Energy Efficient Clustering Approach for Data Aggregation and Fusion in Wireless Sensor Networks

Rajesh K. Yadav¹, Daya Gupta¹ and D.K. Lobiyal²

¹ Computer Science & Engineering Department, Delhi Technological University, New Delhi, India, Emails: rkyadav@dce.edu, dggupta@dce.ac.in
² School of Computer & System Sciences Jawaharlal Nehru University, New Delhi, India, Email: lobiyal@gmail.com

Abstract: Wireless Sensor Network (WSN) is a network which formed with a huge number of sensor nodes which are deployed in an application environment to monitor the physical entities in a target area. Clustering is a widely used mechanism in wireless sensor networks to reduce the energy consumption by sensor nodes in data transmission. Partitioning the network into optimal number of clusters and selecting an optimal set of nodes as cluster heads is an NP-Hard problem. The NP-Hard nature of clustering problem makes it a suitable candidate for the application of evolutionary algorithms and particle swarm optimization (PSO). In this paper we first present particle swarm optimization based clustering (PSOBC) with discrete search space and then propose an energy efficient hybrid clustering protocol (EEHCP) for multilevel heterogeneous wireless sensor networks. The performance metrics of proposed methods evaluated and results are compared with well known clustering algorithms to validate the reduction in energy consumption, stability period and prolongs network life time.

Keywords: Wireless Sensor Network, PSO, discrete search space, data aggregation and clustering.

1. INTRODUCTION

A Wireless Sensor Network (WSN) consists of several spatially distributed autonomous devices (sensor nodes) with sensing and communication capabilities that cooperatively sense physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants at different locations & used in applications such as environmental monitoring, homeland security, critical infrastructure systems, communications, manufacturing etc. WSNs are data-driven networks that usually produce a large amount of information that needs to be routed across the networks. As sensor nodes are energy-constrained devices and the energy consumption is generally associated with the amount of gathered data [1-3]. Since energy conservation is a key issue in WSNs, Data fusion and Data aggregation is exploited in order to save energy. A strategy to optimize the routing task for the available processing capacity can be provided by the intermediate sensor nodes along the routing paths. Data aggregation is defined as the process of aggregating the data from multiple sensors to eliminate redundant transmission and provide fused information to the base station. The main goal of data aggregation
algorithms is to gather and aggregate data in an efficient manner so that lifetime of the network increases by
decreasing the number of packets to be sent to sink or base station [4-5], intern reduces the communication costs
and energy consumption. Grouping of sensor nodes into clusters has been widely used by researchers to satisfy
the scalability, high energy efficiency and prolong network lifetime objectives [6-7]. In clustering the whole
sensor network is partitioned into multiple groups of sensor nodes. Each group is called a cluster and each
cluster has a leader called cluster head that perform special tasks such as data aggregation and fusion. Clustering
has number of other benefits and corresponding objectives, In addition to supporting network scalability and
decreasing energy consumption through data aggregation.

Particle Swarm Optimization (PSO) is an optimization technique in which natural species social behaviors
are considered for the purpose of computation [8]. It is a swarm intelligence technique which is based on
population that performs optimization process with the objective of optimizing a fitness function. This approach
makes use of a swarm for the purpose of search on every particle and records the fitness value of each particle.
Then the particles are linked with their matching velocity. It will help the particle to make a move to a proper
location by considering the optimized fitness function’s cost [9-10].

The rest of this paper is organized as follows: Section 2 presents related works. In Section 3 the proposed
PSO based clustering and Hybrid clustering algorithms explained. In Section 4 the performance of the proposed
approaches is analyzed. Finally the conclusions are presented in Section 5.

