International Journal of Wireless Information Networks (2019) 26:96–112
https://doi.org/10.1007/s10776-019-00425-y

Balanced Density-Based Clustering Technique Based on Distributed Spatial Analysis in Wireless Sensor Network

Walaa Abdellatief[1] · Osama Youness[1] · Hatem Abdelkader[2] · Mohiy Hadhoud[1]

Received: 15 December 2015 / Accepted: 18 March 2019 / Published online: 3 April 2019

Abstract
Clustering in wireless sensor networks (WSNs) is an important stage for the communication between sensor nodes. Many clustering techniques were proposed with different characteristics. The main goal of them is to facilitate a power-aware communication between a large number of deployed nodes. One of the important factors which affect the clustering process is the distribution of the nodes. In many real situations, the distribution of nodes is random. This type of distribution produces a network with different density sub-regions. A different number of nodes in each sub-region of the network means different communication load and therefore different energy consumptions. This work proposes a distributed density-based clustering technique called "spatial density-based clustering for WSNs." It aims to achieve balanced energy consumption all over the constructed clusters. This is done with the help of a simple initial spatial analysis for the distribution of the nodes before the clustering process. This analysis divides the network to sub-regions according to their density level. Clusters formed in each sub-region will use a suitable size according to the measured density level. Simulation results show that the proposed technique achieves less power consumption and therefore longer network lifetime when compared with other clustering techniques.

Keywords Wireless sensor network · Clustering · Energy consumption · Load-balance · Density · Topological structure · Spatial analysis

1 Introduction

Nowadays, the research in the field of WSNs has taken a significant attention, in particular with the appearance of the Internet of Things. Sensor nodes used in these networks have distinctive characteristics which give it a chance to be used in a wide range of applications. These nodes are able to detect and monitor changes in physical phenomena such as pressure, temperature, and moisture. Furthermore, thanks to the evolution of the industry, sensors are infinitely small-sized and low-cost devices. These characteristics facilitate embedding them in everything around us and allow applications to use hundreds or even thousands of them. WSN applications include, but not limited to, industrial monitoring and control [1, 2], transportation and logistics [3, 4], smart buildings [5, 6], health care [7, 8], environmental monitoring [9–12], intelligent agriculture [13, 14], animal tracking [15, 16], military applications [17–19].

Despite the numerous advantages of WSNs, there are some challenges in their design. First of these challenges is the limited lifetime of the sensors because they are powered by small batteries. For this reason, researchers are always motivated to search for energy-aware communication protocols. Another challenge is the choice of a deployment method to be adopted by the application. Generally, there are two methods of deployment in WSNs. The first one is the planned deployment where the location of nodes is predetermined according to a prior study for the region of interest. In this type of

✉ Walaa Abdellatief
walaaali285@gmail.com

Osama Youness
osama_youness@hotmail.com

Hatem Abdelkader
hatem6803@yahoo.com

Mohiy Hadhoud
mmhadhoud@yahoo.com

[1] Information Technology Department, Faculty of Computers and Information, Menoufia University, Shebin El Kom, Egypt

[2] Information Systems Department, Faculty of Computers and Information, Menoufia University, Shebin El Kom, Egypt

deployment, the minimum number of nodes that can achieve the goals of the application is used. The second method of deployment is the random deployment. It is used in remote and dangerous places such as battlefields, forests, and volcanoes where human incursion is difficult. In this type of deployment, a large number of nodes are randomly and independently scattered via aircraft passing over the area of the network. Random deployment is an easy and less-expensive deployment method but has its own challenges. One of the important challenges in this method is the clustering process. Clustering is an essential stage in randomly deployed applications as the number of distributed nodes may vary from dozens to thousands which need to communicate together in an energy-restricted infrastructureless network. This process is used to divide deployed sensors into groups. Each group has a selected Cluster Head (CH) node responsible for managing the communication between the members of the group.

The importance of the clustering process come from its effect on the whole communication cycle. It helps in the subsequent stages of the communication such as data aggregation and routing. Data aggregation, which is used in such networks where energy is critical, helps in minimizing the number of data transmissions. By grouping sensors into clusters, CHs take the responsibility of aggregating sensed data from many source nodes in its cluster. Furthermore, routing techniques have to find a reliable path from the massive number of available paths to the final sink, mainly the Base Station (BS) of the network. To facilitate finding this route, data are routed between CHs only. For these reasons, aggregation and routing techniques can achieve the desired effect if sensors in the network are clustered in a proper manner. Consequently, clustering adds to the general network efficiency and has a great interest in research.

Our concern in this paper is clustering in randomly deployed WSNs. We aim to study the effect of randomness on the distribution of node densities before clustering. We study the spatial distribution of nodes to differentiate between dense and sparse sub-regions. This can help to configure clusters emerged in each of these sub-regions to suit the distribution of nodes inside it. Dense sub-regions are recognized by a large number of nodes distributed in a small area, i.e. neighbors are near to each other. Nodes located at the center of these areas are able to achieve more connectivity with less power consumption as it can detect more near neighbors. Therefore, these nodes are recommended to be selected as CHs. On the other hand, sparse sub-regions mean a small number of nodes distributed over a large area. Nodes in these areas have fewer neighbors and need to use more power to communicate with them because they reside away from it.

The proposed protocol takes all these points into consideration. It chooses mostly centered nodes in each sub-region as a CH. It also allows CHs in sparse sub-regions to use larger communication range to be able to communicate with more neighbors. Achieving these conditions produce balanced-size clusters and therefore balanced energy consumption distributed all over the area of the network.

The remaining of the paper is organized as follows: Sect. 2 covers the related work of WSNs clustering techniques. Sections 3 and 4 provides preliminaries and assumptions needed for the proposed solution. In Sect. 5, we define the problem to be handled in this paper and the reasons for it. Section 6 reviews the detailed description of the proposed protocol. The complexity analysis of the proposed protocol is shown in Sect. 7. The specification of the simulation setup and the evaluation results are displayed in Sect. 8 and Sect. 9 respectively. We conclude this work and present future ideas in Sect. 10. Finally, "Appendix" is added at the end of the paperwhere we show in detail the steps of calculating the communication range used in the proposed protocol.

2 Related Work

Clustering has been extensively studied to improve the communication performance of WSNs. Heinzelman et al. in [20] proposed the Low Energy Adaptive Clustering Hierarchy (LEACH) protocol. It was the first dynamic clustering technique addressed mainly for WSNs. LEACH provides a simple mechanism to organize a large number of nodes into clusters and periodically changes the roles of each node in the network. It consists of two stages, which are the setup stage and the steady-state stage. These stages are repeated for a number of rounds during the whole lifetime of the network. The concept of rounds is used to give a chance for all nodes to become a CH. This helps in distributing the energy consumption between all nodes. The operation of LEACH starts with the determination of a percentage of nodes to become a CHs which is p. This percentage value also controls the time to wait for any previously selected CH to become a CH again. This period of time is a number of rounds equal to or greater than $1/p$. For example: if $p = 0.1$ for a network contains 200 nodes, the number of CHs in each round equals $200 \times 0.1 = 20$ node. Anyone of the selected 20 CH has to wait for at least $1/0.1 = 10$ rounds to be selected as a CH again. Nodes that are not selected as a CH in the current round r belongs to a set named G. The determination of a node s to become a CH or a member node depends on a generated random number. This random number is compared with a threshold $T(s)$ computed using Eq. (1). If the random number is less than the threshold value, the node becomes a CH in the current round. After being selected, CHs broadcast an advertisement message to all other nodes. Member nodes join the cluster of the closest CH.

$$T(s) = \begin{cases} \frac{p}{1-p\cdot\left(r mod \frac{1}{p}\right)} & if\ node\ \in G \\ 0 & otherwise \end{cases} \tag{1}$$

The problem of LEACH is that it does not ensure even distribution of clusters as it does not take into consideration the distribution of nodes and the size of produced clusters.

In [21], Younis and Fahmy proposed the Hybrid Energy-Efficient Distributed (HEED) clustering protocol. They produced two metrics to prolong the lifetime of the network. These metrics are the residual energy and the intra-cluster communication cost. In HEED, the probability of a node to be selected as a CH mainly depends on its residual energy. Nodes with higher residual energy are more candidate to become a CH. HEED protocol also takes the intra-cluster communication cost into consideration when member nodes choose their CH. After the selection and the announcement of CHs, member nodes select to join a cluster based on proximity or cluster density. Experiments show that HEED is effective in prolonging the lifetime of the network than LEACH, but it is an iterative technique which consumes more overhead until CHs are selected.

