Energy Based Clustering Self Organizing Map Protocol For extending Wireless Sensor Networks lifetime and coverage

Neda Enami*, Reza Askari Moghadam*
* *Engineering Department*
Payam Noor University of Tehran, Iran
Email:nedaenami@yahoo.com,askari@pnu.ac.ir

Abstract

Today, Cluster based routing protocols are well known schemes for extending Wireless Sensor Networks lifetime. However, there are several energy efficient cluster-based protocols in the literature; most of them use the topological neighborhood or adjacency as main parameter to form the clusters. This paper presents a new centralized adaptive Energy Based Clustering protocol through the application of Self organizing map neural networks (called EBC-S) which can cluster sensor nodes, based on multi parameters; energy level and coordinates of sensor nodes. We applied some maximum energy nodes as weights of SOM map units; so that the nodes with higher energy attract the nearest nodes with lower energy levels. Therefore, formed clusters may not necessarily contain adjacent nodes. The new algorithm enables us to form energy balanced clusters and equally distribute energy consumption. Simulation results and comparison with previous protocols(LEACH and LEA2C) prove that our new algorithm is able to extend the lifetime of the network, while it can insure more network coverage in it's lifetime through distributed death of nodes in network space.

Key Words: Energy Based Clustering, Self Organizing Map Neural Networks & Wireless Sensor Networks.

1. Introduction

The most important difference of Wireless Sensor Network (WSNs) with other wireless networks may be constraints of their resources, especially energy which usually arise from small size of sensor nodes and their batteries which is a prerequisite to WSNs main applications. The main and most important reason of WSNs creation was continuous monitoring of environments where are too hard or impossible for human to access or stay. So there is often low possibility to replace or recharge the dead nodes as well. The other important requirement is that we need a continuous monitoring so the lifetime and network coverage of these networks are our great concerns. Therefore, even extending the network lifetime without having enough coverage on whole network is often not desirable. As a result, as energy conservation is the main concern in WSNs, but also it should be gained with balanced distribution in whole network space. Balanced distribution of energy in whole network will lead to balanced death of nodes in all regions preventing from lacking network coverage in a rather large part of the network. Energy conservation should be gained by wisely management of energy sources. Several energy conservation schemes have been proposed in the literature while there is a comprehensive survey of energy conservation methods for WSNs and the taxonomy of all into three main approaches (duty-cycling, data reduction, and mobility based approaches) (Anastasi et al., 2009). Also these methods can be divided according to the layer of protocol stack with which they are involved such as several MAC protocols that have been proposed in the literature and survey studies on them as in (Demirkol at al., 2006; Langendoen, 2008)

Today, radio communications are the most energy consuming task of WSNs (Pottie and Kaiser, 2000; Raghunathan et al., 2002). So many research studies focused on energy efficient routing protocols to address this problem. Routing protocols can be divided based on different considerations like application, network structure or protocol operation but the most commonly used classification usually divide them into three general categories based on the underlying network structure: flat, hierarchical (cluster based) and location-based routings (Al-karaki et al., 2004). In flat networks, each node typically plays the same role and sensor nodes collaborate to perform the sensing task as SPIN (Heinzelman et al., 1999; Kulik et al., 2002), Direct Diffusion (Intanagonwiwat et al., 2000) etc. In location based routing, sensor nodes are addressed by means of their locations. The sensing area is divided into small virtual grids. All nodes in same virtual grid are equivalent for routing and only one node need to be active at a time. The most famous protocols from this category are GAF (Xu et al., 2001a) and GEAR (Xu et al., 2001b) routing protocols. Hierarchical or Cluster based routing protocols, as potentially the most energy efficient organization, have shown wide application in the past few years (Vlajic and Xia, 2006; Wei, 2007) and numerous clustering algorithms, have been proposed for energy conservation such as LEACH (Heinzelman et al., 2000, 2002), PEGASIS (Lindsey and Raghavendra, 2002), EECS (Ye et al., 2005), HEED(Younis and Fahmy, 2004), EEUC (Chengfa et al., 2005), LEACH-C (Heinzelman et al., 2002) and LEA2C (Dehni et al., 2005) etc. Hierarchical routing is mainly two-layer routing where one layer is used to select cluster heads and the other for routing (Al-Karaki and Kamal, 2004). In clustering protocols, geographically close nodes are organized into groups and each group is referred to as a cluster. Higher-energy nodes called *Cluster Heads* (CHs) play the coordination and communication tasks and other nodes in the clusters called *normal (simple) nodes* only do the sensing job and transmit their data packets to CHs. Because the data from adjacent sensor nodes usually have high correlation, CHs should also aggregate and/or fuse these received data packets to decrease the number of transmitted messages to Base Station (Wei et al., 2008).

In this paper we present a novel Energy Based Clustering protocol through using of Self organizing map neural networks (called EBC-S). Our work is closely related to LEACH-Centralized (Heinzelman et al., 2002) according to the Base Station cluster formation method it uses which requires global knowledge about all nodes energy and positions. EBC-S is also related to LEA2C (Dehni et al., 2005) protocol which is another SOM-based clustering protocol. LEA2C handled the NP-hard problem of optimal number of clusters by a two-phase method; SOM followed by Kmeans and it shows a considerable profit compared with another LEACH like protocol, called EECS (Ye et al., 2005). The difference of our proposed protocol with previous one is that it is able to adaptively cluster the nodes not only based on their topological closeness (coordinates) but also based on their energy levels in each set-up phase by using SOM capability on multi dimensional data classification. The formed clusters may not necessarily contain adjacent nodes anymore. As the result of forming clusters with near equal energy level, we better can balance the energy consumption in whole network during the data transmission phase and extend the lifetime of the network in the terms of first dead time and insures more network coverage during network life time. Simulation results show the profit of our protocol over LEACH and LEA2C.

