Selection

3.3.3 Dynamic Clustering

Initially our approach is to build the static clusters just based on the grids and location, hence it doesnot take into account others factors like entropy, correlation into account. But by the time after some tranmission of data by the member nodes in the cluster ot the cluster head, the cluster head aggregates that data. This aggregated data is then broadcasted to all the nodes in nearby (not necessarily only the cluster members will receive that broadcasted data). The broadcasting range of this signal depends upon one factor h. This h is the predefined or user specified value. The value of the h can vary depending upon the topology, grid size, type of the environment, type of the data sensed. The broadcasted signal strength is adjusted such that all the nodes within the area of (s+ h) *(s+ h) are covered that is all the nodes within this area will receive broadcasted data. All the nodes lying in range (s+h) and (s-h) from the centers will receive more than one broadcasted message, one freom their own cluster head and others from the cluster head of the adjacent clusters. Each of these nodes will calculate the joint entropy based on that data. The minimum value case of the joint entropy is considered, the node will now belong to the particular cluster with respect to which it is having the minimum joint entropy. There can be the case where the value of the entropy will be same, in that case to resolve the conflict we will take the distance of that node into consideration. Cluster head with the minimum distance to that node is choosen as the new cluster head. These nodes with send the message to both old and new clusterhead notifying their cluster change.

3.4 Pseudo-code

1. **Input:** s = r/root(8)
2. **Output:**
3. Divides the Geographical Area 'A' into grids of size s*s.
4. for each G_j = 0 to k //k=no. Of grids
5. for each $N_i \in G_j$
6. $d_i = ((X_i - X_c)^2 + (Y_i - Y_i)^2))^{1/2}$;
7. end;
8. //base station broadcast selection message to each grid
9. select(Nid);
10. for each G_j = 0 to k
11. calculates average entropy for Cluster Head;
12. //Broadcast entropy value
13. entropy(CH_iavg);
14. for N $\in G_j$ h region
15. for each $G_j \in$ adjacent grid of G_j
16. calculates joint entropy;
17. select the least entropy grid;
18. //send Request message to selected grid
19. REQ(G_s);
20. end;

```
                end;
      end;
      for each G_j = 0 to k
          for each received REQ
              CH_j sends reply;
              node included as the member of G_j ;
          end;
    end;
end;
```

CHAPTER 4

EXPERIMENT AND RESULT

4.1 Simulation Tools

The simulation tool used for implementing the project and simulating the proposed algorithm to study the results is Matrix Laboratory (MATLAB). The main reasons for choosing MATLAB were:

- MATLAB allows efficient mathematical modelling.
- MATLAB allows easy graphical representation of output results.
- It is a high-level language that provides support for numerical computation, visualization and application development.
- It provides an interactive environment suitable for the purpose of iterative exploration and problem solving.
- It provides a vast library of mathematical functions which can be used for a variety of mathematical models like linear algebra, statistics, Fourier analysis, etc.
- One of the most important benefits of MATLAB in presenting the output results is that it provides built-in graphics for visualizing data and tools for creating custom plots.
- It provides tools for developing the applications with custom graphical interfaces.

The code base was very computationally intensive and was based on mathematical and probabilistic models. So, a tool which was efficient as well as feature rich was required, thus MATLAB was chosen over others.

4.2 Results

In our proposal main focus is on reduction of number of messages for cluster formation so that clustering becomes more energy efficient.

In case of dynamic clustering as proposed in the research paper Entropy-based Correlation Clustering for Wireless Sensor the number of messages are approximately (n)*(n-1) where ‘n’ is number of nodes deployed in geographical region ‘A’.

In our proposed method these messages reduced to linear number.

Mathematical Calculations :

By Considering the following figure we find out three types of clusters based on no. of nodes present in "h" region.

