Research Journal of Applied Sciences, Engineering and Technology 5(3): 975-981, 2013
DOI:10.19026/rjaset.5.5049
ISSN: 2040-7459; E-ISSN: 2040-7467

Submitted: June 23, 2012 Accepted: July 28, 2012 Published: January 21, 2013

Research Article

A Clustering Algorithm for Wireless Sensor Network Management

Wenbo Zhang and Yongxin Feng
School of Information Science and Engineering, Shenyang Ligong University, Shenyang, 110159, China

Abstract: In typical clustering algorithms for wireless sensor network, the remained energy of the elected cluster heads is not considered. Moreover, the service failure ratio that the cluster heads provided for nodes in clusters is not considered either. So a new clustering algorithm-MSCBRE (Maximum Succeed Communication Based on Remainder Energy) is proposed for wireless sensor network management. In this algorithm the service failure ratio is set to the ration of the remained energy to the initial energy of nodes. Based on this the elected cluster heads can provide multiple coverage to key nodes to ensure reliable communication. The experiment results show that the new algorithm is correct and effective.

Keywords: Clustering, MSCBRE, network management, wireless sensor networks

INTRODUCTION

Recent advancements in integrated circuits have fostered the emergence of a new generation of tiny, inexpensive low power sensors (Kurosawa *et al*., 2004). A sensor network is a set of nodes in which a battery, a sensoring and a wireless communication device are embedded (Agah *et al*., 2006). Sensor networks are a special case of ad-hoc networks with objects generally densely deployed either very close or inside a studied phenomenon (Younis and Fahmy, 2004; Shen *et al*. 2008). Sensor nodes are deployed over hostile or remote environments to monitor a target area. Therefore, their unrepeatable batteries imply energy to be the most important system resource. These objects are expected to work and collaborate as long as possible in order to send their collected data to one or more sink stations. These sinks, also called monitoring stations, are considered to have unlimited battery and aim to collect information from sensor nodes in multi-hop manner. The lifetime of the network is the time during which the surface coverage is maintained (Zhang *et al*., 2006). A point of the target surface is said to be covered if it is in the sensoring range of an active sensor which can report to a sink. It means that the sensor network can accomplish its surveillance task while the set of connected components contain monitoring stations covers the target area. To extend the network lifespan, some nodes are placed into sleeping mode to save their energy (Yao and Gehrke, 2002). The issue consists in these nodes deciding themselves whether to turn off or not so that the whole area remains to be covered and the subset of active nodes could be connected (Heinzelman *et al*., 2002).

Due to their economic and computational feasibility, a network of hundreds and thousands of sensors has the potential for numerous applications in both military and civil applications such as combat field surveillance, security and disaster management. These sensing devices are capable to monitor a wide variety of ambient conditions such as: temperature, pressure, motion etc. The sheer number of these devices and their ad-hoc deployment in the area of interest brings numerous challenges in networking and management of these systems (Mhatre *et al*., 2004). Sensors are typically disposable and expected to last until their energy drains. Therefore, energy is a very scarce resource for such sensor systems and has to be managed wisely in order to extend the life of the sensors for the duration of a particular mission (Jiang *et al*., 2009).

Wireless sensor network clustering algorithm is recognized as a valid method of self-organization. Currently, many protocols applications in the wireless sensor network are depended on the logic network architecture of sub-clusters, so an important research content of wireless sensor network is putting the wireless sensor networks into clustering with a reasonable method (Sankarasubramaniam *et al*., 2003).

LEACH algorithm is very typical for the clustering algorithm, which makes nodes in the network cluster according to a certain rule and selects the head of the cluster to take on the role of data control center, makes data in cluster integrate at home then forward to the SINK nodes, thus reducing the amount of data

Corresponding Author: Yongxin Feng, School of Information Science and Engineering, Shenyang Ligong University, Shenyang 110159, China

transmission and the number of transponders, the network will reduce energy consumption (Eduardo *et al*., 2009). But the election of cluster head is random and cannot guarantee the rationality and it requires cluster head node for long-distance transmission and makes the data sent directly to SINK, which in this large-scale networks is not easily achieved. Although HEED algorithm is different from LEACH algorithm on the selection criteria of cluster head and the mechanism of competition of the cluster head, clustering speed has a certain improvement, taking into account the cluster of communications expenses after cluster, the remaining energy of the nodes as a parameter introduce to the algorithms, making the election cluster better suited to play the task of transmitting data, making the network topology more reasonable, so that the entire network of power consumption will be more evenly, but the algorithm did not consider wireless sensor networks in the link failure or services failure, the cluster head node cannot provide services to the node within the cluster (Yah *et al*., 2008). So designing an algorithm, which not only considers the cluster head uniform distribution and balances link but also provides the best service at the circumstance of services failure, is very valuable (Giuseppe *et al*., 2009).

