Scalable Grid-Based Data Gathering Algorithm for Environmental Monitoring Wireless Sensor Networks

YUVARAJ PADMANABAN, (Student Member, IEEE),
AND MANIMOZHI MUTHUKUMARASAMY

School of Electrical Engineering, Vellore Institute of Technology (VIT), Vellore 632014, India

Corresponding author: Manimozhi Muthukumarasamy (mmanimozhi@vit.ac.in)

ABSTRACT Proper utilization of the available low-power is essential to extend the lifetime of the battery-operated wireless sensor networks (WSNs) for environmental monitoring applications. It is mandatory because the batteries cannot be replaced or recharged after deployment due to impracticality. To utilize the power properly, an appropriate cluster-based data gathering algorithm is needed which reduces the overall power consumption of the network significantly. So, in this paper, a grid-based data gathering algorithm called energy-efficient structured clustering algorithm with relay (EESCA-WR) is proposed. In this algorithm, the grids have a single grid leader (GL) and multiple grid relays (GRs). The count of GRs in a grid is variable based on the geographic location of the grid with respect to the destination sink (DS). By doing this, we ensure that the reduction in power consumption is achieved because of the multi-hop short-distance data communications. Also, the GLs are rotated in the right intervals in hybrid modes to minimize the usage of control messages considerably. A hybrid GL selection policy, a threshold-based GL rotation policy, and the policy of allotting dedicated relay-clusters in every grid make the proposed algorithm unique and better for homogeneous and heterogeneous wireless sensor networks. Performance evaluation of the proposed algorithm is carried out by varying the length of the field, the node-density, the grid-count, and the initial energy. Experimental results show that EESCA-WR is extremely scalable, energy-efficient with a minimum number of control messages, and can be used for large scale WSNs.

INDEX TERMS Clustering algorithm, EESCA, grid-based data gathering, heterogeneous, network lifetime, wireless sensor networks.

I. INTRODUCTION

A. WSN FOR ENVIRONMENTAL MONITORING

Environmental monitoring is essential to forecast and control big disasters in the remote areas of the world. Monitoring sediment transport processes in coastal areas [1], monitoring factory environmental quality [2], examining forest fire [3]–[5], predicting the erection of a volcano [6], [7], and forecasting the snow-melt [8] are some of the best examples of environmental monitoring applications [9]. Various sensors are available to sense the physical parameters like temperature, pressure, and vibrations from the environment. Modern sensors are tiny, handy, and compatible with the wireless communication devices. Generally, wireless communication is preferred in the environmental monitoring applications

where a fixed wired communication structure is not feasible. A sensor node (or simply a node) is a device that is capable of gathering sensory information, performing some processing, and communicating with other nodes. When plenty of sensor nodes are deployed in the above-mentioned large-area monitoring fields to make wireless sensor networks (WSNs) [10], obtaining the individual data from all the sensor nodes is not useful because the data from the same region is highly correlated. Also, the nodes which are deployed far from the destination sink (DS) lose their energies soon due to large communication distances. To avoid the redundant data as well as the long-distance communications, a bunch of data from the same region (grid) can be clubbed into a single data which state the aggregated information about the grid. The aggregated data can be sent to the DS. When direct communication happens between a grid leader (GL) and the DS, the GL alone has to communicate the aggregated data to the DS.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhangbing Zhou.

This methodology can be used for small-area networks. But, when the distance between the grids and DS is more, the data should be routed via other grids using multi-hop communication. When the routing is dynamic, every time the route should be discovered which needs more number of control messages. Instead, the routing can be done by dedicating some nodes as relays permanently. They won't involve in sensing the information from the fields. They just receive the information from some nodes and send the received information to some other nodes based on the algorithm. Thus, for clubbing the nodes properly and transmitting the beneficial data to the distant DS through the appropriate route, a good clustering-based data gathering algorithm is preferred. A well-developed algorithm is useful for improving the life-span of the WSN since the wireless sensor nodes are supplied with tiny batteries with limited energy. The paper aims to propose a scalable energy-efficient grid-based data gathering algorithm which minimizes the control messages substantially.

The rest of the work is organized as follows: Section II elaborates on the related works and our contributions; Section III describes the proposed algorithm in detail; Section IV presents the extensive simulation results obtained, and Section VI concludes the paper.

II. RELATED WORKS

Evidence of extensive research has been found in the literature to make the network energy-efficient. Among those, clustering is a widely accepted technique and it has good scalability also. The cluster head (CH) selection process is carried out using probabilistic and non-probabilistic ways [11], [12] in the past. Low energy adaptive clustering hierarchy (LEACH) [13] is one of the important algorithms which proposed the probabilistic approach for CH selection. Then, modified versions of LEACH [14] are introduced. They followed the same approach with minor updates. After that, hybrid algorithms like [15], [16] considered the residual energy along with the probabilistic approach. However, in the aforementioned approaches, there is a possibility of selecting the nodes with low residual energy as CHs and the cluster count is variable. This leads to the load imbalance among the nodes. To address this problem, the algorithms [17], [18] selected the CH in non-probabilistic ways by considering the parameters such as residual energy, the node proximity, and the distance between the CHs and DS. Even though the load is balanced and the network lifetime is improved in these approaches, the control overhead is more to make the clusters. Also, the clusters which are located far from the DS lose their connectivity soon. To overcome this problem, some algorithms used single-hop communication for intra-cluster communication and multi-hop communication for the inter-cluster communication [19]. In these approaches, the CHs in the intermediate regions act as the routers for the distant clusters. However, it increases the load of the intermediate CHs and the lifetime of the network is reduced. Meanwhile, grid-based data gathering algorithms are developed [19], [20]

to have fixed clusters. In these approaches also, the data is routed via the CHs. Later, some algorithms like scalable energy efficient clustering hierarchy (SEECH) [21] allotted some nodes as relays to route the aggregated data from the CHs to the DS. But, in this approach, the relay node has to receive multiple messages from their own CHs, and from other clusters which results in the reduction of the network lifetime. All the above-mentioned algorithms are proposed for the homogeneous WSNs. On the other hand, the stable election protocol (SEP) [22] is introduced for the two-level heterogeneous WSNS. In this algorithm, the nodes are supplied with different initial energies to check the performance of the network. Later, the distributed energy-efficient clustering (DEEC) algorithm [23] is suggested for multi-level heterogeneous WSNs. The authors of the SEP, DEEC, and other successive algorithms [24] claimed that most of the algorithms developed for homogeneous networks are not suitable for heterogeneous networks.

Motivated by the literature, in this paper, a grid-based data gathering algorithm called energy-efficient structured clustering algorithm with relay (EESCA-WR) is proposed. The major contributions of this research work are as follows.