2. LEACH Protocol

Low Energy Adaptive Clustering Hierarchy (LEACH) by Heinzelman et al. (2000) is the most famous clustering protocol which had been a basis for many further clustering protocols. The most important goal of LEACH is to have local Base Station (Cluster Heads) to reduce the energy cost of transmitting data from normal nodes to a distant Base Station. In LEACH, nodes organize themselves into local clusters with one node acting as cluster head. All non-cluster head nodes (normal nodes) transmit their data to the cluster heads. Cluster head nodes do some data aggregation and/or data fusion function on which should be transmitted to Base Station. Cluster head nodes are much more energy intensive than normal nodes. So choosing fix cluster heads, will end up in their early death. One solution can be random rotation of cluster head among nodes to balance the energy level of the network. The operation of LEACH is divided into rounds. Each round begins with a set-up (clustering) phase when clusters are organized, followed by a steady- state (transmission) phase when data packets are transferred from normal nodes to cluster heads. After data aggregation, cluster heads will transmit the messages to the Base Station. The election of cluster head is done with a probability function: each node selects a random number between 0 and 1 and if the number is less than *T(n)*, the node is elected as a cluster head for current round:

$$T(n) = \begin{cases} \dfrac{P}{1 - P\left(r \bmod \dfrac{1}{P}\right)} & \text{if} \quad n \in G \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

Where, P is the cluster head probability, r is the number of current round and G is the set of nodes that have not been cluster-heads in last $1/P$ round. The strength of LEACH is in its CH rotation mechanism and data aggregation. But one important problem with LEACH is that it offers no guarantee about placement and/or number of cluster head nodes in every round. Therefore using a centralized clustering algorithm would produce better results. LEACH-Centralized (LEACH-C) is a Base Station cluster formation algorithm. It uses the same steady state protocol as LEACH. During the steady state phase, each node sends information about its current position and energy level to BS. The assumption usually is that each node has a GPS receiver. The BS have to insure the evenly distribution of energy among nodes. So it determines a threshold for energy level and selects the nodes (with higher energy than this threshold) as possible cluster heads. The problem of determining the optimal number of cluster heads is an NP-Hard problem. LEACH-C makes use of Simulated Annealing (Murata and Ishibuchi, 1994) algorithm to address this problem. After determining the cluster heads of current round, BS sends a message containing cluster head ID for each node. If a node's cluster head ID matches its own ID, the node is a cluster head; otherwise it's a normal node and can go to sleep until data transmission phase. LEACH-C is more efficient than LEACH (LEACH-C delivers about 40% more data per unit energy than LEACH) because the BS has global knowledge of the location and energy level of all nodes in the network (Heinzelman et al., 2000). Also LEACH-C always insures the existence of K optimal number of cluster heads in every set-up phase while LEACH can not ensure that (Heinzelman et al., 2000, 2002).

3. SOM Based Routing Protocols

Today, Neural Networks can be applied as effective tools in all aspects of reducing energy consumption such as duty cycling, data driven and mobility based approaches in WSNs (Enami et al., 2010). Dimensionality reduction, obtained simply from the outputs of the neural-networks clustering algorithms, leads to lower communication costs and energy savings (Kulakov et al., 2005).

The Self-Organizing Map (SOM) is an unsupervised neural network structure consists of neurons organized on a regular low dimensional grid (Vesanto et al., 1999). Each neuron is presented by an n- dimensional weight vector where n is equal to the dimensions of input vectors. Weight vectors (or synapses) connect the input layer to output layer which is called *map* or *competitive layer*. The neurons connect to each other with a neighborhood relation as shown in figure 1. Every input vector activates a neuron in output layer (called winner neuron) based on its most similarity. The similarity is usually measured by Euclidian distance of two vectors.

$$D_j = \sum_{i=1}^{n} \left\| W_{i,j} - x_i \right\|^2 \qquad (2)$$

Where x_i is the i^{th} input vector, $W_{i,j}$ is the weight vector connecting input i to output neuron j and D_j is the sum of Euclidian distance between input sample x_i and it's connecting weight vector to j^{th} output neuron which is called a *map unit*.

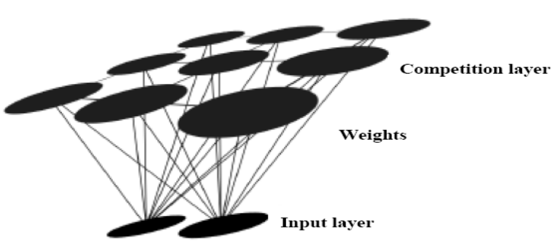


Figure 1. SOM topology structure (Yun et al., 2007)

The important difference of a SOM training algorithm with other vector quantization algorithms is that not only the best matching units (the winner neuron) but also its topological neighbors would be updated. Close observations in input space would activate two close units of the SOM. The learning phase continues until the stabilization of weight vectors.

$$W_{i,j}^{new} = W_{i,j}^{old} + h_{i,j}(x_{.i} - W_{i,j}^{old}) \qquad (3)$$

Where $x_{.i}$ is the input sample, Wi,j^{old} is the previous weight vector between input vector x_i and weight vector connected to output neuron j , $h_{i,j}$ is the neighborhood function and Wi,j^{new} is the updated weight vector between input neuron i and output neuron j.

There are different applications for SOM neural networks in WSNs routing protocols. These applications can be divided into three general groups: deciding optimal route, selection of cluster heads and clustering of nodes. The authors in (Aslam et al., 2010) used Kohonen SOM neural networks for clustering and their analysis to study unpredictable behaviors of network parameters and applications. Clustering of sensor nodes using Kohonen Self Organizing Map (KSOM) is computed for various numbers of nodes by taking different parameters of sensor node such as direction, position, number of hops, energy levels, sensitivity, latency, etc. Authors in (Shahbazi et al., 2008) proposed a new method for routing in WSNs in which each wireless node use a SOM neural network to decide about containing the data packet and participate in routing or dropping the packet. As soon as a packet arrived, its feature vector is extracted and this vector is sent to SOM of that node. If the node wins the competition against other nodes, it is allowed to send the packet and participate in routing. Otherwise it should drop the packet. SIR (Barbancho et al., 2007) is another QoS-driven SOM based routing protocol in which a SOM neural network is introduced in every node to manage the routes that data have to follow. Also Cordina and Debono (2008) proposed a new LEACH like routing protocol in which the election of Cluster Heads is done with SOM neural networks where SOM inputs are intended parameters for cluster heads. SOM cluster the nodes according to their cluster head qualities. However a minimum separation filter should be applied on SOM output then to ensure a minimum separation distance between selected CHs. Results show a 57 % profit of this protocol over LEACH (in terms of first dead time). Low Energy Adaptive Connectionist Clustering (LEA2C) by Dehni et al. (2005) is another LEACH-C (Heinzelman et al., 2000, 2002) like SOM-Based clustering protocol. The cluster formation is done by Base Station in the similar way said for LEACH-C. LEA2C uses a two phase clustering method, SOM followed by Kmeans. The inputs to SOM are the coordinates of sensor nodes in network space. LEA2C apply the connectionist learning by the minimization of the distance between the input samples (sensor nodes coordinates) and the map prototypes (referents) weighted by an especial neighborhood function. After set-up phase, the cluster heads of every cluster are selected according to one of the three criterions, max energy node, nearest node to BS and nearest node to gravity center of each cluster. Then the transmission phase starts and normal nodes send their packets to their CHs and on to the BS. In the case of using max energy factor for cluster head selection, the protocol would have a cluster head rotation process after every transmission phase. The transmission phase continues until the occurrence of first dead in the network. After that, the reclustering (set-up) phase will repeat. The simulation results show the profit of LEA2C over another LEACH-based protocol, called EECS (Ye et al., 2005) (In terms of 50 percent longer lifetime and insuring the network coverage during 90 percent of its total lifetime).

