Journal of Computers Vol. 29 No. 4, 2018, pp. 203-216
doi:10.3966/199115992018082904016

Energy Savings through a Clustering-based Routing Protocol for Wireless Sensor Networks

Omar Moussaoui, Mohammed Jabri

MATSI Lab, ESTO, University Mohamed IB.P 473, 60000 Oujda, Morocco
Omar.moussaoui78@gmail.com, jabrimed@hotmail.fr

Received 10 March 2017; Revised 25 August 2017; Accepted 06 October 2017

Abstract. A Wireless Sensor Network (WSN) provides a significant contribution in emerging fields such as ambient intelligence and ubiquitous computing. In WSNs, the optimization and the load balancing of network resources are a critical concern to provide intelligence for a long duration. Since the clustering of sensor nodes can significantly enhance the overall system scalability and energy efficiency of the network, this paper presents a distributed energy-efficient clustering-based hierarchy protocol to achieve the network longevity in WSN. Indeed, the proposed algorithm partitions the network evenly into different clusters based on the size of cluster, the transmission power, and the energy level of nodes with one node acting as a cluster head (CH). Additionally, the routing paths between CHs are dynamically formed using a geographical and energy aware neighbors' selection to achieve inter-clusters communication. Finally, the intra-cluster scheduling is the last major issue related to the setup phase. Simulation results show clearly that our scheme reduces the overall energy consumption of the network and improves its lifetime compared to other energy-saving-based mechanisms.

Keywords: clustering, energy-efficiency, hierarchical routing protocol, network lifetime, wireless sensor networks

1 Introduction

Wireless Sensor Networks (WSNs) probe and collect environmental information in order to provide ubiquitous sensing, computing, and communication capabilities [1]. Although the WSNs are valuable assets for data collection in remote and hostile locations, a network can only survive as long as the battery capacity of its components. Furthermore, because of the difficulties of battery recharging in such conditions, the availability of a sensor is dependent on initial battery capacity and energy consumption efficiency. Thus, prolonging the lifetime of a battery powered network is a primary consideration in many WSNs [2].

Various works to achieve an efficient use of energy have been studied. Among such works, routing protocols based on clustering approach are viewed as superior since they can potentially reduce energy consumption in multiple ways [3]. For example, a clustering hierarchy can reduce the amount of query packets via inter-cluster query dissemination that reduces the amount of data packets by aggregating collected data. A clustering hierarchy may also control the minimal number of active nodes to cover the target area by putting redundant sensor nodes to sleep. Moreover, a clustering is particularly useful for applications that require scalability to hundreds or thousands of nodes. The scalability in this context implies the need for load balancing and efficient resource utilization.

The clustering-based routing schemes partition the network into virtual cells governed by a set of cluster heads (CHs) selected among the sensor nodes, while the other nodes are grouped with these CHs. The CHs are responsible for coordination among the nodes within their clusters (intra-cluster coordination) as well as communicating with other CHs (inter-cluster communication). The other sensor nodes only have to transmit their information to their respective CH, which aggregates the received information and forwards it to the base station (BS). However, hierarchical-clustering scheme operates within some constraints. The most relevant is the CH re-election process. In order to alleviate the large amount of energy consumption required by a CH, frequent reconfigurations of clusters are needed. Furthermore, partitioning the network in clusters, as well as optimally choosing a CH, is an NP-hard problem [4]. The existing solutions to this problem are based on heuristic approaches and no attempts to retain the stability of the network topology. We believe that a good clustering scheme should preserve its structure as much as possible when the nodes are moving and/or the topology is slowly changing. Otherwise, the re-computation of the cluster heads and frequent information exchange among the participating nodes will result in high computation overhead.

We are motivated through the current work to limit some constraints that impact negatively the use of energy in WSN, particularly by reducing energy consumption due to data communication. We first eliminate the redundant data locally at the level of each CH by gathering neighboring nodes in the same cluster and using a data aggregation mechanism. Then, we adapt a load balancing mechanism within the entire network in order to prolong its lifetime as much as possible.

In this paper, we propose a novel routing scheme based on clustering [5] called DECHP (Distributed Energy-efficient Clustering-based Hierarchy Protocol), which utilizes a fully distributed approach to set up clusters and routing paths, performs the rotation of cluster heads, and carries out other energy intensive tasks. The proposed algorithm partitions the network into different clusters based on:

The cluster size (number of sensors). This equilibrates the cluster in terms of sensor nodes present in the cluster by defining a threshold of the sensor nodes that a cluster can regroup (depending on CH capacity to handle traffic).

The distance between nodes constituting the cluster. This allows the improvement of the communication quality by reducing interferences, wireless fading, and energy consumption.

The energy level of each node. This leverages the network lifetime by balancing the energy capacity across clusters.

As soon as clusters are established, one node is elected from each cluster as a CH. Moreover, elected CHs use a geographical and energy aware neighbor CH's selection to join the BS. All non-cluster head nodes in each cluster transmit their data to the CH. Thus, the CH receiving data from all the cluster members performs signal processing functions on data (data aggregation) and transmits the aggregated data to its upper level CH. This process continues until the data reaches the BS. However, a reconfiguration procedure is usually required in wireless sensor networks since the CH is limited by its residual energy level. Unlike other existing clustering schemes where the CH reconfiguration is invoked periodically which introduces high computation overhead, DECHP is adaptively invoked to only change the CHs by taking into account their remaining energy levels. Namely, the cluster creation is made only at the system activation. The CH's change is done when necessary, according to the remaining energy level.

The remainder of this paper is organized as follows. Section 2 surveys the related work. Then, we introduce network, radio, and data correlation models used by DECHP in Section 3. A detailed description of DECHP is presented in Section 4 and simulations are given in Section 5. Finally, we conclude this paper in Section 6.

2 Background and Related Work

Based on network structure, routing protocols in WSNs can be coarsely divided into two categories: flat routing and hierarchical routing.

In a flat topology, all nodes perform the same tasks and have the same functionalities in the network. Data transmission is performed hop by hop usually using the form of flooding. The typical flat routings in the WSNs include Sensor Protocols for Information via Negotiation (SPIN) [6], Directed Diffusion (DD) [7], Gradient-Based Routing (GBR) [8], etc. In small-scale networks, flat routing protocols are relatively effective. However, it is often unfavorable in large-scale networks where resources are limited, but all sensor nodes generate more data processing and bandwidth usage.