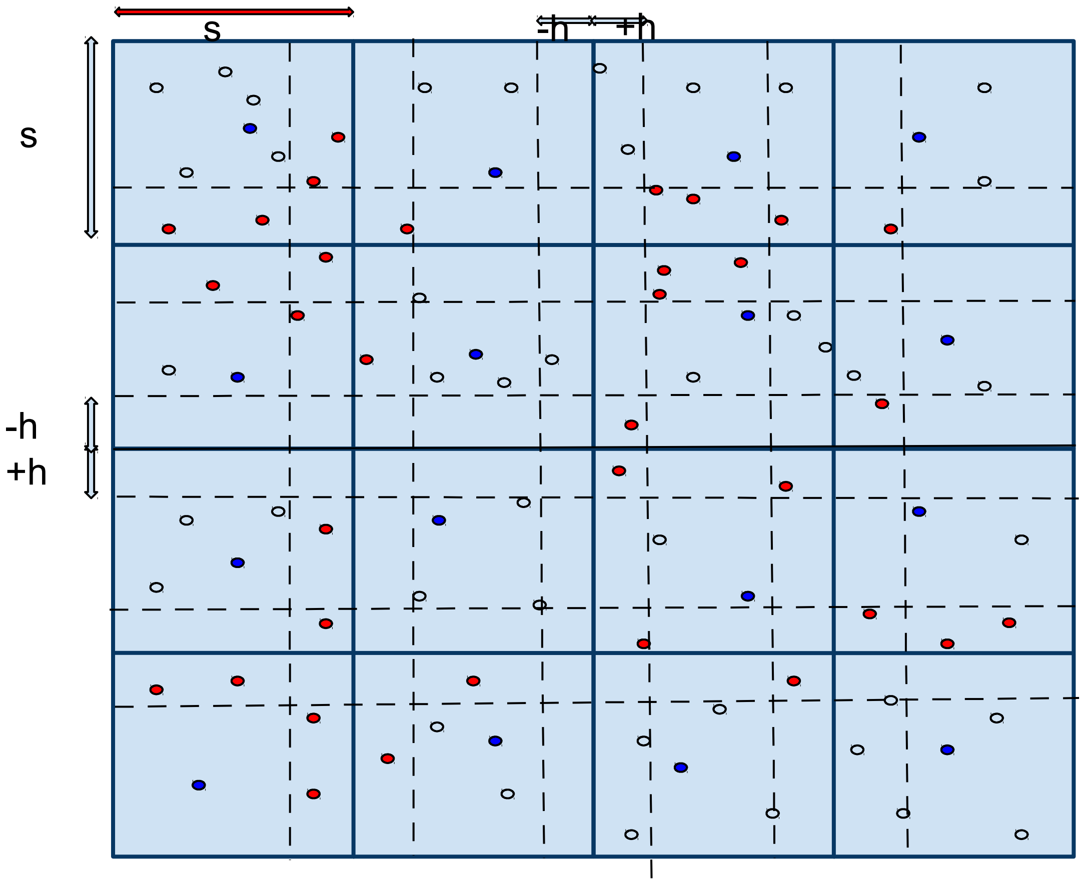
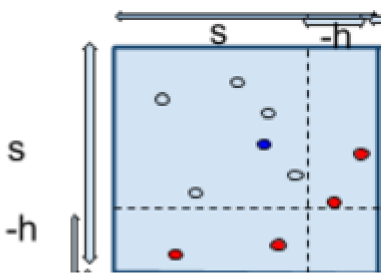


Figure 4.1: Dynamic Clustering Using h Value

These are as follows :

1. No. of nodes for which dynamic clustering will takes place in this grid type are n_1.



Area of cluster :

$$A_g = s \times s$$

Area occupied by 'h' region:

$$A\,n_1 = 2 \times s \times h - h \times h$$

Number of nodes present in a grid:

$$m = \left((n \times A_g) \div (A \times k) \right)$$

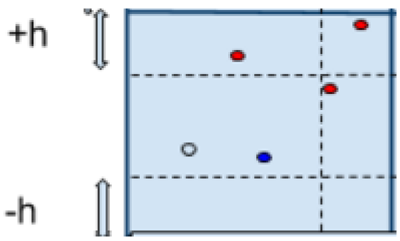
Where

k=number of grids.

$$n_1 = (m \times A\,n_1) \div A_g$$

$$n_1 = (n \times (2 \times s \times h - h \times h) \div A \times k)$$

2. No. of nodes for which dynamic clustering will takes place in this grid type are n_2.



Area of cluster:

$$A_g = s \times s$$

Area occupied by 'h' region:

$$An_2 = 3 \times s \times h - 2 \times h \times h$$

Number of nodes present in a grid:

$$m = \left((n \times A_g) \div (A \times k) \right)$$

Where

k=number of grids.

