Clustering Algorithm based on Fuzzy Weight for Wireless Sensor Networks

Teng Gao
School of Mechanical Engineering & Automation
Dalian Polytechnic University
Dalian, China
gaoteng@dlpu.edu.cn

Jin-Hua Ding   De-Quan Wang
School of Mechanical Engineering & Automation
Dalian Polytechnic University
Dalian, China
jxmcu@dlpu.edu.cn   wangdq@dlpu.edu.cn

Jin-Yan Song
School of Information Engineering
Dalian Ocean University
Dalian, China
thesjyyan@163.com

Zhen-Yuan Si
Department of Automation
Henan Senyuan Electric Co., Ltd.
Changge, Henan, China
sikairui@126.com

Abstract—Because of the non-uniform cluster-head election in the classical clustering algorithm, some nodes maybe exhaust energy prematurely so that it is not fit for large-scale wireless sensor networks(WSNs). In this paper, a distributed clustering routing algorithm based on fuzzy weight attribute degree(FWAD) is proposed. The direct methodology of fuzzy engineering theory is adopted to assign relevant weight to each parameter by taking all parameters into account synthetically, which makes each node calculates its own attribute value. The attribute value will be mapped to the time coordinate axis so that the node can broadcast cluster head information by means of timer triggering, meanwhile, the density method is adopted to avoid collisions and to ensure the symmetrical distributing of the cluster-head. Multi-hop is adopted to forward aggregate data to the sink. Simulations denote that FWAD algorithm has longer lifetime and better expansibility than LEACH-like algorithms.

Keywords- Wireless Sensor Networks; Routing; Cluster; Fuzzy Weight; Attribute

I. INTRODUCTION

The recent technological advances make wireless sensor networks (WSNs) technically and economically feasible to be widely used in both military and civilian applications, such as monitoring of ambient conditions related to the environment, precious species and critical infrastructures[1]. A sensor network consists of multiple wireless sensor nodes and a stationary sink node connected to the Internet[2]. Since sensor nodes are generally battery-powered devices, their operational lifetimes are limited[3]. Therefore, saving energy and extending lifetime has been the focus of current research. Researchers have developed many measures to save energy from all aspects, from which we have our sight on cluster-based routing protocol.

In a clustering architecture, cluster-head nodes can be used to process and send the information to the sink while member nodes can be used to perform the sensing in the proximity of the target and transmit the information to corresponding cluster-head. This means that creation of clusters and assigning special tasks to cluster heads can greatly contribute to overall system scalability, lifetime, and energy efficiency [4]. However, the disadvantage of cluster-based routing is obvious. That is, the cluster-head is so vital that it becomes the bottleneck of the whole network. Therefore, the selection of the cluster-head will influence the performance of the whole network. The existing clustering algorithms differ on the criterion for the selection of the cluster heads. According to the current research results, the selection method of the cluster-head can be divided into several categories below.

The classical routing protocols that based on k-means clustering such as LEACH[5] select cluster-head only based on a random acquired value. If this value is less than a certain threshold, the nodes will be the cluster-head. Whereas, because of the randomicity of the selection of the cluster-head, the uneven traffic flow in different cluster-heads may occur. Thus, some cluster-head will exhaust energy, and the performance of the network is affected.

Some routing protocols based on a certain attribute are proposed in [6] and [7]. The attributes include residual energy, neighbors' number, communication cost in intra-cluster, the distance to the sink and ID, etc. Because only one attribute is taken into account in these protocols, the selected cluster-head is not the most suitable node. Although the rationality of the cluster-head selection is improved to a certain extent, the uneven distribution of cluster-head or the unbalanced load still exist.

Based on an overall consideration of various factors, multi-attribute cluster-head selection protocol such as HEED[8], WCA[9] use several attributes to determine the cluster-head. The multi-attribute cluster-head selection can result in a better partition of cluster. The two protocols adopt successive screening method to determine the cluster-head, by which the finite iteration must be implemented. The major drawback of the former is that distributed algorithm makes each node not know global information so that some nodes may not join any clusters. While the latter need iterate many times if many attributes

are used to gain a better performance, which will increase time complexity and consume more energy.

