

# An Energy Efficient Protocol to Mitigate Hot Spot Problem Using Unequal Clustering in WSN

Sunil Kumar Singh<sup>1</sup> · Prabhat Kumar<sup>1</sup> · Jyoti Prakash Singh<sup>1</sup>

Published online: 18 April 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

**Abstract** In multihop scenarios, the sensor nodes nearer to the base station (BS) are overloaded because they handle their own data as well as the information obtained from far away nodes. This induces a higher energy depletion rate in nodes near to the BS causing early death of these nodes resulting in hot spot/energy hole problem in wireless sensor network (WSN). This paper proposes a novel strategy using unequal fixed grid-based cluster along with a mobile data mule for data collection from the cluster head (CH). A CH is selected in such a manner that the cumulative transmission distance for member nodes within the cluster is minimum. The paper has attempted to optimize the values for CH change time or round number ( $f$ ) and also established a relationship between different size clusters by using a factor ( $r$ ), as they are playing an important role in the overall performance improvement of the WSN. Integrating a mobile data mule in the protocol enhances its efficiency of handling hot spot problem and makes it more energy effective. Two different WSN-scenarios have been considered based on the movement pattern of the data mule. The results obtained through simulation in both scenarios prove the success of our scheme in terms of energy efficiency, load balancing and network lifetime as compared to the existing protocols. The paper also providing a balance trade-off between delay and high overheads by using a single mule with simple predefined path. It also minimizes the hot spot problem as it sustains more than 3000 rounds, which is far better than the existing methods.

---

✉ Sunil Kumar Singh  
sunil.cse14@nitp.ac.in

Prabhat Kumar  
prabhat@nitp.ac.in

Jyoti Prakash Singh  
jps@nitp.ac.in

<sup>1</sup> Department of Computer Science and Engineering, National Institute of Technology Patna, Patna, Bihar, India

**Keywords** Wireless sensor network · Cluster head · Grid · Unequal clustering · Data mule · Hot spot

## 1 Introduction

A WSN is a self-organized network which contains a large number of tiny sensor nodes deployed in the sensing area [1]. Sensor nodes sense the environment, process the data and send it to the BS either directly or through intermediate nodes. All the nodes are capable of acting as sensing node as well as relay node. Environment monitoring, healthcare, agriculture, military and smart home are few important applications of WSN. The battery power of the sensor node is a major constraint in these networks. The proper utilization of this power is necessary to prolong network lifetime [20]. Searching of various paths from sensing node to the BS and selection of a path among them for transmitting the sensed data in an efficient manner is collectively known as routing [9]. Various protocols have been designed for this purpose depending on the objectives of the WSN. In [31], these protocols are categorized into four groups, namely (1) communication model, (2) network structure, (3) network topology and (4) reliable routing. Network structure based routing protocols are further sub-categorized into two classes, namely flat protocols and hierarchical protocols.

Hierarchical routing protocols are designed to build network energy efficient that can improve the overall network lifetime. In these protocols, deployed nodes are split into groups termed as clusters. A node within the cluster works as the CH and the remaining as cluster members (CMs). CMs communicate with the CH and CH further interacts with the BS using single-hop or multihop transmission. The energy depletion rate of the CH is always high with respect to CMs due to the extra load which result is unequal energy consumption in the network. The CH rotation is a common method deployed to balance this uneven energy depletion. To optimize the load balance in a WSN, various optimization techniques have been used to find the optimal number of clusters and CHs [13, 27, 30]. Earlier, most of the clustering protocols have used single hop communication from CH to the BS, which consumes a lot of energy because the required energy for signal propagation increases proportionally to the square of the propagation distance upto a certain distance after that it will be power of four [7, 33]. Recently, multihop communication gets more attention for CH to the BS communication in WSN due to energy efficiency [16]. In spite of various advantages, the major drawback of a multihop routing protocol is the hot spot problem. In multihop mechanism, the nodes closer to the BS are prone to dissipate their energy rapidly as compared to the other nodes due to forwarding of more packets. Hence, these nodes will generally die early and network may get isolated. This is known as hot spot problem [5]. The approaches like unequal clustering techniques [12] and mobile data mule or mobile sinks [26] are mainly used to resolve this issue. However, most of the unequal clustering schemes proposed for the solution of the hot spot uses smaller cluster size near the BS and the size of the cluster increases as we go far from the BS. The size of a cluster is inversely proportional to its distance from the BS. The clusters near the BS accommodate a higher number of nodes which help in effective load sharing. The proposed scheme for size allocation and size variation of clusters results in reduction of the frequency at which a particular node becomes CH. This helps in maintaining the overall connectivity and prevents network isolation. In this manner, the hot spot problem is minimized. One major issue regarding unequal clustering is cluster size increasing or decreasing ratio, which is clearly not discussed in existing schemes. We have resolved this

issue by finding the relationship between various sizes of the clusters, discussed in Sect. 5.2.

This research also endeavours at determining a near-optimal time to change the CH as discussed in Sect. 5.1. Hence, in this paper we have taken a CH change factor  $f$  and fixed it after 7th round as discussed in Sect. 5.1. The path traversed by data mules have been proposed in several research works [2, 4, 8]. However, most of them make it necessary for the mule to visit every node resulting in unwanted delay. Successive research works introduced the concept of using multiple mules for data collection to decrease this delay. However, this increases the overheads. Our proposed trajectory for the mule attempts at providing a balance in this trade-off between delay and high overheads.

Data mules [34] are dedicated mobile nodes deployed in the network to support the BS in data collection. They primarily operate as a carrier in between sensor nodes and the static BS and they have no resource issues. Data mule collects data from individual sensor node or CHs or relay nodes, when it comes within their transmission range. It collects data in two different contact schemes: mule motion contact and sojourn point (SP) contact. In mule motion contact, a mule is always moving and periodically transmits beacon messages when it detects any node, contact has been made and it accumulates data from the node. In SP contact, mule stays at a predetermined location for a fixed amount of time called sojourn time and in that duration mule collects data [17]. When it reaches near the BS it transfers all its gathered data to the BS. We have considered SP of contact in this work for collecting data from CH.

The early death of sensor nodes near the BS owing to the hot spot and unfair load distribution is a serious issue in static WSN. And then the hot spot problem during the multihop routing needs to be discussed. An energy efficient multihop routing for uniform energy sharing is needed to increase the lifetime of WSN. We are concerned with resolving the hot spot problem in WSN. This can be done by minimizing the communication overheads of the BS's nearby nodes. In this scheme, we propose an energy efficient protocol, which provides the solution for hot spot problem and maximizes the lifetime of the network. The scheme comprises of unequal rectangular clusters with a data mule, which moves with a constant speed for collecting data from CHs. The operating cost of cluster formation, CH selection and its rotation is minimized as our protocol is centralized in nature and these roles are performed by the BS. The BS uses fixed clustering approach and partition the network into rectangular grids of varying size called clusters. The size of the cluster increases while moving towards the SP. The BS handles the movement path and time schedule of the mobile data mule. The data mule moves in a pre-defined path and halts for a predefined time at SP to collect data. It transfers the collected data to the BS when reaches in the BS range. The unequal clustering and use of data mule in our protocol makes it more energy effective and results in the mitigation of the hot spot problem.

The remaining part of the paper is organized in the following manner: Sect. 2 summarizes the current related literatures. The proposed scheme is explained in detail with system model in Sect. 3. Section 4 gives an overview of the simulation environment and Sect. 5 discussed the evaluation of the proposed work with the help of simulation results. The overall discussion about the contribution and findings is presented in Sect. 6 Finally, Sect. 7 conclude the paper with some future works.

## 2 Related Works

The paper considers unequal grid based clustering with a mobile data mule. Hence in the related works first we will discuss few basic clustering techniques and after that unequal clustering will be talked about. Further, we will explore research works related to grid-based clustering as well as data mule based protocols.

Hienzelman et al. [15] introduced first cluster based routing protocol LEACH which is distributed in nature. Each node generates a number between 0 and 1 on a random basis at the beginning of each round. The nodes, whose generated number is less than a threshold value, declare themselves as CH for the round. Nodes send join request to their nearest CH based on the received signal strength (RSS). CH sends an acknowledgment message and TDMA time slot for data transmission in reply. The major drawbacks of the protocol are unequal size of clusters in different rounds, no consideration of energy level of nodes in the CH selection process and single-hop transmission between the CH and the BS. They extended their previous works and proposed LEACH-Centralized (LEACH-C) [14] protocol to overcome the limitations of LEACH. The number of CHs for each round is fixed in LEACH-C. The protocol reduces the overhead of CH selection from the nodes as it is centralized in nature. It offers better performance than LEACH but also suffers from problems such as single-hop transmission and centralized CH selection, which is not good for the large area network. A lot of successors of LEACH have been proposed for performance improvement till date [36]. Bandyopadhyay and Coyle [6] proposed Energy-Efficient Hierarchical Clustering (EEHC) algorithm. It is a layer based clustering protocol in which network is segmented into a hierarchy of layers. The data transmission goes from lower layer towards the BS. The CHs at the lowest layer transmit data to the apparent upper layer CHs. This process persists till data is received by the BS. Hybrid Energy-Efficient Distributed Clustering (HEED) [39] is a distributed protocol which uses intra-cluster communication cost and residual energy of the node as parameter for CH selection. In this protocol, CH selection takes place after a defined number of rounds instead of each round. Jin et al. [19] enhanced the EEHC protocol and introduced a new protocol Energy-Efficient Multilevel Clustering (EEMC). This protocol reduces energy consumption and latency of the network by selecting CH based on the residual energy of the node and the distance from the BS. The main demerit of this protocol is that the BS requires information of all the nodes for the selection of CH in each round. All the nodes send information to the BS and receive the reply message. This increases the message overhead and energy depletion of the network. MR-LEACH [10] is circular layer based protocol in which the BS is positioned at the center. Each lower layer CH sends data to its apparent upper layer nearest CH for communicating to the BS. Most of the multihop routing protocols discussed above suffer from the hot spot problem. Goa et al. [11] have proposed a fuzzy multiple criteria decision making (MCDM) CH selection scheme to increase the lifetime of the network. In this scheme, CH selection is optimized by the use of hierarchical fuzzy integral and trapezoidal fuzzy AHP with the help of energy status, QoS impact and location as the main factors. This protocol improves the lifetime of the network but its perform poor for large area network due to single hop communication. Mantri et al. [28] proposed a Two Tier Cluster based Data Aggregation Algorithm (TTCDA) to reduce the overall communication cost by minimizing the number of transmission from sensor nodes to the BS. They have used additive and divisible data aggregation functions at the CH and number of packets

generated by a node with variable rate, are maintained by spatial and temporal correlation. This protocol performs better in terms of bandwidth utilization and throughput but it suffers from hot spot problem due to two tier cluster communication. This work has been extended by using a mobile BS and heterogeneous nodes and they named it mobile sink and heterogeneous nodes aware cluster-based data aggregation algorithm (MHCDA) [29]. It minimizes the communication and computation cost by aggregating the packets at a predefined region. The main drawback of this protocol is hot spot problem which affects the network lifetime.

