



# Hierarchical clustering and routing protocol to ensure scalability and reliability in large-scale wireless sensor networks

Harmanpreet Singh<sup>1</sup> · Damanpreet Singh<sup>1</sup>

Accepted: 3 February 2021 / Published online: 26 February 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

## Abstract

Cluster-based routing protocols have been proven efficient in prolonging the life cycle of wireless sensor networks (WSNs). Periodic and multi-hop clustering are the most popular techniques which provide the required energy-efficient communication and scalability in large-scale WSNs. In clustering, WSN is divided into number of clusters, and cluster head is selected in each cluster. However, in the existing clustering protocols, CH's near base station undergoes large number of receiving, aggregating and transmitting operations in comparison with far away CHs. This imbalance of load on CHs and lack of structured multi-level clustering framework leads to early death of WSNs. Moreover, resolving the issues of scalability and data reliability along with load balancing is a very tedious task. In this paper, a hierarchical clustering and routing (HCR) protocol is proposed to formulate a load-balanced approach for clustering while taking care of energy efficiency, reliability and scalability. Firstly, a hierarchical layered framework is created to split the WSN into virtual circular layers for efficient transmission of data in hierarchical fashion. Subsequently, an ant lion optimizer is employed for the selection of CHs to ensure reliable, energy balanced and scalable cluster formation. Simulation results demonstrate that HCR protocol outperforms existing state-of-the-art clustering protocols in terms of network lifetime, balanced clustering, throughput and energy efficiency.

**Keywords** Wireless sensor networks · Energy efficiency · Clustering · Antlion optimizer · Data reliability

---

✉ Harmanpreet Singh  
harmansran@outlook.com

<sup>1</sup> Department of Computer Science and Engineering, SLIET, Sangrur, Punjab 148106, India

## 1 Introduction

Wireless sensor networks (WSNs) comprise of tens to thousands of sensor nodes (SNs), deployed to sense and transmit the collected data to Base Station (BS) [1]. WSNs are being employed in a wide range of practical applications like pollution monitoring, natural disaster detection, smart healthcare monitoring, military surveillance, intrusion detection and target tracking, etc. [2]. Due to limited energy resources and remote deployment of WSNs, it may not be possible to recharge or change the battery of dead SNs from time to time. This limitation has inspired industrialists and researchers to design low-power hardware devices and energy-efficient protocols respectively [3].

Energy-efficient routing protocols provide the productive utilization of battery power required to extend the lifetime of WSNs. As compared to non-clustering protocols, clustering-based routing (CBR) protocols have proved to efficiently improve the network lifetime by mitigating the energy consumed in collisions, over-hearing and idle listening [4]. In clustering, SNs are grouped into clusters and one node from each cluster acts as cluster head (CH). A time slot for data transmission is assigned to each SN by their respective CHs. CHs collect the sensed data from SNs, aggregate the collected data and transmit the aggregated packets to BS directly or via another CHs. Network lifetime in CBR protocols is divided into periodic rounds, where each round comprises of two stages: setup stage and steady-state stage [5]. In setup stage, clustering is performed to select the appropriate CHs for current round, and in steady-state stage, transmission of sensed data takes place. Periodic re-clustering after each round is performed to rotate the role of CH for even distribution of load among SNs.

Huge difference in the energy consumption of CHs and SNs has driven the researchers to avoid premature death of WSN by equally distributing the energy load in CBR protocols. Low Energy Adaptive Clustering Hierarchy (LEACH) protocol was the first attempt to address this problem. In LEACH, a probabilistic function is employed to select CHs. SN once selected for CH role cannot become CH again for next  $k$  rounds, where  $k$  is the optimal number of required CHs [6]. Some more stochastic approaches like LEACH have been proposed in the literature [7–9], which do not consider residual energy as a parameter to select the CH. Despite reducing the control overhead to save energy, these protocols suffer from issues like scalability and premature death of WSN due to energy unaware selection of CHs. Event-driven-based data transmission protocols like [10, 11] aim to further reduce the overhead by only sending the sensed data when sudden change in environment occurs. But these protocols are restricted to specific application environment like intrusion detection or volcano monitoring.

In contrast to above techniques, authors of [12–14] have considered energy left in each SN to compete in the CH selection process. SNs compete with their neighbor nodes to elect themselves as CH. Competition is performed based on higher residual energy to select only those SNs as CH which are capable enough to run throughout the current round. In addition to residual energy, distance between CH and BS is minimized in [15] to reduce the transmission cost of CHs,

whereas [16] has selected CHs based on higher node degree. These competition-based approaches are simpler to implement but are not appropriate for large-scale WSNs due to their high message passing complexity.

It is proved in the literature that CH selection is a non-deterministic polynomial hard (NP-hard) problem because there are  $k^N C$  possible combinations to select  $k$  CHs from  $N$  SNs [17]. Researchers have explored Evolutionary Optimization Techniques (EOT) like ant colony optimization (ACO) [18], biogeography-based optimization (BBO) [19], particle swarm optimization (PSO) [20], differential evolution (DE) [21] and genetic algorithms (GA) [22] to solve various NP-hard problems including CH selection problem. EOT aims to find an optimal set of CHs by minimizing or maximizing the objectives defined in the fitness function. Intra-cluster distance is minimized in [6] and [23] using simulated annealing algorithm and PSO, respectively. Main goal for minimizing the intra-cluster distance is to reduce the energy consumed in transmissions between SNs and their CH. In addition to intra-cluster distance, [24] has also minimized the CH to BS distance using GA to further reduce the overall communication cost. All these protocols extend the network lifetime by minimizing the energy consumed during data transmission. However, they have not considered the balancing of energy load among CHs.

Some studies [25, 26] have tried to stabilize the energy load of clusters by generating the clusters with similar intra-cluster distances. The motive is to equalize the consumption of energy while transmitting the data from SNs to their CHs. However, minimizing the variation in the intra-cluster distances may lead to formation of clusters with varying node degree. Subsequently, energy of CHs with high node degree will deplete fast, resulting in the premature death of WSN.

In one of the recent works, PSO-based clustering protocol is proposed [27] considering intra-cluster distance, node degree and residual energy for the selection of CHs. The protocol minimizes the orphan nodes that are not connected to any CH for energy-efficient communication. Some authors [28–30] performed the selection of CH based on CH to BS distance, residual energy of SNs and intra-cluster distance using chemical reaction optimization (CRO) [31], BBO and PSO, respectively.