MWBC[10] and AOW-LEACH[11] combine a few factors such as residual energy, communication cost, neighbor nodes number, etc. to one value to avoid the randomicity in the cluster-head selection. However, all the algorithms above use trial and error method to determine the weight of each factor, which will influence the performance of the whole protocol.

From the analysis and comparison mentioned above, multi-attribute cluster-head selection can obtain rational partition of clusters. Therefore, we consider the residual energy, neighbors' number, the cost that communicate in intra-cluster, and the distance to the sink as the attributes to propose a cluster-based routing algorithm using fuzzy multi-attribute weights to determine cluster-head selection, by which the energy can be saved and the lifetime of the whole network will be extended.

II. THE WEIGHT OF ATTRIBUTE

Aimed at saving energy and balancing load, the residual energy is the most crucial attribute during the cluster-head selection while the cost that communicate in intra-cluster and the neighbors' number also influence cluster-head selection. In this paper, the direct method [12] based on the abutting object relative membership degree in engineering fuzzy theory and intelligence decision-making is adopted in order to confirm the proportion of each attribute during cluster-head selection.

Definition Compare the member O_k with another member O_l on duality about weightiness in the object set O. When O_k is more important than O_l,

$$0.5 < \beta_{kl} \leqslant 1;$$

when O_l is more important than O_k,

$$0 \leqslant \beta_{kl} < 0.5, \text{ and } \beta_{kl} = 1-\beta_{lk};$$

when O_k has the same importance as O_l,

$$\beta_{kl} = 0.5, \text{ especially, } \beta_{kk} = 0.5,$$

in which the β_{kl} is named relatively weightiness fuzzy value between the object O_l and O_k. Especially, if the object sequencing about weightiness is $O_1 \prec O_2 \prec \ldots \prec O_m$, $\beta_{k1,k1+1}$(k1=1,2,…,m-1) is defined abutting object relatively weightiness fuzzy value.

TABLE I.   RELATIONSHIPS BETEEN MOOD OPERATOR AND FUZZY VALUE

mood operator	fuzzy scale value	memberships value
similar	0.5	1
ratherish	0.55	0.905
slightly	0.6	0.667
relatively	0.65	0.538
obviously	0.7	0.429
markedly	0.75	0.333
quite	0.8	0.25
very	0.85	0.176
extremely	0.9	0.111
violently	0.95	0.053
uncomparable	1	0

According to the sequencing on weightiness, residual energy (E_r) $\succ$ intra-cluster communication cost (*Cost*) $\succ$ neighbors number (*Deg*) can be obtained from attributes. That is, residual energy has the unexampled importance than intra-cluster communication cost while the latter is more important than neighbors' number ratherish. The relevant fuzzy value that $\beta_{Er,Cost}$ is 1 and $\beta_{Cost,Deg}$ is 0.55 can be found out based on the Tab.I.

Based on the definition about relative significance fuzzy scale value, provided that the object E_r is more important than *Cost*, $\beta_{Er,Cost}$ is the corresponding significance degree when just comparing object E_r and *Cost*, which benchmark is E_r, the more important one between these two attributes. Because the $\beta_{Er,Er}$ is 0.5, if the only two objects E_r and *Cost* are still compared, the degree that E_r belongs to significance is β'_{Er}=1, and the one of *Cost* is β_{Cost}= 1.5- $\beta_{Er,Cost}$. Therefore, the relationship of the significance degree between E_r and *Cost* is

$$\frac{\beta'_{Er}}{\beta'_{Cost}} = \frac{1}{1.5 - \beta_{Er,Cost}}. \tag{1}$$

The non-normalization weight may be figured out:

$$\omega'_{Er} = 1 \quad \omega'_{Cost} = \omega'_{Er}(1.5\text{-}1) = 0.5$$

$$\omega'_{Deg} = \omega'_{Cost}(1.5\text{-}0.55) = 0.475$$

The object weight vector that obtained after normalization and reverting suffix is：

$$\omega = (\omega_{Er}, \omega_{Cost}, \omega_{Deg}) = (0.5063, 0.2532, 0.2405) \tag{2}$$

On account of the diffidence in the unit of each attribute, normalization procedure will be implemented.

