Volume 2, No. 3, May-June 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

ISSN No. 0976-5697

Cluster Based Data Gathering in Wireless Sensor Networks

Narendran.M*
Department of Information Technology,
Anna University of Technology, Coimbatore
Tamilnadu, India
naren0812@hotmail.com

Shobanalakshmi.R
Department of Information Technology,
Anna University of Technology, Coimbatore
Tamilnadu, India
Shobanalakshmi.it@gmail.com

Maivizhi.C
Department of Information Technology,
Anna University of Technology, Coimbatore
Tamilnadu, India
caspermayil@gmail.com.

Abstract: Use of wireless sensor network for real life applications have rapidly increased. Since sensor nodes operate on batteries, energy efficient mechanisms for gathering sensor data are indispensable to prolong the lifetime of a sensor network as long as possible. A sensor node consumes energy: observing its surroundings, transmitting data, and receiving data. Energy consumption is one of the major concerns in wireless sensor networks since it impacts the network lifetime. Clustering is one of the key techniques which reduce energy consumption is wireless sensor networks. In this paper we proposed an information similarity concepts with optimal clustering algorithm for partition the adjacent nodes which will sense similar target into one cluster, thus cluster heads are elected. Our method can effectively reduce redundant data transmission and the whole energy consumed in the network. Our simulation results show that our clustering mechanism reduced the energy consumption and prolong the life time of network.

Keywords: Wireless sensor network, clustering algorithm, Information similarity. Data aggregation

I. INTRODUCTION

A wireless sensor network is includes of many sensor nodes, each having a microprocessor, sensor, actuator and wired/wireless transceiver. Wireless sensor networks can be used in range of applications such as environment monitoring, military tactics, inventory inspection and medical care. It is also a key tool in ubiquitous computing, next generation mobile communications, ITS (Intelligent Transportation System) and home networking. However when battery-powered wireless sensor nodes are placed in a specific field, it is critical to replace their batteries or supply additional energy. Furthermore, if one sensor node consumes fully its energy, part of the network may disconnect. Thus it is very important to reduce the energy wastage of a sensor node as much as possible. Wireless Sensor Network is collection of a large number of sensor nodes that are densely deployed either inside the object being sensed or very close to it. Since sensor nodes are severely controlled by the amount of battery power available, advanced techniques that improve energy efficiency to prolong the network lifetime are highly required. Data collection is a common but critical operation in many applications of WSNs, where data aggregation and hierarchical mechanism are commonly used techniques. Data aggregation can reduce the data redundancy and reduce the communication load. Clustering mechanisms are especially effective in increasing network scalability and lifetime, which have been extensively exploited. LEACH which is the first clustering protocol has encouraged the design of many other protocols. Which follow a similar concept; it uses randomization to rotate the cluster heads to evenly distribute the energy load among the sensors in the network. LEACH-C is a variant of LEACH that chooses cluster heads informally but the base station makes sure that only nodes with "enough" energy are participating in the cluster head selection. LEACH's clustering algorithm adopts that sensor nodes are homogeneous and equal. In certainty, however, their battery capacities are different, and the amount of energy consumed in gathering data also differs among cluster heads, depending on the

number of cluster members and their situations in the region. Energy usage also differs among cluster members due to the distance to a cluster head. Some sensor nodes might also be arranged later for denser observations. Subsequently, residual energy is different among sensor nodes. In addition, the finest percentage of cluster-heads has to be determined in advance, considering the topology of a sensor network. Therefore, LEACH cannot adjust to such changes in sensor networks as the addition, removal, and transfer of sensor nodes, although the percentage of cluster-heads considerably affects the efficiency of data collecting method.

Finally, for structured clusters to cover an entire sensor network, each cluster-head must broadcast its own advertisement to all the other nodes, another inefficient use of energy. To tackle the problem, in they proposed two different of LEACH. LEACH-C (LEACH-centralized) is a centralized protocol, which takes into account the residual energy in choosing sensor node for cluster-heads. The other is a distributed but minimum efficient implementation of LEACH.

Based on their similarity, a cluster is created, merged, or deleted. By repeating meetings, an appropriate set cluster is eventually formed so that same set objects are accommodated in the same cluster.

The paper consists of the following sections. Sections II review the related works. Section III introduces a clustering algorithm for energy-efficient data collecting in sensor networks. Results of simulation experiments are given in Section IV. Finally, Section V conclusion is summarized.

