Journal of
Applied Sciences

An Energy-Efficient Node-Clustering Algorithm in Heterogeneous Wireless Sensor Networks: A Survey

Norah Tuah, Mahamod Ismail and Kasmiran Jumari
Department of Electrical, Electronic and Systems Engineering,
Faculty of Engineering and Built Environment,
Universiti Kebangsaan Malaysia, UKM Bangi, Selangor, 43600, Malaysia

Abstract: Wireless Sensor Networks (WSNs) have been widely used in many fields, including military, environmental, home and other commercial applications. A WSN is made up of a large number of sensor nodes which sense data and submit them to the sink node. One of the critical challenges in building such a network is the limited lifetime of the batteries in the sensor nodes. Many proposed algorithms have been published to overcome this problem. This study reviews some of these proposed algorithms involving a node-clustering scheme. The node-clustering algorithm with its terminology and attributes is also explained. In addition, the concept of Heterogeneous Wireless Sensor Networks (HWSNs) is explained in detail. Finally, a review and comparison of various energy-efficient clustering algorithms for HWSNs is presented.

Key words: Energy-efficient, node-clustering algorithm, heterogeneous wireless sensor networks

INTRODUCTION

A sensor network is an infrastructure composed of sensing (measuring), computing and communication elements that enable administrators to observe and react to events and phenomena in a specified environment (Sohraby *et al.*, 2007). Wireless Sensor Networks (WSNs) are used in air-traffic control, battlefield management, commercial application, food safety, intrusion detection and many other fields (Puccinelli and Haenggi, 2005; Akyildiz and Vuran, 2010). Each sensor node in a WSN is composed of a sensor, an A/D converter, a processor with memory (e.g., a StrongARM SA-1100 processor), RF circuits, a DC converter and a battery, as shown in Fig. 1. Each sensor node has a very limited energy supply and may not be rechargeable. In some cases, nodes may be self-powered. Energy is used for sensing, data processing and communication. Communication is the most energy-intensive process, consuming more energy than the computation process in the operating system (Akyildiz and Vuran, 2010). Therefore, a sensor node must be designed intelligently taking all its aspects into consideration, including architecture, algorithms, protocols and circuitry. The current study is focused on the design of an algorithm for sensor nodes that can prolong the lifetime of a WSN.

A WSN is a large-scale network that must be arranged in an appropriate manner, especially for data aggregation. The data aggregation process is intended to balance the load and thereby to extend the network lifetime. The node-clustering algorithm has been proven to be an effective way of organizing a network into a connected hierarchy. In a node-clustering algorithm, several nodes are aggregated to form a group. This scheme usually operates in two phases: node cluster setup and cluster maintenance (Zhang *et al.*, 2009). In the node-cluster setup phase, Cluster Heads (CHs) are selected among the nodes in the network using various selection schemes, as shown in Table 1. After the CHs have been selected, other nodes that are affiliated with each CH form clusters. Nodes that are not a CH are called non-CH nodes. Each CH, acting as a router, transmits data collected from non-CH nodes to the sink node. In the cluster maintenance phase, the cluster configuration may be changed after the initial cluster is set up because of node movements or topology changes. An example of a node-cluster structure is shown in Fig. 2.

The cluster head node will lose more energy than a non-cluster head node because it transfers data over longer distances. Hence, to distribute the load uniformly among all the nodes, the network must re-cluster itself periodically, selecting energy-abundant nodes to serve as CHs. Thus, the network will achieve energy efficiency, reduce channel contention and reduce packet collisions, resulting in better network throughput under high load (Younis and Fahmy, 2004).

Node clustering has been studied comprehensively in the data processing and wired network literatures.

Corresponding Author: Norah Tuah, Department of Electrical, Electronic and Systems Engineering,
Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor, Malaysia Tel: 60389216590

Table 1: Different cluster-head (CH) selection schemes.

| CH selection scheme	Characteristics
Stochastic CH election (Heinzelman *et al.*, 2002)	A node is randomly selected without inclusion of the remaining energy level available in each node
Weighted clustering algorithm (Xuegong *et al.*, 2010)	A node is randomly selected based on its energy level and distance; the number of times the node was selected as the CH is also considered
Deterministic CH election (Handy *et al.*, 2002)	A node is randomly selected with the inclusion of the remaining energy level available in each node.
CH election techniques for coverage preservation (Soro and Heinzelman, 2009)	A node was selected based on a set of coverage-aware cost metrics that favor nodes deployed in densely populated network areas
Centralized and Distributed CH selection (Tillapart *et al.*, 2004)	The sink node determines nodes' metric (CH) of all nodes in the cluster by taking into account the node's remaining energy and the total sum of squared distance to every node in the same cluster
CH selection using Fuzzy logic (Kim *et al.*, 2008)	A node was selected based on fuzzy if-then rule
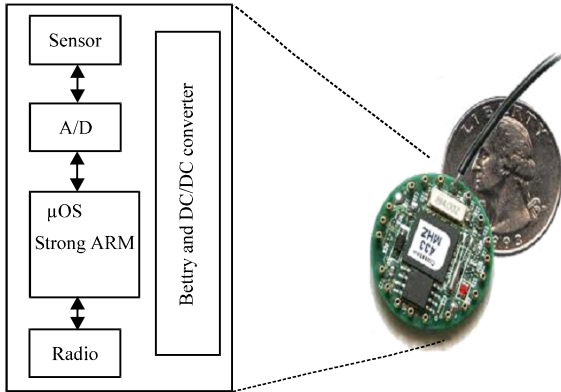
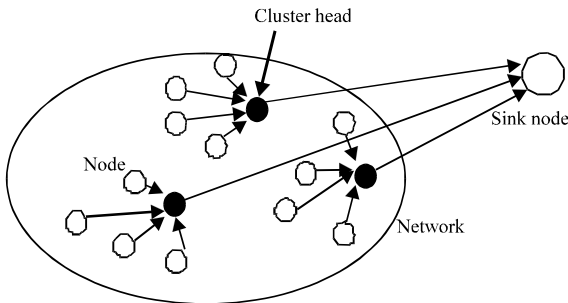Fig. 1: Sensor node architecture (Sohraby *et al.*, 2007)

Fig. 2: Example of a node-clustering structure in a WSN

Unfortunately, the clustering approaches developed and implemented in these areas cannot be used directly in WSNs because of the unique deployment and operational characteristics of these networks. The first study of cluster-based networks was carried out by (Heinzelman *et al.*, 2002). They proposed a Low-energy Adaptive Clustering Hierarchy (LEACH), a routing protocol with self-organizing and adaptive clustering behavior that randomly distributes the energy load among the sensors in a network. LEACH uses localized coordination to enable scalability and robustness for active networks. Moreover, it also uses data fusion in the network to reduce the amount of information that must be sent to the sink node. This study was followed by several attempts to implement a node-clustering algorithm in a WSN, including Hybird Energy-efficient and Distributed (HEED) (Younis *et al.*, 2006), Power-efficient Gathering in Sensor Information System

(PEGASIS) (Lindsey and Raghavendra, 2002), Tree-based Energy Efficient Protocol for Sensor Information (TREEPSI) (Satapathy and Sarma, 2006) and Advanced LEAH (ALEACH) (Ali *et al.*, 2008).