The normalization expression of residual energy is E_r /E_{max}, in which E_{max} is the original energy of each node. A function relationship exists between residual energy and cluster-head selection, that is, the node whose residual energy is higher has more chances to be a cluster-head.

The intra-cluster communication cost is obtained by calculating received signal strength information (RSSI), which is normalized as *RSSIave/TSSV*, where TSSV is the transmission signal strength value. *TSSV* will be same in the broadcasting phase of each node. The *RSSIave* denotes the average strength value of all the wireless signals that have been received. The bigger the value is, the lower the cost is. The *RSSIave* is also proportional to the probability that the node can be selected as the cluster-head.

Neighbors' number is an important attribute in the energy-efficient sensor networks. Theoretically, the closer the neighbors' number approach the optimal number, the greater probability a node becomes a cluster-head with[13]. Therefore, the evaluation value of the number of neighbors is described as:

$$Deg = \frac{|D_i - D_0|}{D_0}. \tag{3}$$

Here D_i represents the number of neighbors of the node, and D_0 is the optimal number of neighbors, which can be obtained according to [14]. When the number of neighbors is equal to the optimal number, the evaluation value of the number of neighbors should reach 1.

The absolute attribute degree value of each node based on OWA operator can be calculated out by the object function below.

$$F = \omega \times (\frac{Er}{E\max}, \frac{RSSI_{ave}}{TSSV}, F(Deg))^T \tag{4}$$

In like manner, the function

$$F = \left(\frac{Er}{E\max}\right)^{\omega_{Er}} + \left(\frac{RSSI_{ave}}{TSSV}\right)^{\omega_{Cost}} + F(\mathrm{Deg})^{\omega_{Deg}} \quad (5)$$

is the absolute attribute degree value based on GOWA plus operator while the expression

$$F = \left(\frac{Er}{E\max}\right)^{\omega_{Er}} \times \left(\frac{RSSI_{ave}}{TSSV}\right)^{\omega_{Cost}} \times F(\mathrm{Deg})^{\omega_{Deg}} \quad (6)$$

figures out the absolute attribute degree value based on GOWA multiple operator.

Obviously, $\omega_{Er} + \omega_{Cost} + \omega_{Deg} = 1$. $\qquad (7)$

III. FWAD ALGORITHM DESCRIPTION

Cluster-head maldistribution will result in unbalanced load in the network and premature death of some nodes, so tackle this issue is primary. The solution is to use the fuzzy weight attribute degree algorithm (FWAD) to establish a clustering routing in network layer.

FWAD algorithm comprises two phases, initialization and operation. There are several time slices in initialization phase for receiving the signal from the sink and implementing flooding to obtain the grads level. Operation phase contains set-up phase and steady-state phase. Four sub-periods such as node broadcasting, cluster-head broadcasting, joining cluster and TDMA schedule broadcasting are within the former while the latter is consisted of a few rounds which contain several frame.