II. RELATED WORK

Many clustering algorithms have been proposed for wireless sensor networks in recent years. We review some of the most related papers:

In LEACH, each node has a certain chances of becoming a cluster head per round, and the task of being a cluster head is rotated between nodes. In the data transmission phase, each cluster head sends an aggregated data to the base station by single hop. The amount of energy used in a radio transmission is proportional to the square of the range of the radio signal propagation [2]. In LEACH, which is well-known and widely referred to, a per-determined percentage of sensor nodes become cluster heads which broadcast their candidacy to the rest of the sensor nodes. Hearing announcements chooses the nearest cluster-head and register itself as a cluster member. Taking inspiration from such genetic systems, much research has been done in the fields of data clustering and graph partitioning [3-5].HEED is one of the operational data-collecting protocols without location support. In HEED, cluster-head selection is based on the candidate's residual energy and a secondary parameter, such as proximity to its neighbors [12]. Hein Zelman, et.al. Introduced a hierarchy clustering algorithm for sensor networks, called Low Energy Adaptive Clustering Hierarchy (LEACH) [2]. LEACH contracts with the heterogeneous energy condition are the node with higher energy should have larger probability of becoming the cluster head [7]. Each sensor node must have an estimate of the total energy of all nodes in the network to compute the probability of becoming a cluster head only by its local information, so the scalability of this scheme will be influenced. Although LEACH is able to increase the network lifespan, but it has two main weaknesses:

- It's possible no or lots of Cluster Heads are nominated.
- Cluster Heads are not selected in a distributed manner; it's possible that too many Cluster Heads are located in a exact area.

In Power-efficient collecting in sensor information system (PEGASIS) is designed to increase the performance of LEACH. Nodes in PEGASIS are formed by chain [7]. Each node communicates only with the adjoining neighbor, and takes turns to be the CH and transmit information to the sink. PEGASIS is most efficient energy consumption than LEACH. With the use of data combination and aggregation techniques, while minimizing the total energy per round [8], if power consumption per node can be balanced as well, a near optimal data collecting and routing scheme can be achieved in terms of network lifetime. Each sensor has an on panel radio that can be used to send the collected data to interested parties. Such scientific development has encouraged practitioners to visualize aggregating the limited capabilities of the individual sensors in a large scale network that can operate unattended [9].

A. *Problem Identification*

A wireless sensor network consisting of a large number of small sensors with low-power transceivers can be an effective tool for collecting data in a variety of surroundings. The data collected by each sensor is communicated through the Network to a single processing center that uses all reported data to determine characteristics of the situation or detect an event. The communication or message passing process must be designed to preserve the limited energy resources of the sensor. The nodes in a sensor networks frequently need to organize themselves into cluster.

Clustering can be used to thin out parts of the networks where an extreme number of nodes may be present. A simplified long range communication network can be setup using only Cluster-Heads and gateways- all other nodes communicate via their cluster-head. In other words, nodes in the areas have a great probability of sensing the same position data. To reduce the energy spent of remote communication, we need partition the adjacent nodes which will sense similar data into one cluster, so as to eliminate the redundant data transmission through similar data collection and help the end-users to search information relevant to a certain kind of object. The cluster head nodes in each sub region act as the union points within the network. During each data collecting cycle, the sensor nodes sensed data to the resultant cluster head node which performs data aggregation. Then the cluster head nonstop transmits the aggregated data to a Base station. The sensor nodes have simple functionality, Later they perform sensing and moderately short range communication.

Sensor nodes can conclude their absolute or relative geometrical positions by using Global Positioning System (GPS) or a position detection system. However, the number of applications where these conventions hold is limited. In our next work, we consider to adopt our method to sensor networks where some of these expectations are not valid. In this paper, we propose a novel clustering algorithm Optimal Clustering algorithm which fully consider the characteristic of target data in monitor area Fig: 1. In Optimal Clustering algorithm, we adopt the strategy of forming cluster first and selecting cluster-head afterward. We optimize the cluster forming by using optimal clustering algorithm to partition the adjacent nodes which will sense similar target into one cluster.

