A Novel Fuzzy Metric to Evaluate Clusters for Prolonging Lifetime in Wireless Sensor Networks

Peyman Neamatollahi, Hoda Taheri, Ehsan Toreini
Dept. Computer Engineering
Young Researchers Club, Mashhad branch, Islamic Azad University, Mashhad, Iran
{neamatollahi, h.taheri}@ieee.org, toreini@gmail.com

Mahmoud Naghibzadeh
Dept. Computer Engineering
Ferdowsi University of Mashhad
Mashhad, Iran
naghibzadeh@um.ac.ir

Mohammad Hossein Yaghmaee
Dept. Computer Engineering
Islamic Azad University, Mashhad branch
Mashhad, Iran
yaghmaee@ieee.org

***Abstract*—The most important consideration in designing protocols for wireless sensor networks is the energy constraint of nodes because in most cases battery recharging is inconvenient or impossible. Therefore, many research have been done to overcome this demerit. Clustering is one of the main approaches in designing scalable and energy-efficient protocols for wireless sensor networks. The cluster heads take the task of data aggregation and data routing to decrease the amount of communication and this prolongs the network lifetime. To the best of our knowledge, in spite of the importance of clustering in extending network lifetime, no metrics have been proposed for evaluating the quality of generated clusters. In this paper, we define some metrics to compare different protocols in terms of clustering qualification. Then we combine these metrics using fuzzy logic to better conclude about the superiority of different clustering schemes. The correction and feasibility is validated in simulations of three different protocols (LEACH, HEED, and HEED-NPF) in Matlab software.**

Keywords-sensor networks; clustering metrics; network lifetime; fuzzy logic;

I. INTRODUCTION

In general, the wireless sensor networks (WSNs) are deployed for monitoring at a large area so the WSNs need many sensor nodes. The sensor nodes have limitation in computation capability, power and memory. Especially the limitation of energy is very important design factor because of remaining energy of each sensor node is directly related to the lifetime of the WSNs. Therefore, the network topology research based on clusters is always the focus in recent years [1]-[3].

In cluster based WSNs, Cluster Head (CH) is responsible for gathering sensed data from its members and sending the information to the Base Station (BS). In this way, the sensor nodes can reduce communication overheads that may be generated if each sensor node reports sensed data to the base station independently. Clustering coordinates the transmissions of sensor nodes with a common schedule in the steady state phase, which removes collisions, idle listening, and overhearing. Therefore, clustering achieves an important improvement in terms of energy consumption. Moreover, each sensor node knows when to transmit; thus, it does not require being awake during the complete Time-Division Multiple Access (TDMA) frame but only at its specific time slot [4]. On the other hand, clustering is particularly crucial for scaling the network to hundreds or thousands of nodes [5]. Therefore, it is one of the basic approaches for designing energy-efficient and highly scalable distributed WSNs. In many applications, cluster organization is a natural way to group spatially close sensor nodes, so that exploit the correlation and eliminate redundancy that often exists among the sensor readings [6]. Aggregation and fusion of sensor node's data at the CHs motivate significant reduction in the amount of data sent to the BS; therefore result in saving energy and bandwidth resources.

In spite of the importance of clustering in prolonging lifetime of the sensor networks, no metrics have been provided to evaluate the quality of clusters formed via different protocols. In this paper we attempt to propose some metrics for this purpose. Then, by means of fuzzy logic we combine these metrics as a fuzzy metric to decide on the superiority of different protocols in terms of their cluster formation. The more the value of the fuzzy metric, the better the protocol performs clustering and consequently the longer lifetime will be gained in wireless sensor networks.

We simulated three different protocols in Matlab software:

- LEACH protocol [7], which is a probabilistic clustering protocol that does not consider energy of sensor nodes

978-1-4244-9834-5/11/$26.00

in CH election. Besides it applies single-hop routing and regular nodes join the closest CH,

- HEED protocol [8], which is an iterative clustering protocol that uses the energy of nodes and a communication cost to elect CHs. It applies multi-hop routing for the communications between CHs and the BS. In this protocol regular nodes join the least cost CH,
- HEED-NPF protocol [9], which is an improvement on HEED protocol that uses fuzzy logic and a non probabilistic approach for CH election. In [9], the authors showed that HEED-NPF acts better than the other two protocols in terms of network lifetime.