A. Initialization

The sink broadcasts a beacon at a certain power, the sensor node who received the beacon signal should limit within a region at the radius of $(d_{max} - d_0/8)$. After a period of delay, the sink broadcasts another beacon at the maximum power of the sensor node, by which the radio wave covers a circle region at the radius of d_{max}. The node that has received this beacon evaluates the distance between the sink and itself based on the RSSI, as well as gain the grads level 0.
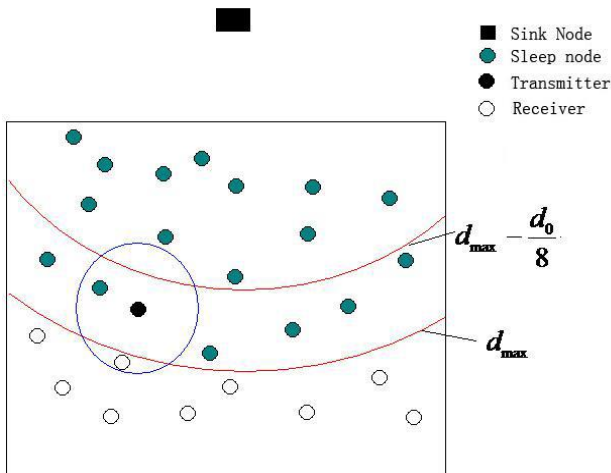

Figure 1. Flooding in initialization phase

After evaluating the distance, the node who received both beacons turns off transceiver and goes into dormancy. The node only received the second beacon wakes up and starts broadcasting its own grads level at the radius of $d_0/8$ at a random time in the certain interval during which all nodes that hold the same grads level will complete grad broadcasting, then goes to sleep again. The nodes, which have never received any signals before, receive this message and set its grads level as 1(received message plus 1), from which the distance between the sink and itself is considered as $(d_{max} + d_0/8)$. After that, the receivers broadcast their grads levels at the same radius of $d_0/8$ at a random time before going into dormancy. The node who receives this message sets their grads level to 2 and broadcast their grads level. The rest may be deduced by analogy until each node in the network has a grads level.

B. Clustering

Each node broadcasts a message $<ID, E_r>$ at a certain power in a period of time T_b, which covers a region at the radius of $d_0/2$. The node receives the messages from neighbors and stores the information into memory after the end of broadcasting, based on which each node calculates out neighbor number, the average residual energy and the intra-cluster communication cost $RSSIave$. Thus, the node can obtain all attributes it wants.

A calculation will be implemented in terms of (4) or (5) or (6) to obtain the absolute attribute degree F by each node. The node whose absolute attribute degree F is bigger has larger probability of being the cluster-head than the smaller one, because the former has great advantage over the latter in the energy efficiency. Then the absolute attribute degree F is mapped onto the time axis. The cluster-head broadcasts by means of the timer triggering.

A timer Ti, whose time span is determined by the absolute attribute degree F, is set for each node. Since the node whose absolute attribute degree is bigger broadcasts cluster-head information at earlier time, so

$$Ti = \{1 - [\lambda \times F + (1 - \lambda) \times rand(0,1)]\} \times T' \quad (8)$$

Here T' is the total time in which all cluster-head broadcast information. A constant λ is set to be 0.9 to avoid the nodes to hold the same absolute attribute degree value and to broadcast cluster-head simultaneously. Timer overtime, a $<ID>$ message is broadcasted at the radius of d_0 by the cluster-head. The nodes that can receive and parse the message correctly will lose the chance of being the cluster-heads if the sender is in the neighbor list. This is the density method, which makes the cluster-head distribute evenly in the network. The cluster-head broadcasting is finished while T' is over.

C. The Establishment of Routing

After receiving the broadcasting of cluster-heads, the member nodes select the nearest cluster-head based on the RSSI and send the join information to it. The distance between the member node and the corresponding cluster-head is evaluated as well. If the distance is longer than the one to the sink, the member will communicate with the sink directly at a fixed timeslice regardless of cluster-head while go to dormancy at the rest time to save energy.

As for a cluster-head, the nearest cluster-head will be selected to join based on the RSSI. The distance between the relational cluster-heads is evaluated in the same way. If the cluster-head is longer to the sink, or there is no cluster-head nearer the sink than itself, this cluster-head communicates directly with the sink. The cluster-head assigns a time slot for each member after receiving all join information, by which a TDMA schedule is schemed. The cluster-heads promulgate the schedule based on the order

of the grad level. As one of the members, the cluster-head communicate with its cluster-head at the appointed time slot, thus the routing in inter-cluster is established.