PRODUCE [21] is an unequal clustering based protocol which is semi-centralized in nature. It provides better performance in small networks but it suffers from scalability problem. It balances the intra-cluster and inter-cluster data load by varying the size of the cluster. The transmission distance restricted up to the threshold transmission distance of the radio energy model. A location-based unequal clustering algorithm (LUCA) [23] proposed by Lee et al. reduces the overall energy consumption in the network. The main idea in this scheme is to keep the larger cluster size farther from the BS and small size clusters nearer to the BS. This location based cluster formation minimizes the intra and inter cluster communication but due to the unequal cluster its size is not optimal which results in a high energy consumption. Low Power Grid based Cluster Routing Algorithm (LPGCRA) [24] is a grid based routing protocol which addresses hot spot problem. It selects the maximum residual energy node within the cluster as the CH without considering its distance from the remaining nodes of the cluster. This outcomes in more energy consumption in intra-cluster communication as the overall transmission distance of the cluster is not downplayed. It also uses single-hop transmission for communicating with the BS. Grid Based Routing (GBR) [3] uses data delivery time as a parameter for the CH selection. The nodes nearest to the BS in each cluster are selected as CH. This increases the overall intra-cluster transmission distance which results in shorter network lifetime. Another drawback of this protocol is that it segments the network into fixed number of clusters without considering the size of the network. This leads to the higher rate of energy depletion in larger network as the transmission distance exceeds from a certain limit. In [18], Authors have proposed a grid based protocol GFTCRA to reduce hot spot problem. It uses overall transmission distance within the cluster as a parameter for CH selected. It also limits the number of CHs for which a CH acts as a relay node in order to balance uneven load of the clusters nearer to the BS.

Shah et al. [22] have proposed three tier architecture for sparse sensor network which is based on functionality of data mule. Data mules are mobile devices which go in the network for the accumulation of data from the sensor nodes. Authors in [41] have presented a message ferrying scheme in order to increase the network lifetime and reduce hot spot problem. It specifies a limited route of the data mule for collecting information from the static nodes. In [37], Authors have discussed the vital topics of data mule such as speed control at the time of communication with the sensor nodes, the velocity with which it prompts in the network. In order to minimize data latency time, Luo et al. [25] have introduced an algorithm which utilizes multiple data mules for data collection that reduces the total tour time of each mule. Prince et al. [32] have proposed an unequal clustering based routing protocol using data mule in order to reduce hot spot problem. The protocol specifies a true way for data mule traversal which reduces the complexity of incorporation of the data mule in the routing protocols. The protocol balances the intra-cluster data and inter-cluster data by varying the cluster size. Singh et al. [35] have presented an odd-even level routing in a square grid based clustering with two mobile mules. Authors have divided the whole sensing area

into levels and further levels are divided into square grids/clusters. Established along the round number, the mule visits all the levels vertically and collects data from the borderline CHs by stopping at the anchor points and eventually transfer the collected data to the BS. If the round number is even, CHs transmit their data from left to right direction to the CHs in their level and vice versa in case of odd round number. The odd-even level routing reduces the hot spot problem by not forwarding data vertically towards the BS which enhanced the lifetime of the nodes nearer to the BS. The main drawback of this approach is scalability due to the centralized nature.

Nevertheless, most of the above protocols have not discussed about the energy efficiency, hot spot problem and load balancing issues together. In our strategy, we have looked at all these matters and likewise attempted to settle them. A comparative analysis of discussed protocols in this section has been presented in Table 1.

The proposed scheme has the following advantages over the existing similar type protocols.

1. It considers an unequal clustering where the size of the cluster linearly increases as it goes towards the SP.
2. It has ability to handle the hot spot problem.
3. The distribution of the CH is better.
4. The CH does not rotate in every round. Hence, a significant amount of energy saves due to minimum overhead.
5. It proposed a relationship between different cluster sizes with the help of a factor ( $r$ ) and also finds the round number after which CH will change ( $f$ ).
6. The mobile data mule collects data from CHs and delivers it to the BS.

### 3 Proposed Work

This section, describes our proposed scheme which ensures the equal energy distribution among sensor nodes in the network and extends the network lifetime. The proposed work is divided into four phases- (1) Cluster Formation, (2) Cluster head selection, (3) Route Discovery and (4) Data Forwarding. All four levels are centrally managed by the BS. Each sensor node has a synchronized clock to know the start time of each round. In the cluster formation phase, the sensing area is divided into unequal rectangular grids termed as cluster in this article. The clusterhead selection phase is primarily responsible for selection of CH. The route discovery phase constructs the multihop routing path from CH to the SP. In data forwarding phase, the TDMA slots are allocated to each node for communication.

#### 3.1 Network Model

The network contains a BS, a data mule and  $n$  number of sensor nodes, which are scattered in the environment in a random manner. Each node has a unique id denoted by  $S_i$ . We have made some assumptions regarding the network model which are presented in this section. The assumptions are quite realistic and can be easily achieved in a deployment.

- The sensor nodes are homogenous and static.
- The BS is fixed and positioned outside the network.
- The BS has information about all sensor nodes regarding their node ids and locations.

**Table 1** Comparative analysis of related work protocols

Protocols	Year	Major contribution	Advantages	Drawbacks
LEACH [15]	2000	CH selection and rotation is distributed using a random number	Less communication overheads, reduces correlated data and each sensor node get equal chance to become the CH	Random selection of CH, position and number of CHs are not optimum, single hop communication from CH to the BS
LEACH-C [14]	2002	centrally, the creates the optimal number of clusters using a simulated annealing algorithm and optimal number of CHs calculated based on the average energy of the network	Less overheads due to centralized nature, optimal numbers of cluster and CHs created and minimum energy dissipation of individual nodes	For location GPS requires which is costly and consumes extra energy, It is also less scalable
EEHC [6]	2003	Multi-level clustering with multihop communication by CHs	Optimal number of CHs are selected, suitable for large sensor network	Hot spot problem near the BS, random selection of CHs
HEED [39]	2004	A distributed energy efficient hybrid clustering approach, where CHs are probabilistically selected based on their residual energy and cluster members are joins on the basis of minimum communication cost	Optimize resource usage, maximize network lifetime	Hot spot problem for more than two-level of hierarchy
EEMC [19]	2008	Optimal number of CHs with multi-level clustering minimizes the energy consumption and delay.	Achieve minimum energy dissipation and delay with a efficient sleep schedule	Uniform distribution and collision
PRODUCE [21]	2008	Distributed and unequal clustering	Reduces hot spot problem	Scalability
MR-LEACH [10]	2010	Entire sensing area is divided into various layers and further divides in equal clusters	Multihop layerwise communication and highly scalability are the main feature	Hot spot is the main issue
LPGCRA [24]	2010	Grid based fixed clustering where CH are selected dynamically	Efficient load balance and lesser hot spot.	Finding relay nodes creates overheads
MCDM [11]	2012	Optimal CH selection using fuzzy multiple criteria decision making and first declaration wins rule	Better in terms of network lifetime	Scalability is the major problem
GR [3]	2013	CH selected based on the least transmission time and highest residual energy and data transmitted to the BS via a virtual chain formed by other CHs	Efficient routing path	Hot spot problem

**Table 1** continued

Protocols	Year	Major contribution	Advantages	Drawbacks
TTCDA [28]	2014	Additive and divisible data aggregation functions at the CH.	Less energy consumption and better bandwidth utilization	Hot spot problem
GFTCRA [18]	2015	A Distributed grid based clustering and routing to solve hot spot problem	Energy efficient, low hot spot and fault tolerance	More delay and complex
MHCDA [29]	2016	Data aggregation at a predefined region in a heterogeneous network with a mobile sink	Less communication and computation cost and better bandwidth utilization	Hot spot problem is a major concern

- The data mule moves in a predefined path with constant velocity and stop at SPs.

The different notations used in this article are tabulated in Table 2.

### 3.2 Cluster Formation Phase

The main objective of this phase is to partition the target area into rectangular grids or clusters. The BS performs the grid (cluster) formation only once after deployment. The grid is fixed for the lifespan of the network. Each grid holds a unique id with fixed number of nodes. Algorithm 1 outlines the working of this phase.

The BS divides the deployment area into  $m$  number of horizontal lanes of equal width ( $w$ ) called levels. Each level is then subdivided into blocks giving the whole network, a grid like structure. The largest cluster which is nearer to the SP is  $w \times w$  meters. To balance the load of CH, each node is selected as CH atleast once during its life cycle. The CHs are selected from each cluster. It may happen that the CHs of two clusters are lying at diagonally opposite ends in the worst case. For this type of situation, we can fix the maximum length and width of a cluster by formulating Eqs. 1, 2 and 3 as shown in Fig. 1.

$$\sqrt{w^2 + (2w)^2} = Z \tag{1}$$

where  $2w \leq w_{next} + w$ , therefore

$$\sqrt{w^2 + (2w)^2} \leq Z \tag{2}$$

$$w = \frac{Z}{\sqrt{5}} \tag{3}$$

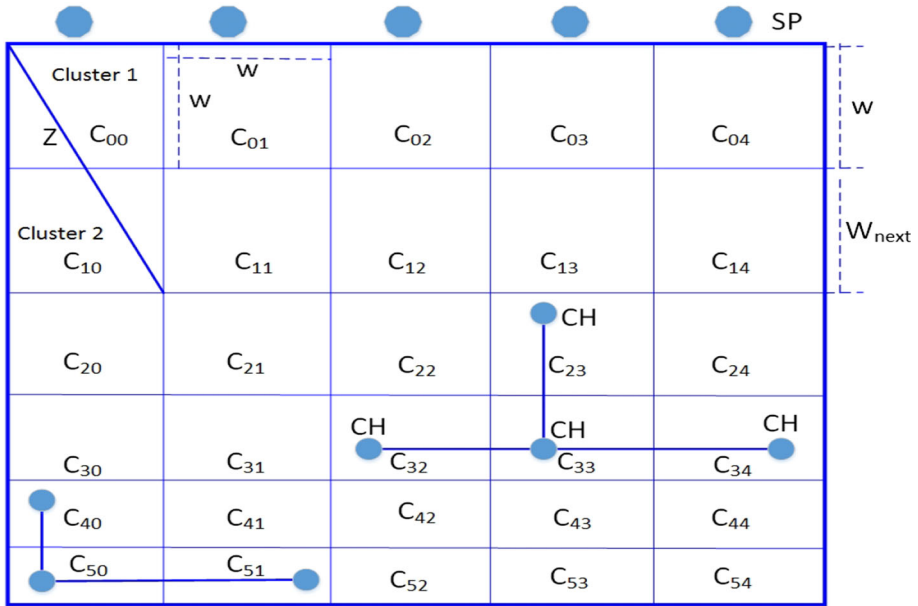
The largest cluster is constructed near to the SP. The width of the cluster decreases with a square exponent formula build with a constant factor  $r$  as we move away from the SPs. The parameter  $r = 0.5$  is evaluated and found 0.5 to be near-optimal for our simulation as discussed in Sect. 5. The bigger size of the cluster near to the SP can accommodate more number of sensor nodes, which decreases the frequency of a particular node to be selected as CH in multiple rounds. This distributes the traffic load and prolongs the lifetime of the nearer nodes and hence hot spot problem does not arise in the network. The difference in widths of any two clusters belonging to consecutive cluster levels has been expressed as