- A large rectangular field is selected for the investigation. It exhibits the algorithm's capability to be utilized in huge-area wireless sensor networks.
- The algorithm is tested by varying the length of the field from 200 m to 800 m. It indicates that the algorithm is extremely scalable.
- Complete useful data percentage (CUDP) is used for analyzing the lifetime of the network. It shows the accurate performances of the algorithm for various sizes of fields.
- A concept of relay-cluster is introduced for balancing the load among the nodes. It is a novel technique to route the data from the grids to the DS easily.
- The GL is rotated in multiple modes. The role of the GL is retained for the same node until an energy-based threshold is reached in mode 1. It is to avoid the re-clustering process every round. In mode 2, the GL role is rotated based on the energy threshold. This methodology reduces control messages.
- The algorithm is designed to suit both homogeneous and heterogeneous WSNs.

III. PROPOSED EESCA-WR ALGORITHM
A. PRELIMINARIES
1) NETWORK ARRANGEMENT
The proposed network field arrangement is shown in Fig. 1. The nodes are randomly deployed in rectangular fields with dimensions ranging from 100 m × 200 m to 100 m × 800 m. The total number of nodes in the fields, n_t is varied from 200 to 1200. The DS is located in the fixed position of $M_f/2, N_f$ as suggested in [25], where M_f, and N_f are the dimensions of the field. The field is segregated into a number of grids with the dimensions $M_g = X_f/2$, and
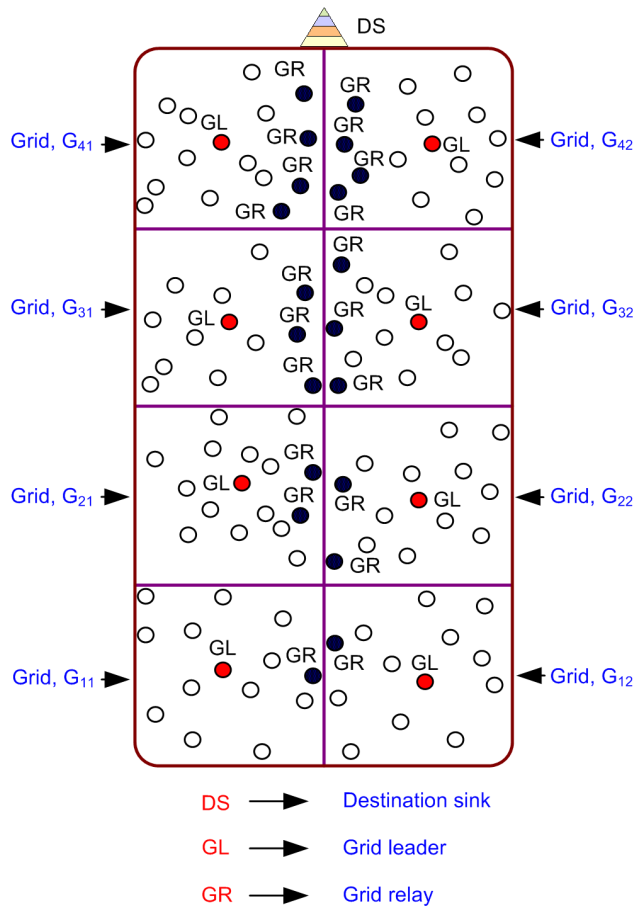
FIGURE 1. Proposed network field arrangement.

$N_g = (N_f * 2)/d_0$, where d_0 is the distance threshold which is calculated by using (3).

2) ROUTING MECHANISM

The data routing mechanism of the proposed algorithm is shown in Fig. 2. The grid normal nodes (GNs) transfer the sensed data to the GL. The GL aggregates the data and passes to a grid relay (GR). The data from the lower layer grids are passed to DS through the GRs in the higher layers. The relay cluster's count in a grid is decided by the position of the grid with respect to the DS. The grids which are far from the DS have only one GR node. Thus, the GR present in a lower grid transmits the aggregated data obtained from the GL to a GR in the next higher layer grid. The higher layer grids have multiple GRs. All the GRs receive one message from either local GL or the lower layer grid's GR and send one message to either higher layer grid's GR or the DS. By doing this, the workload of the nodes is balanced and the relay nodes are allowed to be in sleep mode most of the time. Also, the nodes which are allotted as GRs act as GRs only throughout the operation. This ensures stable connectivity between the grids and the DS.
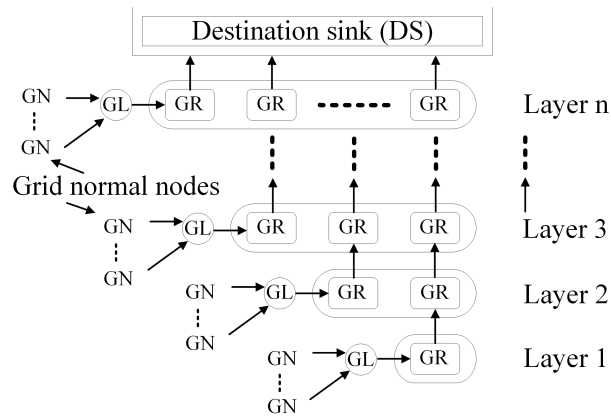

FIGURE 2. An illustration of routing mechanism of EESCA-WR.

3) ASSUMPTIONS

The assumptions made in this work are listed as follows.

- The nodes are stationary after deployment.
- The information is sent to the DS every round.
- The nodes can identify their location coordinates.
- To compare the results with other clustering schemes, it is assumed that each node can transmit its data packet to the DS directly.
- The nodes use power control to adjust the transmit power.

4) NETWORK ENERGY MODEL

The energy of a node is dissipated when the node senses the environmental parameter, transmits the data to other nodes, receives the data from other nodes, and aggregates the received messages from other nodes when acting as a GL. Among these, the sensing takes negligible energy. The energy consumption of a node when transmitting and receiving any data is calculated using the simple first-order radio model which is used in [13].

The energy required for data transmission, E_{tx} is given as:

$$E_{tx}(k, d) = \begin{cases} kE_{elec} + k\epsilon_{fs}d^2 & for \ d \leq d_0 \\ kE_{elec} + k\epsilon_{mp}d^4 & for \ d > d_0 \end{cases} \quad (1)$$

The energy required for data reception, E_{rx} is given as:

$$E_{rx} = kE_{elec} \quad (2)$$

where k is the number of message bits, d is the distance of communication in meters, E_{elec} is the energy required to run transmit, T_x or receive, R_x circuitry, ϵ_{fs} is the energy consumption in the free space, ϵ_{mp} is the energy consumption in multipath, and d_0 is the distance threshold which can be calculated as:

$$d_0 = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}}. \quad (3)$$

5) PROBLEM STATEMENT

The ultimate aim of this research work is to develop a grid-based data gathering algorithm with multi-hop routing which produces an energy-efficient network for gathering the data over a prolonged period. As the sensor nodes are energy constrained, the reduction of the overall energy consumption of the network is essential. In addition to this, the workload should be evenly shared among the nodes to avoid the possibility of the early death of the nodes due to the overburden. Then, the algorithm should achieve good scalability. Finally, a fast response is preferable to reduce the delay in operation.

B. ALGORITHM DETAILS

The lifetime of the network depends on the lives of individual nodes. So, in the proposed algorithm, the load is evenly balanced among the nodes. The lifetime of the network is maximized by

- minimizing the transmission distance for GNs.
- minimizing the number of associated GNs with the GL.
- allotting the responsibility for each GR to receive the data from one node only (either from GL of the same grid or from the GR of the lower layer grid) and to send the data to one node only (either to the GR of the higher layer grid or to the DS).