On the other hand, in a hierarchical topology, nodes perform different tasks in the WSNs and are typically organized into lots of clusters according to specific requirements or metrics [9]. Generally, each cluster comprises a leader referred to as a cluster head (CH) and other member nodes (MNs). The CHs can be organized into further hierarchical levels. The typical clustering routing protocols in WSNs include Low-energy Adaptive Clustering Hierarchy (LEACH) [10], Hybrid Energy-Efficient Distributed clustering (HEED) [11-12], Position-based Aggregator Node Election protocol (PANEL) [13], Power-

Efficient Gathering in Sensor Information Systems (PEGASIS) [14], The Adaptive Threshold sensitive Energy Efficient sensor Network protocol (APTEEN) [15], Concentric Clustering Scheme (CCS) [16], etc. Table 1 summarizes the categories and differences of the main clustering routing protocols in WSNs according to a variety of clustering attributes. Furthermore, we compare the prominent clustering routing approaches in WSNs based on a few important metrics in Table 2.

Table 1. Classification of Prominent Clustering Routing Protocols in WSNs

Protocol Name	Cluster Characteristics	CH Characteristics	Clustering Process	Proceeding of the Algorithm
LEACH	- Cluster count: variable - Intra-cluster routing: single-hop - Inter-cluster routing: single-hop	- Existence: CH based - Difference of capabilities: homogeneous - Mobility: stationary - Role: relay/aggregation	- Control manners: distributed - Execution nature: probabilistic - Convergence time: constant - Objectives: load balancing	cluster construction
HEED	- Cluster count: variable - Intra-cluster routing: single-hop - Inter-cluster routing: single/multiple-hop	- Existence: CH based - Difference of capabilities: homogeneous - Mobility: stationary - Role: relay/aggregation	- Control manners: distributed - Execution nature: iterative - Convergence time: constant - Objectives: load balancing	cluster construction
PANEL	- Cluster count: fixed - Intra-cluster routing: single-hop - Inter-cluster routing: multiple-hop	- Existence: CH based - Difference of capabilities: homogeneous - Mobility: stationary - Role: relay/aggregation	- Control manners: distributed - Execution nature: probabilistic - Convergence time: constant - Objectives: load balancing and reliability	cluster construction
PEGASIS	- Cluster count: variable - Intra-cluster routing: multiple-hop - Inter-cluster routing: single-hop	N/A	- Control manners: distributed - Execution nature: probabilistic - Convergence time: constant - Objectives: load balancing	data transmission
APTEEN	- Cluster count: variable - Intra-cluster routing: single-hop - Inter-cluster routing: multiple-hop	- Existence: CH based - Difference of capabilities: homogeneous - Mobility: stationary - Role: relay/aggregation	- Control manners: distributed - Execution nature: probabilistic - Convergence time: constant - Objectives: proactive and reactive scenes	data transmission
CCS	- Cluster count: variable - Intra-cluster routing: multiple-hop - Inter-cluster routing: multiple-hop	- Existence: CH based - Difference of capabilities: homogeneous - Mobility: stationary - Role: relay/aggregation	- Control manners: distributed - Execution nature: probabilistic - Convergence time: constant - Objectives: lifetime extension	data transmission

Table 2. Comparison of Prominent Clustering Routing Protocols in WSNs

Protocol name	Energy Efficiency	Cluster Stability	Scalability	Delivery Delay	Load Balancing	Algorithm Complexity
LEACH	very low	moderate	very low	very small	moderate	low
HEED	moderate	high	moderate	moderate	moderate	moderate
PANEL	moderate	low	low	moderate	good	high
PEGASIS	low	low	very low	very large	moderate	high
APTEEN	moderate	very low	low	small	moderate	very high
CCS	low	low	low	large	very bad	moderate

Clustering routing is becoming an active branch of routing technology in WSNs on account of a variety of advantages, such as increased scalability, data aggregation/fusion, reduced load, reduced energy consumption, increased robustness, etc. LEACH, which is considered one of the most popular hierarchical routing algorithms, is an application-specific data dissemination protocol that uses clustering to prolong the network lifetime. LEACH shows good characteristics of clustering with little overhead.

Based on pre-determined probability and cluster head history, some of the nodes elect themselves as CH. Consequently, the elected CHs collect data from cluster members in their vicinity and transfer the aggregated data to the BS until the next election round begins. However, LEACH does not guarantee a good cluster head distribution and the remaining energy is not reflected in the next head election procedure. Due to its drawbacks, many variant protocols are introduced in order to improve LEACH [17].

PEGASIS is another clustering-based routing protocol that enhances network lifetime by increasing local collaboration among sensor nodes. In PEGASIS, the nodes are organized into a chain using an aggressive algorithm so that each node transmits to and receives from only one of its neighbors. In each round, a randomly chosen node from the chain will transmit the aggregated data to the BS, thus reducing the per round energy expenditure compared to LEACH protocol. However, PEGASIS introduces excessive delay for distant nodes on the chain. Moreover, the single leader can become a bottleneck.

LEACH and, its follow-up work, PEGASIS are built on the assumption that the base station is fixed and located far from sensors. Furthermore, they assume that every sensor node can reach the BS directly. Hence, the assumptions cited above severely limit the applicability of such protocols.

Targeting to the clusters that are formed with respect to existing network topology is the issue that makes our scheme different from existing sensor network clustering algorithms. If the cluster head selection is based on node identification, node connectivity, or randomness, it does not guarantee that cluster head location is reasonable in terms of spatial attributes. Moreover, less attention is paid to the issue of clustering the network for maximum leverage of data aggregation. Therefore, our algorithm allows nodes to organize themselves into groups of locally, regular, and non-overlapping clusters. Indeed, grouping the nodes with respect to the regions of close proximity and similar deployment density promotes both efficient data aggregation and efficient compression of sensor data. This grouping would also assist transmission power control since intra-cluster communication requires less transmission power in dense clusters. Additionally, our scheme presents some specifications that make change compared to the other protocols cited above. Obviously, we avoid the periodical re-election of CHs which is responsible of the energy overconsumption and the convergence time of the periodic set-up phase. We also maintain through our algorithm the topology stability of the network using a load balancing mechanism. Finally, we ensure that all these aspects can significantly reduce energy consumption.