In this paper, we implemented the fuzzy metric on all the three protocols. Simulation results demonstrate that HEED-NPF overcomes the other protocols in terms of better cluster formation. As a result, the fuzzy metric is correctly chosen because when a protocol achieves a higher value of fuzzy metric it means that it better forms the clusters which results in more network lifetime extension.

The rest of the paper is organized as follows: Section 2 introduces the referenced protocols used in the simulation, Section 3 describes the metrics for evaluating the superiority of clusters, Section 4 presents the combination of the mentioned metrics using fuzzy logic, Section 5 shows the simulation results in Matlab software, and conclusion is at the end.

II. REFERENCED PROTOCOLS

A. LEACH: Low Energy Adaptive Clustering Hierarchy

LEACH [7] minimizes energy dissipation in sensor networks due to its constructing of clusters. LEACH operation is performed in two phases: a setup phase and a steady state phase. In the setup phase, a sensor node selects a random number between 0 and 1. If this number is less than the threshold *T(n)*, the node becomes a CH. *T(n)* is computed as:

$$T(n) = \begin{cases} \frac{p}{1-p\times\left(r \bmod \frac{1}{p}\right)} & if\ n \in G \\ 0 & otherwise \end{cases} \quad (1)$$

where *r* is the current round; *p,* the desired percentage for becoming CH; and *G* is the collection of nodes not elected as a CH in the last $1/p$ rounds. After being elected, every CH announces to all of the network's sensor nodes that it is the new CH. When each node receives this announcement, it chooses a cluster to join based on the signal strength of the announcement. The sensor nodes then inform their appropriate CH to join them. Afterwards, the CHs, according to a TDMA approach, assign a time slot to each node so that its data can be sent to its CH during this period. During the steady state phase, the sensor nodes can perform sensing and transmit data to the CHs. The CHs also aggregate the data received from the nodes in their cluster before sending these data onto the BS via a single-hop fashion.

B. HEED: Hybrid Energy Efficient Distributed

Younis and Fahmy [8] proposed an iterative clustering protocol, named HEED. HEED is different from LEACH in the way CHs are elected. Both electing the CHs and joining clusters are done based on the hybrid combination of two parameters. The primary parameter depends on the node's residual energy. The alternative parameter is the intra-cluster communication cost. Each node computes a communication cost depending on whether variable power levels, applied for intra-cluster communication, are permissible or not. If the power level is fixed for all of the nodes, then the communication cost can be proportional to (i) node degree, if load distribution between CHs is required, or (ii) 1/node degree, if producing dense clusters is required. The authors define *AMRP* the average of the minimum power levels needed by all *M* nodes within the cluster range to access the CH u, i.e. $AMRP(u) = \frac{\sum_{i=1}^{M} MinPwr(i)}{M}$. If variable power levels are admissible, *AMRP* is used as the cost function. In this approach, every regular node elects the least communication cost CH in order to join it. On the other hand, the CHs send the aggregated data to the BS in a multi-hop fashion.

C. HEED-NPF: HEED protocol using Non-Probablistic approach and Fuzzy logic

This protocol improves HEED protocol using fuzzy logic and a non-probabilistic approach for CH election. The protocol uses node's remaining energy and two fuzzy descriptors to elect CHs. The fuzzy variables used in the fuzzy if-then rule are defined as follows:

- Node degree: number of neighbors a node has
- Node centrality: a value that shows how central the node is in the cluster.

Node centrality is calculated as the average of the squared distances of neighbor nodes from a particular node. Since, typically transmission energy is proportional to the squared distance, the lower value of the node centrality results in a lower amount of energy required by other nodes to send data to that node as a CH. The output of fuzzy logic is the fuzzy cost. With an increase in node degree and decrease of node centrality, the cost of a node being elected as a CH is decreases. A smaller cost means that the node has a higher priority of being elected a CH.

Non probabilistic CH election is implemented by introducing delay inversely proportional to the residual energy of each node. Note that when a node has higher energy, its *delay time* is less than that of nodes with a lower amount of energy. As a result, because its *delay time* has expired sooner, the node has a higher priority of becoming selected as a *tentative_CH*. Simulation results in [9] demonstrate that HEED-NPF approach performs better than LEACH and HEED protocols in terms of network lifetime and also cluster formation.