Most of the clustering algorithms named above assume that all sensor nodes are equipped with the same quantity of links, computational resources and energy; this is the case in homogeneous sensor networks. However, this kind of network cannot be assumed because different sensor nodes do not always have the same communication and sensing capabilities. The fact that sensor nodes use the same platform does not guarantee that they will have exactly the same physical properties (Das and Ammari, 2009). In contrast with homogeneous networks, there are also heterogeneous networks, called here Heterogeneous Wireless Sensor Networks (HWSNs). Recently, researchers have shown an increased interest in heterogeneous network research because of its ability to extend the lifetime of networks (Katiyar *et al.*, 2011), improve the reliability of data transmission (Yarvis *et al.*, 2005) and decrease the latency of data transfer (Yu *et al.*, 2007). A typical heterogeneous WSN consists of a large number of normal nodes and a few heterogeneous nodes (Duarte-Melo and Liu, 2002; Yu *et al.*, 2007). The normal node, whose main tasks are to sense and to issue data reports, is inexpensive and source-constrained. The heterogeneous nodes which provide data filtering, fusion and transport, are more expensive and more capable. A considerable amount of experimental research on HWSNs has been done by Du *et al.* (2008), who provided examples of two real sensor network deployments that use heterogeneous nodes for processing and transport tasks. Furthermore, many clustering algorithms have been proposed for an HWSN environment, for example, Stable Election Protocol (SEP) (Smaragdakis *et al.*, 2004), Distributed Energy Efficient Clustering (DEEC) (Qing *et al.*, 2006), Energy Efficient Hetrogenous Clusters (EEHC) (Kumar *et al.*, 2009) and others. This study presents a deeper explanation of this entire algorithm and a comparison with other algorithms.

NODE-CLUSTERING ALGORITHM

Recently, researchers have devoted considerable attention to node-clustering algorithms because of their
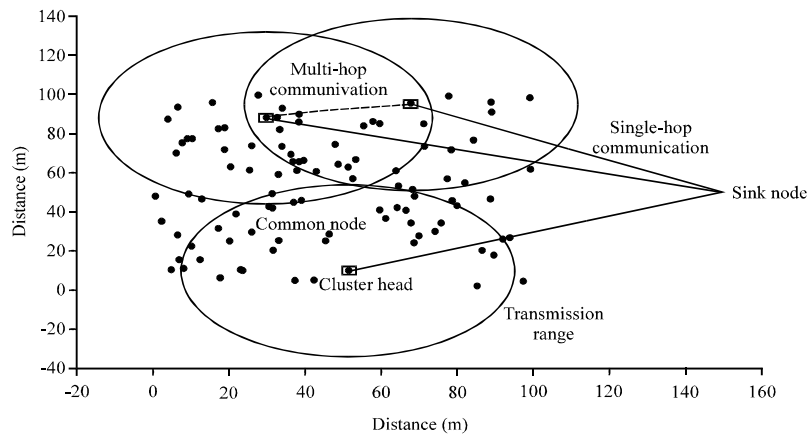
Fig. 3: Architecture of a wireless sensor network

ability to prolong the lifetime of WSNs. All the necessary terms involved in designing node-clustering algorithms are explained in this section.

Terminology

Homogeneous vs. heterogeneous: In a homogeneous network, all nodes in the WSN have the same storage, computation, communication, sensing and energy capabilities. The communication links between the sensor nodes are symmetric. As a result, in such a network, a pair of neighboring sensors can communicate directly; otherwise, the network is heterogeneous. The combination of both is called a hybrid network. The heterogeneous network is a fair assumption for practical implementations.

Optimum number of clusters: The number of clusters required in a network is closely related to achieving optimal energy consumption. In some published approaches, the set of CHs is predetermined and thus the number of clusters is preset. Randomly picking CHs from the deployed sensors usually yields a variable number of clusters. Typically, use of at least 10 clusters is recommended (Heinzelman *et al.*, 2002).

Stability: Stability is influenced by how the numbers of clusters in the network are counted. If the cluster counts vary and a node's membership evolves over time, this is called an adaptive scheme. Otherwise, the clustering scheme is called fixed because the number of clusters does not change and sensor nodes do not switch from one cluster to another in the network (Abbasi and Younis, 2007).

Cluster head (CH): The CH acts as a router that receives sensing data from all members of the cluster and transmits data to the sink node using single- or multi-hop data

transmission. Multi-hop data transmission is the appropriate technique for sink nodes located far away from the sensing area. Otherwise, single-hop data transmission is better. Sometimes the CH can be designated as a mobile or immobile node (Abbasi and Younis, 2007).

Non-cluster head node: Each node can become the CH in any round. A node that is not selected as the CH becomes a non-CH node which then selects its corresponding cluster based on its signal strength with the current CH.

Transmission range: This term refers to a disk of a given radius, including its boundary and center. Transmission range is usually the threshold that differentiates two kinds of communication schemes in the network, namely free space and multi-path fading models.

Single-hop vs. multi-hop: In contrast to multi-hop transmission, data in single-hop mode are transmitted from one node to another without using a third node as a router.

Mobile vs. immobile nodes: A node can either move within the network (mobile node) or be stationary (immobile node).

Figure 3 illustrates all the terminological definitions given above.

HETEROGENEOUS WIRELESS SENSOR NETWORKS (HWSNs)

Types of heterogeneous resources: Yarvis *et al.* (2005) list three types of resource heterogeneity in sensor nodes. The first, computational heterogeneity, means that at least one node contains a more powerful microprocessor and larger memory than a normal node. The second type is

known as link heterogeneity and means that at least one heterogeneous node has higher bandwidth and a longer-distance network receiver than a normal node. The third type, energy heterogeneity, means that at least one heterogeneous node is line-powered or that its battery is replaceable.

For the most part, energy heterogeneity is the most vital type of resource heterogeneity. A sensor network generally suffers a negative impact when both computational and link heterogeneity have consumed large quantities of energy resources. Therefore, energy heterogeneity must be given top priority when creating a heterogeneous environment in a WSN.

