International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 07 | July -2017 www.irjet.net p-ISSN: 2395-0072

Energy efficient clustering algorithm based on Expectation Maximization for homogeneous WSN

Neha Sharma[1], Kanika Sharma[2]

[1] *M.E Student, Electronics and Communication Engineering, NITTTR, Chandigarh, India*
[2]*Assistant Professor, Electronics and Communication Engineering, NITTTR, Chandigarh, India*

Abstract - *The wireless sensor network comprises of large number of nodes deployed over an area that together operate to perform a task. The nodes have sensing, computing and communication features. The main focus is on enhancing the energy efficiency of network and improving its lifetime. Since the nodes are constrained in power, energy, computation it is necessary to deal with these efficiently. This paper includes a new method of energy efficiency and lifetime improvement by changing the way the CHs are selected and clustering is done. Unequal clustering approach is used where in as the distance of cluster from BS decreases the size of cluster also decreases. This helps in balancing the load of the network. Paper includes CH selection using EM- expectation maximization algorithm that gives improved results over LEACH, PEGASIS and PLEACH protocols. Simulation is done in MATLAB and results show that the proposed algorithm has outperformed existing ones by significantly decreasing the number of dead nodes and energy consumption per round.*

Key Words: Wireless Sensor Network, Base station, Cluster head, Base station, Energy Efficiency.

1. INTRODUCTION

Wireless sensor networks (WSNs), are also termed as actuator networks are small devices that are meant for collection of physical data from the surroundings, then the collected data is routed through the network to sink node or base station (BS).The Nodes have sensing, computation, and wireless communications capabilities over short distances [1]. These are used for the applications like environment monitoring, military surveillance and controlling processes in industrial process control. Even if the single node is not having sufficient levels of energy, the combined power of the whole network is enough for carrying out the required sensing our large geographic area [2]. These nodes are deployed randomly and do their task unattended also nodes have self organizing property in case of any node failure scenario. For computation purposes the nodes are equipped with on board microprocessors either 16bit or 3 bit [3]. A large number of nodes deployed over an extended area improve the output of the network. But since the nodes are battery run and are irreplaceable conserving energy becomes an important aspect [4].Thus the main aim of WSN is to maintain the correct data transfer over longer distances for larger durations by using appropriate management techniques. Various techniques have been developed over the years relating to this and the work is still in progress.

LEACH (low energy adaptive clustering hierarchy) is one of the first type of hierarchical protocols, is self organizing and adaptive in nature. It works on the assumption that all the nodes are identical, sink node is fixed, node adopt same antenna and the channel is symmetric.

In LEACH sensor nodes are organized into clusters than a particular Cluster head is selected for each cluster on the basis of threshold set for it. Here the sensor nodes are organized into clusters than a particular Cluster head is selected for each cluster on the basis of threshold set for it. The protocol operates in two phases set up and steady state for selection of CHs and data transmission respectively. TDMA access is used to reduce the inter and intra cluster collisions. It is used in constant tracking applications since the CH is selected dynamically at a time interval. Most of communication is restricted within the cluster hence it provides scalability in network. The traffic load is controlled by the cluster head through its data aggregation step. Although the protocol brings in improvements in terms of lifetime but has disadvantage that it cannot be applied to larger area, CHs may get concentrated in some part of network, its clustering involves extra amount of overheads.

PEGASIS (power efficient gathering in sensor information systems) is an enhancement over LEACH and is a near optimal chain based protocol. To increase the lifetime nodes make communication only with the neighboring nodes in turns. The second round starts when the round of all the nodes communicating with BS ends. By draining the power uniformly in the network this provides the advantage of reduced power requirements. Since there is local coordination among the nodes bandwidth consumption decreases leading to improved lifetime. The overall improvement is also on the account of fact that the overhead due to dynamic clustering gets reduced. But the disadvantage that arises is that the protocol needs dynamic adjustments to be done in the network topology.

Hence in this paper a new CH selection technique using EM algorithm is proposed along with unequal clustering. This overcomes the issues of above mentioned protocols and shows improvements in objectives. The remaining of paper is organized as follows: Section II includes clustering, its importance and various techniques

involved; Section III includes description of EM, uneven clustering, algorithm used for clustering; Section IV includes the simulation results, comparative analysis of proposed algorithm with other algorithms and the Section V concludes the paper.