After thorough review of the existing literature, three main research gaps have been identified. *Firstly*, there is a lack of scalability approach for large-scale WSNs, which also focuses on load balancing. Although, multi-hop routing [27–30] and unequal clustering protocols [32–36] provide scalability, but due to uneven formation of clusters [37] and imbalanced load on CHs, respectively [38], they suffer from hot spot problem. Some works [25, 26] have tried to equalize the intra-cluster distances of all clusters for balancing the energy consumption of CHs. However, considering node degree along with intra-cluster distance for balancing the energy load of CHs and SNs will result in better utilization of limited energy resources. *Secondly*, reliable delivery of data while ensuring scalability results into contradicting objectives. Due to which, one of these issues has always been left behind while designing the routing protocols for WSNs. *Finally*, two separate problems are defined for clustering and routing in existing CBR protocols, which increases the computational complexity and latency in cluster formation.

Specifically, in this paper, hierarchical clustering and routing (HCR) protocol is proposed to enhance the network lifetime of large-scale WSN by creating balanced

clusters. To reduce the computational complexity and control overhead, a hierarchical layered framework (HLF) is designed to provide the joint solution for clustering and routing. To achieve the objectives of HCR protocol, a multi-objective fitness function is derived according to the constraints of HLF. Following are the major contributions of this paper:

- **Hierarchical clustering:** HLF is designed which divides the large-scale WSN into circular layers based on the number of SNs and CHs, required at each layer. The motive is to distribute the energy load evenly among hierarchical layers and to perform joint clustering and routing.
- **Balanced inter-cluster and intra-cluster routing load:** Equal number of CHs from succeeding layer are assigned to each CH of their preceding layer for balancing the inter-cluster routing load, and for balancing the intra-cluster load, node degree of CHs is equalized along with intra-cluster distance.
- **Scalability at network level:** HCR protocol utilizes HLF to estimate the number of layers and SNs required at each layer to make sure that the network is fully connected. Further, inter-cluster distance is maximized between the clusters of each layer to cover all the SNs.
- **Energy efficiency:** Energy consumed in data transmissions from SNs to CHs and from CHs to BS is conserved by minimizing the inter-cluster and intra-cluster routing distances.
- **Optimization of CH selection process:** A novel EOT namely ant lion optimizer (ALO) is applied to optimize the selection of CHs.

The rest of the paper is organized as follows: The network model and energy model are described in Sect. 2. Section 3 describes the proposed methodology of HCR protocol. Section 4 demonstrates simulation results of HCR protocol in comparison with BERA [29], PSO-ECHS [30] and PSO-C [23]. Finally, Sect. 5 concludes the paper.

## 2 System model

### 2.1 Network model

Consider a WSN having area  $a * a$  square units and  $N$  number of SNs are deployed randomly over some geographic location for the realization of HCR protocol. BS is located in the middle of WSN and has unrestricted computational ability, storage and battery power. Further, all the SNs have similar storage, transceiver and battery power. It is assumed that BS knows the location of all SNs, which can be obtained from localization techniques or received signal strength indicator value. The communication between SNs and their CHs is performed in a round-robin scheduling methodology, and inter-cluster communication is done using CSMA/CA technique to avoid any packet collisions. Various notations used in the paper are presented in Table 1.

**Table 1** Network parameters

Notation	Meaning
$a * a$	WSN having length 'a' units
$N$	SNs deployed in the WSN
$K$	Total CHs in WSN
$BS$	Base station node
$S$	Set representing SNs
$S_{ch}$	Set representing CHs
$D_{s_i}^{ch_j}$	Distance between $j$ th CH and $i$ th SN
$D_{ch_i}^{ch_{i+1}}$	Distance between CHs of $i$ th layer and ( $i + 1$ )th layer
$D_{ch_i}^{S_{BS}}$	Distance between BS node and $j$ th CH
$E_{ch_i}$	Energy of $i$ th CH
$E_{s_i}$	Energy of $i$ th SN

### 2.2 Energy model

First-order radio model [6] is considered in this paper to measure the energy consumption of SNs. This model considers the energy dissipated in aggregation, transmission and reception of data packets. Energy consumed in transmissions ( $E_{trans}$ ) and receptions ( $E_{recv}$ ) of  $k$  bits over distance  $X$  is shown in Eqs. 1 and 2 as follows:

$$E_{trans} = \begin{cases} k * E_{elec} + k * \epsilon_{fs} * X^2 & \text{if } X > d_o \\ k * E_{elec} + k * \epsilon_{mp} * X^4 & \text{if } X \leq d_o \end{cases} \tag{1}$$

$$E_{recv} = k * E_{elec} \tag{2}$$

where  $E_{elec}$  is energy consumed by electrical circuit,  $\epsilon_{fs}$  is energy dissipated in free space communication and  $\epsilon_{mp}$  is energy dissipated in multi-path communication. Thus, energy consumed in each round by CHs of inner layers and outermost layer is subsequently evaluated using Eqs. 3 and 4, respectively, as follows:

$$E_{CH} = m * k * E_{elec} + k * E_{elec} + k * \epsilon_{fs} * X^2 + 2 * k * E_{elec} \tag{3}$$

$$E_{CH} = m * k * E_{elec} + k * E_{elec} + k * \epsilon_{fs} * X^2 \tag{4}$$

### 2.3 Hierarchical layered framework

HLF is designed in this paper to partition the WSN into circular layers considering number of SNs and CHs required at each layer. Although few layering-based frameworks are present in the literature, they all are based on uniform node density or

predetermined node deployment strategy. However, in random deployment of SNs, determining the position of layers is a very tedious process.

It is proved in [41] that after level 3 hierarchy, the reduction in energy consumption is negligible. So, following function is derived to divide the WSN upto level 3 hierarchy based on the number of SNs:

$$l = \begin{cases} 1, & \text{if } N_{X>d_o} < N/3 \\ 2, & \text{if } N/3 < N_{X>d_o} < 2 * N/3 \\ 3, & \text{if } N_{X>d_o} > 2 * N/3 \end{cases} \quad (5)$$

where  $N_{X>d_o}$  represents the SNs with distance greater than  $d_o$  and  $l$  is number of hierarchical levels. Concentric layers are formed such that immediate succeeding layer will have two times the SNs as compared to preceding layer. So, number of SNs in each layer for 2-level and 3-level hierarchy can be evaluated using Eq. 6 and Eq. 7, respectively, as:

$$N^j = \begin{cases} N/3 & \text{for } j = 1 \\ (2 * N)/3 & \text{for } j = 2 \end{cases} \quad (6)$$

$$N^j = \begin{cases} N/7 & \text{for } j = 1 \\ 2 * N/7 & \text{for } j = 2 \\ 4 * N/7 & \text{for } j = 3 \end{cases} \quad (7)$$

where  $N^j$  is the number of SNs in  $j$ th layer. Similarly, number of CHs in immediate succeeding layers will be two times more than preceding layer for balancing the energy load throughout the WSN. Figure 1 shows the division of WSN into  $h$  layers using HLF.