Ye et al. in [22] proposed the Energy Efficient Clustering Scheme (EECS) protocol. It is based on a competition mechanism to select CHs from a set of randomly chosen candidate nodes. The CHs are selected as nodes with higher residual energy between its neighbors. Remaining nodes join clusters based on a cost function. This cost is a tradeoff between the distance from the node to the CH and the distance from the CH to the BS. Tests show that EECS provide longer lifetime than LEACH but the initial randomly selected candidate nodes can produce a non-uniform CHs distribution.

In [23], Yan et al. proposed the Hybrid, Game-Theory based, Distributed (HGTD) clustering protocol. HGTD protocol utilized the concept of game theory to solve the problem of CHs selection. Many nodes compete for the CH role is the same as many players participate in a strategic game, and each one wants to take the leader role. The node with its neighbors are considered as a set of players in a game, and each one of them has to choose a strategy to be a CH or not. This strategy is based on an equilibrium probability which depends on (1) the overhead cost for the node to serve as a CH, (2) the amount of consumed energy for successfully transmitting the data of the members to the BS, and (3) the number of neighbors. Each node calculates its equilibrium probability and generates a random number. If the random number is less than its equilibrium probability, it will be selected as a CH; else it will be a member node. In HGTD, nodes with more neighbors and longer distance to the BS has a fewer opportunity to become a CH. However, HGTD provides a low energy dissipation; it consumes a great time because of many iterations until CHs are selected.

In [24], Heinzelman et al. proposed the LEACH-Centralized (LEACH-C) protocol to overcome the problems of LEACH. In this protocol, the selection of CHs done centrally, not distributed. The BS takes the responsibility of choosing the optimal CHs according to information collected from all nodes. All nodes send their energy and location to the BS at the beginning of the set-up phase. The BS calculates the average network energy then nodes having energy above the average network energy are selected as CHs.

Other authors also use the same concept of centralized clustering and produced performance enhancement using different optimization methodologies such as genetic algorithm and particle swarm optimization.

In [25], Rahmanian et al. proposed a clustering technique based on the Genetic algorithm. The objective function of this Genetic-based technique minimizes the total distance from member nodes to their associated CHs and the distance from the CHs to the BS.

In [26], Latiff et al. used the particle swarm optimization to produce an energy-efficient clustering technique called (PSO-C). The objective function of PSO-C minimizes the average distances between member nodes and their associated CHs and the ratio of the total initial energy of all nodes to the total energy of candidate CHs.

Although it is proven that all these centralized techniques achieve longer lifetime than LEACH, it consumes too much overhead during the setup phase because all nodes have to send their information to the BS.

Discussion Through the previous work; we notice that the problem of clusters formation is solved using distributed or centralized mechanism. Some techniques use a distributed mechanism such as LEACH, HEED, EECS, and HGTD. These techniques depend on a probabilistic mechanism where a random number is generated and compared with a function which utilizes some different parameters such as the residual energy, distance to the BS, or the node degree. The problem of probabilistic techniques is that it does not guarantee uniform distribution of clusters because it does not consider the topological characteristic of the deployed network. Some of these techniques use several iterations trying to optimize their CHs selection such as HEED and HGTD. However, this increases the overhead in the network. Other techniques use a centralized mechanism such as LEACH-C, PSO-C, and Genetic-based clustering. This type of techniques can provide optimal clusters but consumes too much overhead especially with the large numbers of nodes used in WSNs. In these techniques, the central BS has to collect a lot of information about every node in every round of the clustering process. Therefore, in this paper, we propose a global, distributed clustering technique. Based on some global information about the network, each node can locally decide its role. This information is assumed to be known, by default, by the BS for any communication and control activities. This information helps to extract topological information about the network. This topological information helps to select CHs in a distributed manner based on a global view of the deployed network.

3 Preliminaries

Any sensor node consumes power in three activities which are sensing, processing, and communication. Processing power is the smallest one and is ignored in evaluations. Any sensor node has a defined value of a sensing range s_r and a communication range c_r. Sensing and communication ranges are a distance measure that defines a circular sensing and communication area with the node at its center. The sensing range determines the area in which the node is able to monitor the change of the sensed phenomenon. The communication range determines the area in which the node can send and receive information. Nodes reside within the communication range of another one are considered to be neighbors. As shown in Fig. 1, s_2 is a neighbor of s_1, but s_3 is not a neighbor. Most often, the communication range is an adaptable value that can be changed by the node and is assumed to be greater than twice the sensing range [27]. Our concern in this work is the communication range value. This value determines the number of neighbors and therefore the communication energy consumption of each node. The number of neighbors that each node can detect is denoted as the node degree $d(s)$ of any node s. For the whole network, the network degree is calculated as the average of degrees of all nodes in the network.

4 Assumptions

In our proposed protocol, we use the following assumptions:

- The location of nodes is fixed after deployment.
- The number of nodes is N.
- Nodes are randomly distributed in a $W \times H$ network area.
- Nodes can adapt their communication range and use different values for different purposes.
- Nodes can estimate the distance to neighbors.
- Nodes can be heterogeneous and use different initial energy values.

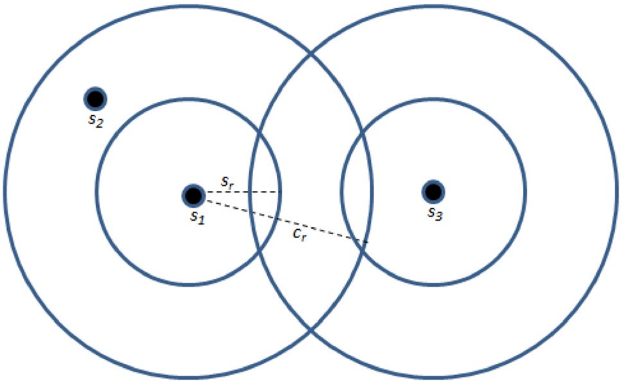

Fig. 1 Communication range and sensing range of sensor nodes

5 Problem Formulation

Our focus in this work is randomly deployed WSN applications such as battlefields and forests applications. Random deployment in these applications leads to regions with different node densities. Actually, there are some reasons which lead to this network figure where some sub-regions are dense, and others are sparse. These reasons are:

1. The method of deploying the sensors:

In randomly deployed applications, nodes are usually deployed using aircraft passing over the area of interest. As sensors are lightweight, it can be easily affected by air while dropping. Moreover, antennas may be poorly placed. Therefore, even if a large number of sensors are deployed, some of them are not practically used because of placement failure.

2. The total number of scattered sensors:

Although it is theoretically advised to increase the number of deployed nodes in random distributions to achieve a uniform density all over the area of the network, we have to keep in mind that this increase is costly. For example, according to Theorem 2 in [28], for a $200 \times 200\text{m}^2$ network, 300 nodes are needed to achieve 98% coverage rate, while 500 nodes reach 99% coverage rate. The cost of 200 additional nodes for only more 1% increase has to be considered. Therefore, it is practically advised to use only the number of nodes that can achieve the coverage requirement of the application.

3. The nature of the application:

Even if enough number of nodes is initially used to get uniform densities all over the area of the network, this distribution changes by the time according to the communication load caused by the application in different regions of the network. For example, nodes located at the border of the network or near to the BS can be used more than other nodes. Different communication load affects the distribution of dead nodes through the lifetime of the network and different density sub-regions of the alive nodes appear.

According to all the previously mentioned reasons, some cases may appear which has a direct impact on the clustering process. These cases are shown in Figs. 2 and 3. The first case shown in Fig. 2, if we apply different communication range values for the node in the middle, the node degree will change. Reducing the communication range may make it

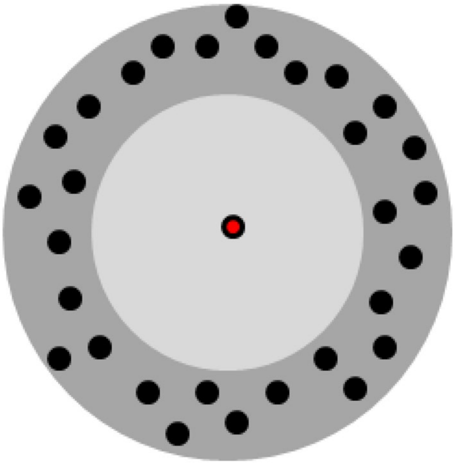

Fig. 2 Effect of node's communication range and node's degree

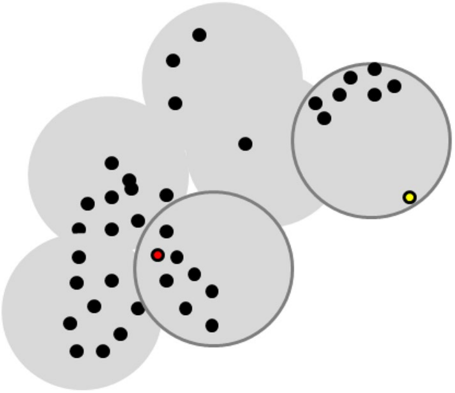

Fig. 3 Effect of random distribution and nodes density

isolated with degree equal to 0, however increasing it with a small value change its degree to 30. Another case is shown in Fig. 3 where the red and yellow nodes have the same degree. However, if the whole network is considered, the red one is located in a middle of the dense region. In such situation, the red node is more suitable to be chosen as a CH than the yellow one. This is because the red one can communicate with a larger number of nodes with less energy consumption.

