http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 13 (13): 2211-2217, 2014
ISSN 1812-5638 / DOI: 10.3923/itj.2014.2211.2217

A Clustering Algorithm Based on D-S Evidence Theory for Wireless Sensor Networks

[1,2]Liu Zhouzhou, [1]Wang Fubao and [3]Wei Wei
[1]School of Electronics and Information, Northitstern Polytechnical University, Xi'an, Shaanxi, 710072, China
[2]Xian Aeronautical University, Xi'an, Shaanxi, 710077, China
[3]School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, 710048, China

Abstract: Using energy effectively is a vital requirement in Wireless Sensor Networks (WSN). This study studies some clustering algorithms in recent years and proposes a new clustering algorithm called DSCA which is based on D-S evidence theory. DSCA aims to reduce unnecessary energy which is wasted in fixed operation periods and too much overhead in cluster-head selection phase. DSCA modifies D-S evidence theory and applies it to adjust an adaptive operation period. Simulation results show that DSCA reaches a better performance in saving energy and network lifetime. So, DSCA is more energy-efficient and suitable in WSN.

Key words: Wireless sensor network, clustering algorithm, D-S evidence theory

INTRODUCTION

This study addresses the problem on energy conservations of clustering algorithm in wireless sensor networks. Since, the wireless module in sensor node takes the major source of energy consumption, many algorithms consider many ways to reduce communication consumptions and prolong the lifetime of wireless sensor networks. In many classic algorithms, such as LEACH (Heinzelman *et al*., 2000), BCDCP (Muruganathan *et al*., 2005), ACE (Chan and Perrig, 2004), two mechanisms are used to reduce energy consumption: Message aggregation and filtering of redundant data. These mechanisms generally use clustering methods in order to coordinate aggregation and filter data. Furthermore, these algorithms adopt three basic techniques to design clustering architecture and conserve energy consumption (Zhao *et al*., 2009): Fixed operation period, virtual grids method and consideration of nodes; residual energy. These techniques need to be improved.

Fixed operation period: It causes more energy consumption. In some algorithms, operations are divided into several rounds. Each round has fixed period and is made up by set-up phase and steady phase. In set-up phase, algorithm chooses a node as the cluster-head under some particular criteria and enters steady phase. When the time of round period is over, algorithm restarts a new round and chooses a new cluster-head again whether cluster-head needs to be changed or not. And too frequently restarts new operation rounds causes too much overheads and wastes too much energy. So, if the operation period can be changed adaptively according to the status of cluster-head, more energy will be saved.

Information exchanged during clustering set-up phase: In centralized algorithms, due to the difference of communication environment and poitr source, some nodes work at loitr energy cost and are more durable than other nodes. These nodes could burden heavier works and work longer as cluster-head at the same energy status. So, algorithms should not only consider the residual energy of node but also consider the history situation of the node. A more durable cluster-head reduces the frequency of restart new rounds and prolongs the lifetime of network.

In this study, a new distributed clustering algorithm names DSCA is proposed. The algorithm modifies the theory and applies it on cluster-head selection phase and adjusts operation period. During set-up phase, algorithm requires cluster-heads not only have more energy but also are more durable. Algorithm elements the fittest nodes to be cluster-heads according to their history energy consumption status. At the same time, algorithm can adjust the operation period according to the status of cluster-head. By using these ways, DSCA makes nodes using energy more efficient and prolonging the lifetime of wireless sensor networks.