In this study, based on the MSCBRE principle, we put forward a new clustering algorithm. Under the circumstance of provided failure rate, the nodes demands and the optimal location of the cluster heads are computed. The experiment results show that the distribution of the cluster heads is more uniform. With the same consumption between this algorithm put forward in this study and the LEACH algorithm, the new algorithm can send more data. So it is easy to see that the new clustering algorithm is better than the traditional algorithm such as the LEACH.

OVERVIEW OF MEXCLP MODEL

MCLP model: The Traditional Maximal Covering Location Problem (MCLP) is positioning M facilities in N candidate nodes. The M facilities provide service for N demand nodes as far as possible. That is to say, these M facilities cover N nodes to maximize the demand which makes the demand quantity can be maximized. MCLP model uses the locations of N candidate nodes, demands, facilities services radius D and facilities number M as input parameters. Through calculating the maximum resolving the best positions of M facilities can be got. The following is the formulation description of the model of M facilities in N nodes:

$$\text{Max} \sum_{k=1}^{N} h_k y_k \tag{1}$$

St.

$$y_k - \sum_{i=1}^{N} r_{ki} X_i \le 0 \qquad k = 1, 2 \cdots, N \tag{2}$$

$$\sum_{i=1}^{N} X_i \le M \tag{3}$$

$$\left. \begin{array}{ll} X_i = 0,1, & i = 1,\cdots,N \\ y_k = 0,1, & i = 1,\cdots,N \end{array} \right\} \tag{4}$$

Among them h_k is equal to the demands produced by Node k:

$$y_k = \begin{cases} 0 & \text{node k isn't covered by facilities} \\ 1 & \text{node k is covered by facilities} \end{cases}$$

$$X_i = \begin{cases} 0 & \text{facilities aren't in the nodes} \\ 1 & \text{facilities are in the nodes} \end{cases}$$

$$r_{ki} = \begin{cases} 0 & d_{\text{ki}} > D(\text{The facilities at node i covere node k}) \\ 1 & d_{\text{ki}} \le D(\text{The facilities at node i don't covere node k}) \end{cases}$$

here, D refers to the effective distance covered service facilities, namely in this range nodes are covered.

MEXCLP model: The MCLP assumed that all facilities are perfectly reliable and are able to serve demands at all times. But this assumption is not practical. Daskin modeled the Maximum Expected Covering Location Problem (MEXCLP) considering the unavailability of servers to serve all demands at all times. In MEXCLP, Daskin associates each facility with a probability p of being inoperative. The model assumes that the probabilities of the facilities not working are independent of each other and are same for all facilities.

The objective of MEXCLP is to maximize the expected demand covered by locating a given number of facilities. In MEXCLP the number of facilities working follows a binomial distribution and the probability of a node k being covered is given by:

- 1-Prob (node k not being covered)
- 1-Prob (m facilities are not working)
- $1\text{-}p^m$

where, m is the number of facilities covering node k.

Let $H_{k,m}$ be the random variable denoting the number of demands at node k covered by a working facility, given that m facilities are capable of covering node k, hence we have:

$$H_{k,m} = \begin{cases} h_k & \text{with probability } 1 - \text{p}^{\text{m}} \\ 0 & \text{with probability } \text{p}^{\text{m}} \end{cases}$$

and E ($H_{k,m}$) = h_k (1 - P^m), ∀ k, m

If the number of facilities covering node k increase from m-1to m, the corresponding increase in the expected coverage of node k is given by when the number of covering node k, change from m-1 to m, the expected increase for coverage of Node k may be expressed as:

$$\Delta E(H_{k,m}) = E(H_{k,m}) - E(H_{k,m-1})$$
$$= h_k p^{m-1}(1-p) \qquad \forall \ m = 1,2,\ldots N$$

We now define the variables and their corresponding indices, utilized in the MEXCLP formulation as follows:

i = Index for potential facility locations
k = Index for demand nodes
N = Number of demand nodes
M = Number of facilities to be located
D = The distance beyond which a demand node is considered "uncobered"
D_{ik} = The distance between potential facility location I and demand node *k*
h_k = Demand of node k
p = probability of a facility failure (0<*p*<1):

$$r_{ik} = \begin{cases} 1, \text{ if } D_{ik} < D \\ 0 \text{ otherwise} \end{cases}$$