3 Models

3.1 Network Model

We consider a system based on the similar model used in LEACH. It consists of a BS, far from nodes, through which the end user can access data from the sensor network. In this network, sensing tasks can be either query-based, event detection, or periodic data reporting. Thereby, we propose a generic framework for various sensing applications. We assume the following properties about the sensor network model:

- The nodes in the network are quasi-stationary. Indeed, this assumption is typical for several wireless sensor networks.
- The nodes use an Omni-directional antenna, which has a fixed number of transmission levels (i.e. every node can control radio transmission power).
- The sensing range of a node is smaller than the communication range.
- The BS has a constant power supply. In this case, no energy constraints are considered, so the BS can transmit with high power to all nodes. There is no need for routing from the BS to any specific node and the nodes cannot always reply to the BS directly due to their power constraints, resulting in asymmetric communication.
- The nodes are aware and able to measure the distance to their 1-hop neighborhood. We consider that both assumptions are reasonable. The first assumption is standard for many neighborhood discovery algorithms. Whereas the second assumption is becoming a common feature of many sensor network applications. Accurate inter-node distance measurements in the sensor network domain have been demonstrated using ultrasound in the system described in [18], MIT Crickets [19], and Medusa MK-2 node [20]. In the radio domain, ultra-wide-band ranging systems such as the one offered by Ubisense [21] have already demonstrated accurate distance measurements with small sensors from factors that

will be suitable for sensor networks.

— The proper transmission control schemes for medium access are available to minimize the effect of dynamic wireless channel conditions. For sensor networks, those control schemes need to provide energy-efficient listening and random back-off mechanisms, low overhead contention control and adaptive rate control schemes, along with mechanisms to reduce hidden node problem [22].

The two key elements considered in the design of DECHP are sensor nodes and the BS node. The sensor nodes are geographically grouped into clusters and capable of operating in two basic nodes: *(i)* the cluster head node and *(ii)* the sensing node. In the case of sensing node, nodes perform sensing tasks and transmit the sensed data to the cluster head. However, in the case of cluster head node, a node gathers data from other nodes within its cluster, performs data fusion, and routes the data to the BS through other cluster head nodes. In addition, CHs can operate at a higher power mode (resulting in higher transmission range) for inter-clusters communication while they use a lower power for intra-cluster communication. The BS, in turn, supervises the entire network.

3.2 Radio Model

A typical sensor node consists of four major components: a data processor unit; a micro-sensor; a radio communication subsystem that consists of transmitter/receiver electronics, and an antennae, amplifier and power supply unit [23]. Although energy is dissipated in all of the first three components, we mainly consider the energy dissipations associated with the radio component since the core objective of our work is to develop an energy-efficient network layer protocol to improve network lifetime. In addition, the energy dissipated during data aggregation in cluster head nodes is also taken into account.

In our analysis, we use the same radio model discussed in [10]. The transmit and receive energy costs for the transfer of a k-bit data message between two nodes separated by a distance of *r* meters is given by equations (1) and (2), respectively.

$$E_T\left(k,r\right)=E_{Tx}*k+E_{amp}(r)*k\cdot \tag{1}$$

$$E_R\left(k\right)=E_{Rx}*k\cdot \tag{2}$$

Where $E_T\left(k,r\right)$ in (1) denotes the total energy dissipated in the transmitter of the source node and $E_R\left(k\right)=E_{Rx}*k$ in (2) represents the energy cost incurred in the receiver of the destination node. The parameters E_{Tx}, E_{Rx} in equations (1) and (2) are the per bit energy dissipations for transmission and reception, respectively. $E_{amp}(r)$ is the energy required by the transmit amplifier to maintain an acceptable signal-to-noise ratio in order to transfer data messages reliably. Like [10], we use both the free-space propagation model as well as the two-ray ground propagation model to approximate the path loss sustained due to wireless channel transmission. Given a threshold transmission distance of r_o, the free-space model is employed when $r\le r_o$ and the two-ray model is applied for cases where $r\succ r_o$. Using these two models, the energy required by the transmit amplifier $E_{amp}(r)$ is given by the equation (3).

$$E_{amp}(r)=\begin{cases}\varepsilon_{FS}r^2, r\le r_o\\ \varepsilon_{TR}r^4, r\succ r_o\end{cases}\cdot \tag{3}$$

Where ε_{FS} and ε_{TR} denote transmit amplifier parameters corresponding to the free-space and the two-ray models, respectively, and r_o is the threshold distance given by the equation (4).

$$r_o=\sqrt{\varepsilon_{FS}/\varepsilon_{TR}}\cdot \tag{4}$$

For all experiments in our work, we assume the same set of parameters used in [10]. That is, $E_{Tx}=E_{Rx}=50nJ/bit$, $\varepsilon_{FS}=10pJ/bit/m^2$, and $\varepsilon_{TR}=0.0013pJ/bit/m^4$. Moreover, the energy cost for data aggregation is set at $E_{DA}=5nJ/bit/message$.

3.3 Data Correlation Model

Since the data collected by neighboring sensors has a lot of redundancy, authors in [10] assume a perfect data correlation so that all individual signals from the members of the same cluster can be combined into a single representative signal. Nevertheless, this assumption cannot hold when the cluster size increases to some extent. Therefore, we developed a complementary exponential data correlation model based on the observation in distributed data compression [24].

Considering the phenomenon of interest as a random process, the correlation between data collected by two sensors is generally a decreasing function of the distance r between them. After aggregating data, most of the redundancy is removed. Hence, the residue can be assumed to be an increasing function of r based on the observation above. The effect of data aggregation is modeled as below.

Suppose a node collects l bits and sends them back to its head at a distance r while the head expends $2lE_{DA}$ Joules to perform data aggregation on the $2l$ bits (collected by itself and its members). The resulting data is assumed to be $l(1 + \eta)$ bits, where η is data aggregation residue ratio and assumed to be complementary exponential, namely,

$$\eta = 1 - e^{-\lambda r}, 0 \le \lambda \le 1 \cdot \tag{5}$$

where λ is a positive number depending on the specific phenomenon of interest. For example, the light, sound, and temperature often show a strong correlation at a short distance, so λ will take lower values for such data. Since η is a monotonic increasing function of r, it varies from zero to one when r increases from zero to infinity. So, this model can approach the perfect-data-correlation assumption by decreasing λ or approach the no-data-aggregation assumption by increasing λ. Thus, different scenarios can easily be set up by varying λ.