4. Proposed Algorithm (EBC-S)

In order to use the effectiveness of cluster-based routing algorithms in increasing of WSNs lifetime, we tried to present a new Energy Based Clustering Self organizing map (EBC-S). The motivation of creating EBC-S was inattention of previous clustering algorithms to energy level of the nodes as a key parameter to cluster formation of the networks. We tried to develop the classic idea for topological clustering and incorporate a topology-energy based clustering method in order to approach to our main goal in WSNs, extending life time of the network with enough network coverage. In our idea, energy based clustering can create clusters with equivalent energy levels. In this way, energy consumption would be better balanced in whole network.

4.1 Algorithm Assumptions

The proposed algorithm is more like LEACH-C and LEA2C protocols. Thus the assumption about BS cluster formation tasks and energy consumptions models of normal and cluster head nodes are the same as previous. The operation of the algorithm is divided into rounds in a similar way to LEACH-C. Each round begins with a cluster setup phase, in which cluster organization takes place, followed by a data transmission phase, throughout which data from the simple nodes is transferred to the cluster heads. Each cluster head aggregates/fuses the data received from other nodes within its cluster and relays the packet to the base station. In every cluster setup phase, Base Station has to cluster the nodes and assign appropriate roles to them. After determining the cluster heads of current round, BS sends a message containing cluster head ID for each node. If a node's cluster head ID matches its own ID, the node is a cluster head otherwise it is a normal node. BS also creates a Time Division Multiple Access (TDMA) table for each cluster and affects this table to CHs. Using TDMA, schedules the data transmission of sensor nodes and also allows sensor nodes to turn off their antennas after their time slot and save their energy. So the energy cost for cluster formation is just for BS and there are no control packets for sensor nodes. We assume that BS has no constraint about its energy resources. Also we assume that BS has total knowledge about the energy level and position of all nodes of the network (most probably by using GPS receiver in each node). The other important assumption of the protocol is random distribution of nodes in network space. The sensor nodes are homogenous, means they have the same processing and communication capabilities and the same amount of energy resources (at the beginning).

4.2. Cluster Setup phase

The protocol uses a two phase clustering method SOM followed by Kmeans algorithm which had been proposed in (Vesanto et al., 2000) with an exact comparison between the results of direct clustering of data and clustering of the prototype vectors of the SOM. We selected SOM for clustering because it is able to reduce dimensions of multi-dimensional input data and visualize the clusters into a map. In our application, dimensions of input data relates to the number of variables (parameters) that we need to consider for clustering. The reason for using SOM as preliminary phase is to make use of data pretreatment (dimension reduction, regrouping, visualization...) gained by SOM (Dehni et al., 2005). Therefore the data set is first clustered using the SOM, and then, the SOM is clustered by kmeans.

The variables that we want to consider as SOM input dataset is x and y coordination of every node in network space and the energy level of them. So we will have a *D* matrix with $n \times 3$ dimensions. Since we are applying two different type variables, first we have to normalize all values. We used a Min-Max normalization method (Visalakshi and Thangavel, 2009) in which min_a and max_a are the minimum and maximum values for attribute a. Min-max normalization, maps a value v in the range of (0, 1) by simply computing:

$$V' = \frac{v - \min_a}{(\max_a - \min_a)} \qquad (4)$$

So by means of above equation, our dataset matrix would be:

$$D = \begin{bmatrix} \frac{xd_1}{xd_{\max}} & \frac{yd_1}{yd_{\max}} & \frac{E_1}{E_{\max}} \\ . & . & . \\ . & . & . \\ . & . & . \\ \frac{xd_n}{xd_{\max}} & \frac{yd_1}{yd_{\max}} & \frac{E_n}{E_{\max}} \end{bmatrix} \tag{5}$$

Where D is the data sample matrix or input vectors of SOM, $XD=(xd_1 \ldots xd_n)$ are X coordinates, $YD=(yd_1 \ldots yd_n)$ are Y coordinates, $E=(E_1 \ldots E_n)$ are energy levels (remained energy) of all sensor nodes of the networks, xd_{max} is the maximum value for x coordinate of the network space, yd_{max} is the maximum value for Y coordinate of network space and E_{max} is the remain energy of maximum energy node of the network(at the beginning it is equal to $E_{initial}$).

In order to determine weight matrix, Base Station has to select m nodes with highest energy in the network. At the beginning, the nodes have equal energy level according to our assumptions. So we can partition the network space to m regions and select the nearest node to center of every region. However due to using two phase SOM-Kmeans method, we usually need to consider a rather large value for m, especially in large WSNs. In this case we can choose the m nodes randomly. We need three variables of these selected (high energy) nodes to apply them as weight vectors of our SOM: their x coordinate, their y coordinate and their energy level. Therefore our weight matrix would be:

$$W = \begin{bmatrix} \frac{xd_1}{xd_{\max}} & . \ . \ . & \frac{xd_m}{xd_{\max}} \\ \frac{yd_1}{yd_{\max}} & . \ . \ . & \frac{yd_m}{yd_{\max}} \\ 1-\frac{E_1}{E_{\max}} & . \ . \ . & 1-\frac{E_m}{E_{\max}} \end{bmatrix} \tag{6}$$

Where W is the weight matrix of SOM, $XD= (xd_1 \ldots xd_n)$ are x coordinates, $YD= (yd_1 \ldots yd_n)$ are y coordinates and $(1-E_1/E_{max} \ldots 1-E_n/E_{max})$ are consumed energy of m selected max energy sensor nodes. As you can see in equation (6), we have a $3 \times m$ weight matrix, so we would also have m map units (clusters). We made a change in third variable (remain energy) of selected nodes. In this way we want to move the nodes with less energy towards max energy nodes in order to form balanced clusters in the terms of energy level. So the SOM topology structure would be as figure.2:

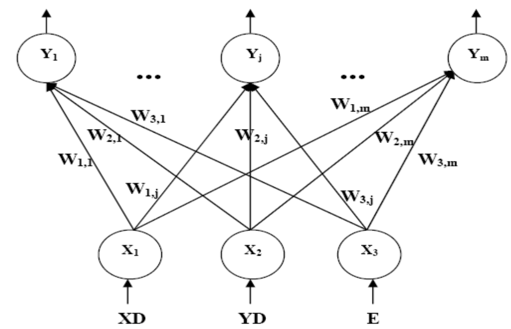


Figure 2. SOM topology structure in EBC-S protocol

In our application, learning is done by minimization of Euclidian distance between input samples and the map prototypes weighted by a neighborhood function $h_{i,j.}$ So the criterion to be minimized is defined by (Dehni et al., 2005):

Where N is the number of data samples, M is the

$$E_{SOM} = \frac{1}{N} \sum_{k=1}^{N} \sum_{j=1}^{M} h_{j,N(X^{(k)})} \left\| w_{.j} - x^{(k)} \right\|^2 \tag{7}$$

number of map units; $N(x^k)$ is the neuron having the closest referent to data sample $N(x^k)$ and h is the Gaussian neighborhood function defined by:

$$h_{i,j}(t) = \exp\left(- \frac{\left\| r_j - r_i \right\|^2}{2\sigma_t^2} \right) \tag{8}$$

Where $\left\| r_j - r_i \right\|^2$ the distance between map unit j and input sample i and σ_t is the neighborhood radius at time t which is defined by:

$$\sigma(t) = \sigma_0 \exp\left(- t/T \right) \tag{9}$$

Where t is the number of iteration, T is the maximum number of iteration or the training length.

The distance between X_k and weight vectors of all map neurons are computed. A neuron $N(X_k)$ which has the minimum distance with input sample X_k, would win the competition phase:

$$N(X_k) = \arg \min_{1 \prec j \le m} \left\| W_{.j} - X_k \right\|^2 \tag{10}$$

The neighborhood radius is a great value at the beginning and it will reduce with increasing of the time of the algorithm in every iteration. After competition phase, SOM should update the weight vector of the winner $N(X_k)$ and all its neighbors which placed at the neighborhood radius of ($R^{N(Xk)}$). If $W_{.j} \in R^{N(X_k)}$ then

$$W_{.j}(t+1) = W_{.j}(t) + \alpha(t) h_{j,N(x_k)}(t)_k (x(t) - W_{.j}(t)) \quad (11)$$

Else

$$W_{.j}(t+1) = W_{.j}(t) \quad (12)$$

Where $h_{j,N(X_k)}(t)$ is the neighborhood function at time t and $\alpha(t)$ is the linear learning factor at time t define by:

$$\alpha(t) = \alpha_0 (1 - t/T) \quad (13)$$

Where α_0 the initial learning rate, t is the number of iteration and T is the maximum training length. The learning phase repeats until stabilization (no more change) of weight vectors. SOM clusters n data samples into m map units (clusters). Now the SOM should be given to K-means algorithm as input.

K-means, partitions the data set into K subsets (clusters) such that all objects in a given dataset are closest to the same centroid. K-means randomly selects K of objects as cluster centroids. Then other objects are assigned to these clusters based on minimum Euclidean distance to their centroids. The mean of every cluster is recomputed as new centroids and the operation will continue until the cluster centers do not change anymore. The criterion to be minimized in K-means is defined by:

$$E_{K-means} = \frac{1}{C} \sum_{k=1}^{C} \sum_{x \in Q_k} \left\| x - C_k \right\|^2 \quad (14)$$

Where C is the number of clusters, Q_k is K^{th} cluster, C_k is the centroid of cluster Q_k.

The best value for K (optimal number of clusters) can be determined with an index. We selected Davies-Bouldin index. DB index actually compute the ratio of intra-clusters dispersion to inter-cluster distances by:

$$I_{DB} = \frac{1}{C} \sum_{k=1}^{C} \max_{l \neq k} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{ce}(Q_k, Q_l)} \right\} \quad (15)$$

With

$$S_c(Q_k) = \frac{\sum_i \left\| x_i - c_k \right\|^2}{|Q_k|} \quad (16)$$

$$d_{cl}(Q_k, Q_l) = \left\| c_k - c_l \right\|^2 \quad (17)$$

Where C is the number of clusters, S_c is the intra-cluster dispersion and d_{cl} is the distance between centroids of two clusters k and l.

Small values of DB index correspond to clusters which are compact, and whose centers well separated from each other. Consequently, the number of clusters that minimizes DB index is taken as the optimal number of clusters.

Now, Base station knows the optimal number of clusters and their member nodes. So the next step before going to transmission phase is selection of suitable cluster heads for each cluster and assigning appropriate roles to each node.

4.3 Cluster Head selection phase

Different parameters can be considered for selecting a CH in a formed cluster. In (Dehni et al., 2005a, 2005b) three criterions have been considered for CH selection:

1- The sensor having the maximum energy level
2- The nearest sensor to the BS
3- The nearest sensor to gravity center (centroid) of the cluster.

When we select the nearest node to BS in a cluster as CH, we insure to consume least energy to transmit the messages to BS. Also the nearest sensor to gravity center (centroid) of the cluster insure least average energy consumption for intra cluster communications while the reduction of CH overhead is not guaranteed. The results from LEA2C showed that the selecting the nodes with maximum energy level (first factor) as cluster head, gives the best results. This profit over two other criterions might be cause of having CH rotation. Because in the case of two other criterions (nearest sensor to BS or cluster cetroid) the selected CHs stay fixed during the transmission phase until next reclustering phase which may last for several rounds and it will cause the rapid depletion of that CHs, while applying these two criterions showed a longer lifetime (last dead) results.

After determining the cluster head nodes, BS assign appropriate roles to all nodes through the method mentioned for LEACH-C protocol before.

