International Journal of Innovative Computing, Information and Control
Volume **16**, Number **2**, April **2020**
 ISSN 1349-4198
pp. 681–700

A NODE DENSITY-BASED APPROACH FOR ENERGY-EFFICIENT DATA GATHERING PROTOCOL IN WIRELESS SENSOR NETWORK ENVIRONMENTS

Waskitho Wibisono, Tohari Ahmad, Royyana Muslim Ijtihadie and Kharisma Monika Dian Pertiwi

Department of Informatics
Institut Teknologi Sepuluh Nopember
Kampus ITS, Keputih Sukolilo, Surabaya 60111, Indonesia
{ waswib; tohari; roy }@if.its.ac.id; kharismamonika@gmail.com

Received July 2019; revised November 2019

Abstract. *Network lifetime is one of the main issues of wireless sensor network (WS-N) development since sensor nodes are usually powered by a battery with limited energy. This characteristic has highlighted the importance of energy-efficient protocol to minimize energy consumption during data-gathering operations in WSN environments. The cluster-based protocol is an effective way to reduce long-distance transmissions and preserve energy by partitioning WSN into some clusters. In this approach, only selected cluster head (CH) nodes will perform data gathering from the cluster node members and transmit the aggregated data to the base station (BS). In this paper, a new approach to build clusters for WSN based on sensor nodes density is proposed. This approach includes the mechanisms to estimate the optimal number of clusters, to select and rotate CH nodes and the algorithm to perform data communications among sensor nodes, CH nodes, and the base station. The proposed approach shows better performance in terms of network lifetime, energy consumption in comparison to the other two well-known WSN communication protocols in data gathering scenarios.*
Keywords: Energy-efficient data-gathering operations, Network lifetime, Cluster-based protocol

1. **Introduction.** The advances of micro-sensor technology have enabled us to develop tiny and low-power sensor platforms with wireless capability known as sensor nodes used in various applications scenarios. In general, WSNs are composed of numerous sensor nodes distributed in the physical environment, equipped with a low power wireless communication system for sensing data from the environment and transmitting the gathered data to the base station (BS). The deployed sensor nodes are usually powered by a battery that highlights network lifetime issues of sensor nodes due to energy constraints [1] operation to prolong network lifetime [2]. In most of the cases, the WSN need to be developed in ad-hoc environments where the sensor nodes form a self-organized ad-hoc network to sense and forward the sensed data to the BS using the ad-hoc routing approaches [3, 4].

Sensor nodes in WSN consume a significant portion of their energy resource for data transmission using wireless communications in data gathering operations [1]. Accordingly, an energy-efficient routing technique for the WSN environment is highly necessary to maintain the network lifetime in data gathering processes. The network lifetime is the number of rounds (can be expressed as time) until the first sensor node is drained (FND) [2] or the number of rounds when a specific number or percentage of sensor nodes die [3]. Due to the limited energy resource, building energy-efficient but reliable routing protocol

DOI: 10.24507/ijicic.16.02.681

for WSN environment is challenging, especially, when a large number of sensor nodes are required to deal with wide ranges of sensor monitoring fields, difficulty in replacing node battery as well as possible damages during deployment or operation [5, 6].

There exist several approaches to develop an energy-efficient routing protocol to support data gathering operations in WSN environments. One prominent example is to build clusters of sensor nodes to form a hierarchical-based communication scheme [7]. The cluster-based approaches divide the WSN into some clusters where each of them is coordinated by a cluster head (CH) node that is responsible for gathering and aggregating data from its cluster members [1, 8]. The CH nodes will be selected among the available sensor nodes in the region. The CH node also acts as a gateway between its members and the base station node (BS). Generally, the CH nodes will create time division multiple access (TDMA) schedule to manage communication between sensor nodes and CH nodes, and between the CH nodes and the BS [1, 9]. It will also be responsible for performing data aggregation to minimize the number of packet data to the BS.

Among many approaches for building cluster-based protocols for WSN environments [7, 10], the Fuzzy C-Means (FCM) [3, 11, 12] is one of the popular clustering techniques for WSN environments. In this approach, the sensor nodes are linked into clusters with the degree of memberships to each cluster. It requires iterative computation processes to compute the degree of memberships of each sensor node to all WSN clusters. The processes will be continued until the clustering processes reach its convergence or where a certain condition is satisfied [3, 13]. The main idea of the approach is to guarantee uniform creation of clusters by minimizing the distance between the cluster members and their CH nodes [11, 12]. Nevertheless, there are several limitations to these FCM-based approaches. Firstly, the clustering results are sensitive to the selection of the initial cluster, where commonly, the random selection in center points can make the iterative process fall into a local optimal solution. Accordingly, some heuristic optimization algorithms can be applied to deal with this challenge [13, 14, 15]. However, these processes may spend significant computation time if a large number of objects and attributes of clustering are involved. These facts can bring another issue when adopting this approach for WSN nodes with limited energy and computation power.

On the other hand, by applying only the clustering approach for the WSN environment without employing CH rotation schemes raise another issue of the network lifetime. A different approach of hierarchical-based routing for WSN environments is the LEACH-based approach, as shown in [9], which is then extended in many research works such as in Leach-B [16], ANTACLUST [17], EM-LEACH [18], and Modified LEACH [1]. In these approaches, the clusters of sensor nodes are formed using a distributed cluster formation approach where each sensor node can make individual decision to become CH candidates or cluster members. The LEACH-based approach divides the process of cluster formation into round-based where, in each round, a threshold value is calculated based on various criteria, e.g., round number, percentage number of cluster head (CH), remaining energy, number of neighbors, the position of sensor nodes. Each node then generates a random for the current round [7]. The sensor node becomes a CH candidate node if the generated random number is less than the threshold value for the corresponding round. Furthermore, the LEACH-based approaches apply CH rotation in each cluster operation round to balance energy among the distributed sensor nodes. However, the main drawback of the LEACH-based approaches is by using random number generation often produces unbalanced CHs distribution in the WSN network. It can lead to unbalance energy consumption to the entire network.

The other well-known approach is PEGASIS [10, 19], which builds hierarchical-based routing by applying the concept of chaining. The greedy approach is applied to build paths between sensor nodes and form the chain. In each sensor node in the chain, data fusion is performed, followed by data transmitting the fused data to the next node in the chain until finally reach the chain leader. In each round, a random leader is selected among sensor nodes in the chain. The leader is responsible for transmitting the aggregated data to the BS. Despite the advance of the PEGASIS approach, some drawbacks need to be addressed. Firstly, the use of the random approach in selecting a leader as CH node can bring long-chain transmission in long multi-hop scenarios. Secondly, the data fusion process must be performed in all sensor nodes in the chain that also consume energy and time cost to complete the process. Lastly, if the number of sensor nodes is large, it will produce a very long chain. Hence, the required processes can breed too much energy consumption.

We have discussed the advantages and highlighted the limitation of the existing hierarchical-based approaches to address the challenges of data gathering operations in WSN environments. This paper proposes a novel approach to build a hierarchical-based approach for WSN environments based on the concept of node density for construction clusters of sensor nodes. We then also extend the approach with a CH rotations scheme to balance energy among the existing CH candidates and prolong the WSN network lifetime.

The main contributions of the paper are:

1) We propose a new approach of clustering the WSN based on node density suitable for WSN sensor nodes with limited energy and computation power by elimi-nating random initial CH selection and minimizing high iterative computation processes in the cluster formation.
2) We propose a new approach to minimize cluster overlapping and maintain effective distribution of clusters in the WSN monitoring field. To address this challenge, the optimal number of clusters is estimated to avoid random selection as shown in LEACH-based and PEGASIS-based approach. The effective distance between the centre of clusters is also estimated to minimize overlapping between clusters and distribute the CH locations in the field.
3) We then integrate the above schemes to build a new hierarchical-based communication scheme for WSN environments to minimize energy consumption and prolong the network lifetime for data-gathering scenarios in WSN environments.

This paper is organized as follows: Section 1 presents the background of cluster-based routing protocol for WSN. Section 2 discusses cluster construction and energy model for WSN environments. In Section 3, we discuss the details of the proposed cluster formation and data transmission algorithms. The simulation setting, parameters, experiment results, and their analysis are discussed in Section 4, and finally, Section 5 concludes the paper.

2. **Background Theory and Related Work.**

WSN sensor nodes have energy-related research issues since commonly; their primary energy source is a battery. Various challenges need to be addressed to build an energy-efficient yet robust communication protocol to support data-gathering operations in WSN environments. In this section, we discuss related work of clustering for WSN, the energy consumption model, problems of cluster head selection in the existing approaches to highlight our research contributions. We start by discussing clustering for WSN environments, as discussed in the following sub-section.