4 DECHP: Distributed Energy-efficient Clustering Hierarchy Protocol

The proposed algorithm (DECHP) operates in two major phases: setup and data communication. In this section, we describe the details of the two phases. As a prelude, we introduce both the clustering construction constraints and the algorithm aims.

4.1 Clustering Constraints

The wireless sensor network formed by nodes and links is represented by an undirected graph G=(V, E), where V represents the set of the nodes v_i and E represents the set of the links e_i. A clustering can be thought as a graph-partitioning problem with some added constraints. As the underlying graph does not show any regular structure, partitioning the graph optimally with respect to certain parameters is an NP-hard problem. Our goal is to divide the whole sensor network into no overlapping clusters, where one node is elected as a CH in each cluster. To this aim, the following requirements must be met:

- The clustering procedure is completely distributed (i.e. each node independently makes its decisions using local information).
- The clustering procedure terminates within a fixed number of iterations (regardless of network diameter).
- When the clustering procedure is finished, each node is either a cluster head or a regular node that belongs to exactly one cluster.
- The clustering procedure should be efficient in terms of processing complexity as well as messages exchange.
- CHs are well-distributed over the sensor field.

4.2 Desired Aims of the Clustering Procedure

Aiming at ensuring that the clustering procedure is suitable for a hierarchical routing infrastructure, we propose the following desirable properties that should be present in the clustering mechanism:

- Each cluster is connected. This is an obvious requirement to localize and restrict cluster traffic within clusters.

- All clusters should have a minimum and maximum size constraint. A maximum size constraint limits the cluster size; hence the CH is able to maintain efficiently the intra-cluster communication. Meanwhile, the cluster size can also be chosen to bind the amount of state that needs to be maintained within the cluster. Ideally, the size χ of all the clusters is the same (a pre-defined threshold). Thus, no cluster is overburdened or under-burdened with processing and storage requirements of cluster maintenance. On one hand, small clusters lead to wasteful resource allocation for nodes. On the other hand, large clusters leverage overheads that result from an increased delay (as in TDMA (Time Division Multiple Access)) involved when the nodes get their shared resources.
- All nodes' transmission range is limited within a certain distance. Besides involving less power consumption when a node has to communicate with other nodes, this enables a CH to communicate with a better channel condition (less attenuation, low interference) with its neighbor nodes (short distance).
- Each node must belong to one, unique cluster in order to reduce the power consumption. In fact, when a node belongs to two clusters, the power consumption is wasted through maintaining the cluster state and carrying out intra-cluster traffic for both clusters.
- The clusters reconfiguration should be delayed as long as possible aiming to reduce both the system updates and communication costs.

4.3 Setup Phase

The main activities in setup phase are: (1) cluster setup and cluster head selection, (2) routing paths between CHs formation, and (3) schedule creation for each cluster. At first, DECHP organizes sensor nodes into local clusters based on some system parameters such as cluster size, transmission power, and energy level of nodes with one node acting as a cluster head. Then, the routing paths between CHs are formed, while schedule creation is the last major issue related to the setup phase.

Clustering procedure. Based on the preceding discussion, DECHP combines the system parameters with certain weighing factors based on system needs. For instance, if we consider the energy level of nodes as a crucial parameter in sensor networks, then its weight should be considered with larger value. In addition, based on specific applications, any or all of these parameters can be used in the metric to group nodes into local clusters. According to our notation, the number of nodes that a CH can handle ideally is χ. Besides ensuring that each CH is not overloaded, this allows system efficiency to maintain the expected level.

Through the proposed algorithm, only nodes with sufficient energy are selected as CHs, while those with low energy extend their lifetime by performing tasks that require low energy costs. Further, each cluster members are mostly adjacent to each other and sense similar data, which are aggregated by the CH. Thus, limiting the amount of data that needs to be sent to the BS. On the other hand, cluster formation procedure is only invoked at the system activation and the CH reconfiguration procedure is delayed as long as possible.

At this point the cluster formation takes place in five steps:

(1) Each node *i* finds its neighbor set NS_i (6) that represents the set of nodes inside its transmission range.

$$NS_i = \left\{ j \in V / d\left(i, j\right) \prec tx_{range}\left(i\right) \right\}. \tag{6}$$

(2) Each node *i* computes its Equivalence Classes (ECs). An EC_i (7) of node *i* is the set of nodes that belong to NS_i. Meanwhile, two nodes (*j*, *k*) belong to the same EC_i, if *j* and *k* are neighbors.

$$EC_i = \left\{ j \in NS_i / \forall k \in EC_i \Rightarrow j \in NS_k \right\}. \tag{7}$$

(3) For every Equivalence Class *EC* of each node, computes the combined weight W_{EC}, defined as follows:

$$W_{EC} = \alpha \left\| EC \right| - \chi \right| + \frac{2\beta}{\left(\left| EC \right|^2 - \left| EC \right| \right)} \sum_{j,k \in EC} d\left(j,k\right) + \gamma \sum_{j \in EC} {}^{1}\!\big/_{C_e\left(j\right)}. \tag{8}$$

Where α, β and γ are weights, |*EC*| is the size of *EC*, and $C_e(j)$ is the energy level currently available at node *j*.

(4) Each node chooses among the set of its equivalence classes the class which minimizes W_{EC} as a cluster.

(5) If a node belongs to several clusters, then this node is placed in the smallest cluster in term of size. This allows balancing evenly the number of nodes in these clusters.

From (8), it is obvious that the main metric defining the cluster selection is W_{EC}. This metric, in fact, is composed by different components reflecting the: (1) number of nodes enclosed by the cluster; (2) distance between nodes; and (3) energy level of each node. The first component is mainly contributing to limit the number of nodes in each cluster, which allows the CH to handle up to limited number of nodes in its cluster. Accordingly, we leverage the efficient of MAC functioning when we know that high node numbers increase the complexity of TDMA scheduling at the CH. The second component is related to the energy consumption, where it is well known that more power is required to communicate to a large distance. The last component measures the energy level currently available in nodes, which in turn depends on the node's initial energy as well as the energy expended according to the current network traffic and length of links used to support it.