2. CLUSTERING AND ITS VARIOUS TECHNIQUES

In grid clustering, the entire area is divided in small grids, where each such grid is a cluster. Clustering refers to division of sensor field in some virtual groups according to pre defined rules, where the member sensor nodes of that particular group execute their own functions different from other group's functions [5]. Clustering involves making a group and then selecting a leader i.e. CH for that group that act as data aggregating center for all member nodes and the collaborated data is transferred further to sink [6]. This aggregation of data reduces the total amount of data that has to be communicated hence reducing the energy consumption. This creates the hierarchy of nodes, i.e. those nodes having low energy will act as normal member nodes of a cluster and one having higher energy are regarded as cluster heads because they have to perform more energy consuming tasks. In high level approaches, there are three important phases of clustering algorithms that includes- 1.Phase of formation of cluster 2.Phase of construction (CHs selection) 3.Phase of Maintenance. The time of construction phase should be less than the time duration of maintenance phase.

Clustering of network has various advantages such as- scalability, data fusion, less energy consumption, balancing of load, network lifetime maximization. The types of clustering process can be classified as even and odd on the basis of size of cluster; centralized or distributed on the basis of control manners of clustering; probabilistic or iterative on the basis of decision making; homogeneous or heterogeneous on the basis of nature of node.

Ali Abed in [7] described a dynamic algorithm for mobile networks; the mobile clustering mechanism (MCM) has been examined and analyzed appropriately. The anticipated method is hierarchical, dynamic and energy efficient algorithm. This system displays numerous clusters, with each cluster having a unique CH and two deputy CHs. The sensors start gathering the data only when the base station approaches in range with the cluster head. The performance of the projected algorithm has been assessed against the present LEACH-Mobile algorithm. This approach displays a large decrease in average communication energy and node death rate. The network lifetime has been extended by assimilating the fresh concepts to the proposed methodology. The purpose of procedure is to progress network lifetime of wireless sensor nodes in the network. In the prevailing procedure (LEACH-M), nodes keep on sensing the data and directing the data to its CH, and this CH directs data only when the BS arises in range with the CH.

V. Loscri et al. in [8] proposed a protocol TL-LEACH which is an improvement over LEACH. It employs primary cluster heads and secondary cluster heads. In such a way that it formed a two level hierarchy in the network. This scheme handles the situation where the base station is located away from the sensor network. The nodes are being categorized into three categories namely simple nodes, primary nodes, secondary nodes.

T.N.Quynh et al. in [9] gave a hierarchical routing algorithm, Energy and load balanced LEACH, which overcomes the disadvantages of LEACH. It considers the residual energy of nodes in selection of CH (cluster head). This protocol also reduces the possibility of selection of two cluster heads that are very close to each other, which is disadvantage of LEACH. EL-LEACH assumes homogeneous environment of the network.

GUPTA et al. in [10] has proposed an improved PEGASIS protocol where for collecting data two parameters are considered namely degree of nodes and their remaining energy. This generates multiple degree graphs that enhance the chain formation technique. The energy cost of network is improved by allowing the same node to be visited again this also reduces the search space for the algorithm. For selection of leader and transmission of data a factor is used that is inversely proportional to both degree and distance. The modified protocol reduces energy dissipation by half amount and improves lifetime by 40% in comparison to PEGASIS.

Razaque et al. in [11] describes that LEACH protocol uses a technique for making of cluster and how to choose a CH that introduces disadvantage in context of isolated nodes. Another algorithm PEGASIS that is near optimum chain based protocol had removed the disadvantage of LEACH but lacks dynamicity provided by LEACH. Thus to overcome the disadvantage of both and combine their advantages, a new approach called PEGASIS-LEACH (P-LEACH) is introduced which helps to reduce the limitations and provide enhancement above PEGASIS and LEACH both, with best amalgamation of chain type and cluster type protocol. For sending the information effectively an optimum routing technique is also followed in this method. The idea behind the stated protocol is to develop an ideal route with less energy consumption for wireless transmission and networking. In case of LEACH the data forwarding task is taken care by set of CH nodes, while the implementation of constructing a chain in PEGASIS, is done using algorithm that is energy efficient. Thus it can be stated that in P-LEACH within the clusters chain construction method is use for transferring data.

