International Journal of Pervasive Computing and Communications

An efficient clustering algorithm for wireless sensor networks

Mariam Alnuaimi Khaled Shuaib Klaithem Alnuaimi Mohammed Abed-Hafez

Received 4 February 2015
Revised 16 April 2015
8 May 2015
Accepted 30 June 2015

An efficient clustering algorithm for wireless sensor networks

Mariam Alnuaimi, Khaled Shuaib and Klaithem Alnuaimi
College of Information Technology, United Arab Emirates University, Al Ain, United Arab Emirates, and
Mohammed Abed-Hafez
College of Engineering-Electrical Engineering Department, United Arab Emirates University, Al Ain, United Arab Emirates

Abstract

Purpose – This paper aims to propose a new node energy-efficient algorithm with energy threshold to replace cluster heads. The proposed algorithm uses node ranking to elect cluster heads based on energy levels and positions of the nodes in reference to the base station (BS) used as a sink for gathered information. Because the BS calculates the number of rounds a cluster head can remain for as a cluster head in advance, this reduces the amount of energy wasted on replacing cluster heads each round which is the case in most existing algorithms, thus prolonging the network lifetime. In addition, a hybrid redundant nodes duty cycle is used for nodes to take turn in covering the monitored area is shown to improve the performance further.

Design/methodology/approach – Authors designed and implemented the proposed algorithm in MATLAB. The performance of the proposed algorithm was compared to other well-known algorithms using different evaluation metrics. The performance of the proposed algorithm was enhanced over existing ones by incorporating different mechanisms such as the use of an energy-based threshold value to replace CHs and the use of a hybrid duty-cycle on nodes.

Findings – Through simulation, the authors showed how the proposed algorithm outperformed PEGASIS by 15 per cent and LEACH by almost 70 per cent for the network life-time criterion. They found that using a fixed pre-defined energy threshold to replace CHs improved the network lifetime by almost 15 per cent. They also found that the network lifetime can be further improved by almost 7 per cent when incorporating a variable energy threshold instead of a fixed value. In addition to that, using hybrid-redundant nodes duty-cycle has improved the network lifetime by an additional 8 per cent.

Originality/value – The authors proposed an energy-efficient clustering algorithm for WSNs using node ranking in electing CHs and energy threshold to replace CHs instead of being replaced every round.

Keywords Clustering protocols, Energy efficiency protocols, Load balancing, Routing protocols, Wireless sensor networks

Paper type Research paper

1. Introduction

Recent improvement in hardware electronic technology enabled manufacturers to develop low-cost, low-power and small-size sensors (Akyildiz *et al.*, 2002; Baronti *et al.*, 2007; Alnuaimi *et al.*, 2011). Hundreds and thousands of these sensors are deployed as wireless sensor networks (WSNs) serving many applications based on the specific

International Journal of Pervasive Computing and Communications
Vol. 11 No. 3, 2015
pp. 302-322

1742-7371
DOI 10.1108/IJPCC-02-2015-0007

requirements of each one (Alnuaimi *et al.*, 2006, 2012; García Villalba *et al.*, 2009). Nowadays, there are many applications of sensor networks covering different fields such as agriculture, medicine, military, environment monitoring, toys, intrusion detection, motion tracking, machine malfunction and many others (Baronti *et al.*, 2007). Sensors can be deployed to continuously report environmental data for long periods of time. WSNs applications are evolving every day for information gathering to better help monitor and control components of such applications. Below that, we shed light on some of the most recent applications of WSNs.

1.1 Smart power grid systems

Smart power grid is an efficient and reliable automation service for electricity flow. One of the recent applications of WSNs is smart grid automation. WSNs are used to capture and analyze data related to power usage, power delivery, power generation and power disturbances and outages. Sensors are used to know energy usage frequency, phase angle and the values of voltage to help utility companies manage electricity in an efficient way. WAMR (Depuru *et al.*, 2011) which stands for wireless automatic meter reading is an example for such applications. WAMR collects real-time energy consumption of the customers and provides the customer with archived old readings. It can also control light, air conditioning, heater and other devices of the house to help customers to manage electricity usage in an efficient way.

1.2 Smart habitat monitoring

Ecologists study the wildlife such as animal origin, migration, behavior, diseases and life processes and their environment. Habitat monitoring applications provide ecologists with data on surrounding environmental conditions such as weather conditions that are affecting birds' migration. They are used in helping settle large scale land use issues affecting animals, plants and people (Szewczyk *et al.*, 2004). Kaur and Sharma (2011) proposed an approach for monitor activities of birds. They proposed this approach to track 350 species of exotic bird migration from Siberia during winter to India. They implemented habitat monitoring system where sensors are attached on bodies of the birds to track each bird's activity and record it.

1.3 Smart cloud

Could computing has gained great attention in the recent years due to its wide deployment and offered services. Cloud services imply using the Internet as a huge repository or a workspace. People can access the Internet anytime anywhere. Song *et al.* (2010) proposed an intelligent smart cloud model. This model provides customized services to users by personalizing the contents through smart processing based on the user behavior. In this model, aspects of the user's behavior were collected by sensors mounted on their devices such as mobile phones and smart tablets.

1.4 Smart health-care delivery

Smart health and care delivery applications are used for patient monitoring and care in remote sites like monitoring patients' facial expression, respiratory conditions or movement and forward these images to doctors in distant hospitals to make better diagnosis. According to Sha *et al.* (2010), a health-care sensor periodically captures vital signs information (e.g. body temperature, blood pressure) and sends it to a gateway device. Once the information is processed by the gateway, it is forwarded to doctors to help them make an initial diagnosis.

In general, a WSN is a collection of nodes with sensing, computation and wireless communication capabilities (Baronti *et al.*, 2007). These nodes or terminals communicate with each other by forming a network of nodes and maintaining connectivity in a distributed way. An unstructured WSN contains a dense collection of sensor nodes that are deployed in an *ad hoc* manner into an area of choice. In such an environment, network maintenance such as managing connectivity and detecting failures is difficult due to the large number of deployed nodes and the large coverage areas. However, such deployments are critical to have in certain harsh environments where the deployment of pre-planned networks can be difficult if not impossible. In a structured WSN, the deployment of the sensor nodes is pre-planned. Typically, a WSN has one or more sinks (or base stations [BS]) which collect data from sensors within the WSN. These sinks are considered the gateways through which a WSN interacts with the outside world.

The denseness and random distribution of WSNs makes it quite difficult to replace or recharge node batteries, especially in applications such as: disasters recovery areas, environment monitoring, border monitoring, battle fields, under water sensing, oil fields and many others. Therefore, energy efficiency and management is a major design goal in these networks.

2. Review of related work

Research related to WSNs is not new, and several related problems have been exposed and addressed within the past few years. This research effort has been categorized into three main areas: clustering algorithms, data dissemination techniques and routing protocols (Radi *et al.*, 2012). In this paper, we are focusing our attention on clustering algorithms; therefore, the introduced related work will mainly address this research area.

Heinzelman (2000) introduced a hierarchical clustering algorithm for sensor networks, the Low Energy Adaptive Clustering Hierarchy (LEACH) was introduced. The idea is to form clusters of the sensor nodes based on the received wireless signal strength. Local cluster heads (CHs) are used by members of the cluster as routers to the sink. The essence of this approach is to save node energy consumption as the transmissions of gathered data to the sink will only be done by CHs rather than by all sensor nodes. LEACH randomly selects a number of sensor nodes as CHs and then rotates this role to uniformly distribute the energy load among the sensors in the network. Each elected CH broadcast an advertisement message to the rest of the nodes in the network informing them of their new role as CHs. All the non-CH nodes, after receiving this advertisement, choose the cluster to which they want to belong to. This decision is based on the signal strength of the received advertised message.