**Table 2** Notations used in the protocol

Symbols	Descriptions
$X$ and $Y$	Dimension of deployment area
$Z$	Threshold transmission range of radio energy model
$n$	Number of sensor nodes deployed in the sensing region
$w$	Length of the each cluster
$m$	Number of levels
$r$	Cluster width decreasing factor of grid
$f$	Round numbers for CH change
$S_i$	Sensor id numbered from 0 to $(n - 1)$
$level_{id}$	Level id numbered from 0 to $(m - 1)$
$C_{level_{id},i}$	Cluster id of the cluster that belongs to $level_{id}$
$CH_{level_{id},i}$	Cluster head of the cluster $C_{level_{id},i}$
$C_{level_{id},i} [ ]$	Array that stores all the node ids belonging to cluster $C_{level_{id},i}$
$numSC_{level_{id},i}$	Number of sensor nodes belonging to cluster $C_{level_{id},i}$
$Var_{CL}$	Number of clusters in a level
$E_{th,level_{id},i}$	Minimum energy for a node of cluster $C_{level_{id},i}$ to become CH
$nodeSet [ ]$	Array that stores ids of the node having $E > E_{avg}$
$dist [ ]$	Array that stores the distance of node from the centroid
$E_{min}$	Minimum energy required for an active node
$msg_{length}$	Length(in bits) of sensed data transmitted by each node
$E_{recv}(msg_{length})$	Energy required to receive $msg_{length}$ bit of data
$E_{trans}(msg_{length}, dist_{trans})$	Energy required to transmit $msg_{length}$ bit of data over a distance $dist_{trans}$
$E_{ampl}$	Amplifier energy of the sensor node for one bit of data
$E_{elect}$	Electronic energy of the sensor node for one bit of data
$e_{freespc}$	Amplifier constant for free space model
$e_{mpfading}$	Amplifier constant for multi-path fading model
$SP_{level_{id}}$	Sojourn point belonging to $level_{id}$

shown in Eq. 4. Based on this relation, the dimension of the sensing area is fixed where, the length is twice of the width. For example: suppose the length and width of the network is 500 and 250 m respectively. Therefore, based on the Eq. CSize, the total number of levels in  $y$  axis is 8 and the cluster width decreasing sequence is 40, 39.5, 38, 35.5, 32, 27.5, 22 and 15.5 where, 40 and 15.5 are the maximum and minimum width of a cluster respectively. In  $x$  axis approximately 12 levels will formed, since cluster length is uniform and size of the cluster length is  $\frac{Z}{\sqrt{5}}$  which is about 40. Hence, the total number of clusters in this network will be  $12 \times 8 = 96$ . In this relationship, we have used square power function



**Fig. 1** Grid formation with maximum  $w$  width to restrict the communication in free space radio model

which falls the value of width by 1 linearly except the first level clusters. We have also looked into this relationship with without power and power three by mathematical computation as well as simulation. In without power case, the difference between two clusters or levels is only 0.5 and with this difference width size is linearly decreasing. Simulation results are also not that much effective, due to this we have not used without power relationship. In front of power three, the deviation between two clusters is very high which is not suited for our system.

$$w_{next} = w - ((level_{id})^2) * r \tag{4}$$

The BS maintains a record of each cluster after complete execution of aforementioned phase. This record contains cluster id, number of sensor nodes in the cluster and its constituent sensor nodes id, location and energy level.

**Algorithm 1** Cluster Formation Algorithm

---

```

1: procedure CLUSTER-FORMATION
Require: Sensing area dimension:(X,Y), Transmission range:Z, Number of deployed
      nodes:n
Ensure: Rectangular shaped clusters with their ids and member nodes
2:    $y=0, w=\frac{Z}{\sqrt{3}}$  (Maximum Width),  $x=\text{minimum width}$ 
3:    $m=\lceil \frac{Y}{w} \rceil$ 
4:   for  $m=0$  to  $(m-1)$  do
5:     Assign  $level_{id}=m$  to the horizontal lane between  $y$  and  $(y+w)$ 
6:      $y=y+w$ 
7:      $i=0, Var_{CL}=0$ 
8:     while  $w_{next} \geq x$  do
9:        $w_{next} = w - ((Var_{CL})^2) * r$ 
10:      Assign  $C_{level_{id},i}$  to the rectangular area between  $w_{next}$  and  $w$ 
11:      Initialize  $C_{level_{id},i}[ctr]$ ,  $ctr=0$  and  $numSC_{level_{id},i}=0$ 
12:      for  $j=0$  to  $(n-1)$  do
13:        if  $S_j$  lie in the cluster  $C_{level_{id},i}$  then
14:          Store  $S_j$  to  $C_{level_{id},i}[ctr]$ 
15:           $ctr++$ 
16:           $numSC_{level_{id},i}=ctr$ 
17:         $i++$ 
18:         $Var_{CL} = i$ 
19:      Total number of clusters =  $Var_{CL} * m$ 

```

---

**3.3 Cluster head (CH) Selection**

Selection of the most suitable node as a CH in each cluster and finding the round in which CH will change are the key functionalities of this phase. These algorithms are executed by BS at the start of each round of operation. The working of this phase is split into two sections: CH change and CH selection. The CH change algorithm determines whether the CH of a cluster will change or not? If yes, the CH selection algorithm is triggered to select new CH for that cluster. Initially, all nodes are having equal energy, hence the BS selects the node as a CH which is nearest to the centroid of the deployed sensor nodes. Thereafter the CH change and CH selections are accomplished by Algorithms 2 and 3 respectively. The BS checks the status of each node by comparing its energy level to  $E_{min}$  given in Eq. 5. If the energy level of a node is less than  $E_{min}$ , it is considered as dead or inactive node. The threshold energy  $E_{th}$  for each cluster is calculated in each round. The  $E_{th}$  energy is the minimum required energy for a node of a cluster, which is going to participate in the CH selection process. Different clusters have different  $E_{th}$  due to the variation of nodes and data packets. The  $E_{th}$  of a cluster for a particular round depends on the size of its own cluster data as well as data forwarded to it by the neighboring clusters. The value of  $E_{th}$  depends on the minimum energy required to receive and transmit data as presented in Eq. 8. The minimal energy needed for receiving data  $E_{recv}$  of a cluster can be derived by estimating the minimal energy required to receive information from its own cluster as well as neighboring cluster as mentioned in Eq. 6. The minimum energy required to transmit in a cluster is based on the distance between the communicating nodes and it can be given by Eq. 7.

**Algorithm 2** Cluster Head Change Algorithm

```

1: procedure CH-CHANGE
Require:  $C_{level_{id},i}$ ,  $CH_{level_{id},i}$ 
Ensure: Determines whether  $CH_{level_{id},i}$  of a cluster  $C_{level_{id},i}$  will change or not
2:   for each  $C_{level_{id},i}$  (where  $level_{id}=0$  to  $(m-1)$ ,  $i=0$  to  $(Var_{CL}-1)$ ) do
3:     for  $i=0$  to  $(numSC_{level_{id},i-1})$  do
4:       if  $E_{resi} \{C_{level_{id},i}[i]\} \leq E_{min}$  then
5:         Update status of node  $C_{level_{id},i}[i]$  as dead
6:         Remove node  $C_{level_{id},i}[i]$  from  $C_{level_{id},i} [ ]$ 
7:          $numSC_{level_{id},i} - -$ 
8:     for  $level_{id}=0$  to  $(m-1)$  do
9:        $numNodes=0$ 
10:      for  $i=0$  to  $(Var_{CL}-1)$  do
11:         $E_{recv}(nbrClst-data) = numNodes * msglength * E_{elect}$ 
12:         $E_{recv}(ownClst-data) = (numSC_{level_{id},i-1}) * msglength * E_{elect}$ 
13:         $numNodes = numNodes + numSC_{level_{id},i}$ 
14:         $E_{trans} = numNodes * msglength * (E_{elect} + e_{freespc} * Z^2)$ 
15:         $E_{th\ level_{id},i} = E_{recv}(nbrClst-data) + E_{recv}(ownClst-data) + E_{trans}$ 
16:      Initialize factor
17:      for  $level_{id}=0$  to  $(m-1)$  do
18:        for  $i=0$  to  $(Var_{CL}-1)$  do
19:          if  $(Energy-of-CH_{level_{id},i} \geq E_{th\ level_{id},i})$  OR  $(Round-number \leq f)$  then
20:            No-Change-in-CH
21:          else
22:            goto Algorithm 3 for CH selection
23:          Change status of selected node as CH and previous CH as CM

```

$$E_{min} = msglength * E_{elect} + msglength * e_{freespc} * Z^2 \tag{5}$$

$$E_{recv} = (data_{ownClst} + data_{nbrClst}) * E_{elect} \tag{6}$$

$$E_{trans} = (data_{ownClst} + data_{nbrClst}) * E_{elect} + (data_{ownClst} + data_{nbrClst}) * e_{freespc} * Z^2 \tag{7}$$

$$E_{th} = E_{recv} + E_{trans} \tag{8}$$

The current CH status depends on the factor  $f$  and  $E_{th}$ . Where,  $f$  is the optimal round number after which CH will change and goes for new CH selection. If either  $(E(CH_{(level_{id},i)}) < E_{th}(level_{id},i))$  or  $(Round\ number \leq f)$  as mentioned in line number 19 in the Algorithm 2 and explains in details in Sect. 5, then CH will change else the current CH remains unchanged. If current CH status has changed then the new CH selection process starts by executing Algorithm 3 for that cluster. The BS updates the status of newly selected CH and previous CH. Algorithm 3 outlines the working of the CH selection process. In this algorithm, the BS first calculates the centroid of the deployed nodes in each cluster by averaging  $x$  co-ordinates and  $y$  co-ordinates of the participating members of the cluster as presented in line number 8 in Algorithm 2. It also calculates the average energy of the active member nodes within the cluster. Line numbers 11–21 of Algorithm 3 describe the complete process for a node to become CH. The set of nodes having higher energy than the average energy as well as  $E_{th}$  of the cluster are selected as initial CH. The node nearest to the centroid is selected as final CH of that cluster. This minimizes the intra-

cluster communication cost due to the minimum transmission distance within the cluster which minimizes the overall energy consumption.