Various stages of the EESCA-WR are described in the following sections where i refers the count of the round, j refers the count of the node, l refers the virtual centerline, k_{opt} refers the optimum cluster-count which is derived in (5), and $neigh$ refers to the neighbour nodes. The other parameters used are self-explanatory.

1) NODE INITIALIZATION

In the beginning, the nodes are initialized to have preliminary information like their locations and the distance between other nodes and DS. Then, the average communication distance (ACD) is calculated as follows:

$$ACD_j = \frac{\sum_{j=1}^{n_g} d_{j \to neigh}}{n_g} \tag{4}$$

where $d_{j \to neigh}$ is the distance of jth node to the other nodes and n_g is the number of nodes in the grid. The node-level initialization process is given in Algorithm 1.

2) GR SELECTION

After the node initialization process, the m nodes which are closer to the virtual vertical centerline of the field elect themselves as GRs. When p layers are there in the field, the pth layer grid has p number of GRs. In the bottom layers, only one GL and one GR present. So, they associate with each other. In higher layers, multiple GRs present. So, the GLs select the GRs which have the shortest communication distances to DS. Similarly, the GRs from the bottom layers associate with the GRs from the higher layers which have closer communication

---

Algorithm 1 Node Level Initialization Process

Require: M_f, N_f
Ensure: $ACD_j,\ d_{j \to l},\ and\ ID_j$
1: Identify X_j, Y_j
2: Identify ID_j using M_f, N_f
3: Calculate $d_{j \to l}$ using M_f, N_f
4: Broadcast $self\_intro\_msg_j$ within ID_j
5: Receive $self\_intro\_msg_{neigh}$
6: Calculate $d_{j \to neigh}$
7: Calculate ACD_j

Algorithm 2 Node Level GR Selection Process

Require: $M_f, N_f,\ and\ k_{opt}$
Ensure: $status_j^i$
1: **if** $i == 1$ **then**
2:     **if** $ID_j > k_{opt}/2$ **then**
3:         $m = ID_j - k_{opt}/2$
4:     **else**
5:         $m = ID_j$
6:     **end if**
7:     Compare $d_{j \to l}$ with $d_{neigh \to l}$
8:     **if** $d_{j \to l} \in \min_m(d_{neigh \to l})$ **then**
9:         Set $status_j^i = GR$
10:    **end if**
11: **else**
12:     **if** $i > 1$ **then**
13:         **if** $status_j^{i-1} == GR$ **then**
14:             Set $status_j^i = GR$
15:         **end if**
16:     **end if**
17: **end if**

---

distance between them. The detailed algorithm for GR selection is shown in Algorithm 2.

3) GL SELECTION

The GL is selected in hybrid modes based on the centrality and the nodes' residual energies as mentioned in [26], [27]. The node-level GL selection process is depicted in Algorithm 3. In mode 1, the node which holds the least ACD in the grid acts as the GL for the initial rounds. The node which has the least ACD is in the center of the grid. So, the communication distance of the GNs gets reduced. In mode 2, the node which has the highest residual energy gets the leadership role every round. This is done to ensure the longevity of the network lifetime.

4) GRID FORMATION

The nodes in a grid take the roles of the GR, GL, and GN based on the algorithm. In every grid, some nodes form the relay-cluster and act as the GRs, one node acts as the GL, and the remaining nodes act as the GNs. After the GRs and GLs are selected and associated with appropriate nodes, the GL does its operation by broadcasting a message,

Algorithm 3 Node Level GL Selection Process

Require: $M_f, N_f, k_{opt}, GLmode_j^i$
Ensure: $status_j^i$
1: **if** $GLmode_j^i == 1$ **then**
2:     Calculate ACD_j
3:     **if** $ACD_j < ACD_{neigh}$ **then**
4:         Set $status_j^i = GL$
5:         **if** $CTNR_j^i \geq 0.4$ **then**
6:             Set $status_j^i = GN$
7:         **else**
8:             Set $status_j^i = GL$
9:         **end if**
10:     **else**
11:         Set $status_j^i = GN$
12:     **end if**
13: **end if**
14: **if** $GLmode_j^i == 2$ **then**
15:     **if** $E_j^i > E_{neigh}^i$ **then**
16:         Set $status_j^i = GL$
17:     **else**
18:         Set $status_j^i = GN$
19:     **end if**
20: **end if**

Algorithm 4 Node Level Grid Formation Process

Require: M_f, N_f, k_{opt}
Ensure: $status_j^i$
1: **if** $status_j^i = GL$ **then**
2:     Broadcast GL_{msg} with in ID_j
3: **end if**
4: **if** $status_j^i = GN$ **then**
5:     Join with the GL in the grid
6: **end if**
7: **if** $status_j^i = GR$ **then**
8:     Join with the GL in the grid or join with the closest possible GR in the higher layer grid
9: **end if**

GL-MSG including its ID and a secret code to the other nodes in the grid. The GNs send join message, $join_{msg}$ to the respective GLs. Thus the grid is formed with minimum control messages. The operation is described in Algorithm 4. After that, a time division multiple access (TDMA) schedule is created for all the nodes for the data transmission. This avoids the data traffic in the network.

C. METHODOLOGIES USED FOR IMPROVING NETWORK LIFETIME

The main motive of this work is to reduce energy consumption among the nodes to enhance the network's lifetime. Literature results show that the grid-count and the communication distance between the nodes affect the energy consumption of the total network. So, an effort has been taken for getting optimum grid-count and reduced communication distance between the nodes.

1) OPTIMIZATION OF GRID-COUNT

In the proposed algorithm, the grid-count is decided based on the dimensions of the grid. The relay-cluster of every grid should be within the reach of the relay-cluster of higher layer grids. It is done mainly to reduce the communication distance between the GRs. In this work, N_g is fixed as $N_f/2$ and M_g is scalable to $(M_f/d_0/2)$m approximately. So, the optimum grid-count, k_{opt} is derived as:

$$k_{opt} = 2 * \frac{M_f}{d_0/2} \tag{5}$$

2) MINIMIZATION OF COMMUNICATION DISTANCE

- When grid size is restricted to $M_f/d_0/2, N_f/2$, obviously the maximum distance between a GN and the GL is limited to:

$$d_{j \to GL} = \frac{\frac{M_f}{d_0/2}}{\sin\left(\tan^{-1}\left(\frac{\frac{M_f}{d_0/2}}{N_f/2}\right)\right)} \tag{6}$$

- Since the GL is in the middle of the grid in the initial stages, the $d_{GL \to GR}$ in the grid is much lesser than the $d_{j \to GL}$. However, in the worst case, the maximum distance between the GL and the GR is limited to (6).
- The distance between a GR to a higher layer GR is limited to:

$$d_{GR \to higher\_layer\_GR} = \frac{\frac{N_f}{4}}{\sin\left(\tan^{-1}\left(\frac{\frac{N_f}{4}}{2M_g}\right)\right)} \tag{7}$$

D. ENERGY CONSUMPTION

The time-lines of the GNs, GLs, and GRs are shown in Fig. 3. The GNs and GRs are active for the required timings and at sleep state for the remaining duration. The GL has to be active throughout the operation since it receives the data regularly from GNs, aggregates the data, and sends the aggregated data to the GR. The overall energy consumption per round is analyzed in the following sections.