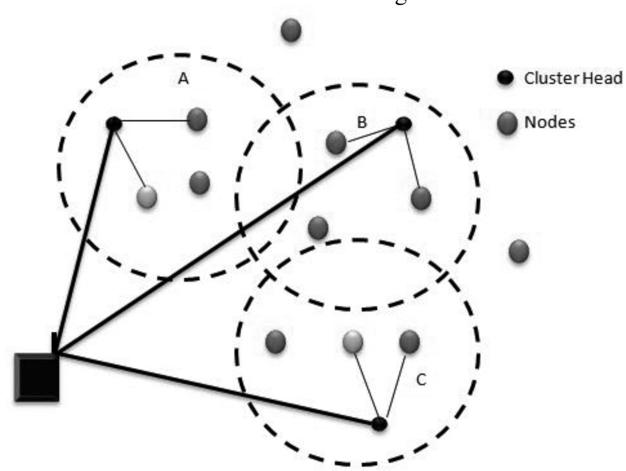

Figure 1: Cluster target in the Monitoring area

Since improving the rate of data aggregation in clusters, our approach can efficiently reduce redundant data transmission and the whole energy consumed in the network.

B. Network Model

Assume n sensor nodes are casually and uniformly distributed over the sensing field R and the sensor network has the following properties:

1. This network is a static efficiently deployed network. It means n sensor nodes are compactly deployed in a two dimensional geographic space, developing a network and those nodes do not move after deployment.
2. All nodes would be approximately time coordinated on the order of seconds.
3. There is one base station, which is deployed at (50, 50) position.
4. Nodes are location-aware, i.e. not prepared with GPS-capable antennae.
5. There are three types of nodes normal, progress and super nodes. Advance and super nodes are well-found with more battery energy than normal node.
6. These nodes are uniformly spread over the region R and they are not mobile.

III. OCA DETAILS

In this paper, we propose and evaluate an optimal clustering algorithm (OCA) for periodical data collecting applications in WSNs. Sensor nodes break due to a loss of battery power, move from one place to another, and are deployed later. Sensor nodes can determine their entire or relative geometrical positions by using Global Positioning System (GPS) or a position detection system [12-14]. However, the number of applications where these expectations hold is limited. In our next work, we consider to accept our method to sensor networks where some of these expectations are not valid.

We use the energy consumption model as in whose model a sensor node consumes Eelec (nJ/bit) in transmitter or receiver integrated circuit and εamp (pJ/bit/m2) in transmitter amplifier [15]. A sensor node spends energy ETx (k, d) or ERx (k) in transmitting or receiving a k-bit message to or from distance d, given by the following equations:

$$\mathbf{ETx\ (k, d) = \varepsilon amp \times k \times d^2 + Eelec \times k} \qquad (1)$$

$$\mathbf{ERx\ (k) = Eelec \times k.} \qquad (2)$$

A sensor node also consumes Efuse (nJ/bit/message) in collecting multiple sensor data into one.

The basic idea of OCA is using clustering algorithm to partition the adjacent nodes which will sense similar target into one cluster, then reducing the amount of redundant data transmission through similar data aggregation. Since the nodes in our optimal clusters will sense related information to a high degree, OCA not only improves the accuracy of information received at the BS but also moderates the remote communication cost of cluster-heads, thus prolongs the network lifetime.

A. Base station collect the information of registered nodes

The operation of Optimal clustering algorithm is divided into number of rounds. Each round starts with a set-up phase when the clusters are organized and cluster head are selected, followed by a steady-state phase when data are transferred from the nodes to the cluster head and on to the BS. The steady-state phase of Optimal clustering algorithm is equal to that of LEACH.

Assumed that each node in the network retains sensing data of the last N times and the set of monitored target in the system is t_j {t1, t2… tn}, where tn is the monitored target n. For any node we can get $Pt_j = Nt_b/N$ from N times monitor results, where NO_b is the number of times t_b occurred. Thus according to the set of monitored target T_b, we gain the probabilistic set NodeB {Pt1 , Pt2… Ptn} for all targets observed by nodes B.