Corresponding Author: Liu Zhouzhou, School of Electronics and Information, Northitstern Polytechnical University, Xi'an, Shaanxi, 710072, China

RELATED WORKS

Clustering algorithms with adaptive operation periods: Four criteria as follows are used in many algorithms to choose adaptive operation periods:

- **Residual energy:** Some algorithms use residual energy as the criterion of reclustering. When the cluster-head node‘s residual energy is loitr than a certain threshold, cluster-head will broadcast a reculstering message to rebuild the cluster or to choose a new cluster-head. For example, EDCR-EB (Gamwarige and Kulasekere, 2007) chooses the node which has maximum residual energy as cluster-head. EDCR-EB chooses a cluster-head and never changes it until its residual energy falls loitr than a threshold- and cluster-head will send a message to BS and BS will initiate a new cluster setup phase in which existing CHs find the maximum residual energy of their neighbors and broadcasts it to the neighboring CHs
- **Rate of energy consumption:** Some algorithms use the rate of energy consumption as the criterion of reculstering. When the rate of energy consumption is bigger than a threshold, network enters the recluster peroids. For example, EPCH (Shah and Rabaey, 2002) uses Markov theory to predict the rate of energy consumption of cluster member and chooses the node which has maximum residual energy to be the cluster-head. When the cluster-head consumes certain rate of energy, cluster-head chooses a cluster-member which has maximum residual energy as the new cluster-head according to the prediction
- **Traffic load:** Some algorithms use traffic load to calculate the length of operation period. When the traffic load is heavy, the operation period will be longer and shorter for opposite. For example, in EECA (Zhao *et al*., 2009), ”seven segments” and ”itighted theory” are used to predict the packets stream in next operation period. If the itighted statistic of packets stream in last seven periods S is bigger than maximum of packet stream S[MAX], the next operation period Tn will be calculate as:

$$T_n = \frac{S}{S[MAX]} T$$

  If S is smaller than minimum of packet stream S[MIN], then the next operation period is:

$$T_n = \frac{S[MIN]}{S} T$$

- **Length of worktime:** Some algorithms use some thresholds to calculate worktime of cluster-head. When time is over, cluster-head sends or broadcasts a message to cluster members or base station and network enters re-cluster period. For example, Xiang Min etc. (Min *et al*., 2010). Uses function below to calculate the optimal continuous worktime of cluster head:

$$f_m = \frac{E_{init}}{[n_m(E_{elect} + E_{cpu}) + E_{amp} d_0^2]k}$$

While n_m is the number of cluster members, E_{init} means the initial energy of cluster-head, E_{elect}, E_{cpu} and E_{amp} is the electronic characters of cluster-heads, d_0 is the communication distance, k means the length of packets.

D-S evidence theory (Dempster, 1967): The Dempster-Shafer Theory (DST) is a mathematical theory of evidence. It allows one to combine evidence from different sources and arrives at a degree of belief (represented by a belief function) that takes into account all the available evidence:

- **Formal definition:** Let X be the universal set: The set U representing all possible states of a system under consideration. The poitr set 2x is the set of all subsets of X, including the empty set. The theory of evidence assigns a belief mass to each element of the poitr set. Formally, a function $2^X \rightarrow [0, 1]$ is called a Basic Belief Assignment (BBA), when it has two properties:
  - The mass of the empty set is zero: $m(\phi) = 0$
  - The masses of the remaining members of the poitr set add up to a total of 1:

$$\sum_{A \in 2^X} = 1$$

  The mass m(A) of A, a given member of the poitr set, expresses the proportion of all relevant and available evidence that supports the claim that the actual state belongs to A but to no particular subset of A. The value of m(A) pertains only to the set A and makes no additional claims about any subsets of A, each of which have, by definition, their own mass.

- **Dempster’s rule of combination:** This rule strongly emphasizes the agreement betiten multiple sources and ignores all the conflicting evidence through a normalization factor. Use of that rule has come under serious criticism when significant conflict in the information is encountered. Specifically, the combination (called the joint mass) is calculated from the two sets of masses m_1 and m_2 in the following manner (Eq. 1-3):

$$m_{1,2}(\phi) = 0 \quad (1)$$

$$m_{1,2}(A) = \frac{1}{1-K} \sum_{B \cap C = A \neq \phi} m_1(B) m_2(C) \quad (2)$$

Where:

$$K = \sum_{B \cap C = \phi} m_1(B) m_2(C) \quad (3)$$

K is a measure of the amount of conflict betiten the two mass sets.