2.1. **Energy-efficient data-gathering operations in WSN environments.**

The limited bandwidth and energy resource of WSN sensor nodes for communications have highlighted the need to build an energy-efficient data-gathering technique for WSN environments. Clustering has been proposed as one of the fundamental approaches to decrease energy consumption and prolong the network lifetime of WSN [7]. The objective of the clustering for WSN is to divide the sensor nodes into clusters where each of them is controlled by a cluster head, responsible as a gateway between the cluster members and the base station (BS) during the data gathering operations [7]. The clustering approach also gives benefits by limiting the radius of communications within inter-cluster nodes and reducing the possibility of interfering with other nodes by adjusting node transmission power [1, 6]. Dividing WSN into numerous clusters of sensor nodes also aims to reduce long-distance transmission from sensor nodes to the BS [9].

A new different approach to avoid direct communications between CH nodes and the BS in the data gathering process is by employing mobile BS. In this new approach, the assumption is made that the BS can travel to pick up the gathered data in each CH by using an unmanned aerial vehicle (UAV) shown in [20, 21]. This approach was proposed to eliminate direct communications between CH nodes and the BS. Nevertheless, this new approach requires additional equipment to carry the BS to visit all clusters and establish communications with the selected CH nodes in each cluster. Since generally, UAV devices also use battery power and consume significant energy to support their mobility, we do not consider the option to have mobile sink communication approaches in this paper.

Several constraints, and assumptions have been applied in the literature [1, 6, 8, 9, 11] to build effective cluster-based communication for WSN environments, as described below:

- All sensor nodes have similar hardware capabilities. Batteries power them with limited energy constraint.
- The battery sources for all sensor nodes are not rechargeable, and the initial battery powers for all sensor nodes are assigned to be equal.
- All sensor nodes are randomly deployed at the sensor monitoring field without any obstacles that may obstruct the communication among the sensor nodes.
- Each sensor node is assumed to have the capability to adjust their transmission power to transmit their data to the BS if necessary or their corresponding CH nodes.
- Each sensor node knows its position or at least knows the distances between sensor nodes based on the received signal strength indicator (RSSI).
- The wireless communication between nodes is symmetrical, where the energy consumption of data transmission between two nodes is equal from both sides and direction.

Minimizing energy consumption in WSN communications and prolonging network lifetime are the main goals of this paper. In the following sub-section, the energy consumption model for WSN communications is discussed in detail.

2.2. Energy consumption model for communications in WSN environments.

In general, the energy consumption model for data communication in WSN environments uses the model as described in [1, 3, 9, 11, 23]. In this approach, the energy consumption will be estimated based on the distance between the sender and receiver. We adopt a similar energy consumption model for this paper.

We denote that E_{elect} is the amount of energy consumed in the circuit (e.g., for filtering and modulation). Furthermore, by considering the multipath fading channel, energy consumption of transmitting circuit and power amplification, the energy consumption to transit k bits message over a distance d can be approximated as follows:

$$E_{Tx}(k,d) = E_{Tx\text{-}elec}(k) + T_{Tx\text{-}amp}(k,d) \begin{cases} E_{elec} * k + \varepsilon_{fs} * k * d^2 & d \leq d_p. \\ E_{elec} * k + \varepsilon_{mp} * k * d^4 & d \geq d_p. \end{cases} \quad (1)$$

$E_{Tx\text{-}elect}$ is the amount of energy consumed in the circuit (e.g., for signal filtering and modulation) for the transmitter, and $E_{Tx\text{-}amp}$ is the amount of energy required for the transmitter while transmitting a one-bit packet, k is bit data and d is the distance between the transmitter and receiver. The power consumption in the free space propagation model (d^2 propagation loss) is denoted as ε_{fs} (pJ/bit/m^2), and ε_{mp} (pJ/bit/m^4) is the power consumption in the multipath fading propagation model (d^4 propagation loss). The cross-over distance between the two propagation models is denoted as d_p, that is used as a transmission threshold [24, 25] as computed using the following equation:

$$d_p = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}} \tag{2}$$

Energy for receiving k-bits of a packet can be estimated as follows:

$$E_{Rx}(k) = E_{Rx\text{-}elec}(k) = E_{elec} * k \tag{3}$$

Using this model, suppose that energy for data aggregation in each cluster head for one-bit packet is E_{DA}, the energy for aggregating k bit data is denoted as follows:

$$E_{DA}(k) = E_{DA} * k \tag{4}$$

2.3. **Clustering for energy-efficient data-gathering operations.** There are various approaches to build hierarchical-based protocols by applying various data clustering algorithms to segment the WSN into numerous clusters of sensor nodes [7]. One of the popular techniques is by adopting Fuzzy C-Means [3, 11, 12] by which sensor nodes are grouped into clusters with different membership degrees. Generally, the FCM approaches apply random initialization of cluster centroids that may cause clusters fluctuations, which then can produce local optimal clustering solutions [14]. Many techniques have been applied to deal with the cluster initialization, such as using the particle swarm optimization (P-SO) [14], genetic algorithm, or ant colony optimizations [13, 15]. Using these approaches generally, requires high iterations that may need long computation time to find the best solution. On the other hand, the WSN nodes generally consist of small devices with limited energy sources and computation capabilities. These related challenges can limit the application of those algorithms for WSN environments due to the related issues of limited energy and computation capability.

Dividing WSN into numerous clusters of sensor nodes works aim to reduce long-distance transmission from sensor nodes to the BS. In cluster-based WSN systems, the CH nodes are responsible for gathering data and aggregating sensing from their member nodes and forward them to the BS. However, using only the clustering approach, without employing rotation among the potential candidates of CH nodes, can deplete the energy of the selected CH nodes faster than the member nodes. This phenomenon may bring another issue of network lifetime and unbalance energy consumption among the available sensor nodes. To address these issues, the LEACH and PEGASIS-based approaches introduce the concept of periodic CH rotations [9, 10, 22] to balance the energy consumption among sensor nodes. The LEACH-based approach [1, 9, 10, 22] is a cluster-based communication technique that applies round-based mechanism [7]. The round-based approach arranges the data gathering and transferring processes into the setup-state and steady-state. During the setup phase, a threshold for the associated round is computed. Then every node generated a random number and compared to the threshold for the current round. The threshold number for each round is computed using Equation (5), as described below [1, 9, 10].

$$T(n) = \begin{cases} \dfrac{p}{1 - p\left(r * \bmod\left(\frac{1}{p}\right)\right)} & \text{if } n \in \hat{N}. \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

In Equation (5), p is the desired percentage of CH nodes, r is the current round, $\hat{N}$ is a set of sensor nodes that have not been elected as cluster heads in the last $\frac{1}{p}$ rounds. This approach will balance energy since every node becomes CH node only once within $\frac{1}{p}$ rounds. If the generated random number in each sensor node is less than the threshold for the current round, the corresponding sensor node becomes a CH candidate for the round.

Despite various advantages of LEACH-based approaches to balance energy consumption and prolonging network lifetime in WSN environments, there some weaknesses need to be addressed to improve the performance of these approaches. They are:

- The LEACH-based approaches use a random number to determine CS status, and it may lead to an unbalanced distribution of CH nodes in the area.
- The number of CH can randomly fluctuate, which leads to energy inefficiency due to overhead in CH rotation.
- The probability-based approach in LEACH does not consider the density of sensor node distribution in selecting CH candidates. Hence nodes located in less populated sensor nodes can be selected as CH candidates.
- LEACH-based approach does not consider the distance between formed CH candidates. Therefore cluster overlapping can occur randomly.

The other well-known approach to deal with energy-efficient data gathering in WSN is the PEGASIS approach that introduces the chain-based cluster for data gathering in WSN environments [10, 19, 26]. The PEGASIS-based approaches apply a chain-based approach where sensor nodes located far away from the BS will only send to the nearest neighbor node and build the chain connection. The process of building the chain will be continued until the whole chain has been built. In this approach, the leader in each round is responsible for forwarding the aggregated data to the BS.

Figure 1 illustrates the PEGASIS approach, where the sensor node s_3 is selected as the chain leader to transmit the aggregated data to the BS. The leader will be randomly selected among the sensor nodes in the chain that will take turns in sending data to the BS. The leader node will be selected at a random position on the chain to balance energy consumption by ensuring that sensor nodes die at random locations [10]. When any sensor node dies, the chain is again reconstructed the same manner by ignoring dead nodes to be included in the chain. The idea of sensor nodes (as leader) dying at random place is to enhance the WSN robustness [7].