When nodes are randomly deployed, similar cases may appear. All of these considerations affects the final number of clusters, their distribution in the network, and the size of each cluster. Factors such as the relative position of the node between its neighbors and the size of each cluster depending on the communication range value should be considered.

6 The SDC Protocol

The SDC protocol is a density-based clustering. However, it takes into consideration other factors besides the node degree. The relative position of the node within its neighbors and the effect of changing its communication range are also considered. In the SDC protocol, we try to discover the sub-regions with different density levels resulting from randomness in deployment. This is done by defining boundary nodes which border these sub-regions. The SDC protocol mainly depends on performing a prior spatial analysis for the network before clustering. The use of spatial analysis methods followed by spatial clustering is widely investigated in the field of data clustering [29]. In the SDC protocol, this analysis is based on some global information provided by the BS to every node in the network. The BS is supposed to know the area of the region of interest. And it is also expected to retain the number of deployed nodes because any node must have a connection to the BS to be considered as a working node in the network. By broadcasting these two values at the beginning of the clustering process, each node can start the SDC protocol. To give an overview of the SDC protocol working mechanism, we preview a flowchart in Fig. 4.

The SDC protocol can be divided into three main steps which are: (1) sub-regions border discovery, (2) cluster formation, and (3) data transmission and clusters adjustment.

In the following part of this section, we present the detailed description of each step along with a reference to the line numbers of these steps in the pseudo code of the protocol shown at the end of the section.

(1) Sub-regions border discovery

The aim of this step is to discover the distribution of the neighbors around each node and to be able to differentiate between high-density and low-density sub-regions. Measuring the number of neighbors around each node is not enough. We also want to know how far these neighbors are. To do that, we use a border discovery mechanism to classify nodes into two types which are interior nodes and boundary nodes. The interior nodes are distributed close to each other and construct a region of high-density. These regions are bordered with a set of the boundary nodes located away at the border and begin a low-density region.

The problem of regions border detection in WSNs has been discussed by many authors such as [30–32]. We choose the protocol presented in [33]. The author uses the 1st hop neighborhood information in a distributed algorithm to differentiate between interior nodes and boundary nodes. The node will be interior if it is enclosed in a triangle of three neighboring nodes. This is described as the interior condition. Otherwise, it will be a boundary node. The difference between interior and boundary nodes is shown in Fig. 5.

Based on the condition that the sum of the three angles drawn from each vertex node of the triangle and centered at the interior node equals to 360 degrees, the interior condition can be tested. Nodes that do not satisfy the interior

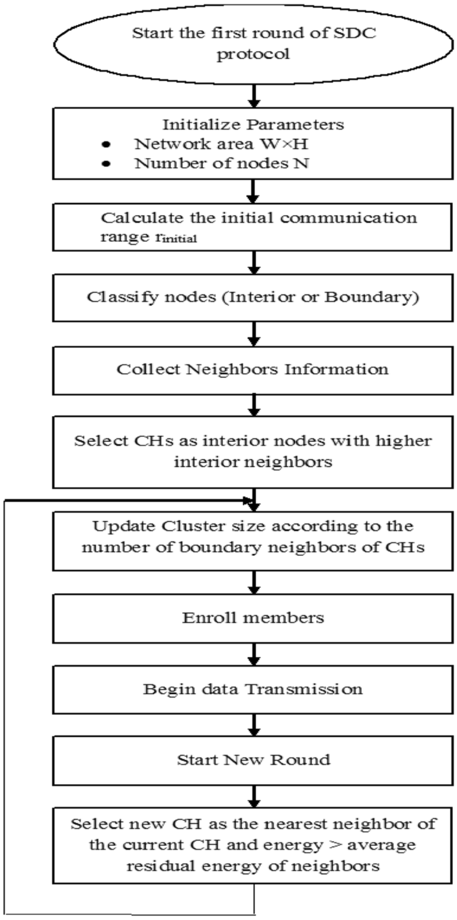
Start the first round of SDC protocol
Initialize Parameters
• Network area W×H
• Number of nodes N
Calculate the initial communication range r_initial
Classify nodes (Interior or Boundary)
Collect Neighbors Information
Select CHs as interior nodes with higher interior neighbors
Update Cluster size according to the number of boundary neighbors of CHs
Enroll members
Begin data Transmission
Start New Round
Select new CH as the nearest neighbor of the current CH and energy > average residual energy of neighbors

Fig. 4 Flowchart of the SDC protocol

condition are classified as boundary node. From Fig. 5, the interior condition can be written as:

$$\angle APB + \angle APC + \angle BPC + = 360 \tag{2}$$

Using distances which can be calcuated according to the received signal strength indicator (RSSI) [34], the interior condition is written as Eq. (3):

$$\cos^{-1}\left(\frac{X^2 + Y^2 - c^2}{2XY}\right) + \cos^{-1}\left(\frac{X^2 + Z^2 - b^2}{2XZ}\right) + \cos^{-1}\left(\frac{Y^2 + Z^2 - a^2}{2YZ}\right) = 360 \tag{3}$$

When each node tests this condition, the result will differ according to the number of neighbors it can detect. When the node uses a large communication range value, the number of its neighbors will increase and therefore its probability to be an interior node will increase. In Fig. 6, we preview the effect of changing the communication range value while testing the interior condition. We also show how the communication range value can be used to differentiate between dense and sparse sub-regions. The figure contains two cases (A) and (B). In the two cases, the same number of nodes are deployed, and the interior condition is tested at all nodes. By using different communication range value in each case, a different set of boundary nodes is defined. In case (A), all nodes use a communication range equals 130 m. It seems that the defined boundary nodes with red color construct the border of the whole area. In case (B), all nodes use communication range equals 65 m. It seems that the appearance of more boundary nodes indicates a low-density sub-region such as the two examples bordered with the green border. On the contrary, the appearance of more interior nodes indicates a high-density sub-region such as the area with the yellow border.

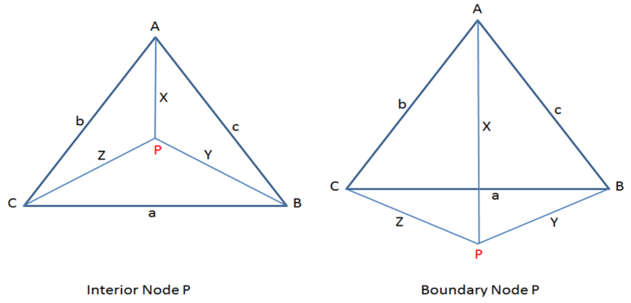

Fig. 5 Difference between interior and boundary node [33]

As the number of neighbors of each node is determined by its communication range, so the question here is: "what is the suitable communication range value that can be used by all nodes in a randomly distributed network which enables them to detect the boundaries of its different densities sub-regions?". The calculation of this value is shown at the end of the paper ("Appendix") in order not to distract the reader. The value calculated in this step is the initial communication range used by all nodes in the SDC protocol. We recall this value as $r_{initial}$ (line 4). According to this value, the set of neighbors of each node is determined, and the interior condition is tested. According to this test, nodes are classified into two types: (1) interior nodes or (2) boundary nodes (lines 3–7). Each node sends a *node_identification(S.id, S.Type)* message to announce its type to all its neighbors (line 8). Upon receiving this message from all neighbors, each node begins to construct a neighboring table which contains the ids and types of all neighbors (lines 9–11). This information will be used in the following step to give an image about the distribution of neighbors around each node.