2. RELATED WORK
A large number of cluster based routing techniques have been proposed in the recent days to increase the
network’s lifetime [6-15]. These algorithms have made the communication very effective and also increase the
throughput of the network. WSNs have many research challenges and network issues when deploying the sensor
nodes to monitor the physical world. Hierarchical routing protocols are appropriate for organizing the nodes to
increase the scalability of the WSNs. The traditional clustering algorithm LEACH [11] uses randomized rotation
with uniform clustering of local cluster heads to increase the scalability and network performance. In LEACH,
initially all the nodes are given the same amount of energy and have the same physical properties. These nodes
organized themselves into clusters with one node out them acting as CH. All the Non-CH nodes transmit their
data to their respective cluster head (CH). The function of CH is to perform data aggregation function or data
fusion function on this data, after performing these functions CH transmits the data to the BS. The cluster head
are re-elected periodically to balance the load of the network. The life time of the network has extended by
utilizing a HEED clustering protocol [7] this formed the clustering and cluster head selection based on the
residual energy of sensor nodes and the cost of communication from source to destination. Paper [12] proposed
Energy Efficient Hierarchical Clustering (EEHC) that increases the lifetime of the sensor network. However,
hierarchical clustering made overload in cluster heads and reduces its power sooner than the other nodes. Paper
[13] proposed a distribution scheme of cluster heads to reduce energy dissipation by avoiding unnecessary
redundancy and compared with existing LEACH it prolongs network lifetime. Paper [14] proposed energy
efficient adaptive multipath routing technique to reduce routing overhead and efficiently utilizes the energy
availability. LEACH-C [15] is a kind of improved LEACH. In LEACH-C, the location information and the
residual energy value of all the nodes will be sent to the base station at the beginning of each round. The authors
in [16] have showed that PSO outperforms both LEACH and LEACH-Centralized in terms of the network span
and the overall throughput. In [17] an Enhanced PSO-Based Clustering Energy Optimization (EPSO-CEO)
algorithm proposed in which clustering and clustering head selection are done using PSO algorithm.

Most of the early routing protocols proposed for wireless sensor networks do not consider heterogeneity in
the network and therefore are not able to take advantage of the heterogeneity present in the network. In [18]
authors proposed stable election protocol (SEP) which consumes energy from the nodes having high energy and
increase the stability period and life time of the network.
Particle swarm optimization (PSO), developed by Dr. Eberhart and Dr. Kennedy in 1995 and inspired by social behaviour of bird flocking or fish schooling is a population based stochastic technique to solve continuous and discrete optimization problems. PSO learned from the scenario and used it to solve the optimization problems. In PSO, each single solution is a “bird” in the search space. We call it “particle”. All of particles have fitness values which are evaluated by the fitness function to be optimized, and have velocities which direct the flying of the particles. The particles fly through the problem space by following the current optimum particles [19-20].

PSO is initialized with a group of random particles (solutions) and then searches for optima by updating generations. In every iteration, each particle is updated by following two “best” values. The first one is the best solution (fitness) it has achieved so far. (The fitness value is also stored.) This value is called pbest. Another “best” value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the population. This best value is a global best and called gbest. When a particle takes part of the population as its topological neighbours, the best value is a local best and is called lbest [21-22].

Suppose, there is a group of K random particles in an n-dimension searching space, the position of the $i^{th}$ particle is $X_i = (x_{i_1}, x_{i_2}, ..., x_{i_n})$, the personal best value of the particle is $pbest_i = (p_{i_1}, p_{i_2}, ..., p_{i_n})$, and the velocity of the particle is $V_i = (v_{i_1}, v_{i_2}, ..., v_{i_n})$. The best value obtained so far by any particle in the population is $gbest = (g_{1}, g_{2}, ..., g_{n})$. After finding the two best values, $pbest$ and $gbest$ the particle updates its velocity and positions as follows

$$v_{i_j} = w_v v_{i_j} + c_1 r_1 (p_{i_j} - x_{i_j}) + c_2 r_2 (g_{j} - x_{i_j}) \tag{1}$$

$$x_{i_j} = x_{i_j} + v_{i_j} \tag{2}$$

Where $w$ is inertia and used to control the trade-off between the global and the local exploration ability of the swarm, $c_1$ and $c_2$ are learning factors, $r_1$ and $r_2$ random numbers between 0 and 1.