## 2.4 Problem formulation

### 2.4.1 Load-balanced cluster formation

Since, routing load of all CHs for inter-cluster data transmission is already balanced by HLF, intra-cluster routing load is balanced in this objective. Intra-cluster load depends on two factors; intra-cluster distance and node degree of CHs. To equalize the load of all clusters, variations in these two factors should be minimized as follows:

$$\text{minimize } f_1 = \sqrt{\frac{\sum_{i=1}^K (CD_i - \overline{CD})^2}{K-1}} + \sqrt{\frac{\sum_{i=1}^K (ND_i - \overline{ND})^2}{K-1}} \quad (8)$$

where  $CD_i$  and  $ND_i$  are the intra-cluster distance and node degree of  $i$ th CH, respectively.

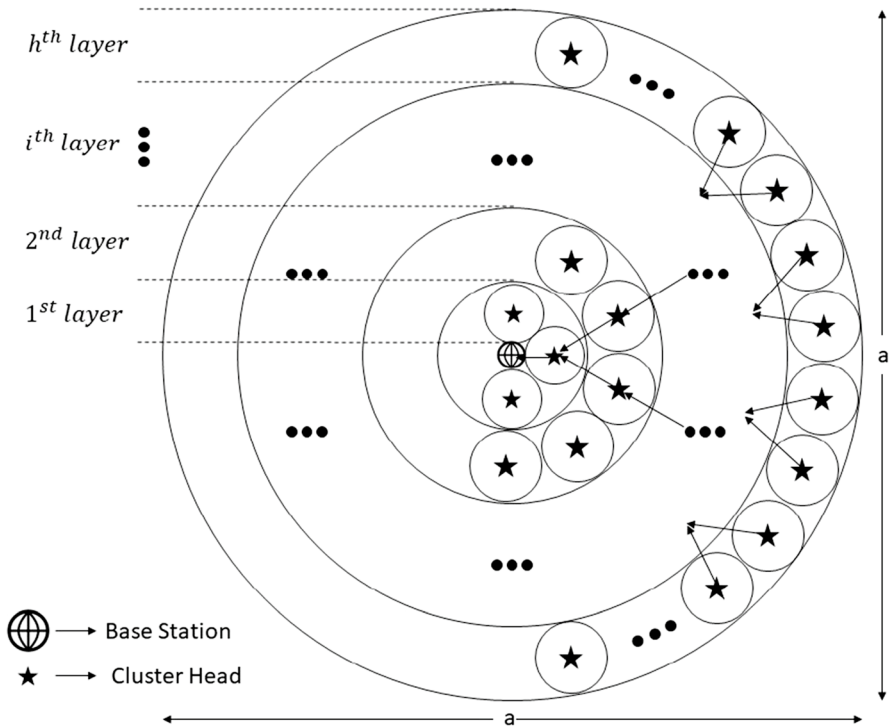


Fig. 1 Hierarchical layered framework

### 2.4.2 Scalable and distributed clusters

Scalability is the ability of an algorithm to provide the same performance regardless of any change in the size or node density of WSN. Since, HLF adjusts itself according to the number of SNs and distance from BS, WSN will always be connected in a multi-level fashion. However, following objective is formed to disperse the CHs within a layer by maximizing the inter-cluster distance of same layer:

$$\text{maximize } f'_2 = \sum_{i=1}^h \sum_{j=1}^{K_i} D_{CH_i^i}^{CH_i^{i+1}} \tag{9}$$

where  $D_{CH_i^i}^{CH_i^{i+1}}$  represents the distance between two adjacent CHs of same level. This objective will force the CHs of same layer to form a shape regular polygon, hence covering the network effectively. Since  $f'_2$  is a maximization objective, it is converted into minimization objective as follows:

$$\text{minimize } f_2 = \frac{1}{f'_2} \tag{10}$$

### 2.4.3 Energy-efficient communication

For energy-efficient data transmission, intra-cluster routing path, inter-cluster routing path and distance between CHs of first layer and BS are minimized as follows:

$$\text{minimize } f_3 = \sum_{i=1}^h \sum_{j=1}^{K_i} D_{CH_{i-1}}^{CH_i^j} + \sum_{l=1}^{K_1} D_{BS}^{CH_l^1} + \sum_{i=1}^K \sum_{j=1}^{n^i} D_{S_j}^{CH_i} \quad (11)$$

where  $D_{CH_{i-1}}^{CH_i^j}$  represents the distance between linked CHs of adjacent layers,  $D_{BS}^{CH_l^1}$  is the distance of BS and  $l$ th CH of innermost layer and  $D_{S_j}^{CH_i}$  is the distance of  $j$ th MN from  $i$ th CH and  $n^i$  is the number of MNs in  $i$ th cluster.

### 2.4.4 Data delivery reliability

Reliability in data delivery means that the routing path is strong enough to handle the transmission from source to destination. In proposed protocol, data delivery reliability is ensured by minimizing the weak links as follows:

$$\text{minimize } f_4 = \sum_{i=1}^N \begin{cases} 1 & \text{if } D_{S_i}^{CH} > AD \\ 0 & \text{if } D_{S_i}^{CH} < AD \end{cases} \quad (12)$$

where  $AD$  is average distance of all the SNs from their CHs. If the distance between any SN and its CH is greater than average distance, it is considered as a weak link. Objective function  $f_4$  gives the count of weak links in the current scenario, and this objective needs to be minimized to ensure data reliability as high as possible.

### 2.4.5 Normalized fitness function

Min-max normalization is applied on all the objectives as they have different ranges. An overall LP minimization fitness function is formulated as follows:

$$\text{minimize } f = \beta_1 * f_1 + \beta_2 * f_2 + \beta_3 * f_3 + \beta_4 * f_4 \quad (13)$$

where  $\beta_1, \beta_2, \beta_3$  and  $\beta_4$  are the weightage parameters such that  $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$ .