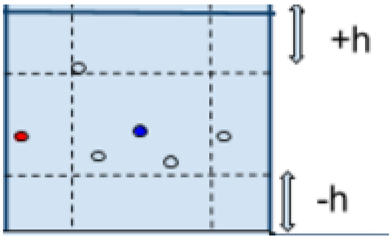
$$n_2 = (m \times An_2) \div A_g$$

$$n_2 = (n \times (3 \times s \times h - 2 \times h \times h) \div A \times k)$$

3. No. of nodes for which dynamic clustering will takes place in this grid type are n_3.

Area of cluster:

$$A_g = s \times s$$



Area occupied by 'h' region:

$$An_3 = 4 \times s \times h - 4 \times h \times h$$

Number of nodes present in a grid :

$$m = ((n \times A_g) \div (A \times k))$$

Where

k=number of grids.

$n_3 = (m \times An_3) \div A_g$

$n_3 = (n \times (4 \times s \times h - 4 \times h \times h) \div A \times k)$

We consider number of grids of type 1 are 'a', number of grids of type 2 are 'b', number of grids of type 3 are 'c'.

Hence total number of nodes present in region 'h' are:

$$n_g = a \times n_1 + b \times n_2 + c \times n_3$$

Calculation of number of messages generated in for cluster formation

Total number of messages generated are:

$$M = m_1 + m_2 + m_3$$

Where

m_1 = number of messages generated during grid formation in ideal case

m_2 = number of messages generated in cluster head selection in ideal case

m_3 = number of messages generated during dynamic entropy based clustering in ideal case

Calculation for m_1 :

$$m_1 = \text{n}$$

Calculation for m_2 :

$$m_2 = \text{n} + k \times m$$

Where

n=number of nodes in region ‘A’
k=number of grids in region ‘A’

Calculation for m_3 :

$$m_3 = (n + (n + n_g) + 2 \times n_g) \times k$$
$$m_3 = (2 \times n + 3 \times n_g) \times k$$

So,

$$M = m_1 + m_2 + m_3$$
$$M = (n) + (n + k \times m) + ((2 \times n + 3 \times n_g) \times k)$$
$$M = 2 \times n + k \times (2 \times n + 3 \times n_g + m)$$

4.3 Comparison

Our results would have to be compared to some other clustering algorithm. So we chose multi-hop scheme to compare our results to. The main focus of our research was to ensure that the clustering algorithm we propose is an energy efficient algorithm. The comparison between the multi-hop scheme and our proposed clustering algorithm is depicted through graphs simulated on MATLAB. The variation in the performance is clearly depicted through these graphs.

Energy consumed in the algorithm is main basis of our comparison. Energy is calculated as a function of total messages passed around in that algorithm. Thus we would be using simulated graphs to highlight how different conditions affect the performance of the algorithm. The conditions that we varied were

- number of nodes in the sensor field
- field size under consideration
- density of deployment

Varying the number of nodes in the sensor field is simulated by variation in the values on the x-axis of the graphs. Field size variation does not visually appear on the graph, but the results are definitely affected by it. Density of deployment is measured in terms of number of nodes in the 100X100 sq unit sensor field. Though we used the deployment density as a criterion, the results show that there is not much difference in the plot of the two algorithms. The default density of nodes is taken as 440 nodes in

100X100 sensor field size; with minimum allowable distance between any 2 random nodes is 5 units. Whenever we vary certain parameter, the other two are assumed to be constant.

4.3.1 Varying The Number Of Nodes

We took three general cases for the variation of number of nodes (n).

4.3.1.1 Field With Less Number Of Nodes (n=18)

When comparing our algorithm with multi hop scheme, we found out that the graph of the energy consumed didn't show a uniform nature. Possibly, the multi-hop scheme performs better than our proposed clustering algorithm, when the number of nodes in the sensor field is less. This is a very justifiable statement, because when the number of nodes is few, then direct-routing or multi-hop scheme is destined to perform better than any clustering algorithm. The reason being that in multi-hop routing, when the number of nodes are less, the total number of messages exchanged is low as compared to the clustering. Any cluster formation and routing requires more number of messages than multi-hop scheme and thus consume more energy. The graph in Figure 4.2 shows that the comparison between the multi-hop scheme and our proposed clustering scheme is not even.