Once the cluster formation is achieved, sensor nodes are grouped into clusters. Each node belongs to only one cluster and each cluster is represented by a CH, which is elected from the node's set. Thereby, simple nodes communicate with others only through the CH of its cluster. On the other hand, the CH is the node that possesses a maximum energy level inside its cluster. As CHs perform functions that consume more energy, it is important to evenly re-elect another CH aiming at distributing the energy consumption among nodes. Indeed, each CH computes, periodically, the energy level average of all nodes within its cluster. After that, if its current energy level is below the average, a node with a maximum of remaining energy level is indicated as a new CH.

An illustrative example of the clustering procedure. In order to demonstrate the feasibility of our clustering algorithm, we draw Fig. 1 and Fig. 2 that represent an example of applying our algorithm on a network with 10 nodes. Further, all numeric values are obtained when executing the proposed algorithm on 10 nodes as shown in Fig. 1 and tabulated in Table 3.

Table 3. Clusters formation

Node ID	Step 1	Step 2	Step 3	Step 4
	NS	EC	W_{EC}	C
1	{1,2,3,4,5}	{1,2,4,5} {1,3}	3.3 6.5	{1,2,4,5}
2	{1,2,4,5}	{1,2,4,5}	3.3	{1,2,4,5}
3	{1,3,6,8}	{3,6,8} {1,3}	3 6.5	{3,6,8}
4	{1,2,4,5,6}	{1,2,4,5} {4,6}	3.3 5.5	{1,2,4,5}
5	{1,2,4,5,7,10}	{1,2,4,5} {5,7,10}	3.3 4	{1,2,4,5}
6	{3,4,6,7,8}	{3,6,8} {6,4} {6,7}	3 5.5 6.5	{3,6,8}
7	{5,6,7,9,10}	{7,9,10} {5,7,10} {6,7}	3.15 4 6.5	{7,9,10}
8	{3,6,8,9}	{3,6,8} {8,9}	3 5.5	{3,6,8}
9	{7,8,9,10}	{7,9,10} {8,9}	3.15 5.5	{7,9,10}
10	{5,7,9,10}	{7,9,10} {5,7,10}	3.15 4	{7,9,10}

Fig. 1 shows the initial configuration of the sensor nodes in the network with individual node IDs. A node can hear/broadcast beacons from the nodes that are within its transmission range. An edge between two nodes represents the fact that nodes are neighbors. For this example, we set the cluster size threshold at $\chi=3$ and we chose the unit distance between nodes arbitrarily. To simplify, we assume that nodes operate initially with the same energy levels. The weights considered in equation (8) are $\alpha=0.5$, $\beta=0.5$ and $\gamma=0$. By taking $\gamma=0$, we avoid considering the energy level resulting from the routing procedures.

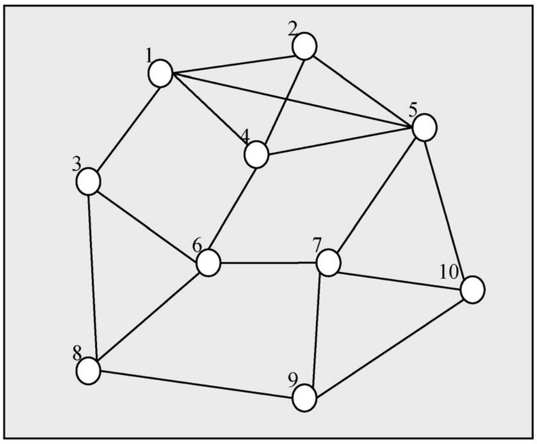

Fig. 1. Initial configuration of nodes

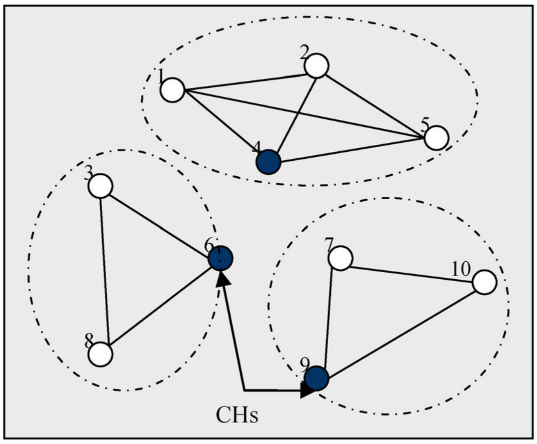

Fig. 2. Clusters identified

By executing the proposed algorithm on Fig. 1, we obtain: (1) the neighboring set of node *i* (NS_i) in Step 1; (2) the equivalence classes for each node, computed in Step 2; (3) the weighted metric (W_{EC}) for every equivalence class of each node, calculated in Step 3; and (4) the best equivalence class of each node (class with minimum weight W_{EC}) is selected as a cluster in distributed fashion in Step 4.

Routing paths between CHs formation. The second major activity within the setup phase is the formation of routing paths between CHs. DECHP uses a geographical and energy aware neighbor cluster heads selection to transfer aggregated data to the BS. Once clusters, as well as cluster head nodes, have been identified, each CH picks a next-hop node among all neighbor cluster heads that are closer to the BS. The CH routes the sensed data progressively towards the BS. Meanwhile, it tries to balance the energy consumption across its entire neighbor cluster heads. CH_i achieves this trade-off by minimizing the learned cost $l(CH_j, BS)$ value of its neighbor cluster head CH_j.

Each cluster head CH_i maintains a state $l(CH_i, BS)$ noted as a learned cost to BS. This value is communicated by a CH to its neighbor's CHs. In cases where the CH_i does not have $l(CH_j, BS)$ state for a neighbor CH_j, it computes the estimated cost $c(CH_j, BS)$ as a default value for $l(CH_j, BS)$. The estimated cost $c(CH_j, BS)$ of CH_j is defined as follows:

$$c\left(CH_j, BS\right) = \mu \times d(CH_j, BS) + \left(1-\mu\right) \times \frac{1}{C_e\left(CH_j\right)}. \qquad \textbf{(9)}$$

Where μ is a tunable weight, $d(CH_j, BS)$ is the distance from CH_j to the BS, and $C_e(CH_j)$ is the current energy level of the node CH_j.

Further, the cluster head CH_i selects the next-hop neighbor CH_{min} and initializes its own $l(CH_i, BS)$ value with $l(CH_{min}, BS) + c(CH_i, CH_{min})$, where $c(CH_i, CH_{min})$ is the cost to transmit a packet from CH_i to CH_{min}. The cost $c(CH_i, CH_{min})$ can also be considered as a combination function of both the remaining energy levels of (CH_i, CH_{min}) and the distance between two neighbors CH_i *and* CH_{min}.

