
M. D. Gbolagade  
Departments of Physical sciences, Computer Science Unit, Al-Hikmah University Ilorin, Nigeria

R. G. Jimoh  
Department of Computer Science, University of Ilorin Ilorin, Nigeria

K. A. Gbolagade  
Department of Computer Science, Kwara State University, Malete, Nigeria

O. V. Mejabi  
Department of Information and Communication Science, University of Ilorin Ilorin, Nigeria

ABSTRACT
Prolonging the network lifetime in wireless sensor networks (WSNs), Clustering has been recognized has one of the significant methods in achieving this, It entails grouping of sensor nodes into clusters and electing cluster heads (CHs) for all the clusters. CH’s accept data from relevant cluster’s nodes and forward the aggregate data to base station. A main challenge in WSNs is the selection of appropriate cluster heads. This work proposes a system that is efficient, scalable and load balanced. The proposed scheme combines two known algorithms namely k-means clustering and genetic algorithms based on the weaknesses identified in the two. The simulated data is obtained through the enhancement of clustering by the cluster head (base station) that helps in locating the nearest node that is important in the data transfer instead of transferring to a node that is not necessary, thereby wasting time and resources. The obtained simulation results indicate that this approach is efficient and last longer in elongating the battery life time than the conventional method by 60%.

Keywords
Efficient, Clustering, Base station, Sensor, Node, Lifetime

1. INTRODUCTION
With the trend of modern day technology in low-power electronic devices integrated with wireless communication capabilities are one of the new areas of research in the field of the wireless sensor networks (WSNs), it is observed that information is the major key in wireless sensor network any deployed area. The invention of wireless sensor networks is a great tool which helps to facilitate data transfer as well as helps to save resources and energy.

Wireless sensor network (WSN) is a group of sensor nodes which is used for monitoring and recording the physical conditions of the environment [3]. The sensed data are then transmitted to sink node and transmit using RF (radio frequency) channel. Wireless sensor network is a network made of numerous small independent sensor nodes. Size of a Sensor node is 35mm, it consists of a battery, radio, sensors, and a minimal amount of on-board computing power. The nodes self-organize their networks, rather than having a pre-programmed network topology. Because of the limited electrical power available, nodes are built with power conservation in mind, and generally spend large amounts.

By networking tiny sensor nodes, it becomes easy to obtain the data about physical phenomena which was very much difficult with conventional ways. Wireless sensor network typically consist of tens to thousands of nodes. These nodes collect process and cooperatively pass this collected information to a central location. Application of WSN are, area monitoring, environmental monitoring, air quality monitoring, forest fire detection, water quality monitoring, natural disaster prevention. Sensing capability provides automatic notification upon detection of some unusual events in the home. WSNs have unique characteristics such as low duty cycle, power constraints and limited battery life, redundant data acquisition, and dynamic network topology, e.t.c.[9]. In sensor networks, data transmission, signal processing and hardware operation are the three purposes for energy consumption. When a sensor network is deployed, one important question of energy-efficient and reliable data transport, affects the overall performance of the application, thus becoming particularly challenging. Limited and unrechargeable energy provision is considered to be the most challenging issue in sensor networks [4].

In this paper, we propose a new technique for the selection of the sensors cluster-heads based on the amount of energy remaining after each round. As the minimum percentage of energy for the selected leader is determined in advance and consequently limiting its performance and nonstop coordination task, the new hierarchical routing protocol is based on an energy limit value threshold preventing the creation of a group leader, to ensure reliable performance of the whole network.

In the rest of this paper, section 2 gives related work, sections 3 briefly discusses about techniques of U LEACH clustering protocols, section 4 presents details about simulation and results; in section 5 Paper is concluded.
2. RELATED WORK

A protocol that minimizes the total consumed energy to each the destination is proposed in [7]. The drawback with this approach is that if all the traffic is routed through the minimum energy path to the destination, the batteries in the nodes in that path can be drained out fast.

EECH [2] proposes appointed cost function that gets into account sensor to CH and CH to sink distances to consider the transmission costs. HEED [12] initiates average attainable power, which happens to be the average minimum transmission power to contact a CH from its neighbor nodes. This is used as an estimate of the prospective intra-cluster communication costs if the node chooses to act as a CH.

The discussion of PEBECS [13] and UCR [5] was based on assume network-wide announcements throughout the cluster formation process. Though, such an assumption not only decreases energy efficiency and also applicable to small-scale networks only.