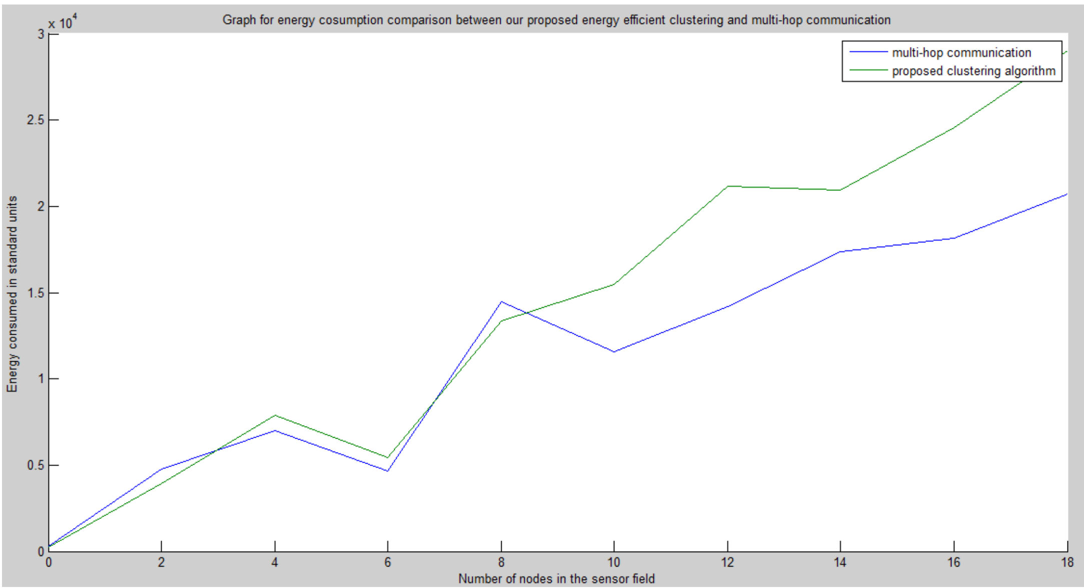


Figure: 4.2: Energy Comparison For Less Number Of nodes

4.3.1.2 Field With Average Number Of Nodes (n=90)

When we considered the average case, we could see that with the increase in the number of nodes, the graph becomes more even. Our proposed algorithm starts performing better when the number of nodes increases and crosses a particular value. This is because the multi-hop scheme leads to a lot of message overhead redundancies which can be avoided in our proposed algorithm.

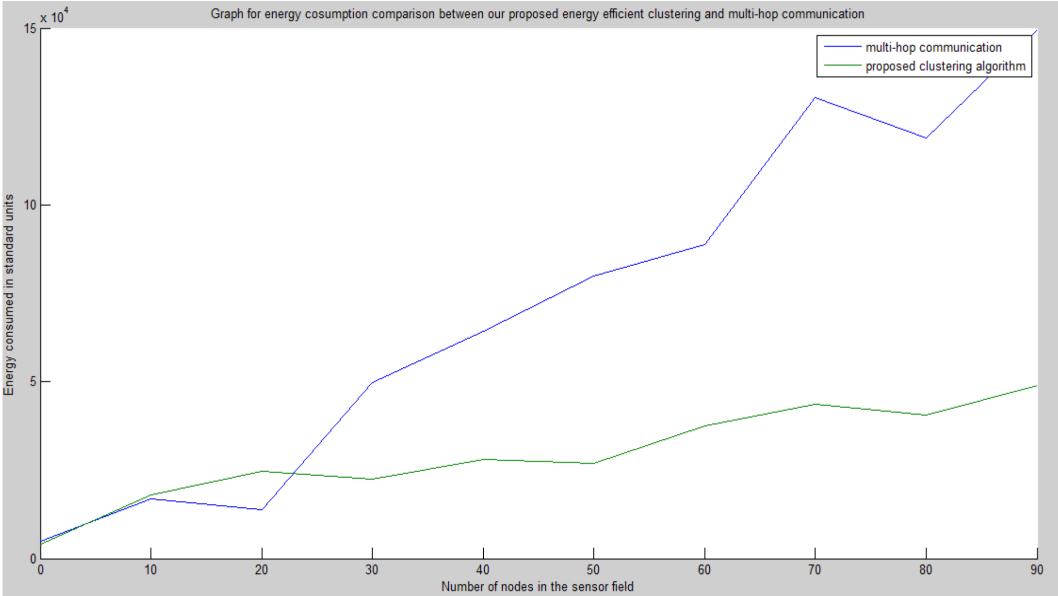


Figure 4.3: Energy Comparison With Average Number Of Nodes

4.3.1.3 Field with high number of nodes (n=900)

As can be seen from the graph shown below, the proposed algorithm consumes far less energy than multi-hop scheme and is thus more energy efficient when there are large number of nodes. As the number of nodes increases, a clear distinction can be seen between the energy consumption in our proposed algorithm and the multi-hoping