III. EVALUATION METRICS

Suppose N nodes $\{S_1, S_2, \ldots, S_N\}$ are dispersed in a sensor area and k number of nodes are CHs where $k << N$ and they are assigned to clusters $\{C_1, C_2, \ldots, C_k\}$ respectively and number of nodes joined these clusters are defined as $\{n_1, n_2, \ldots, n_k\}$. Considering the mentioned assumptions, we introduce four metrics to evaluate the quality of clusters formed in the network.

A. Distribution

Distribution is expressed as the average of distances between each two CH nodes. The total average is determined as follows:

$$\text{Distribution} = \frac{\sum_{j=1}^{k}((\sum_{i=1,i\neq j}^{k} \text{distance}(S_j,S_i))/(k-1))}{k} \quad (2)$$

The higher the value of Distribution, the better the CHs distribute in the network. As a result:

- The average distance between each node and its CH becomes less; hence, the nodes consume less energy to send their data to their respective CHs. Therefore, it causes more scalability in number of nodes.
- Fair distribution of CHs in the area has noticeable advantage. Generally sensed data received in CHs have redundancy. Therefore, CHs (through aggregation and fusion) can considerably reduce the size of data before sending these data on to the BS.
- A loss of network resources due to radio wave interferences between CHs will be decreased.
- In protocols with multi-hop routing, the better the CHs distribute in the network, the greater the number of hops. Well distribution of CHs results in reducing the amount of energy consumption due to message exchanges.

B. Centrality

Centrality is defined as the average of the distances of neighbor nodes from a particular CH node. To measure this metric, we should compute the total average for all CHs as follows:

$$\text{Centrality} = \frac{\sum_{j=1}^{k}((\sum_{i\in \text{cluster}_j} L_i)/n_j)}{k} \quad (3)$$

where L_i denotes the distance between node i and its CH. For CH nodes, L_i equals to zero. The lower the value of Centrality, the closer the CH nodes are located to the center of their respective clusters. There are some benefits when CHs locate close to their respective members:

- Member nodes can communicate with their respective CH node by consuming less energy. Since, typically transmission energy is proportional to squared distance, the lower value of Centrality results in a lower amount of energy required by other nodes to send data to that node elected as CH.
- Radio wave interferences between CHs are reduced. Suppose that CH nodes are not located in the center of their respective clusters. It is then possible that some CHs are situated in each other's transmission range and this interference results in a loss of network resources.

C. Cluster Load

To measure this metric, we have the following standard deviation:

$$\text{Cluster Load} = \sqrt{\frac{\sum_{i=1}^{k}(n_i-(N/k))^2}{k}} \quad (4)$$

Note that N/k can be the average number of nodes in each cluster. When the value of cluster load is low it means that nodes are uniformly dispersed in the clusters. Therefore, working-load is fairly distributed among CHs which causes more network lifetime extension.

D. Proximity

Let m be the number of nodes joined the closest CH, the proximity metric is expressed as follows:

$$\text{Proximity} = \frac{m}{N-k} \quad (5)$$

The higher the value of proximity indicates that the more percentage of regular nodes join the closest CH. If the nodes join the closest CH, then:

- They consume less energy to send data to their respective CHs.
- Their wave interferences with the nodes in other clusters reduce considerably.

IV. FUZZY METRIC

Sometimes it is not possible to hold all metrics ideal because they may have tradeoff with each other; an increase in the value of one metric may result in a decrease in the value of the other. Generally, the protocols that considerably meet the metrics consume lower amount of energy and hence lengthen the network lifetime. Also, these protocols have more scalability property than the other protocols.

Fuzzy Logic [10]-[11] is useful for making real-time decisions without needing complete information about the environment. Merging different environmental parameters according to predefined rules and then making a decision based on the result is another important application of fuzzy logic. In this section we introduce the fuzzy system used to combine the four metrics mentioned in the previous section. The model of a fuzzy logic control consists of a fuzzifier, fuzzy rules, a fuzzy inference engine (concluding 81 rules), and a defuzzifier. The most commonly used fuzzy inference technique, called the Mamdani Method [12], has been employed because of its simplicity. To obtain a fuzzy metric, we used four fuzzy sets and the fuzzy if-then rules. The fuzzy variables used in the fuzzy if-then rule are as follows:

- Distribution
- Centrality
- Cluster Load
- Proximity

The higher the value of fuzzy metric means the protocol could form better clusters in the field. Based on the fuzzy variables, we can define fuzzy if-then rules like Table I.