---

**Algorithm 3** Cluster Head Selection Algorithm

---

```

1: procedure CH-SELECTION
Require:  $C_{level_{id},i}$ ,  $CH_{level_{id},i}$ 
Ensure: Determines whether  $CH_{level_{id},i}$  of a cluster  $C_{level_{id},i}$  will change or not
2:   for each  $C_{level_{id},i}$  (where  $level_{id}=0$  to  $(m-1)$ ,  $i=0$  to  $(Var_{CL-1})$ ) do
3:      $E=0$ ,  $x=0$ ,  $y=0$ 
4:     for  $i=0$  to  $(numSC_{level_{id},i}-1)$  do
5:        $x=x+x_{ordinate-of-node}C_{level_{id},i}[i]$ 
6:        $y=y+y_{ordinate-of-node}C_{level_{id},i}[i]$ 
7:        $E=E+E_{energy-of-node}C_{level_{id},i}[i]$ 
8:       Centroid $_{level_{id},i}(x,y) = (\frac{x}{numSC_{level_{id},i}}, \frac{y}{numSC_{level_{id},i}})$ 
9:        $E_{avg} = \frac{E}{numSC_{level_{id},i}}$ 
10:      Initialize nodeSet[ ], count=0
11:      for  $i=0$  to  $(numSC_{level_{id},i}-1)$  do
12:        if Energy-of- $C_{level_{id},i}[i] \geq E_{avg}$  then
13:          nodeSet[count]= $C_{level_{id},i}[i]$ 
14:          count++
15:      for  $i=0$  to  $(count-1)$  do
16:         $dist[i] = \sqrt{(x_{nodeSet[i]} - x_{centroid})^2 + (y_{nodeSet[i]} - y_{centroid})^2}$ 
17:         $dist_{min} = \min_{0 \leq i < count} \{dist[i]\}$  and let  $C_{level_{id},i}[i]$  be the corresponding
node
18:        if  $E\{C_{level_{id},i}[i]\} \geq E_{th}level_{id},i$  then
19:          Select node  $C_{level_{id},i}[i]$  as the  $CH_{level_{id},i}$  of cluster  $C_{level_{id},i}$ 
20:        else
21:           $dist_{min} = next_{dist_{min}}$  and let  $C_{level_{id},i}[i]$  be the corresponding node &
goto 17

```

---

**3.4 Route Discovery Phase**

The main objective of this phase is to discover the data forwarding route of the network. Algorithm 4 explains the complete working procedure of this phase.

---

**Algorithm 4** Route Discovery algorithm

---

```

Require:  $C_{level_{id},i}$ ,  $CH_{level_{id},i}$ 
Ensure: Determines the data forwarding route for each  $CH_{level_{id},i}$ 
1: procedure ROUTE-DISCOVER
2:    $p=level_{id}$ ,  $q=i+1$ , search=0
3:   if  $CH_{(p,q)} = NULL$  &&  $q \geq (Var_{CL-1})$  then
4:     if search == 0 OR 1 then
5:       if search == 2 then
6:          $p=level_{id}$ ,  $q=q+1$ , search=0
7:     else
8:       if  $q \leq (Var_{CL-1})$  then
9:         Create a routing path between  $CH_{level_{id},i}$  and  $CH_{(p,q)}$ 
10:      else
11:        Create a routing path between  $CH_{level_{id},i}$  and  $SP_{level_{id}}$ 

```

---

The BS performs this process in each round after setup phase. It selects  $CH_{(level_{id},i)}$  of the cluster  $C_{(level_{id},i)}$ . It searches the  $CH_{(level_{id},i+1)}$  of the apparent cluster  $C_{(level_{id},i+1)}$  towards the  $SP_{(level_{id})}$ . If  $CH_{(level_{id},i+1)}$  is present, it selects this CH to forward the data of  $CH_{(level_{id},i)}$ . Otherwise, it first searches the apparent upper level  $CH_{(level_{id+1},i+1)}$  and then the apparent lower level  $CH_{(level_{id-1},i+1)}$ . This process is repeated till all the CHs have multiple data forwarding path. The BS maintains a routing path table of the current round after the execution of this phase.

### 3.5 Data Forwarding Phase

The steady phase allocates the TDMA slots to CMs, CHs and data mule for collision free communication during transmission and reception. The BS finds the cluster, which has maximum number of nodes and determines the corresponding number of nodes ( $max\_numSC_{level_{id},i}$ ).  $t_{CM}$  is the maximum time required for communication for transmission and reception between the CM and its CH. Hence, the time assigned to each cluster for intra-cluster communication is equal to  $t_{CM} * (max\_numSC_{level_{id},i} - 1)$ . For instance, if the number of CMs is 10 then 9  $t_{CM}$  time slots are assigned to each cluster. A cluster having 6 CMs utilize 5 time slots and remaining slots are idle for that cluster. After the intra-cluster time allocation, inter-cluster time allocation takes place. If the maximum time taken in transmission and reception of data, including HELLO message exchange between two CHs is  $t_{CH}$ , time assigned to each CH for transmitting the received data is equal to  $t_{CH}$ . The data mule traverses with a velocity  $V$  m/s and waits for time  $t_{wait}$  Sec at each SP to collect the data. The time period of a complete round  $t_{round}$  is calculated based on the above mentioned time units

$$t_{round} = t_{CPkt} + t_{CM} * (max\_numSC_{level_{id},i} - 1) + (t_{CH} * Var_{CL}) + \frac{(Y - m)}{V} + t_{wait} * m \tag{9}$$

Time  $t_{CPkt}$  is allotted to each round as an overhead for the exchange of control packets before the actual data transmission occurs. The time slots  $t_{round}$ ,  $t_{CPkt}$ ,  $t_{CM}$ ,  $t_{CH}$  and  $t_{wait}$  are constant values for the complete network lifetime.

The BS determines the time period of the round by using Eq. 9. It creates a table for each cluster containing the information such as cluster id,  $t_{round}$ , number of nodes within the cluster, node id belonging to the cluster, their corresponding location, their TDMA slots, status (active or inactive), role (CH or CM), relay CH node ids and its location. In the first round BS sends a copy of these tables to the corresponding CHs. The CH stores it and forwards it to each CM. Later on, when  $t_{CPkt}$  time lapses, transmission of sensed data takes place according to the allotted time slots. In the subsequent rounds, nodes wait for a  $t_{CPkt}$  time period for the control packet. If it receives a control packet, it updates its cluster table and data transmission starts based on the updated table. The BS sends control packet to a cluster if its CH changes or any node dies. In case the CH changes, BS sends control packet to newly selected CH which transmits this packet to its CMs. Nodes update the role and time slot of the previous CH and the new CH in the table after receiving the packet. The time slots allotted to the new CH and the previous CH are interchanged in the updated table. The BS also sends the control packet to the CHs for which the previous CH is acting as a relay node. These CHs update relay node id and location in their tables.

After the elapse of  $t_{CPkt}$  time period, CMs transmit data to the CH in the allotted time-slot concurrently in each cluster. During the time period  $t_{CM} * (max\_numSC_{level_{id},i} - 1)$ ,

CH collects the data from all the CMs. All the CHs of the vertical lane, which are farthest from the BS, transmit HELLO message to their relay CH node. After getting the acknowledgment message, they start sending data concurrently to their relay nodes in the allotted time interval  $t_{CH}$ . During the time period  $t_{CH} * Var_{CL}$ , data are collected by the CHs of vertical lane nearest to the BS. These CHs first establish the connection with the data mule by exchanging HELLO message as it reaches to the SP of the corresponding level and then transmit data to it. Data mule collects the data at all SPs in sequence and delivers it to the BS.

## 4 Simulation Environment

In order to evaluate the performance of our scheme, we have simulated it in OMNeT++ [38] on Intel Core i7 3.6 GHz CPU and 8GB RAM running on Microsoft Windows 8.1 professional platform. A deployment area of  $500 \times 250 \text{ m}^2$  is chosen during simulation where 500 and 750 normal nodes are randomly deployed. The mule moves in the simulation area from one end of the deployment area to the BS. We have simulated 10 different mule movement paths and selected two best paths. The results of mule moving across the boundary and moving in the centre of simulation area are found better than other mule movements. So, the results of mule movements of the above two cases are reported in the article. For each scenario the experiment was performed 40 times and the average result is reported here. The values of the parameters used during simulation are mentioned in Table 3.

## 5 Simulation Results and Analysis

The results of the proposed algorithms are compared against well known existing works LPGCRA [24], GBR [3] and GFTCRA [18] protocols. Comparisons and performances of the protocol in both the scenarios are discussed in the subsequent sections based on the simulation results.

First of all, we evaluated the near-optimal frequency of the CH change for both the scenarios. We have set the values of round number after which CH will change and the width decreasing factor of grids as  $f = 7$  and  $r = 0.5$  respectively through out the simulations. The justification for taking these values and their testified graphs with respect to first node dies (FND) is discussed below.