The decision variables of the problem are:
x_i = number of facilities placed at location *i*

$$y_{jk} = \begin{cases} 1 \text{ If demand node k is covered} \\ \quad \text{by at least j facilities} \\ 0 \text{ otherwise} \end{cases}$$

The MEXCLP can be formulated as follows:
Max:

$$\sum_{k=1}^{N} \sum_{j=1}^{M} (1-p)\, p^{j-1} h_k y_{jk} = \sum_{k=1}^{N} \sum_{j=1}^{M} w_j h_k y_{jk}$$

Subject to:

$$\sum_{j=1}^{M} y_{jk} - \sum_{i=1}^{N} r_{ik} X_i \le 0 \qquad \forall k = 1,\ldots N$$

$$\sum_{i=1}^{N} x_i \le M$$

Among them:
x_i □ Z+ $\forall i = 1,\ldots N$

y_{jk} $\{0,1\}$ $\forall j = 1,\ldots M, \quad k = 1,\ldots, N$

The objective function maximizes the total expected coverage. The inner summation in the objective function represents the number of demands that are covered by at least *j* facilities in which the term (1-*p*) pj^{-1} represents the weight associated with the number of demands covered by at least *j* facilities for any demand node *k*.

THE PROPOSITION OF MSCBRE CLUSTERS ALGORITHM

Although MEXCLP model takes into account the case of service failure, it does not take into account the problem of limited energy in the network of the wireless sensor nodes and it cannot be directly used for wireless sensor networks. In this study, MEXCLP is improved and a new clustering algorithm (Maximum Succeed Communication Based on Remainder Energy, MSCBRE) is proposed.

In MEXCLP, supposed the probability p is known and its value is same to all service facilities. However, in the wireless sensor network, the surplus energy of sensor nodes is not similar; here the probability of the service failure of the nodes has been revised. The excess energy parameter and the influence factor have been joined in the service failure rate. The failure probability P_j of the node j has been revised as following:

$$p_j = p - \alpha p + \alpha\,(1 - Er_j \,/\, Ei_j) \qquad (5)$$

In MSCBRE, supposed that the situation if each node works is mutually independent. Therefore by the theory of probability, the probability that the node *k* can be covered and served by facilities is expressed as following:

$$\text{P\{K is the probability of work facilities covered \}}$$
$$= 1 - \text{P\{ the probability of M a facilities are not working \}}$$
$$= 1 - p^m$$

However, taking into account the service failure rate of each facility node is not the same, the probability that *m* facilities cannot succeed to provide the service cannot use the expression ρ^m, but the expression $\prod_{j=1}^{m} \rho_j$. If we use the variable Y_{j_k} indicated whether the node *k* is covered by facilities *j*, when the node *k* is covered by the facility *j*, its value is 1; otherwise its value is $1/p_j$. Then we can put the expression $\prod_{j=1}^{m} \rho_j$ as $\prod_{j=1}^{M} \rho_j \, Y_{j_k}$. This is because when the node k is not covered by the facility *j*, Y_{j_k} = $1/p_j$, p_j Y_{j_k} = p_j * $1/p_j$ = 1.

Obviously when the node k is not covered by the facility j, p_j Y_{jk} has no effect on the value of $\prod_{j=1}^{m} \rho_j$., so in $\prod_{j=1}^{m} \rho_j$. Adding M-m facilities in which Node k is not covered by does not have the influence on the result of $\prod_{j=1}^{m} \rho_j$. Therefore the following expression is tenable:

$$p^m = \prod_{j=1}^{m} p_j = \prod_{j=1}^{M} p_j y_{jk}$$

Therefore, the probability of the node k demand success and is serviced by work facilities are expressed as:

$$\begin{aligned} & P\{\text{ probability of woking facilities services }\} \\ & = 1 - p^m \\ & = 1 - \prod_{j=1}^{M} p_j y_{jk} \end{aligned} \tag{6}$$

It supposes that node k works the facility cover place the demand quantity is used the expression H_{km}, it can be expressed as follows:

$$H_{k,m} = \begin{cases} h_k & \text{probability } 1 - p^m \\ 0 & \text{probability } p^m \end{cases}$$