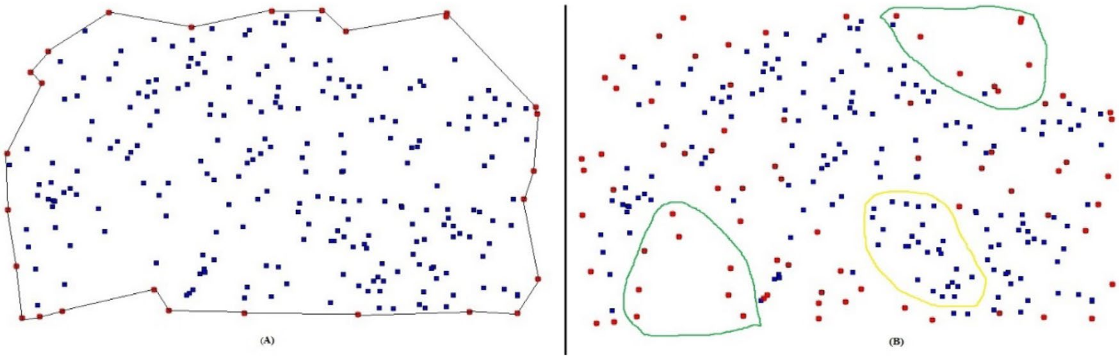

Fig. 6 The effect of changing the communication range value when testing the interior condition

(2) Clusters formation

In the SDC protocol, we depend on counting the number of interior and boundary neighbors around each node. This helps us to differentiate dense sub-regions from sparse ones and choose the best CHs according to this. The appearance of more interior neighbors around a node indicates that it is located in a dense sub-region. Nodes relatively located at the center of dense sub-regions are better candidates for being a CH because they achieve more connectivity with less communication range. On the other hand, the appearance of more boundary neighbors around a node indicates that it is located in a sparse sub-region. These regions are characterized by a small number of nodes distributed over large areas. CHs chosen for these areas have to use more communication range to balance its size i.e. to be able to communicate with more members and increase its connectivity in the network.

According to this, we set the conditions upon which each node can determine to be a CH or not. The first condition is that a CH must be an interior node. The second condition depends on the distribution of the neighbours of the node or the energy of the node. To be able to examine the distribution of the neighbours of the node, we have a new parameter. We named it as the Inner Degree *(IN_degree)* which is the ratio of the number of interior neighbors to all neighbors of the node as shown in Eq. (4).

$$IN\_degree = \left(\frac{Number\ of\ interior\ neighbors}{Number\ of\ all\ neighbors} \right) \tag{4}$$

Nodes calculate its *IN_degree* from the neighboring table information and broadcast it with its energy to its neighbors as a *node_cost(S.id, S.IN$_{Degree}$, S.Energy)* message (lines 12–13). Upon receiving the cost from all neighbors, each node updates its neighboring table by adding the IN_{degree} and energy of all neighbors to it (lines 14–16).

Now, the neighboring table can be used to help each node to decide its role in the network. The interior node with the highest IN_{degree} or highest energy in case of equal inner degrees will change its status to be a CH. By this way, the final selected CHs are highly centered in its sub-region. This choice produces a less and balanced intra-cluster communication cost. After that, we propose to adapt the communication range value of the selected CHs according to the distribution of neighbors around them. The aim of this adaptation is to produce relatively balanced-size clusters and achieve balanced energy consumption all over the network. The adaptation of the communication range of CHs is done according to Eq. (5). All nodes start the process of cluster formation with a communication range value equals to $r_{initial}$ calculated in the first step of the SDC protocol. According to this value, the cluster radius is determined, i.e. the number of members that can join each cluster. Using Eq. (5), CHs which have more boundary neighbors will use larger communication range to be able to reach them.

$$ClusterRadius = r_{initial} + \left(r_{initial} \times \frac{Number\ of\ boundaries\ neighbors}{Number\ of\ all\ neighbors} \right) \tag{5}$$

After the adaptation of the communication range value of the selected CHs, these nodes announce itself as a CHs and broadcast a *head(S.id, S.status)* message to all nodes in its range (lines 17–21). After CHs selection, non-selected nodes start to join the cluster of the nearest CH by sending a *join(S.id, S.CH)* message to that CH. Nodes that does not hear from any CH announce itself as a CH (lines 22–29).

(3) Data transmission and clusters adjustment

After the creation of the first clusters, data transmission starts. CHs use a Time Division Multiple Access (TDMA) scheduling to allocate the time for each member node.

All nodes in the network start the clustering process with an initial energy value. During the lifetime of the network, nodes consume power for sensing, processing, and communication activities. CHs consume more power than members of the cluster because it is responsible for aggregating and sending data sensed by all members in the cluster to the BS.

Therefore, staying as a CH for a long time will accelerate the death of this node. For this reason, CHs have to change their role in order to give the chance for other nodes to become a CH and distribute the communication load among other nodes of the network. In the SDC protocol, rounds of clusters adjustment process start after data transmission. New CHs are proposed to be selected as the nearest node to the current CH with higher residual energy. The reason for that is to preserve the same structure of constructed clusters with CHs mostly centered in the middle as possible and have higher energy than other members. This is done by calculating a cost value and the new CH is the node with the maximum cost. This cost function is shown Eq. (6) where β and γ are factors stated by the application, $E_{Resdiual}$ is the residual energy and D_{CH} is the distance to the current CH.

$$Cost = \beta E_{Resdiual} - YD_{CH} \quad (6)$$

Current CH broadcasts an *Announce_new_CH(S(i).id, New_CH_id)* message to its neighbors to announce the new selected CH (lines 30–38).

7 Complexity Analysis

In this section, the overhead complexity of the SDC protocol is calculated. In the first step of the spatial analysis, the BS broadcasts a message with the value of the *N* and the value of the $W \times H$. The cost of receiving this message is 1. Upon receiving this message, each node calculates the initial communication range. According to this value, each node sends messages to its *n* neighbours to estimate the distance to them., Therefore, the cost of the initial spatial analysis is $(1 + nN)$ messages.

In the first round of the proposed protocol, each node in the network broadcasts a *node_identification* and *node_cost* messages to its neighbors. Therefore, if *N* nodes are deployed, a (2*N*) messages are sent. Based on this analysis, if K CHs are selected, a (*K*) *head* messages and $(N - K)$ join messages are sent. Therefore, the cost of the first round is $(2N + K + N - K) = (3N)$. In the remaining subsequent rounds, new CHs are chosen as the nearest neighbor of the current CH with higher residual energy. In these rounds, the *N* nodes sends their cost to select K new CHs. After that, these new CHs will send *Announce_new_CH*. Therefore, the cost in these rounds is $(N + K) = (N + K)$. Clearly, the total overhead complexity of the control messages in all rounds is a value of order N. So, the overhead complexity of the proposed protocol is *O*(*N*).

Pseudo code for the Spatial Density-based Clustering Technique (SDC)

```
Parameters
   Network Area A= W × H
   Number of Nodes = N
   $r_{initial}$= Clauculate_initial_communication_range(A,N)
Sub-regions border discovery and clusters formation
 Repeat:
   for i=1; i≤N ; i++ do
     if (S(i).Interior_Condition() == true)
       S(i).Type = Interior
     else
       S(i).Type = Boundary
     end if
     Broadcast node_identification(S(i).id, S(i).Type)
     while Receive node_identification(S.id, S.Type)                    // from all neighbors
       Construct_neighboring_table(S.id, S.Type)
     end while
     S(i).IN_degree = Calculate_Inner_degree(neighboring_table);
     Broadcast node_cost(S(i).id, S(i).IN_degree , S(i).Energy)
     while Receive node_cost(S.id, S.IN_degree, S.Energy)               // from all neighbors
       Update_neighboring_table(neighboring_table , S.id , S.IN_degree)
     end while
     if (S(i).Type == Interior &&
         S(i).IN_degree > S.IN_degree of all neighbors || S(i).Energy > S.Energy of all neighbors )
       S(i).Status = CH
       Update_communication_range(S(i).$r_{initial}$)
       Broadcast head(S(i).id,S(i).Status)
     end if
     if Receive head(S.id,S.Status)
       S(i).CH=S.id
       S(i).Status = Member
       Send join(S(i).id, S(i).CH)
     else
       S(i).Status = CH
    end if
 end for
 Start data transmission
 Clusters adjustment
 Repeat
     for i=1; i≤ N ; i++ do
       Calculate $S(i).Cost = \beta\, S(i).Energy - \Upsilon\, S(i).D_{CH}$
       Broadcast S(i).Cost
       while Receive S.Cost                                           // from all neighbors
         If (Max S.Cost)
         Broadcast Announce_new_CH(S(i).id, New_CH_id)
       end while
     end for
```

Table 1 Values of the simulation parameters

Simulation parameters	Values
Initial energy	1–3 J
E_{elec}	50 nJ/bit
ε_{fs}	10 pJ/bit/m^2
ε_{mp}	0.0013 pJ/bit/m^4
E_{DA}	5 nJ/bit/signal
D_0	87 m
Packet size	4000 bytes

8 Simulation Setup

The simulation was done to compare the performance of the SDC protocol with LEACH, HEED, and HGTD. Simulation is performed by MATLAB. Presented results have been averaged over 50 simulation runs.