LEACH uses single-hop routing where each node can transmit directly to the CH which, in turn, transmits directly to the sink, regardless of the distance. This technique might work well in dense WSNs but not in large-scale networks with large distances between nodes due to a direct proportional energy consumption relationship with distance. LEACH elects CHs randomly regardless of their energy level and thus it is not suitable for networks deployed at a large scale. Furthermore, the idea of dynamic clustering brings extra overhead, e.g. head changes, advertisements, etc., which may diminish any gain realized in energy consumption. It also assumes that nodes always have data to send, and nodes located close to each other have correlated data. In addition, it is not obvious how the number of the predetermined CHs is going to be uniformly distributed through the network. Therefore, there is a possibility that the elected CHs will be more concentrated in one part of the network over other parts. As a consequence of this, some nodes will not have any CHs in their neighborhood and will not be covered. Finally, the protocol assumes that all nodes begin with the same amount of energy capacity in each election round, assuming that being a CH consumes approximately the same amount of energy for each node. To mitigate some of these problems Multi-hop LEACH was proposed by Heinzelman *et al.* (2000). Multi-hop LEACH is another extension of the LEACH routing protocol to increase energy efficiency through using multi-hopping to reach the BS of the WSN (Biradar *et al.*, 2011; Radi *et al.*, 2012). Cluster heads receive data from all nodes at a single-hop and send to the BS through intermediate cluster heads. However, some of the aforementioned problems are still considered open research issues and have not been solved yet.

Grid Schemes Power-efficient GAthering in Sensor Information Systems (PEGASIS) approach (Lindsey and Raghavendra, 2002) which is a chain-based algorithm showing an improvement over the LEACH protocol. PEGASIS forms chains from sensor nodes instead of forming multiple clusters. Only one node is selected from that chain to transmit to the BS or sink. Gathered data move from a node to other neighboring nodes, aggregated and eventually sent to the BS. Each node uses the signal strength to measure the distance to all neighboring nodes, and then adjusts the signal strength so that only one node can be heard. Therefore, the chain will consist of those nodes that are closest to each other and form a path to the BS. The aggregated form of data will be sent to the BS by any node in the chain and the nodes in the chain will take turns in sending to the BS. Unlike LEACH, PEGASIS avoids cluster formation and uses only one node in a chain to transmit to the BS instead of using multiple nodes, thus saving energy consumed by the rest of the nodes within the network. However, PEGASIS adds more delay for distant nodes on the chain, especially if the wrong direction to the base-station is taken. In addition to that, the chain leader can become a bottleneck for the whole chain and it is also assuming all nodes in the network can reach the base-station.

In addition to the above two, several other research issues have been considered by researchers recently related to large-scale WSNs. Du *et al.* (2011) proposed a mixed unequal clusters size algorithm (MNUC) to prolong the life time of the network. It is addressing the problem of hot spot where nodes have to do more processing- and transmission-related work compared to other parts of the network. Therefore, their energy will be drained more quickly than the others. The idea of the algorithm is to form clusters with unequal sizes. Nodes closer to the BSs will be gathered into smaller sizes clusters and nodes that are far will have bigger clusters sizes. Nodes that are closer to the BS will be used more, but the transmission's ranges will be less.

Universal LEACH (ULEACH) (Kumar *et al.*, 2012) was proposed as an improvement over LEACH. Selection of CHs in ULEACH is based upon initial and residual energy of nodes. Data are sent in a multi-hop approach from farthest node to CHs and from CHs to master CHs. This algorithm incorporates some features of HEED (Younis and Fahmy, 2004) and PEGASIS into LEACH. Although it utilized the multi-hop data transmission approach, it does not take into account the distance of the master CHs from the BS. Therefore, there might be more delay in delivering the data if the master CHs are far from the BS which will also result in an additional transmission cost.

T-LEACH, proposed by Hong *et al.* (2011), is an improvement on LEACH. It is a threshold-based CH replacement scheme for clustering protocols of WSNs. T-LEACH minimizes the number of CH selection by using threshold of residual energy. However, it is still using random head selection process of LEACH without specifying any criteria to choose CHs.

Despite recent achievements in these three areas of research, challenges still exist in these active areas of research, gaining the focus of many researchers who are working on areas such as quality of service, security, energy harvesting and prolonging the network lifetime by conserving energy on deployed nodes. In this paper, a new clustering algorithm based on node ranking is being proposed with energy thresholds. The proposed algorithm prolongs the network lifetime and uses energy thresholds to replace CHs instead of the random replacement methods used by others. The BS calculates the number of rounds a CH can remain through when selecting each CH; this feature is shown to have a positive effect on prolonging the network lifetime, as it reduces the amount of energy wasted on replacing CHs.

This paper is organized as follows. In Section 3, recent WSN applications are introduced. The current challenges of clustering algorithms are discussed in Section 4. The proposed clustering algorithm is described in Section 5. In Section 6, performance evaluation of the proposed algorithm is compared to other existing well-known ones. Section 7, concludes the paper.

3. Current challenges of clustering algorithms

Based on the literature review discussed in Section 2, there are several challenges which need to be considered while clustering a large-scale WSN. These issues are summarized below.

3.1 Selection of CHs

After dividing WSNs into clusters, it is very important to choose the best CH for each cluster. The optimal selection of the CH that is reachable by all member nodes in the cluster increases the life time and reliability of the network. There are several approaches for CH selection, such as selecting the node with the maximum current energy among the cluster members. Another way is to select the node which can be reached by all nodes with the least energy. Moreover, it is necessary to rotate the role of CHs among nodes to avoid overloading a few nodes with more responsibilities than others and deplete their energy fast. There are several approaches for CH rotation. One approach is to use a time stamp to initiate the process of electing another CH. The other approach is to use the remaining energy level to start the process of electing another CH. For example, a CH might trigger a new CH election process if its remaining energy level goes below a specified threshold. Frequent CH rotation results in more clustering overhead and network interruption. On the other hand, less frequent rotation may cause some nodes to die faster than others. The study of the optimal selection and rotation of CHs is essential for prolonging the life time of the network and increasing its reliability (Hong *et al.*, 2011; Younis *et al.*, 2006).

3.2 Cluster size

Most existing clustering protocols assume a fixed cluster communication range in distance, which implies that all clusters have the same physical size. This assumption results in unfair load balancing where CHs that are closer to the observed event will carry more traffic and their energy will be drained faster than distant CHs. According to Shu *et al.* (2005), a larger cluster size is suggested to CHs that are having less data to forward to distribute load evenly among CHs. However, this requires the nodes to know their locations based on the place of the occurred event and the location of the BS. Selecting appropriate clusters' sizes to minimize energy consumption within a WSN, not just based on the communication range, but considering other factors such as the denseness of the WSN, the location of the BS, the application requirement with respect to reliability and the frequency of data collection is still an open research issue.