## 2.5 Ant lion optimizer

Ant lion optimizer (ALO) is a recently proposed nature-inspired evolutionary optimization technique based on the hunting behavior of antlion. Antlion has a unique way of hunting prey by creating circular cone-shaped pits in sand. Then, antlions hid themselves at the center bottom of the pit and wait for the prey (usually ants) to fall in the pit. ALO has a good balance between exploration and exploitation phases. Population in ALO comprises of two sets, one each for ants and antlions. Each ant and antlion



in ALO technique is considered as a possible solution of CHs positions. Elite antlion represents the optimal positions of CHs w.r.t. objectives defined in the fitness function. Ants rotate around antlion and elite antlion in search of better solution. Positions of antlions and elite antlion are updated accordingly when a better solution is found. Mathematical modeling of ALO for CH selection problem is done as follows:

**2.5.1 Initialization and parameter settings**

Antlions and ants are randomly initialized in the search area of  $a * a$  sq. meter. Each antlion ( $A$ ) or ant ( $T$ ) represents a complete solution of CH positions. Population of  $m$  antlions for selecting  $K$  number of CHs is represented as follows:

$$\text{Population} = \begin{bmatrix} A_{1,1}, A_{1,2}, \dots, A_{1,j} \dots, A_{1,K} \\ A_{2,1}, A_{2,2}, \dots, A_{2,j} \dots, A_{2,K} \\ \dots \dots \\ \dots \dots \\ A_{m,1}, A_{m,2}, \dots, A_{m,j} \dots, A_{m,K} \end{bmatrix}_{m * K} \tag{14}$$

where  $A_{i,j}$  represents  $j$ th candidate CH node having two dimensions for its x and y coordinates. Similarly, population of  $m$  ants is created randomly. Lower bound and upper bound are set according to the network area as 0 and  $a$ , respectively.

**2.5.2 Evaluate fitness and select elite antlion**

After initialization, fitness value of all the antlions and ants is calculated using fitness function defined in Sect. 2.4. Antlion with best fitness is selected as elite antlion ( $A_{elite}$ ). Since, fitness function for selection of CHs is derived as a minimization problem as shown in Eq. 13, antlion having least fitness cost is considered as an elite antlion.

**2.5.3 Random walk of ants**

An antlion is selected for each ant using roulette wheel selection method based on the fitness of antlion. Movement of ants is monitored by both the selected antlion and elite antlion. Random walk of  $i$ th ant can be formulated as follows:

$$R_i(\text{ant}) = \frac{R_i^A + R_i^E}{2} \tag{15}$$

where  $R_i^A$  is the random walk of ant around antlion selected using roulette wheel and  $R_i^E$  is the walk of ant around elite antlion. Random walk of an ant is given as:

$$R_{\text{ant}} = [cs(2r(t_1) - 1), s(2r(t_2) - 1), s(2r(t_{\text{max}}) - 1)] \tag{16}$$

where  $cs$  represent cumulative sum of the uniformly distributed random numbers. Accordingly,  $r(t_i)$  is calculated based on the random numbers generated as follows:

$$r(t) = \begin{cases} 1 & \text{if rand} > 0.5 \\ 0 & \text{if rand} < 0.5 \end{cases} \quad (17)$$

To mimic the behavior of ants falling in pits, lower and upper bounds of ant movement boundary are decreased based on iterations. The reduction in search space for an ant represents the exploitation behavior of the algorithm.

#### 2.5.4 Catching ants and rebuilding traps

After the random movement of ants, new position of antlion is selected based on the current position of ant revolving around it. Position of antlion is updated to the position of ant when fitness value of ant becomes better than the current fitness of antlion. Similarly, position of an elite antlion is updated when fitness of any antlion becomes better than fitness of elite antlion as follows:

$$A_i^{t+1} = \text{ant}_i^t \text{ if } f(\text{ant}_i^t) > f(A_i^t) \quad (18)$$

$$A_{\text{elite}}^{t+1} = A_i^{t+1} \text{ if } f(A_i^{t+1}) > f(A_{\text{elite}}^t) \quad (19)$$

where  $A_i^{t+1}$  is the new position of  $i_{th}$  antlion for  $(t+1)_{th}$  iteration,  $A_{\text{elite}}^{t+1}$  is the new position of elite antlion for  $(t+1)_{th}$  iteration,  $\text{ant}_i^t$  is the current position of ant revolving around  $i_{th}$  antlion in  $t_{th}$  iteration and  $f()$  represents the fitness value.

The process of ants random walk and position updating of antlions continues until optimal solution is found or maximum number of iterations are reached. It may be possible that there are no SNs placed at the optimal positions generated at the end of ALO-based CH selection. So, the SNs nearest to the chosen positions are selected for the role of CH.

### 3 Hierarchical clustering and routing (HCR) protocol

HCR protocol presents a joint solution for multi-hop routing and multi-level clustering. Firstly, hierarchical layered framework (HLF) is designed in HCR protocol for partitioning of the WSN into hierarchically aligned circular layers as shown in Fig. 1. Basic design goal of HLF is to balance the CHs routing load at each layer. Then, an ALO-based CH selection algorithm is run to find the most favorable solution of CHs for the current round such that scalable, energy-efficient, reliable and load-balanced clusters are formed. Layers are formed by HLF in such a way that successive layer has double the SNs than preceding layer. To balance the CHs load throughout the WSN, number of CHs in successive layer is also kept double than preceding layer.

In HCR protocol, transmitting, receiving and aggregating operations of CHs are distributed equally for even energy consumption of CHs at each layer. Basically, a CH undergoes three kinds of data transmission and receiving operation in HCR protocol; 1) MNs transmit the sensed data to their assigned CH, 2) CHs of preceding

layer receive data packets from the CHs of succeeding layer and 3) CHs transmit the aggregated packet to the CH of preceding layer. Only the CHs of outermost layer do not undergo receiving operation from outer CHs. Since CHs have to perform various operations, its energy will be depleted fast. Hence, a round-based policy is followed in this paper to rotate the job of CH among SNs of same hierarchical layer. Each round consists of three phases: information gathering phase, cluster formation phase and data transmission phase. Working flow of HCR protocol is shown in Fig. 1 and various phases followed in HCR protocol are discussed as follows:

### 3.1 Information gathering phase

Information gathering phase is run at the beginning of each round. In information gathering phase, BS broadcasts a message requesting the residual energy information from all SNs. Location information is not requested from SNs as it is assumed that BS knows the locations of all SNs, which can be obtained from RSSI values or localization techniques. Based on number of SNs and their location, BS utilizes HLF to partition the WSN into virtual circular layers aligned in hierarchical fashion.

### 3.2 Cluster formation phase

In cluster formation phase, BS runs an ALO-based algorithm for the selection of CHs to select the appropriate CHs at each layer. Four objectives have been derived for the appropriate selection of CHs, one each for load-balanced cluster formation, scalable and distributed clusters, energy-efficient communication, and data delivery reliability as explained in Sect. 2.4. After the selection of CHs using above fitness function, a role message is broadcasted by BS containing the role of all SNs as either CH or MN. SNs on receiving this message will change their status to either CH node or MN. MNs then send a join request message to get time slot for data transmission from their respective CH to avoid any intra-cluster collisions. On receiving the allotted time, SNs go into sleep state until their turn comes up to send data.