Fuzzy sets of input and output variables are described in Fig. 1 and Fig. 2. After aggregating the results achieved from each rule, a defuzzification method is required to get the crisp value. Defuzzification is performed using the COA method, which returns the Center Of Area under the fuzzy set achieved aggregating conclusions.

TABLE I. Fuzzy Rules

Rule	Distribution	Centrality	Cluster Load	Proximity	cost
1	H	L	L, M, H	L, M, H	VH
	H	H	L, M, H	L, M, H	RH
…					
	M	L	L, M, H	L, M, H	HM
	M	M	L, M, H	L, M, H	M
	M	H	L, M, H	L, M, H	LM
….					
	L	L	L, M, H	L, M, H	RL
81	L	H	L, M, H	L, M, H	VL

Legend: *VLow=very Low, RLow=rather Low, L=Low, LM=medium Low, M=Medium, HM=medium High,RH=rather High,H=High, VH=very High.*

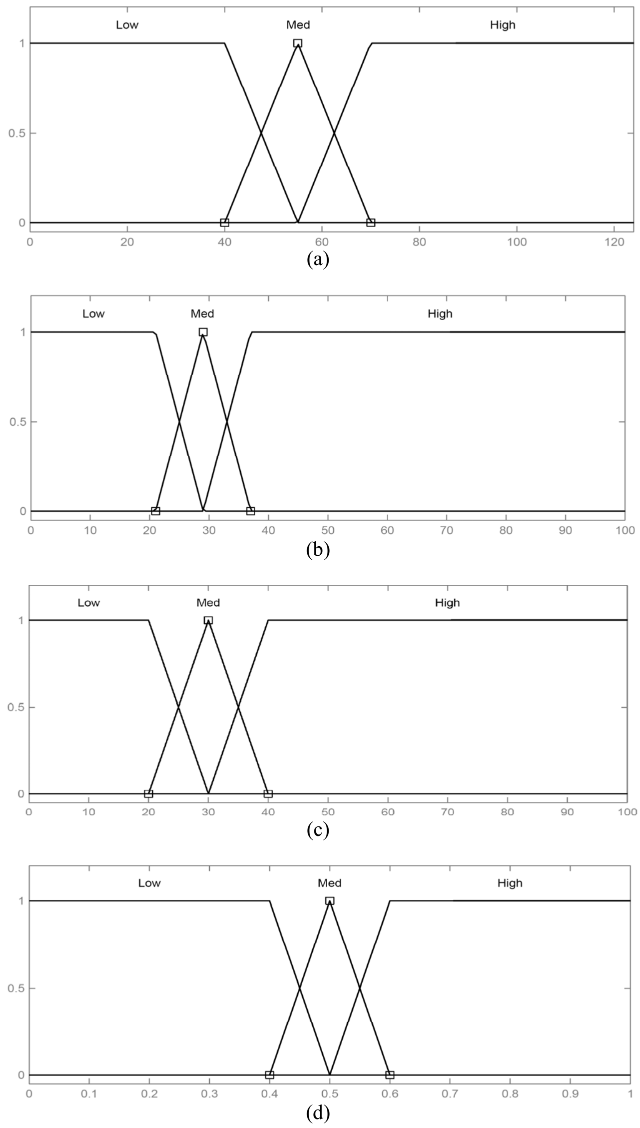
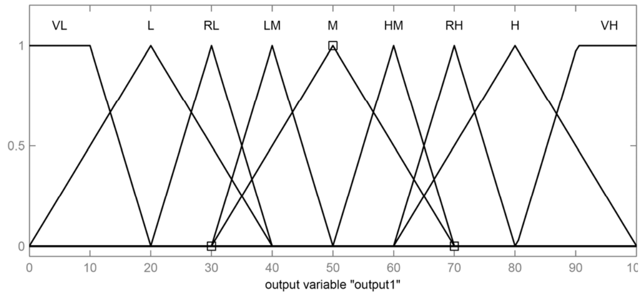


Figure 1. Fuzzy sets of input variables: (a) Distribution, (b) Centrality, (c) Cluster Load, (d) Proximity



(e)

Figure 2. Fuzzy set of output variable: fuzzy metric

V. SIMULATION RESULTS

100 nodes were initially positioned at random locations over 100*100 m^2 area and the BS is located outside the supervised area at the coordinate (50, 175). Each node has the initial energy of 2 Joules. Similar to assumptions and system parameters in [9], we compared LEACH, HEED and HEED-NPF protocols in Matlab from the aspect of evaluation metrics.