3. PROPOSED WORK

3.1. Proposed PSO based Clustering Algorithm

3.1.1 Fitness Function

Success of our proposed algorithm will depend greatly on the formulation of fitness function. So we are defining a fitness function that includes all optimization criteria. Our aim is to minimize the intra-cluster communication energy and energy loss due to cluster head and base station communication, so we can define the fitness of a particle $i$ as

$$F(P_i) = E_1(P_i) + \mu \cdot E_2(P_i) \tag{3}$$

$$E_1(P_i) = \sum_{k=1}^{K} \sum_{n_{kh} \in C_k} \frac{f(n_{kh}, CH_k) - E_{\text{min}}}{E_{\text{max}} - E_{\text{min}}} \tag{4}$$

$$E_2(P_i) = \sum_{k=1}^{K} g(CH_k, BS) \cdot \frac{E_{\text{min}}}{E_{\text{max}} - E_{\text{min}}} \tag{5}$$
\[
f(n_j, CH_k) = \begin{cases} 
    s^2(n_j, CH_k) & \text{if } s(n_j, CH_k) \leq d_0 \\
    s^4(n_j, CH_k) & \text{if } s(n_j, CH_k) > d_0
\end{cases}
\]
(6)

\[
g(CH_k, BS) = \begin{cases} 
    d_{CH,BS}^2 & \text{if } d_{CH,BS} \leq d_0 \\
    d_{CH,BS}^4 & \text{if } d_{CH,BS} > d_0
\end{cases}
\]
(7)

\[
s(n_i, CH_k) = \min_{s_i, CH_k}^{k=1,2,\ldots,K}
\]
(8)

Where, \(d_{ij}\) is the distance between node \(i\) and node \(j\); \(s\) is a function that finds the minimum distance cluster head for a given node; \(f\) is a function whose value for a given node is proportional to the energy consumed in communication between the node and its cluster head; similarly \(g\) signifies the energy loss due to cluster head and base station communication; \(E_{max}\) and \(E_{min}\) are the maximum and minimum energy loss in the network. \(C_j\) is the \(k\)th cluster in a solution or particle.

\(E_1\) and \(E_2\) are two normalized functions that represent the energy dissipated in intra-cluster communication and due to communication between sink and CHs respectively. F is fitness function and our aim is to minimize this function.

\(\mu\) is a controlling parameter that controls the distance between base station and cluster heads. The higher the value of \(\mu\) will be the closer will be the CHs from BS. \(K\) is the optimal number of cluster heads. For each particle or solution we will choose \(k\) random nodes as cluster heads and remaining nodes will join the cluster whose CH is at minimum distance from it. Then we will evaluate the value of fitness function for each particle and will calculate \(pbest\) and \(gbest\). Then we will update the velocity vector and position vector according to equation (1) and (2).

3.1.2. A new operator \(\oplus_{NW}\)

We have defined a new operator \(\oplus_{NW}\) that when applied on a location with respect to a network, it returns a valid sensor node location in the network. In each iteration of our algorithm location of CHs updated in each particle or solution. Keeping this into consideration operator \(\oplus_{NW}\) is defined as follows: Suppose \(\hat{a} = (a_1, a_2)\) is any location with respect to a sensor network \(NW\) then \(\oplus_{NW} \hat{a}\) will return a valid location in network \(NW\). The operator \(\oplus_{NW}\) will first check if \(\hat{a}\) is a valid location than it return \(\hat{a}\) as it is; if not then it returns nearest valid location in the network \(NW\) toward base station with highest residual energy. After calculating new velocity and position using equation (1) and (2) operator is applied to the calculated positions to get valid new positions.

3.1.3. Working of proposed PSO Algorithm

1. create and initialize a K-dimension swarm of \(P\) particles by choosing \(K\) CHs with residual energy higher than average energy of network for each particle.