At this point, a cluster head CH_i has learned a cost state $l(CH_j, BS)$ or an estimated cost function $c(CH_j, BS)$ for each neighbor cluster head CH_j. Accordingly, a cluster head CH_i that receives packets will pick the next hop among neighbor cluster heads that are closer to the BS, while minimizing the learned cost value $l(CH_j, BS)$.

Globally, the learned cost is a combination of consumed energy and distance. Hence, minimizing the learned cost value is a trade-off between routing towards the next-hop closest to the BS and balancing energy usage. By using the learned cost, DECHP distributes the burden of routing evenly among all cluster heads.

Scheduling creation. Schedule creation is the last issue related to the setup phase. The proposed algorithm uses Time Division Multiple Access (TDMA) scheduling scheme to minimize the collision between sensor nodes trying to transmit data to its cluster head. CHs act as local control centers to coordinate the data transmissions in their cluster. The CH node sets up a TDMA schedule and transmits it to all nodes in the cluster. Besides ensuring no collisions, the TDMA schedule allows the radio components of each non-cluster head node to be turned off at all times except during their transmit time, which reduces the energy consumed by individual sensors. Here, all nodes in the cluster know the TDMA schedule. Therefore, the setup phase is complete and the data communication phase can begin.

4.4 Data Communication Phase

The data communication phase consists of three major activities: (1) data gathering, (2) data fusion, and (3) data routing. By using the TDMA schedule described above, each sensor node can transmit the sensed information to its cluster head. Since sensor nodes are geographically grouped into clusters, these transmissions consume minimal energy due to small spatial separations between the cluster head and the sensing nodes. Once data from all sensor nodes have been received, the CH performs data aggregation on collected data following the correlation model described above, which reduces the amount of data to be sent to the BS. The aggregated data, along with information required by the BS to properly identify and decode the cluster data, are then routed back to the BS via the routing paths created between CHs. We also assume that aggregated data from a given CH undergoes further processing when traversing the routing path to the BS.

Another key issue that needs to be addressed here is the radio interference caused by neighboring clusters that could hinder the operation of any given cluster. DECHP utilizes Code-Division Multiple Access (CDMA) codes to counteract this problem. Thus, nodes belonging to the same cluster use the identical spreading code in order to transmit their data to the CH.

5 Simulation and Results

In order to evaluate the advantages of DECHP, we have constructed a set of simulations using NS-2 (Network Simulator). We compare our algorithm with other clustering-based routing protocols such as LEACH, LEACH-C, and PEGASIS. Performances are measured by quantitative metrics of average energy dissipation, total amount of data received by the BS, and the number of survival nodes. The communication energy consumption model is adopted from LEACH with the same parameters. Moreover, we assume the perfect-data-correlation model like LEACH and PEGASIS by setting λ in equation (5) at 0. According to the importance of keeping the cluster size as close as possible to the threshold χ, the weight α in equation (8) was taken at high level. On the other hand, distances' sum and battery power were taken at lower values of weights. The values used for simulation were α=0.5, β=0.25 and γ=0.25. Furthermore, we set μ in equation (9) at 0.5 and the cluster size threshold χ at 25. We note also that these values are arbitrary at this time and should be adjusted according to system requirements.

Throughout the simulations, we consider several random network configurations with 500 nodes where each node is assigned an initial energy of 2J. The data message size for all simulations is fixed at 500 bytes, of which 25 bytes represent the length of the packet header.

In the first experiment, we simulate different network topologies on an area of *100m*100m* with a BS distant from the nearest node of about 75*m*. Fig. 3 shows the number of survival nodes over time. We clearly depict that DECHP outperforms the system lifetime of LEACH, LEACH-C, and PEGASIS. We argue this by the fact that all cluster heads, in both LEACH and LEACH-C, transmit data directly to the BS, which in turn causes significant energy losses in the cluster head nodes. Both DECHP and PEGASIS alleviate this problem by having one node forwarding data to the BS. Nevertheless, we can note that DECHP exceeds in terms of system lifetime PEGASIS, since in DECHP CHs use geographical and energy-aware routing scheme to transmit aggregated data to the BS. This is well confirmed by the fact that utilizing an aggressive algorithm in PEGASIS results in a gradual increase in neighbor distances, which in turn increases the communication energy cost for those PEGASIS nodes that have distant neighbors. Note that increasing neighbor distances will have a significant effect on PEGASIS' performance when the area of the sensor field is increased.

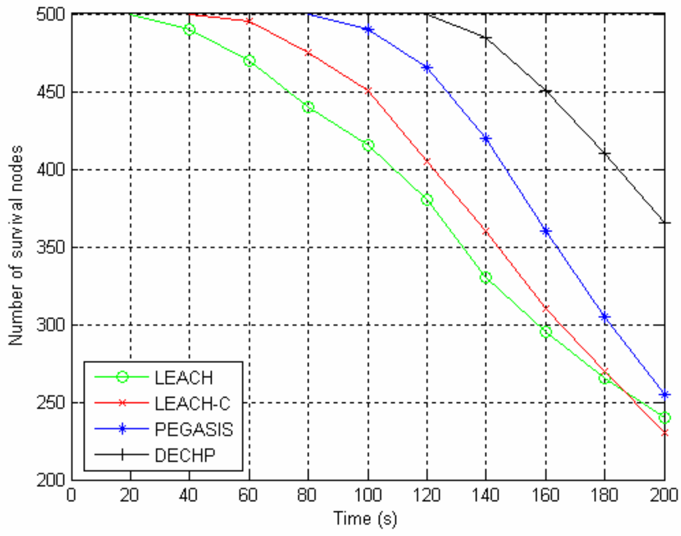

Fig. 3. System lifetime

The improvement gained through DECHP is further exemplified in Fig. 4. This plot presents the average energy dissipation of each protocol by report to the simulated time for a *100m*100m* network scenario. It is clearly seen that DECHP has a much more desirable energy expenditure curve than those of LEACH, LEACH-C, or PEGASIS. This improvement is expected to be more significant for networks with larger dimensions due to the same reasons cited in case of Fig. 3.