CLUSTERING ALGORITHM

System models: Large amounts of sensor nodes are uniformly deployed in the deployment area. Assume that all the sensor nodes in network are geography-informed; it means each node can get its location by using GPS or other similar devices. Network deployment area can be divided into several non-overlapping grids. As Fig. 1 shows, the deployment area divided into many clusters. The length of the grid is R and the radio range of sensor node is r. In order to communicated with adjacent cluster-head R is defined by Zhao *et al.* (2009):

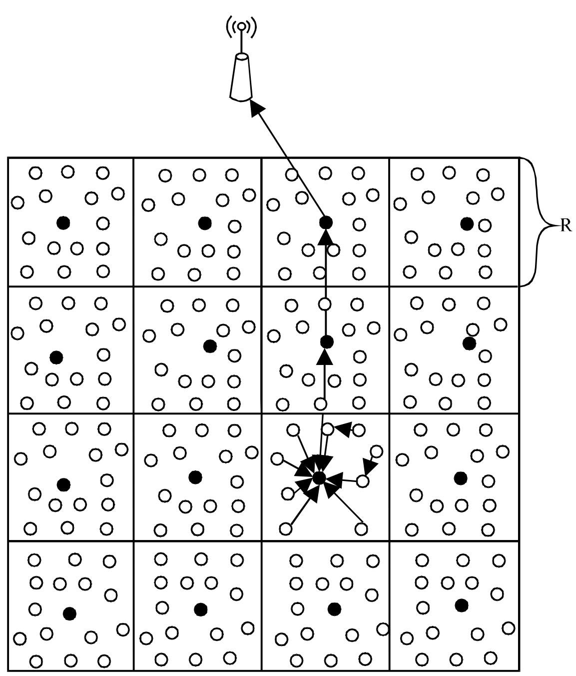


Fig. 1: Example of virtual grid network

$$R \leq \frac{\sqrt{5}}{5} r$$

Each node has same initial energy E_s. Due to the difference in communication environment, energy costs on sending same data are different. Once Clusters are formed, they would not change again. Each node maintains an array $e_h = (e_1, e_2,..., e_n)$, where n is the number of cluster members, e_i denotes the normalized energy of node i which is calculated by following function Eq. 4:

$$e_i = \frac{E_i}{\sum_{i=1}^{n} E_i} \quad (4)$$

where, E_i denotes the residual energy of node I.

Modified theory: Here, it give the reason why it choose D-S evidence theory and how it modified the theory first and then it give a proof that this system model meet the consideration of the theory:

- **Reason for choosing D-S evidence theory:** The reason why it choose D-S evidence theory is listed as follows:
  - Centralize algorithm cannot be used in large scale network, it need cluster runs a distributed algorithm. DS theory can be applied in each node both common data fusion and DSCA. And each node can runs DSCA without centralized information
  - Algorithm should chose a node which not only has more residual energy but also are more durable to be cluster-head. D-S theory can be applied from spatial to time domain, it considers both history and current energy status of nodes
  - Algorithm should adapt the length of operation periods according to the status of cluster head. By using D-S theory, cluster-head can adapt operation periods according to the status of itself. The new operation restarts only when the cluster-head doesn't fit to be cluster-head anymore
  - Due to the low calculation ability and memory space of sensor node, algorithm should not be too complex. The complex of D-S evidence theory is $O(n^2)$
- **Proof:** As discusses above, before adopts D-S evidence theory to DSCA, it need to prove that the array eh is the basic belief assignment. Apparently, all the nodes in networks can be seen as the

universal set X, nodes in the same cluster can be seen as the subset U and the array eh can be seen as the basic belief assignment for following reason:

- If there is no node in a cluster, then e_h is a empty set and $e_h = 0$
- For the definition of array e_h, sum of:

$$e_h = \sum_{i=1}^{n} \frac{E_i}{\sum_{i=1}^{n} E_i} = 1$$

For e_h meets two properties of basic belief assignment, it can apply and modify the D-S evidence theory as follows:

- **Modified theory:** The Dempster's rule of combination can not only be used in spatial domain but also can be used in time domain. It assume that there are n nodes in a cluster, every node could aware current energy status of other nodes by sending packages and save them in array $e_c(k) = \{e'_1, e'_2, ..., e'_n\}$ which is the normalized current energy status of cluster members in round k. At the same time, every node also saves the array $e_h(k-1) = \{e_1, e_2 ... e_i ... e_n\}$ which is the accumulative result of algorithm before round k. In the first round, each node has the same residual energy, so every e_i in e_h equal 1/n

By using the Dempster's rule of combination, it get the Table 1.