3.3 Ensuring connectivity

Maintaining connectivity is an important objective of clustering protocols. Every node in a network must be a member of a cluster. It is recommended, as much as possible, that all nodes within a cluster are able to communicate with their CH directly to avoid multi-hoping which is usually results in more energy consumption. However, in certain cases, where the cluster size is larger than the communication range of nodes or when nodes died due to the depletion of their energy, multi-hop communication cannot be avoided. To strike a balance between choosing the most appropriate cluster size while maintaining proper connectivity within each cluster, i.e. intra-cluster communication, is used to indicate the success of cluster formation. There is another type of connectivity called inter-cluster communication which is between different clusters. Two main approaches were proposed in the literature, relaying data through CHs or relaying data through gateways. According to Amis *et al.* (2000), Banerjee and Khuller (2001) and Basagni (1999), nodes on clusters' boundaries are used as gateways to relay data among CHs (shown in Figure 1). Network density has to be sufficiently high to ensure that enough gateways are present at the intersection areas between clusters. On the other hand, according to Younis and Fahmy (2004), Heinzelman *et al.* (2002), the CH overlays data only through CHs (shown in Figure 1 as a dotted line). An advantage of the second relay approach is that it enables all non-cluster nodes to sleep while not sensing or

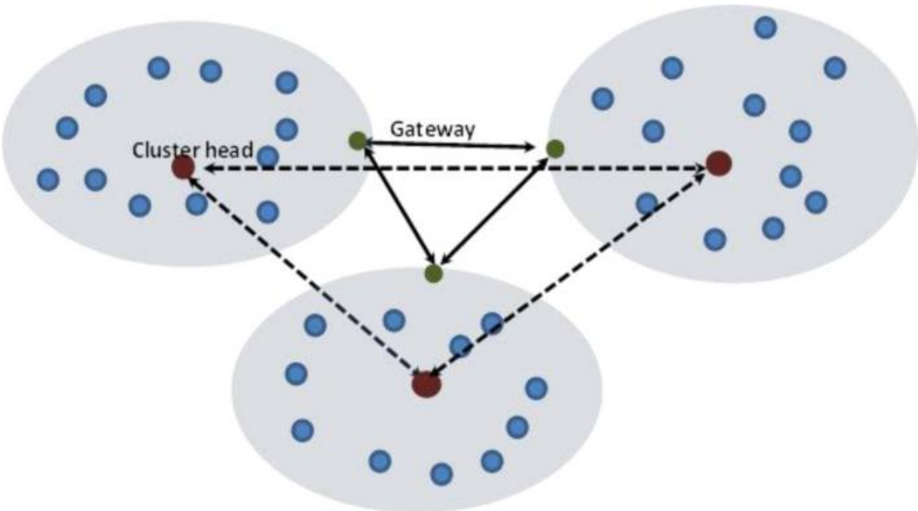

Figure 1.
Routing via gateway nodes and cluster heads

transmitting data. Selecting efficient intra- and inter-cluster transmission ranges to ensure connectivity and prolong the network lifetime is an important issue in clustering which need to be considered when designing a clustering algorithm.

3.4 Clustering the network in the presence of duty-cycle

Allowing sensors to sleep when they are not active contributes significantly to prolonging their battery lifetime. This is because listening consumes a significant amount of energy that is comparable to reception. Therefore, node's duty cycle should be taken into consideration when designing clustering techniques. Incorporating node's duty cycle in the design of the clustering can be done in one of two ways, depending on the type of the application. In the first approach, non-CH nodes can be allowed to sleep when they are not sensing any data or when they are not communicating with their CHs. This approach is appropriate for applications where sensors are sending updates on a periodic predefined time. The second approach is used if the application requires the sensors to continuously monitor the field for unexpected events, then a CH can determine which of its cluster members are sending redundant data and advise them to sleep (Younis *et al.*, 2006).

4. Node ranking clustering algorithm

Because data transmission can account for up to 70 per cent of the power consumed in typical sensor nodes (Kumar *et al.*, 2012), substantial energy can be reduced by reducing the distance traveled and the amount of data transmitted to the BS. Distance of the nodes from the BS and inter-node distances can have a high impact on saving nodes' energy, thus prolonging the network lifetime which can be defined either as the time for the first node to die, the time for the last node to die or the time for a certain percentage of nodes in the WSN to die (Blough and Santi, 2002) Moreover, in dense deployments of sensor nodes in a WSN, nodes can cooperate to send data and, therefore, distribute the consumption of energy between them.

In this paper, we propose a node ranking clustering algorithm (NRCA). The difference between this algorithm and other algorithms is that this algorithm uses a more efficient mechanism to select CHs. It is consider more efficient, as it prolongs the network lifetime further by decreasing communication overheads for the frequent election of CHs which as a result decreases the energy consumed on nodes compared to other algorithms. This is achieved by the proper election and replacement of CHs which involves measuring the distance and current energy level of nodes, using energy thresholds and calculating the number of sensing rounds CHs can serve before being replaced. In this algorithm, nodes are ranked based on their current energy level (En) and their positions (Dn) in reference to the BS. This ranking is used for choosing CHs which are also ranked into levels based on their position, Euclidean distance, from the BS. Therefore, each node is assigned a rank Rn (En, Dn), reflecting its candidacy for being elected as a CH. In the next subsection, we introduce the proposed algorithm in more details.

4.1 Description of NRCA

In most previously proposed clustering algorithms, a node is elected as a CH either randomly or based on having the highest residual energy in a cluster. This selection might lead to inefficiencies (Nikolidakis *et al.*, 2013). For example, and as was previously shown by Nikolidakis *et al.* (2013), node A in Figure 2 has higher residual energy than the other nodes M and S belonging to the same cluster as A. Thus, this node is typically elected as the new CH. As a result, this causes M and S in the same cluster to send data through A to the BS taking a longer path as the location of A is toward the opposite direction from the BS. The additional distance the data need to travel to get to the BS will result in more energy consumption. Also, there can be forgotten nodes or disconnected nodes which are not covered by any of the CHs chosen, due to being far from any reachable CHs. Moreover, the frequent replacement of CHs in each round wastes more energy. These three problems can be avoided in our proposed algorithm where data can be sent through the correct path or direction with respect to the BS and by the BS maintaining a global knowledge of all nodes in the WSN area to ensure that all alive nodes are connected through the proper choice of CHs. In addition, we propose the use of an energy threshold technique in making decisions to replace CHs which prolongs the life time of nodes closer to the BS. This, in effect, prolongs the overall network lifetime as nodes closer to the BS are more critical than those far-away nodes in maintaining connectivity to the sink.

In the proposed algorithm, the BS is placed in a fixed position and has unlimited energy. Thus, no constraints are assumed with regards to power consumption due to data processing and communication. Through the initial step of the below algorithm, the BS becomes aware of the locations of all sensor nodes either via collecting their GPS coordinates or any other mechanism (Koutsonikolas *et al.*, 2010).