LEACH (Low-Energy Adaptive Clustering Hierarchy) [8, 9] is one of the first and popular hierarchical routing algorithms for sensor networks. The idea is to form clusters of sensor nodes based on received signal strength and use cluster heads as routers to the sink. Each node choose itself as a cluster head (CH) based on a probabilistic scheme and broadcast its accessibility to all the sensor nodes present in the deployed area. Although the density of LEACH is low, the algorithm is not energy efficient due to irregular distribution of the CHs. LEACH generates homogeneous load distribution and in addition, the localized coordination scheme used in LEACH produces good scalability for cluster formation even though LEACH also has some flaws. Because of probabilistic criterion, good Cluster Heads (CH) distribution cannot be a certain and which can leads to some nodes not having any CH in their range. The CHs are a likely to have a long communication range for the quick receiver of data in Base Stations directly which is not always a pragmatic assumption.

3. METHODOLOGY

Clustering has been proved to be an effective method to increase the lifetime of WSN [11]. To use this effectiveness of clustering the researcher tried to present a new technique for clustering the sensor nodes. The main motivation for proposing this system is applying the goodness of two effective population algorithm K-Means and the Genetic Algorithm. The researcher tried to develop a clustering method depending on the energy of the nodes in order to extend the lifetime of WSN.

The model consist of two major phase i.e. the cluster formation and the cluster head selection. The base station executes the algorithm to cluster the sensor nodes and assign the proper roles to them (cluster head). The K-means algorithm helps to setup clusters, it will take the set of node input with their location and then evaluate the distance between them to form the clusters and then use the mean value of their distance because K-means uses clustering means. The genetic algorithm (GA) on the other hand takes each cluster as a new population and performs child generation, mutation, and crossover. It evaluates the new population with fitness function and selects the best population with the best nodes for cluster head in each given cluster; the fitness function selection is based on energy level and the distance from the sink. The sink node is the base station that has unlimited energy supply because it is responsible for getting all the information needed from the cluster head.

3.1 K-Means Clustering

K-Means is the simplest algorithm used for clustering which is unsupervised clustering algorithm. This algorithm partitions the data set into k clusters using the cluster mean value so that the resulting clusters intra-cluster similarity is high and inter-cluster similarity is low. K-Means is iterative in nature.

The K-Means algorithm has the following steps:

STEP [1] Arbitrarily generate k points (cluster centres), k being the number of clusters desired.

STEP [2] Calculate the distance between each of the data points to each of the centres, and assign each point to the closest centre.

STEP [3] Calculate the new cluster centre by calculating the mean value of all data points in the respective cluster.

STEP [4] With the new centres, repeat step 2. If the assignment of cluster for the data points changes, repeat step 3 else stop the process.

The distance between the data points is calculated using Euclidean distance as follows. The Euclidean distance between two points or tuples,

$$m_1 = (m_{11}, m_{12}, \ldots, m_{1n})$$

$$m_2 = (x_{21}, m_{22}, \ldots, m_{2n})$$

$$Dist(m_1, m_2) = \sqrt{\sum_{i=1}^{n} (m_{1i} - m_{2i})^2}$$

The effectiveness of clusters is evaluated based on uniformity of node distribution.

Intra-cluster Distance In intra-cluster the distance between the cluster nodes to its cluster centres to determine whether the clusters are compact is calculated using the below formula [1].

$$intra = \frac{1}{K} \sum_{I=1}^{K} \sum_{Z_{i} \in C_i} \| m - Z_{i} \|$$

Where $N$ is the number of nodes in the network, $K$ is the number of clusters, and $Z_i$ is the cluster centre of cluster $C_i$.

Inter-Cluster Distance

This is the distance between clusters[1]. We calculate this as the distance between cluster centres, and take the minimum of this value, defined as

$$inter = \left( \| Z_i - Z_j \| \right)^2$$

Where $I = 1, 2 \ldots K-1$ and $j = 1 + 1 \ldots K$

Then we pick only the minimum value.
3.2 Channel Propagation Model

In the wireless channel, the electromagnetic wave propagation can be modelled as falling off as a power law function of the distance between the transmitter and receiver. The free space model which considered direct line-of-sight and two-ray ground propagation model which considered ground reflected signal also, were considered depending upon the distance between transmitter and receiver. If the distance is greater than \( d_{\text{crossover}} \), two-ray ground propagation model is used. The crossover is defined as follows using the formula below:

\[
\frac{4\pi r^2}{d} = \frac{4\pi r^2}{d_{\text{crossover}}}
\]  

(6)

Where, \( L \leq 1 \) is system loss factor. In the equation above, \( h_t \) is the height of the receiving antenna; \( h_r \) is the height of the transmitting antenna and is the wave length of the carrier signal. Then the transmitting power will be attenuated based on the formula given below:

\[
P_d(d) = \left\{ \begin{array}{ll}
\frac{P_t G_t G_r \lambda^2}{(4\pi d)^2} & \text{if } d < d_{\text{crossover}} \\
\frac{P_t G_t G_r \lambda^2}{d^2} & \text{if } d \geq d_{\text{crossover}}
\end{array} \right.
\]  

(7)

Where, \( P_t \) is the received power at distance \( d \), \( P_t \) is the transmitted power, \( G_t \) is the gain of the transmitting antenna and \( G_r \) is gain of the receiving antenna.