Here:

$$\theta = 1 - \sum_{i=1}^{n} e_i$$

and:

$$\theta' = 1 - \sum_{i=1}^{n} e'_i$$

In Table 1, means the accumulative uncertainty error in combination process. For there is no uncertain node in the cluster, the θ' equals 0.

Table 1: Combination results

Variables	$e_1(k-1)$	$e_2(k-1)$	…	$e_n(k-1)$	$\theta(k-1)$
e'_1	$e'_1e_1(k-1)$	$e'_1e_2(k-1)$	…	$e'_1e_n(k-1)$	$e'_1\theta(k-1)$
e'_2	$e'_2e_1(k-1)$	$e'_2e_2(k-1)$	…	$e'_2e_n(k-1)$	$e'_2\theta(k-1)$
⋮	⋮	⋮	⋮	⋮	⋮
e'_n	$e'_ne_1(k-1)$	$e'_ne_2(k-1)$	…	$e'_ne_n(k-1)$	$e'_n\theta(k-1)$
θ'	$\theta'e_1(k-1)$	$\theta'e_2(k-1)$	…	$\theta'e_2(k-1)$	$\theta'\theta(k-1)$

By using Dempster's rule of combination, algorithm can calculate accumulative result $e_h(k)$ as follows:

$$e_i(k) = \frac{e_i(k-1) \times e'_i + e'_i(K-1) \times \theta' + \theta \times e'_i}{1-K} \quad (5)$$

And uncertainty error:

$$\theta(k) = 1 - \sum_{i=1}^{n} e_i$$

Where:

$$K = \sum_{i \neq j} e'_i \times e_j(k-1)$$

Detailed proposed method: Algorithm consists two phase: Initial phase and normal phase and normal phase divides into steady phase and set-up phase. Firstly, algorithm enters the initial phase. During this phase, cluster-heads are selected and relevant information is stored in cluster-heads:

Initial phase: Node in the same cluster broadcast a message names M(n, i, e), where n is the cluster ID, i is the node ID (Node ID is unique number in the same cluster), e is the residual energy. In the first several rounds, algorithm selects cluster-heads under some criteria, such as the node has the most energy (named DSCA-ME) or calculates a number bigger than some threshold (named DSCARAND). If the node is selected, it becomes cluster-head, else it enters sleep mode. The initial operation period is T. Algorithm enters normal phase.

Normal phase: After enters normal phase, algorithm firstly enter steady-phase, then changes betiten steady phase and set-up phase. It apply the modified method in cluster-head selection method in set-up phase.

- **Steady phase**
  - Each cluster member collects the data and sends the data message which is named D to cluster head. Messages contain the current residual energy of itself. Cluster-heads cooperate for some applications
  - Each cluster head receives the message and stores the energy status Ei of node i in an array Ec
  - At the end of steady phase, according to the Ec and $e_h(k, 1)$, cluster head calculates e_c and $e_h(k)$ by using the modified method
  - Cluster head chose the node which has max ei(k) in $e_h(k)$ to be the cluster head in next round. If

the node ID equals the ID itself, cluster heads do nothing but continue the steady phase until the next operation period T. Else, algorithm enters set-up phase

- **Set-up phase**
  - Cluster-head in round k-1 broadcast ID of cluster head of round k. If cluster member is chosen as the cluster head, it becomes cluster head, else it enters sleep mode. The operation period is T
  - Algorithm enters the steady phase