The following steps give a description of the algorithm and CHs' selection process:

- Similar to the initial step done by Younis and Fahmy (2004), Lindsey and Raghavendra (2002) and Nikolidakis *et al.* (2013), each node at the set-up phase broadcasts a message of its energy level and location to its neighbors. Therefore, each node sets up a neighbor information table recording the energy levels and positions of its neighbors and broadcast this information to its neighbors. This is conducted by all nodes in the network until information about all nodes in the network is received by the BS. This will provide the BS with a global knowledge of the network, pseudo code is shown in Algorithm 1:

Algorithm 1. The pseudo-code for NRCA

```
if is_the_network_clustered = false
  for every node u ∈ Node-List do
    u advertise its position and its energy level to the BS
    For every node i ∈ Node-List do
      Sort nodes according to their geographical location
      //Partition sorted nodes into groups according to their communication range.
```

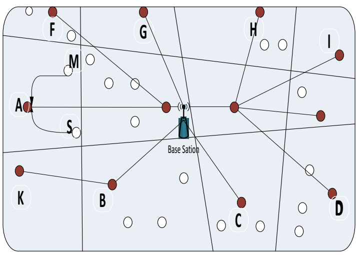

Figure 2. WSN clustering example of sending data to the BS in the wrong direction

```
    If distance between i and i + 1 < communication range then
        add i and i + 1 to cluster_list
        else
        create new cluster_list
        add i + 1
    end if
    i + 1
End for
    for every node u Node-List do
    rank(u) = BS ranks u based on its energy's level (En) and Euclidean position
    (Dn) from the BS and
    end for
    for every node u ∈ Node-List_same _region do
        if (rank(u) > rank(u + 1)) then
        canBeClusterHead = true
        add node to Candidate_Cluster_heads_list
        end if
        u + 1
    end for
    for every cch ∈ Candidate_Cluster_heads_list do
        Candidate Cluster Head are ranked into levels based on their position from
        the BS
    end for
    for every node ch ∈ Cluster_heads_list do
        Calculate number of rounds cch can serve as a cluster head
        Broadcast msgs that its a cluster head
        u joins the ch
```

The BS divides the area into smaller partitions called clusters based on the assumed communication range of the nodes and their positions, i.e. geographical locations, by geographical partitioning or grouping nodes into groups. The size, distance between any two farthest nodes within a cluster, of each cluster should be less than the predefined communication range. Therefore, no node will be out of coverage. Pseudo code is shown in Algorithm 1.

- The BS calculates the number of rounds (a round is a random time slot where CHs election phase and data transmission phase occurred) CHs can serve based on their residual energy and an initial a pre-defined energy threshold, then relays this information to each CH.
- Cluster heads close to the BS will have higher energy threshold value; however, CHs that are farther from the BS will have a low energy threshold value.
- Cluster heads are replaced only when their energy level drops below the pre-defined or calculated energy threshold.
- Cluster heads, which are located the closet to the network BS, are referred to as the first-level CHs. The CHs that are located at more distant positions from the BS are considered as second level, third level […] etc.
- Higher CHs' levels transmit to lower CHs' levels to reach the BS with the least energy consumption.

- If there is a change in the network topology, due to nodes being considered dead or having residual energy below a certain threshold, the BS determines the next appropriate CH in each cluster while considering the new changes.

4.2 Used energy model

In this paper, the energy model adopted is the same used by Younis and Fahmy (2004), Lindsey and Raghavendra (2002), Heinzelman *et al.* (2000) and Nikolidakis *et al.* (2013) and as shown in Table I where Eelec is the radio-dissipated energy which is assigned a value of 50 nJ/bit to run the transmitter or receiver circuitry. The Eamp is the used energy for the transmitting amplifier and assigned a value of 100 pJ/bit/m2. ETx(k, d) is the energy that a node dissipates for the radio transmission of a message of k bits over a distance d and expressed by equation (1):

$$\mathrm{ETx(k)} = \mathrm{Eelec} \times \mathrm{k} + \mathrm{Eamp} \times \mathrm{k} \times \mathrm{d}^2 \tag{1}$$

In the same way, the equation of the energy dissipated by a node for the reception ERx(k) of a message of k bits which is due to running the receiver circuitry Eelec (k) can be expressed by equation (2):

$$\mathrm{ERx(k)} = \mathrm{Eelec} \times \mathrm{k} \tag{2}$$

4.3 Cluster head selection process

After the forming of clusters, the BS assigns a CH for each cluster based on the proposed NRCA. Nodes in each cluster are ranked based on how far they are from the BS and on their current energy level. Nodes with the maximum residual energy and minimum distance will be chosen as a CH based on equations (3) and (4):

$$\mathrm{NodeRanking(En, Dn)} \tag{3}$$

where

$$\mathrm{(Dn(i))} = \mathrm{Min(D(i, BS))},\ \mathrm{(En(i))} = \mathrm{Max(ResidualEnerg)} \tag{4}$$

$$|\mathrm{D(i, BS)}| = \sqrt{(\mathrm{Xi} - \mathrm{Xbs})^2 + (\mathrm{Yi} - \mathrm{Ybs})^2} \tag{5}$$

Notation	Description
$N = 100$	Total number of sensor nodes
Eo = 0.5 J/node	Initial energy of each node
Eelec = 50 nJ/bit	Per bit energy consumption
EDA = 5 nJ/bit	Energy for data aggregation
Eamp = 100 pJ/bit/m^2	Amplifier transmitting energy
Sensing field = 100 × 100 m	Area of the sensing field
Communication range	40 m

Table I. Parameters used in the simulation

Residual (En) is the current energy level of the node i, D(i, BS) is the Euclidean distance of node i to the BS. Given a particular deployment region of interest, Xi and Yi are the X and Y positions of node i, respectively. Xbs and Ybs are the X and Y positions of the BS, respectively.

A CH in each cluster will be changed when its energy level reaches a pre-defined threshold or a calculated value and not every round. This will make it possible for a node, i, to continuously play the role of a CH for multiple rounds and thus save any energy to be wasted for control and exchanged messages used in replacing it:

$$T(i) = \frac{Residual(En(i))}{Average(En)} \times \frac{Average(Dn)}{D(i, BS)} \tag{6}$$

$$Average(Dn) = \frac{\sum_{1=1}^{n} D(i, BS)}{n} \tag{7}$$

$$Average(En) = \frac{\sum_{1=1}^{n} Residual(En(i))}{n} \tag{8}$$

Equation (6) shows how to calculate the energy threshold value used for all nodes. T(i), is calculated based on its residual energy, Residual (En(i)), average residual node energy within its cluster, the Euclidean distance between it and the BS D(i, BS) and the associated average Dn. In the first round, all nodes have the same energy level. Consequently, ranking will depend solely on the distance. If a node is closer to the BS, it has a greater probability of becoming a CH. In the next rounds, the residual energy of each candidate node in the network is different. Therefore, the selection of CHs will depend both on the residual energy and Euclidean distance. According to equation (6), nodes close to the BS will be changed faster as their threshold values will be higher. This is because they are critical to the network and needed more to aggregate the data to the BS. However, nodes that are far from the BS will have a lower threshold and will be changed less frequently. The number of rounds a node, i, can stay as a CH, CountRound (i). is calculated based on the node residual energy and the calculated threshold value as shown in equation (9):

$$CountRound\ (i) = \frac{Residual(En(i)}{T(i)} \tag{9}$$

5. Performance evaluation

To evaluate the performance of the proposed algorithm against two other well-known algorithms (LEACH and PEGASIS), we used MATLAB for simulating the considered algorithms. Table I shows the parameters used in this simulation environment which are a standard parameters used by all researcher in this field. The simulated area is 100 × 100 m. Every node was given an initial energy of 5 J. The energy for data aggregation is 5 nJ/bit. The energy to run the radio is 50 nJ/bit. The amplifier transmitting energy is 100 nJ/bit/m2. In our performance evaluation, we focused our attention on the main two algorithms, LEACH and PEGASIS, which were used as the baseline for all researchers in the field. According to Alnuaimi *et al.* (2013), we showed how PEGASIS outperformed HEED; therefore, HEED was not selected. SEP was also not considered here, as it uses heterogeneous nodes with different initial energy levels. Using simulation, we have considered several metrics to evaluate the performance of NRCA as follow.

5.1 Clusters formation and CHs selection

As we can see from Figure 3, PEGASIS forms a chain starting with the furthest node from the BS. A leader node is elected randomly in each round, and it assumes all nodes can reach the BS. The leader node is the one responsible for transmission all sensed data to the BS in each round. As shown, the leader node is far from the BS, so it consumes more energy to send the data to the BS, especially if it is the furthest node.