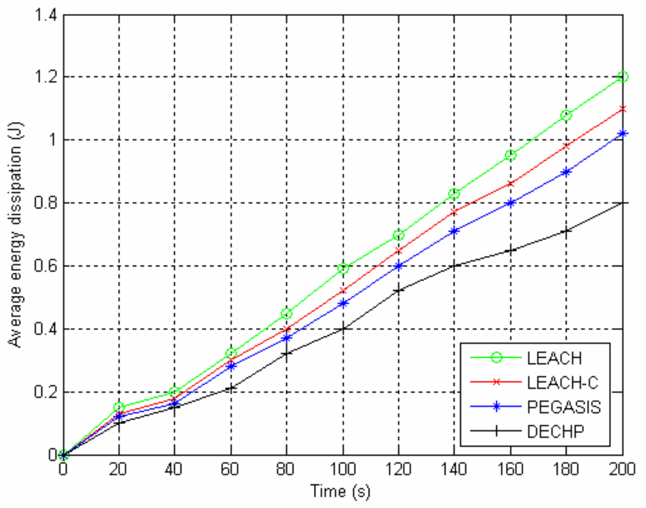

Fig. 4. Average energy dissipation

Next, we analyze the number of data messages received by the BS when considering the four routing protocols. For this experiment, we also simulate different network topologies on the same area 100m*100m where each node begins with an initial energy of 2J. Fig. 5 shows the total number of data messages received by the BS as a function of average energy dissipation. The plot illustrates clearly the capacity of DECHP to deliver more data messages than the other protocols.

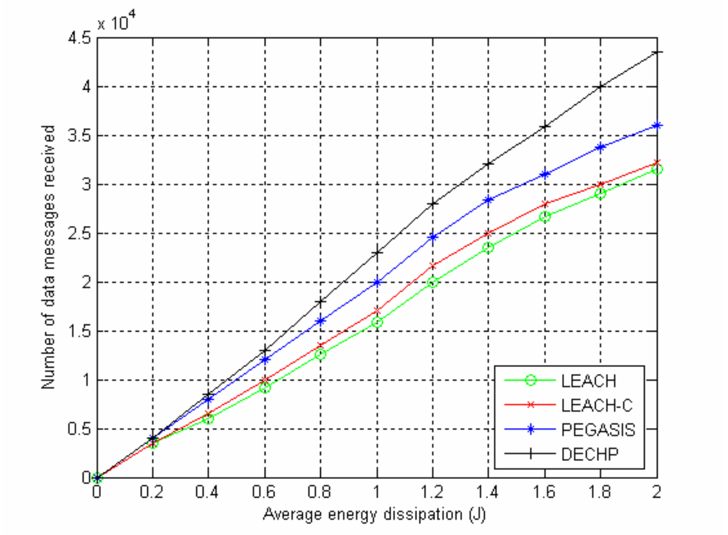

Fig. 5. Total amount of data received at the BS as a function of average energy dissipation

In the final experiment, we evaluate the performance of routing protocols by reporting the area of the sensor field. For this simulation, 500 nodes are randomly placed in a square where the base station is located at least at *100m* away from the closest sensor node. The results were obtained over several different network topologies (network area).

Fig. 6 shows the average energy dissipation of four protocols as a function of the network area. Clearly, DECHP outperforms both LEACH and LEACH-C as the network's area increases. We argue this by the fact that in LEACH (with its two versions), CHs are not uniformly placed across the whole sensor field, which led to an unfair cluster placement. In fact, the CHs in LEACH and LEACH-C can become concentrated in a certain region of the network, in which case nodes from the "cluster head deprived" regions will dissipate a considerable amount of energy while transmitting their data to a faraway cluster head. DECHP alleviates this problem by evenly allocating CHs across the sensor field. Another factor that leverages DECHP performances over LEACH and LEACH-C is the utilization of the balanced clustering approach. DECHP distributes the load evenly among the cluster heads, while in LEACH and LEACH-C some cluster heads are overloaded and the others can serve only a handful of nodes. On the contrary, we depict from this figure, a significant energy savings when using DECHP compared to PEGASIS.

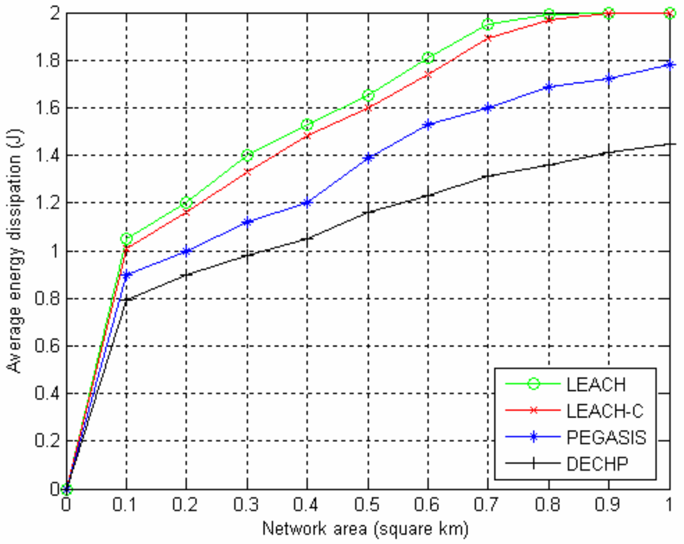

Fig. 6. Average energy dissipation over varying network areas

Finally, Fig. 7 illustrates the effectiveness of DECHP for wireless sensor applications that cover a large network area. As shown in this figure, approximately 70 percent of the DECHP nodes were still alive for a network with an area of 1 km2, while other protocols encounter significant sensor node deaths. From these analyses, it is obvious that DECHP offers a significant performance gain for networks with large coverage areas. In addition, we note that DECHP prominence over the other clustering-based protocols decreases as the sensor field area becomes smaller.