It is assumed that the locations of the nodes and the BS are fixed. BS is located at the center of the area. Simulation parameters are shown in Table 1. To evaluate the energy consumption in the network, we use the same radio model of [35]. We assume that each node sends one packet to the BS in each round. CHs aggregate data from all members of each cluster and send it to the BS. Circuits of nodes consume $E_{elec}$to send or receive data. Amplifier consumes energy ε_{fs} of the free space model for small distances and ε_{mp} of the multipath fading model for large distances. Nodes also consume E_{DA} to aggregate data before being sent to the BS. Values of the simulation parameters are the same as used in [20] and many of its variants [36–39].

Equations of the energy consumption activities are calculated for each node as follows:

1) Transmit a L-bit message to a sink located at distance d

$$E_{TX} = E_{elec} \cdot L + \varepsilon_{fs} \cdot L \cdot d^2 \; when \, d < D_0 \tag{7}$$

$$E_{TX} = E_{elec} \cdot L + \varepsilon_{mp} \cdot L \cdot d^4 \; when \, d \geq D_0 \tag{8}$$

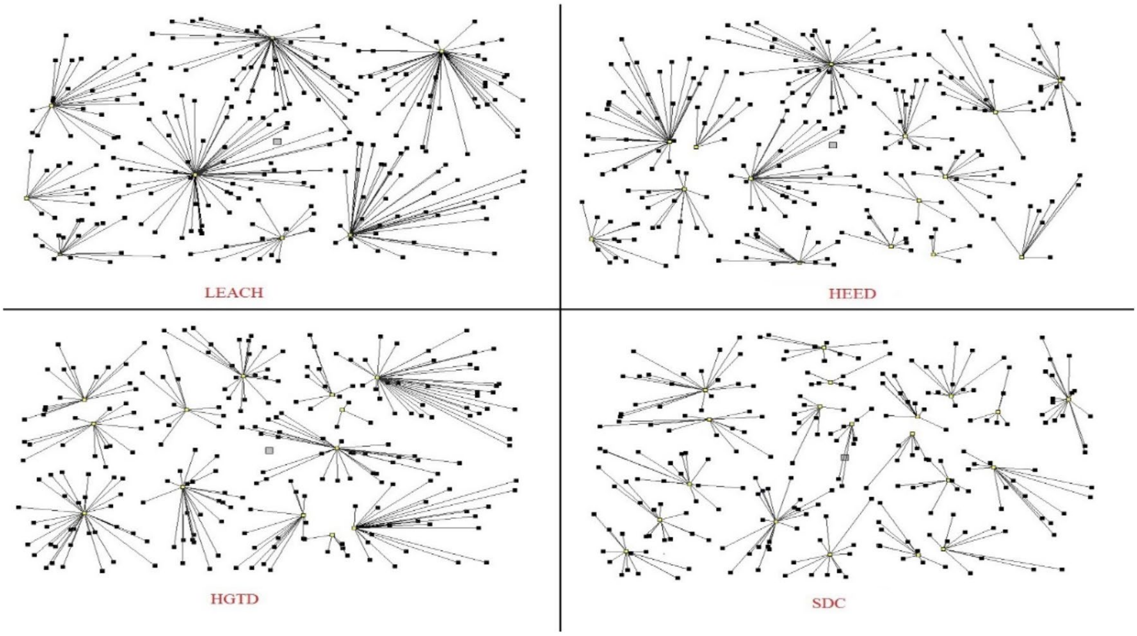

Fig. 7 Comparison between LEACH, HEED, HGTD, and SDC clusters

Table 2 Mean and standard deviation of clusters size

Protocol	100 nodes		150 nodes		200 nodes		250 nodes		300 nodes	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
LEACH	20.23	16.04	20.03	17.13	20	15.94	20.08	13.85	20.03	13.65
HEED	17.78	14.5	18.13	13.08	18.96	14.65	18.12	14.81	18.69	14.57
HGTD	15.72	8.7	15.95	10.11	16.05	10.91	16.23	9.99	16.34	9.68
SDC	9.18	2.12	11.92	3.11	14.63	3.84	16.81	3.26	18.76	3.64

Table 3 Mean and standard deviation of energy consumption inside clusters

Protocol	100 nodes		150 nodes		200 nodes		250 nodes		300 nodes	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
LEACH	5.67	5.38	14.6	7.22	16.18	7.56	18.65	8.92	25.42	14.54
HEED	3.88	3.71	12.74	6.3	12.84	6.56	15.87	11.9	19.2	11.29
HGTD	3.88	2.95	11.56	6.28	11.96	6.97	14.23	9.66	16.65	11.127
SDC	2.64	0.93	8.71	2.6	8.93	3.01	9.19	3.96	10.18	3.55

2) Receive a L-bit message.

$$E_{RX} = E_{elec} \cdot L \tag{9}$$

3) Aggregate a L-bit message.

$$E_{Agg} = E_{DA} \cdot L \tag{10}$$

The energy consumption in the network is the total value of energy lost in all nodes. Nodes lose different energy amount according to its rule in the network. If the node is a member (Me) node, energy lost during the transmission of the sensed data to the CH is shown in Eq. (11). If the node is a CH, energy lost during the aggregation and the transmission of the collected data to the BS is shown in Eq. (12). The total energy consumption in the network is shown in Eq. (14) where N is the total number of nodes in the network, L is the length of the packets, and K is the number of CHs.

$$E_{Me} = E_{elec} \cdot L + \varepsilon_{fs} \cdot L \cdot d_{CH}^2 \tag{11}$$

$$E_{CH=}E_{DA} \cdot L + E_{elec} \cdot L + \varepsilon_{mp} \cdot L \cdot d_{BS}^4 \tag{12}$$

9 Simulation Results

Simulation experiments were done to show the effectiveness the SDC protocol when compared with other techniques. In Fig. 7, we preview the distribution of produced clusters after applying each protocol for one round to the same deployed network. In SCD, it is evident that member nodes belong to any CH are the nearest set to it. Other techniques do not have control on the distance between the CHs and members. Therefore, we can see some clusters with members located at far distances. These clusters suffer from high intra-communication energy consumption. We can also notice that there is a variation between the sizes of clusters in all techniques other than SDC. These notifications will be proven by measurements in the following part.

We evaluate the performance of the four techniques for a 200 × 200 m^2 network by measuring a set of parameters at different number of deployed nodes ranging from low-density (which tends to be non-uniform) to high-density (which tends to be uniform) networks. These parameters are:

(1) The mean and the standard deviation of clusters sizes.

$$E_{total} = N \cdot E_{elec} L + N.\varepsilon_{fs} \cdot L \cdot d_{CH}^2 + N.E_{DA} \cdot L + N.E_{elec} \cdot L + K \cdot \varepsilon_{mp} \cdot L \cdot d_{BS}^4 \tag{13}$$

$$E_{total} = L\left(2N \cdot E_{elec} + N \cdot E_{DA} + N \cdot \varepsilon_{fs} \cdot d_{CH}^2 + K \cdot \varepsilon_{mp} \cdot d_{BS}^4\right) \tag{14}$$

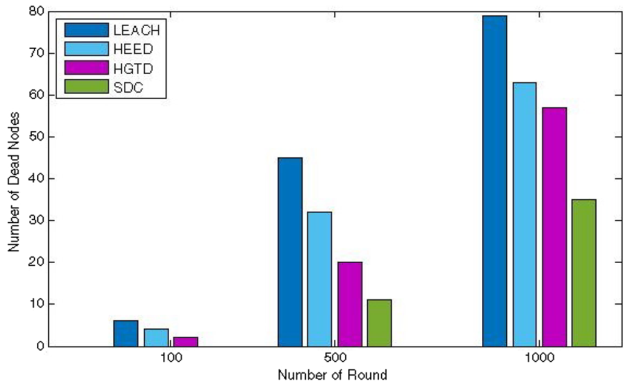

Fig. 8 The number of dead nodes during the rounds of the communication (150 nodes)

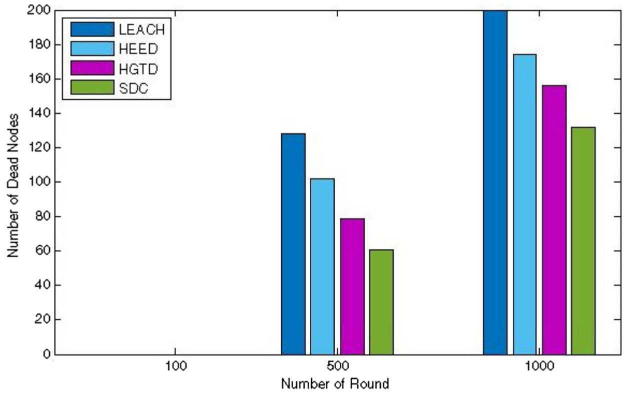

Fig. 9 The number of dead nodes during the rounds of the communication (300 nodes)