In intra-cluster, the members, who go to dormancy at the rest time to save energy, communicate directly with the cluster-head at the appointed time slot.

D. Data Transmission

The interval in data transmission is much longer than the time in setup phase to reduce the energy dissipation further. Compared with the LEACH-like algorithms, FWAD algorithms have longer time in data transmission.

At data transmission phase, the member nodes send information to the cluster-head according to the schedule, and then go into dormancy. While the cluster-head must keep working state to receive the information come from its members, and send the aggregated data to the next hop at its own time slot. The cluster-head that communicates directly with the sink implements data fusion after a frame, and the send the aggregated data to the sink.

IV. SIMULATION AND ANALYSIS

A. Simulation Platform Selection

NS2 is adopted as the simulation platform. The related models and parameters are described in [5]. In this scenario, 500 nodes are deployed in 200m×200m area and the sink node is located at (100,275).

LEACH-like algorithms such as LEACH, DCHS and AOW-LEACH are simulated and compared with FWAD family algorithms in the same scenario.

FWAD family algorithms comprise FWAD_1, FWAD_2 and FWAD_3, which are developed based on different absolute attribute degree value F from (4), (5) and (6), respectively.
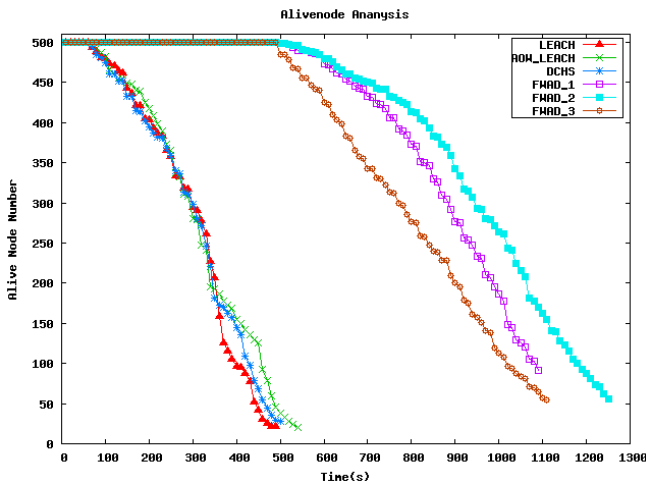
B. Performance Analysis

Figure 2.   Alive nodes comparison diagram

Fig .2 denotes the relation between nodes number alive and runtime, from which it is obvious that FWAD family algorithms enhance 4 times approximately more than LEACH, AOW_LEACH and DCHS on the total runtime of the whole network. The time of the first dead node is at 500th second in FWAD family algorithms while 80th

second is available in LEACH-like algorithms, which is a great improvement. The main reason of the results is that the power that can cover the circle region at the radius of $d_0/2$ is used to broadcast information to neighbors, in addition to the multi-hop routing in inter-cluster and the longer interval of data transmission is adopted to reduce and balance the energy consumption of the whole network, which make the energy of each node be saved and be used efficiently, so that the lifetime of the network is extended.
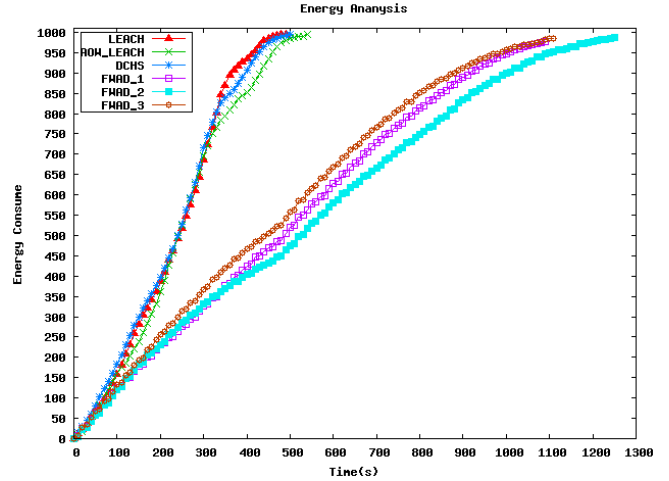