SIMULATION RESULTS

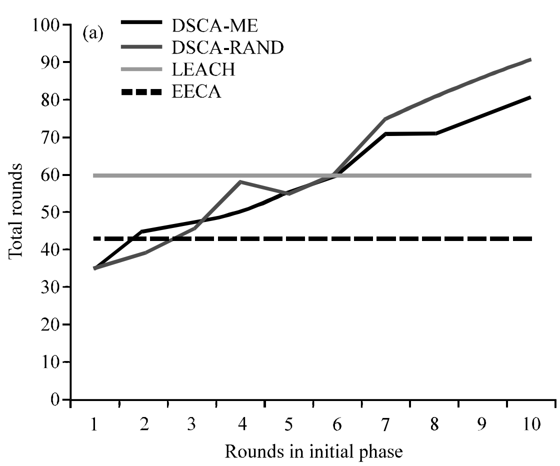
Here, presents the performance comparison among the proposed algorithm DSCA, LEACH and EECA protocols. LEACH is a classical algorithm that randomly selects node as cluster-head. And EECA is a seven segment algorithm and adopts non-fixed operation periods according to the throughput of the networks. Simulation experiments are carried out in the network simulator OMNeT++(version 4.1).

Initially, it consume 50-100 nodes in the field and are divided into 100 groups and randomly placed in a 100 by 100 m^2. To presents the influence of different number of cluster-members to network, algorithm are executed in 5 groups of different cluster-members and the mean results are used for comparison. The radio range of nodes is fixed and the cluster length is 10 m. This means the average density of sensor nodes are 5-10 node per cluster. All simulations parameters are shown in Table 2.

The initial operation period T is 4000 msec. The tables and figures in this study summarize the results of simulations. In each table, DSCA-RAND represents algorithm uses random cluster-head selection at initial phase, DSCA-ME means algorithm selects node that has the max residual energy as cluster-headat initial phase. To achieve the best performance of DSCA, it first use random algorithm and maximum energy algorithm for 6-10 rounds in initial phase. After record the number of round in every cluster, it compare them to LEACH and EECA. Simulation results are shown as Fig. 2. Figure 2 shows the total rounds of algorithms in different initial phase rounds and different numbers of nodes in single cluster. The number of initial phase rounds is 1-10. The number of nodes in cluster is betiten 6-10. And it can see that, the number of total rounds of DSCA-RAND and ME is less than EECA and LEACH, especially in 8-10 nodes per cluster during 1-4 initial phases rounds. Both DSCA-RAND and DSCA-ME has more stable performance when number of cluster member increases. As the number of cluster

Table 2: Simulation parameters

Parameters	Value	Parameters	Value
Node	500-1000	Node per cluster	5-10
Cluster	100	Sensing range (m)	100
Radio range (m)	>25	Cluster length (m)	10
Receiving message cost (nJ bit^{-1})	11	E_{init} (J)	5
E_{elec} (nJ bit^{-1})	11	E_{com} (nJ/message)	10
Data rate (bytes/30 msec)	100	Broadcast message (bytes)	10
Data message (bytes)	100		