1) ENERGY CONSUMPTION IN GNs

The role of a GN is to sense the environmental parameter and send that to the GL. Thus, the energy consumption of a GN is given as:

$$E_{GN} = kE_{elec} + k\epsilon_{fs}d_{to\_GL}^2 \tag{8}$$

2) ENERGY CONSUMPTION IN GLs

The GL does the collection of all the information from the GNs, aggregates the data to make a piece of single information, and passes it to the GR for further transmission.
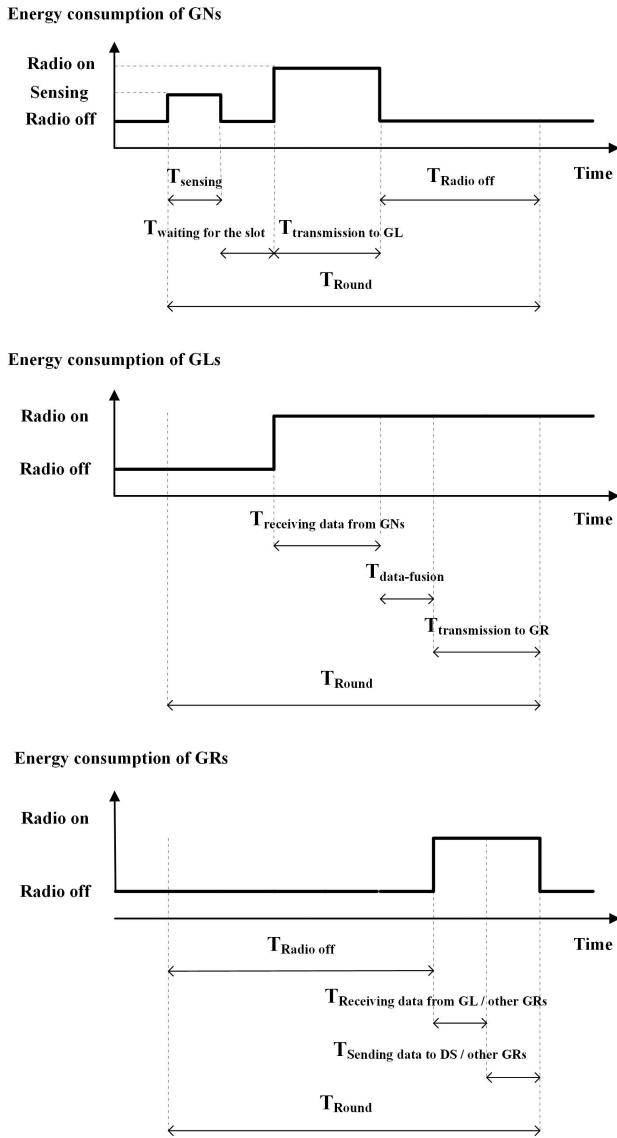
Energy consumption of GNs

Energy consumption of GLs

Energy consumption of GRs

FIGURE 3. Time lines of GNs, GLs, and GRs.

Thus, the energy consumption of a GL is given as:

$$E_{GL} = kE_{elec}\left(n_g - 1\right) + kE_{da}\left(n_g\right) + kE_{elec} + k\epsilon_{fs}d_{to\_GR}{}^2$$
(9)

where n_g is the number of the nodes in the grid.

3) ENERGY CONSUMPTION IN GRs

A GR in the relay-cluster collects the information either from the GL or from a lower layer GR. Thus, all the GRs receive and send only one data per round. So, the energy consumption of a GR in the grid closest to the DS is given as:

$$E_{GR} = 2kE_{elec} + k\epsilon_{fs}d_{to\_DS}{}^2$$
(10)

The energy consumption of a GR in other layer grids is given as:

$$E_{GR} = 2kE_{elec} + k\epsilon_{fs}d_{to\_Higher\_layer\_GR}{}^2$$
(11)

4) ENERGY CONSUMPTION IN A GRID

In a grid, the GNs sense the physical parameters needed from the environment and send the detail to the GL. The GL receives all the details from GLs and aggregates the data into a single data and transmits it to the nearest GR. The GRs in the relay-cluster collect the information from the GL and a lower layer GR if any. So, the energy consumption of higher layer grids is given as:

$$\begin{aligned} E_{grid} = {} & kE_{elec} + k\epsilon_{fs}d_{to\_GL}{}^2 \\ & + kE_{elec}\left(n_g - 1\right) + kE_{da}\left(n_g\right) \\ & + E_{elec} + k\epsilon_{fs}d_{to\_GR}{}^2 \\ & + GR_{tot}(2kE_{elec} + k\epsilon_{fs}d_{to\_DS}{}^2) \end{aligned}$$
(12)

where GR_{tot} is the number of GRs in a grid.
The energy consumption of other layer grids is given as:

$$\begin{aligned} E_{grid} = {} & kE_{elec} + k\epsilon_{fs}d_{to\_GL}{}^2 \\ & + kE_{elec}\left(n_g - 1\right) + kE_{da}\left(n_g\right) \\ & + kE_{elec} + k\epsilon_{fs}d_{to\_GR}{}^2 \\ & + GR_{tot}(2kE_{elec} + k\epsilon_{fs}d_{to\_Higher\_layer\_GR}{}^2) \end{aligned}$$
(13)

Even though the relays in the higher layer grids are more compared to the lower layer grids, every relay node receives only one message every round. Hence the energy consumption of all the grids is almost the same.

5) OVERALL ENERGY CONSUMPTION OF THE NETWORK

The energy consumption of the network depends on the number of GRs and k_{opt}. So, the energy consumption of a total network in a round is given as:

$$\begin{aligned} E_{round} = {} & k_{opt}\{kE_{elec} + k\epsilon_{fs}d_{to\_GL}{}^2 \\ & + kE_{elec}\left(n_g - 1\right) + kE_{da}\left(n_g\right) \\ & + kE_{elec} + k\epsilon_{fs}d_{to\_GR}{}^2 \\ & + GR_{totH}(2kE_{elec} + k\epsilon_{fs}d_{to\_DS}{}^2) \\ & + GR_{totL}(2kE_{elec} + k\epsilon_{fs}d_{to\_Higher\_layer\_GR}{}^2)\} \end{aligned}$$
(14)

where GR_{totH} is the number of GRs in higher layer grids and GR_{totL} is the number of GRs in higher layer grids.