M. Alnuaimi et al. in [12] discussed the different problems in clustering when done in large scale WSN, various clustering protocols, and classification of these protocols on the basis of approach used for forming the cluster and the technique by which the data is aggregated at the sink node. Authors classify the clustering protocols into three category i.e.

hierarchal based (LEACH), chain or grid based (PEGASIS) and weighted based (HEED) clustering techniques. The implementation of border monitoring for intruder identification in rectangular shaped area is considered in simulation using MATLAB for different scenario based on the position of the base station. assumption is made that the line of the border area is split into rectangular shaped region each such region having one sink node and it is observed that when the sink node is at the center and at the bottom right corner LEACH, SEP and HEED performed better. But when placed at middle of the bottom line and at the bottom left corner lifetime of these protocols decremented by 12%.While PEGASIS maintained its lifetime in all the scenarios. Due to random selection of cluster head LEACH has shorter lifetime. HEED performs better as compare to LEACH because it utilizes residual energy of nodes of cluster head selection. Because of having advanced nodes SEP have longest network lifetime.

3. PROPOSED WORK

The proposed work includes two methods that are used for enhancing the network lifetime and increasing energy efficiency. The two scheme used are EM algorithm for cluster head selection using uneven clustering.

3.1 Expectation Maximization-

Expectation maximization is an iterative method to find maximum likelihood estimates of parameters in statistical models. The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and a maximization (M) step, which computes parameters maximizing the expected log likelihood found on the *E* step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step.

A likelihood function (often simply the likelihood) is a function of the parameters of a statistical model given data Likelihood functions play a key role in statistical inference, especially in estimating a parameter. In informal contexts, "likelihood" is often used as a synonym for "probability." In statistics, a distinction is made depending on the roles of outcomes vs. parameters. Probability is used before data are available to describe possible future outcomes given a fixed value for the parameter. Likelihood is used after data are available to describe a function of a parameter for a given outcome.

The likelihood of a parameter value, θ, given outcomes x, is equal to the probability assumed for those observed outcomes given those parameter values, that is

$$£(\theta \mid x) = P(x \mid \theta)$$

For many applications, the natural logarithm of the likelihood function, called the log-likelihood, is more convenient to work with. Because the logarithm is a monotonically increasing function, the logarithm of a function achieves its maximum value at the same points as the function itself, and hence the log-likelihood can be used in place of the likelihood in maximum likelihood estimation and related techniques. The EM algorithm proceeds from the observation that the following is a way to solve these two sets of equations numerically. One can simply pick arbitrary values for one of the two sets of unknowns, use them to estimate the second set, then use these new values to find a better estimate of the first set, and then keep alternating between the two until the resulting values both converge to fixed points.

Despite the fact that EM can occasionally get stuck in a local maximum as we estimate the parameters by maximizing the log-likelihood of the observed data, but the following reasons still make it beneficial-

a) The ability to simultaneously optimize a large number of variables
b) The ability to find good estimates
c) The ability to create both the traditional “hard” clusters and not-so-traditional “soft” clusters.

About EM returning both hard and soft clusters, hard clusters mean a disjoint partition of the data. This is normally what classifiers do. Soft clusters mean allowing for a data point to belong to two or more clusters at the same time, the level of membership in a cluster being expressed by the probabilities of the classes at the data point.

Applications of EM algorithm-

a) EM is frequently used for data clustering in machine learning and computer vision.
b) EM is also widely used in medical image reconstruction.
c) EM has application in structural engineering.

Wei Meng et al., in [13] described an efficient algorithm for estimating maximum likelihood (ML) called EM algorithm is presented with acoustic sensors for energy base multisource. The algorithm divides the energy of sensors into sub components where each component further acts as a new source, whose parameters such as its location, decay factor etc. are calculated. An iterative incremental method is used to increase the speed of this algorithm. EM algorithm is used

for estimating the parameters and has less computations requirements. EM algo is highly sensitive the value taken as initialization value and has chances of its getting trapped in local optimum value to avoid this proper step size of increment should be taken into consideration.

Robert D. Nowak et al., in [14] described two things are considered one is clustering and other is estimation of density in WSNs. The components that are sensed by sensor

are considered to be modeled as Gaussian components and EM algorithm is used for estimation of such components. Distributed EM involves sending a small set of values iteratively from node to node. In comparison to EM algorithm this does not need forward or backward message transfer in its implementation. This algorithm converges rapidly as it involves use of multiple EM steps repeatedly. The algorithm results in estimation of data without the need of it to be transmitted or even processed at some other centralized location. Results prove that the DEM has less power requirements, efficient encoding, feasible in dense environment.