Figure 4 shows the cluster formation and CHs election in LEACH. As can be seen, CHs are elected randomly in each round, so a CH can be the farthest node from the BS in its cluster (as shown in Cluster A) or it can be the node with the least energy. In both cases, the election leads to inefficiencies.

On the other hand, Figure 5 shows the NRCA clusters formation and CHs selection. Nodes with highest energy and closest to the BS in each cluster will be selected as CHs. For example, as shown in Figures 3 and 4, for Custer A, the node closest to the BS was chosen as a CH, while, in LEACH, the farthest node in the same cluster was chosen. Therefore, energy consumed to send data to the BS is reduced in NRCA. Moreover, there are no disconnected or forgotten nodes and thus no clusters are formed with only one node.

5.2 Network lifetime

Network lifetime is defined here as the time interval from the time the sensor network starts its operation until the death of the last node in the network. From Figure 6 and

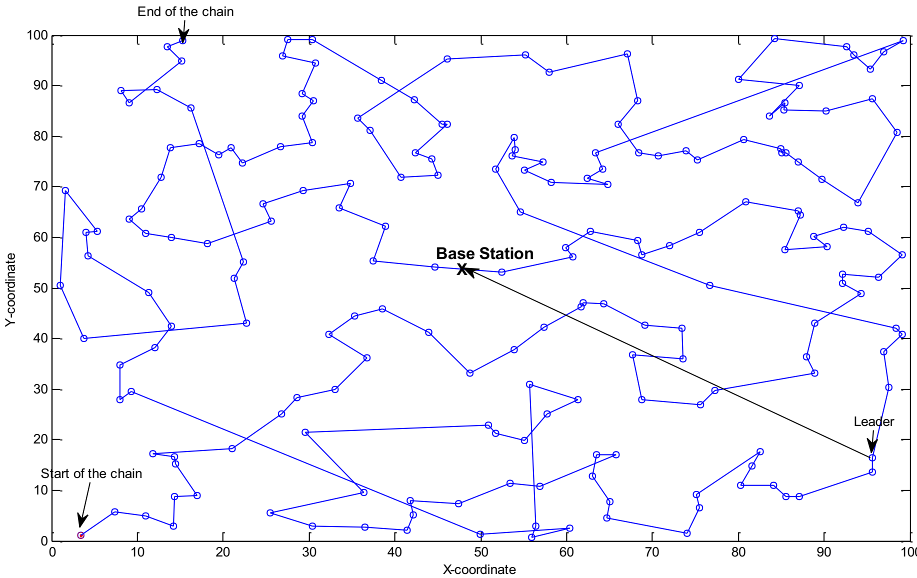

Figure 3.
PEGASIS chain formation

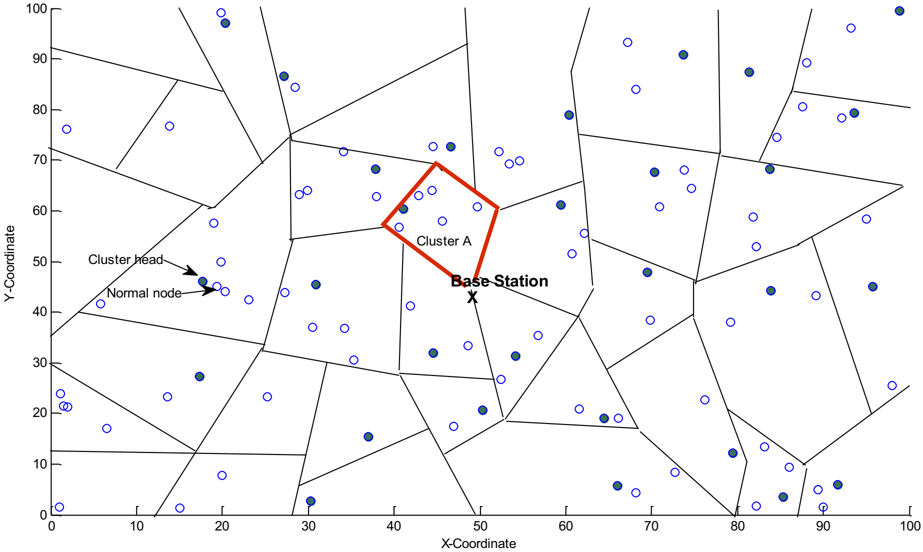

Figure 4.
LEACH clusters formation and cluster heads election at first round

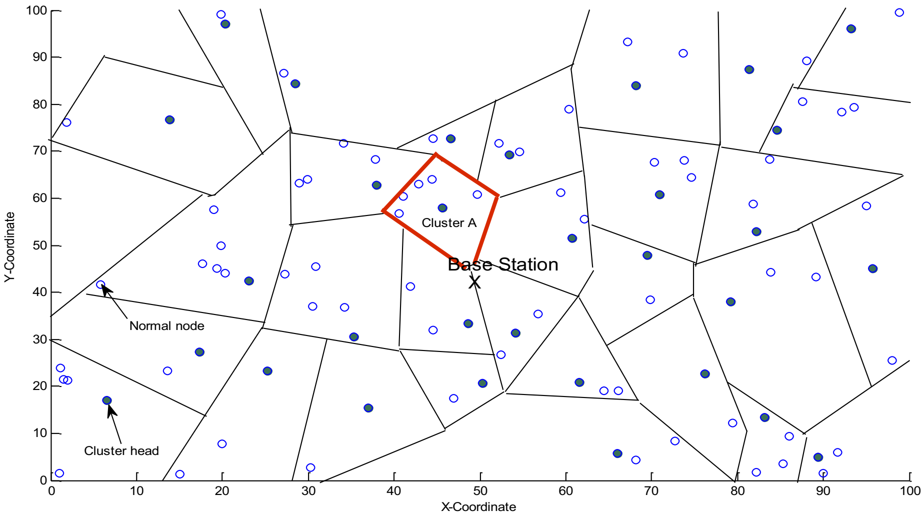

Figure 5.
NRCA clusters formation and cluster heads election at first round

Table II, we can see that the last node in the simulated WSN died in LEACH at round 2,230, making it the least achiever with the shortest network lifetime among the other considered protocols. On the other hand, we can see NRCA has the longest network lifetime followed by PEGASIS as their last nodes died at rounds 3,200 and 2,774, respectively. Table I and Figure 4 show how NRCA outperformed PEGASIS by 15 per cent and LEACH by almost 70 per cent for the network life-time criterion. In this scenario, no threshold was chosen so heads will be changed every round.