### 3.3 Data transmission phase

In data transmission phase, CHs of outermost layer collect data from the SNs of their cluster known as intra-cluster data collection, aggregate the collected data into one packet and send the aggregated packet to closest CH located in the adjacent inner layer. CHs of inner layers receive data packets from both their MNs and outer layer CHs, aggregate it and transmit it to their adjacent inner layer until the data are received at BS. Routing load of CHs is equally divided throughout the WSN such that each pair of CHs will send data packet to one CH of their adjacent inner layer. CHs use CDMA technique for data transmission to avoid inter-cluster collisions. The detailed flowchart of different phases followed in HCR protocol is shown in Fig. 2.

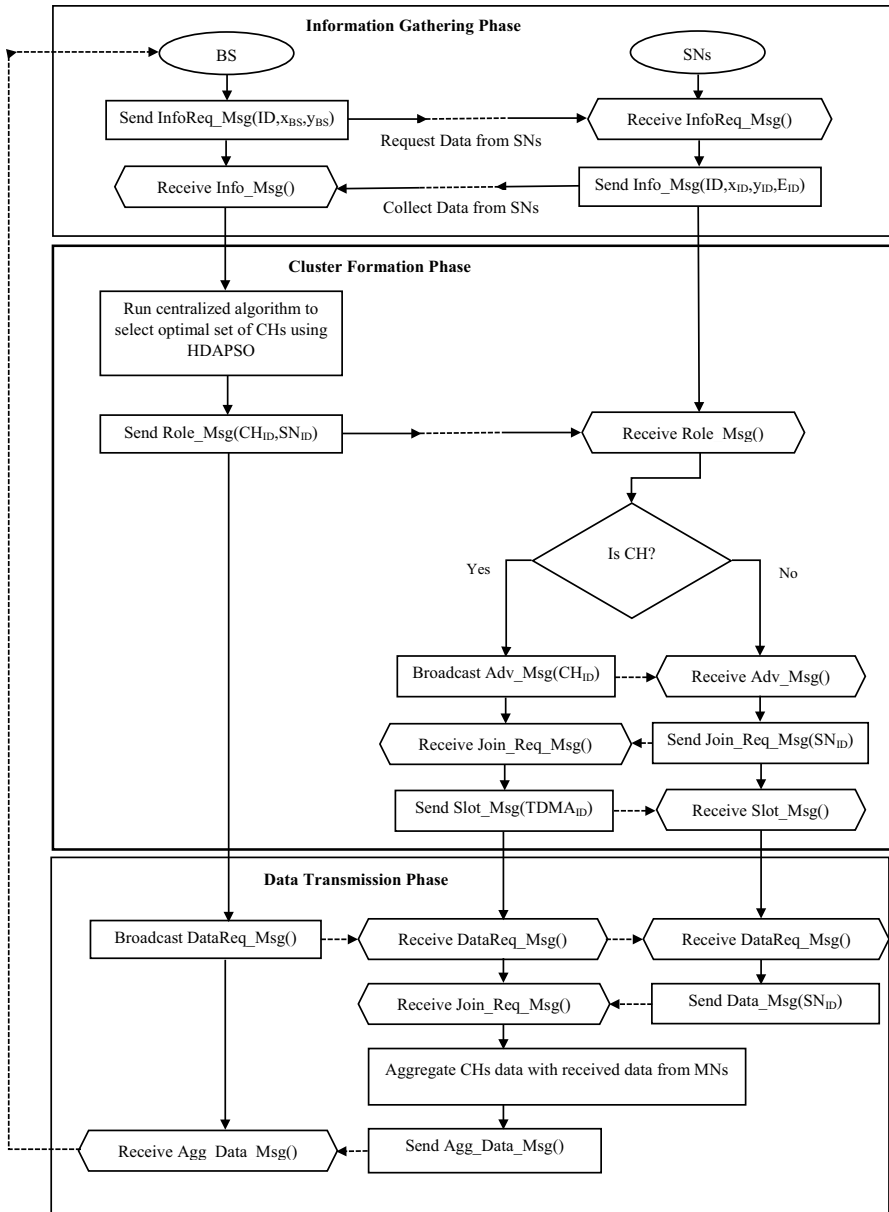


Fig. 2 Flowchart of working phases in proposed HCR protocol

### 4 Performance evaluation

In this section, performance of proposed HCR protocol is evaluated and compared with popular clustering protocols namely PSO-ECHS, PSO-C and BERA.

Experimentation is performed for WSNs deployed over large geographic area to test the robustness of HCR protocol in handling large-scale WSNs. Energy consumption, network lifetime, balanced clustering and throughput are the performance metrics used in this paper to validate the simulation results.

#### 4.1 Simulation parameters

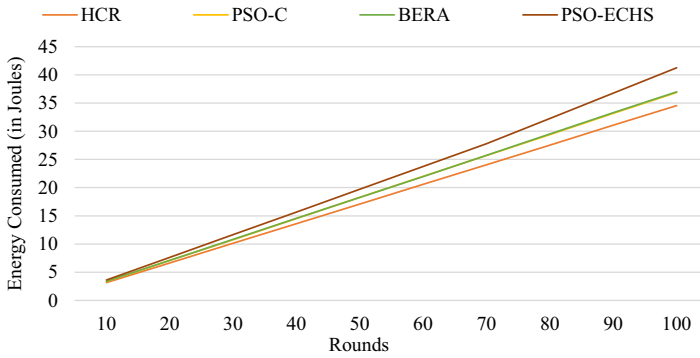
The simulation of the HCR protocol and other competent protocols is performed on MATLAB under diverse network conditions. 500 SNs are randomly deployed over the WSN area, and BS is deployed at center of the WSN. Three different network sizes (WSN1 for  $300 \times 300 \text{ m}^2$ , WSN2 for  $500 \times 500 \text{ m}^2$  and WSN3 for  $700 \times 700 \text{ m}^2$ ) and five random network topologies for each size are considered for simulation. Average of all network topologies for a particular network size is taken for performance comparison. Data packet and control data packet size are set as 6400 and 200 bits, respectively. Energy model and parameters are taken same as in competent protocols. ALO is run for 100 iterations with population size of 30 to find the best possible CH positions.