The impact of heterogeneous resources: The main benefit of a heterogeneous network is extended network lifetime. Average energy consumption can be decreased by forwarding packets from normal nodes to the sink node (Yu *et al.*, 2007). In addition, if each hop in the network can significantly reduce its end-to-end delivery time, this can improve the reliability of data transmission because there will be fewer hops between the normal sensor nodes and the sink. Consequently, use of computational or link heterogeneity can decrease the latency of data transfer in the network.

OBJECTIVES OF HETEROGENEOUS CLUSTERING

Nodes are formed into clusters to attain certain objectives, as explained below:

Enhanced network lifetime: The lifetime of the network can be extended by re-energizing the sensor nodes. This will create heterogeneity because some of these nodes are equipped with more energy than the normal nodes already in use. Consequently, all the normal nodes can transmit sensed data to the sink node through the nearest heterogeneous node. This will make the average energy consumption for forwarding data to the sink node much less than in the homogeneous sensor network.

Load balancing: LEACH guarantees, in the homogeneous case, that any unstable region in the network will be of limited size. No such guarantee exists for heterogeneous networks. As a solution, a weighted probability has been introduced to increase the stable region in a network which is implemented as a heterogeneous network (Smaragdakis *et al.*, 2004).

Decreasing the latency of data transfer: HWSNs with computational heterogeneity may decrease the processing

latency for immediately adjacent nodes. Moreover, if link heterogeneity is implemented in the HWSNs, this will reduce the waiting time in the transmission queue.

Improving the reliability of data transmission: In a heterogeneous network, the number of hops to transmit data from sensor nodes to the sink node is less than in a homogeneous network (Yu *et al.*, 2007). For this reason, HWSNs can achieve much higher end-to-end delivery speeds.

CLUSTERING ATTRIBUTES

Clustering attributes are a vital part of categorizing and differentiating the various clustering algorithms which have been proposed for HWSNs. There are three types of attributes of clustering algorithms (Abbasi and Younis, 2007).

Cluster properties: Important cluster attributes include the optimal number of clusters, the stability of the network and the data transmission technique used for intra-cluster or inter-cluster connectivity. Normally, direct communication is selected for nearby data communication and multi-hop data communication for long-distance communication.

Cluster head capabilities and selection: The CH is a particularly powerful node in a clustering algorithm. It can be either mobile or immobile. When a CH is mobile, its cluster's sensor membership changes dynamically and clusters will need to be continually maintained. If the CH is immobile, a good choice of CH yields a stable cluster which facilitates intra- and inter-cluster network management. Each proposed algorithm has its own types of node heterogeneity, as was explained in the previous section. Furthermore, sometimes the CH acts as a data aggregation or relay node. All these tasks vary depending on the network implementation. Finally, the CH selection scheme varies depending on many factors such as the initial, residual and average energy of all the nodes in the network.

Clustering process: The methodology, the objective of node clustering and the cluster-head selection method are important factors that must be considered during the clustering process. For HWSNs, various methodologies can be selected, including adaptive clustering, energy multi-level clustering and stability-oriented clustering. Consequently, the objectives of clustering must be considered with respect to load balancing, network

connectivity and similar tasks. Furthermore, a cluster-head selection method is needed to determine the CH using stochastic or deterministic methods or a combination of both.

ENERGY-EFFICIENT NODE-CLUSTERING ALGORITHM FOR HWSNs

Energy consumption model: Energy consumption by the sensor nodes in WSNs can be calculated using the energy model proposed by Heinzelman *et al.* (2002). The energy consumption for sending k bits of data over a distance d can be calculated using Eq. 1. The distance threshold d_0 will differentiate the type of data communication. The first equation is the free-space model, in which the transmission power is attenuated to d^2 for $d < d_0$. The second equation is the multi-path fading model, in which the transmission power is attenuated to d^4 for $d > d_0$:

$$E_{Tx}(k,d) = \begin{cases} kE_{elec} + k\varepsilon_{fiss-amp}d^2 & : d < d_0 \\ kE_{elec} + k\varepsilon_{two-ray-amp}d^4 & : d > d_0 \end{cases} \qquad (1)$$

where, $\varepsilon_{fiss-amp}$ and $\varepsilon_{two-ray-amp}$ are the two-channel model parameters for the energy needed for power amplification. Consequently, the energy consumed when receiving k bits of data can be calculated using Eq. 2:

$$E_{rx}(k) = E_{Rx-elec}(k) = kE_{elec} \qquad (2)$$

where E_{elec} represents the energy consumed in transmitting or receiving a bit of data.

NODE-CLUSTERING ALGORITHM FOR HWSNs

Node-clustering algorithms are a well-known technique for reducing energy consumption in a sensor network as well as increasing the lifetime of the network. Currently, several types of node-clustering algorithm exist each with different clustering attributes as explained earlier. The following section will discuss some of these techniques.

Energy-efficient stability-oriented clustering: The first design for a heterogeneous WSN was reported by Smaragdakis *et al.* (2004). This study examined the effect of a heterogeneous network with a cluster-based clustering implementation scheme. Furthermore, it reported a simulation of the LEACH routing protocol for heterogeneous WSNs. The LEACH routing protocol was developed for homogeneous network implementation.

According to Smaragdakis *et al.* (2004), when the node population starts to decrease, the number of cluster heads chosen per round becomes unstable (fewer than intended) and therefore there is no guarantee that a constant number of cluster heads (equal to $n \times P_{opt}$) will be selected per round per epoch. Consequently, a new clustering algorithm was proposed to increase the stable region and thereby decrease the unstable region and improve the feedback quality of the wireless clustered sensors in the presence of a heterogeneous network. Some percentage of the population of sensor nodes is equipped with additional energy resources; this is a source of heterogeneity which may result from initialization. Hence, two types of nodes, called normal and advanced nodes, were designed. The selection of each cluster head depends on a threshold calculation as shown in Eq. 3 for a normal node and Eq. 4 for an advanced node:

$$T(S_{nrm}) = \begin{cases} \dfrac{Pnrm}{1 - pnrm(rm\,od\dfrac{1}{pnrm})} & if\,S_{nrm} \in G' \\ 0 & otherwise \end{cases} \qquad (3)$$

$$T(Sadv) = \begin{cases} \dfrac{Padv}{1 - padv(rm\,od\dfrac{1}{padv})} & if\,S_{adv} \in G' \\ 0 & otherwise \end{cases} \qquad (4)$$

where, P_{nrm} and P_{adv} are the weighted probabilities for normal and advanced nodes, r is the current round identifier and G' (G'') is the set of normal (advanced) nodes that have not become the cluster head. In the clustering process, a normal node (S_{nrm}) or an advanced node (S_{adv}) will choose a random number between zero and one. If the number selected is less than a threshold value T(S), the node becomes a CH for the current round. Otherwise, the node is not selected as a CH, but possibly in the next round it will have a chance to become the new CH. The simulation revealed that network stability was achieved and that the lifetime of the network could have been increased.