And $E(H_{k,m}) = h_k (1 - P^m)$, $\forall$ k, m

By the formula (6) shows:

$$E(H_{k,m}) = h_k (1 - p^m) = h_k (1 - \prod_{j=1}^{M} p_j y_{jk})$$

Therefore, in the network all nodes succeed the service expected value be possible to express as follows:

$$\sum_{k=1}^{N} E(H_{k,m}) = \sum_{k=1}^{N} h_k (1 - p^m) = \sum_{k=1}^{N} h_k (1 - \prod_{j=1}^{M} p_j y_{jk})$$

Then the formula (5) can get MSCBRE model is as follows:

$$\text{Max.} \sum_{k=1}^{N} (1 - \prod_{j=1}^{M} p_j y_{jk}) h_k \tag{7}$$

$$\text{St. } p_j = p - \alpha p + \alpha (1 - Er_j / Ei_j) \tag{8}$$

Among them:

$$y_{jk} = \begin{cases} 1 & \text{If the node k is covered by j a facility} \\ \dfrac{1}{p_j} & \text{If the node k isn't covered by j a facility} \end{cases}$$

where,

P : The cluster head node failure rate of the service that is the failure rate of the network

α : The adjustment factor, it regulates energy in the influence degree clusters algorithm.

Ei_j : Initial energy of the node j

Er_j : Residual energy of node j

M : The number of Cluster-heads nodes

N : The number of nodes

MSCBRE CLUSTERING ALGORITHM SIMULATION

In this study, clustering algorithm for MSCBRE cluster head selection is simulated. The input parameters of MSCBRE clustering algorithm include the location of the node, the node's residual energy and node communication radius and so on. In this study simulation's experiment scene data is as follows: Nodes randomly deployed in the area 20×30 m; The node correspondence radius is 8; The adjustment factor takes 0.5; In the network nodal point number is 40; cluster head node number is 5; In the wireless sensor network's service failure rate is 0.9; Each node's initial energy is 100; In the network each node's position, the demand number and excess energy as shown in Table 1.

Carrying on the simulation to the MSCBRE algorithm, the cluster head selected in this simulation node is: 1, 2, 3, 39, 40. Their distribution relationship is shown in Fig. 1.

From the chart, it can be seen that the cluster head nodes of the MSCBRE algorithm is distributed evener and cluster head is in the place where there are a large number of nodes. The cluster head nodes of the MSCBRE algorithm are the nodes that have relatively much surplus energy. In the figure above left bottom dotted portion's 5 nodes respectively are 2, 7, 17, 26 and 20. Their location and residual energy are 2[(6, 2), 92], 7[(9, 5), 21], 17[(13, 2), 92], 20[(9, 7), 65], 26[(0, 5), 77]. By the above data it can be seen that the residual energy of node 2 and the node 17 is the same, but because node correspondence radius supposition is 8, the node 17 cannot cover the node 26 correspondences, the node 2 can actually cover the node 17, so the node 2 is more suitable than node 17 as the cluster head. It can be seen although MSCBRE is a centralized clusters algorithm, the cluster heads formed through it which is more suitable for wireless sensor networks.

Table 1: Clusters simulation data tables

Serial number	Location	Demand number	Excess energy	Serial number	Location	Demand number	Excess energy
1	(8, 27)	24	17	21	(24, 30)	9	94
2	(6, 2)	8	92	22	(1, 16)	1	71
3	(29, 28)	24	77	23	(4, 37)	25	65
4	(36, 28)	16	66	24	(2, 36)	9	93
5	(32, 16)	24	14	25	(25, 10)	28	17
6	(4, 23)	1	25	26	(0, 5)	28	77
7	(9, 5)	23	21	27	(6, 33)	20	3
8	(13, 24)	18	12	28	(32, 28)	11	14
9	(33, 34)	26	17	29	(39, 33)	14	49
10	(34, 36)	8	74	30	(27, 19)	17	49
11	(33, 21)	6	15	31	(39, 29)	12	96
12	(36, 10)	7	76	32	(29, 28)	8	2
13	(22, 39)	27	47	33	(25, 12)	1	78
14	(20, 39)	3	34	34	(31, 3)	19	20
15	(25, 10)	29	78	35	(9, 21)	14	20
16	(10, 24)	22	10	36	(37, 5)	8	40
17	(13, 2)	8	92	37	(39, 7)	19	3
18	(32, 8)	10	67	38	(4, 18)	14	23
19	(8, 20)	19	45	39	(34, 3)	26	6
20	(9, 7)	1	65	40	(31, 21)	9	89