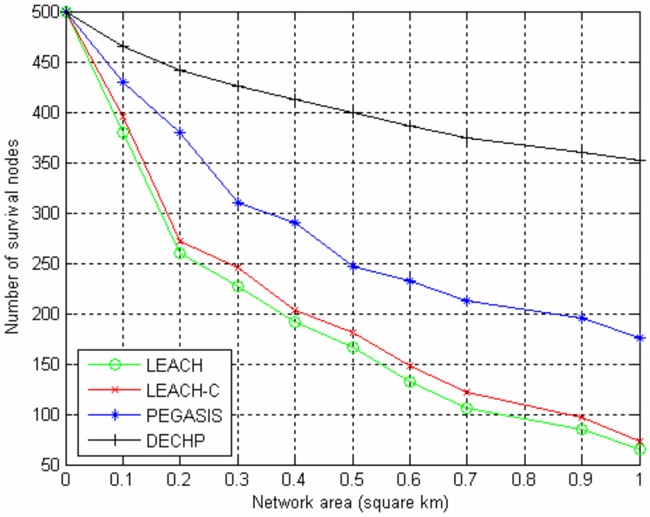

Fig. 7. Number of survival nodes as a function of network area

6 Conclusion

In this paper, we have proposed a novel routing protocol based on clustering called DECHP (Distributed Energy-efficient Clustering-based Hierarchy Protocol). Our protocol uses clustering procedure to organize sensor nodes into an energy-efficient hierarchy and relies on a geographical and energy aware neighbors' selection algorithm to establish dynamic routing paths between CHs. The performance of the proposed DECHP protocol is assessed by simulation and compared to other clustering-based protocols (LEACH, LEACH-C, and PEGASIS). The obtained results clearly show that DECHP outperforms the other cluster-based protocols by: *(i)* avoiding the periodical reelection of CHs and set-up phase; *(ii)* uniformly placing CHs throughout the whole sensor field; *(iii)* performing balanced clustering; and *(iv)* using a geographical and energy aware routing path between CHs to transfer aggregated data to the BS. Further, the performance gain of DECHP increases considerably when the area of the sensor field increases. Consequently, we can conclude that DECHP provides an energy-efficient routing scheme for a vast range of sensing applications.

In future work, we attempt to further improve the DECHP protocol to overcome some limitations such as the management of mobility and quality of service (QoS). We plan to integrate the QoS based on service differentiation mechanism in order to enhance the intra-cluster and inter-cluster communication models and support real time applications. Furthermore, we are considering the management of the nodes' mobility to extend DECHP use to mobile applications.

References

[1] A. Bröring, J. Echterhoff, S. Jirka, I. Simonis, T. Everding, C. Stasch, S. Liang, R. Lemmens, New generation sensor web enablement, Sensors 11(3)(2011) 2652-2699.

[2] International Electrotechnical Commission, Internet of things: wireless sensor networks. <http://www.iec.ch/whitepaper/internetofthings/>, 2014.

[3] X. Liu, A survey on clustering routing protocols in wireless sensor networks, Sensors 12(2012) 11113-11153.

[4] S. Basagni, I. Chlamtac, A. Fargo, A generalized clustering algorithm for peer-to-peer networks, in: Proc. Workshop on Algorithmic Aspects of Communication, Satellite Workshop of ICALP, 1997.

[5] O. Moussaoui, A. Ksentini, M. Naïmi, M. Gueroui, A novel clustering algorithm for efficient energy saving in wireless sensor networks, in: Proc. the 7th IEEE International Symposium on Computer Networks (ISCN'06), 2006.

[6] J. Kulik, W.R. Heinzelman, H. Balakrishnan, Negotiation based protocols for disseminating information in wireless sensor networks, Wireless Networks 8(2002) 169-185.

[7] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, Directed diffusion for wireless sensor networking, IEEE/ACM Transactions on Networking 11(1)(2003) 2-13.

[8] C. Schurgers, M.B. Srivastava, Energy efficient routing in wireless sensor networks, in: Proc. Military Communications Conference on Communications for Network-Centric Operations: Creating the Information Force, 2001.

[9] S. Sharma, A survey of hierarchical routing protocols in wireless sensor networks, International Journal of Advanced Research in Computer Science and Software Engineering 3(12)(2013) 140-146.

[10] W.R. Heinzelman, A.P. Chandrakasan, H. Balakrishnan, An application-specific protocol architecture for wireless micro sensor networks, IEEE Transactions on Wireless Communications 1(4)(2002) 660-670.

[11] O. Younis, S. Fahmy, HEED: a hybrid, energy-efficient, distributed clustering approach for ad-hoc sensor networks, IEEE Transactions on Mobile Computing 3(4)(2004) 366-379.

[12] A. Kazerooni, H. Jelodar, J. Aramideh, LEACH and HEED clustering algorithms in wireless sensor network: a qualitative study, Advances in Science and Technology Research Journal 9(25)(2015) 7-11.

[13] L. Buttyan, P. Schaffer, PANEL: position-based aggregator node election in wireless sensor networks, International Journal of Sensor Networks (2010) 1-16.

[14] S. Lindsey, C. Raghavendra, K. M. Sivalingam, Data gathering algorithms in sensor networks using energy metrics, IEEE Transactions on Parallel and Distributed Systems 13(9)(2002) 924-935.

[15] A. Manjeshwar, D.P. Agrawal, APTEEN: a hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks, in: Proc. the 2nd International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing, 2002.

[16] S. Jung, Y. Han, T. Chung, The concentric clustering scheme for efficient energy consumption in the PEGASIS, in: Proc. the 9th International Conference on Advanced Communication Technology, 2007.

[17] H. Dhawan, S. Waraich, A comparative study on LEACH routing protocol and its variants in wireless sensor networks: a survey, International Journal of Computer Applications 95(8)(2014) 0975-8887.

[18] A. Harter, A. Hopper, A new location technique for the active office, IEEE Personal Communication 4(5)(1997) 42-47.

[19] N.B. Priyantha, A. Chakraborty, H. Balakrishnan, The cricket location support system, in: Proc. 6th ACM Mobicom, 2000.

[20] A. Savvides, H. Park, M.B. Srivastava, The n-hop multilateration primitive for node localization problems, Proceedings of Mobile Networks and Applications 8(2003) 443-451.

[21] Ubisense, Enterprise location intelligence solutions. <https://ubisense.net/en> 2017 (accessed 24.07.2017).

[22] A. Woo, D. Culler, A transmission control scheme for media access in sensor networks, in: Proc. ACM/IEEE International Conference on MOBICOM, 2001.

[23] V. Raghunathan, C. Schurgers, S. Park, M.B. Srivastava, Energy-aware wireless microsensor networks, IEEE Signal Processing Magazine 19(2)(2002) 40-50.

[24] S. Pradhan, J. Kusuma, K. Ramchandran, Distributed compression in a dense microsensor network, IEEE Signal Processing Magazine 19(2)(2002) 51-60.