Energy-efficient adaptive clustering: In an energy-efficient adaptive clustering algorithm, sensor nodes organize themselves into local clusters, with one node in each cluster acting as the cluster head node. Certain factors must be considered in selecting which node will become the cluster head. Careful selection can prolong the lifetime of the Heterogeneous Wireless Sensor Network (HWSN). Qing *et al.* (2006), Changmin and Hong (2007) and Elbhiri *et al.* (2010) revealed

that consideration of the residual and average energy of the network during cluster-head selection can increase the lifetime of the HWSN. Moreover, distance (Jiun-Jian *et al.*, 2009) and communication cost (Liu *et al.*, 2011) attributes can be added to the clustering algorithm to enhance the lifetime of the HWSN.

Qing *et al.* (2006) proposed a clustering algorithm for two-level HWSNs called Distributed Energy-efficient Clustering (DEEC). The first-level nodes are designated as the advanced nodes and the second level consists of the normal nodes. The total initial energy of the HWSN is given by Eq. 5:

$$E_{total} = N(1-m)E_0 + NmE_0(1+a) = NE_0(1+am) \qquad (5)$$

where, N is the number of nodes, E_0 is the initial energy of the normal nodes and m is the proportion of advanced nodes which possess a times more energy than normal nodes. In addition, a multi-level HWSN was considered and its total initial energy was calculated using Eq. 6:

$$E_{total} = \sum_{i=1}^{N} E_O(1+a_i) = E_O\left(N + \sum_{i=1}^{N} a_i\right) \qquad (6)$$

where, a_i is times more energy for node i.

In the clustering algorithm sequence, the nodes were distributed uniformly and one node in each cluster was selected as the cluster head node. The selection of the cluster head was based on the initial and residual energy left in the node at each round. The nodes with high residual energy become cluster heads more often than lower-energy ones. To organize the nodes into an HWSN, a weighted probability equation has been defined for normal and advanced nodes and is stated as Eq. 7 for a two-level HWSN:

$$P_i = \begin{cases} \dfrac{P_{opt}E_i(r)}{(1+am)\overline{E(r)}} & \text{if } s_i \text{ is a normal node} \\ \dfrac{P_{opt}(1+a)E_i(r)}{(1+am)\overline{E(r)}} & \text{if } s_i \text{ is an advanced node} \end{cases} \qquad (7)$$

where, p_{opt} is the average probability of becoming a cluster head, $E_i(r)$ is the residual energy of node s_i in round r and $\overline{E(r)}$ is the average energy in round r. The weighted probabilities for multi-level HWSNs are given by Eq. 8:

$$p(S_i) = \frac{p_{opt}N(1+a_i)E_i(r)}{(N+\sum_{i=1}^{N}a_i)\overline{E(r)}} \qquad (8)$$

The $\overline{E(r)}$ value was estimated to make sure that the DEEC protocol can guarantee that all nodes die at approximately the same time. Each non-cluster head node sends L bits of data per round to the cluster head.

Changmin and Hong (2007) have implemented a similar clustering algorithm concept as a DEEC protocol. Their work showed that consideration of the ratio between remaining energy and the average energy of the network can extend the lifetime of the wireless network.

Incorporation of energy into cluster-head selection has been studied by Elbhiri *et al.* (2009) for two-level HWSNs. Their work was based on a DEEC scheme with certain improvements involving the use of a threshold residual energy value, as shown in Eq. 9. This equation is intended to ensure that advanced and normal nodes have the same probability to become a cluster head. As a result, cluster-head selection will be more balanced and more equitable:

$$Th_{REV} = E_0\left(1 + \frac{\alpha E_{disNN}}{E_{disNN} - E_{disAN}}\right) \qquad (9)$$

where, E_{disNN} is the energy dissipated by a normal node in each round and E_{disAN} is the energy dissipated by an advanced node. Finally, the weighted probability that each node will be selected as a cluster head is given by Eq. 10:

$$P_i = \begin{cases} \dfrac{p_{opt}E_i(r)}{(1+am)\overline{E(r)}} & \text{for nml nodes}, E_i(r) > Th_{REV} \\ \dfrac{(1+a)p_{opt}E_i(r)}{(1+am)\overline{E(r)}} & \text{for adv nodes}, E_i(r) > Th_{REV} \\ \dfrac{(1+a)p_{opt}E_i(r)}{(1+am)\overline{E(r)}} & \text{for adv, nml nodes}, E_i(r) \le Th_{REV} \end{cases} \qquad (10)$$

where, c is a real positive variable which directly controls the number of cluster heads.

Jiun-Jian *et al.* (2009) implemented an algorithm similar to the SGCH algorithm but in a heterogeneous environment. In this algorithm, every node is divided into several clusters according to the distances between nodes. There are three types of nodes, referred to as normal, moderately advanced and highly advanced nodes. In the clustering process, the node in each cluster with the maximum energy remaining will be selected as the cluster head. Subsequently, each node forwards its sensed data to the nearest cluster head. The cluster head will then transmit the data to the sink node.

This network performs an environmental monitoring task and its sensor nodes monitor a variety of objects. A recent study by Liu *et al.* (2011) involved nodes distributed randomly with mobile or stationary properties. They assumed that the node communication links are symmetric and that nodes do not have any location information, but can calculate the distance between nodes according to the signal strength received. The distance between nodes i and j was calculated using Eq. 11:

$$d_{i,j} = \alpha \sqrt{\frac{K \times E_i^{tran}}{E_{j,i}^{rec}}} \qquad (11)$$

where, α is a distance-energy gradient with a value varying from 1-6 according to the physical environment in which the sensor networks operate; E_i^{tran} is the transmission energy for broadcasting and $E_{j,i}^{rec}$ is the received signal strength (received energy). In the clustering algorithm, the node establishes a routing table of neighboring nodes based on received data and saves all relevant information for all nodes within its communication range. The nodes with more residual energy should be given greater opportunity to become cluster heads and all nodes take turns at being cluster head nodes. The threshold equation for the probability of selecting sensor nodes as a cluster head is given here as Eq. 12:

$$T(i) = \begin{cases} \dfrac{p_i}{1 - p_i \left(r \bmod \dfrac{1}{p_i}\right)} [(\alpha \bar{\omega}(E)_i + \beta \bar{\omega}(C)_i)] + \left(r_i \, div \dfrac{1}{p_i}\right)(1 - (\alpha \bar{\omega}(E)_i + \beta \bar{\omega}(C)_i) & \text{if } i \in G \\ 0 & \text{otherwise} \end{cases} \qquad (12)$$

where, α and β are coefficients, r is the round identifier, $\bar{\omega}(E)_i$ is an energy factor and $\bar{\omega}(C)_i$ is a communication factor. The algorithm operates in an iterative fashion. After completion of one round, a new node must be selected as the cluster head for the next round. To reduce the energy consumed when broadcasting and the communication cost, an energy consumption prediction mechanism for Regular Data Acquisition (RDA) nodes has been developed. Each node will determine whether its current residual energy is close to the residual energy predicted in the last round using Eq. 13:

$$\gamma = \left| 1 - \frac{E_{jr\_predicted}}{E_{jr}} \right| \qquad (13)$$

where, the value of γ is less than the amplifier energy, $E_{jr,predicted}$ is the predicted residual energy of node j at the beginning of round r and E_{jr} is the current energy of node j.