Assumed that the probabilistic set of each target observed by two adjacent nodes A and B is Node A {Pt1 , Pt2… Ptn} and Node B {Pt1 , Pt2… Ptn}, then the explanation of information similarity for two nodes is:

$$\mathbf{S_{ab}=1-1/n \sum | P_{atk} - P_{btk} |} \qquad (3)$$

As you can see, information similarity is used to measure the probability of detecting the same target for two adjacent nodes. When the occurrence of targets have area characteristic, the higher probability means the neighbors would have an increasing possibility of obtaining similar data at one point.

$$\mathbf{S=2/\ k\ (k\text{-}1) \sum S_{ab}} \qquad (4)$$

B. Searching IS based clusters

We can avoid the maximum redundant data transmission in the network through the high rate of data collection in clusters. The above process is a problem of complicate information similarity method. The information similarity a concept is fully based on the number of nodes extant in the network and number of sensor's sense the same type data. Based on the data range and sensed data the clustering process done by base station. A cycle of data collecting, called a round, consists of four phases: (i) cluster-head election, (ii) cluster formation, (iii) registration, and (iv) data transmission. The behavior of sensor nodes in the third and fourth phases is identical to the cluster set-up phase through the data communication phase.

The sensor voting is that the importance of a sensor should be replicated from all its neighbors instead of being done from its local properties alone [16]. Each sensor calculates this correctness that one of its neighbors becomes its cluster head and casts a vote for that neighbor.

Sensors collect votes from their neighbors and calculate the total votes received. The more votes a sensor collects, the more importance it obtains in the whole network. During the clustering phase, sensors participate with one another in terms of the total votes each has expected. Because sensors are location-unaware, topology and residual energy become the two primary issues in electing the cluster heads. A sensor uses the following rules to calculate the poll for each of its neighbors:

R1) The sum of the polls a node gives to all its neighbors is 1.0.

R2) The neighbor whose ratio of residual energy to distance from the node is greater should gain more vote than the neighbor whose proportion of residual energy to distance is less.

C. Cluster Formation Phase

Out of various planning proposed for energy efficient data collecting, clustering is shown to be scalable and energy

.efficient architecture for wireless sensor networks, in the presence of large number of nodes and many of its advantages are listed in [17].After the cluster heads have been nominated, the next step in the clustering phase in a round is the cluster formation step. As discussed earlier, each cluster head can handle a certain number of communication networks. This constrains the number of sensor nodes that can communicate with the cluster head, thus existing in a cluster. The balanced clustering formation can overcome this shortcoming by having an upper/lower bound on the size of each cluster.

Cluster formation, hence, is one of the critical problems in sensor network applications and can drastically affect the network's communication energy dissipation. Clustering is performed by allocating each sensor node to a specific master node. All communication to (from) each sensor node is passed out through its corresponding master node. Observably one would like to have each sensor to communicate with the closest master node to conserve its energy; however master nodes can usually handle a specific number of communication channels. Therefore there are a maximum number of sensors that each major node can handle. This does not allow each sensor to communicate to its closest main node, as that master node might have already reached its service capacity.

D. Cluster Head Selection

Now most energy-efficient clustering algorithms proposed use probability based approach of clustering and weight based method of cluster-head selection (such as node density, residual energy and communication cost of intra-cluster, and so on), which lay emphasis on the even distribution of energy consumed across the network. They describe the data aggregation in cluster-heads with an inherent assumption, i.e. all the data received at cluster-heads are similar, so that cluster heads could collection them into one signal sent to the remote base station.

LEACH's stochastic cluster-head selection is disposed to producing unbalanced energy level reserves in nodes and thus increase the total energy dissipated in network. To ensure an even energy load sharing over the whole network, additional parameters including the residual energy level of candidates relative to the network and their hotness value should be considered to improve the process of cluster-head selection. In our method, after finding several optimal clusters based on information similarity, now's problem is to solve the strategy of cluster-head selection. Since the energy distribution in the network is unstable, we select cluster-heads by synthetically considering two essential parameters of candidate including the energy reserved and the total sum of normal distances between all the intra-cluster nodes and the candidate. Our strategy efforts not only to balance the total energy dissipation of the sensors, but also to minimize the amount of energy for the non-cluster head nodes to transmit their data to the cluster head.

IV. SIMULATION RESULTS

In this section, the performance of the recommended method will be evaluated. Here, we compare clustering in LEACH algorithm with a recommended method. We evaluate the performance of optimal clustering algorithm using Network simulator with the 50-node network in a play field of size 100m x 100m. The base station is located at point (50,200) and the initial energy per node is the random number of [2J, 5J]. For simplicity, we assume the probability of signal smash and interference in the wireless channel is ignorable.