4.4 Transmission phase

After formation of clusters and selecting their related cluster heads, now it's time to send data packets sensed at normal nodes to their related cluster heads and after applying data aggregation functions to received packets by CHs, send messages on to the base station. The energy consumption of all nodes is computed. As in (Heinzelman et al., 2000; Dehni et al., 2005) the energy consumed for transmission of k bits of data over a d distance is computed by:

$$E_{Tx}(k,d) = E_{Tx}(l) + E_{Tx\_amp}(k,d) \tag{18}$$

$$E_{Tx}(k,d) = \begin{Bmatrix} k.E_{elec}(k,d) + k.\varepsilon_{friss}d^2 & if\ d < d_{crossover} \\ k.E_{elec}(k,d) + k.\varepsilon_{two-ray-amp}.d^4 & else \end{Bmatrix} \tag{19}$$

The energy consumption for receiving k bits of data is computed by:

$$E_{Rx}(k,d) = E_{Rx-ele}(k) = k.E_{elec} \tag{20}$$

Where E_{elec} is the energy of electronic transmission/reception, k is the size of message in bit, d is the distance between transmitter and receiver, E_{tx_amp} is amplification energy, ε_{friss} is amplification factor, $d_{\text{crossover}}$ is a threshold distance in which transmission factors change. Also energy consumption of data aggregation of CHs is:

$$E_{DA} = 5nJ / bit / msg \tag{21}$$

After every transmission phase, we count a new round and would have a cluster head rotation (in the case of using maximum energy criterion) as described in last section. But how often should we have a reclustering phase? Since our goal is to create clusters with equal energy levels, we should have a threshold for reclustering phase according to variation of energy level of the nodes. The best time for reclustering can be when a relative reduction occurs in energy level of nodes. So the energy level of m selected highest energy nodes are checked regularly. These nodes are cluster heads of last setup phase. The condition can be the depletion of a predefined percent of their energy level. This threshold energy level is defined experimentally. In this paper, 20 percent depletion of initial energy for first time reclustering phase and 5 percent depletion for next times are used. When the reclustering threshold is satisfied, BS sends a reclustering message to whole network. So, we can summarize the algorithm into following steps:

1- Initialization: random deployment of N homogeneous sensors in a given space and with the same energy level.

2- Cluster set-up phase:

2.1- clustering of WSN through SOM and K-mean clustering method by using sensor coordinates and remained energy as SOM inputs and selecting of m nodes with maximum energy level as the weights of SOM map units using Eqs. (5) to (17). The value for m can be different for every scene and experimental.

2.2- selection of cluster heads for every cluster with one of the 3 criteria mentioned (maximum energy sensor, nearest sensor to BS and nearest sensor to gravity center of the cluster).

2.3- assigning roles to every node (CH or Normal node) by BS.

3- Data Transmission Phase

3.1- Data transmission from normal nodes to CHs. Energy consumption of nodes is then computed using energy model and Eqs (18) to (20)

3.2- Data aggregation and or fusion of received packets and sending results to BS by CHs. energy consumption of CHs is then computed using Eqs (18) to (21)

3.3- CH selection if the CHs had been chosen according to maximum energy criteria

3.4- Repeat the steps 3-1 to 3-4 until the average energy level of m selected maximum energy nodes show a 20 percent reduction for first time reclustering and 5 percent for next times.

4- Repeat the steps 2 to 3 until all sensors in the network die.

5. Simulations and Results

MATLAB is used to simulate and compare the proposed algorithm (EBC-S) with previous works. To compare proposed protocol results with previous similar protocol (LEACH and LEA2C) we used the energy models as in Esq. (18) to (21) and scenes according to table1. SOM toolbox proposed by HUT researchers has been used to simulate proposed algorithm (Vesanto et al., 1999).

The other parameter that should be defined in simulation is the value for m (number of maximum energy level nodes that we use as SOM weights). This number is selected experimentally and its value is in relation with optimal number of clusters that we expected to have. In first scene (100 nodes), we assume *m=16 or 20* and in second scene (400 nodes) *m=50 or 80*.

Table1: Parameters of simulation

Parameter	Scene1	Scene2
N	100	400
Area	100 ×100	
Location of BS	(50,200)	
$d_{crossover}$	87m	
Initial Energy	0.5	
E_{elec}	50nJ/bit	
ε_{fs}	10pJ/bit/m^2	
ε_{mp}	0.0013pJ/bit/m^4	
EDA	5 nJ/bit/signal	
Packet size	4000 bits	

The EBC-S protocol performance was evaluated with three criterions for cluster head selection used by Dehni et al. (2005). The results show that selection of maximum energy node as CH, always give the best performance far enough from two other criterions (nearest sensor to BS or nearest sensor to GC). So the best performance of EBC-S (with CH maximum energy) has been compared with two other previous protocols; LEACH and LEA2C for two scenes with characteristics mentioned in table1. The comparison was done through using of three metrics: the number of round (time) when first node dies (First dead time), the number of round (time) when half of nodes die (Half dead time) and the number of round (time) when last node dies (Last dead time).the results are shown in table (2 ,3)

Table 2: comparison of algorithms results (first scene)

	Algorithm	First death	Half death	Last death
Number of nodes=100 (First scene)	LEACH	576	781	1857
	LEA2C(maximum energy)	626	738	977
	EBCS(maximum energy)	862	878	897
	EBCS(nearest to BS)	47	996	1206
	EBCS(nearest to GC	47	834	1558

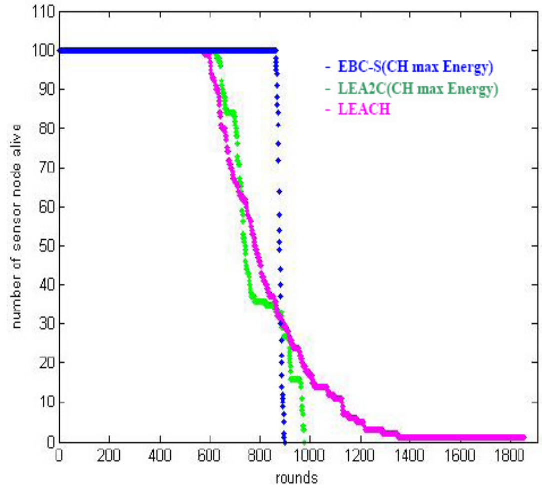
Table3. Comparison of algorithms results (second scene)

	Algorithm	First death	Half death	Last death
Number of nodes=400 (Second scene)	LEACH	713	958	2184
	LEA2C(maximum energy)	867	1045	1087
	EBCS(maximum energy)	959	999	1053
	EBCS(Nearest to BS)	18	1120	1421
	EBCS(Nearest to GC)	22	1057	1712