IV. SIMULATION RESULTS AND ANALYSIS

In this section, the performance of EESCA-WR is compared with LEACH [13], EESCA [26], and SEECH [21] algorithms. LEACH uses a probabilistic threshold to select the CHs. The CHs directly send the data to the DS. Since the CHs are selected randomly, the algorithm does not consider the residual energy of the nodes which results in frequent data packet losses. Also, the CHs directly send the data to DS. Hence, this algorithm is not suitable for large monitoring fields. EESCA selects the CH in hybrid modes. Mode 1 uses the node centrality and mode 2 uses the residual energy as parameters. The data from the lower layer clusters is routed via the CHs of the higher layer clusters. This affects the load balancing of the network when the algorithm is used in

large fields. SEECH uses the residual energy and node degree as the main parameters for the CH selection. As mentioned in section 2, SEECH isolates the relay role from the CH and appoints some nodes for taking that role. But, in this approach, the relay nodes have to receive messages from multiple clusters. It results in a reduction of the network lifetime. Both the LEACH and SEECH rotate the CH every round whereas the EESCA rotates the CH only after the CH loses its 40% of the initial energy. It reduces the control messages used for re-clustering.

A. SIMULATION PARAMETERS

Simulations are carried out in MATLAB R2018a. The simulation parameters used are as follows: Packet size, $k = 4000$ *bits*, Transmit energy, $E_{tx} = 50$ *nJ/bit*, Threshold distance, $d_0 = 87$ *m*, Multipath energy, $\epsilon_{mp} = 0.0013$ *pJ/bit/m^4*, Free space energy, $\epsilon_{fs} = 10$ *J/bit/m^2*, and Data aggregation energy, $E_{da} = 5$ *nJ/bit/messsage*. The DS is fixed at the location $(M_f/2, N_f)$. The width of the field, N_f is fixed as 100 *m* and the length of the field, M_f is varied from 200 *m* to 800 *m* whereas the number of nodes, n_t is varied from 400 to 800 for checking the scalability. The initial energy supplied to the node, E_0 is ranged from 0.25 *J* to 1 *J* to evaluate the impact of the initial energy on the lifetime of the network. In LEACH, the communication occurs between the GL and the DS directly which is highly improbable when the length of the field is more than the communication range of the nodes. In this work, the communication range of the nodes is assumed as unlimited for comparison. In SEECH, the internal parameters used are as follows: needed CH candidates $K_{CHC} = 8$, needed CHs, $K_{CH} = 3$, needed relays, $K_R = 10$, and specific radius, $RNG = 55$ *m* as prescribed in scene 1 of SEECH. To evaluate the network lifetime, First node die (FND) and Last node die (LND) are the common parameters used in most of the algorithms. In addition to those, the parameters quarter node die (QND), half node die (HND), and 90% node die (90%ND) are used in this work. When very few nodes in the same region are alive for many rounds, the LND is more. But, the data obtained from those nodes cannot give the overall information about the field. So LND cannot be a suitable indicator of the effectiveness of the algorithm. So, In addition to the above-mentioned parameters, a new parameter CUDP is introduced to evaluate the performance of the algorithms. The CUDP is described as follows:

$$CUDP = \frac{FND}{LND} \times 100 \qquad (15)$$

By using the CUDP, the effectiveness of the algorithm can be tested successfully. When the CUDP is more, the algorithm can be considered as energy-efficient.

B. SIMULATION RESULTS

1) IMPACT OF THE LENGTH OF THE FIELD

When the width of the field is maintained constant, and the length of the field is increased, the nodes have to send the data
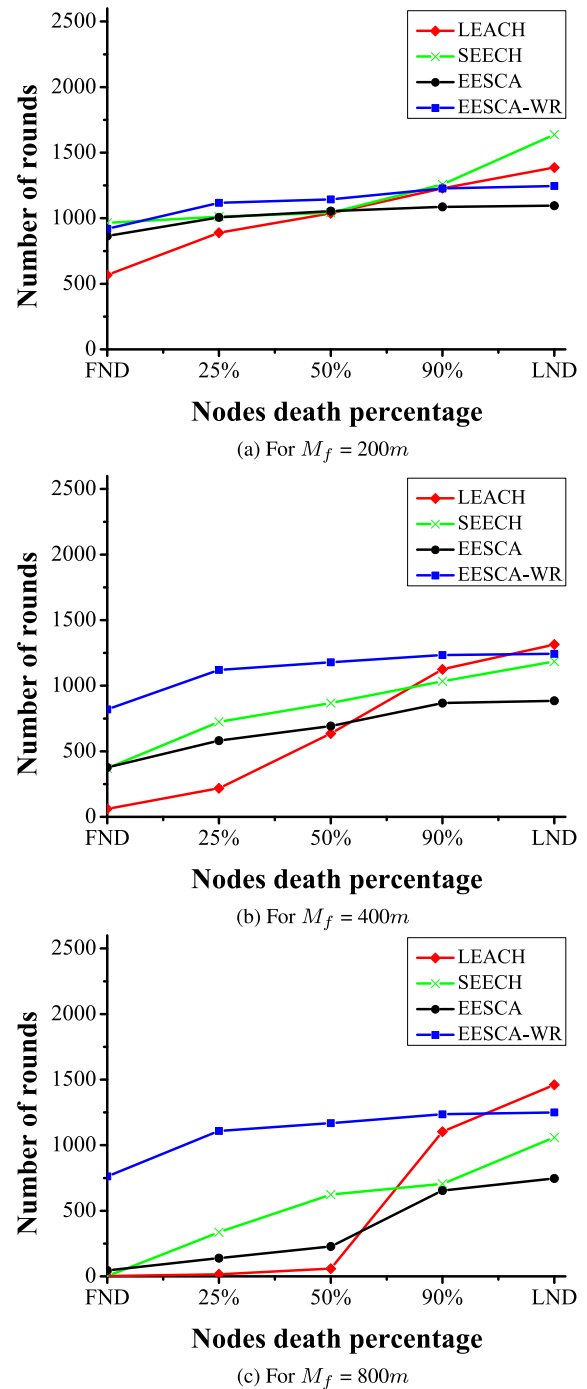

(a) For $M_f = 200m$

(b) For $M_f = 400m$

(c) For $M_f = 800m$

FIGURE 4. Impact of length of the field.

to the longer distances. This impacts on the network lifetime. Fig. 4 shows the simulation results to analyze the impact of the length of the field. The simulations are carried out by keeping the following simulation parameters: n_t is 200, 400, and 1200, N_f is 200 m, 400 m, and 800 m, and $E_0 = 0.5$ *J*. In EESCA-WR, k_{opt} is set as 10, 20, and 40 respectively. From the results, it can be observed that EESCA-WR is doing better than other algorithms. EESCA-WR achieves approximately 1.8, 1.3, and 0 times more CUDP for

Scene 1, 14.2, 2.1, and 1.6 times more CUDP for scene 2, and 892, 646, and 10.1 times more CUDP for scene 3, compared to LEACH, SEECH, and EESCA. The FND is almost equal in SEECH and EESCA-WR in scene 2. Similarly, EESCA gets better CUDP in Scene 1. These results indicate that SEECH and EESCA are doing better when the field is small. Also, in scene 1, it is observed that LEACH, SEECH, and EESCA perform reasonably well. But, in scenes 2 and 3, when the area of the field is increased, only EESCA-WR works better compared to other algorithms.