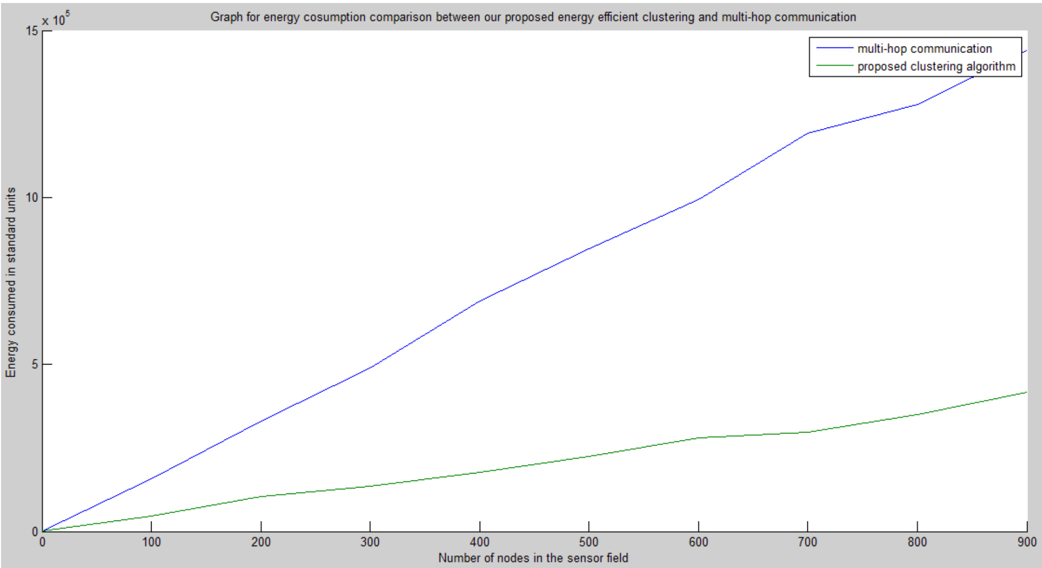


Figure 4.4: Energy Comparison With large Number Of Nodes

4.3.2 Varying The Field Size

The default field size was 100X100 sq units. Now, we increased it to 1000X1000 sq. units. The simulation led to the following new graphs.

4.3.2.1 For Less Number Of Nodes (n=18)

For less number of nodes and large field size, the distinction between the energy consumption of the two approaches is very clear. Due to less message overhead, the multi-hop algorithm performs better than our proposed algorithm when the number of nodes are few.