Energy-efficient multi-level clustering: Multi-level clustering is achieved by multi-level cluster head node selection during clustering algorithm execution. According to Hasnaoui *et al.* (2010) and Soni and Katiyar (2011) use of this technique can reduce and balance energy consumption and hence prolong the lifetime of a WSN. This can also be achieved using multi-hop data transmission, in which each cluster head

acts as a relay node during data transmission (Kumar *et al.*, 2010; Han, 2010). In addition, the network can be divided into two or more zones or tiers to prolong its lifetime, as suggested by Li *et al.* (2007a) and Mehrotra and Leong (2009).

Hasnaoui *et al.* (2010) carried out studies in a two-level HWSN of a multi-level clustering algorithm known as the Oriented Cluster Formation (OCFLE) algorithm. OCFLE is an oriented cluster formation algorithm based on the position of the CHs in the sensing area and on the minimum distance between a non-cluster head node and the sink node using a suitable intermediate CH.

In the clustering process, the cluster head node was selected based on the ratio between the residual energy of each node and the average energy of the network. A two-level hierarchical design was introduced in which suitable intermediate CHs for data transmission were developed. The CH with the most energy, called MaxCH, is the most likely to be chosen as the intermediate CH used to transmit aggregate data to the Base Station (BS). MaxCH collects all data coming from all CHs, compresses them into a single signal and sends them directly to the base station. The multi-level clustering network concept is illustrated in Fig. 4.

Recently, Soni and Katiyar (2011) have developed a clustering algorithm for two-level HWSNs which introduces the concept of Minimum Reachability Power (MRP). This concept takes into consideration the transmission range among the nodes in the network. Eq. 14 expresses the probability of cluster head node selection during the clustering process:

$$P_1^n(u) = N_{CH1}^{opt} \left[\phi \frac{E_u(t)}{\sum_{v \in S(t)} E_v(t)} + (1 - \phi) \left(\frac{\frac{1}{P_u}}{\sum_{v \in S(t)} \frac{1}{P_v}} \right) \right]$$

$$P_1^a(u) = N_{CH1}^{opt} \left[\phi \frac{E_u(t) * \gamma}{\sum_{v \in S(t)} E_v(t)} + (1 - \phi) \left(\frac{\frac{1}{P_u}}{\sum_{v \in S(t)} \frac{1}{P_v}} \right) \right] \qquad (14)$$

where, P_1^n is the probability of becoming CH for a normal node and $P_1^n(u)$ the probability of becoming CH for an advanced node, ϕ is a parameter which determines the weighting of energy and distance factors, N_{CH1}^{opt} is the optimal number of CHs for level 1, $E_u(t)$ is the energy of the nodes, γ is the energy multiplier for advanced nodes compared to normal nodes, $\sum_{v \in S} E_v(t)$ is the total remaining energy of the network and P_u and P_v are the MRP values. Eq. 15 shows the probability calculation for level-2 cluster-head selection:
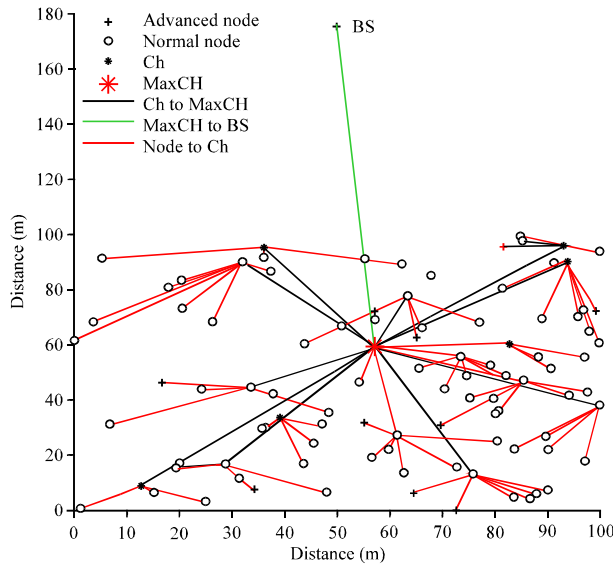
Fig. 4: Dynamic oriented cluster formation using OCFLE (Hasnaoui *et al.*, 2010)

$$P_2^n(u) = \sqrt{N_1} \left[\phi \frac{E_u(t)}{\sum_{v \in S(t)} E_v(t)} + (1-\phi) \left(\frac{\frac{1}{p_u}}{\sum_{v \in S(t)} \frac{1}{p_v}} \right) \right]$$

$$P_2^\alpha(u) = \sqrt{N_1} \left[\phi \frac{E_u(t) * \gamma}{\sum_{v \in S(t)} E_v(t)} + (1-\phi) \left(\frac{\frac{1}{p_u}}{\sum_{v \in S(t)} \frac{1}{p_v}} \right) \right] \quad (15)$$

where, N_1 is the cardinality of the active set. After a level-T clustering topology has been formed, the regular nodes start transmitting their sensed data to their CHs. Level-T CHs aggregate the sensed data and send them to level-(T-1) CHs and so forth. Finally, all level-1 CHs transmit the aggregated data to the sink node.

Kumar *et al.* (2010) extended the clustering algorithm to include inter-cluster transmission considerations. Their research is an extension of their previous work on the EEHC algorithm. A new mechanism was added to the clustering algorithm, as shown in Fig. 5. The diagram shows that the sensed data were transmitted to the sink node using multi-hop data transmission. The path metrics and their route costs were formed into an array matrix called an "adjacency matrix." In each cluster, the cluster head receives the messages and then aggregates the received messages with its own message. The aggregated message is then forwarded to the next-hop receiver (sink or CH), based on the routing information calculated using the optimal path approach.