### 5.1 Selection of Near-Optimal CH Change Factor ( $f$ )

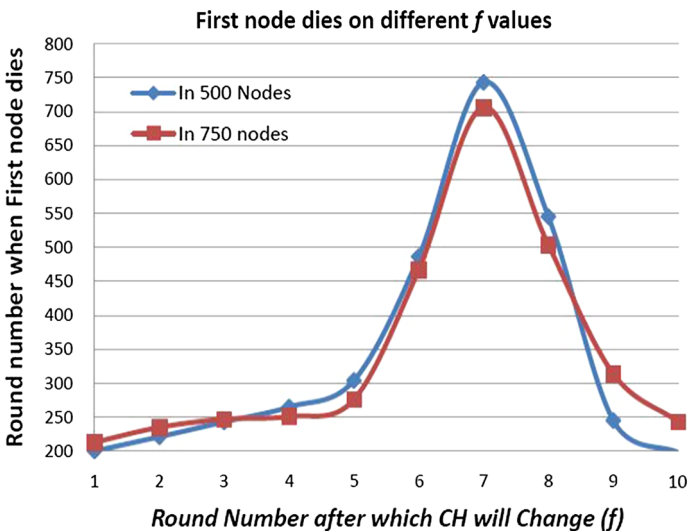
The  $f$  represents the round number after which the current CH should leave its role and a new CH should take over. We repeated the experiment with  $f$  values ranging from 1 to 10 and noted the round number when we first observed the dead node. The round after which first node died is highest at  $f$  value 7 for both scenarios as shown in Fig. 2. The nodes are dying at a very slow rate and uniformly for  $f$  value 7 as can be seen from Tables 4 and 5. Upto round number 1750, only 173 nodes died for  $f$  value 7 whereas for all other case it is higher than that. This is because the CH change at this rate facilitates the nodes to drain their energy at almost equal rate. However after frequency value of 7, the number of dead nodes in early round starts increasing as shown in Tables 4 and 5. This indicates that for

**Table 3** Simulation parameters

Simulation parameters	Value
Deployment area	500 m × 250 m
Z	90 m
Average number of clusters	96
Average number of nodes	500 and 750
Initial energy of node	0.5 J
$E_{elect}$	50 nJ/bit
$\epsilon_{fs}$	10 pJ/bit/m <sup>2</sup>
$\epsilon_{mp}$	0.0013 pJ/bit/m <sup>4</sup>
Data packet length	1000 bits
Control packet length	100 bits
Communication range of nodes	87 m
Antenna type	Omni directional
MAC protocol	IEEE 802.11
Traffic type	CBR (constant bit rate)
Data mule speed	10 m/s

our current setting a frequency value of 7 for CH change is better. Although, after 2000 rounds of operations for frequency of 7, the dead nodes increases dramatically compared to other frequency value. This indicates that the energy of nodes are decreasing uniformly for frequency value of 7 compared to other frequency value. In other cases, some nodes are dying too early due to their depletion of energy indicating that energy drainage is non-uniform.

Tables 4 and 5 represents the average number of dead nodes after certain round of operations for the network, where the mule moves in the middle of the network. In a random deployment, the node distribution is the key issue for varying dead nodes in a



**Fig. 2** FND at different *f* factor with respect to round number



round. In both the cases, we have considered the average dead nodes of more than 20 different randomly deployed sensor nodes in the target area.

## 5.2 Selection of Near-Optimal Width Decreasing Factor ( $r$ )

Though our clusters are in unequal size and its width is decreasing in a square exponent as mentioned in Eq. 4, but for this, we should have to take the optimal value of  $r$ . In order to efficiently balance the load in clusters and across the network,  $r$  should be efficiently set. To analyze the influence of  $r$  parameter on network lifetime, we have considered the range of  $r$  from 0 to 1 and simulated it on 10 different values between 0 and 1. The simulation is done with respect to round number and range of  $r$  to find out the round number when first node has died in the network. As we can see from the Fig. 3 Range at  $r = 0.5$  and  $0.6$  the FND after 740 rounds but the  $0.5$  is optimal and best in all cases and for other values like  $0.2$ ,  $0.4$  and  $0.8$ , the FND within 700 rounds. Hence, we have set  $r = 0.5$  in our scheme.

## 5.3 Simulation Results of WSN-Scenario#1

In WSN-Scenario#1, the data mule moves in the middle of the deployment area in a predefined straight line path as shown in the Fig. 4. This straight line path is defined by connecting all the SPs positioned in the middle of each level. The position of the sojourn points  $SP_0$  to  $SP_5$  are (125, 20), (125, 60), (125, 100), (125, 140), (125, 180) and (125, 220) respectively.

### 5.3.1 Comparison Results in Terms of Dead Nodes

One of the primary motivation in WSN is to prolong the life of operations. Hence, number of dead nodes after a certain round of operations is one of the primary metric evaluated by researchers.

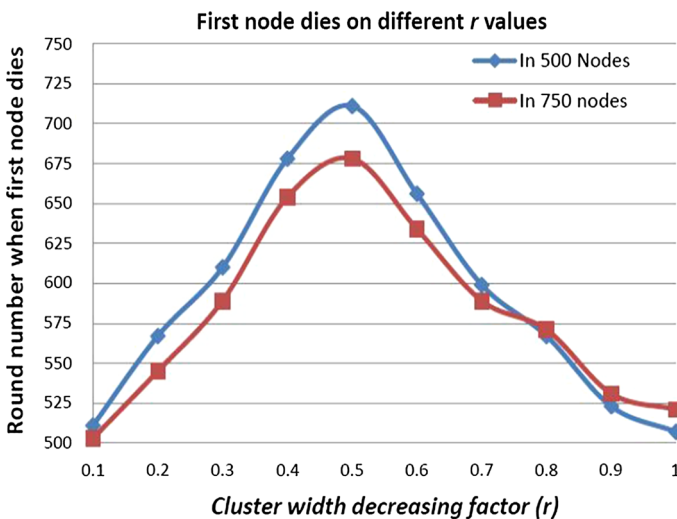
The number of dead nodes versus round number of the proposed work is compared against existing literatures LPGCRA [24], GBR [3] and GFTCRA [18]. The results are plotted in Figs. 5 and 6. As can be seen from the Figs. 5 and 6 that our protocol has the least number of dead nodes throughout the operation of the network compared to other protocols. The first nodes dies (FND) and last node dies (LND) are the key parameters for measuring lifetime of the network. FND may lead to a isolation in the network and LND indicates the total lifetime of the network. It is observed from several simulations that the

**Table 4** Number of Dead Nodes with 500 nodes in the network

Round number	$f = 3$	$f = 4$	$f = 5$	$f = 6$	$f = 7$	$f = 8$	$f = 9$	$f = 10$
500	11	8	6	3	0	4	13	18
750	44	26	12	9	3	23	31	54
1000	108	72	45	35	27	41	65	88
1250	167	131	95	64	57	61	93	112
1500	228	202	161	130	88	75	97	154
1750	273	267	239	206	173	187	231	263
2000	308	320	319	323	333	376	402	450

**Table 5** Number of dead nodes with 750 nodes in network

Round number	$f = 3$	$f = 4$	$f = 5$	$f = 6$	$f = 7$	$f = 8$	$f = 9$	$f = 10$
500	15	11	4	0	0	0	19	24
750	41	25	11	5	1	9	21	52
1000	85	59	42	27	17	22	36	49
1250	185	104	79	60	42	58	89	143
1500	284	201	126	103	81	96	154	208
1750	402	329	240	160	152	175	214	278
2000	480	461	506	555	580	578	598	603

**Fig. 3** The impact of  $r$  on FND with respect to round number

first node died after the 721 rounds in our protocol, whereas in case of GFTCRA, GBR and LPGCRA, the first node died after round numbers 596, 5 and 7 respectively with 500 nodes as shown in the Table 6. The optimized CH rotation, cluster size selection based on threshold transmission distance and the centroid distance for CH selection are the key factors for the better performance. Uneven clustering is a major lawsuit for the balanced energy depletion of the nodes. This increases the lifetime of the nodes near the BS and minimizes the hot spot problem. It also conserves transmission energy loss of nodes and increases the network lifetime.

The same effect is presented in a different representation as shown in Tables 6 and 7 which shows the round number when a fraction of the total nodes are dead. It is seen that our proposed scheme has the maximum rounds of operations before a fraction say 25, 50, and 65% of total nodes are dead as compared to the other existing protocols.

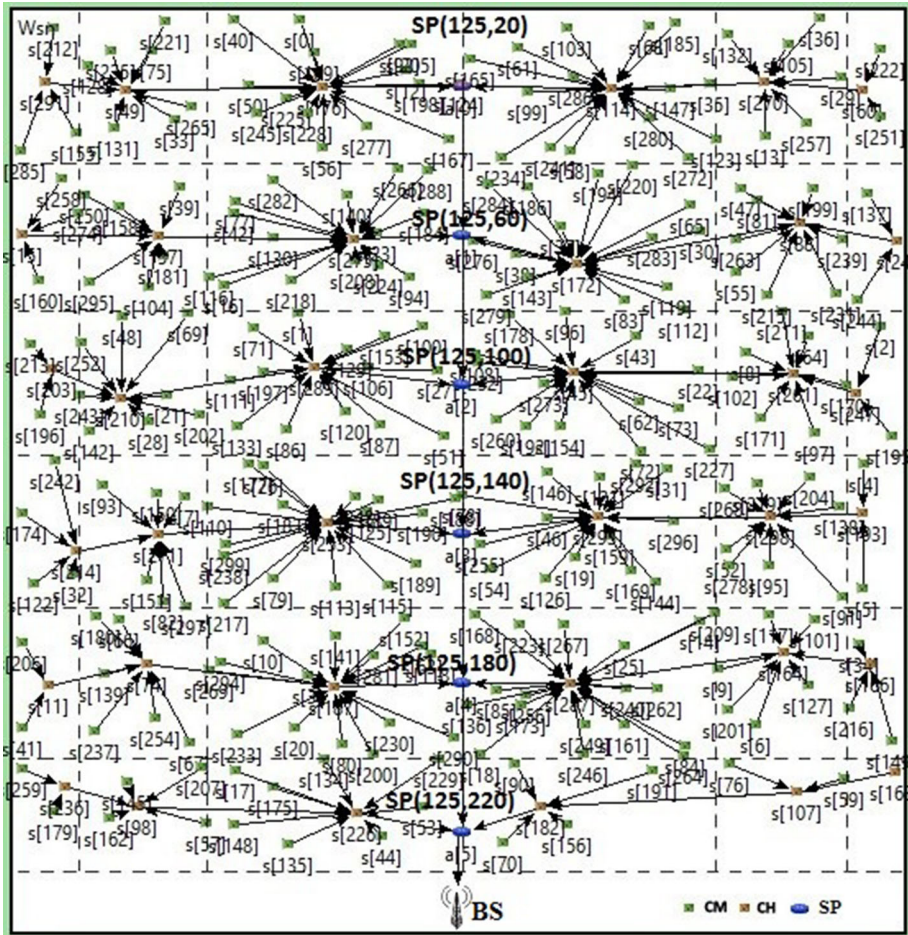


Fig. 4 WSN-scenario#1

### 5.3.2 Comparison Results in Terms of Residual Energy of the Network

Figures 7 and 8 illustrate the average residual energy (mJ) of the network against the different round numbers for the deployment of 500 and 750 nodes respectively. The comparison results clearly show that the residual energy of the network in our proposed protocol is more than the GFTCRA and even more than the GBR in both deployment cases. This is due to considering the centroid distance and  $E_{th}$  energy for CH selection in our scheme, whereas GFTCRA only considered transmission range and residual energy, and LPGCRA and CBR considered only the residual energy of the sensor nodes. As the number of nodes increases the residual energy decreases as shown in Figs. 7 and 8.