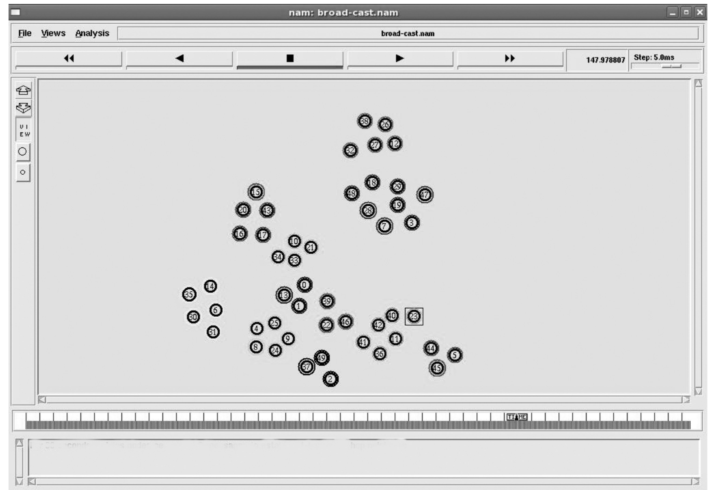

Figure 2: Sensor nodes Deployment

We compare the performance of optimal clustering algorithm with three important clustering algorithms. Fig.2 shows the relationship between the number of node death and the network runtime (rounds).

Table I. SIMULATION PARAMETERS

Parameters	Value
Data Packet Size	1000 bytes
Packet header size	50 bytes
Round (T)	10 TDMA
$\mathbf{E_{elec}}$	100nJ/bit
$\mathbf{E_{fs}}$	10pJ/bit/m^2
$\mathbf{E_{mp}}$	0.0013pJ/bit/m^4
$\mathbf{E_{fusion}}$	5nJ/bit/signal
Threshold	75m

We see that, in addition to LEACH, the curve of the other three algorithms is moderate, which determines their methods of cluster head selection enable the balanced energy dissipation among the sensor nodes. The other three algorithms, however, by using strategy of probabilistic based clustering, are disposed to to partition the adjacent nodes with similar sensed data into several clusters, thus increase the energy consumption of wasteful message transmission. Impact on network lifetime when the network size with the same density differs. We observe that optimal clustering algorithm keeps the good behavior as the network grows larger, which implies the effect of energy savings from elimination of redundant data transmission gets further important.

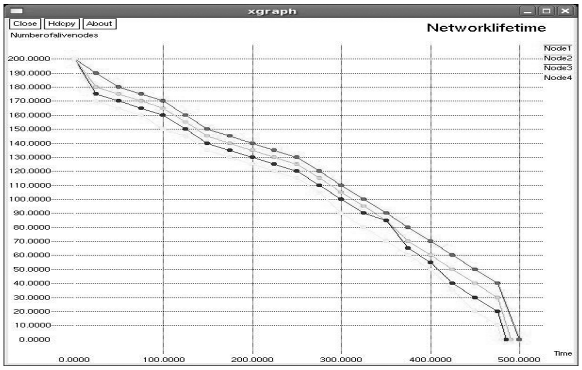

Figure 3: Network Lifetime

We have generated a sample sensor network with 50 sensor nodes. These sensors are randomly spread on a small region. Similarly, we have randomly located some Cluster head nodes in this region .We have applied our algorithm to form *k* clusters, such that the total square of the distances between sensors in a cluster and the matching cluster head node, over all clusters is minimized. The number of cluster head nodes depends on the capacity constraint of each of them. It is observable to see that Optimal Clustering algorithm has the highest rate of data aggregation, which ensures the maximum reduction of the total amount of data received at the BS.