2) IMPACT OF THE NODE DENSITY

When n_t is increased in the same area of the field, the GLs have to receive more messages from the GNs within the grid. So, it reduces the E_{res} of the GLs quickly. Also, the rotation of the GLs takes place frequently and more control messages are needed in mode 2. To evaluate the impact of node density on each approach, the parameter n_t is varied from 400 to 800. Fig. 5 shows the results of each approach for the simulation setting: n_t is 400, 600, and 800 respectively, $M_f = 400m$ and $E_0 = 0.5 J$. In EESCA-WR, k_{opt} is set as 20. The results show that FND is significantly more compared to other algorithms. EESCA-WR achieves approximately 14.2, 2.2, and 1.3 times more CUDP in all the scenes, compared to LEACH, SEECH, and EESCA.

3) IMPACT OF THE INITIAL ENERGY

The impact of the initial energy is tested by providing E_0 of 0.25 J, 0.5 J, and 1.0 J for the nodes by keeping other parameters constant. The performance of the algorithms is shown in Fig. 6. The simulation parameters set are as follows: n_t is 400, M_f is 400 m and $E_0 = 0.25 J, 0.5 J, 1 J$. In EESCA-WR, k_{opt} is set as 20. Results show that EESCA-WR achieves a better network lifetime. On average, EESCA-WR achieves approximately 14.2, 2.1, and 1.6 times more rounds compared to LEACH, SEECH, and EESCA. But, when E_0 is varied, the performance is not affected because the network lifetime is linearly proportional to the E_0.

4) IMPACT OF THE GRID-COUNT

When k_{opt} is more, the intra-grid distance for communications is reduced. But, in turn, the control overhead is more. On the other hand, When k_{opt} is less, the GLs have to receive more information from normal nodes and the inter-grid distance is more. To compute the impact of k_{opt} on EESCA-WR, the sensor field is partitioned into 12, 16, and 20 grids. In this comparison, $k_{opt} = 20$ is more appropriate for EESCA-WR for the following simulation parameters: n_t is 400, M_f is 400 m and $E_0 = 0.5 J$. In EESCA-WR, k_{opt} is set as 12, 16, and 20. Fig. 7 shows the simulations for this comparison. The expression for deriving the optimum grid-count is given in (5). According to that, having a shorter communication distance is a better way to achieve less energy consumption. In EESCA-WR, the GNs only need to transmit the data locally to their GL located at a short distance. The energy
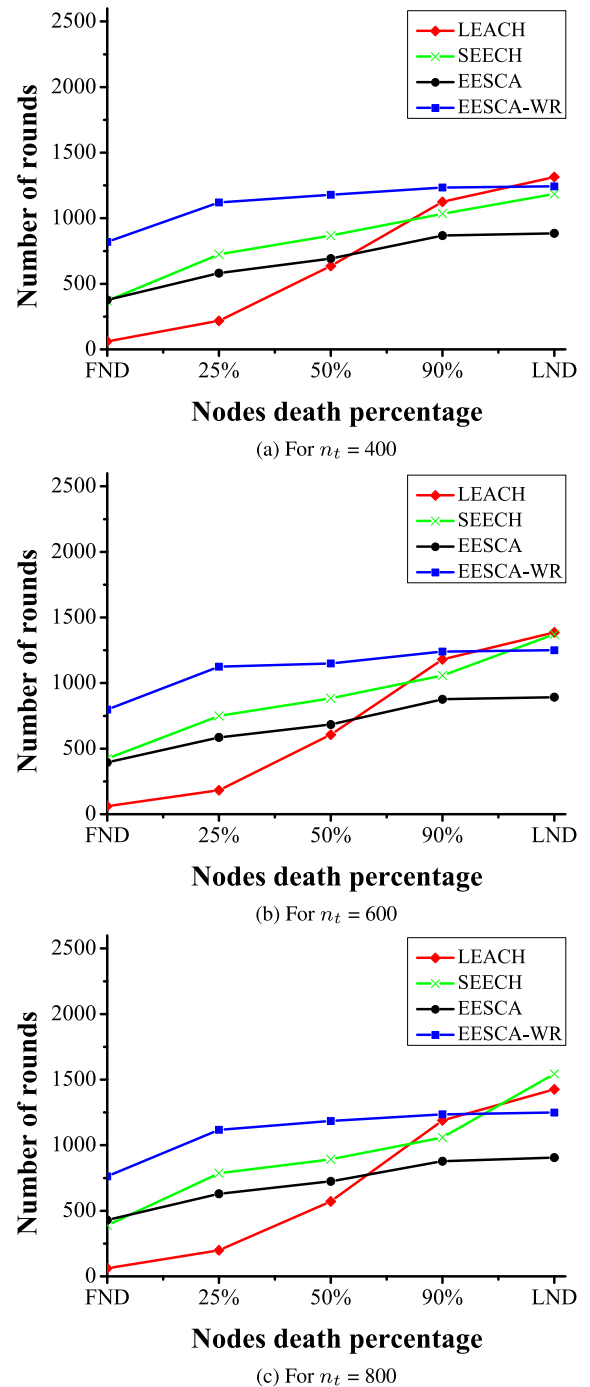

(a) For $n_t = 400$

(b) For $n_t = 600$

(c) For $n_t = 800$

**FIGURE 5.** Impact of node density.

consumed by them is thus reduced and the distance between the GL and the GR is also less.
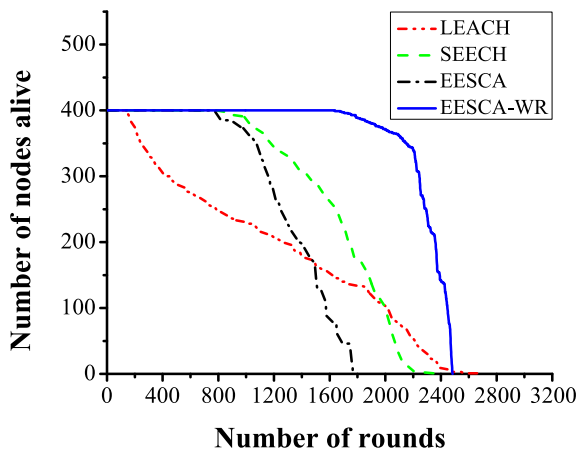
5) AVERAGE RESIDUAL ENERGY