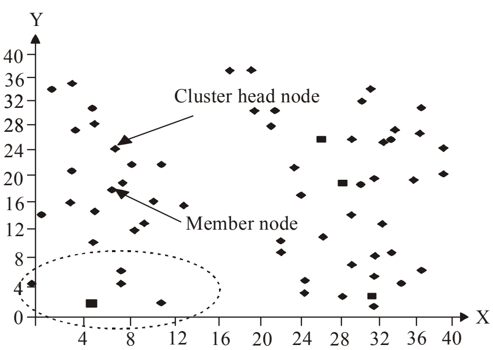

Fig. 1: Cluster head choice result diagram

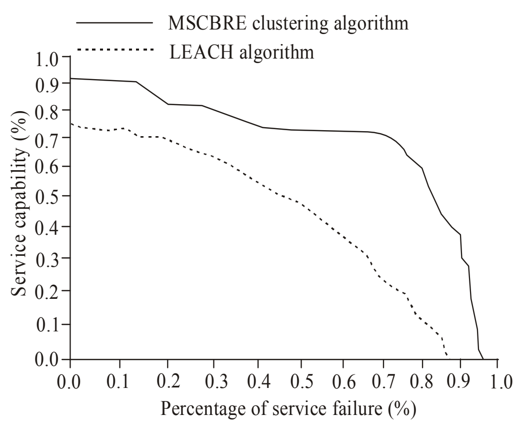

Fig. 2: Capability curve of the algorithms

Furthermore using network simulation tool NS-2 to test algorithm capability of MSCBRE and LEACH, 100 nodes are randomly distributed within range 100×100 m. Those two clustering algorithm are used separately to generate clusters, supposed that in each cluster any two member nodes can communicate each other.

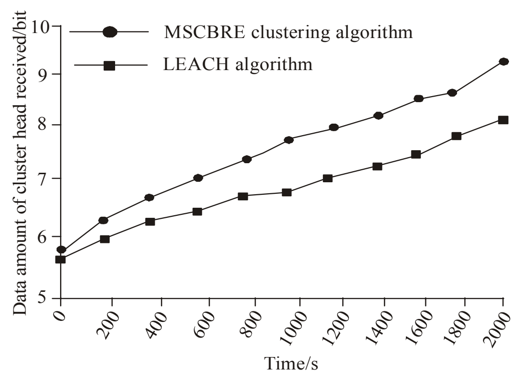

Fig. 3: Data amount received by sink node

Figure 2 are capability curve of LEACH algorithm and MSCBRE algorithm. We can see that, when the service failure rate of the cluster head is increased, the service capability of LEACH algorithms and MSCBRE algorithm are all reduced. But LEACH algorithm capability decline very significant with the service failure rate increase. In the case of service failure rate reaching 80%, the cluster head nodes are almost impossible to provide normal services. At the circumstance of MSCBRE failure rate of 78%, the cluster head nodes can still provide service; it is because the design of the algorithm gives full consideration to the possibility of failure service. Therefore, the cluster head nodes can provide service at the circumstance of service failure rate still higher.

Figure 3 shows the data amounts received by the sink node with time changing. Comparing MSCBRE algorithm with LEACH algorithm, it can be seen that the amount of data received by the sink in MSCBRE is more than the amount of data in LEACH.

The relationship between the number of cluster head polling rounds and the number of the remainder

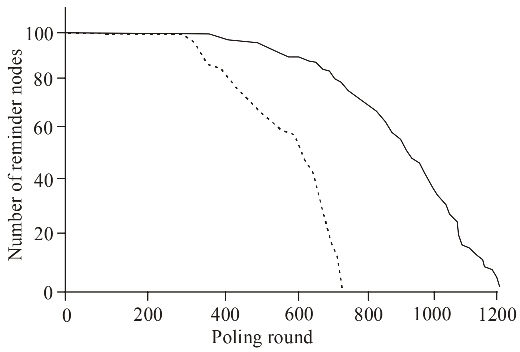

Fig. 4: Relationship between the polling rounds and remainder nodes

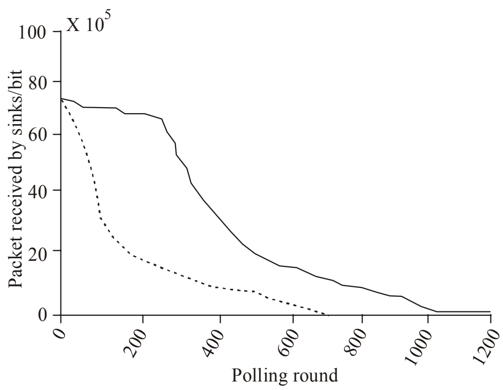

Fig. 5: Relationship between the amount of data received by sink and the polling round

nodes is shown in Fig. 4. From this figure it can be seen that all the nodes are dead when cluster head polling round exceeding 778 used LEACH in this simulation. However, using MSCBRE, the remainder nodes can survive 1200 round. This is because that in MSCBRE the remainder energy of nodes is significant to selectingthe cluster heads. The energy consumption of nodes in wireless sensor network is more even. That cannot consume overmuch energy to lead the sensor nodes accelerative to die.