3.2 Unequal clustering-

This refers to the clustering type in which each cluster has varying size. The size of cluster goes on decreasing as the distance from cluster to base station increases. This is done to avoid large burden on the cluster that is nearer to the sink node. Since the CH that is nearer to the sink will have high amount of data to be transferred and will consume more energy in data aggregation, computation and transmission. Thus by keeping small size cluster nearer to sink its CH would be less burdened for processing. Thus this appropriate choice of clustering will lead to uniform cluster distribution in the network and homogeneous energy consumption.

Jiguo Yu et al in [15] propose a cluster based routing protocol for wireless sensor networks with non uniform distribution. Here author has made use of energy aware distributed clustering scheme. Even size clusters were formed here using the competition range as the major factor. The routing algorithm forwards data packets to higher energy cluster heads. Simulation results show that this method has stable network and enhanced lifetime.

3.3 Proposed algorithm-

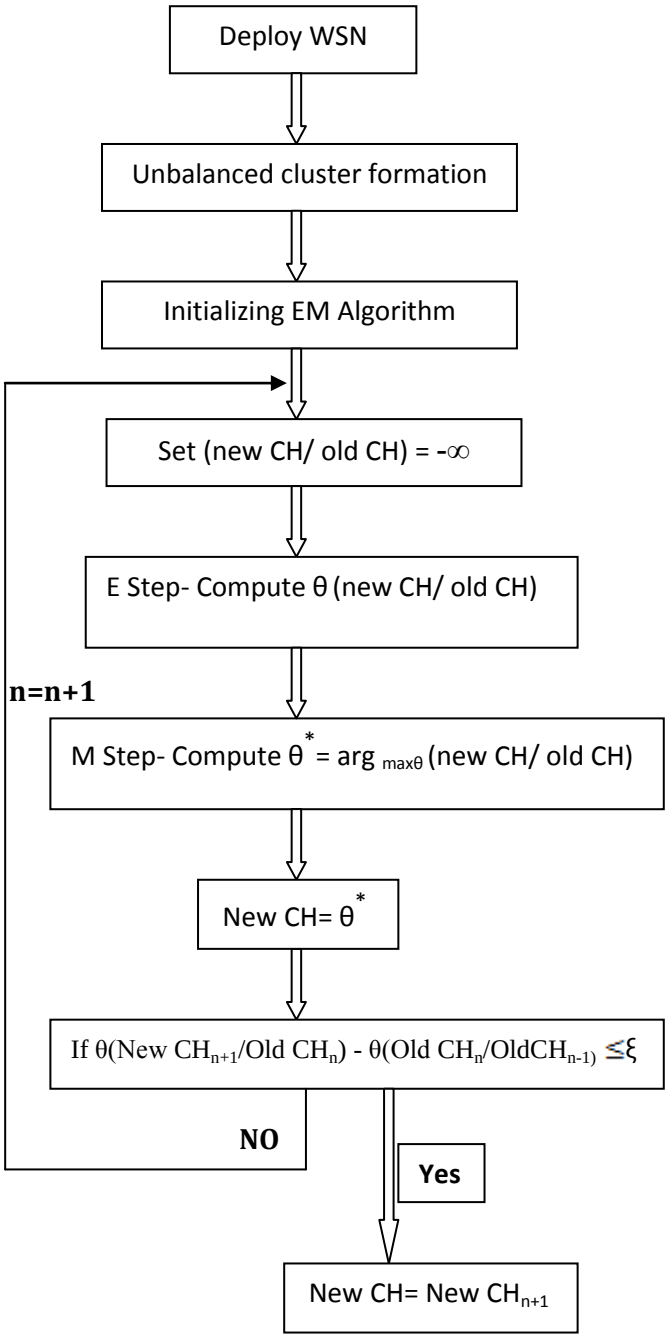

Fig 1. Flow diagram of algorithm

The algorithm consists of following steps-

Step1- This involves deploying the wireless sensor network that is the field area, nodes, sink node etc.
Step2- Cluster formation is done on the basis of distance of cluster head from sink node i.e. uneven clustering.
Step3- Initialization EM algorithm parameters.
Step4- Expectation of new CH is calculated on basis of old CH value.