#### 4.2 Energy consumption comparison

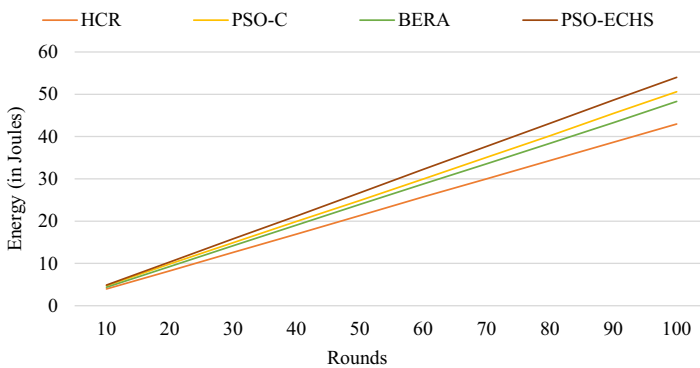
Simulation results of energy consumed by HCR protocol and its competent protocols are plotted in Fig. 3a–c. Each protocol is run for five different network topologies and average energy consumed in first 1000 rounds is considered for comparison. Further, results are compared for three different network sizes to evaluate the effect of node density on CH selection and energy consumption. Figure 3a–c demonstrates the superiority of HCR protocol in comparison with competent protocols for conserving the energy irrespective of the network sizes. BERA has shown close performance to HCR protocol for dense networks but lacks far behind in sparse networks. The major reason for the conservation of energy in HCR protocol is the adoption of layering-based framework for balanced and structured clustering.

#### 4.3 Network lifetime comparison

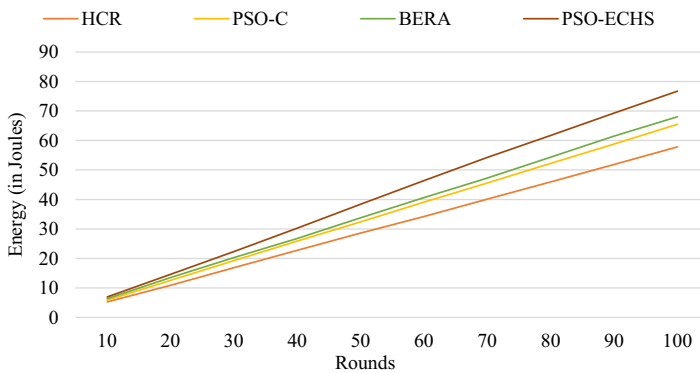
Network lifetime is time period for which the network can perform at its full potential [39]. In this paper, first node die (FND) is considered as a metric to evaluate the lifetime of network. Figure 4 illustrates the simulation result for different network areas and it is perceived that HCR has much better lifetime as compared to competent protocols. While the difference in network lifetime of HCR and BERA is not significant in sparse WSNs but HCR has performed for much longer in dense networks as shown in Fig. 4. This is due to the balanced load of CHs in HCR protocol, which results in longer functioning of the WSN. PSO-C and BERA have not employed any objective for balanced energy consumption due to which they have short life span. PSO-ECHS on the other hand has shown worst network lifetime because it employs a parametric function for cluster formation which results in the formation of unequal and inefficient clusters.



(a) Comparison based on energy consumption for WSN#1(300\*300 m<sup>2</sup>)



(b) Comparison based on energy consumption for WSN#2(500\*500 m<sup>2</sup>)



(c) Comparison based on energy consumption for WSN#3(700\*700 m<sup>2</sup>)

**Fig. 3** a Comparison based on energy consumption for WSN#1(300\*300 m<sup>2</sup>) b Comparison based on energy consumption for WSN#2(500\*500 m<sup>2</sup>) c Comparison based on energy consumption for WSN#3(700\*700 m<sup>2</sup>)

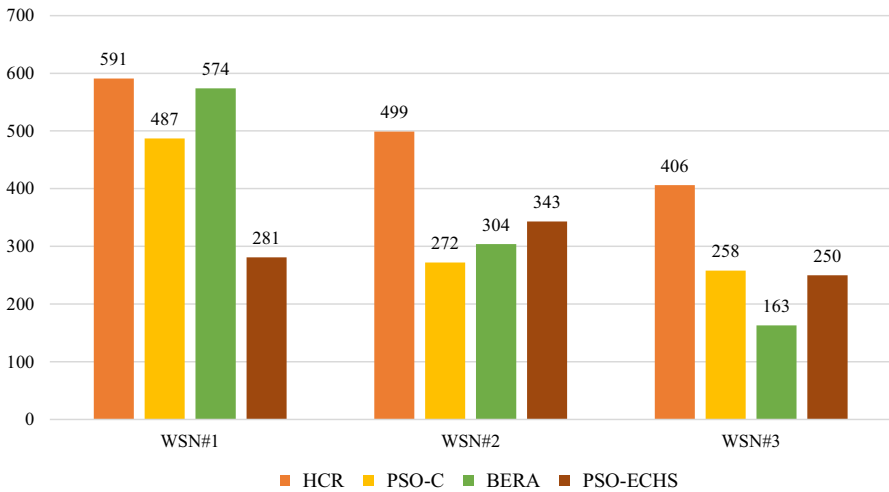


Fig. 4 Network lifetime comparison

### 4.4 Balanced clustering comparison

In this paper, focus is on balancing the CHs load in terms of node degree and data transmission distance. CHs with same node degree will dissipate equal amount of energy in receiving operations, and clusters will consume same amount of energy in transmission operations if intra-cluster distance of all clusters is similar. Figures 5 and 6 illustrate the comparison of various protocols based on variations in the node degree and intra-cluster distance, respectively. HCR protocol has shown least variations for both the parameters which strengthen its ability to create

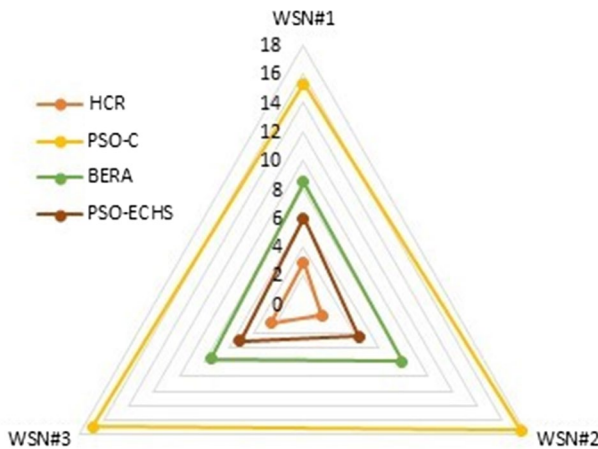


Fig. 5 Comparison of variations (standard deviation) in node degree of CHs

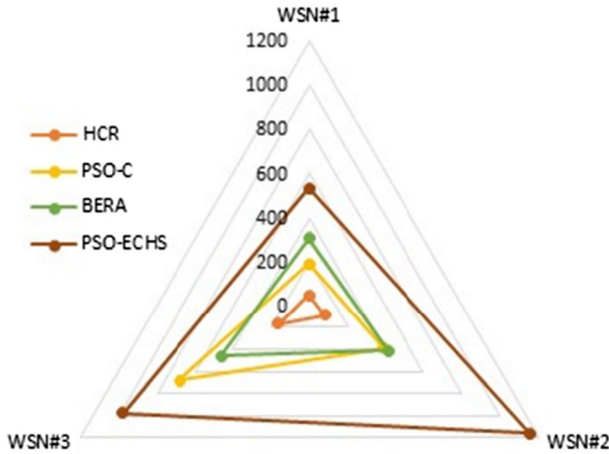


Fig. 6 Comparison of variations (standard deviation) in intra-cluster distance

balanced clusters. This is due to the novel fitness function used in HCR protocol for CH selection and cluster formation.