At first, we investigate the quality of cluster formation of three protocols with different metrics and also the fuzzy metric. Then, we show two figures indicating the superiority of protocols in terms of network lifetime. Finally, we conclude about the relation between the quality of cluster formation and network lifetime.

In Fig. 3 four metrics (distribution, centrality, cluster load and proximity) are displayed. Both HEED and HEED-NPF protocols are well distributed the CHs in the network (Fig. 3(a)). The reason is that these two protocols use cluster radius for inter and intra-cluster communications while in LEACH protocol, nodes advertise their decision to be elected as CH to all the other nodes in the area. On the other hand, CHs are elected randomly; hence in LEACH protocol it is more possible that two CHs be located in close vicinity. Therefore, compared to HEED-NPF and HEED protocols, LEACH does not guarantee well distribution of CHs in the area. In Fig. 3(b) the value of centrality of LEACH protocol is more than that of the others which means that in this algorithm, CHs are not located in the center of their clusters totally compared to HEED and HEED-NPF protocols. The reason is that all the nodes in the area hear the advertisement of all CHs and each node joins the closest CH. Therefore, the probability that CHs be located in the center of their respective clusters is less than that of HEED and HEED-NPF protocols which apply cluster radius. Fig. 3(c) depicts that the value of cluster load (the standard deviation of number of nodes per cluster) for HEED protocol with 1/node degree and HEED-NPF protocol is totally more than this value for the others. The reason is that their purpose is to form dense clusters. It means that regular nodes do not uniformly join the clusters. But the cluster load value for HEED protocol with node degree is lower than that of the others because its purpose is to uniformly balance working-load among CHs. Fig. 3(d) shows that proximity value of LEACH protocol equals to one because in this algorithm, nodes join the closest CH. HEED-NPF and HEED with AMRP cost act better than HEED with its two other costs. The reason is that they consider the distance factor in computing their costs for CH election.

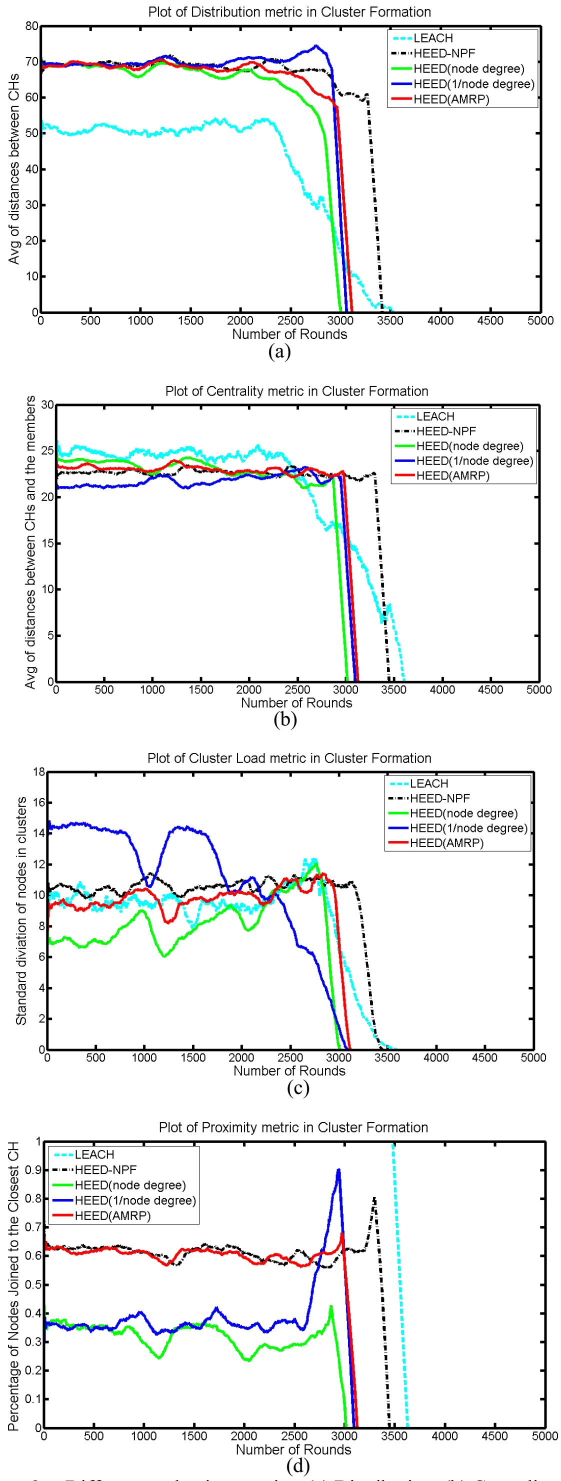


Figure 3. Different evaluation metrics: (a) Distribution, (b) Centrality, (c) Cluster Load, (d) Proximity.