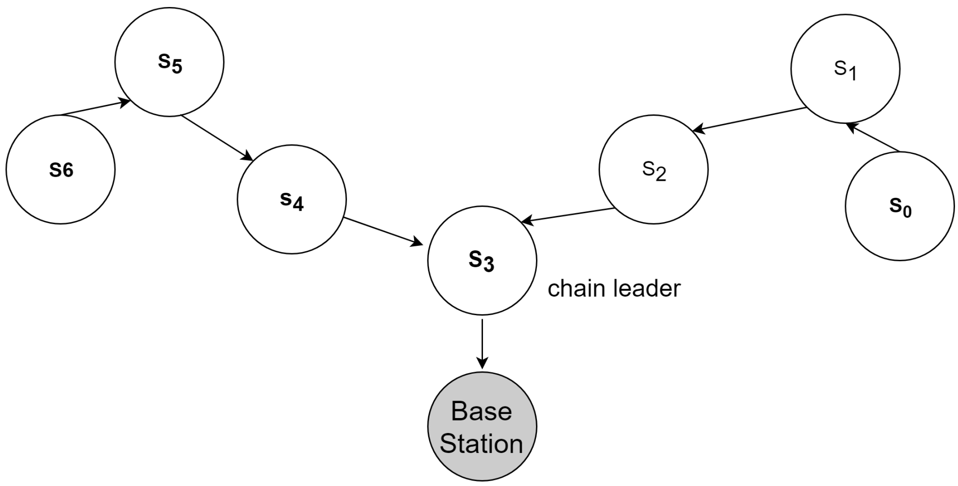

FIGURE 1. PEGASIS chain with the sensor node s_3 as the chain leader

The PEGASIS approach shows better performance compared to LEACH in some aspects [7, 10], in particular by decreasing the number of data transmission volume through chain data aggregation and balance energy distribution in the network by random chain leader selection. However, there are some disadvantages of chain-based cluster approach as applied in PEGASIS, especially problems related to the process of chain maintenance when the topology is changing, especially when some nodes die during the operation. Furthermore, hop-by-hop transmission along the chain will produce significant delay and energy for data fusion or aggregation in each sensor node along the chain. The random leader of CH selection in both the LEACH and PEGASIS-based approaches also introduces another issue related to unbalance energy consumption due to random characteristic. To deal with the above issues of both LEACH, we proposed a density-based clustering protocol for WSN where nodes located within central of cluster density will be selected as CH candidates and avoid random selection, as shown in the LEACH-based approach. In each round, to minimize overlapping between clusters, two CH nodes or more must not be located in the same cluster. Hence the radius of the cluster must be defined. Furthermore, to balance energy among CH rotating scheme to balance energy among sensor nodes in each cluster is also proposed. We discussed the proposed approach in the following section.

3. **Proposed WSN Clustering and Cluster Rotation Approach Based on Node Density.** WSN generally consists of large numbers of sensor nodes, randomly distributed in a sensor-monitoring field. The hierarchical-based protocol for WSN is known as one of the main approaches to minimize energy consumption during data transmission in WSN environments. The main idea of the approach is to reduce the number of long-distance data transmissions by constructing clusters of sensor nodes. Each cluster is controlled by a CH node to perform data gathering and transfer the aggregated data to the BS. The processes of cluster formation and data transfer are done within the round-based mechanism, as discussed in the following sub-section.

3.1. **Round-based mechanism.** We apply the round-based clustering algorithm as used in all LEACH-based clustering approaches [1, 3, 9, 11]. The round is used as a unit that consisted of the cluster formation phase and the data transmission phases. The cluster formation phase concerns the formation of the clusters and cluster heads selection. When the clusters have been established, all sensor nodes and selected CH nodes can start the data-transferring phase. Figure 2 illustrates the round-based clustering approach, as implemented in this paper.

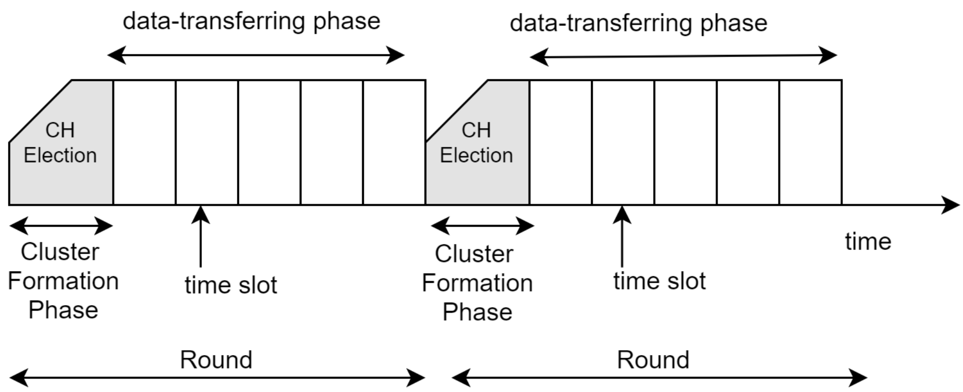

Figure 2. Round-based mechanism with TDMA scheduling

All sensor nodes will organize themselves into clusters with one node in each cluster selected as the CH node. For the selection of CH nodes, all sensor nodes that are elected as CH nodes are assumed capable of transmitting data to the base station (BS) if required. It is done by adjusting their transmission powers. This assumption is similar to the clustering approaches applied for WSN environments, as shown in [1, 3, 9, 11]. In each cluster, member sensor nodes will transmit their data to the corresponding CH. The CH will then perform data aggregation before forwarding the data to the remote BS. All member nodes are required to adjust their transmission power based on the received signal strength from CH advertisements that also represent the distance between the member nodes and their CH nodes. The CH nodes then also arrange the time division multiple access (TDMA) schedule for each of its cluster members to send the data to the CH nodes [1, 9]. The CH node will then aggregate the incoming data and send it to the BS. In this scenario, being a CH requires higher energy consumption than a regular member node. Hence, there must be a smart mechanism to select and rotate CH nodes to balance energy consumption and prolong the network lifetime. We discuss our approach to deal with these challenges in the following sub-section.

3.2. **Cluster formation and cluster head election phase.** During this phase, clusters of sensor nodes are built where CH candidates are selected among the available sensor nodes in each round. Before we start the CH selection process, first, we need to determine the number of optimum clusters since the total energy consumption of network in the WSN is highly influenced by the number of its clusters [9, 23]. The following sub-section discusses the related processes.

3.2.1. *Determining number of clusters.* Let $SN^t = s_1, s_2, s_3, \ldots, s_Z$ be Z number of alive sensor nodes at round r_t, placed in two-dimensional spaces of a WSN sensor monitoring field where the base station (BS) located at the center of the area with a circular radius of wireless transmission R as illustrated in Figure 3. Let k be the number of clusters, the average number of sensor nodes in each cluster will be $\frac{Z}{k}$. The energy consumption of a CH consisted of energy consumption for receiving and aggregating data as well as forwarding the aggregated data to the BS. We follow an approach as described in [23] that if the area of the network is πR^2, and there are k clusters in the network, the following equation can determine the average of cluster areas.

$$Avg_C = \frac{\pi R^2}{k} \tag{6}$$

If a cluster $\hat{C}$ is also considered as a circle with its CH located at the center of the cluster, the coverage area of the cluster can be approximated as $\pi \hat{r}^2$. Thus, the optimal radius of the cluster can be estimated as follows:

$$\hat{r} = \frac{R}{\sqrt{k}} \tag{7}$$

To find the optimal number of k, by considering energy efficiency, we extend an approach to find the optimal cluster number as described in [23]. Let Z be the number sensor node, ε_{fs} is the power consumption in the free space propagation model, E_{DA} is the power required for CH to perform data aggregation per bit data, and let $SN_{x,y}$ be the standard deviation of all sensor node positions. The number of clusters in the WSN network can be obtained using the following equation:

$$k = \sqrt{\frac{Z\varepsilon_{fs}R^2}{2(E_{elec} + E_{DA})}} + \frac{Z + SN_{xy}}{d_p} \tag{8}$$

where d_p is the cross-over distance between the two propagation models as the transmission threshold, computed by using Equation (2).