Figure 1: Radio energy dissipation model

3.3 Radio Energy Dissipation

We assumed a simple model for the radio hardware energy dissipation where the transmitter dissipates energy to run the radio electronics and the power amplifier, and the receiver dissipates energy to run the radio electronics as shown in Figure 3.1. Using this radio model, to transmit \( y \)-bit message at distance \( d \) the radio expends:

\[
E_{\text{TX}}(y, d) = E_{\text{TX-elect}}(y) + E_{\text{TX-amp}}(y, d)
\]  

(8)

\[
E_{\text{TX}}(y, d) = E_{\text{elec}} * (y) + E_{\text{amp}} * y * d^2
\]  

(9)

The receiving this message it expends:

\[
E_{\text{RX}}(y, d) = E_{\text{Rx-elect}} * (y)
\]  

(10)

\[
E_{\text{RX}}(y, d) = E_{\text{elec}} * (y)
\]  

(11)

3.4 Genetic Algorithm

In the method, GA is used to maximize the lifetime of the network by means of rounds. Binary representation of the cluster is used and each sensor node corresponds a bit. Cluster heads (CH) are represented as “1” and non-CH nodes are represented as “0”. The representation of a cluster is called a Chromosome or Genome, a collection of bits. Initially the GA starts with a population, a pre-defined number of chromosomes, consists of randomly generated individuals. Then GA evaluates each chromosome by calculating its fitness. Fitness of a chromosome depends on some fitness parameters that are explained in Section 3.5 after evaluating the fitness of each chromosome in the population, GA selects the best fit chromosomes by using a specific selection method based on their fitness values and then applies two operators, Crossover and Mutation, respectively. These operations are carried out to produce a new population better than the previous one for the next generation.

3.5 Fitness Function

The aim of this study is to maximize the lifetime of the network in other to achieve this, fitness function must be considered. The fitness function has 3 parameters. These Parameters are:

- RFND: The round which first nodes dies,
- RLND: The round which last node dies,
- C: The cluster distance.

The cluster distance is the sum of the distances from the member nodes to the CH and the distance from the cluster heads to the Base station (BS). For a cluster with \( k \) member nodes the cluster distance is denoted as follows [5].

\[
C = \sum_{i=1}^{k} d_{ih} + d_{hs}
\]  

(12)

Where \( d_{ih} \) is the distance from node \( i \) to the cluster head \( h \) and \( d_{hs} \) is the distance from the cluster head \( h \) to the BS node \( s \).

The fitness function, \( F \), is a function of all parameters described above and used in the genetic algorithm. It is defined as follows:

\[
F = \sum_{i} \left( f_i \times w_i \right) \forall f_i \in \{C\}
\]  

(13)

The \( w \) value is an application-dependent weight of a fitness parameter that indicates which parameter is more effective for the function. We can make a fitness parameter more important than the other by changing its weight or we can give them equal importance by setting the weights equal.

3.5.1 Selection

In selecting the best nodes that fits for the cluster head

This process determines which of the chromosomes from the current population will create new child chromosomes by doing crossover and mutation. The new child chromosomes join the existing population. The new population with new child chromosomes will be the basis for the next selection. The chromosomes which have better fitness values have bigger chance to be selected. In the proposed method, Roulette-Wheel selection method is used.

3.5.2 Crossover

Crossover is a genetic operator that generates two new child chromosomes (nodes) from two parent chromosomes. The easiest way to do this is to choose a random crossover point and the two parent chromosomes exchange information after that point.

Crossover is done after the selection process and depends on a probability defined initially before GA starts. The probability that the crossover will take place depends on the crossover rate.
3.5.3 Mutation
After a crossover is performed, mutation takes place. This is to prevent falling all solutions in population into a local optimum of solved problem. Mutation changes every bit of the new child chromosome with a probability called mutation rate.

Original chromosome: 11100101
Mutated chromosome: 11000111

Figure 1 show the table of mutation.

4. SIMULATION RESULTS

4.1 The U-LEACH protocol
In this section, we describe the network model. Assume that they are N sensors nodes, which are evenly dispersed within a MxM square region. The nodes always encompass with data to transmit to a base station, which is frequently far from the sensing area. Without loss of generality, we assume in first simulation that the base station is located at the center of square region, and in the second simulation we assume that the base station is in the peak or square region (99m X 99m). The network is structured into a clustering hierarchy, and the cluster-heads execute fusion function to reduce correlated data produced by the sensor node within the clusters. The cluster heads transmit the aggregated data to the base station directly, we believe that the nodes are in static mode eradicate the frequent change of topology.