Han (2010) examined multiple-hop routing in a clustering algorithm for HWSNs and proposed a three-level heterogeneous network. The nodes organize themselves into local clusters, with one node acting as
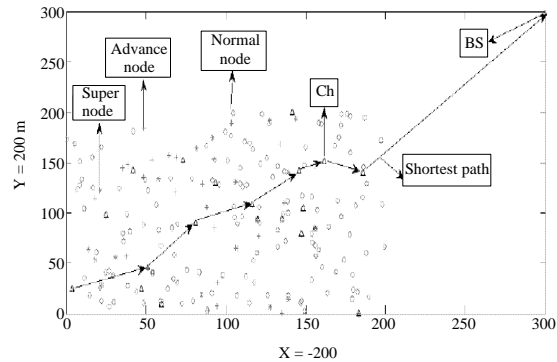

Fig. 5: General operation of the algorithm with BS at (300, 300) (Kumar *et al.*, 2010)

the cluster head in each cluster. The node with the most remaining energy will be selected as the cluster head. The CH node sets up a TDMA schedule and transmits it to the nodes in the cluster. Non-cluster head nodes will join a CH based on the distance calculation given in Eq. 16:

$$C_{join} = *E_{remaining} + (1 -)*(1 - \arctan(d) * \frac{2}{\pi}) \quad (16)$$

where, d is the distance and $E_{remaining}$ the remaining energy of the cluster-head candidate. Each CH finds its own optimum path to the sink node using a minimum spanning tree algorithm. This technique can be used for multiple-hop data transmission which can achieve short communication distances between the cluster heads and the sink node.
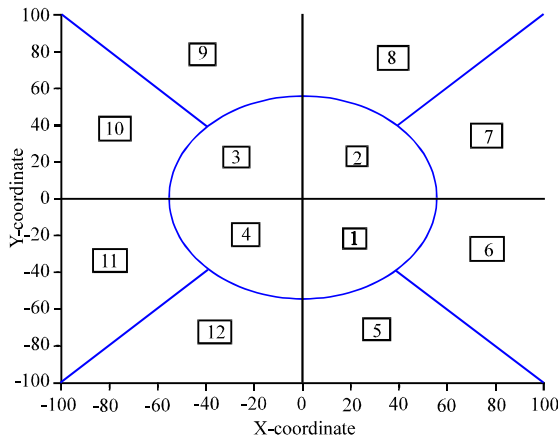
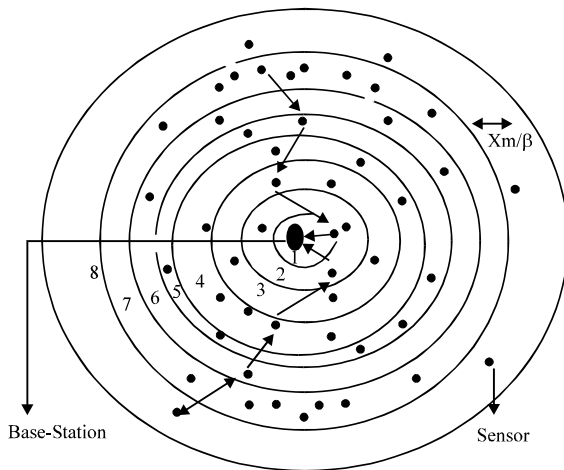Fig. 6: Schematic diagram of different-sized zones

Fig. 7: Illustration of routing in NSEEAR (Mehrotra and Leong, 2009)

Li *et al.* (2007b) investigated a zone-based clustering algorithm for two-level HWSNs called zone-based REECR (ZREECR). Their work was based on the Residual Energy and Energy Consumption Rate (REECR) routing protocol. In the clustering process, the cluster head nodes were evenly spread throughout the network which can minimize the distance over which the non-cluster head nodes need to send their data. The algorithm uses a zone-based routing protocol which divides the network field into several zones of different sizes depending on their distance and orientation from the sink node. Figure 6 shows an example of a two-layer zone of the network.

Cluster head nodes closer to the sink node have smaller zones because they consume less energy for intra-cluster data processing and can preserve more energy for inter-cluster relay traffic. Consequently, the choice of the cluster head in each zone is based on residual energy and on energy consumption rate, as shown in Eq. 17:

$$p_i = \frac{E_i^{\alpha}(t)}{V_i^{\beta}(t)} \qquad (17)$$

where, α and β are weighting coefficients, $E_i(t)$ is the current residual energy of node i and $V_i(t)$ is the energy consumption rate of node i. Hence, all the data collected in the cluster head will be forwarded to the base station over multiple hops, in which every cluster head chooses to forward its data to the closest cluster head in the direction of the base station. As a result, this algorithm is approximate, but not more energy-efficient than REECR. Mehrotra and Leong (2009) conducted a simulation of a two-level HWSN in various network tiers, as shown in Fig. 7. The lowest tier is closest to the base station. Every sensor node belongs to one of these tiers.

This kind of network is guaranteed to have two properties: first, to maintain a short distance between transmitter and receiver and second, to choose relay nodes on the basis, not only of each node's distance from the transmitter, but also of its residual energy. In the clustering process, this algorithm uses a node-by-node adaptive approach which can use either direct transmission or the multi-hop scheme depending on the residual energy limitations of each node. The communication process in the network is governed by two sets of rules. The first set of rules applies when the residual energy of a node is greater than 30% of its initial energy while the second set of rules applies when the residual energy of a node is less than 30% of its initial energy. This technique can increase the efficiency of the algorithm and help to prolong its period of stability.

All the algorithms discussed in this section are summarized in Table 2. Table 3 then classifies these algorithms based on the attributes discussed. Finally, Table 4 presents a comparison of the cluster-head node selection process used for the various algorithms.

PERFORMANCE EVALUATION METRICS

Some of the parameters that can be used to evaluate the performance of node-clustering algorithms are:

Average energy dissipated: The average energy usage by a sensor node in the network over time for various purposes such as transmitting, receiving and aggregating data.