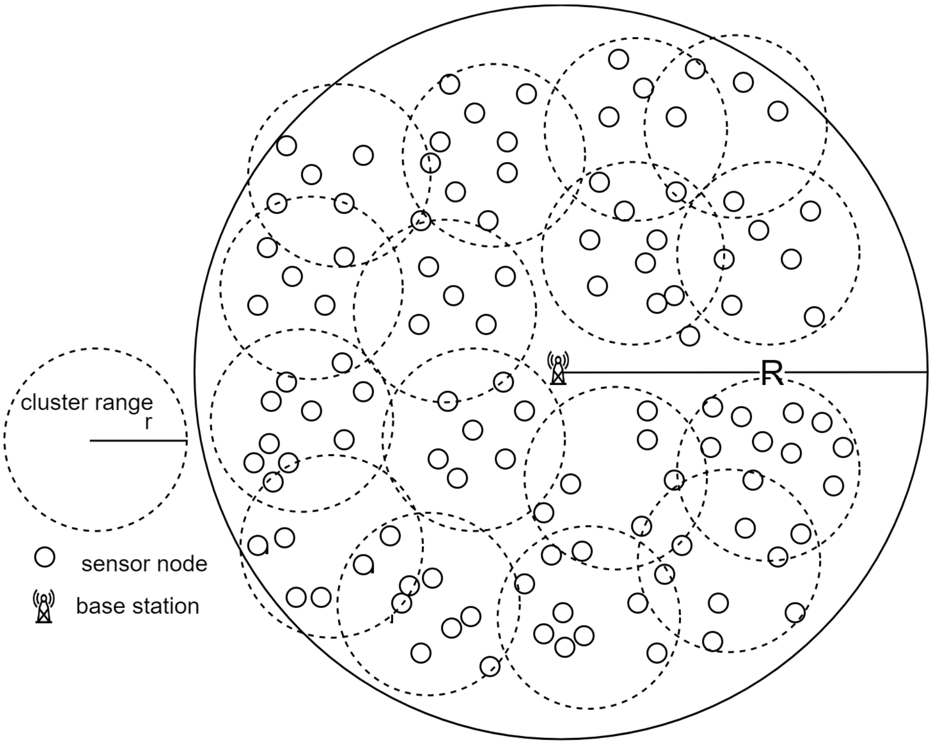

FIGURE 3. A WSN divided into some clusters with the base station located at the center of the area

3.2.2. *Cluster head election using the node density approach.* Our idea of selecting CH nodes is based on each sensor node density. In the cluster formation phase of each round, each sensor node s_i computes the number of alive sensor nodes within the radius of cluster $\hat{r}$ as defined in Equation (7). We then define $N_i^t = \{s_1, s_2, s_3, \ldots, s_j\}$ where $s_j \in SN^t$, as a set of neighboring nodes for the sensor node s_i at round r_t. The node density for a sensor node s_i, is defined as λ_i^t. It represents the total number of neighbors of the sensor node s_i, where their remaining energy is greater than the minimum energy threshold T, to perform data processing and communications at round $\hat{r}_t$. The higher node density indicates that the sensor node has many neighbors. If the node densities of sensor nodes around the sensor node s_i are also high, it means that surrounding sensor nodes have a similar number of neighbors for the corresponding round.

Let CH_t be all sensor nodes which have been selected as CH nodes at the round $\hat{r}_t$ and $s_{i:xy}$ be the location of sensor node s_i in a two-dimensional space. We define $\min(\Delta(s_{i:xy}, CH_{t:xy}))$ as the minimum distance of s_i location towards the locations of all sensor nodes which have been included in CH_t for the round. To build a cluster and prolong the network lifetime, CH selection and rotation processes are required. Two CH nodes at the same time should not become the CH node for the same cluster. Hence, the distance among all selected CH nodes should be bigger than the radius cluster $\hat{r}$.

Once a sensor node s_i is selected as CH candidate for the current round $\hat{r}_t$ for any cluster, to balance energy consumption, CH rotation scheme must be applied by assigning the next available round for CH election process for the corresponding node. Let δ value represent the percentage number of alive sensor nodes within a cluster assigned to be available for CH selection and rotation processes. Given $|N_i^t|$ as the number of alive neighbors of node

s_i for the current round $\hat{r}_t$, the next available round for the sensor node to be eligible for CH selection denoted as $s\hat{r}_{i,t}$, is computed using the following equation:

$$s\hat{r}_{i,t} = \hat{r}_t + \delta|N_i^t| \quad (9)$$

Algorithm 1: CH nodes selection and rotation in each round

```
Result: List of sensor nodes being selected as CH nodes for the current round and
        the next available rounds for these selected nodes, to be included in the CH
        selection processes;
#Step 1 Initialization for a new round r̂_t;
SN^t ← a set of alive nodes for the current round r̂_t;
r̂ ← get optimal radius of clusters using Equation (7);
kCluster ← get optimal number of clusters using Equation (8);
while there still exist alive nodes in SN^t do
    r̂_t ← current clustering round;
    #Step 2 Clear list of CH nodes and node densities from the previous round;
    clear ĈH_t; clear Λ_i^t;
    #Step 3 Estimate node densities for the current round;
    foreach s_i ∈ SN^t do
        clear λ_i^t;
        foreach s_j ∈ N_i^t do
            if the neighbor's energy e_j^t ≥ T then
                λ_i^t ← λ_i^t + 1;
            end
        end
        add λ_i^t into Λ_i^t;
    end
    #Step 4 Select CH nodes using node density approach;
    ŜN^t = sort(SN^t, Λ_i^t, Descending) #get sorted noded based on node density;
    i ← 1; ch ← 1; #index number for member nodes and CH nodes;
    while (ch ≤ kCluster and i ≤ ŜN^t) do
        # check if the sensor node is available to become CH node for the current
         round;
        if ŜN^t[i : sr̂_{i,t}] ≤ r̂_t then
            if min(Δ(ŜN^t[i : x, y], CH_{t:x,y})) > r̂ then
                add ŜN^t[i] into CH_t;
                ch ← ch + 1;
                #Step 5 Set the next available round for this sensor node using
                 Equation (9);
                ŜN^t[i : sr̂_{i,t}] ← r̂_t + δ|N_i^t|;
            end
        end
        i ← i + 1;
    end
end
```

Algorithm 1 describes the steps to select CH nodes in each round. It is then followed by assigning the next available round for the currently selected CH nodes to be involved again in the CH selection processes. At the beginning of the CH selection process (Step 1), the approximate radius of clusters is calculated using Equation (7), then the number of optimal clusters is also calculated using Equation (8). If there are still alive sensor nodes (Step 2), the CH nodes selection is initiated by clearing CH nodes status from the previous round. Then the total number and status of alive neighbors of each existing sensor node are estimated as the node density of this sensor node the current round (Step 3).

In Step 4, the process of selecting CH nodes for the current round is initiated by sorting available nodes based on their node density status. The CH selection process is then continued by selecting sensor node $s_i \in SN^t$ with the highest node density status as a CH candidate. If this sensor node is currently eligible to be involved in the CH selection process for the current round $\hat{r}_t$, i.e., $s\hat{r}_{i,t} \leq \hat{r}_t$ and the distance of this sensor node toward the currently selected CH nodes is greater than the radius cluster threshold $\hat{r}$ (Equation (7)), the sensor nodes can be assigned as CH nodes for the current round. In Step 5, the next available rounds for the selected CH nodes are computed using Equation (9). At the end of the algorithm, the list of CH nodes for the current round, and the next available rounds for these selected CH nodes, to be included in the CH selection processes are obtained. Upon finishing these CH selection processes, the data-transferring phase is performed. The details of this phase are discussed in the next sub-section.

3.3. **Data transferring phase.** During this state, the operation is divided into rounds of units where member nodes send the data to the closest CH nodes during the allocated time division multiple access (TDMA) slots. Initially, each selected CH node broadcasts an advertisement message to declare its CH status. Each normal node will join CH node based on the strength of advertisement message signals where closer CH node will produce a stronger signal. Each sensor node that receives this message will respond to the request selecting it as the CH node for the current round.

The flowchart, as shown in Figure 4, illustrates the data-transferring phase upon completion of CH nodes selection in each round. Since the energy consumption model, as described in Section 2.2, highly depends on the distance between the transmitter and receiver, the closest CH candidate will be selected by each member node as its CH node for the current round. In this state, all CH nodes must be awake to receive transmitted data from member nodes in the cluster, followed by data aggregation before transmitting to the base station. For member nodes located closer to the BS than any other available CH nodes, they will send the data directly to the BS to preserve energy. For all of the sensor nodes, where their energy less than 1% from the initial energy allocation, will broadcast unregistered message to other nodes and will be classified as the dead nodes. In the next round, only alive nodes will be considered in the cluster construction processes. These processes will be continued while there are still exist alive nodes in the WSN monitoring field.