Figure 3.   Energy analysis chart

Energy analysis demonstrates that FWAD family algorithms consumed similar energy at each round during the network operation while the different energy dissipation at each round in the LEACH-like algorithms especially after the first node death, as described in Fig .3.
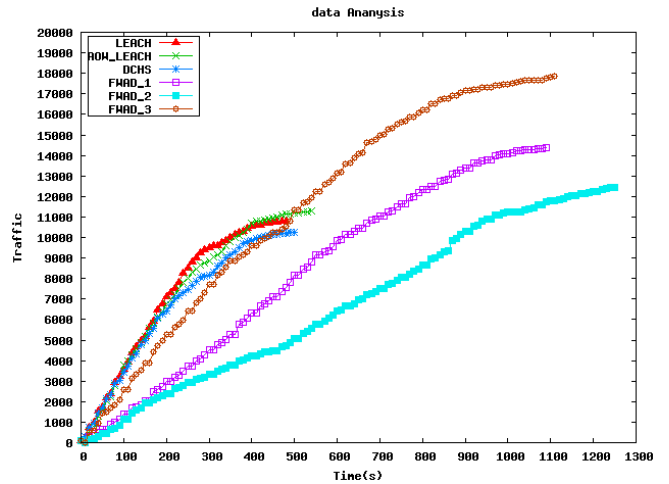

Figure 4.   Traffic comparison diagram

The traffic that is received by the sink is shown in Fig .4, from which it is indicated that the traffic of FWAD family algorithms are much less than that of LEACH-like algorithms. The reason for the great differences of traffic is that cluster-head node only implements data aggregation once before data is sent to the sink in LEACH-like algorithms while multiple data aggregations are run during the process of data being delivered to the sink in FWAD.

C. Parameters

1) Node broadcasting time T_b: Each node broadcasts information to the neighbors at the radius of $d_0/2$ at the beginning of each round, the time span of which is the pivotal factor that may influence the usage of the energy. If T_b is too large, each node will add waiting time so that unwanted energy in idle state is consumed. However, the parameter is connected with the network size. If the network size is too large, T_b must be enlarged to avoid the collisions on account of broadcasting in the limited time.

2) Cluster-head broadcasting time T': Cluster-head broadcasting is transmitted by radio in turn based on the time order that is mapped by the absolute attribute degree of its own. T', which is the total time span in the process of the broadcasting of cluster-heads, is also a significant factor that influences energy efficiency. If T' is too small, the density method will not be implemented. The large T' will result in energy consumption in the waiting time, which makes the lifetime of the entire network shorten.

D. Complexity Analysis

Assumed that there are n nodes in the network, and the nodes broadcast n $<ID, E_r>$ messages during the cluster-head selection, followed by k cluster-head broadcasting if k cluster-heads are selected in the network. Even if only one cluster-head can communicate directly with the sink, n-1 join messages will be broadcast by all nodes. Furthermore, k cluster-heads will broadcast k TDMA schedule subsequently. Thus, the total message spending in the phase of cluster forming is $n+k+n-1+k=2n$ in the whole network, which denotes that the message complexity of FWAD family algorithms in the setup phase is $O(n)$.

All nodes finish broadcasting within T_b, while the timer of each node will stop when cluster-head broadcasting interval T' is over. Therefore, the total time span of the algorithm FWAD in setup phase is $T_b+ T'$.

V. CONCLUSION

Cluster-based routing algorithm is an important research issue, which will also influence the operational efficiency of network. On the basis of analysis and comparison of some classical algorithms, a cluster-based routing algorithm FWAD is proposed. The fuzzy weight absolute degree is introduced into to make the most factors that can influence energy efficiency become an organic whole to determine the selection of the cluster-head, which is the main innovation and improvement of the classical algorithms. Moreover, FWAD algorithm supports data fusion both in inter-cluster and intra-cluster, which can eliminate the redundant data effectively so as to reduce the traffic and save the energy. In addition, FWAD algorithm selects the nearest path to forward the aggregated data to the sink at the type of multi-hop by comparing the distance between node and the corresponding cluster-head and that between node and the sink. The simulation results show that the lifetime and energy efficiency of FWAD family algorithms is better than the classical algorithm.