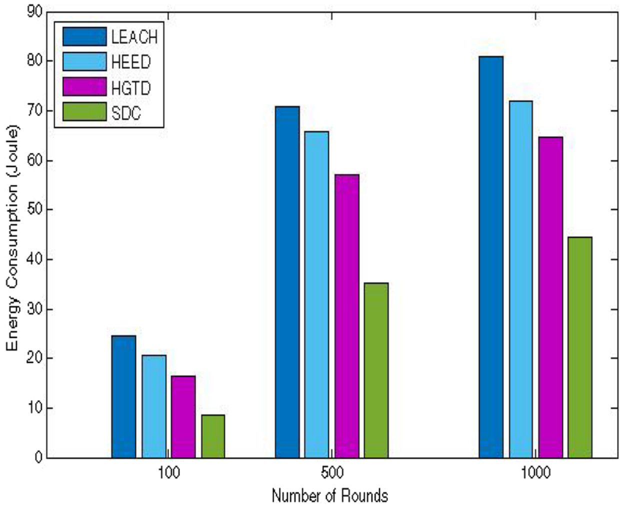

Fig. 10 Energy consumption during the rounds of the communication (150 nodes)

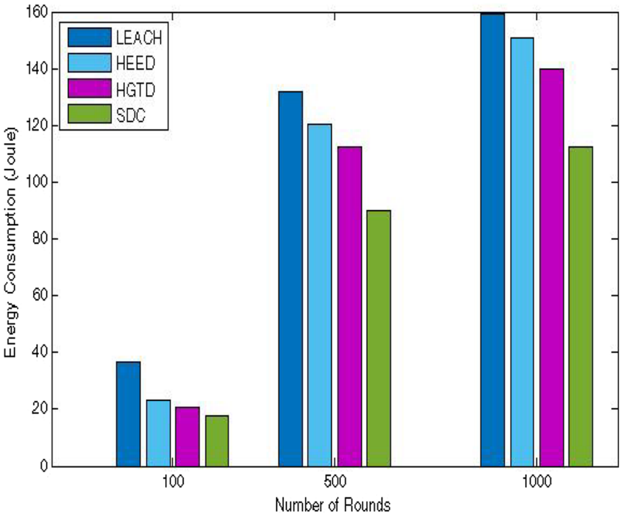

Fig. 11 Energy consumption during the rounds of communication (300 nodes)

(2) The mean and the standard deviation of the energy consumption inside clusters.
(3) The total energy dissipation at different rounds.
(4) The number of dead nodes at different rounds.
(5) The total lifetime of the network for a different number of nodes.

The first and the second parameters are measured to prove the achieved load balance of the SDC protocol. we measure the mean cluster size and the standard deviation between the sizes of the constructed clusters in five scenarios where a 100, 150, 200, 250, and 300 nodes are deployed. We also measure the mean energy consumption inside each cluster and the standard deviation between the energy consumption values of clusters. These values are shown in Tables 2 and 3.

In Table 2, it appears that the sizes of the SDC clusters have a little dispersion, i.e. clusters have approximately similar sizes. This is due to the communication range adaptation step. On the contrary, LEACH, HEED, and HGTD suffer from a high dispersion between the sizes of their clusters which leads to unbalanced energy consumption between clusters. We also notice that the density of the network does not have a noticeable effect on the sizes of the resulting clusters in LEACH, HEED, and HGTD. For example, the mean cluster size is approximately 20 in LEACH protocol even if a 100, 150, 200, 250, or 300 nodes are deployed. The same situation in HEED and HGTD with approximately an 18 and 16 mean cluster size respectively in different density deployments. However, the mean cluster size in SDC increases with the increase of the network density. In LEACH, HEED, and HGTD, the large cluster size in a low-density network leads to high intra-cluster communication cost because distances between nodes are expected to be high. For this reason, these techniques require high network density to achieve acceptable performance, however, the efficiency of the SDC protocol is the same regardless of the network density.

We measure the mean energy consumption inside each cluster in Table 3. We find that the energy consumption of the SDC clusters also has smaller dispersion than other techniques. The unbalanced energy consumption in LEACH, HEED, HGTD protocols causes fast energy dissipation and therefore the fast death of CHs inside large clusters. We also notice that at the same deployment scenario, the SDC protocol achieves smaller mean energy consumption inside its clusters than other techniques. For example, if 200 nodes are deployed, a mean energy consumption equals to 8.93, 11.96, 12.84, and 16.18 J for SDC, HGTD, HEED, and LEACH respectively. This means that SDC achieves lower

Table 4 The saving rate of the SDC protocol energy consumption over LEACH, HEED, and HGTD protocols

Protocol	150 nodes			300 nodes		
	100 round	500 round	1000 round	100 round	500 round	1000 round
LEACH	65.23464	50.38444	44.95216	51.68562	31.50752	29.33091
HEED	58.92205	46.7253	38.08713	23.18781	24.95918	25.27592
HGTD	48.49891	38.43833	31.02617	15.26391	19.76575	19.52299

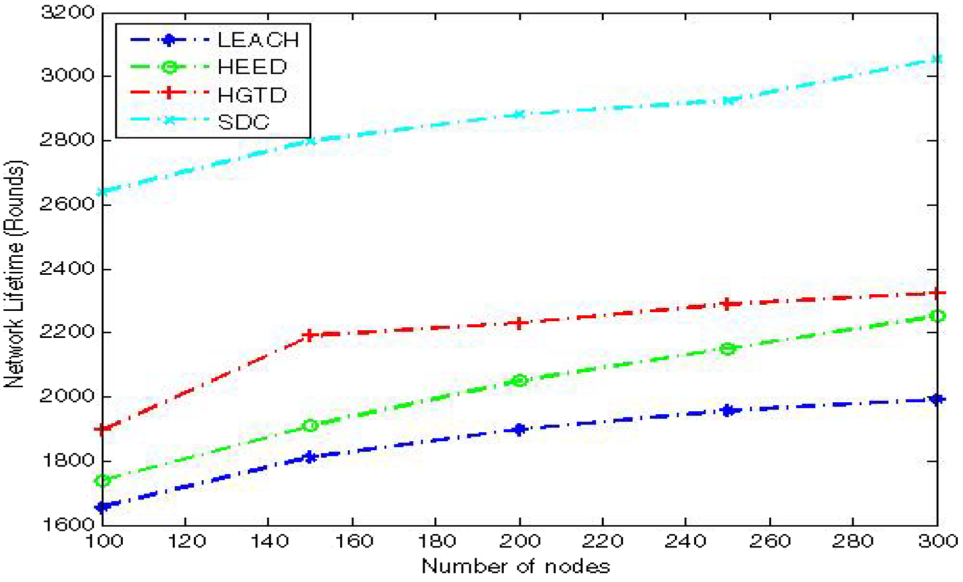

Fig. 12 The lifetime of the network at different number of nodes

intra-cluster communication cost because CHs are highly centered between members of the cluster.

In Figs. 8 and 9, we preview the number of dead nodes at different rounds for a 150 and 300 deployed nodes respectively. We notice that the number of dead nodes in the SDC protocol is less than other techniques at the same time. For example, after 500 round in a network with 150 deployed nodes, the SDC protocol achieves about 76, 66, and 45% decrease rate for the number of dead nodes than LEACH, HEED, and HGTD respectively. And about 34, 24, and 15% decrease rate in case of 300 deployed nodes. From these values, we can also notice that the SDC protocol outperform other techniques in case of low-density networks.

In Figs. 10 and 11, we show the total energy consumption in the network for a 150 and 300 deployed nodes respectively. These values are calculated according to Eq. (13). The SDC protocol achieves a less energy consumption rate than other techniques at the same round of communication.

The values of the achieved saving rate of energy consumption by the SDC protocol over other techniques are shown in Table 4 for the two networks.

According to the saved energy during the rounds of the communication, the SDC protocol achieves longer lifetime than other techniques. In Fig. 12, we measure the total lifetime for a 100, 150, 200, 250, and 300 deployed nodes as the number of rounds till the last node dead. We notice that the SDC protocol achieves significant longer lifetime with about 36, 32, and 24%, when 150 nodes are deployed, and 34, 30, and 23% increase when 300 nodes are deployed than LEACH, HEED, and HGTD respectively.

From all the measured results, we conclude that heuristic techniques which depend on random probability such as LEACH, HEED, and HGTD are not efficient. All of these techniques suffer from a significant dispersion in the sizes of the constructed clusters. It also does not produce even cluster distribution in all cases. Therefore, its performance clearly decreases when the number of nodes decreases. The initial spatial analysis performed in the SDC protocol helps to provide more uniform clusters distribution with a guarantee of load-balancing regardless of the network density.