4. **Simulation and Performance Evaluation.** We have built a simulation under MATLAB environment to evaluate the performance of our proposed approach called DCLUST (Node **D**ensity-based **CLU**stering for wireless **S**ensor ne**T**works). To see the effectiveness of the proposed DCLUST approach in data-gathering operations, we compare the result of the experiments with two other well-known PEGASIS [10], and LEACH [22] approaches. Four performance metrics are being evaluated during the simulations. They include the network lifetime in terms of rounds of events when the first node die (FND), and

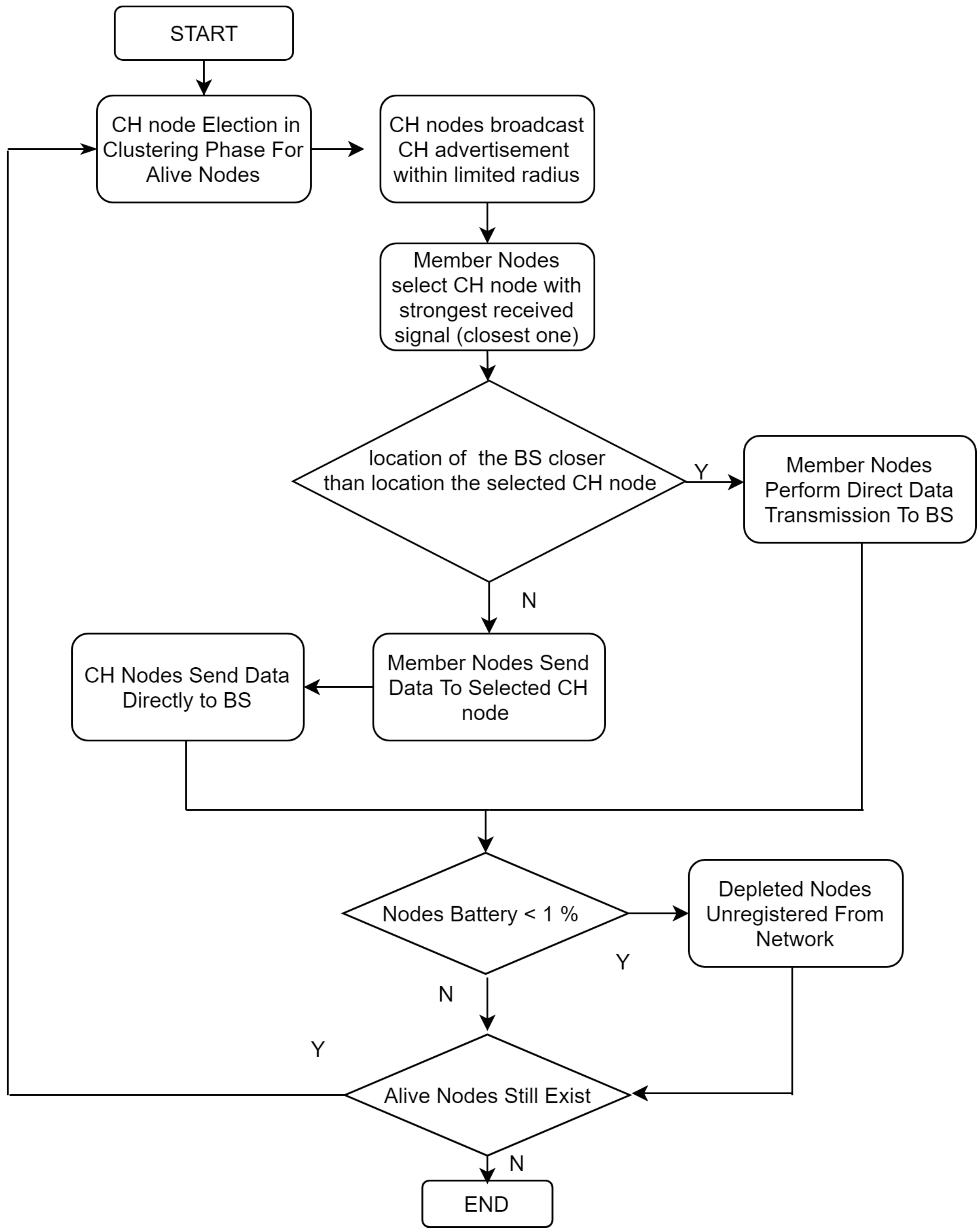

Figure 4. Flowchart of data transferring scheme

when the last node die (LND), the difference of rounds between the half node die (HND), and the last node die (LND) events to verify network lifetime and stability, total energy dissipation, and the average remaining energy of sensor nodes during the data-gathering operations.

Table 1 describes the simulation parameters of the experiments wherein each simulation round, 4000 bits of packet data, will be transferred from sensor nodes to their respective CH nodes during the data-gathering operations. The BS is stationery placed at the center of each WSN monitoring area. In the beginning, all sensor nodes have the same initial battery energy values at 2 joules. The p value that represents the LEACH-based percentage for CH (See Equation (5)) is assigned at 0.05.

Table 1. Simulation parameters

Simulation Parameters	Value
Number of Nodes	100, 150
WSN Monitoring Areas	100 m^2, 200 m^2, 250 m^2
R	50 m, 100 m, 125 m
p (LEACH)	0.05
E_{elec}	50 nJ/bit
ε_{fs}	10 pJ/bit/m^2
E_{DA}	5 nJ/bit/message
ε_{amp}	0.0013 pJ/bit/m^4
δ	0.7
Initial Energy	2 joule
Size of Packet	4000

We apply the energy consumption model, as described in Section 2.2. During the simulation, the energy for circuit processing per bits (E_{elec}) is 50 nJ/bit, energy for communication in the free space propagation model ε_{fs} (d^2 propagation loss) is 10 pJ/bit/m^2 while ε_{amp} value that represents the power consumption in the multipath fading propagation mode (d^4 propagation loss) is assigned at 0.0013 pJ/bit/m^4. The energy for data aggregation for the CH nodes E_{DA} is 5 nJ/bit/message. The δ parameter for CH rotation of the DCLUST approach is set at 0.7 (See (Equation (9)). The minimum remaining energy threshold T for sensor nodes is 1% from the sensor node's initial battery assignment. The experiments are conducted in scenarios where 150 and 200 sensor nodes are randomly deployed in different sizes of WSN monitoring fields, i.e., 100 m^2, 200 m^2 and 250 m^2, to see the performance of the proposed approach in different node densities compared to the LEACH and PEGASIS approaches.

4.1. **Operational nodes and network lifetime.** We observe the number of operational nodes (live nodes) in each round and then analyze the result to obtain the network lifetime of the WSN. The experiments are conducted by randomly deploying 100 and 150 wireless sensor nodes in different sizes of WSN monitoring fields to see the impact of different node densities toward the performance of the proposed approach. During the simulation, each sensor node will generate packets at a constant rate, which then are transmitted to the BS.

We start by evaluating the round of events when the first node die (FND) of the proposed DCLUST approach compared to the other two well-known LEACH, and PEGASIS approaches. Prolonging the FND round is a fundamental issue in the WSN research since it leads to the network stability issues [2, 6]. Table 2 shows the rounds of FND and LND events when 100 and 150 sensor nodes are randomly deployed and perform data-gathering processes in different sizes of WSN monitoring fields.

The table shows that the proposed DCLUST approach gives the most extended FND round rounds in all sizes of WSN monitoring areas, i.e., 100 m^2, 200 m^2, and 250 m^2 compared to the other two approaches. In the first experiment scenario, 100 sensor nodes were randomly deployed in 100 m^2 WSN area, the proposed DCLUST produced the best FND round at 3906, followed by LEACH at round 3711 and PEGASIS at round 3852. The experiments were continued by using 150 nodes for the same area. Although all approaches, especially the PEGASIS approach, got earlier rounds of FND events than the first experiments with 100 nodes, the proposed DCLUST approach still gave better results compared to the other two approaches. In the next experiments, the WSN monitoring

Table 2. Rounds of when the first node die (FND) and last node die (LND)

WSN areas	Rounds of FND events					
	100 Nodes			150 Nodes		
	DCLUST	LEACH	PEGASIS	DCLUST	LEACH	PEGASIS
100 m^2	3906	3711	3852	3761	3650	1623
200 m^2	2866	1646	529	2661	2656	865
250 m^2	1523	777	160	1949	1465	258
WSN areas	Rounds of LND events					
	100 Nodes			150 Nodes		
	DCLUST	LEACH	PEGASIS	DCLUST	LEACH	PEGASIS
100 m^2	6639	5014	6075	7472	5395	6129
200 m^2	7539	3758	4990	7849	3991	4980
250 m^2	6081	3734	4616	6628	3908	4577

area was changed to 200 m^2. In these experiments, the proposed DCLUST approach also gave better results compared to the other two approaches. The rounds of FND events for 100 and 150 sensor nodes were 2866, and 2661 compared to the LEACH approach that produced FND rounds at rounds 1646 and 2656, followed by the FND event of the PEGASIS approach that occurred at round 529 and 865.