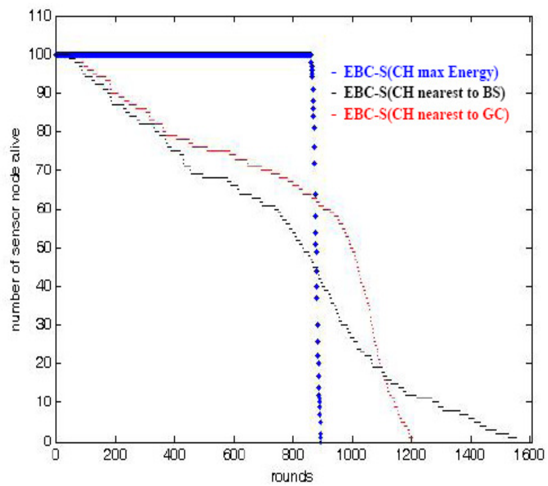
In figures (3.a, 4.a) you can see the advantages of the proposed protocol compared with others. The results on figures (3.a, 4.a) show that the proposed algorithm can insure total survival (network coverage) during 95% of network lifetime in first scene and 90% in second scene.

As shown in figure (3.a), the new algorithm can increase the lifetime of the network up to 50% over LEACH and 38 % over LEA2C protocols (for the first scene and with maximum energy CH criterion).Also results shown on figure (4a) prove that the new algorithm increase the lifetime of the network up to 27% over LEACH and 11% over LEA2C protocols (for the second scene and with maximum energy CH criterion).

In figures (3.b, 4.b) the performance of using two other CH selection criterions (nearest node to Gravity Center of the cluster and nearest node to Base Station) have been compared to maximum energy criterion. As you can see, the performances of two other criterions are very near to each other while they are too far from maximum energy criterion performance.



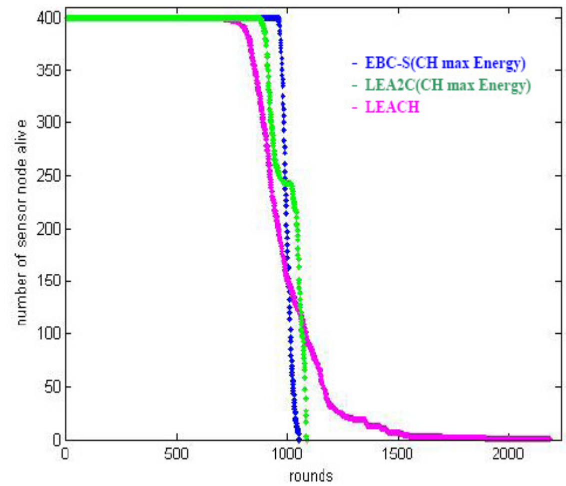
a



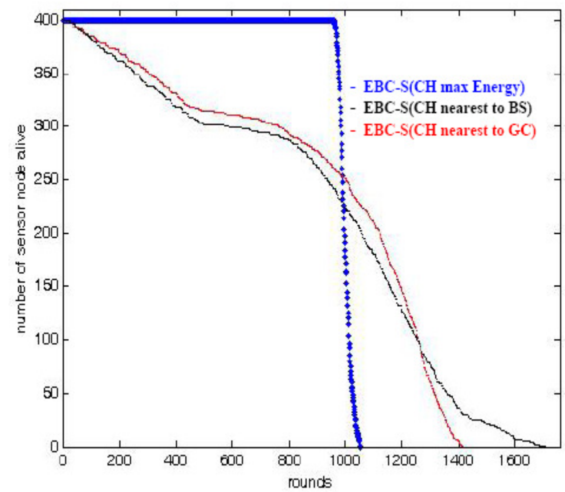
b

Figure 3. Number of alive nodes VS time (a) comparing in LEACH, LEA2C and EBC-S (proposed algorithm) (b) comparing in EBCS (proposed algorithm) with different CH criterions (First Scene)

In figure (5) you can see the cluster formation situation and dispersion of cluster nodes in LEACH and EBCS protocols. As it is shown in EBCS unlike LEACH (and LEA2C), the boundary of clusters is unlimited and each cluster does not necessarily contain adjacent nodes.



a



b

Figure 4. Number of alive nodes VS time (a) in LEA2C with different CH criterions and (b) in EBCS (proposed algorithm) with different CH criterions and in LEACH (Second scene)

Since loss of alive nodes in one region of the network will cause lack of network coverage (sensing) on that region, we applied another test on proposed algorithm. We tried to compare the network coverage of EBCS (proposed protocol) and LEACH with the same number of dead nodes (36 dead nodes) and between EBC-S and LEA2C protocols with the same number of dead nodes (50 dead nodes/half dead time) while all started with 100 sensor nodes. In order to define a reasonable metric, we used dividing the network space into virtual grids.

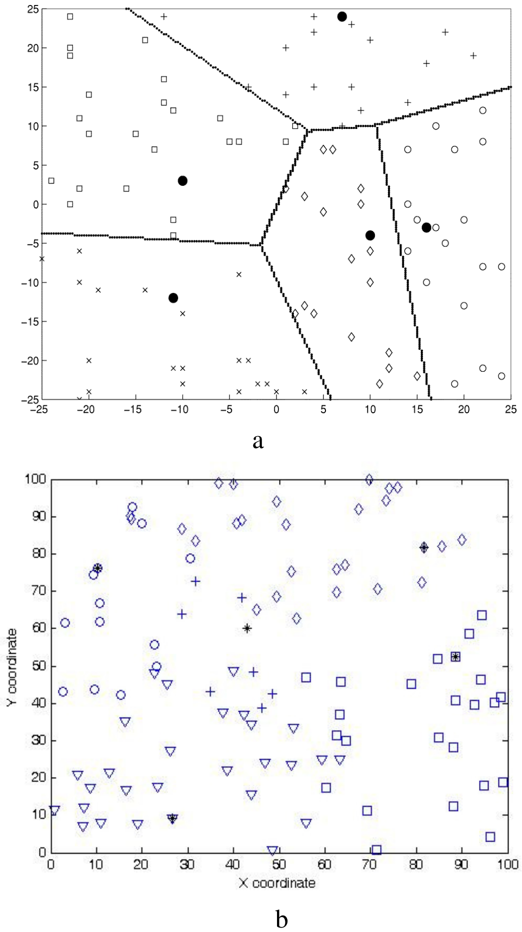


Figure 5. The cluster formation in (a) LEACH and (b) EBCS protocols. All nodes marked with a given symbol belong to the same cluster, and the cluster head nodes are marked with •in LEACH and with ✱ in EBCS.