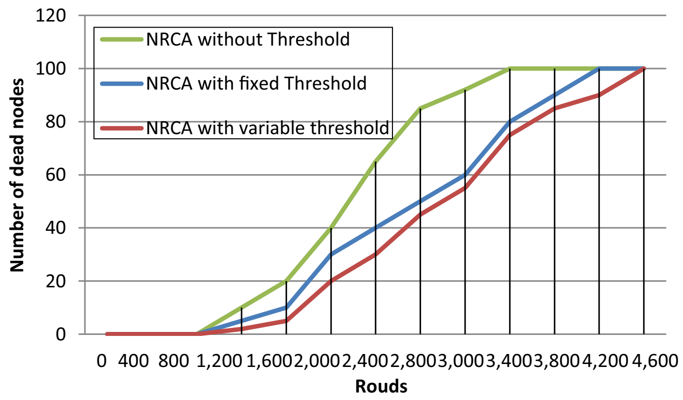

Figure 6.
Simulation results for the network life time

Protocols	Measurements Round first node dies	Round last node dies
LEACH	821	2230
NRCA	1179	3200
PEGASIS	1086	2774

Table II.
Simulation results for the network life time

5.3 Connectivity and coverage

There exists a connectivity between the CH and nodes in the cluster if and only if the physical Euclidean distance between the CH and any node in the cluster is less than or equal to the transmission range of the CH. The more CHs nodes there are, the better coverage or connectivity the network will have and the less distance and energy will be needed to send data. Better coverage also implies minimal or no forgotten or disconnected nodes. From Table III, we can see that NRCA has the least number of disconnected nodes compared with the other two protocols and, as such, its performance with respect to connectivity and coverage is considered better.

5.4 Varying the placement of the BS

In this simulation sub-section, we change the placement of the sink node or BS to see the effect of that on the performance of the algorithms. At position 1 (P1), we placed the BS at the center or middle of the WSN area, (x = 50, y = 50), and at position 2 (P2), we placed

Protocol	No. of disconnected node in first round	No. of disconnected node in round 100	No. of disconnected node in round 500	No. of disconnected node in round 1,000	No. of disconnected node in round 2,000
LEACH	8	13	17	25	31
NRCA	0	2	3	5	10
PEGASIS	3	7	13	17	27

Table III.
Number of disconnected nodes per selected rounds

the BS on the border line of the area where the WSN is being deployed, i.e. (x = 50, y = 0).

From the results in Table IV, we can notice that the change of the BS placement has the least effect on NRCA. PEGASIS came after with a minor effect. On the other hand, LEACH has been affected more by this. It performed better when the BS was placed at the center of the WSN area. This is due to LEACH treating all the nodes without discrimination and randomly selecting the CHCHs.

5.5 Varying the number of nodes

In this simulation, we vary the number of nodes, while keeping the deployment area fixed to see if changing the density of the nodes has any impact on the performance of the algorithms. The position of the BS was fixed at P1. We simulated 100 nodes, 200 nodes and 500 nodes and looked at when the first and last nodes died as shown in Table V.

From Table V, we can see an improvement in the network lifetime of all protocols as the number of nodes increases. The first node in LEACH died at round 821 when the number of nodes is 100, at 830 when the number of nodes is 200 and at round 846 when the number of nodes is 500. Similar performance was observed for the other protocols. As the number of nodes increased, the density increased, making the transfer of data to the sink node less costly in most cases due to shorter transmission distances.

5.6 Received data by the BS

As shown in Figure 7, received data by the BS in NRCA was more than when using the other two algorithms. Data imply both control data sent in CH selections or network setup and the sensed data which are sent through sensors (control data were < 10 per cent). This was due to NRCA choosing the most appropriate nodes as CHs based on both energy and the correct path to the BS. It is also due to minimizing the number of overhead messages needed for CHs selections and replacing processes.

Table IV. Simulation results of changing the placement of the BS

	Measurements			
	Round first node dies		Round last node dies	
Protocols	Middle	Border	Middle	Border
LEACH	821	801	2,350	2,058
NRCA	1,185	1,179	3,302	3,292
PEGASIS	1,086	1,022	3,290	3,240

Table V. Simulation results for different number of nodes

	Measurements					
Protocols	Round first node dies			Round last node dies		
Number of nodes	100	200	500	100	200	500
LEACH	821	830	846	2303	2303	2374
NRCA	1,179	1,185	1,200	3,220	3,329	3,442
PEGASIS	1,086	1,090	1,109	2,974	3,174	3,255

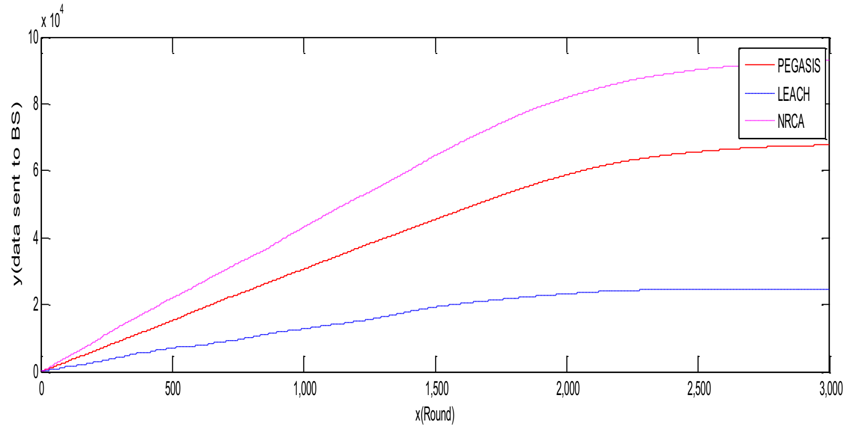

Figure 7.
Received data by the BS per round

5.7 Energy consumed

As shown in Figure 8 the energy consumed per round in NRCA is less than that in LEACH and PEGASIS, with LEACH consuming the most. The amount of energy wasted on the frequent replacement of CH nodes by allowing them to serve as CHs in several rounds as long as their energy did not drop below the specified threshold level, was the main factor in achieving this.

5.8 Using fixed threshold percentages of remaining energy to replace CHs

In this simulation, we varied the fixed threshold values of the remaining energy to replace the CHs. We used 40, 30, 20 and 10 per cent of the remaining energy. Table VI shows the result. As we can see at the beginning, using 40 per cent as energy threshold to replace the CHs performed better than the rest. When almost 50 per cent of the nodes died, we achieved almost equal results for all threshold values. However, as the remaining nodes decrease, the 10 per cent threshold value performed better than the other values.

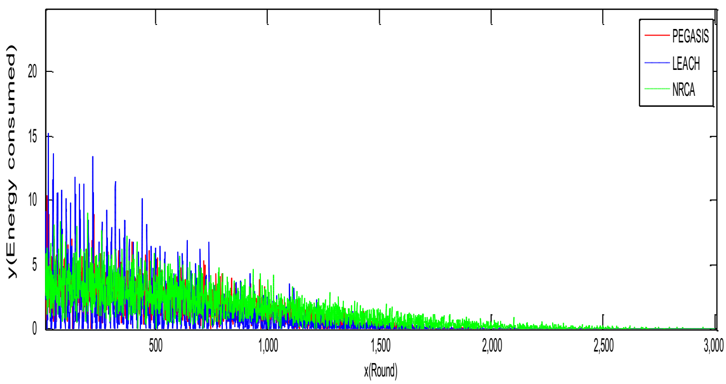

Figure 8.
Energy consumed per round

5.9 Using equation (6) to replace CHs (variable threshold)

Using the result we achieved in the previous simulation section, we apply the formula as defined in equation (6) for calculating the energy threshold in replacing CHs. In this formula, we sought after the nodes closest to the sick to live the longest as they are critical to the network and used by other nodes in the network to forward the data to the BS. Using equation (6) implies that CHs close to the BS will have a higher replacement energy threshold value, i.e. it will be replaced more frequently than farther CH nodes which will have lower replacement energy threshold values. Figure 9 shows the obtained results when simulation experiment are run using the variable energy thresholds calculated based on equation (6) versus using a fixed pre-defined threshold versus replacing heads in each round. As can be seen from Figure 9, the last node died when using NRCA without threshold was at round 3,200 and with fixed threshold at round 4,020, while, with the use of variable threshold values based on equation (6), the last node died at round 4,320. This shows how NRCA with variable and fixed threshold values outperformed NRCA without a using threshold in terms on network life time. Also we can see that NRCA with variable threshold values outperformed NRCA with a fixed one in terms of network lifetime by almost 7 per cent.