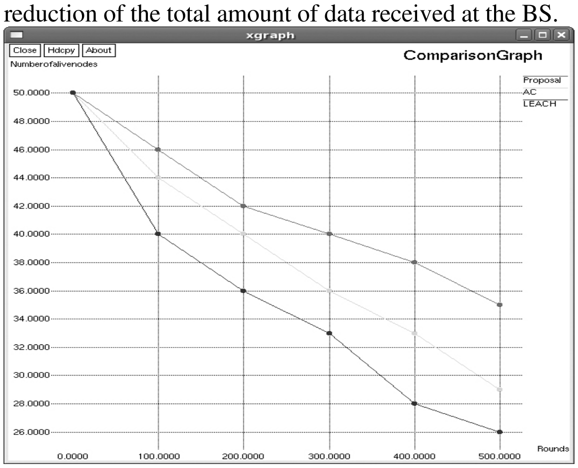

Figure 4: Comparison Graph

V. CONCULSION

In WSNs, there usually occur several kinds of monitored target. The regionalism feature of target occurrence demands the design of clustering algorithms consider the degree of information similarity in clusters and realize energy efficiency from the point of the reduction of redundant data transmission in the whole network. In this paper, we propose Optimal clustering algorithm, which efficient improve the rate of data collection in the network, thus greatly reduction the energy consumed of cluster-head for remote transmission. Moreover, Optimal clustering algorithm select cluster-heads by unnaturally considering the residual energy and the average distances of intra-cluster transmission, which reduced the whole energy rate of clusters. Simulation results show that, compared with probability bases clustering algorithm, Optimal clustering algorithm has the better performance of prolonging the network lifespan under the special environment of applications.

VI. ACKNOWLEDGMENT

The authors would like to thank the reviewers for their detailed comments.

VII. REFERENCES

[1] Junpei Kamimura, Naoki Wakamiya, Masayuki Murata," Energy-Efficient Clustering Method for Data Gathering in Sensor Networks " January 2004.

[2] W.R.Heinzelman, A. Chandrakasan, and H.Balakrishnan,"Energy-Efficient communication protocol for wireless micro sensor networks", in proceedings of HICSS-33, pp.3005-3014, January 2000.

[3] N. Labroche, N. Monmarch´e, and G. Venturini, "A new clustering algorithm based on the chemical recognition system of ants," in Proceedings of ECAI 2002, pp. 345–349, July 2002.

[4] A. E. Langham and P. W. Grant, "Using competing ant colonies to solve k-way partitioning problems with foraging and raiding strategies," in Proceedings of the 5th European Conference on Artificial Life, September 1999.

[5] I.Akyilidiz et al.," Wireless sensor networks: a survey, "Computer Networks, Vol 38, pp.393-422, March 2002

[6] Ying Liang Hang Li "An Energy-Efficient Clustering Algorithm for Wireless Sensor Sensor Network" . March 2006

[7] S. Lindsey and C. Raghavendra, "Pegasis: Power-efficient gathering in sensor information syetems," IEEE Trans. on Parallel and Distributed Systems, vol. 13, no. 9, pp. 924-932, 2002.

[8] K.Ramanan and E.Baburaj "Data gathering algorithms for Wireless sensor networks: a survey" International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC) Vol.1, No.4, December 2010.

[9] Ameer Ahmed Abbasi, Mohamed Younis "A survey on clustering algorithms for wireless sensor networks" Computer Communications 30 (2007) 2826–2841. June 2007.

[10] J.Liu,C.Lin, "Power-efficiency clustering method with power-limit constraint for sensor networks performance," Computing and Communications conference, 2003.Conference Proceedings of the 2003 IEEE International, Vol.9,No.2,pp.129-136, 2003

[11] O.Younis and S.Fahmy,"HEED: A hybrid, energy-efficient, distributed clustering approach for adhoc sensor networks,"IEEE Trans.Mobile Comput., Vol.3.4, pp.366-379, Oct-Dec.2004

[12] A. Savvides, C. Han, and M. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in Proceedings of Mobicom 2001, pp. 166–179, July 2001.

[13] A. Nasipuri and K. Li, "A directionality based location discovery scheme for wireless sensor networks," in Proceedings of the first ACM international workshop on Wireless Sensor Networks and Applications, pp. 105–111, September 2002.

[14] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in Proceedings of the 9th annual

international conference on Mobile computing and networking, pp. 81–95, September 2003.

[15] S. Lindsey, C. Raghavendra, and K. Sivalingam, “Data gathering in sensor networks using the energy*delay metric,” in IPDPS Workshop on Issues in Wireless Networks and Mobile Computing, pp. 924–935, April 2001.

[16] Jalil Jabari Lotf, Mehdi Nozad Bonab, Siavash Khorsandi, “A Novel Cluster-based Routing Protocol with Extending Lifetime for Wireless Sensor Networks”, IEEE, 2008.

[17] O. Younis, M. Krunz and A. Ramasubramanian, "Node Clustering in Wireless Sensor Networks: Recent Developments and Deployment Challenges", IEEE Network, May-June-2006, pp. 20-25.