2. Repeat

3. for each particle \(i = 1, 2, 3 \ldots \), \(P\) do

4. if \(F(X_i) < lbest_i\) then

5. \(\text{lbest}_i = X_i\)

6. end
Energy Efficient Clustering Approach for Data Aggregation and Fusion in Wireless Sensor Networks

7. if $F(X_i) < gbest$ then
8. $gbest = X_i$
9. end
10. end
11. for each particle $i = 1, 2, 3 \ldots, P$ do
12. update velocity $V_i$ using equation (1)
13. update position vector $X_i$ using equation (2)
14. apply $\oplus_{NW}$ operator to updated position
15. end
16. until the maximum number of iteration reached

3.2. Proposed EEHCP Protocol

3.2.1. Optimal Number of Clusters

Optimal number of clusters $K_{opt}$ can be found using simple analysis as in [11]. Let us assume an area of $M \times M$ square meters with base station situated at the centre and $N$ sensor nodes are distributed over this area. The energy dissipated in the cluster head node assuming its distance from BS less than $d_0$ during a round can be given by the following formula:

$$E_{CH} = \left(\frac{N}{k} - 1\right)L.E_{elec} + \frac{N}{k}L.E_{DA} + L.E_{elec} + L.e_{fs}.d_{toBS}^2$$  \hspace{1cm} (9)

Where $k$ is the number of clusters, $E_{DA}$ is the data aggregation cost of a bit per report to the base station, and $d_{toBS}$ is the average distance between the cluster head and the base station.

The energy used in a non-cluster head node is equal to:

$$E_{nonCH} = L.E_{elec} + L.e_{fs}.d_{toCH}^2$$  \hspace{1cm} (10)

Here $d_{toCH}$ is the average distance between a cluster member and its cluster head.

Thus energy dissipated in a cluster per round:

$$E_{cluster} \approx E_{CH} + \frac{N}{k}E_{nonCH}$$  \hspace{1cm} (11)

The total energy dissipated in the network is equal to:

$$E_{tot} = L \left(2.N.E_{elec} + N.E_{DA} + e_{fs} \left(k.d_{toBS}^2 + N.d_{toCH}^2\right)\right)$$  \hspace{1cm} (12)

According to [7]:

$$d_{toCH}^2 = \int_{x=0}^{x=x_{max}} \int_{y=0}^{y=y_{max}} \left(x^2 + y^2\right) \rho(x,y) \, dx \, dy = \frac{M^2}{2\pi k}$$  \hspace{1cm} (13)

And
Differentiating $E_{tot}$ with respect to $k$ and equating to zero, the optimal number of constructed clusters can be found:

$$K_{opt} = \sqrt{\frac{N}{2\pi}} \sqrt{\frac{M}{d_{obs}} \frac{\epsilon_{fs}}{\epsilon_{mp}}}$$

(15)

### 3.2.2. Network Deployment Model

Most of the routing protocols for heterogeneous WSNs consider the random deployment of sensor nodes in the network space but to take advantage of heterogeneity present in the network in term of energy level we partitioned the network space into two zones: $\alpha$- zone and $\beta$ – zone.

This protocol assumes three levels of heterogeneity with three types of nodes: $\alpha$-nodes, $\beta$-nodes and $\omega$-nodes $\alpha$-nodes have the highest energy among all the nodes, $\beta$-nodes have energy less than $\alpha$-nodes but greater than $\omega$-nodes and $\omega$-nodes are normal nodes. We assume that a small fraction of total nodes are $\alpha$- or $\beta$-nodes while most of the nodes are $\omega$-node. As in SEP, we consider that the base station is not mobile and coordinates of the base station are known. The area in network space near the base station is $\beta$-zone and $\beta$-nodes are deployed in this region while the $\alpha$-nodes are deployed in the farthest region, i.e. $\alpha$-zone and $\omega$-nodes are randomly deployed in the whole network space.

### 3.2.3. Cluster Head Selection

Let us assume that $A$ is fraction of total nodes that are $\alpha$-nodes and $B$ is the fraction of total nodes that are $\beta$-nodes. The energy of a $\alpha$-node is $\alpha$ times higher than an $\omega$ node while the energy of a $\beta$-node is $\beta$ time higher than an $\omega$ node. If initial energy of an $\omega$-node is $E_0$ then total energy of all the nodes will be:

$$E_{total} = N (1 - A - B) E_0 + N.A.E_0 (1 + \alpha) + N.B.E_0 (1 + \beta)$$

(16)

Optimal probability of cluster head selection in case of homogeneous network is given by [4]:

$$P_{opt} = \frac{K_{opt}}{N}$$

(17)

Where $K_{opt}$ is optimal number of cluster heads as calculated above and $N$ is total number of nodes in network.