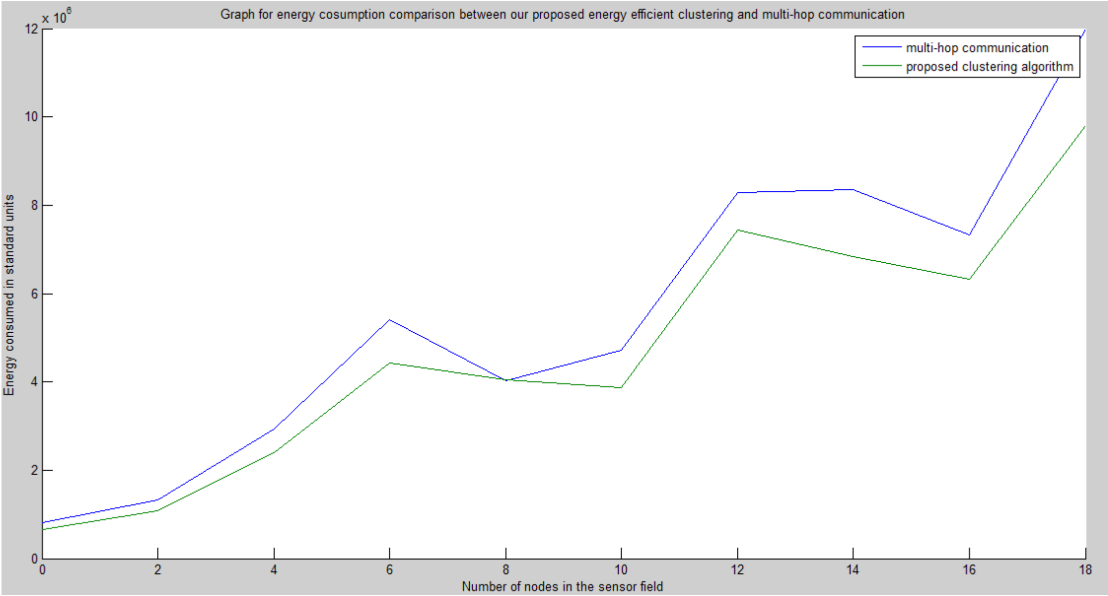


Figure: 4.5: Energy Comparison Varying Field Size With nodes = 18

4.3.2.2 For More Number Of Nodes (n=90)

With increase in the number of nodes, the distinction becomes clearer, slowly but surely. The difference between the energy consumption levels of the two approaches gradually increases. And it will become more and more distinctive as the number of nodes increases.

4.3.3 Variation In The Density Of Deployment

We took the average case in varying the density of deployed nodes. We decreased the density of nodes from 440 nodes in 100X100 field to 110 nodes in 100X100 field. Taking the average case of number of nodes (i.e. taking n=90) we found out that there is not much difference in the energy consumed by the two algorithms. Our proposed algorithm though performs better as the number of nodes increases, but the difference in the energy consumption is not so high as compared to when the density was high.

This is because when the nodes are scattered far away any clustering algorithm performs similar to direct or multi-hop routing.