### 5.4 Simulation Results of WSN-Scenario#2

In WSN-Scenario#2, the data mule moves along with the vertical boundary of the deployment area towards the BS in a predefined straight line as shown in the Fig. 9. This

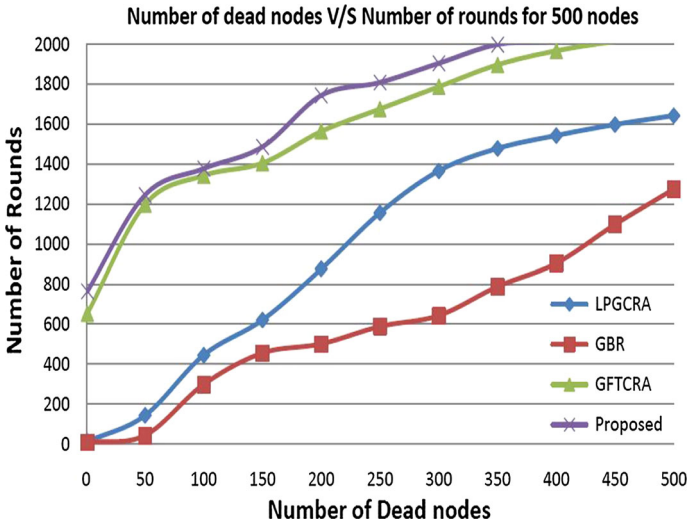


Fig. 5 Number of dead nodes v/s number of rounds for 500 nodes

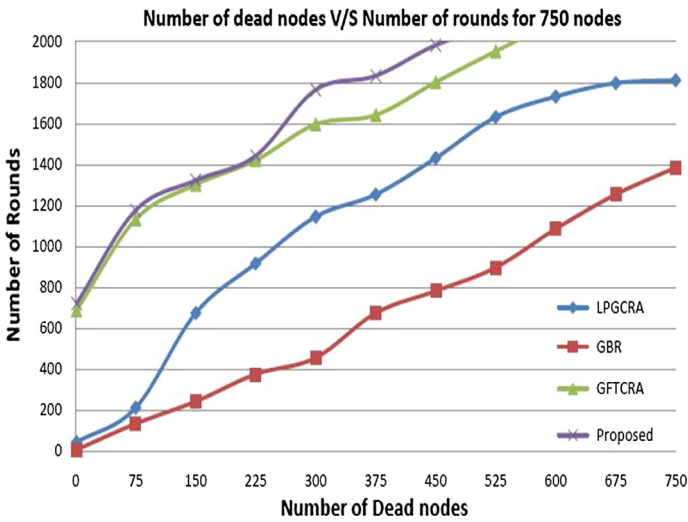
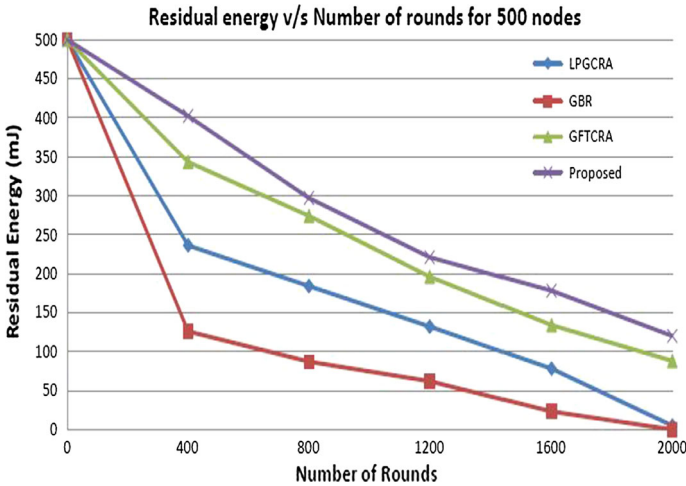


Fig. 6 Number of dead nodes v/s number of rounds for 750 nodes

Table 6 Round numbers for dead node % in 500 nodes deployment

Protocols	First dead	25% dead	50% dead	65% dead
LPGCRA	7	1102	1576	1805
GBR	5	11	16	510
GFTCRA	596	1311	1660	1903
Proposed	721	1489	1878	2008



**Fig. 7** Residual energy (mJ) v/s number of rounds for 500 nodes

path is defined by connecting all the SPs located at the border of each level. The position of the sojourn points  $SP_0$  to  $SP_5$  are (250, 20), (250, 60), (250, 100), (250, 140), (250, 180) and (250, 220) respectively. In this scenario we have again simulated with 500 and 750 nodes which are randomly deployed in the area of 500 X 250 meters. The size of the cluster increases from left to right as we move towards SP. The SPs are located at the middle of the right edge of each cluster. The data load of the clusters increases while moving towards the SP but due to the large cluster size and more number of sensor nodes, a particular node will not become CH frequently. From Table 7 we can see that the round numbers against the FND are more in case of 500 nodes as compared to 750 nodes. This is mainly because the network bears higher overheads and has more data to handle in case of higher nodes. However, in this case its performance is far better than other three protocols.

#### 5.4.1 Comparison Results in Terms of Dead Nodes

The performance in terms of dead nodes versus round number in both types of node distribution: 500 and 750 slightly changes from the WSN-Scenario#1. Figures 10 and 11 demonstrate the comparison graph of the dead nodes with respect to the number of rounds with other three similar existing schemes. It is seen that the proposed protocol performs better and it has less dead nodes even in the higher round numbers as compared to existing protocols LPGCRA, GBR and GFTCRA. In this scenario, data mule moves at a uniform speed and collects data from the borderlined CH directly nearer to the BS, which cuts down

**Table 7** Round numbers for dead node % in 750 nodes deployment

Protocols	First dead	25% dead	50% dead	65% dead
LPGCRA	7	1183	1628	1832
GBR	4	13	19	842
GFTCRA	570	1361	1668	1882
Proposed	668	1641	1896	1991



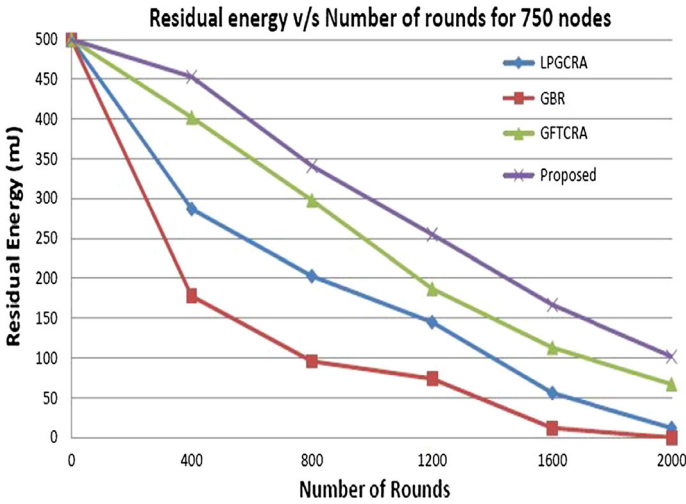


Fig. 8 Residual energy (mJ) v/s number of rounds for 750 nodes

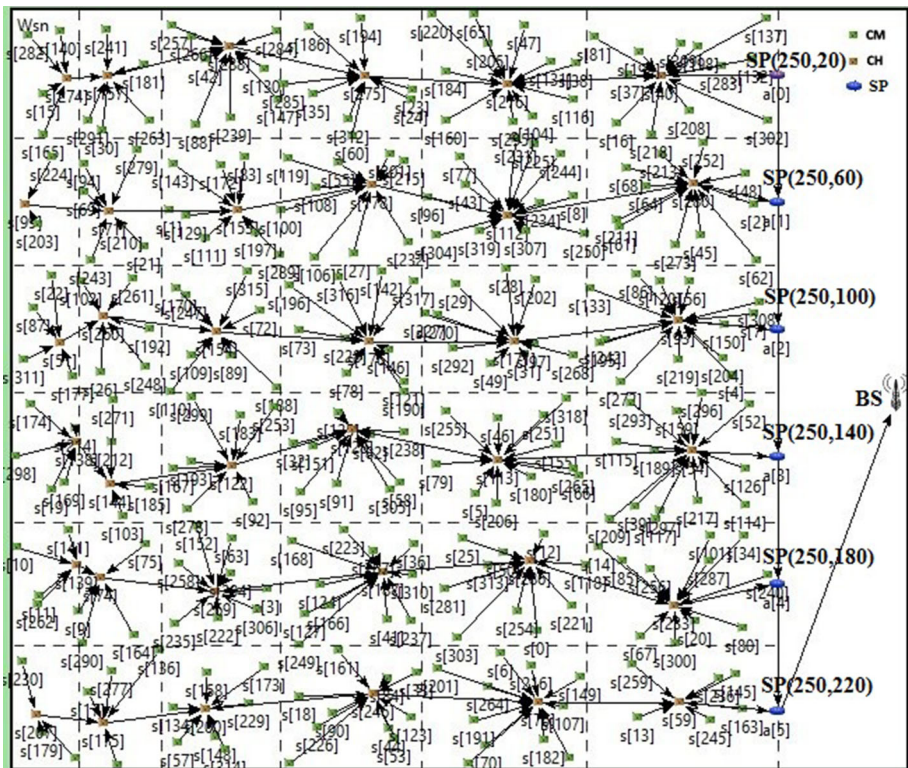


Fig. 9 WSN-Scenario#2

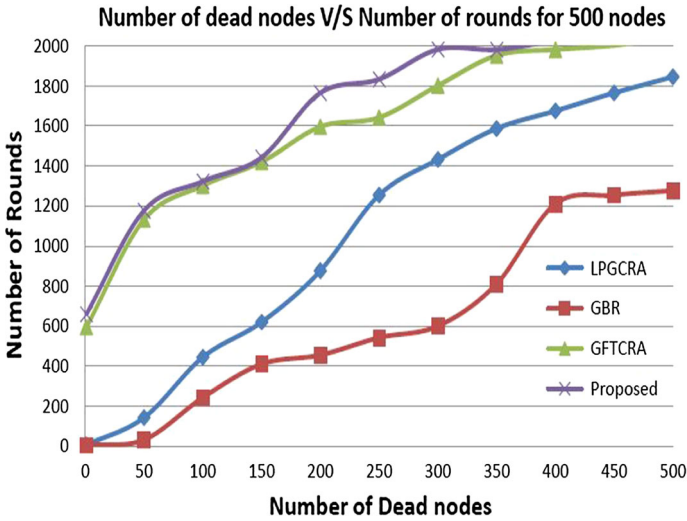


Fig. 10 Number of dead nodes (mJ) v/s number of rounds for 500 nodes

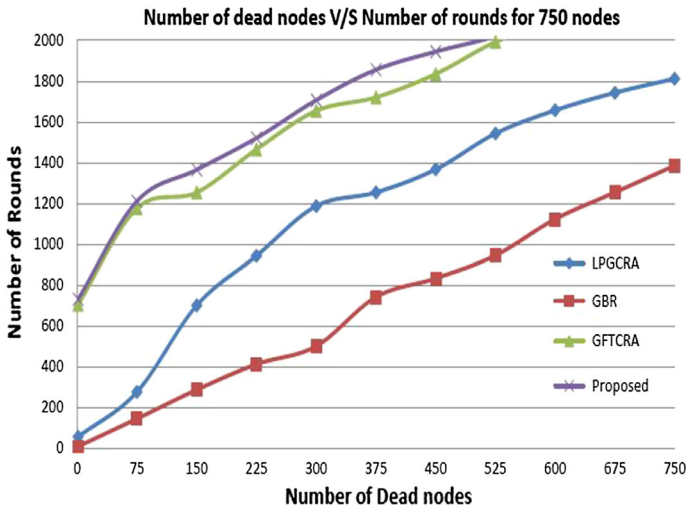


Fig. 11 Number of dead nodes v/s number of rounds for 750 nodes

the load of nodes which are nearer to the BS. This increases the overall lifetime of the network and it is evident for the reduction of the hot spot problem in our protocol.