### 4.5 Throughput comparison

Throughput in WSNs is measured in terms of total raw packets generated in its lifetime [26]. Figure 7 shows the throughput of all the protocols for different network sizes. HCR protocol has highest throughput as compared to its competent protocols and maintains its integrity regardless the size of WSN. Throughput is directly

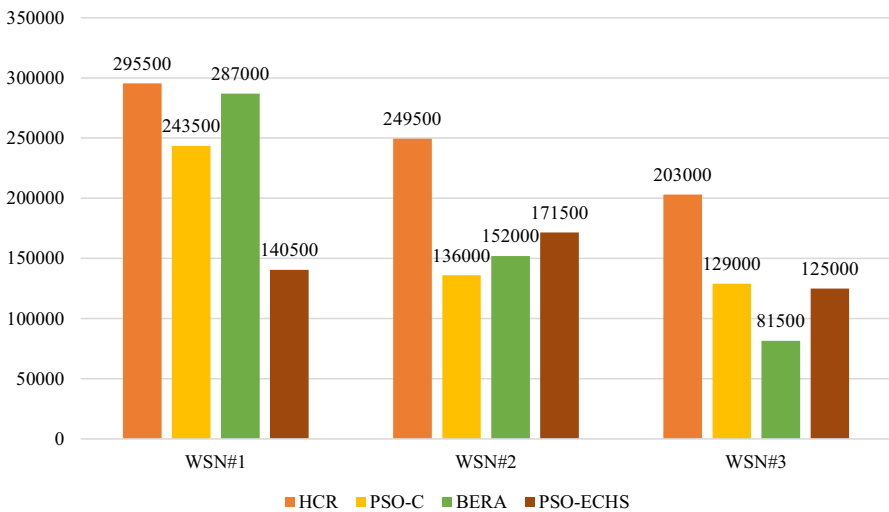


Fig. 7 Throughput comparison



**Table 2** Evolutionary optimization techniques comparison

Algorithm	Convergence rate comparison		Best fitness cost comparison	
	Mean	SD	Mean	SD
ALO	81.25	24.71	1.8164	0.0152
BBO	315.67	92.54	1.8551	0.0363
DE	188.36	75.66	2.2972	0.0348
PSO	126.58	56.67	1.9048	0.0559
ABC	377.45	108.65	1.8649	0.0207

proportional to the lifetime of WSN, and same behavior is observed in Fig. 7. It is due to the combination of HLF and defined fitness function, that HCR protocol has shown superior performance in diverse network conditions.

#### 4.6 Convergence comparison

ALO is deployed in this study due to its high convergence rate and less parametric tuning as compared to other EOT. Further to test the robustness of ALO for achieving the stable optimal solution in comparison with PSO, BBO, ABC and DE, 20 independent runs are performed for the objective function defined in HCR protocol. Each run is of 500 iterations and population size is set at 30. Table 2 illustrates the performance comparison of different EOTs in terms of convergence rate and optimal fitness cost. ALO converges toward an optimal solution in minimal number of iterations and has shown stability in convergence rate in successive runs. Also, the fitness cost attained by ALO to obtain an optimal solution is best among other EOT.

## 5 Conclusions

This paper presents a load-balanced hierarchical clustering and routing protocol to provide a load-balanced and energy-efficient communication in large-scale WSNs. The problem of network lifetime optimization is addressed in HCR protocol by balancing the inter-cluster and intra-cluster routing loads of CHs. HLF is designed in HCR to reduce the delay in the formation of clusters and routing paths by providing a joint solution for clustering and routing. HLF divides WSN into concentric virtual layers such that number of SNs and CHs in succeeding layer is two times its preceding layer. Based on the constraints of HLF, a novel fitness function is derived to choose the best favorable set of CHs in each layer such that load-balanced, reliable, energy-efficient and scalable clusters are created. ALO-based CH selection algorithm is run to select CHs based on the derived fitness function. Results obtained from the simulations establish that HCR protocol outperforms other competent protocols in terms of energy efficiency, network lifetime, load balancing, throughput and convergence. The scalable and load-balanced methodology of HCR can be further extended for in.