In the last experiments with the WSN area of 250 m^2, the proposed DCLUST still outperforms the other two approaches in terms of FND events. The FND rounds produced by LEACH, and PEGASIS approaches for both 100 and 150 sensor nodes occurred at rounds 777 and 1465 for the LEACH, and 160 and 258 for the PEGASIS approach. On the other hand, the FND rounds produced by DCLUST were 1523 and 1949. It has highlighted the excellent performance of the proposed DCLUST approach in prolonging WSN network lifetime in data gathering scenario processes.

The results of the experiments, as shown in the second part of Table 2, denote the rounds of LND events, i.e., when the last sensor nodes die. In a 100 m^2 WSN area, the proposed DCLUST approach gave longer LND rounds at 6639 for 100 nodes and 7472 for 150 sensor nodes compared to the LEACH approach that produced the LND rounds 5014 for 100 nodes and 5395 for 150 nodes. The proposed DCLUST approach also outperformed the PEGASIS approach that gave LND rounds at 6075 for 100 nodes, and at the round 6129 for 150 nodes. In the next experiments for WSN area 200 m^2, the proposed DCLUST approach also produced the most extended rounds of LND events at rounds 7539 and 7849 for 100 and 150 sensor nodes, compared to the LEACH approach at rounds 3758 and 3991, and at rounds 4990 and 4980 for the PEGASIS approach. The similar results are also shown in the last experiments for the WSN area of 250 m^2 WSN, where the proposed DCLUST approach gave the most extended LND event rounds at 6081 and 6628, followed by PEGASIS at rounds 4616 and 4577 and the LEACH approach at rounds 3734 and 3908. Since prolonging FND and LND events are among the critical factors of WSN performances, the results of FND and LND experiments in various simulation scenarios as shown in Table 2, have highlighted the excellent performances of the proposed DCLUST approach in prolonging the network lifetime of WSN network during the data gathering scenarios.

To have more comprehensive views of the experiment results, we show the number of alive nodes (operational nodes) produced by DCLUST, LEACH and PEGASIS approaches from the beginning of the simulation until the LND events round as shown in Figure 5.

The figure shows that the proposed DCLUST approach outperforms the LEACH and PEGASIS approaches in terms of FND and LND rounds in all scenarios, as also have been reported in Table 2. On the other hand, the figure also shows that the PEGASIS approach produced slightly better results of alive nodes in the middle of the simulation rounds toward the proposed DCLUST approach. Nevertheless, it can be seen that after that, the PEGASIS approach lost its alive nodes relatively faster than the other two approaches. Then the differences of rounds between the HND event rounds until the round of LND events are reported in Table 3 to have more detail views about these results. The table shows that the proposed DCLUST approach produced longer rounds

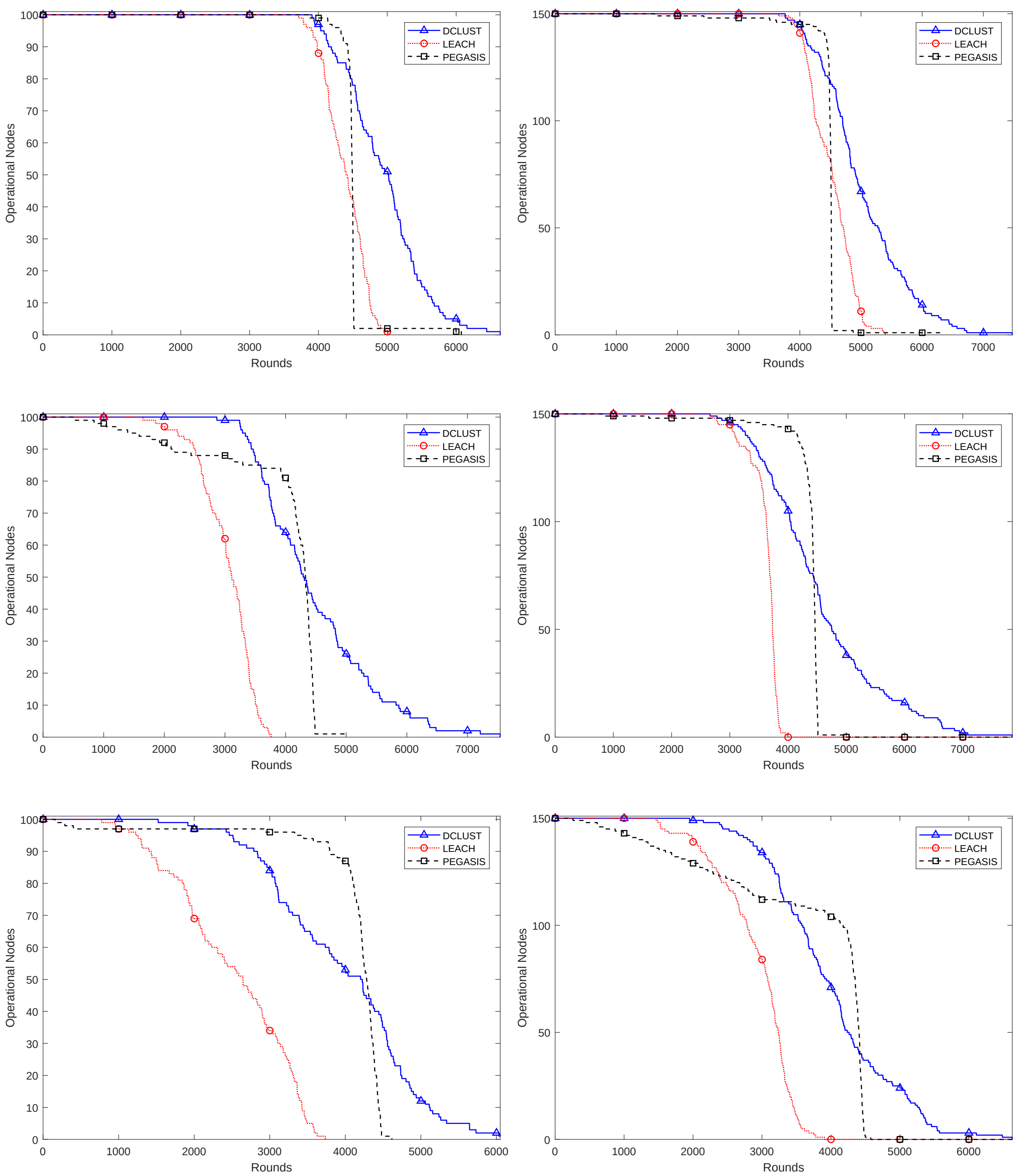

FIGURE 5. Number of operational nodes in each round until the last sensor node dies (LND) in 100 m^2 (top), 150 m^2 (middle) and 250 m^2 (bottom) of WSN areas

Table 3. Difference of rounds between half node die (HND) and last node die (LND)

WSN areas	Difference of rounds between HND and LND events					
	100 Nodes			150 Nodes		
	DCLUST	LEACH	PEGASIS	DCLUST	LEACH	PEGASIS
100 m^2	2093	868	1605	2879	1188	1645
200 m^2	3808	1020	874	4014	415	609
250 m^2	2965	1810	478	3314	1280	1586

between the HND and LND events than the other LEACH and PEGASIS approaches in all simulation scenarios. The longer differences of rounds between these two events can represent the more extended stability of the network structure. This fact has also emphasized the excellent performance of the proposed DCLUST approach in maintaining the network lifetime of WSN nodes during the data-gathering operations.

4.2. **Energy consumption during data-gathering operations.** In this section, the energy consumption of the proposed approach compared with the other two approaches are observed. The energy consumption model of the WSN includes energy for transmitting, receiving, and aggregation mode, as discussed in Section 2. Figure 6 consisted of three sub-figures that depict the average of the remaining energy of current operating nodes in each particular round. In the top figure and middle figure, the average remaining energy of 100 and 150 sensor nodes in 100 m^2 and 200 m^2 WSN areas are depicted. These figures show that the proposed DCLUST approach outperforms the other two approaches in maintaining the remaining energy of sensor nodes. In 250 m^2 area of sensor monitoring, which means that the node density in the network is lower than the first two experiments, the PEGASIS approach shows better average remaining energy at the beginning of the simulation rounds until the round at around 3000. The proposed DCLUST approach outperforms it until the last round.