We evaluate the performance of U-LEACH protocol using C. net runtime, and compare it’s performance with LEACH, using the same initial values and following the same scenario and energy model. The simulation parameters are taken in accordance with the designed network scenario and used energy model. 50 static sensor nodes and one sink (base station) are distributed uniformly in the square region of observation O with M = 500m and W = 500 m. In the simulations, all nodes have initial battery energy of 0.5 J. The Data packet size is taken to be 512 bits see table (1). The simulation time can be varied to study the effect of the proposed scheme under different scenarios as per the requirements. In proposed work, the whole results are obtained by taking 100 seconds simulation time. Two ray ground was used for the radio propagation model. Anode is measured dead if its energy level reaches 10 to 0 joules. The general simulation parameters are:

Table 1 indicates the parameters used for the deployments of sensor nodes across the simulation area.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area network</td>
<td>500*500</td>
</tr>
<tr>
<td>Simulation Time</td>
<td>100s</td>
</tr>
<tr>
<td>Sensor Model</td>
<td>MICA2</td>
</tr>
<tr>
<td>Initial Energy</td>
<td>0.5j</td>
</tr>
<tr>
<td>Data packet length</td>
<td>512 n j / bits</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>50 and 100</td>
</tr>
</tbody>
</table>

Figure 2 shows the Random deployments of sensors in simulation area.

In our simulation, Energy Consumption of the entire network was taken as the main performance evaluation parameter. The overall energy consumption of network includes the energy consumed by all the nodes in sending and transmitting the data. Also, the network lifetime is used as the second parameter for evaluating our model. This is computed by considering the time when first node of the network dies out. It is the difference of total energy of the network and the summation of average used energy of nodes and their energy deviation.

Figure 3. Show the Sample Trace File for communication/residual energy

4
First, the researcher evaluates the performance of our proposed algorithm and compares LEACH. In the simulation results as reported in Table 2, shows the performance in terms of network throughput energy consumption and energy utilization efficiency of the respective algorithms with different simulation time. Our proposed algorithm energy consumption is low and therefore, this makes energy depletion in each node to be less and results in longer network lifetime.

Table 2. Shows the table of the energy consumption

<table>
<thead>
<tr>
<th>Time</th>
<th>Energy proposed</th>
<th>Energy leach</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.223</td>
<td>0.250</td>
</tr>
<tr>
<td>40</td>
<td>0.245</td>
<td>0.276</td>
</tr>
<tr>
<td>60</td>
<td>0.267</td>
<td>0.280</td>
</tr>
<tr>
<td>80</td>
<td>0.281</td>
<td>0.292</td>
</tr>
<tr>
<td>100</td>
<td>0.299</td>
<td>0.311</td>
</tr>
</tbody>
</table>

In our simulation we have clustered the network in same number of clusters, likewise the cluster heads was placed randomly or separated with some minimum distance. Results show that if the cluster heads are separated with some minimum distance it gives the better performance.

Thus we plot a graph of energy against time then we have the below graph.

4.2 Results Analysis

The results indicate that energy consumption is much lesser in our proposed algorithm compared with LEACH if the cluster heads are separated with minimum distance with regards to Figure 4, which shows the graph of the energy consumption for the data transmission in a network. It is inferred that the slope of curve in our graph gives the degree of optimization of the network lifetime and, a better algorithm will have the greater slope, while figure 5.indicate the energy dissipations of both U LEACH and LEACH. We were able to deduce that the proposed algorithm improve lifetime and stability of nodes. U-LEACH has longer stability period than LEACH this indicates that U-LEACH is more efficient than LEACH.

5. CONCLUSION

This research work has been able to achieve some far reaching objectives by optimizing the lifetime of wireless sensor network to become more mobile, scalable, efficient and balance through the use of two known population algorithms that is K-means and genetic algorithms. The process has a model that consists of two major phase i.e. the cluster formation and the cluster head selection. The base station executes the algorithm to cluster the sensor nodes and assigns the proper roles to them (cluster head). The k-means algorithm helps to setup clusters, it will take the set of node input with their location and then evaluate the distance between them to form the clusters and then use the mean value of their distance because K-means uses clustering mean. The genetic algorithm on the other hand takes each cluster as a new population and performs child generation, mutation, and crossover. It evaluates the new population with fitness function and selects the best population with the best nodes for cluster head in each given cluster; the fitness function selection is based on energy level and the distance from the sink. It follows that if this system is applied in network data selection is based on energy level and the distance from the sink. It follows that if this system is applied in network data transmission in a network, it will be noted that wireless sensor network nodes will last longer than they use to and more energy will be gained and there will be no data lost and the quality of information that will be gotten will still remain the same if not better than the former.

6. REFERENCES