## References

1. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) A survey on sensor networks. *IEEE Commun Mag* 40:102–105. <https://doi.org/10.1109/MCOM.2002.1024422>
2. Yick J, Mukherjee B, Ghosal D (2008) Wireless sensor network survey. *Comput Netw* 52:2292–2330. <https://doi.org/10.1016/j.comnet.2008.04.002>
3. Wang F, Liu J (2011) Networked wireless sensor data collection: issues, challenges, and approaches. *IEEE Commun Surv Tutor* 13:673–687. <https://doi.org/10.1109/SURV.2011.060710.00066>
4. Abbasi A, Younis M (2007) A survey on clustering algorithms for wireless sensor networks. *Comput Commun* 30:2826–2841. <https://doi.org/10.1016/j.comcom.2007.05.024>
5. Arboleda, L., Nasser, N.: Comparison of clustering algorithms and protocols for wireless sensor networks. In: *Proceedings of Canadian Conference on Electrical and Computer Engineering*. pp. 1787–1792 (2006)
6. Heinzelman WB, Chandrakasan AP, Balakrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans Wirel Commun* 1:660–670. <https://doi.org/10.1109/TWC.2002.804190>
7. Bandyopadhyay, S., Coyle, E.: An energy efficient hierarchical clustering algorithm for wireless sensor networks. In: *Proceedings of the Twenty-Second Annual Joint Conference of the IEEE Computer and Communications INFOCOM*. pp. 1713–1723 (2003)
8. Kumar D, Aseri TC, Patel RB (2009) EEHC: Energy efficient heterogeneous clustered scheme for wireless sensor networks. *Comput Commun* 32:662–667. <https://doi.org/10.1016/j.comcom.2008.11.025>
9. Kang SH, Nguyen T (2012) Distance based thresholds for cluster head selection in wireless sensor networks. *IEEE Commun Lett* 16:1396–1399. <https://doi.org/10.1109/LCOMM.2012.073112.120450>
10. Manjeshwar, A., Agrawal, D.: TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. In: *Proceedings of International Parallel and Distributed Processing Symposium*. p. 30189a (2001)
11. Manjeshwar A., Agrawal, D.P., Manjeshwar, A.: APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. In: *Proceedings of International Parallel and Distributed Processing Symposium*. pp. 195–202 (2002)
12. Cheng B-C, Yeh H-H, Hsu P-H (2011) Schedulability analysis for hard network lifetime wireless sensor networks with high energy first clustering. *IEEE Trans Reliab* 60:675–688. <https://doi.org/10.1109/TR.2011.2135650>
13. Taheri H, Neamatollahi P, Younis OM, Naghibzadeh S, Yaghmaee MH (2012) An energy-aware distributed clustering protocol in wireless sensor networks using fuzzy logic. *Ad Hoc Netw* 10:1469–1481. <https://doi.org/10.1016/j.adhoc.2012.04.004>
14. Younis O, Fahmy S (2004) HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Trans Mob Comput* 03:366–379. <https://doi.org/10.1109/TMC.2004.41>
15. Chamam A, Pierre S (2010) A distributed energy-efficient clustering protocol for wireless sensor networks. *Comput Electr Eng* 36:303–312. <https://doi.org/10.1016/j.compeleceng.2009.03.008>
16. Ye, M.Y.M., Li, C.L.C., Chen, G.C.G., Wu, J.: EECS: an energy efficient clustering scheme in wireless sensor networks. In: *Proceedings of 24th IEEE International conference on Performance, Computing, and Communications*. pp. 535–540 (2005)
17. Saleem M, Di Caro GA, Farooq M (2011) Swarm intelligence based routing protocol for wireless sensor networks: survey and future directions. *Inf Sci (Ny)* 181:4597–4624. <https://doi.org/10.1016/j.ins.2010.07.005>
18. Dorigo M, Birattari M, Stutzle T (2006) Ant colony optimization. *IEEE Comput Intell Mag* 1:28–39. <https://doi.org/10.1109/MCI.2006.329691>
19. Simon D (2008) Biogeography-based optimization. *IEEE Trans Evol Comput* 12:702–713. <https://doi.org/10.1109/TEVC.2008.919004>
20. Kennedy J (2010) Particle Swarm Optimization. In: Sammut C, Webb GI (eds) *Encyclopedia of machine learning*. Springer, US, Boston, MA, pp 760–766
21. Storn R, Price K (1997) Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim* 11:341–359. <https://doi.org/10.1023/A:1008202821328>
22. Holland JH (1973) Genetic algorithms and the optimal allocation of trials. *SIAM J Comput* 2:88–105. <https://doi.org/10.11393/ncr/i2004-10001-9>

23. Latiff, N.M.A., Tsimenidis, C.C., Sharif, B.S., Kingdom, U.: Energy-Aware Clustering for Wireless Sensor Networks Using Particle Swarm Optimization. In: Proceedings of 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07). pp. 5–9 (2007)
24. Rahmanian, A., Omranpour, H., Akbari, M., Raahemifar, K.: A novel genetic algorithm in LEACH-C routing protocol for sensor networks. In: Proceedings of 24th Canadian Conference on Electrical and Computer Engineering, CCECE. pp. 1096–1100 (2011)
25. Kuila P, Jana PK (2014) A novel differential evolution based clustering algorithm for wireless sensor networks. *Appl Soft Comput* 25:414–425. <https://doi.org/10.1016/j.asoc.2014.08.064>
26. Kuila P, Gupta SK, Jana PK (2013) A novel evolutionary approach for load balanced clustering problem for wireless sensor networks. *Swarm Evol Comput* 12:48–56. <https://doi.org/10.1016/j.swevo.2013.04.002>
27. Elhabyan RSY, Yagoub MCE (2015) Two-tier particle swarm optimization protocol for clustering and routing in wireless sensor network. *J Netw Comput Appl* 52:116–128. <https://doi.org/10.1016/j.jnca.2015.02.004>
28. Srinivasa Rao PC, Banka H (2017) Novel chemical reaction optimization based unequal clustering and routing algorithms for wireless sensor networks. *Wirel Netw* 23:759–778. <https://doi.org/10.1007/s11276-015-1148-0>
29. Lalwani P, Banka H, Kumar C (2018) BERA: A biogeography-based energy saving routing architecture for wireless sensor networks. *Soft Comput* 22:1651–1667. <https://doi.org/10.1007/s00500-016-2429-y>
30. Rao PCS, Jana PK, Banka H (2017) A particle swarm optimization based energy efficient cluster head selection algorithm for wireless sensor networks. *Wirel Netw* 23:2005–2020. <https://doi.org/10.1007/s11276-016-1270-7>
31. Lam AYS, Li VOK (2010) Chemical-reaction-inspired metaheuristic for optimization. *IEEE Trans Evol Comput* 14:381–399. <https://doi.org/10.1109/TEVC.2009.2033580>
32. Sabor N, Abo-Zahhad M, Sasaki S, Ahmed SM (2016) An unequal multi-hop balanced immune clustering protocol for wireless sensor networks. *Appl Soft Comput J* 43:372–389. <https://doi.org/10.1016/j.asoc.2016.02.016>
33. Jiang CJ, Shi WR, Xiang M, Tang XL (2010) Energy-balanced unequal clustering protocol for wireless sensor networks. *J China Univ Posts Telecommun* 17:94–99. [https://doi.org/10.1016/S1005-8885\(09\)60494-5](https://doi.org/10.1016/S1005-8885(09)60494-5)
34. Mohajerani A, Gharavian D (2015) An ant colony optimization based routing algorithm for extending network lifetime in wireless sensor networks. *Wirel Netw* 22:2637–2647. <https://doi.org/10.1007/s11276-015-1061-6>
35. Mao, S., Zhao, C., Zhou, Z., Ye, Y.: An improved fuzzy unequal clustering algorithm for wireless sensor network. In: Proceedings of 6th International ICST Conference on Communications and Networking in China (CHINACOM). pp. 206–214 (2012)
36. Shokouhifar M, Jalali A (2014) A new evolutionary based application specific routing protocol for clustered wireless sensor networks. *AEU - Int J Electron Commun* 69:432–441. <https://doi.org/10.1016/j.aeue.2014.10.023>
37. Sha K, Gehlot J, Greve R (2013) Multipath routing techniques in wireless sensor networks: a survey. *Wirel Pers Commun* 70:807–829. <https://doi.org/10.1007/s11277-012-0723-2>
38. Arjunan S, Pothula S (2016) A survey on unequal clustering protocols in wireless sensor networks. *J King Saud Univ-Comput Inf Sci*. <https://doi.org/10.1016/j.jksuci.2017.03.006>
39. Chen Y, Zhao Q (2005) On the lifetime of wireless sensor networks. *IEEE Commun Lett* 9:976–978. <https://doi.org/10.1109/LCOMM.2005.11010>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.