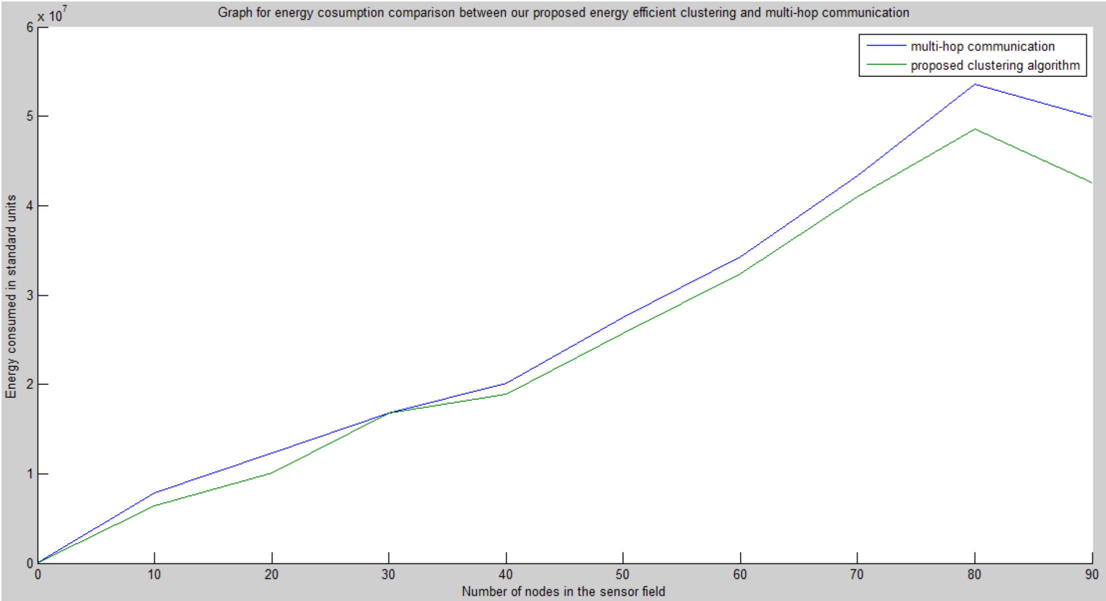


Figure 4.6: Energy Comparison Varying Field Size With nodes = 90

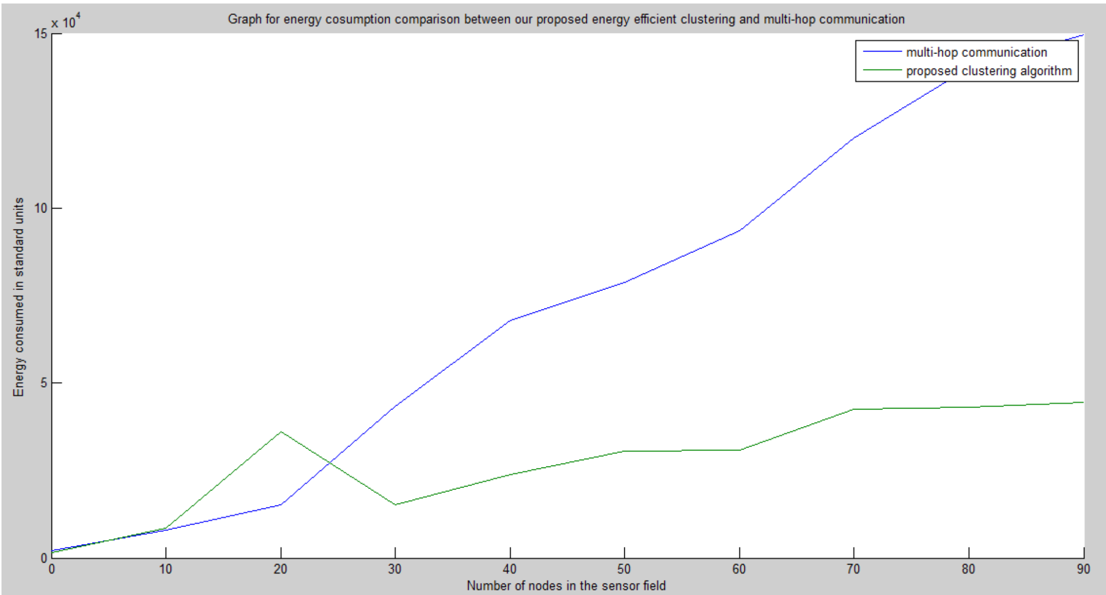


Figure 4.7: Energy Comparison Varying Density Of Deployment

CHAPTER 5

CONCLUSION

Our proposed algorithm shows better results in terms of energy efficiency and hence in the enhancement of the network lifetime.Some of the following points need to be taken care of:

Selection of appropriate h value: h value can be varied from the lowest to the highest possible value. In clustering organization. Efficient in the terms of messages passed, energy consumed as well as the lifetime of the network. This optimal h value will be different for different clustering networks. This value can be found only by multiple simulations of the proposed clustering environment with varying values of h. Thus, h remains the most important issue to be addressed by our algorithm.

Extra memory requirements: For each node there will be some parametres which were not present in any of the algorithms proposed earlier. These parametres are h and GRID_ID. While h is needed to identify the boundary nodes, GRID_ID is used for identifying to which grid does the particular node belong to. These values will be decided and distributed by the Base Station(BS). Though each node is provided a specific and limited memory, the extra requirement of memory can be an issue.

Initial heavy computation by the Base Station: Initially the Base Station has to perform a number of computations and also transmissions to establish our network. Though Base Station has a lot of computation and transmission power, but with increase in the size of a Wireless Sensor Network, it would become tiresome for the Base Station to keep up with heavy computations and transmissions.

FUTURE SCOPE

The proposed algorithm is designed to improve upon the random update clustering method. As the outline of the proposed algorithm is simple, there is a scope of improvement in the algorithm. The main area of improvement can be finding the optimal value of h parameter. This can be done by repeating simulations of this algorithm with varying values of h. Thus, finding out which is the most optimal value of h, which leads to maximum efficiency in the clustering environment. Efficiency is usually measured in the lifetime of the Wireless Sensor Network. So, the future scope mainly inculcates a plausible algorithm to find an optimal value of h.

REFERENCES

[1] Fan Wang, Sining Wu, Kunpeng Wang, and Xiaopeng Hu " Energy-Efficient Clustering Using Correlation and Random Update Based on Data Change Rate for Wireless Sensor Networks" IEEE Sensors Journal, vol. 16, no. 13, July 1, 2016.

[2] Petr Musilek, Pavel Kromer, Tomas Barton " E-BACH: Entropy-Based Clustering Hierarchy for Wireless Sensor Networks" IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2015.

[3] Nguyen Thi Thanh Nga, Nguyen Kim Khanh, Son Ngo Hong "Entropy-based Correlation Clustering for Wireless SensorNetworks in Multi-Correlated Regional Environments" IEIE Transactions on Smart Processing and Computing, vol. 5, no. 2, April 2016.