In Figure 7, the total energy consumption in each round is depicted to have more comprehensive views about energy consumption in different simulation scenarios. The first top figure depicts the total energy consumption for both 100 and 150 sensor nodes in a 100 m^2 WSN area. The figure shows that from the beginning of the simulation round, the proposed DCLUST approach produced the lowest total energy consumption compared to the other two approaches. The middle figure depicts the total energy consumption for a 200 m^2 WSN area, similar to the previous results; the proposed DCLUST approach produced the lowest total energy consumption for both sensor node distributions. The bottom figure shows the total energy consumption in an area with the lowest node density of our experiments, i.e., 250 m^2. In this scenario, the LEACH approach consumed the highest total energy consumption. The PEGASIS approach produced slightly lower total energy consumption at the beginning of the simulation round compared to the proposed DCLUST approach. However, after the simulation round at 4000, the proposed DCLUST approach had lower total energy consumption than the PEGASIS approach for both sensor node distributions. This fact is similar to the results shown in Figure 5 wherein a 250 m^2 WSN area, the PEGASIS approach, rapidly lost their sensor nodes compared to the proposed DCLUST approach. Accordingly, the PEGASIS approach produced earlier LND events compared to the proposed DCLUST approach.

The results and analysis of the experiments have highlighted the excellent performances of the proposed DCLUST approach in prolonging the network lifetime of WSN during the data-gathering operations compared to the other two well-known approaches in different

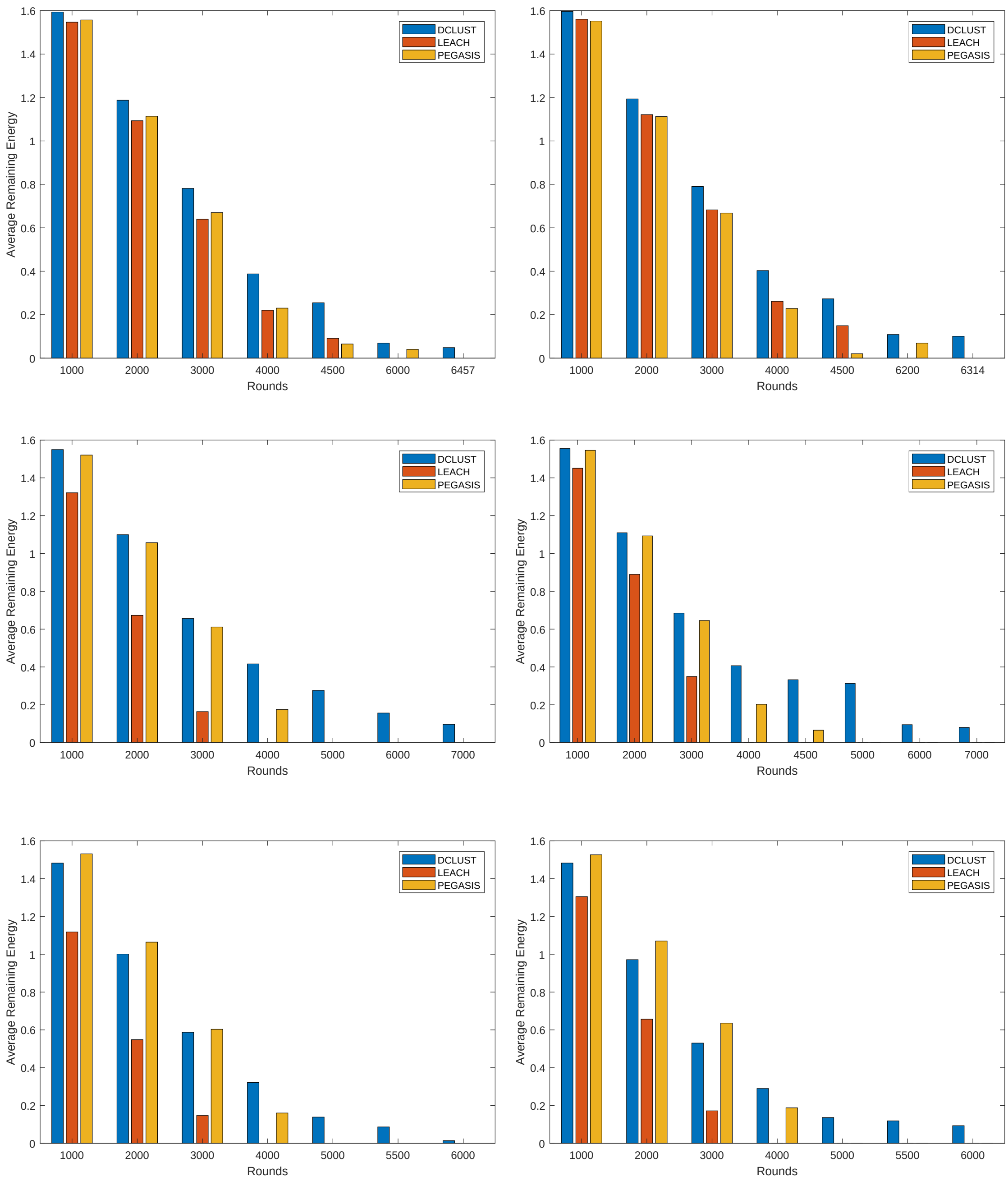

Figure 6. Average remaining energies of sensor nodes during data gathering operations in 100 m^2 (top), 150 m^2 (middle) and 250 m^2 (bottom) of WSN areas for both 100 and 150 sensor nodes distributions

scenarios of node densities. DCLUST approach can extend the FND and LND events as the critical issues of data-gathering protocol compared to the other two approaches. DCLUST approach is also able to extend the difference of rounds between the HND and LND events compared to the other two approaches that highlight its robust performances in maintaining the network lifetime during the data-gathering operations. The proposed DCLUST approach is also able to minimize energy consumptions in terms of average