Fig. 4 explains the impact of fuzzy metric on three protocols. It is mentioned before that fuzzy metric is the fuzzy combination of the above four metrics. This figure demonstrates that HEED-NPF performs better in terms of cluster superiority as the value of fuzzy metric is more than this value for the other two protocols. The values for other costs in HEED protocol were similar, so they are removed.

We claimed that if a protocol well clusters the network, it surely prolongs network lifetime. Therefore, according to the fuzzy metric, HEED-NPF should extend network lifetime more than HEED and it does similar compared to LEACH. In the continuance of this study's simulation, the correction of our claim is validated through figures 5 and 6.

Fig. 5(a) shows the total number of nodes that remain alive over the simulation round for three protocols. As it is shown in this figure, HEED-NPF protocol increases network lifetime more than HEED and LEACH protocols.

To better compare network lifetime for the mentioned communication costs, we can summarize the time the first node dies (FND), only one half of the nodes which are alive (HNA) and the last node dies (LND) in the Fig. 5(b). However, LND is not a very useful metric, because after half of the sensor nodes die, the wireless sensor network becomes almost useless in most of the cases. Therefore, we mostly pay attention to the metrics FND and HNA in order to evaluate our simulation results. In this figure it is obvious that HEED-NPF is better in important metrics for evaluating network lifetime than other two protocols.

Fig. 6 illustrates the superiority of clusters formed through the mentioned protocols. Note that, in HEED protocol, the sensor nodes near the BS die earlier than other nodes in the field. This is because these nodes tolerate a tremendous load while forwarding data from other nodes to the BS in a multi-hop fashion. However in the LEACH protocol, the nodes far away from the BS die earlier. The reason for this is that CH nodes send their data directly to the BS via a single-hop fashion. Therefore, a node located further away from the BS consumes more energy. As it is shown in this figure, although some nodes are dead in HEED and LEACH protocols, in HEED-NPF all of the nodes are alive which demonstrates the longer lifetime of this protocol in terms of FND.

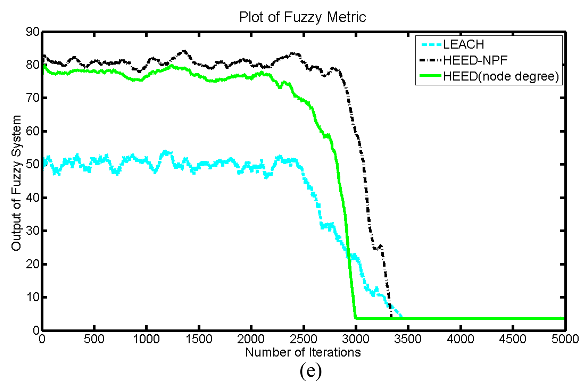
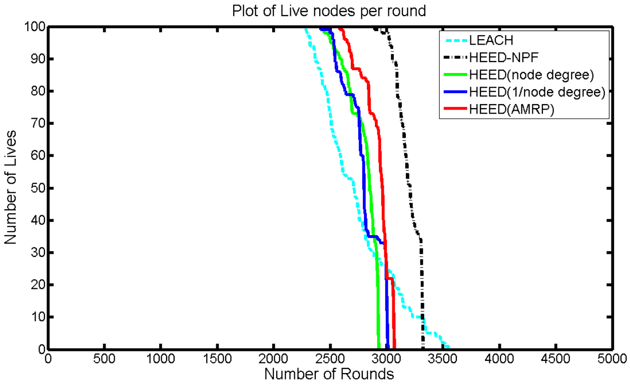
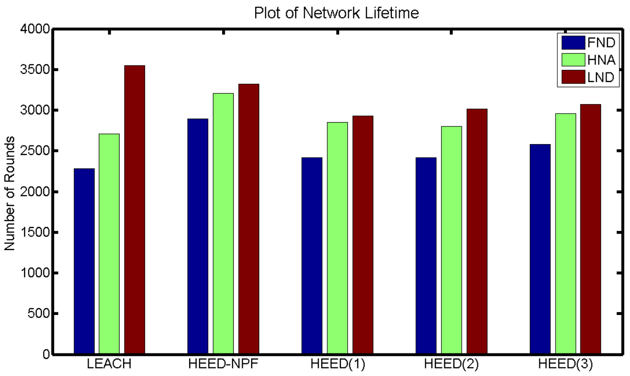


Figure 4. Fuzzy metric



(a)



(b)

Figure 5. Comparison of three protocols in terms of (a) live nodes, (b) different definitions of network lifetime.