Now optimal probability of a node to be cluster head on the basis of residual energy can be calculated as:

$$\left(P_{\alpha}\right)_{i} = \frac{P_{opt} E_{i}(r)}{(1 + A\alpha + B\beta) \bar{E}(r)}$$

(18)

$$\left(P_{\beta}\right)_{i} = \frac{(1 + \beta) P_{opt} E_{i}(r)}{(1 + A\alpha + B\beta) \bar{E}(r)}$$

(19)

$$\left(P_{\beta}\right) = \frac{(1 + \beta) P_{opt} E_{i}(r)}{(1 + A\alpha + B\beta) \bar{E}(r)}$$

(20)
Here $E_i(r)$ is residual energy of $i^{th}$ node in $r^{th}$ round and $\overline{E}(r)$ is the average energy in the $r^{th}$ round.

Depending on the weighted probabilities the threshold values can be calculated as follows

$$T_\omega = \begin{cases} P_\omega & \text{if } \omega \in G'' \\ 1 - P_\omega (1 - r \mod \frac{1}{P_\omega}) & \text{if } \omega \in G' \\ 0 & \text{otherwise} \end{cases} \quad \text{(21)}$$

$$T_\beta = \begin{cases} P_\beta & \text{if } \beta \in G'' \\ 1 - P_\beta (1 - r \mod \frac{1}{P_\beta}) & \text{if } \beta \in G' \\ 0 & \text{otherwise} \end{cases} \quad \text{(22)}$$

$$T_\alpha = \begin{cases} P_\alpha & \text{if } \alpha \in G \\ 1 - P_\alpha (1 - r \mod \frac{1}{P_\alpha}) & \text{if } \alpha \in G' \\ 0 & \text{otherwise} \end{cases} \quad \text{(23)}$$

Where $G$, $G'$ and $G''$ are the sets of $\alpha$, $\beta$, $\omega$-nodes that have not been the cluster head in last epoch respectively.

Each node generates a number in interval [0, 1] randomly. If this random number is less than corresponding threshold the node will become cluster head. Once the cluster head is selected, it broadcasts an advertisement message to all the nodes, a node that receives such message decides on the basis of received signal strength that to which cluster head it will associate for the current round.

### 3.2.4. Data Transmission

In this protocol we use two techniques for data transmission:

a) Single-hop direct transmission

b) Multi-hop transmission through cluster heads

In single hop transmission a node $n$ near the base station will directly send the data to base station if

$$d_{n \text{ to BS}} < \frac{d_0}{k'} \text{ and Residual energy } E_n(r) \geq \overline{E}(r)$$

Here $d_{n \text{ to BS}}$ is distance between $n^{th}$ node and base station; $k'$ is a parameter which is used to control the single hop transmissions; $E_n(r)$ is residual energy of $n^{th}$ node in $r^{th}$ round; $\overline{E}(r)$ is average energy of whole network in $r^{th}$ round.

If the above two conditions do not satisfy simultaneously, the node will send data to cluster head for further processing. Each cluster head create a schedule based on TDMA and nodes send data to their cluster head at
their respective time slots as described in the schedule. We used Dijkstra’s shortest path algorithm to find the shortest route from a cluster head to base station through other cluster heads. For this we used the distance of nodes from each other and base station as weights.

4. SIMULATION RESULTS

For simulation we assume a square network field of size 100m × 100m with 100 sensor nodes deployed uniformly in it. We assume that sink is at the centre of the field. Performance of proposed algorithm compared with LEACH and its popular variant LEACH-C. For the simulation of second proposed algorithm same network field and number of sensors taken as in first but deployment is nonuniform. 20m × 20m area centred at (50, 50) is normal-zone and the area of width 20m surrounding the normal zone is β-zone and β-sensors have been deployed in this zone randomly. Similarly the remaining area of width 20m that surrounds the β-zone is α-zone and α-sensors are deployed in the α-zone remaining ω-sensors are deployed randomly in whole network field. MATLAB is used for the simulation, using different values of α, β, A and B. Simulation parameters are shown in the Table 1.