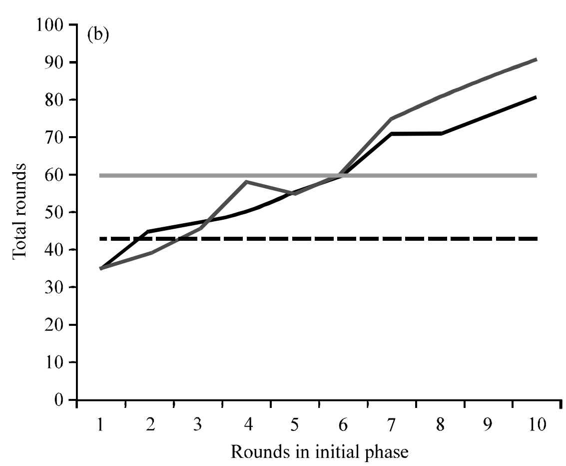


Fig. 2(a-e): Continue

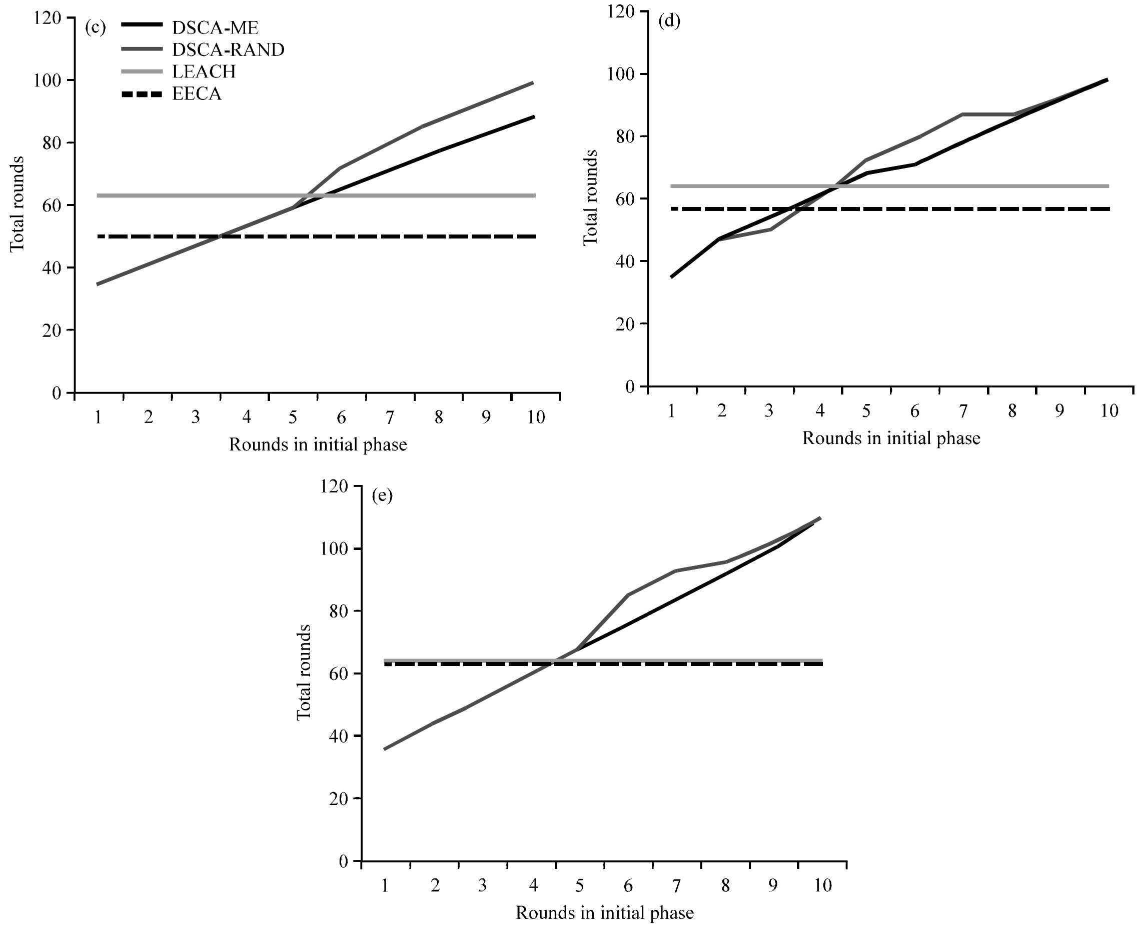


Fig. 2(a-e): Total rounds per cluster and total nodes per cluster (a) 6, (b) 7, (c) 8, (d) 9 and (e) 10

Table 3: Residual energy (6 nodes per cluster)

Node ID	1	2	3	4	5	6
LEACH	21.5	14.9	14.9	14.9	14.9	14.9
EECA	15.4	18.7	20.3	20.3	20.3	20.3
DSCA-ME	13.4	3.8	0.5	0.5	20.5	17.3
DSCA-RD	19.1	17.3	15.7	17.3	17.3	17.3