### 5.4.2 Comparison Results in Terms of Residual Energy of the Network

The average residual energy of the entire network is plotted in Figs. 12 and 13 against the number of rounds by simulating the proposed scheme with LPGCRA, GBR and GFTCRA algorithms. The energy consumption in this scenario is higher as compared to the WSN-Scenario#1 mainly because of more overheads and data is transmitted by additional nodes

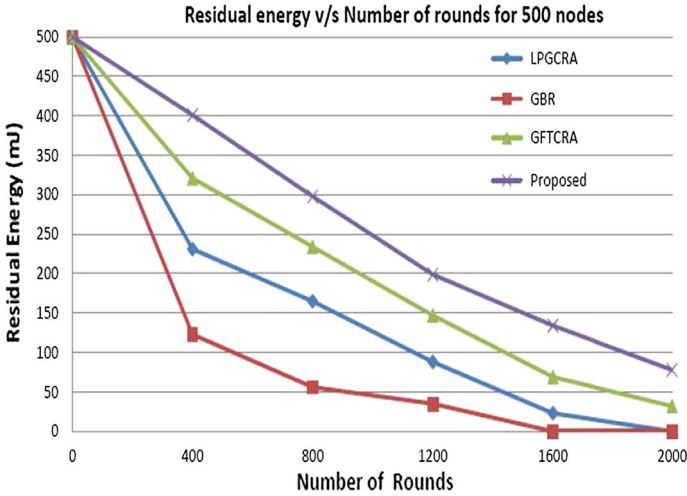


Fig. 12 Residual energy (mJ) v/s number of rounds for 500 nodes

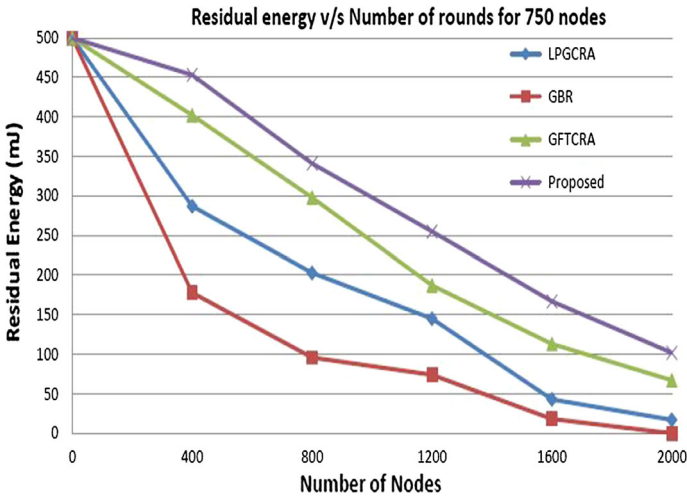


Fig. 13 Residual energy (mJ) v/s number of rounds for 750 nodes

From Figs. 12 and 13, it is clearly observed that our scheme performs far better than the LPGCRA, GBR and GFTCRA algorithms even beyond 2000 rounds. This means that energy consumption per round of the network is less as compared to other existing protocols. The hot spot problem is also reduced in this scheme due to survival of more number of the nodes that are near the BS.



## 6 Discussion

The major contribution of this research is determining the size of unequal clusters in order to divide a large area sensor network. We established the relationships between different sized clusters as presented in Sect. 3.2. In this way, we have formed larger size clusters near the BS which prevented the early death of sensor nodes and minimized the occurrence of hot spot problem. The other major finding of this research is the near-optimal time to change the CH as evaluated in Sect. 5.1. Since CH have not been changed in every round, overheads are reduced and a significant amount of energy has been saved. Several path planning schemes of data mules have been presented in research [2, 4, 8], but most of them tried to make the mule visit each and every node. But this work has used a single mule and at the same time, communication is done only with vertical boundaries CH. This reduces the mule movement complexity and increases the lifetime of sensor nodes. It is evident from the simulations for scenario#1 in the results section that a mule saves more energy if it moves through the middle of the network as compared to borderline movement. Although, we have considered a BS which is located outside the network, our scheme should work efficiently even for scenarios where the BS is placed in the centre of the network. In such cases, the same scheme can be applied after dividing the network into two or four equal parts.

The proposed protocol performed better for WSN-Scenario#1 than WSN-Scenario#2 in terms of dead nodes, average residual energy and network lifetime. In both scenarios, hot spot is also slimmed as our protocol sustains upto more than 3000 rounds, whereas other protocols sustain below 2000 rounds. Also the proposed protocol provides better performance over the existing protocols like LPGCRA, GBR and GFTCRA in both the scenarios.

The worst case time complexity of our routing scheme is  $O(p)$  per CH which is similar to the [18], where  $p$  is the maximum number of paths of the CH. In route discovery algorithm,  $p$  processing time is required for picking all alternative paths. However, we have increased the lifetime of the network as compared to [18] by keeping the routing complexity same. Another complexity that we have calculated is for overheads due to the flow of control messages which is found to be  $O(M)$  as discussed in Lemma 2. The complexity is similar to the [40] which proves the simplicity of our scheme. For a reliable data communication in a multihop routing, the connectivity between CH with other CHs or relay nodes with alternative paths is very essential. The Lemma 1 proved that our proposed multihop routing is reliable and adequate to handle cases where relay nodes or CHs' fail.

**Lemma 1** *Each CH ensures that its connectivity with other adjacent CHs in the network is in between  $\geq (1 - \frac{1}{2^3})$  and  $(1 - \frac{1}{2^2})$ .*

*Proof* As mentioned earlier, the target area is partitioned into number of unequal size grids. The CHs are appointed among the normal nodes of each grid. The connectivity of a CH with all its adjacent CHs is certain with this type of grid partition. The maximum distance between two consecutive grids is the threshold distance of the free space model. Hence, suppose a  $CH_i$  positioned at any corner of the four neighbouring grids will be free to communicate with one of its neighbor CH, say  $CH_j$ , which is also situated at the corner. That means, both CHs are placed at diagonally opposite corner of the radio range  $Z$  in the worst case as shown in the Fig. 1. A  $CH_i$  is always within the communication range of maximum 4 CHs and minimum 2 CHs(for corner). However, maximum 3 neighbour CHs can forward their data towards the SP.

A CH is either dead or alive in the network hence, its probability will be  $\frac{1}{2}$ . So, the

probability of minimum one alive CH for forwarding data towards the BS is  $\geq (1 - \frac{1}{2^3}) \approx 0.87$  which is very high. If the CH is at any corner of the network, then the maximum neighbour in CHs are 2 and the probability of the connectivity is  $> = (1 - \frac{1}{2^2}) = 0.75$  which is also very high.

This proves that our grid partitioning method has better percentage of CHs connectivity.

**Lemma 2** *The complexity of the overhead due to control messages is  $O(M)$  in the network.*

*Proof* In this scheme, the BS is doing maximum work like network partition, CH selection and route setup. So, the nodes only receive the broadcast packets of BS and send their status (residual energy) information through data packets to the BS in each round. Suppose,  $M$  sensor nodes are placed in the network, so the total broadcast packets received by nodes are  $M$ . In each round, member nodes broadcast a join message in each cluster and each CH broadcasts a CH message, a schedule message and a route setup message. If in a network  $N$  number of CHs are selected, then the total number of join message is  $M - N$  and number of messages between the CHs is  $N$ . So, the maximum control messages in the network is:  $M + (M - N) + N + N + N = 2M + 2N$ . Therefore, The complexity of the overhead due to control messages is  $O(M)$  in the network.

## 7 Conclusion and Future Work

In this paper, we have proposed a centralized unequal clustering based routing protocol for mitigating the hot spot problem of the multihop routing. A data mule, which moves in a predefined path and time interval, is used to gather the data from the network and deliver it to the BS. The simple, straight path of the data mule reduces the complexity of the protocol as compared to other protocols. Due to centralized nature of the protocol, maximum of the tasks are performed by the BS. This ultimately preserves the energy of nodes and prolongs the overall network lifetime. The well organized size allocation and size variation of clusters reduces the frequency at which a particular node becomes CH, which results uniform energy consumption in the whole network. The simulation results show that the proposed protocol is better than the existing protocols LPGCRA, GBR and GFTCRA based on number of dead nodes, average energy depletion and lifetime of the network in both the scenarios. The near-optimal value of  $r$  and  $f$ , which we have suggested in this paper is also a major cause of the performance enhancement in all aspects. The simulation results clearly indicate the lifespan of the whole network is more than 3000 rounds, which evident that the hot spot problem is reduced significantly in the network. Further, the results shows that the protocol performs better when the data mule moves in the middle section of the network as compared to movement at the outskirts of the network towards the BS.

In future, we will attempt to implement this scheme in distributed nature and analyze the performance at different data mule speeds and different data generation rates of the sensor nodes.