5.10 Hybrid node duty-cycle (redundant nodes duty-cycle selection)

In large-scale dense WSNs, sensors are often deployed in large quantities to increase reliability and to extend the coverage (Zhou *et al.*, 2006). As a result, there are many redundant sensor nodes collecting redundant data in such networks. However, to

Table VI. Network lifetime using different threshold values

% of alive nodes	Percentage of left energy to change CH (threshold of the left energy) 10%	20%	30%	40%
First died	843	850	858	861
90%	1,230	1,238	1,245	1,257
50%	1,987	1,982	1,979	1,978
10%	3,580	3,540	3,522	3,506
Last node died	4,020	4,010	3,990	3,970

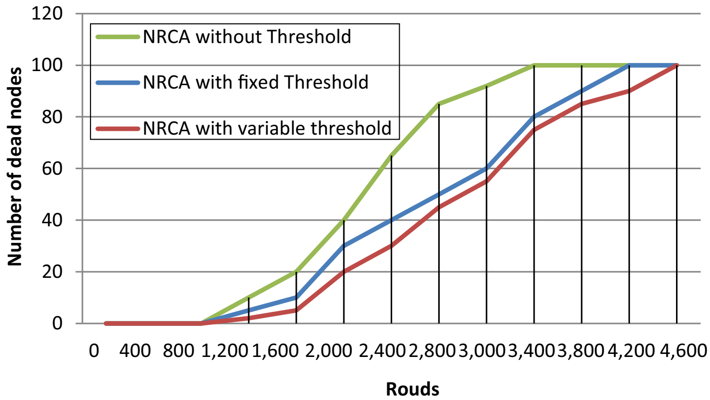

Figure 9. NCRA with fixed, variable and without threshold

increase the network lifetime and distribute the load among nodes, redundant nodes should take turn in covering the monitored area whenever possible. Initially, all nodes are in a working mode and for nodes monitoring the same coverage area, there can be useless redundant data being collected and communicated through the network consuming energy. Therefore, we propose to apply hybrid node duty-cycles, where nodes take turn in monitoring a particular coverage area based on certain conditions. In our technique, we used hybrid duty-cycle scheme where we combined both synchronous and asynchronous schemes. To determine which node should stay active or go to sleep within a cluster, each node will communicate with its direct neighbors and detect nodes which are within the same pre-defined detection range (sensing or coverage range). Nodes, covering the same detection range will then agree on which node stays active based on energy. If the energy of an awake working node is less than a certain threshold, for example, 10 per cent of the initial energy, the working node will send a broadcast massage to wake up sleeping nodes within the detection range before it goes to sleep. For reliability purposes, sleeping nodes will wake up to enter into the detecting mode in the event if a period of time Ts passed without receiving any instructions from the awake node. This technique is efficient when monitoring a continuous event. From Figure 10, we can see that the last node died in NCRA without using nodes duty-cycle died at round 4,320, while, with duty-cycle, it was at 4,660. This shows that using duty-cycle strategy improved the performance by almost 8 per cent.

6. Conclusions

In this paper, we proposed an energy-efficient clustering algorithm for WSNs using node ranking in electing CHs. We compared the performance of our proposed algorithm against two well know algorithms in terms of network life time. Through simulation we showed how the proposed algorithm outperformed PEGASIS by 15 per cent and LEACH by almost 70 per cent for the network lifetime criterion. We found that using energy threshold to replace CHs improved the network lifetime by almost 15 per cent. We also found that using variable energy threshold values to replace CHs improved the network lifetime further by almost 7 per cent over the use of a fixed value. In addition to that, using hybrid redundant nodes duty-cycle has improved the network lifetime by 8 per cent.

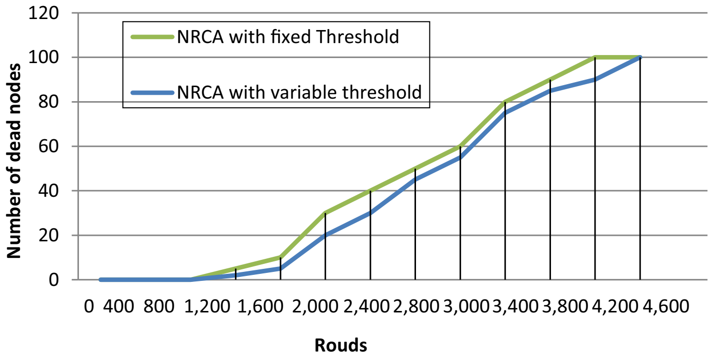

Figure 10. NRCA with and without duty-cycle

References

Amis, A.D., Prakash, R., Vuong, T.H.P. and Huynh, D.T. (2000), "Max-min D-cluster formation in wireless ad hoc networks", *Proceeding of IEEE INFOCOM*, *Richardson, TX, March*.

Akyildiz, I.F., Su, W., Sankarasubramaniam, Y. and Cayirci, E. (2002), "Wireless sensor networks: a survey", *Computer Networks*, Vol. 38 No. 4 pp. 393-422.

Alnuaimi, K., Alnuaimi, M., Mohamed, N., Jawhar, I. and Shuaib, K. (2012), "Web-based wireless sensor networks: a survey of architectures and applications", *The 6th International Conference on Ubiquitous Information Management and Communication*, 20-22 February.

Alnuaimi, M., Sallabi, F. and Shuaib, K. (2011), "a survey of wireless multimedia sensor networks challenges and solutions", *IEEE International Conference on Innovations in Information Technology (IIT'11)*, Abu Dhabi, pp. 191-196.

Alnuaimi, M., Shuaib, K. and Jawhar, I. (2006), "Performance evaluation of IEEE 802.15.4 physical layer using MatLab/Simulink", *IEEE International Conference on Innovations in Information Technology*, 19-20 November, pp. 1-5.

Alnuaimi, M., Shuaib, K., Nuaimi, K. and Abdel-Hafez, M. (2013), "Performance analysis of clustering protocols in WSN", *Wireless and Mobile Networking Conference (WMNC), 2013 6th Joint IFIP*, Dubai, *23-25* April, IEEE, pp. 23-25.

Banerjee, S. and Khuller, S. (2001), "A clustering scheme for hierarchical control in multihop wireless networks", *Proceeding IEEE INFOCOM*, *Anchorage, AK, 22 April*.

Baronti, P., Pillai, P., Chook, V., Chessa, V. and Gotta, A. (2007), "Wireless sensor networks: a survey on the state of the art and the 802.15.4 and ZigBee standards", *Computer Communications*, Vol. 30 No. 7, pp. 1655-1695.

Basagni, S. (1999), "Distributed clustering algorithm for ad hoc networks", *Proceeding International Symposium Parallel Architectures, Algorithms, and Networks*, *Washington, DC*.

Biradar, R.V., Sawant, S.R., Mudholkar, R.R. and Patil, V.C. (2011), "Multi-hop routing in self-organizing wireless sensor networks", *International Journal of Computer Science Issues*, Vol. 8 No. 1.

Blough, D.M. and Santi, P. (2002), "Investigating upper bounds on network lifetime extension for cell-based energy conservation techniques in stationary ad hoc networks", *Proceedings of the 8th Annual International Conference on Mobile Computing and Networking*, *New York, NY, 22-26 September, ACM*, pp. 183-192.