10 Conclusion and Future Work

In this paper, the effect of randomness in randomly distributed wireless sensor networks is studied. A global, distributed clustering technique named Spatial Density-based Clustering (SDC) is proposed. It depends on performing an initial spatial analysis for the distribution of nodes in the network. This analysis helps to differentiate between dense and sparse sub-regions of the network. According to this analysis, mostly centered nodes in each sub-region are selected as CHs. These nodes are also allowed to update its communication range based on the type of its sub-region to produce balanced clusters all over the network. In the subsequent rounds of the SDC protocol, the residual energy besides to the location of the CH is considered while selecting new CHs. Compared to heuristic techniques such as LEACH, HEED, and HGTD which depend on the random selection of CH, SDC provides even distribution of produced clusters. Unlike these techniques, the number of SDC clusters is controlled by the initial distribution of the total number of nodes in the network. Extensive simulations were conducted, and results show that the SDC protocol significantly improves the energy consumption rate and therefore extends the network lifetime. Results show that the SDC protocol performs well in low and high-density networks. Other techniques need high-density networks to achieves acceptable performance.

The concept of using spatial analysis presented in this paper can be used in solving other problems in WSNs such as hole detection and target tracking. The topological structure defined after the spatial analysis, which helps us to define sparse and dense sub-regions, also defines empty sub-regions. These sub-regions are probably suffering from coverage holes. In addition to that, the defined boundary cycles between nodes can help to track moving targets in the network.

Appendix: Communication Range Calculation for Sub-Regions Borders Discovery in Randomly Deployed WSNs

In this section, we calculate the suitable communication range value that can be used by all nodes in a randomly distributed network which enables them to detect the boundaries of different densities sub-regions. This value must yield enough number of boundary nodes which can border each sub-region. When we use a high communication range value, the resulting boundary nodes will be the nodes located at the border of the whole area. By minimizing this value, more boundary nodes appear which helps in defining sub-regions of different densities in the network.

According to the interior condition stated in the sub-regions border discovery step of the SDC protocol, each node needs to recognize at least three. To determine the required communication range value for this, we use the probability equation proposed in [40]. The author utilized the spatial distribution of the network and the communication range to define the degree of the node. He derived an equation which helps to calculate the required communication range that ensures a specific degree $d(s)$ for any node s. The equation was derived for a network with N nodes randomly distributed in a simulation area $A = W \times H$. He assumed that nodes have a uniform random distribution, so for large N, a constant node density can be defined as $\rho = \frac{N}{A}$. For all sensor node that has a communication range r covers an area: $A = \pi r^2$, the probability that each node has at least n neighbors, i.e. nodes has a minimum degree $d_{min} \geq n$, is given by:

$$P\left(d_{min} \geq n\right) = \left(1 - \sum_{i=0}^{n-1} \frac{\left(\rho\pi r^2\right)^i}{i!} \cdot e^{-\rho\pi r^2}\right)^N \tag{15}$$

In the SDC protocol, the probability that the number of neighbor nodes is greater than or equal to 3 must be maximized for a number of N nodes and area size $(W \times H)$, where r changes between 1 to the maximum value of (W, H). To do that, $P\left(d_{min} \geq 3\right)$ must be greater than a specified threshold (Th) as follows:

$$\left(1 - \sum_{i=0}^{2} \frac{\left(\rho\pi r^2\right)^i}{i!} \cdot e^{-\rho\pi r^2}\right)^N > Th \tag{16}$$

$$\sum_{i=0}^{2} \frac{(\rho\pi r^2)^i}{i!} \cdot e^{-\rho\pi r^2} > 1 - \sqrt[N]{Th}. \tag{17}$$

Table 5 Values of initial communication range value for different network sizes

n	$W=H$	$r_{initial}$
50	400	56
50	600	85
50	800	113
100	400	40
100	600	60
100	800	80

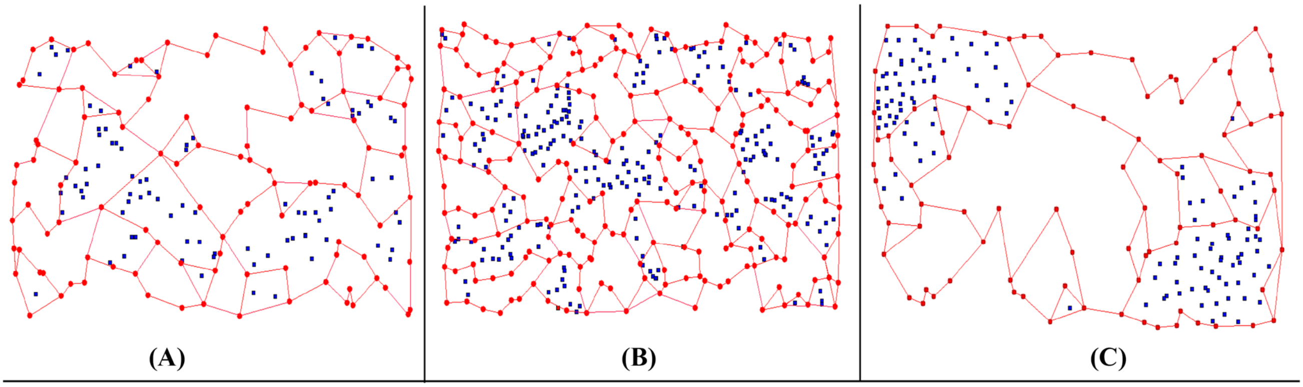

Fig. 13 Topological structure of different random deployment scenarios under the $r_{initial}$ value. **a** Low-density uniformly distributed network, **b** High-density uniformly distributed network, **c** non-uniformly distributed network

The optimal value of *Th* is 1 to 100% achieve the event of having a number of neighbours greater than or equals 3. Therefore, a close value to 1 has been chosen for *Th* which is 0.999999. Equation (14) supposes the network has homogeneous or constant density value all over the network. Practically, in random distribution, this can be achieved by using an extremely large number of nodes to cover all points of the area. When the number of deployed nodes does not achieve coverage for all points of the area, i.e. no constant density value can be measured all over the network, the average network degree value is most appropriate to be used to express the network density. Therefore, we make a mediation for the resulting communication range value calculated from Eq. (16) to achieve an average network degree equal 3. We recall this value as $\mathrm{r}_{\mathrm{initial}}$ which is the initial communication range to be used by all nodes in the SDC protocol. Table 5 shows a sample of $\mathrm{r}_{\mathrm{initial}}$ values to be applied for different network areas and different number of nodes in the SDC protocol.

To prove that the calculated communication range value correctly borders the sub-regions of different densities in the network, we show three different deployment scenarios in a $500 \times 300\,m^2$ network area. The result is shown in Fig. 13. The first scenario (a) presents a uniformly deployed low-density network using 250 nodes. The second scenario (b) presents a uniformly deployed high-density using 450 nodes. And the last scenario (c) presents a non-uniformly deployed network using 300 nodes. After deployment, nodes adapt their communication range to the $r_{initial}$ value which equals 27 m, 20 m, and 25 m for the three scenarios respectively. After that, the interior condition is tested and boundary nodes are colored with red. In this figure, we allow boundary node to connect with other boundary neighbours to produce boundary cycles around defined sub-regions of the network. We can notice that the produced boundary cycles can correctly define the distribution of dense and sparse sub-regions for the three cases.