## References

1. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: A survey. *Computer Networks*, 38(4), 393–422.
2. Alsalih, W., Hassanein, H., & Akl, S. (2010). Placement of multiple mobile data collectors in wireless sensor networks. *Ad Hoc Networks*, 8(4), 378–390.
3. Amrutha, K. M., Ashwini, P., Raj, D. K., Kavitha, R. G., & Mundada, M. R. (2013). Energy efficient clustering and grid based routing in wireless sensor networks. In *International conference on advances in computing* (pp. 69–74). Springer.
4. Anastasi, G., Conti, M., & Di Francesco, M. (2009). Reliable and energy-efficient data collection in sparse sensor networks with mobile elements. *Performance Evaluation*, 66(12), 791–810.
5. Asharioun, H., Asadollahi, H., Wan, T.-C., & Gharaei, N. (2015). A survey on analytical modeling and mitigation techniques for the energy hole problem in corona-based wireless sensor network. *Wireless Personal Communications*, 81(1), 161–187.
6. Bandyopadhyay, Seema., & Coyle, Edward J. (2003). An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *INFOCOM 2003* (Vol. 3, pp. 1713–1723). IEEE.
7. Khurana Batra, P., & Kant, K. (2016). LEACH-MAC: A new cluster head selection algorithm for wireless sensor networks. *Wireless Networks*, 22(1), 49–60.
8. Di Francesco, M., Das, S. K., & Anastasi, G. (2011). Data collection in wireless sensor networks with mobile elements: A survey. *ACM Transactions on Sensor Networks (TOSN)*, 8(1), 7.
9. Fadel, E., Gungor, V. C., Nassef, L., Akkari, N., Abbas Maik, M. G., Almasri, S., et al. (2015). A survey on wireless sensor networks for smart grid. *Computer Communications*, 71, 22–33.
10. Farooq, Muhammad Omer, Dogar, Abdul Basit, & Shah, Ghalib Asadullah. (2010). Mr-leach: Multi-hop routing with low energy adaptive clustering hierarchy. In *24 international conference on sensor technologies and applications (SENSORCOMM)*, (pp. 262–268). IEEE.
11. Gao, T., Jin, R. C., Song, J. Y., Xu, T. B., & Wang, L. D. (2012). Energy-efficient cluster head selection scheme based on multiple criteria decision making for wireless sensor networks. *Wireless Personal Communications*, 63(4), 871–894.
12. Guiloufi, A. B. F., Nasri, N., & Kachouri, A. (2016). An energy-efficient unequal clustering algorithm using ‘sierpinski triangle’ for wsns. *Wireless Personal Communications*, 88(3), 449–465.
13. Harold Robinson, Y., Golden Julie, E., Balaji, S., & Ayyasamy, A. (2016). Energy aware clustering scheme in wireless sensor network using neuro-fuzzy approach. *Wireless Personal Communications*, 1–19.
14. Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on wireless communications*, 1(4), 660–670.
15. Heinzelman, Wendi Rabiner, Chandrakasan, Anantha, & Balakrishnan, Hari. (2000). Energy-efficient communication protocol for wireless microsensor networks. In *Proceedings of the 33rd annual Hawaii international conference on System sciences, 2000*. (pp. 01–10). IEEE.
16. Huang, J.-H., Zhao, Z.-M., Yuan, Y.-B., & Hong, Y.-D. (2017). Multi-factor and distributed clustering routing protocol in wireless sensor networks. *Wireless Personal Communications*, 95(3), 2127–2142.
17. Jain, S., Shah, R. C., Brunette, W., Borriello, G., & Roy, S. (2006). Exploiting mobility for energy efficient data collection in wireless sensor networks. *Mobile Networks and Applications*, 11(3), 327–339.
18. Jannu, S., & Jana, P. K. (2015). A grid based clustering and routing algorithm for solving hot spot problem in wireless sensor networks. *Wireless Networks*, 22(6), 1901–1916.
19. Jin, Y., Wang, L., Kim, Y., & Yang, X. (2008). Eemc: An energy-efficient multi-level clustering algorithm for large-scale wireless sensor networks. *Computer Networks*, 52(3), 542–562.
20. Khan, J. A., Qureshi, H. K., & Iqbal, A. (2015). Energy management in wireless sensor networks: A survey. *Computers & Electrical Engineering*, 41, 159–176.
21. Kim, Jung-Hwan, Chauhdary, Sajjad Hussain, Yang, Wen-Cheng, Kim, Dong-Sub, & Park, Myong-Soon. (2008). Produce: a probability-driven unequal clustering mechanism for wireless sensor networks. In *22nd international conference on advanced information networking and applications* (pp. 928–933). IEEE.
22. Muthu Krishnan, A., & Ganesh Kumar, P. (2015). An effective clustering approach with data aggregation using multiple mobile sinks for heterogeneous WSN. *Wireless Personal Communications*, 90(2), 423–434.
23. Lee, S., Choe, H., Park, B., Song, Y., & Kim, C. (2011). Luca: An energy-efficient unequal clustering algorithm using location information for wireless sensor networks. *Wireless Personal Communications*, 56(4), 715–731.

24. Liu, W.-d., Wang, Z.-d., Zhang, S., & Wang, Q.-q. (2010). A low power grid-based cluster routing algorithm of wireless sensor networks. In *2010 international forum on information technology and applications (IFITA)* (Vol. 1, pp. 227–229). IEEE.
25. Luo, J., & Hubaux, J.-P. (2005). Joint mobility and routing for lifetime elongation in wireless sensor networks. In *Proceedings IEEE 24th annual joint conference of the IEEE computer and communications societies* (Vol. 3, pp. 1735–1746). IEEE.
26. Majma, M. R., Pedram, H., & Dehghan, M. (2014). Igbdd: Intelligent grid based data dissemination protocol for mobile sink in wireless sensor networks. *Wireless Personal Communications*, 78(1), 687–714.
27. Mann, P. S., & Singh, S. (2017). Energy-efficient hierarchical routing for wireless sensor networks: A swarm intelligence approach. *Wireless Personal Communications*, 92(2), 785–805.
28. Mantri, D., Prasad, N. R., & Prasad, R. (2014). Two tier cluster based data aggregation (ttcda) for efficient bandwidth utilization in wireless sensor network. *Wireless Personal Communications*, 75(4), 2589–2606.
29. Mantri, D. S., Prasad, N. R., & Prasad, R. (2016). Mobility and heterogeneity aware cluster-based data aggregation for wireless sensor network. *Wireless Personal Communications*, 86(2), 975–993.
30. Mishra, R., Jha, V., Tripathi, R. K., & Sharma, A. K. (2017). Energy efficient approach in wireless sensor networks using game theoretic approach and ant colony optimization. *Wireless Personal Communications*, 95(3), 3333–3355.
31. Pantazis, N. A., Nikolidakis, S. A., & Vergados, D. D. (2013). Energy-efficient routing protocols in wireless sensor networks: A survey. *IEEE Communications Surveys & Tutorials*, 15(2), 551–591.
32. Prince, B., Kumar, P., Singh, M. P., & Singh, J. P. (2016). An energy efficient uneven grid clustering based routing protocol for wireless sensor networks. In *IEEE conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (pp. 1624–1628).
33. Aravindh, P., Sasikala, S. D., & Sangameswaran, N. (2015). Improving the energy efficiency of leach protocol using VCH in wireless sensor network. *International Journal of Engineering Development and Research*, 3(2), 918–924.
34. Shah, R. C., Roy, S., Jain, S., & Brunette, W. (2003). Data mules: Modeling a three-tier architecture for sparse sensor networks. In *2003 IEEE international workshop on sensor network protocols and applications, 2003. Proceedings of the first IEEE* (pp. 30–41), May 2003.
35. Singh, S. K., Kumar, P., & Singh, J.P. (2016). An energy efficient odd-even round number based data collection using mules in WSNs. In *2016 international conference on wireless communications, signal processing and networking (WiSPNET)* (pp. 1255–1259). March 2016.
36. Singh, S. K., Kumar, P., & Singh, J. P. (2017). A survey on successors of leach protocol. *IEEE Access*, 5, 4298–4328.
37. Sugihara, R., & Gupta, R. K. (2010). Optimal speed control of mobile node for data collection in sensor networks. *IEEE Transactions on Mobile Computing*, 9(1), 127–139.
38. Varga, A., & Hornig, R. (2008, March). An overview of the OMNeT++ simulation environment. In *Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops* (p. 60). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
39. Younis, O., & Fahmy, S. (2004). Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Transactions on mobile computing*, 3(4), 366–379.
40. Jiguo, Y., Qi, Y., Wang, G., & Xin, G. (2012). A cluster-based routing protocol for wireless sensor networks with nonuniform node distribution. *AEU—International Journal of Electronics and Communications*, 66(1), 54–61.
41. Zhao, W., Ammar, M., & Zegura, E. (2004). A message ferrying approach for data delivery in sparse mobile ad hoc networks. In *5th ACM international symposium on mobile ad hoc networking and computing* (pp. 187–198). ACM.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Sunil Kumar Singh** is currently a Ph.D. scholar in the department of Computer Science and Engineering at National Institute of Technology Patna, India. He received the M.Tech and B.Tech degrees in Computer science and Engineering and Information Technology, both from Kalyani Government Engineering College, Kalyani, India in 2010 and 2007, respectively. He has more than 20 publications in various National/International Journals and Conferences. His current research areas are Wireless Sensor Network (WSN) and Mobile Ad-hoc Network (MANET).



**Dr. Prabhat Kumar** is currently the Head of Computer Science and Engineering Department at National Institute of Technology Patna, India. He is also the Professor-In-charge of the IT Services of the Institute and has experience of more than ten years in Network Planning and Management. He is, currently, the State Student Coordinator of Bihar, India for Computer Society of India (CSI). He holds a Ph.D. in Computer Science and M.Tech. in Information Technology. He has over 70 publications in various International Journals and Conferences (viz. IEEE, ACM, Springer and Elsevier). He is also the reviewer of several reputed journals indexed in SCI, SCIE and Scopus. He has chaired sessions at several international conferences held in India and abroad. He is also in the Program Committee of various National/International Conferences. He is an associate member of IEEE, life member of CSI, International Association of Engineers (IAENG) and global member of Internet Society. He has delivered expert talks and guest lectures at various prestigious institutes. His research area

includes Wireless Sensor Networks, Internet of Things, Social Networks, Operating Systems, Software Engineering, E-governance etc.



**Jyoti Prakash Singh** completed his Ph.D. in 2015 from University of Calcutta. He did his M.Tech in Information Technology and B.Tech in Computer Science and Technology in 2005 and 2000 respectively. He is currently an Assistant Professor in the Department of Computer Science and Engineering in National Institute of Technology Patna, India. He has co-authored seven books in the area of C programming, Data Structures, Ad-hoc networks and Operating systems. Apart from this, he has more than 50 research publications in various national and international journals and conference proceedings. His research interests include text mining, social network, sensor network, information security, and data mining. He is a senior member of IEEE, Life member of Computer Society of India (CSI), Indian Society of Technical Education (ISTE), member of ACM, International Association of Engineers (IAENG), International Association of Computer and Information Technology (IACIT).