As you can see in figure (6), the network space is divided into 25 sensing regions, each region at least contain one sensor. The region, in which there is at least one alive sensor, is assumed to be an active region otherwise it is a dead region. We computed the number (percent) of active regions in network space for each algorithm.

Figures (6.a, 6.b) are from LEACH algorithm (Heinzelman et al., 2000) and EBCS when 36 nodes are dead from 100 initial nodes and BS is placed at (x=0, y=-100). Figures (7.a, 7.b) are from LEA2C and EBC-S algorithm when 50 nodes from 100 initial nodes are dead while BS is placed at (x=50, y=200).

The comparison of active regions percentage show that LEACH has 84%network coverage while EBC-S still has 92% network coverage left.

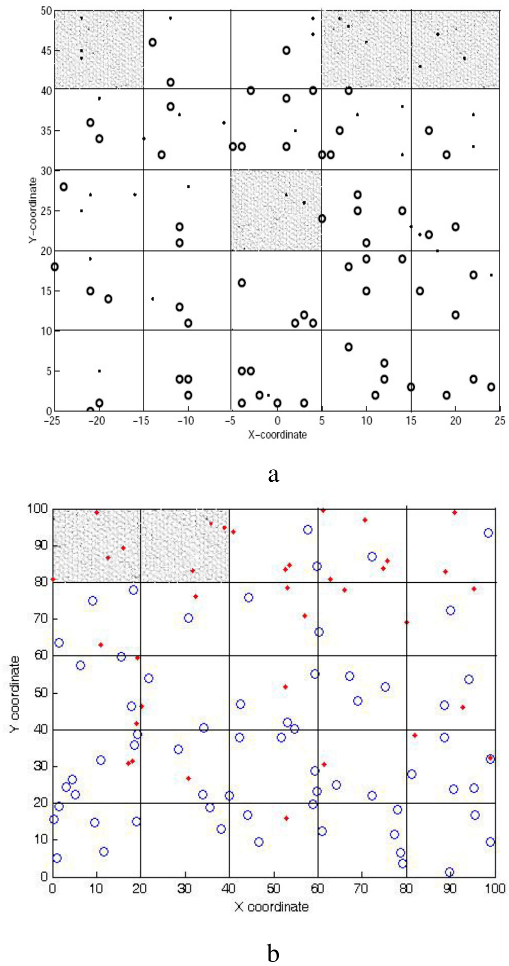
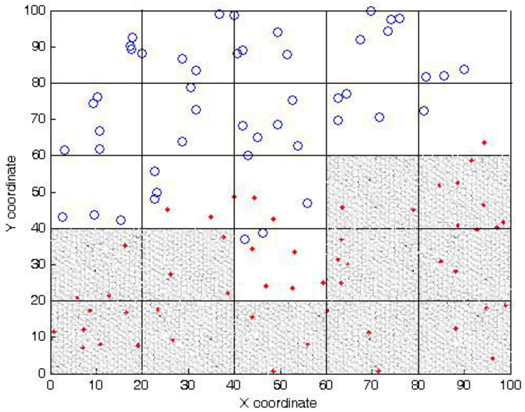
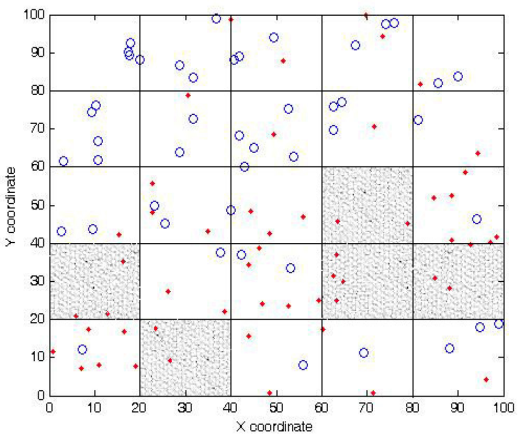


Figure 6.Network coverage: (a) LEACH, (b) EBC-S when 36 nodes are dead out of 100 nodes and where nodes with circles represent alive nodes and red dots represent dead nodes.

Moreover, the comparison of network coverage between LEA2C and EBCS in half dead time in figure.7 show that in EBC-S, nodes dies more randomly than in LEA2C. Also network coverage in LEA2C is 56% while there is still 80% network coverage left in EBC-S in half dead time.



a



b

Figure 7. Network coverage at half dead time: (a) LEA2C, (b) EBC-S where nodes with circles represent alive nodes and red dots represent dead nodes.

6. Conclusions

In this paper we proposed a new Energy Based Clustering protocol through SOM neural networks (called EBC-S) which applies energy levels and coordinates of nodes as clustering input parameters and uses some nodes with maximum energy levels as weight vectors of SOM map units. Nodes with maximum energy attract nearest nodes with lower energy in order to create energy balanced clusters. The clustering phase performs by a two phase SOM-Kmeans clustering method. The simulation results show 50% Profit of new algorithm over LEACH and 38% profit over LEA2C (in first scene) and 27% profit over LEACH and 11% profit over LEA2C (in second scene) in the terms of increasing first dead time while ensuring total coverage during 90% up to 95% of network life time in two scenes. Also the way of cluster formation in EBCS is different from other algorithms besides it shows 8% more network coverage over LEACH and 24% more network coverage over LEA2C in the same conditions. As future works, the following research areas would improve the protocol results:

- Combination of proposed algorithm with multi-hoping routing protocols.
- Applying other useful parameters for clustering
- Applying different structures for SOM and Kmeans algorithms
- Applying different criterions for Cluster Head selection of the protocol.
- Applying different neighborhood functions to optimize SOM clustering

Acknowledgment

Authors would like to thank Dr. Juha Vesanto (juuso) for his support about SomToolBox, Mario Cordina for his sincere suggestions about the simulation, Dr. Mojtaba Hoseini for his kindly feedbacks on this work and Dr. Wendi Heinzelman for her guidance about LEACH.