References

1. D. Ko, Y. Kwak and S. Song, Real time traceability and monitoring system for agricultural products based on wireless sensor network, *International Journal of Distributed Sensor Networks*, Vol. 10, p. 832510, 2014.
2. B. Qiao, K. Ma, An enhancement of the ZigBee wireless sensor network using bluetooth for industrial field measurement, in: International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications (IMWS-AMP), IEEE MTT-S, 2015, pp. 1–3.
3. A. N. Knaian, A wireless sensor network for smart roadbeds and intelligent transportation systems, in, PhD thesis, Citeseer, 2000.
4. S.S.N. Dessai, Development of Wireless Sensor Network for Traffic Monitoring Systems, *International Journal of Reconfigurable and Embedded Systems (IJRES)*, Vol. 3 2014.
5. H. Grindvoll, O. Vermesan, T. Crosbie, R. Bahr, N. Dawood, G. M. Revel, A wireless sensor network for intelligent building energy management based on mulit communication standards-A case study, *Journal of Information Technology in Construction*, 2012.
6. T. A. Nguyen and M. Aiello, Energy intelligent buildings based on user activity: A survey, *Energy and Buildings*, Vol. 56, pp. 244–257, 2013.
7. M. Aminian and H. Naji, A hospital healthcare monitoring system using wireless sensor networks, *J. Health Med. Inform*, Vol. 4, p. 121, 2013.
8. M.-T. Vo, T. T. Nghi, V.-S. Tran, L. Mai, C.-T. Le, Wireless sensor network for real time healthcare monitoring: network design and performance evaluation simulation, in: 5th International Conference on Biomedical Engineering in Vietnam, Springer, 2015, pp. 87–91.
9. K. K. Khedo, R. Perseedoss, A. Mungur, A wireless sensor network air pollution monitoring system, *International Journal of Wireless & Mobile Networks*, pp. 31–45, 2010.
10. X. Jiang, G. Zhou, Y. Liu, Y. Wang, Wireless sensor networks for forest environmental monitoring, in: 2nd International Conference on Information Science and Engineering (ICISE), IEEE, 2010, pp. 2514–2517.
11. G. A. Sánchez-Azofeifa, C. Rankine, M. M. d. E. Santo, R. Fatland, M. Garcia, Wireless Sensing Networks for Environmental Monitoring: Two case studies from tropical forests, in: IEEE 7th International Conference on E-Science (e-Science), IEEE, 2011, pp. 70–76.
12. J. Valverde, V. Rosello, G. Mujica, J. Portilla, A. Uriarte, T. Riesgo, Wireless sensor network for environmental monitoring: application in a coffee factory, *International Journal of Distributed Sensor Networks*, 2012.
13. M. Srbinovska, C. Gavrovski, V. Dimcev, A. Krkoleva and V. Borozan, Environmental parameters monitoring in precision agriculture using wireless sensor networks, *Journal of Cleaner Production*, Vol. 88, pp. 297–307, 2015.
14. J. Li, E. Xu, Development on Smart Agriculture by Wireless Sensor Networks, in: Proceedings of 1st International Conference on Industrial Economics and Industrial Security, Springer, 2015, pp. 41–47.
15. A.-M. Badescu and L. Cotofana, A wireless sensor network to monitor and protect tigers in the wild, *Ecological Indicators*, Vol. 57, pp. 447–451, 2015.
16. I. E. Radoi, J. Mann, D. Arvind, Tracking and monitoring horses in the wild using wireless sensor networks, in: 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) IEEE, 2015, pp. 732–739.
17. A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang, V. Naik, V. Mittal, H. Cao, M. Demirbas and M. Gouda, A line in the sand: a wireless sensor network for target detection, classification, and tracking, *Computer Networks*, Vol. 46, pp. 605–634, 2004.
18. S. M. Diamond, M. G. Ceruti, Application of wireless sensor network to military information integration, in: 5th IEEE International Conference on Industrial Informatics, IEEE, 2007, pp. 317–322.
19. S. H. Lee, S. Lee, H. Song, H. S. Lee, Wireless sensor network design for tactical military applications: remote large-scale environments, in: Military Communications Conference MILCOM IEEE, 2009, pp. 1–7.
20. W. R. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in: Proceedings of the 33rd annual Hawaii international conference on System sciences, IEEE, 2000, pp. 10.
21. O. Younis, S. Fahmy, Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach, in: Twenty-third

AnnualJoint Conference of the IEEE Computer and Communications Societies INFOCOM IEEE, 2004.

22. M. Ye, C. Li, G. Chen, J. Wu, EECS: an energy efficient clustering scheme in wireless sensor networks, in: PCCC 2005. 24th IEEE International Performance, Computing, and Communications Conference, IEEE, 2005, pp. 535–540.
23. L. Yang, Y.-Z. Lu, Y.-C. Zhong, X.-G. Wu and S.-J. Xing, A hybrid, game theory based, and distributed clustering protocol for wireless sensor networks, *Wireless Networks*, Vol. 22, pp. 1007–1021, 2016.
24. W. B. Heinzelman, A. P. Chandrakasan and H. Balakrishnan, An application-specific protocol architecture for wireless microsensor networks, *IEEE Transactions on Wireless Communications.*, Vol. 1, pp. 660–670, 2002.
25. A. Rahmanian, H. Omranpour, M. Akbari, K. Raahemifar, A novel genetic algorithm in LEACH-C routing protocol for sensor networks, in: Electrical and Computer Engineering (CCECE), 2011 24th Canadian Conference on, IEEE, 2011, pp. 001096–001100.
26. N.A. Latiff, C C. Tsimenidis, B.S. Sharif, Energy-aware clustering for wireless sensor networks using particle swarm optimization, in: 2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, IEEE, 2007, pp. 1–5.
27. G. Zhou, T. He, S. Krishnamurthy, J.A. Stankovic, Impact of radio irregularity on wireless sensor networks, in: Proceedings of the 2nd international conference on Mobile systems, applications, and services, ACM, 2004, pp. 125–138.
28. L.-H. Yen, C. W. Yu and Y.-M. Cheng, Expected k-coverage in wireless sensor networks, *Ad Hoc Networks*, Vol. 4, pp. 636–650, 2006.
29. C.T. Reviews, e-Study Guide for: Introducing Geographic Information Systems by Michael Kennedy, ISBN 9780470398173, Cram101, 2012.
30. Y. Wang, J. Gao, J.S. Mitchell, Boundary recognition in sensor networks by topological methods, in: Proceedings of the 12th annual international conference on Mobile computing and networking, ACM, 2006, pp. 122–133.
31. M. I. Ham and M. A. Rodriguez, A boundary approximation algorithm for distributed sensor networks, *International Journal of Sensor Networks*, Vol. 8, pp. 41–46, 2010.
32. I. Khan, H. Mokhtar, M. Merabti, A survey of boundary detection algorithms for sensor networks, in: Proceedings of the 9th Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting, 2008.
33. S. Shukla and R. Misra, Angle based double boundary detection in wireless sensor networks, *Journal of Networks*, Vol. 9, pp. 612–619, 2014.
34. I. J. Fialho and G. J. Balas, Design of nonlinear controllers for active vehicle suspensions using parameter-varying control synthesis, *Vehicle System Dynamics*, Vol. 33, pp. 351–370, 2000.
35. T.S. Rappaport, Wireless communications: principles and practice, Prentice Hall PTR New Jersey, 1996.
36. D. Zhixiang, Q. Bensheng, Three-layered routing protocol for WSN based on LEACH algorithm, in: IET Conference on Wireless, Mobile and Sensor Networks, (CCWMSN07)., IET, 2007, pp. 72–75.
37. N. D. Tan, L. Han, N. D. Viet and M. Jo, An improved LEACH routing protocol for energy-efficiency of wireless sensor networks, *SmartCR*, Vol. 2, pp. 360–369, 2012.
38. C. Fu, Z. Jiang, W. Wei and A. Wei, An Energy Balanced Algorithm of LEACH Protocol in WSN, *International Journal of Computer Science*, Vol. 10, pp. 354–359, 2013.
39. S. Kohli, Implementation of homogeneous LEACH protocol in three dimensional wireless sensor networks, *International Journal of Sensors Wireless Communications and Control*, Vol. 5, 2015.
40. C. Bettstetter, On the minimum node degree and connectivity of a wireless multihop network, in: Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing, ACM, 2002, pp. 80–91.

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

Walaa Abdellatief received the B.Sc. and M.Sc. degrees in computers and information from Faculty of Computers and Information, Menofia University, Egypt, in 2005 and 2010, respectively. Currently, she is a Ph.D. student in Information Technology Department, Faculty of Computers and information, Menofia University, Egypt. Her research interest includes computer networks, wireless networks and technologies, communication protocols, and sensor networks.

Osama Youness received the M.Sc. degrees in computers and information from Faculty of computers and information, Menofia University, Egypt, in 2006. He received his Ph.D. in Computing Science from Newcastle University, UK, in 2013. His thesis work focused on modeling, analysis and optimization of mobile ad hoc networks. His main research interests lie with the performance engineering of computer systems, wired and wireless networks, mobile communication, cloud computing, information and network security, and artificial intelligence.

Hatem Abdelkader obtained his B.S. and M.Sc. (by research) both in Electrical Engineering from the Alexandria University, Faculty of Engineering, Egypt in 1990 and 1995 respectively. He obtained his Ph.D. degree in Electrical Engineering also from Alexandria University, Faculty of Engineering, Egypt, in 2001 specializing in neural networks and applications. He is currently a professor in Information systems department, Faculty of Computers and Information, Information systems department, Faculty of Computers and Information, Menofia University, Egypt since 2004. He has worked on a number of research topics and consulted for a number of organizations. He has contributed more than 30+ technical papers in the areas of neural networks, Database applications, Information security and Internet applications.

Mohiy Hadhoud obtained his Ph.D. from the Department of Electronics and Computer Science, University of Southampton, Southampton, England, in 1987. In 2001 he become a Professor of Digital communications and signal processing. His research interests are in the areas of digital signal processing, adaptive filtering and image processing, multimedia applications, Information security and data hiding, face recognition, and image and video colorization. He, in cooperation with colleagues and students, has published more than 182 papers in international journals, conferences, local journals and local conferences.