Step 5- Maximization step involves maximizing the value achieved in previous step to get better CH selection.
Step6- The maximized value is assigned as new CH.
Step 7- If the comparative difference of the two values taken for adjacent rounds is less than the set threshold then new CH is set for next round otherwise process is repeated from Step 4.

4. SIMULATION RESULTS

The evaluation of proposed algorithm is done for 200 nodes deployed over a sensing region. The network area is taken 600*600. Initial energy is taken to be 5J for all the nodes representing homogeneous environment. Since the node consumes energy in each round the energy gets depleted after each iteration. The node whose energy gets completely zero is called dead node and we try to reduce the number of dead nodes so as to achieve higher lifetime of the network. Second parameter considered is decrease in energy consumption per round. The proposed algorithm is simulated in MATLAB with the parameters as mentioned below in table1-

Table1-Parameters

PARAMETER	VALUE
Network area	600*600
Base station location	150*50
No. of nodes	200
No. of rounds	180
Transmitter/receiver electronic circuit	$E_{elc}= E_{tx}= E_{rx}= 50nJ/bit$
Base station location	150*50
Energy of nodes	50J

The simulation shows that there is decrease in number of dead nodes per round with the proposed algorithm. The proposed algorithm has outperformed other three namely LEACH, PEGASIS and P-LEACH protocols as shown in chart 1.

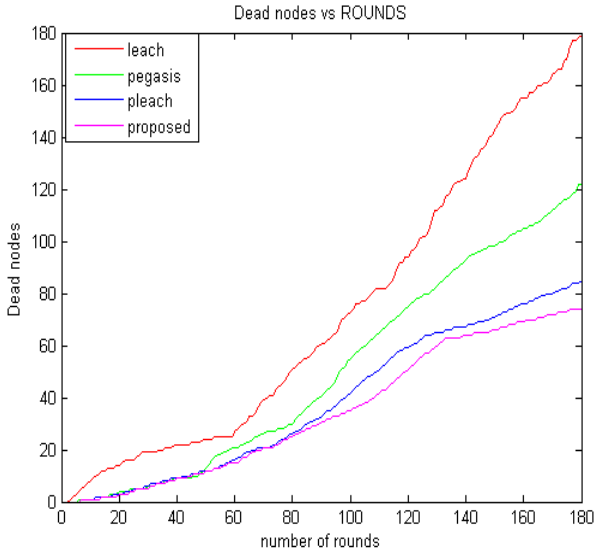

Chart 1-Number of dead nodes

The simulation is also carried for amount of energy consumed per round and is as shown in chart 2. There is decrease in energy consumption per round with the proposed scheme. The proposed algorithm has outperformed the other three namely LEACH, PEGASIS and P-LEACH protocols.

Table 2-Comparison of Number of dead nodes per round

Algorithm	No. of dead nodes
LEACH	180
PEGASIS	120
PLEACH	82
EM	67
% Improvement	18.29%

The table shows the number of dead nodes in all scenarios and makes comparison among them. The percentage of improvement achieved using the proposed algorithm is as mentioned above in table.

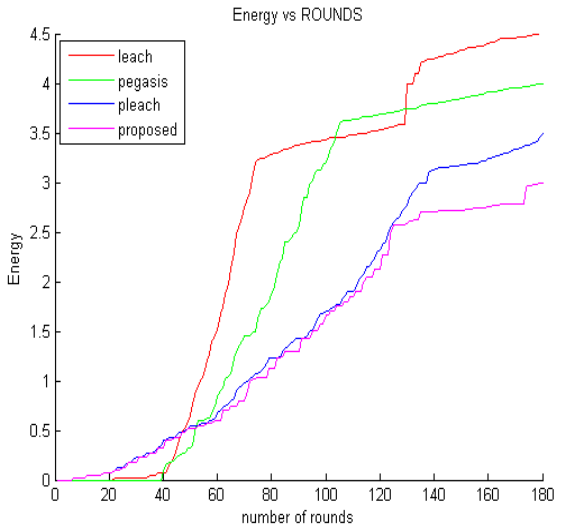

Chart 2- Energy consumption per round

Table3-Comparison of energy consumption per round

Algorithm	Energy consumption(Joules)
LEACH	4.5
PEGASIS	3.9
PLEACH	3.5
EM	2.8
% Improvement	20%

The table shows the energy consumption of nodes in all scenarios and makes comparison among them. The percentage of improvement achieved using the proposed algorithm is as mentioned above in table.