7. References

Al-karaki J.N, Kamal A.E. Routing Techniques in Wireless Sensor Networks: A Survey, IEEE Wireless Communication, 2004, p.6-28

Anastasi G, Conti M, Passarella A. Energy Conservation in Wireless Sensor Networks: a survey. In: Ad Hoc Networks, volume 7, Issue 3, Elsevier; 2009, p.537-568

Aslam N, Philips W, Robertson W, Siva Kumar SH. A multi-criterion optimization technique for energy efficient cluster formation in Wireless Sensor networks. In: Information Fusion, Elsevier; 2010

Barbancho J, Leon C, Molina F.J, Barbancho A. Using artificial intelligence in routing scheme for wireless networks. In: Computer Communications 30, Elsevier; 2007, pp. 2802-2811

Chengfa L, Mao Y, Guigai C. An Energy-Efficient Unequal Clustering Mechanism for Wireless Sensor Networks. In: Proc. Of IEEE MASS. 2005

Cordina M, Debono C.J. Increasing Wireless Sensor Network Lifetime through the Application of SOM neural networks. In: ISCCSP, IEEE, Malta, 2008, p. 467-471

Dehni L, Kief F, Bennani Y. Power Control and Clustering in Wireless Sensor Networks. In: Proceedings of Med-Hoc-Net 2005: Mediterranean Ad Hoc Networking Workshop, France.

Dehni L, Krief F, Bennani Y. Power Control and Clustering in Wireless Sensor Networks. In: Challenges in Ad Hoc Networking, vol 2005, p.31-40.

Dehni L, Krief F, Bennani Y. Power Control and Clustering in Wireless Sensor Networks. In: Proceedings of Med-Hoc-Net 2005, Mediterranean Ad Hoc Networking Workshop, France.

Demirkol I, Ersoy C, Alagoz F, MAC Protocols for Wireless Sensor Networks: a Survey. In: IEEE Communications Magazine, 2006

Enami N, Askari Moghadam R, Haghighat A. A Survey on Application of Neural Networks in Energy Conservation of Wireless Sensor Networks. In: Recent Trends in Wireless and Mobile Networks, WiMo 2010 Proceedings, Ankara, Turkey, 2010. p. 283–294.

Heinzelman W, Chandrakasan A, Balakrishnan H. Application-specific protocol architecture for wireless microsensor networks. In: IEEE Transactions on Wireless Communications, 2002, p. 660 - 670.

Heinzelman W, Chandrakasan A, Balakrishnan H. Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In: Proc. 33rd Hawaii Int'l. Conf.Sys. Sci, 2000

Heinzelman W, Kulik J, Balakrishnan H. Adaptive Protocols for Information Dissemination in Wireless Sensor Networks. In: Proc. 5th ACM/IEEE Mobicom, Seattle, WA, 1999, p. 174–85.

Heinzelman W. Application-Specific Protocol Architectures for Wireless Networks. In: PhD Thesis, Massachusetts Institute of Technology, 2000

Intanagonwiwat C, Govindan R, Estrin D. Directed Diffusion: a Scalable and Robust Communication Paradigm for Sensor Networks. In: Proc. ACM Mobi-Com 2000, Boston, MA, 2000, p. 56–67.

Kulakov A, Davcev D, Trajkovski G. Application of wavelet neural-networks in wireless sensor networks. In: Sixth International Conference on Software Engineering, Artificial Intelligence, 2005, pp. 262–267

Kulik J, Heinzelman W. R, Balakrishnan, H. Negotiation-Based Protocols for Disseminating Information in Wireless Sensor Networks. In: Wireless Networks, vol. 8, 2002, pp. 169–85.

Langendoen K, Medium Access Control in Wireless Sensor Networks. In: Book Chapter in "Medium Access Control in Wireless Networks, Volume II: Practice and Standards", Nova Science Publishers, 2008.

Lindsey S, Raghavendra C. PEGASIS: Power-Efficient Gathering in Sensor Information Systems. In: IEEE Aerospace Conf. Proc., vol. 3, 9–16, 2002, pp. 1125–30.

Murata T, Ishibuchi H. Performance evaluation of genetic algorithms for flowshop scheduling problems. In: Proc. 1st IEEE Conf. Evolutionary Computation, vol. 2, 2008, pp. 812–817.

Pottie G, Kaiser W, Wireless Integrated Network Sensors. In: Communication of ACM, Vol. 43, N. 5, pp. 51-58, May 2000.

Raghunathan V, Schurghers C, Park S, Srivastava M. Energy-aware Wireless Microsensor Networks. In: IEEE Signal Processing Magazine, March 2002, pp. 40-50.

Shahbazi H, Araghizadeh M.A, Dalvi M. Minimum Power Intelligent Routing In Wireless Sensors Networks Using Self Organizing Neural Networks. In: IEEE International Symposium on Telecommunications, 2008, p. 354–358

Vesanto J, Alhoniemi E. Clustering of Self Organizing Map. In: IEEE Transactions on Neural Networks, Vol. 11, No. 3, 2000, pp. 586-600.

Vesanto J, Himberg J, Alhoniemi E, Parhankangas J. Self-Organizing Map in MATLAB: The SOM toolbox. In: Proc of MATLAB DSP Conference, Finland, 1999, p.35-40

Visalakshi N. K, Thangavel K. Impact of Normalization in Distributed K-Means Clustering. In: International Journal of Soft Computing 4, Medwell journal, 2009, p. 168-172

Vlajic N, Xia D. Wireless Sensor Networks: To Cluster or Not to Cluster? In: Proc. of WoWMoM, 2006

Wei D, Kaplan SH, Chan H.A. Energy Efficient Clustering Algorithms for Wireless Sensor Networks. In: IEEE Communication Society workshop proceeding, 2008

Wei D. Clustering Algorithms for Sensor Networks and Mobile Ad Hoc Networks to Improve Energy

Efficiency, PhD thesis, 2007, University of Cape Town.

Xu Y, Heidemann J, Estrin D. Geography informed Energy Conservation for Ad-hoc Routing. In: Proc. 7th Annual ACM/IEEE Int'l. Conf. Mobile Comp and Net, 2001, pp. 70–84.

Ye M, Li C. F, Chen G. H, Wu J. EECS: An Energy Efficient Clustering Scheme in Wireless Sensor Networks. In: Proceedings of IEEE Int'l Performance Computing and Communications Conference (IPCCC), 2005, p. 535-540.

Younis O, Fahmy S. HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks. In: IEEE Transactions on Mobile Computing, vol. 3, no. 4, 2004, p. 660-669

Yu Y, Estrin D, Govindan R. Geographical and Energy-Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks. In: UCLA Comp. Sci. Dept. tech. rep., UCLA-CSD TR-010023(2001)

Yun S.U, Youk Y.S, Kim S.H, Study on Applicability of Self-Organizing Maps to Sensor Network. In: International Symposium on Advanced Intelligent Systems, Sokcho, Korea, 2007