Depuru, S.S.S.R., Wang, L. and Devabhaktuni, V. (2011), "Smart meters for power grid: challenges, issues, advantages and status", *Renewable and Sustainable Energy Reviews*, Vol. 15 No. 6, pp. 2736-2742.

Du, G., Shi, Q., Tang, Y. and Sun, X. (2011), "A mixed non-uniform clustering algorithm for wireless sensor networks", *Communication Technology (ICCT), 2011 IEEE 13th International Conference*, Jinan, 25-28 September, IEEE, pp. 661-665.

García Villalba, L.J., Sandoval Orozco, A.L., Cabrera, A.T. and Barenco Abbas, C.J. (2009), "Routing protocols in wireless sensor networks", *Sensors*, Vol. 9 No. 11, pp. 8399-8421.

Heinzelman, W., Chandrakasan, A. and Balakrishnan, H. (2002), "An application- specific protocol architecture for wireless microsensor networks", *IEEE Transaction on Wireless Communication*, Vol. 1 No. 4.

Heinzelman, W.B. (2000), "An application-specific protocol architectures for wireless networks", PhD. Dissertation, MIT, Cambridge.

Heinzelman, W.B., Chandrakasan, A.P. and Balakrishnan, H. (2000), "Energy-efficient communication protocol for wireless microsensor networks", *Proceedings of the 33rd Hawaii International Conference on System Sciences*, *Washington, DC*, Vol. 8.

Hong, J., Kook, J., Lee, S., Kwon, D. and Yi, S. (2011), "T-LEACH: the method of threshold-based cluster head replacement for wireless sensor networks", *Information Systems Frontiers*, Vol. 11 No. 5, pp. 513-521.

Kaur, R. and Sharma, M. (2011), "An approach to design habitat monitoring system using sensor networks", *International Journal of Soft Computing and Engineering (IJSCE)*, Vol. 1, pp. 5-8.

Koutsonikolas, D., Das, S.M., Charlie Hu, Y. and Stojmenovic, I. (2010), "Hierarchical geographic multicast routing for wireless sensor networks", *Wireless Networks*, Vol. 16 No. 2, pp. 449-466.

Kumar, N., Sandeep, P.B. and Mishra, P. (2012), "U-LEACH: a novel routing protocol for heterogeneous wireless sensor networks", *International Conference on Communication, Information & Computing Technology (ICCICT*), Mumbai, 19-20 October.

Lindsey, S. and Raghavendra, C.S. (2002), "PEGASIS: power efficient gathering in sensor information systems", *Proceedings of the IEEE Aerospace Conference*, *Big Sky, MT*.

Nikolidakis, S.A., Kandris, D., Vergados, D.D. and Douligeris, C. (2013), "Energy efficient routing in wireless sensor networks through balanced clustering", *Algorithms*, Vol. 6 No. 1, pp. 29-42.

Radi, M., Dezfouli, B., Abu Bakar, K. and Lee, M. (2012), "Multipath routing in wireless sensor networks: survey and research challenges", *Sensors*, Vol. 12 No. 1, pp. 650-685.

Sha, C., Wang, R.-C., Huang, H.-P. and Sun, L.-J. (2010), "A type of healthcare system based on intelligent wireless sensor networks", *The Journal of China Universities of Posts and Telecommunications*, Vol. 17 No. S1, pp. 30-39.

Song, S.-M. and Yoon, Y.-I. (2010), "Intelligent smart cloud computing for smart service", *Grid and Distributed Computing, Control and Automation,* Springer Berlin Heidelberg, Berlin, pp. 64-73.

Szewczyk, R., Osterweil, E., Polastre, J., Hamilton, M., Mainwaring, A. and Estrin, D. (2004), "Habitat monitoring with sensor networks", *Communications of the ACM*, Vol. 47 No. 6.

Shu, T., Krunz, M. and Vrudhula, S. (2005), "Power balanced coverage time optimization for clustered wireless sensor networks", *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, *Chicago, IL*, *25-28 May, ACM, New York, NY*, pp. 111-120.

Younis, O. and Fahmy, S. (2004), "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks", *IEEE Transactions on Mobile Computing*, Vol. 3 No. 4, pp. 366-379.

Younis, O., Krunz, M. and Ramasubramanian, S. (2006), "Node clustering in wireless sensor networks: recent developments and deployment challenges", *IEEE Network*, Vol. 20 No. 3, pp. 20-25.

Zhou, Z.X., Wang, X. and Pan, J. (2006), "An energy-efficient data-dissemination protocol in wireless sensor networks", *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM*), Buffalo, NY, pp. 10-19.

Further reading

Tyagi, S. and Kumar, N. (2013), "A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks", *Journal of Network and Computer Applications*, Vol. 36 No. 2, pp. 623-645.

Yu, J.G., Qi, Y.Y., Wang, G.H. and Gu, X. (2012), "A cluster-based routing protocol for wireless sensor networks with nonuniform node distribution", *AEU-International Journal of Electronics and Communications*, Vol. 66 No. 1, pp. 54-61.

Appendix. List of acronyms and abbreviations

WSN = Wireless sensor networks
BS = Base station
LEACH = Low energy adaptive clustering hierarchy
PEGASIS = Power-efficient Gathering in sensor information systems
HEED = Hybrid energy-efficient distributed clustering
SEP = Stable election protocol
MNUC = Mixed unequal clusters size algorithm
HGMR = Hierarchical geographic multicast routing
EADC = Energy-aware dynamic clustering algorithm
ULEACH = Universal LEACH
MCH = Master cluster head
HABRP = Hierarchical balanced energy-efficient routing protocol
En = Energy level
Dn = Distance from base station
CCH = Candidate cluster head
T = Energy threshold
CH = Cluster head
NRCA = Node ranking clustering algorithm

About the authors

Mariam Alnuaimi is a PhD student at the College of Information Technology at the United Arab Emirates University. She has a Master degree from British University in Dubai (BUiD) majoring in Information Technology Management and a Bachelor degree in Network Engineering from College of information Technology in United Arab Emirates University. Her current research interests are wireless sensor networks, energy efficiency protocols and load balancing. Mariam Al Nuaimi is the corresponding author and can be contacted at: mariam.alnuaimi@uaeu.ac.ae

Khaled Shuaib received his PhD in Electrical Engineering/Communication Networks from the Graduate Center of the City University of New York, 1999 and his MS and BE in Electrical Engineering from the City College of New York, 1993 and 1991 respectively. Since September 2002, Khaled has been with the College of Information Technology (CIT), at the UAEU where he is currently a Professor.

Klaithem Alnuaimi is a PhD student at the College of Information Technology at UAEU. She is researching cloud computing load balancing and storage issues. She published several papers in the area since 2011. She obtained her bachelor from UAEU majoring in Software Engineering in 2010. Her current research interests are cloud computing, load balancing and storage optimization in the cloud.

Mohammed Abdel-Hafez is an Associate Professor in the Department of Electrical Engineering, United Arab Emirates University, UAE. He received his BSc, MSc and PhD degrees all in Electrical and Electronic Engineering from Eastern Mediterranean University (EMU) in June 1992, August 1994 and November 1997, respectively.

This article has been cited by:

1. Mariam Alnuaimi, Khaled Shuaib, Klaithem Alnuaimi, Mohammed Abdel-Hafez. 2015. Data Gathering in Delay Tolerant Wireless Sensor Networks Using a Ferry. *Sensors* **15**:10, 25809-25830. [CrossRef]