Table 2: Comparison of various types of energy-efficient node-clustering algorithm

Clustering approach	Type of heterogeneity	Level of heterogeneity	Hierarchy network	Location awareness	Energy efficiency	Balanced clustering
SEP						
Smaragdakis *et al.* (2004)	Energy	Two	No	Not required	Moderate	Yes
DEEC						
Qing *et al.* (2006)	Energy	Two/multi-level	No	Not required	High	-
DEBC						
Changmin and Hong (2007)	Energy	Two/multi-level	No	Not required	Moderate	-
DDEEC						
Elbhiri *et al.* (2010)	Energy	Two	No	Not required	Moderate	-
Based on SGCH						
Jiun-Jian *et al.* (2009)	Energy	Three	Yes	Yes	Moderate	
EEPCA						
Liu *et al.* (2011)	Energy and data-acquisition-regularity	Multi level	no	yes	moderate	-
OCFLE						
Hasnaoui *et al.* (2010)	Energy	Two	Yes	Yes	High	-
PAMC						
Soni and Katiyar (2011)	Energy	Two	Yes	Yes	Moderate	-
EEMCP						
Kumar *et al.* (2010)	Energy	Three	Yes	Yes	High	-
LEACH-HPR						
Han (2010)	Energy	Three	Yes	Yes	Moderate	-
ZREECR						
Li *et al.* (2007a)	Energy	Two	Yes	Yes	Low	-
NSEEAR						
Mehrotra and Leong (2009)	Energy	Two	Yes	Yes	Moderate	-

Table 3: Classification of energy-efficient node-clustering algorithms based on clustering attributes

Clustering approach	Cluster properties		CH to node sink connectivity	Cluster head capabilities		Clustering process	
	Cluster count	Intra-cluster topology		Mobility	Role	Methodology	Objective of node grouping
SEP							
Smaragdakis *et al.* (2004)	Adaptive	Single-hop	Direct link	Not mobile	Aggregation	Distributed	Load balancing
DEEC							
Qing *et al.* (2006)	Adaptive	Single-hop	Direct link	Not mobile	Aggregation	Distributed	Prolong network lifetime
DEBC							
Changmin and Hong (2007)	Adaptive	Single-hop	Direct link	Not mobile/mobile	Aggregation	Distributed	Prolong network lifetime
DDEEC							
Elbhiri *et al.* (2010)	Adaptive	Single-hop	Direct link	Not mobile	Aggregation	Distributed	Prolong network lifetime
Based on SGCH							
Jiun-Jian *et al.* (2009)	Fix	Single hop	Direct link	Not mobile	Aggregation	Centralized	Prolong network lifetime
EEPCA							
Liu *et al.* (2011)	Adaptive	Single hop	Direct link	Mobility/not mobile	Aggregation	Distributed	Prolong network lifetime
OCFLE							
Hasnaoui *et al.* (2010)	Adaptive	Single hop	Multi-hop	Not mobile	Aggregation/relaying	Distributed	Prolong network lifetime
PAMC							
Soni and Katiyar (2011)	Adaptive	Single-hop	Single-hop	Not mobile	Aggregation	Distributed	Prolong network lifetime
EEMCP							
Kumar *et al.* (2010)	Adaptive	Single Hop	Multi-hop	Not mobile	Aggregation/relaying	Distributed	Prolong network lifetime
LEACH-HPR							
Han (2010)	Adaptive	Multi-hop	Multi-hop	Not mobile	Aggregation	Distributed	Prolong network lifetime
ZREECR							
Li *et al.* (2007b)	Fix	Fixed (1-hop)	Multi-hop	Not mobile	Aggregation	Centralized	Prolong network lifetime
NSEEAR							
Mehrotra and Leong (2009)	Adaptive	Single hop/multi-hop	Direct link	Not mobile	Aggregation/relaying	Distributed	prolong network lifetime

Network lifetime: The time interval from initial node activation until network death.

Number of cluster heads per round: The number of cluster heads chosen in each round which represents the number of nodes which will send data directly to the sink node. Comparison of cluster head selection approaches is shown in Table 4.

Total number of nodes alive: The number of nodes that remain alive in each round.

Number of data messages/throughput: The number of data messages received by the sink node. If the volume of data messages received is too big for the sink node, it will make the sink node die and energy consumption will decrease.

Table 4: Comparison of cluster-head selection approaches

| Clustering approach | CH selection based on | | | |
	Initial energy	Residual energy	Energy consumption	Average energy of network
SEP				
Smaragdakis *et al.* (2004)	No	Yes	No	No
DEEC				
Qing *et al.* (2006)	No	Yes	No	Yes
DEBC				
Changmin and Hong (2007)	Yes	Yes	No	Yes
DDEEC				
Elbhiri *et al.* (2010)	Yes	Yes	No	No
Based on SGCH				
Jiun-Jian *et al.* (2009)	Yes	Yes	No	Yes
EEPCA				
Liu *et al.* (2011)	No	Yes	No	Yes
OCFLE				
Hasnaoui *et al.* (2010)	No	Yes	No	Yes
PAMC				
Soni and Katiyar (2011)	No	No	No	Yes
EEMCP				
Kumar *et al.* (2010)	No	Yes	No	No
LEACH-HPR				
Han (2010)	No	Yes	No	No
ZREECR				
Li *et al.* (2007b)	No	Yes	Yes	No
NSEEAR				
Mehrotra and Leong (2009)	No	Yes	No	No

CONCLUSIONS AND FUTURE TRENDS

A survey of the current state-of-the-art in energy efficiency for various types of node-clustering algorithms has been discussed. A number of suggested future research directions are described below.

Hotspot problem: Multi-hop data communication is a good technique for achieving energy efficiency; however, it has a "hotspot" problem which can force a number of nodes in a certain area in the network to die very early.

Coverage: Good coverage ensures that all the nodes in the network provide the data required by the sink.

Optimization of energy efficiency: The energy efficiency of a node-clustering algorithm can be optimized using various optimization techniques such as genetic algorithms and ant-colony algorithms.

REFERENCES

Abbasi, A.A. and M. Younis, 2007. A survey on clustering algorithms for wireless sensor networks. Comput. Commun., 30: 2826-2841.

Akyildiz, I.F. and M.C. Vuran, 2010. Wireless Sensor Networks. John Wiley and Sons, USA.

Ali, M.S., T. Dey and R. Biswas, 2008. Aleach: Advanced leach routing protocol for wireless microsensor networks. Proceedings of the International Conference on Electrical and Computer Engineering, December 20-22, 2008, Dhaka, pp: 909-914.

Changmin, D. and F. Hong, 2007. A distributed energy balance clustering protocol for heterogeneous wireless sensor networks. Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing, WiCom 2007. September 21-25, 2007, Shanghai, pp: 2469-2473.

Das, S.K. and H.M. Ammari, 2009. Routing and Data Dissemination. In: Wireless Sensor Network: A Networking Perspective, Zheng, J. and A. Jamalipour (Eds.). John Wiley and Sons, USA., pp: 67-139.

Du, J.L., L.J. Chen and L. Xie, 2008. Application platform of heterogeneous wireless sensor networks. Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing, October 12-14, 2008, Dalian, pp: 1-4.

Duarte-Melo, E.J. and M. Liu, 2002. Analysis of energy consumption and Lifetime of heterogeneous wireless sensor networks. IEEE Global Telecommun. Conf., 1: 21-25.