The average residual energy after 500 rounds for the algorithms is given in Fig. 8. The simulation parameters used are: n_t is 400, M_f is 400 m, and $E_0 = 0.5 J$. In EESCA-WR, k_{opt} is set as 20. From the results, it can be observed that in EESCA-WR, the consumption of energy is evenly distributed
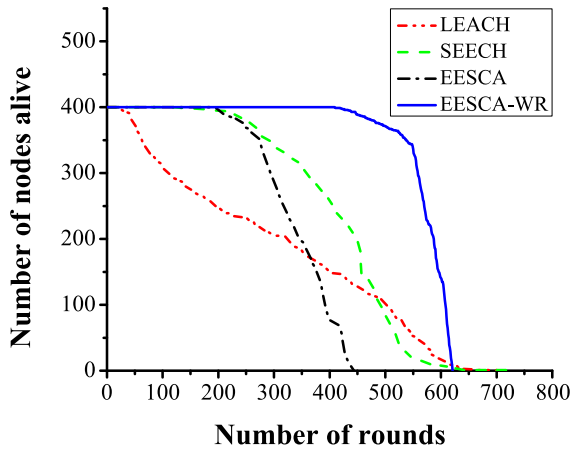
(a) For $E_0 = 0.25$ J
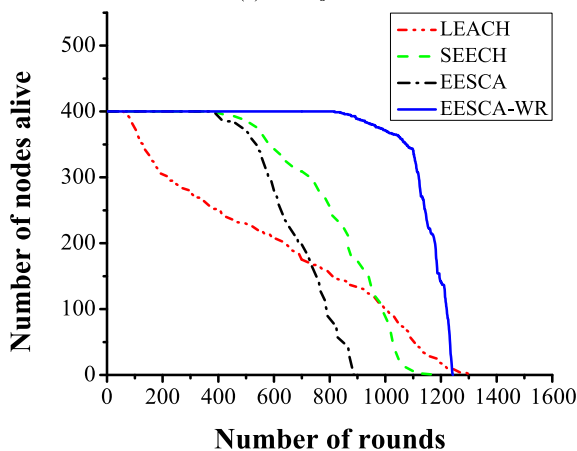
(b) For $E_0 = 0.5$ J

(c) For $E_0 = 1$ J

FIGURE 6. Impact of initial energy.

FIGURE 7. Impact of grid-count.

FIGURE 8. Residual energies after 500 rounds.

FIGURE 9. FND, QND, HND, 90%ND, and LND of EESCA-WR in the presence of heterogeneity for m=0:0.5, and a=0, 3.

among the nodes compared to other algorithms. Similarly, the average energy consumption is very less compared to LEACH, SEECH, and EESCA algorithms.

6) HETEROGENEITY TEST

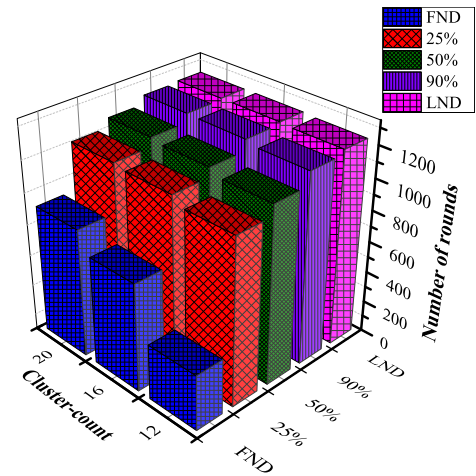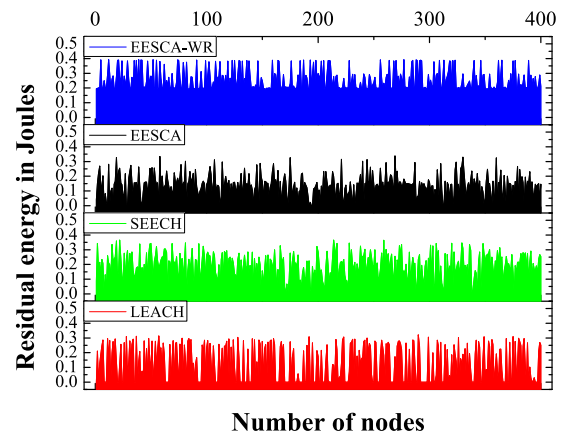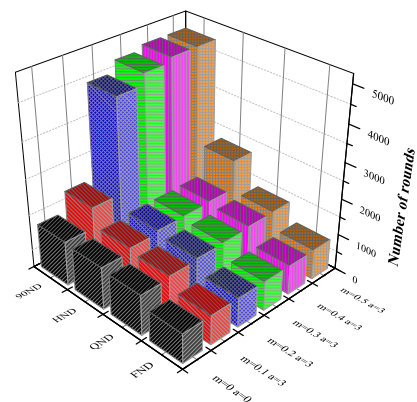The nodes are equipped with different initial energies in heterogeneous WSNs. To test the heterogeneity, the nodes are segregated into two categories. The normal nodes have the initial energy of E_0. The m number of advanced nodes have a times more energy compared to the normal nodes. So, there are mN advanced nodes with the initial energy of $E_0(1 + a)$. The simulation results show that the QND, HND, 90ND are
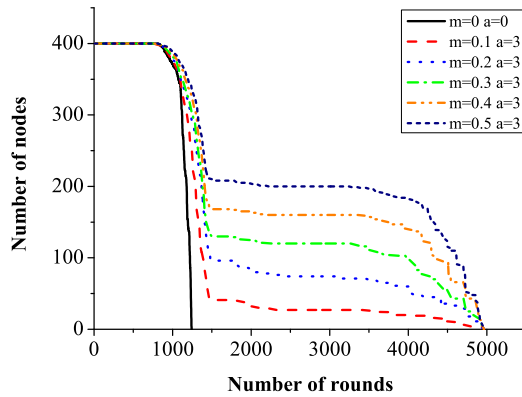
FIGURE 10. Network lifetime of EESCA-WR in the presence of heterogeneity for m=0:0.5, and a=0, 3.

improved when the heterogeneity is introduced. Fig. 9 and Fig. 10 show the impact of heterogeneity in network lifetime.

V. CONCLUSION
The limited energy provided with the sensor nodes should be carefully utilized to prolong the lifetime of the WSNs used for environmental monitoring applications. To achieve a good network lifetime by reducing the overall energy consumption, in this work, an energy-efficient structured clustering algorithm with relay (EESCA-WR) is proposed. The proposed algorithm incorporates the following aspects:
- An optimum grid-count is kept to reduce the communication distance between the nodes.
- All the nodes have an equal chance to serve as GL to ensure load balancing.
- A fixed relay-cluster is maintained to establish the constant route path in the network.

This arrangement reduces the control messages considerably and ensures fast operation. The conventional algorithm LEACH doesn't use the relay concept and performs better for smaller fields. EESCA uses fixed clusters and performs well for the small fields. On the other hand, SEECH uses the relays for routing the data from CH to DS but it overloads the relays and does reasonably well for semi-large sized fields. The proposed algorithm EESCA-WR uses the relays in a better way and it is suitable for both small and large fields because of its characteristics. Simulation results show that EESCA-WR performs well compared to LEACH, SEECH, and EESCA in terms of four parameters, network lifetime, scalability, average energy consumption, and fast response. Moreover, the EESCA-WR is tested with heterogeneous nodes and performs well. Thus, the proposed algorithm is proved to be suitable for large-scale environmental monitoring applications. This work can be extended for multilevel heterogeneous networks also and can be tested using various network simulators in the future.

ACKNOWLEDGMENT
The authors would like to thank Dr. K. Venkata Lakshmi Narayana (Senior Member, IEEE) for extending the support and encouragement to carry out this work.