[4] Kun Xiao, Baolin Sun, Chao Gui, Hua Chen " A Novel Energy Entropy Based on Clusterhead Selection Algorithm for Wireless Sensor Networks" International Journal of Grid and Distributed Computing vol. 9, no. 3 (2016).

[5] "MathWorks for Matlab". [Online]. Available : https://in.mathworks.com

[6]J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey,"Computer Network, vol. 52, no. 12, pp. 2292–2330, 2008.

[7] H. F. Rashvand and J. M. A. Calero, "Distributed sensors," in Distributed Sensor Systems: Practice and Applications. New York, NY, USA: Wiley,2012,pp. 1–29.

[8] IEEE/WIC/ACM International Conference on Web Intelligence and
Intelligent Agent Technology, 2015.

[9] Reenkamal Kaur Gill, Priya Chawla, Monika Sachdeva, " Study of LEACH Routing Protocol for Wireless Sensor Networks". International Conference on Communication, Computing & Systems (ICCCS–2014).

[10] Meena Malik, Dr. Yudhvir Singh, Anshu Arora, " Analysis of LEACH Protocol in Wireless Sensor Networks". International Journal of Advanced Research in Computer Science and Software Engineering, vol 3, Issue 2,February 2013.

[11] Stephanie Lindsey, Cauligi S. Raghavendra, "PEGASIS: Power-Efficient Gathering in Sensor Information Systems",Computer Systems Research Department The Aerospace Corporation P. O. Box 92957 Los Angeles, CA 90009-2957.

[12] Rajesh Chaudhary Dr. Sonia Vatta, “Review Paper on Energy-Efficient Protocols in Wireless Sensor Networks”, IOSR Journal of Engineering (IOSRJEN), vol. 04, Issue 02, February 2014.

[13] Tamanna, Anshu Sharma, “Analyze and implementation of TEEN Protocol in Wireless Sensor Network”, International Journal of Innovative Research in Science, Engineering and Technology, vol. 5 Issue 3, March 2016.

[14] Supriya Dhauta,Ripul Rishi, “ Review on LEACH-Homogeneous and Heterogeneous Wireless Sensor Network”,International Journal of Innovative Research in Computer and Communication Engineering, vol.3, Issue 5, May2015.

[15] Mohd. Waes Siddiqui,Ms.Vandana Dubey ,“A Study Of Hybrid Energy Efficient Distributed (HEED) Protocols For Heterogeneous Wireless Sensor Networks”, 3rd International Conference on Recent Innovations in Science Engineering and Management, 27 February 2016.

[16] Sasikumar M, Dr. R. Anitha, “Performance Evaluation of Heterogeneous-HEED Protocol for Wireless Sensor Networks”, International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, Issue 2, February 2014.

[17] Stanislava Soro, Wendi B. Heinzelman, “Prolonging the Lifetime of Wireless Sensor Networks via Unequal Clustering ”, University of Rochester,Rochester, NY 14627.

[18] Yumna Fatma Farooqi, A.K.Daniel, “ Clustering Nodes in Wireless Sensor Network:A Survey”, International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, Issue 4, April 2016.

[19] N. Gautam and J.Y. Pyun, “Distance aware intelligent clustering protocol for wireless sensor networks,” IEEE Journal of Communications and Networks, vol.12, no.2, April 2010, pp. 122–129.

[20] A.U. Khattak, G.A. Shah and M. Ahsan, “Two-tier cluster based routing protocol for wireless sensor networks,” in Proceedings of IEEE/IFIP 8th International Conference on Embedded and Ubiquitous Computing (EUC), HongKong, December 2010, pp. 410–415.

[21] M. Ye, C. Li, G. Chen and J. Wu, “An energy efficient clustering scheme in wireless sensor networks,” Ad Hocand Sensor Wireless Networks, vol. 3, April 2006, pp.99–119.

[22] Y. Sangho, H. Junyoung, C. Yookun and J. Hong,“PEACH: Power-efficient and adaptive clustering hierarchy protocol for wireless sensor networks,”Computer Communications, vol. 30, no.14-15, October 2007.
[23] S. Banbyopadhyay and E.J. Coyle, “An energy efficient hierarchical clustering algorithm for wireless sensor networks”, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications IEEE Societies (INFOCOM 2003), vol.3, April 2003.