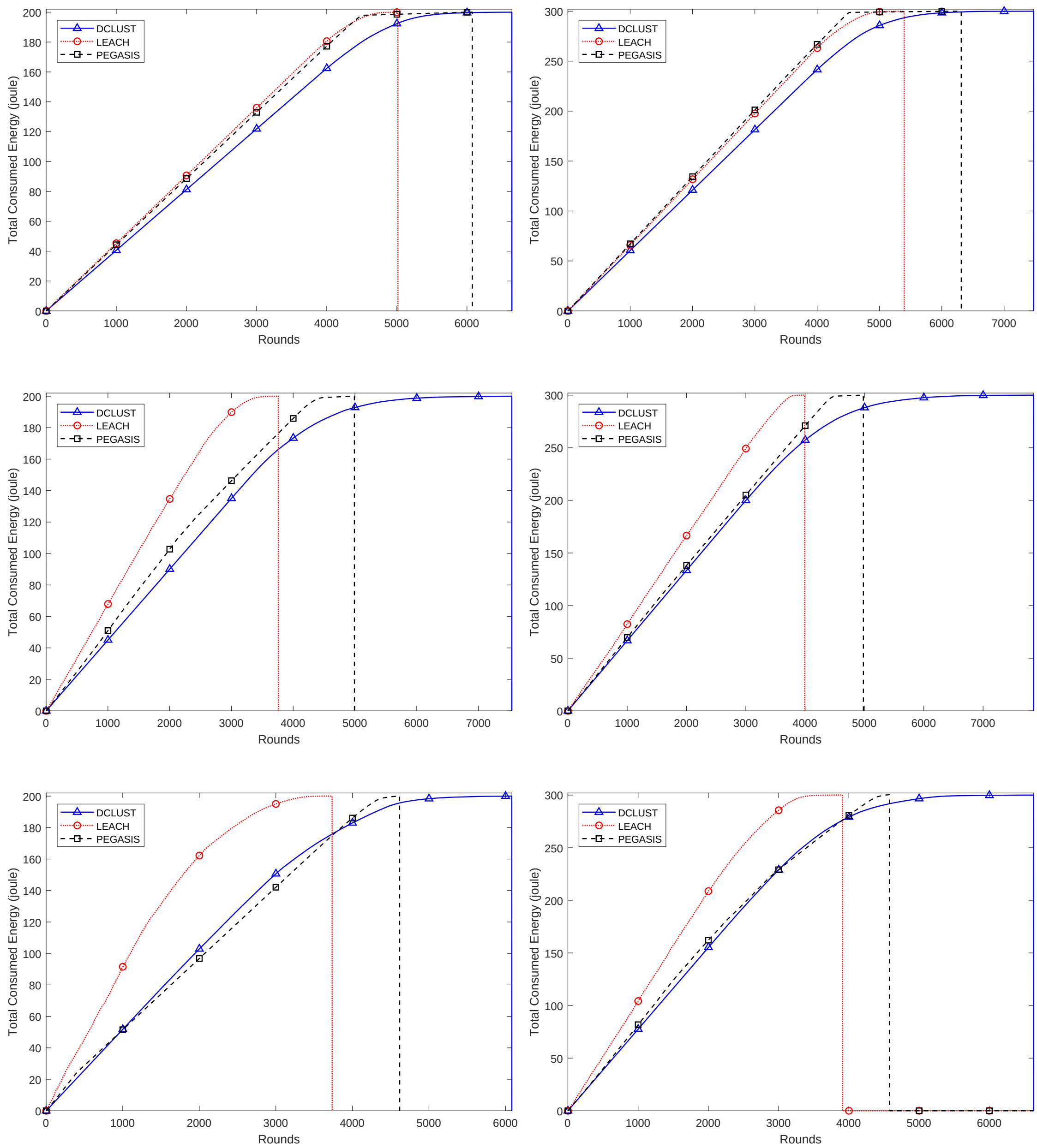

FIGURE 7. Total energy consumption in each round of 100 (left) and 150 (right) sensor nodes during data gathering operations in 100 m^2 (top), 150 m^2 (middle) and 250 m^2 (bottom) of WSN areas

remaining energy and total energy consumptions during the data-gathering operations. These results strongly highlight the contributions of the proposed work.

5. **Conclusion.** Sensor nodes with limited battery capacity bring network lifetime issues in WSN developments. This paper proposes the node density approach to build a cluster-based communication technique to minimize energy consumption during the data-gathering operations in WSN environments. The optimal cluster parameters are estimated to construct energy-efficient cluster operations. These include the number of clusters, the optimal radius of clusters, and the mechanism of CH rotation in the clusters.

Then the algorithm to perform data transmission from sensor nodes to CH and BS is also proposed.

The simulation results show that the proposed DCLUST approach can obtain better performances in terms of the number of alive nodes, postponing the occurrences for FND and LND events. The proposed approach is also able to extend the rounds between the rounds of HND and LND events compared to the other two well-known approaches. The proposed approach also produces reductions in the total number of energy consumption during the data-gathering operations that highlight the ability of the proposed approach to preserve the energy of sensor nodes during the data-gathering operations. Future work will focus on extending the proposed work in different scenarios, such as using a mobile sink scenario or priority-based data communication scheme for WSN environments.

Acknowledgment. This work is supported by Institut Teknologi Sepuluh Nopember through the LBE Research Grant for 2019. We also gratefully acknowledge the helpful comments and advices of the reviewers, which have improved the paper.

REFERENCES

[1] M. Elshrkawey, S. M. Elsherif and M. E. Wahed, An enhancement approach for reducing the energy consumption in wireless sensor networks, *Journal of King Saud University – Computer and Information Sciences*, vol.30, no.2, pp.259-267, 2017.

[2] T. Huynh and W. J. Hwang, Network lifetime maximization in wireless sensor networks with a path-constrained mobile sink, *International Journal of Distributed Sensor Networks*, vol.2015, pp.1-13, 2015.

[3] S. Su and S. Zhao, An optimal clustering mechanism based on Fuzzy-C means for wireless sensor networks, *Sustainable Computing: Informatics and Systems*, vol.8, pp.127-134, 2018.

[4] J. N. Al-Karaki and A. E. Kamal, Routing techniques in wireless sensor networks: A survey, *IEEE Wireless Communications*, vol.11, no.6, pp.6-28, 2004.

[5] B. Mamalis, D. Gavalas, C. Konstantopoulos and G. Pantziou, Clustering in wireless sensor networks, in *RFID and Sensor Networks: Architectures, Protocols, Security, and Integrations*, CRC Press, 2009.

[6] F. Fanian and M. K. Rafsanjani, Memetic fuzzy clustering protocol for wireless sensor networks: Shuffled frog leaping algorithm, *Applied Soft Computing*, vol.71, pp.568-590, 2018.

[7] X. Liu, A survey on clustering routing protocols in wireless sensor network, *Sensors*, vol.12, no.8, pp.11113-11153, 2012.

[8] S. A. Pert, H. Bagci and A. Ya, MOFCA: Multi-objective fuzzy clustering algorithm for wireless sensor networks, *Applied Soft Computing*, vol.50, pp.151-165, 2015.

[9] W. R. Heinzelman, A. P. Chandrakasan and H. Balakrishnan, An application-specific protocol architecture for wireless micro sensor networks, *IEEE Transactions on Wireless Communications*, vol.1, no.4, pp.660-670, 2002.

[10] S. Misra and R. Kumar, An analytical study of LEACH and PEGASIS protocol in wireless sensor network, *Proc. of International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2017.

[11] D. C. Hoang, R. Kumar and S. K. Panda, Realization of a cluster-based protocol using fuzzy C-means algorithm for wireless sensor networks, *IET Wireless Sensor Systems*, vol.3, no.3, pp.163-171, 2013.

[12] D. C. Hoang, R. Kumar and S. K. Panda, Fuzzy C-Means clustering protocol for wireless sensor networks, *IEEE International Symposium on Industrial Electronics*, pp.3476-3482, 2010.

[13] H. A. Arnaldo and B. R. C. Bedregal, A new way to obtain the initial centroid clusters in Fuzzy C-Means algorithm, *The 2nd Workshop-School on Theoretical Computer Science*, pp.139-144, 2013.

[14] J. Zhang and L. Shen, An improved fuzzy c-means clustering algorithm based on shadowed sets and PSO, *Computational Intelligence and Neuroscience*, vol.2014, pp.1-10, 2014.

[15] A. Naik, S. C. Satapathy and K. Parvathi, Improvement of initial cluster center of c-means using teaching learning based optimization, *The 2nd International Conference on Communication*, pp.428-435, 2012.

[16] M. Tong and M. Tang, LEACH-B: An improved LEACH protocol for wireless sensor network, *Proc. of the 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, pp.1-4, 2010.
[17] J. Kamimura, N. Wakamiya and M. Murata, A distributed clustering method for energy-efficient data gathering in sensor networks, *International Journal of Wireless and Mobile Computing*, vol.1, no.2, pp.113-120, 2006.
[18] S. Al-Sodairia and R. Ounia, Reliable and energy-efficient multi-hop LEACH-based clustering protocol for wireless sensor networks, *Sustainable Computing: Informatics and Systems*, vol.20, pp.1-13, 2018.
[19] S. Lindsey and C. S. Raghavendra, PEGASIS: Power-efficient gathering in sensor information systems, *Proc. of IEEE Aerospace Conference*, pp.1125-1129, 2002.
[20] M. A. Sayeed and R. Shree, Optimizing unmanned aerial vehicle assisted data collection in cluster based wireless sensor network, *ICIC Express Letters*, vol.13, no.5, pp.367-374, 2019.
[21] A. Uddin, A. Mansour, D. L. Jeune, M. Ayaz and E. Aggoune, UAV-assisted dynamic clustering of wireless sensor networks for crop health monitoring, *Sensors*, vol.18, no.2, pp.1-24, 2018.
[22] R. M. Bani-Hani and A. A. Ijjeh, A survey on LEACH-based energy aware protocols for wireless sensor networks, *Journal of Communications*, vol.8, no.3, pp.192-206, 2013.
[23] M. Arghavani, M. Esmaeili, M. Esmaeili, F. Mohseni and A. Arghavani, Optimal energy aware clustering in circular wireless sensor networks, *Ad Hoc Networks*, vol.65, no.C, pp.91-98, 2017.
[24] H. Jing and H. Aida, Cooperative clustering algorithms for wireless sensor networks, in *Smart Wireless Sensor Networks*, Y. K. Tan (ed.), IntechOpen, 2010.
[25] Y. Zhou, N. Wang and W. Xiang, Clustering hierarchy protocol in wireless sensor networks using an improved PSO algorithm, *IEEE Access*, vol.5, pp.2241-2253, 2016.
[26] N. Ramluckun and V. Bassoo, Energy-efficient chain-cluster based intelligent routing technique for wireless sensor networks, *Applied Computing and Informatics*, vol.2018, pp.1-13, 2018.