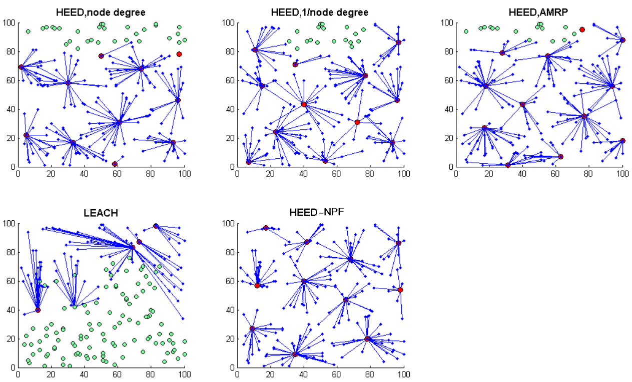


Figure 6. Clusters formed in three protocols in round number 2700. Nodes with bright color are dead nodes.

Consequently, by means of the simulation results we validated the correction of our fuzzy metric. As a result, it is a good measure to decide about the superiority of protocols in terms of cluster formation; the better a protocol forms the clusters, the more it extends network lifetime.

VI. CONCLUSION

In wireless sensor network applications, effective clustering algorithm can reduce energy consumption which can increase network scalability and lifetime.

In this paper, we proposed multi metrics to evaluate the superiority of cluster formation in cluster based WSNs from different aspect of view. Then we combine these metrics which are distribution, centrality, cluster load and proximity by using fuzzy logic to present a fuzzy metric. By means of this fuzzy metric we can conclude about the superiority of different protocols in terms of cluster formation. Regarding to the importance of the quality of cluster formation, if a protocol acts better in terms of fuzzy metric compared to the other protocols, it can more prolong the network lifetime.

REFERENCES

[1] M. W. Park, J. Y. Choi, Y. J. Han, T. M. Chung, "An Energy Efficient Concenteric Clustering Scheme in Wierless Sensor Networks," *Fifth International joint Conference on INC, IMS and IDC,* 2009.

[2] J. M. Kim, S. H. Park, Y. J. Han, T. M. Chung, "CHEF: Cluster Head Election mechanism using Fuzzy logic in Wireless Sensor Networks," *ICACT'08*, 2008.

[3] W. Zhenhua, H. Xiaodong, Z. Hong, L. Chang'an, "Reasearch on Clustering Starategy for Wireless Sensor Network Based on Fuzzy Theory," *LNCS 4864*, pp. 596-604, 2007.

[4] K. Akkaya, M. Younis, "A survey on routing protocols for wireless sensor networks," *Elsevier Journal of Ad Hoc Networks*, vol. 3, no. 3, 325–349, 2005.

[5] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.

[6] O. Younis and S. Fahamy, "Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach," in *Proc. IEEE INFOCOM*, Mar. 2004, vol. 1, pp. 629–640.

[7] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," *Proc. of the Hawaii International Conference on System Sciences*, pp. 1–10, January 2000.

[8] O. Younis, S. Fahmy, "HEED: A Hybrid, Energy-Efficient, Distributed clustering approach for Ad Hoc sensor networks," *IEEE Transactions on Mobile Computing,* vol. 3, no. 4, pp. 366–379, 2004.

[9] H. Taheri, P. Neamatollahi, "Improving on HEED Protocol of Wireless Sensor Networks using Non Probabilistic Approach and Fuzzy Logic (HEED-NPF)," *IST'10*, Dec. 2010.

[10] Zadeh, L. A. 1965. Fuzzy Sets. Information and Control. 8, 3 (Jun. 1965), 338-353.

[11] Zadeh, L. A. 1973. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. IEEE Trans. On Systems, Man and Cybernetics, SMC-3, 1 (Jan. 1073), 28-44.

[12] Mamdani, E.H. and Assilian, S. 1975. An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies, 7, 1, 1-13.