From Fig. 5 it is obvious that in the original condition the difference of the amount of data between MSCBRE and LEACH is tiny because the service (data forwarding) failure rate is low. With the improvement of the simulation the nodes energy decline gradually. So the service failure rate increased. In LEACH the amount of data received by the sink decline obviously. However in MSCBRE though the amount of data received reduce, the network performance is still better than in LEACH. This is because that during selecting the cluster heads in MSCBRE, the most important condition is the cluster heads can provides better service to member nodes. Even if a cluster head serve unsuccessfully, the sensor data also can be sent to the sink because of multiple coverage.

CONCLUSION

Based on the MSCBRE principle a new clustering algorithm is put forward in this study. Under the circumstance of provided failure rate, the nodes demands and the optimal location of the cluster heads are computed. The experiment results show that the distribution of the cluster heads is more uniform. With the same consumption between this algorithm put forward in this study and the LEACH algorithm, the new algorithm can send more data. So it is easy to see that the new clustering algorithm is better than the traditional algorithm such as the LEACH.

ACKNOWLEDGMENT

This study was supported in part by NSFC under Grant No. 60802031 and Liaoning province innovation group project (LT2010091, LT2011005).

REFERENCES

Agah, S., K. Basu and M. Asadi, 2006. Intrusion detection in sensor networks: A non-cooperative game approach [C]. Proceeding of ICWN'06, pp: 29-36.

Eduardo, F.N., S.R. Heitor, A.V. Leandro, H.A.B.F. De Oliveira and A.L.L. De Aquino, 2009. A reactive role assignment for data routing in event-based wireless sensor networks [J]. Comput. Netw., 53(12): 1980-1996.

Giuseppe, A., M. Conti, M.D. Francesco and A. Passarella, 2009. Energy conservation in wireless sensor networks: Asurvey [J]. Ad Hoc Netw., 7(3): 537-568.

Heinzelman, W.B., A.P. Chandrakasan and H. Balakrishnan, 2002. An application-specific protocol architecture for wireless microsensor networks [J]. IEEE T. Wirel. Commun., 1(4): 660-670.

Jiang, C., Y. Dao-Rain and Z. Ying-Hui, 2009. Towards clustering algorithms in wireless sensor networks : A survey [C]. Proceeding of IEEE Conference on Wireless Communications and Networking Conference, Piscataway, IEEE Press, NJ, pp: 2009-2014.

Kurosawa, K. and Y. Desmedt, 2004. A New Paradigm of Hybrid Encryption Scheme [C]. Advances in Cryptology: Springer-Verlag, Berlin, pp: 426-442.

Mhatre, V. and C. Rosenberg, 2004. Design guideline for wireless sensor networks: Communication, clustering and aggregation [J]. Ad Hoc Netw. J., 1(2): 45-63.

Sankarasubramaniam, Y., O.B. Akan and I.F. Akyildiz, 2003. ESRT: Event-to-sink reliable transport for wireless sensor networks [J]. Proc. ACM MOBIHOC, Annapolis, MD, 6:77-188.

Shen, Y., W. Qi and H. Dai, 2008. An energy efficient clustering algorithm for surveillance sensor networks [J]. J. Softw., 19(9): 2432-2441.

Yah, J., L. Wang, Y. Kim and X. Yang, 2008. EEMC : An energy efficient multi-level clustering algorithm for large-scale wireless sensor networks [J]. Comput. Netw., 52(3): 542-562.

Yao, Y. and J.E. Gehrke, 2002. The cougar approach to in-network query processing in sensor networks [J]. Sigmod Record, 31(3): 9-18.

Younis, O. and S. Fahmy, 2004. HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks [J]. IEEE T. Mobile Comput., 3(4): 366-379.

Zhang, R., H. Zhao and M.A. Labrador, 2006. A grid-based sink location service for large-scale wireless sensor networks. International Conference on Comm. Mob. Comput., Vancouver, British Columbia, Canada, 7: 689-694.