<table>
<thead>
<tr>
<th>Description</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial energy of normal nodes</td>
<td>$E_0$</td>
<td>0.5 J</td>
</tr>
<tr>
<td>No. of nodes as fraction of total nodes</td>
<td>$A$</td>
<td>0.1, 0.2</td>
</tr>
<tr>
<td>No. of nodes as fraction of total nodes</td>
<td>$B$</td>
<td>0.1, 0.2</td>
</tr>
<tr>
<td>Energy coefficient for nodes</td>
<td>$\alpha$</td>
<td>2</td>
</tr>
<tr>
<td>Energy coefficient for nodes</td>
<td>$\beta$</td>
<td>1, 1.5</td>
</tr>
<tr>
<td>Data aggregation energy</td>
<td>$E_{DA}$</td>
<td>5 nJ/bit/signal</td>
</tr>
<tr>
<td>Electronic circuitry energy</td>
<td>$E_{elec}$</td>
<td>50 nJ/bit</td>
</tr>
<tr>
<td>Free space co-efficient</td>
<td>$\varepsilon_{fs}$</td>
<td>10 pJ/bit/m2</td>
</tr>
<tr>
<td>Multi-path co-efficient</td>
<td>$\varepsilon_{mp}$</td>
<td>0.013 pJ/bit/m4</td>
</tr>
<tr>
<td>Optimal percentage of CHs</td>
<td>$P_{opt}$</td>
<td>0.1</td>
</tr>
<tr>
<td>Initial Energy of nodes</td>
<td>$E_{\alpha}$</td>
<td>$E_0(1+\alpha)$</td>
</tr>
<tr>
<td>Initial Energy of nodes</td>
<td>$E_{\beta}$</td>
<td>$E_0(1+\beta)$</td>
</tr>
<tr>
<td>Data packet size</td>
<td>$P_{pkt}$</td>
<td>30 bytes</td>
</tr>
<tr>
<td>Total no. of nodes</td>
<td>$N$</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 1, 2 and 3 shows the no. of alive nodes in each round of LEACH, LEACH-C and proposed protocol PSOBC for different positions of base station. Simulation results show a considerable improvement in network lifetime.

The above figures 4-6 shows a comparison of HEED, LEACH, SEP and EEHCP for different numbers and energy levels of α, β and ω nodes. In each case EEHCP outperform LEACH and SEP and HEED.

5. CONCLUSION

In this work we proposed a PSO based solution to clustering problem in which a new operator is defined and used it with PSO algorithm to make it work with discrete search space. Simulation results show a considerable increment in Network lifetime as compared to LEACH and LEACH-C. In the second proposed algorithm (EEHCP) a hybrid cluster head selection mechanism applied which uses the heterogeneity in sensor node for an intelligent deployment of nodes in the network and the residual energy of nodes in particular data transmission round to
Energy Efficient Clustering Approach for Data Aggregation and Fusion in Wireless Sensor Networks

Figure 1: Alive Nodes per Round for BS position (50, 0)

Figure 2: Alive Nodes per Round for BS position (0, 50)
Figure 3: Alive Nodes per Round for BS position (200, 200)

Figure 4: No. of Alive Nodes per Round for $A = 0.1, \alpha = 2, \beta = 1, B = 0.1$
Energy Efficient Clustering Approach for Data Aggregation and Fusion in Wireless Sensor Networks

Figure 5: No. of CHs per Round for $A = 0.1$, $\alpha = 2$, $\beta = 1$, $B = 0.1$

Figure 6: No. of Alive Nodes per Round for $A = 0.2$, $\alpha = 2$, $\beta = 1.5$, $B = 0.1$
weight the optimal probability of cluster head selection. The simulation results show that proposed algorithm outperforms SEP, LEACH and HEED and prolong the stability period and Network throughput.

REFERENCES