REFERENCES
[1] A. J. Watt, M. R. Phillips, C. E.-A. Campbell, I. Wells, and S. Hole, "Wireless sensor networks for monitoring underwater sediment transport," *Sci. Total Environ.*, vol. 667, pp. 160–165, Jun. 2019.
[2] Q. Han, P. Liu, H. Zhang, and Z. Cai, "A wireless sensor network for monitoring environmental quality in the manufacturing industry," *IEEE Access*, vol. 7, pp. 78108–78119, 2019.
[3] A. A. A. Alkhatib, "A review on forest fire detection techniques," *Int. J. Distrib. Sensor Netw.*, vol. 10, no. 3, Mar. 2014, Art. no. 597368.
[4] P. Kułakowski, E. Calle, and J. L. Marzo, "Performance study of wireless sensor and actuator networks in forest fire scenarios," *Int. J. Commun. Syst.*, vol. 26, no. 4, pp. 515–529, Apr. 2013.
[5] Y. E. Aslan, I. Korpeoglu, and Ö. Ulusoy, "A framework for use of wireless sensor networks in forest fire detection and monitoring," *Comput., Environ. Urban Syst.*, vol. 36, no. 6, pp. 614–625, Nov. 2012.
[6] R. Lara, D. Benitez, A. Caamano, M. Zennaro, and J. L. Rojo-Alvarez, "On real-time performance evaluation of volcano-monitoring systems with wireless sensor networks," *IEEE Sensors J.*, vol. 15, no. 6, pp. 3514–3523, Jun. 2015.
[7] W.-Z. Song, R. Huang, M. Xu, B. Shirazi, and R. LaHusen, "Design and deployment of sensor network for real-time high-fidelity volcano monitoring," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 11, pp. 1658–1674, Nov. 2010.
[8] S. Malek, F. Avanzi, K. Brun-Laguna, T. Maurer, C. Oroza, P. Hartsough, T. Watteyne, and S. Glaser, "Real-time alpine measurement system using wireless sensor networks," *Sensors*, vol. 17, no. 11, p. 2583, 2017.
[9] Z. Dong, S. Meyland, and M. Karaomeroglu, "A case study of an autonomous wireless sensor network system for environmental data collection," *Environ. Prog. Sustain. Energy*, vol. 37, no. 1, pp. 180–188, Jan. 2018.
[10] M. Farsi, M. A. Elhosseini, M. Badawy, H. A. Ali, and H. Z. Eldin, "Deployment techniques in wireless sensor networks, coverage and connectivity: A survey," *IEEE Access*, vol. 7, pp. 28940–28954, 2019.
[11] B. Mamalis, D. Gavalas, C. Konstantopoulos, and G. Pantziou, "Clustering in wireless sensor networks," in *RFID and Sensor Networks: Architectures, Protocols, Security and Integrations*, Y. Zhang, L. T. Yang, and J. Chen, Eds., 2009, pp. 324–353.
[12] M. M. Afsar and M.-H. Tayarani-N, "Clustering in sensor networks: A literature survey," *J. Netw. Comput. Appl.*, vol. 46, pp. 198–226, Nov. 2014.
[13] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.
[14] P. Yuvaraj, K. Vikram, and K. V. L. Narayana, "A review on state of art variants of leach protocol for wireless sensor networks," *Sensors Transducers*, vol. 186, no. 3, pp. 25–32, 2015.
[15] O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 366–379, Oct. 2004.
[16] M. Ye, C. Li, G. Chen, and J. Wu, "EECS: An energy efficient clustering scheme in wireless sensor networks," in *Proc. 24th IEEE Int. Perform., Comput., Commun. Conf. (PCCC)*, Apr. 2005, pp. 535–540.
[17] P. Ding, J. Holliday, and A. Celik, "Distributed energy-efficient hierarchical clustering for wireless sensor networks," in *Proc. Int. Conf. Distrib. Comput. Sensor Syst.* Berlin, Germany: Springer, 2005, pp. 322–339.
[18] M. Yu, K. Leung, and A. Malvankar, "A dynamic clustering and energy efficient routing technique for sensor networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3069–3079, Aug. 2007.
[19] Y.-K. Chiang, N.-C. Wang, and C.-H. Hsieh, "A cycle-based data aggregation scheme for grid-based wireless sensor networks," *Sensors*, vol. 14, no. 5, pp. 8447–8464, 2014.
[20] J. Qiao and X. Zhang, "Compressive data gathering based on even clustering for wireless sensor networks," *IEEE Access*, vol. 6, pp. 24391–24410, 2018.
[21] M. Tarhani, Y. S. Kavian, and S. Siavoshi, "SEECH: Scalable energy efficient clustering hierarchy protocol in wireless sensor networks," *IEEE Sensors J.*, vol. 14, no. 11, pp. 3944–3954, Nov. 2014.
[22] G. Smaragdakis, I. Matta, and A. Bestavros, "Sep: A stable election protocol for clustered heterogeneous wireless sensor networks," Dept. Comput. Sci., Boston Univ., Boston, MA, USA, Tech. Rep., 2004.
[23] L. Qing, Q. Zhu, and M. Wang, "Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks," *Comput. Commun.*, vol. 29, no. 12, pp. 2230–2237, Aug. 2006.

[24] S. Verma, N. Sood, and A. K. Sharma, "A novelistic approach for energy efficient routing using single and multiple data sinks in heterogeneous wireless sensor network," *Peer-to-Peer Netw. Appl.*, vol. 12, no. 5, pp. 1110–1136, Sep. 2019.

[25] P. Yuvaraj and M. Manimozhi, "Impact of base station location in clustering schemes of WSN," in *Proc. Int. Conf. Microelectronic Devices, Circuits Syst. (ICMDCS)*, Aug. 2017, pp. 1–5.

[26] Y. Padmanaban and M. Muthukumarasamy, "Energy-efficient clustering algorithm for structured wireless sensor networks," *IET Netw.*, vol. 7, no. 4, pp. 265–272, Jul. 2018.

[27] Y. P. and K. V. L. Narayana, "EESCA: Energy efficient structured clustering algorithm for wireless sensor networks," in *Proc. Int. Conf. Comput., Analytics Secur. Trends (CAST)*, Dec. 2016, pp. 523–527.

MANIMOZHI MUTHUKUMARASAMY received the B.E. degree in electronics and instrumentation engineering and the M.E. degree in process control and instrumentation from Annamalai University and the Ph.D. degree from the Vellore Institute of Technology (VIT), Vellore. She is currently an Associate Professor with the Department of Control and Automation, VIT. Her research interests include the areas of process control and instrumentation, sensors and signal conditioning, and state estimation.

● ● ●

YUVARAJ PADMANABAN (Student Member, IEEE) received the Diploma degree in electrical and electronics engineering from the T. P. E. V. R. Government Polytechnic, the B.E. degree in electronics and instrumentation engineering from M. K. University, and the M.E. degree in instrumentation engineering from Anna University. He is currently pursuing the Ph.D. degree with the Vellore Institute of Technology (VIT), Vellore. His research interest includes the development of energy-efficient wireless sensor networks for environmental monitoring.