REFERENCES

[1] D. Tang, T.T. Li, J. Ren, and J. Wu, "Cost-Aware SEcure Routing (CASER) Protocol Design for Wireless Sensor Networks," IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 4, APRIL 2015, pp. 960-973.

[2] K. Sho and H. Hiroaki, "Intermittent Wireless Multihop Transmission Protocol in Mobile Wireless Sensor Networks," 8th International Conference on Signal Processing and Communication Systems(ICSPCS 2014), Dec, 2014, pp. 1-8.

[3] V. Cagri Gungor and Gerhard P. Hancke. Industrial Wireless Sensor Networks: Applications, Protocols, and Standards. CRC Press. Series Industrial Electronics, USA, 2013, 119-136, ISBN: 978-1-4665-0051-8.

[4] J.N. Al-Karaki, A.E. Kamal, A Taxonomy of Routing Techniques in Wireless Sensor Networks, Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems. CRC Press LLC, 2005, pp. 116-139.

[5] W.B. Heinzelman, A.P. Chandrakasan and H. Balakrishnan, "Energy-efficient communication proocol for wireless microsensor networks," Proc. of the 33rd Annual Hawaii Int'l Conf. on System Sciences, 2000, pp. 3005-3014.

[6] MJ. Handy, M. Haase and D. Timmermann, "Low Energy Adaptive Clustering Hierarchy with Deterministic Cluster-Head Selection," Proc of the 4th IEEE Conf. on Mobile and Wireless Communications Networks, 2002, pp. 368-372.

[7] M. Chatterje, S. K. Das and D. Turgut, "An on-demand weighted clustering algorithm for Ad hoc networks," Proc. of IEEE GLOBECOM 2000, November 2000, pp. 1697-1701.

[8] O. Younis and S. Fahmy, "Heed: A hybrid, energy-efficient, distributed clustering approach for ad-hoc sensor networks," IEEE Trans. on Mobile Computing, vol. 3, no. 4, 2004, pp. 660-669, doi: 10.1109/TMC.2004.41.

[9] M. Chatterjee, S. K. Das, and D. Turgut, "WCA: A weighted clustering algorithm for mobile Ad hoc networks," Journal of Cluster Computing, vol. 5, no. 2, 2002, pp. 193-204.

[10] H.Q. Huang, D.Y. Yao, J .Shen, K. Ma and H.T. Liu, "A Multi-weight Based Clustering Algorithm for Wireless Sensor Networks," Journal of Electronics & Information Technology, vol. 30, no. 6, 2008, pp. 1489-1492.

[11] B. Cai and X.D. Chen, "Clustering Algorithm Based on Automatic On-Demand Weighted for Sensor Networks," Microelecttronics & Computer, vol. 25, no. 11, 2008, pp. 129-132.

[12] X. C. Huang, "A Study on Theories and Methodologies for Fuzzy Multi-objective Decision Makings with their Applications," PhD dissertation, School of civil engineering, Dalian University of Technology, Dalian, China, 2003.

[13] T. Gao, R.C. Jin, J.Y. Song, T.B. Xu and L.D. Wang, "Energy-efficient Cluster Head Selection Scheme based on Multiple Criteria Decision Making for Wireless Sensor Networks," Wireless Personal Communications, vol. 63, no. 4, 2012, pp. 871-894.

[14] T. Gao, R.C. Jin, T.B. Xu and L.D. Wang, "Energy-Efficient Hierarchical Routing for Wireless Sensor Networks," Ad Hoc & Sensor Wireless Networks, vol. 11, no. 1-2, 2011, pp. 35-72.