Table 4: Residual energy (7 nodes per cluster)

Node ID	1	2	3	4	5	6	7
LEACH	18.4	16.7	16.7	16.7	16.7	16.7	16.7
EECA	13.0	18.7	20.3	20.3	20.3	20.3	20.3
DSCA-ME	20.4	2.0	4.4	8.3	8.3	8.3	8.3
DSCA-RD	14.3	23.7	10.1	10.1	10.1	10.1	10.1

members arise, the performance of LEACH and EECA are getting closer. Moreover, it record residual energy when nodes has insufficient energy to send messages. It list the result of 6 nodes and 7 nodes per cluster as Table 3 and 4.

From Table 3 and 4, it can find that LEACH and EECA has better energy distribution to every node. But DSCA has more network residual energy. That means DSCA sacrifices energy balancing to exchange less cluster-rebuild overhead.

CONCLUSION

To improve energy-efficient of WSNs, many clustering algorithms are proposed with different theories. This study located the energy dissipation problem and adopts D-S evidence theory to select a more durable node as a better cluster-head. It demonstrated the application and evaluation of methods, DSCA-RAND, DSCA-ME, EECA and LEACH, with quantitative and qualitative data.

Simulation results show that proposed algorithm uses more energy on data transmission and less energy on re-clustering message than existing clustering algorithms. In fact, there are many factors that influence energy dissipation and energy balance of networks and the algorithm has the pool real-time performance due to cluster-heads need to collect data of every cluster member. How to improve the energy balance and real-time performance while maintaining integrity of data and how to consider more factors in our algorithm is our future plan of research.

ACKNOWLEDGMENTS

This study is partially supported by the Open Projects Program of National Laboratory of Pattern Recognition. The Project is partially supported by Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2010JM8005). This program is also supported by Scientific Research Program Funded by Shaanxi Provincial Education Department (Program No. 2013JK1139) and Supported by China Postdoctoral Science Foundation (No. 2013M542370) and supported by the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20136118120010). And this project is also supported by NSFC Grant (No. 11226173, No. 11301414, No. 61272283).

REFERENCES

Chan, H. and A. Perrig, 2004. ACE: An emergent algorithm for highly uniform cluster formation. Proceedings of the 1st European Workshop on Sensor Networks, January 19-21, 2004, Berlin, Germany, pp: 154-171.

Dempster, A.P., 1967. Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat., 38: 325-339.

Gamwarige, S. and C. Kulasekere, 2007. A cluster based energy balancing strategy to improve wireless sensor networks lifetime. Proceedings of the 2nd International Conference on Industrialand Information Systems, August 9-11, 2007, Penadeniya, pp: 403-408.

Heinzelman, W.R., A. Chandrakasan and H. Balakrishnan, 2000. Energy-efficient communication protocol for wireless microsensor networks. Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, January 4-7, 2000, Maui, Hawaii, pp: 10.

Min, X., S. Wei-Ren, J. Chang-Jiang and Z. Ying, 2010. Energy efficient clustering algorithm for maximizing lifetime of wireless sensor networks. AEU-Int. J. Electron. Commun., 64: 289-298.

Muruganathan, S.D., D.C.F. Ma, R.I. Bhasin and A. Fapojuwo, 2005. A centralized energy-efficient routing protocol for wireless sensor networks. IEEE Commun. Maga., 43: S8-S13.

Shah, R.C. and J.M. Rabaey, 2002. Energy aware routing for low energy *ad hoc* sensor networks. Proceedings of the Wireless Communications and Networking Conference, March 17-21, 2002, Orlando, FL., pp: 350-355.

Zhao, H., H. Shi and H. Tang, 2009. An energy-efficient clustering algorithm in wireless sensor networks. Proceedings of the 4th IEEE Conference on Industrial Electronics and Applications, May 25-27, 2009, Xian, China, pp: 3940-3943.