Elbhiri, B., R. Saadane and D. Aboutajdine, 2009. Stochastic distributed energy efficient clustering (SDEEC) for heterogeneous wireless sensor networks. ICGST-CNIR J., 9: 11-17.

Elbhiri, B., R. Saadane, S. El Fkihi and D. Aboutajdine, 2010. Developed distributed energy-efficient clustering (DDEEC) for heterogeneous wireless sensor networks. Proceedings of the 5th International Symposium on I/V Communications and Mobile Network (ISVC), September 30- October 2, 2010, Rabat, pp: 1-4.

Han, L., 2010. LEACH-HPR: An energy efficient routing algorithm for heterogeneous WSN Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS), October 29-31, 2010, Xiamen, pp: 507-511.

Handy, M.J., M. Haase and D. Timmermann, 2002. Low energy adaptive clustering hierarchy with deterministic cluster-head selection. Proceedings of the 4th International Workshop on Mobile and Wireless Communications Network, September 9-11, 2002 Stockholm, Sweden, pp: 368-372.

Hasnaoui, M.L., A.B. Hssane, A. Ezzati and S. Benalla, 2010. An oriented cluster formation for heterogenous wireless sensor networks. J. Comput., Vol. 2.

Heinzelman, W.B., A.P. Chandrakasan and H. Balakrishnan, 2002. An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wireless Commun., 1: 660-670.

Jiun-Jian, L., D. Chen-Yi and W. Yi-Jie, 2009. The steady clustering scheme for heterogeneous wireless sensor networks. Proceedings of the Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, July 7-9, 2009, Brisbane, pp: 336-341.

Katiyar, V., N. Chand and S. Soni, 2011. Improving lifetime of large-scale wireless sensor networks through heterogeneity. Proceedings of the International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT), March 23-24, 2011, India, pp: 1032-1036.

Kim, J.M., S.H. Park, Y.J. Han and T.M. Chung, 2008. CHEF: Cluster head election mechanism using fuzzy logic in wireless sensor networks. Proceedings of the 10th International Conference on Advanced Communication Technology, February 17-20, 2008, Gangwon-Do, pp: 654-659.

Kumar, D., T.C. Aseri and R.B. Patel, 2009. EEHC: Energy efficient hetergenous clustered scheme for wireless sensor networks. Comput. Commun., 32: 662-667.

Kumar, D., T.C. Aseri and R.B. Patel, 2010. A novel energy efficient multihop communicaiton protocol (EEMCP) for clustered heterogenous wireless sensor networks. J. Global Res. Comput. Sci. Vol. 1.

Li, X., D. Huang and J. Yang, 2007a. Energy efficient routing protocol based on residual energy and energy consumption rate for heterogeneous wireless sensor networks. Proceedings of the Control Conference, CCC 2007, July 26-31, 2007, Hunan, Chinese, pp: 587-590.

Li, X., D. Huang and Z. Sun, 2007b. A routing protocol for balancing energy consumption in heterogeneous wireless sensor networks. Mobile Ad Hoc Sensor Networks, 4864: 79-88.

Lindsey, S. and C. Raghavendra, 2002. PEGASIS: Power-efficient gathering in sensor information systems. Aerospace Conf. Proc., 3: 3-1125-3-1130.

Liu, T., J. Peng, J. Yang and C. Wang, 2011. Energy efficient prediction clustering algorithm for multilevel heterogeneous wireless sensor networks. Energy, 13: 71-75.

Mehrotra, U. and W.Y. Leong, 2009. Nseear: A energy adaptive routing protocol for heterogeneous wireless sensor networks. Proceedings of the 35th Annual Conference of IEEE Industrial Electronics, November 3-5, 2009, Porto, pp: 2647-2652.

Puccinelli, D. and M. Haenggi, 2005. Wireless sensor networks: Applications and challenges of ubiquitous sensing. IEEE Circuits Syst. Mag., 5: 19-31.

Qing, L., Q. Zhu and M. Wang, 2006. Design of a distributed energy efficient clustering algorithm for heterogenous wireless sensor networks. Comput. Commun., 29: 2230-2237.

Satapathy, S.S. and N. Sarma, 2006. TREEPSI: Tree based energy efficient protocol for sensor information. Proceedings of International Conference on Wireless and Optical Communications Networks, April 2006, Bangalore, pp: 4-4.

Smaragdakis, G., I. Matta and A. Bestavros, 2004. Sep: A stable election protocol for clustered heterogenous wireless sensor networks. Proceedings of the Internation Workshop on SANPA (2004), August, 2004, USA.

Sohraby, K., D. Minoli and T. Znati, 2007. Wireless Sensor Network: Technology, Protocols and Applications. 1st Edn., John Wiley and Son, USA.

Soni, S. and V. Katiyar, 2011. Prolonging the lifetime of wireless sensor networks using multi-level clustering and heterogeneity. Proceedings of the WSEAS International Conference on Software Engineering, Parallel and Distributed Systems, February 20-22, 2011, Cambridge, UK.

Soro, S. and W.B. Heinzelman, 2009. Cluster head election techniques for coverage preservation in wireless sensor networks. Ad Hoc Networks, 7: 955-972.

Tillapart, P., T. Thumthawatworn, P. Pakdeepinit, T. Yeophantong, S. Charoenvikrom and J. Daengdej, 2004. Method for cluster heads selection in wireless sensor networks. Proc. Aerosp. Conf., 6: 3615-3623.

Xuegong, Q., C. Yan and M. Fuchang, 2010. A routing algorithm based on the energy threshold's intra-cluster and inter-cluster multi-hop transmission. Proceedings of the International Conference on Computer Application and System Modeling, October 22-24, 2010, Taiyuan, pp: V10-364-V10-367.

Yarvis, M., N. Kushalnagar, H. Singh, A. Rangarajan, Y. Liu and S. Singh, 2005. Exploiting heterogeneity in sensor networks. IEEE Annu. Joint Conf. Comput. Commun. Soc., 2: 878-890.

Younis, O. and S. Fahmy, 2004. HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans. Mobile Comput., 3: 366-379.

Younis, O., M. Krunz and S. Ramasubramanian, 2006. Node clustering in wireless sensor networks: Recent developments and deployment challenges. IEEE Network, 20: 20-25.

Yu, L., N. Wang, W. Zhang and C. Zheng, 2007. Deploying a heterogeneous wireless sensor network. Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing, September 21-25, 2007, Shanghai, pp: 2588-2591.

Zhang, C., E. Hou and N. Ansari, 2009. Node Clustering. In: Wireless Sensor Networks: A Networking Perspective, Zheng, J. and A. Jamalipour (Eds.). John Wiley and Sons, USA., pp: 173-209.