

A Density-based Energy-efficient Clustering Algorithm for Wireless Sensor Networks

Zhanyang Xu^{1,2}, Yue Yin² and Jin Wang²

¹*Institute of Information Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*

²*School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China*

Abstract

Clustering is an efficient method adopted in various routing algorithms for wireless sensor networks. In this paper, we propose a Density-based Energy-efficient Clustering Algorithm (DECA). In DECA, we define the density of each node and regard it as an important evaluation metric. Together with nodes' residual energy under consideration, each cluster head is selected based on the density of nodes. We design an intra-cluster algorithm as well as a multi-hop inter-cluster routing algorithm. Moreover, we discuss the optimal number of clusters. Simulation results show that cluster heads are evenly distributed and our proposed routing algorithm do consume much less energy than some existed algorithms. The network lifetime is also largely prolonged.

Keywords: *wireless sensor networks; clustering; density; multi-hop*

1. Introduction

Wireless sensor networks (WSNs) [1] are composed of hundreds or thousands of sensors that work cooperatively to monitor the environmental conditions of the sensor field. Sensor nodes collect sensed data and pass them to the base station. WSNs have various applications in many fields such as military, agriculture and health care etc.

Study on efficient routing algorithms is an important and challenging research issue. The essence of routing algorithms is to find an optimal path that enables the efficient exchange of information between source node and base station, and to ensure correct transmission of data along the path. As the battery, capability of computing, storage and data processing of a sensor are limited, how to reduce the energy consumption while prolonging the network lifetime stays the key problem.

Most routing algorithms are based on two categories of network structure: planner and hierarchical. For planner routing algorithms, as all nodes have equal roles, traffic is evenly distributed across the network. However, the network lacks scalability, and both data transmission and the route discovery and maintenance procedure cost much resource. For hierarchical routing algorithms, clustering [2] methodology is adopted which divides network into clusters and makes cluster heads responsible for data aggregation. It has following advantages:

- 1) Instead of probable long-distance data transmission to the base station, sensors only need to send data to cluster heads. Then cluster-heads eliminate redundancy of received data and send the aggregated information to the base station. Thus the amount of data communication is reduced;

2) Clustering topology is conducive to the application of distributed algorithms, which is especially suitable for large-scale deployed network;

3) Since most of the nodes close their communication module for relative long time, it can significantly prolong the lifetime of the entire network.

In this paper, we propose a Density-based Energy-efficient Clustering Algorithm (DECA) for WSNs. DECA ensures the even distribution of cluster heads due to the evaluation of density. Moreover, the residual energy of each cluster head is under consideration after each round of cluster head selection. Such improvements can alleviate the energy hole problem. An intra-cluster algorithm as well as a multi-hop inter-cluster routing algorithm is designed. Both save energy to some extent.

The rest of the paper is organized as follows. Section 2 introduces some related work of clustering algorithms. In Section 3 we first present relevant network and energy model. Then we show the method of cluster head selection in our DECA, discuss about the optimal number of clusters, and describe the details of its inter-cluster and intra-cluster routing algorithms. Performance evaluation is given in Section 4 and Section 5 concludes this paper.

2. Related Work

Low Energy Adaptive Clustering Hierarchy (LEACH) [3] is a classical clustering algorithm. In a periodical way, it randomly chooses the cluster heads. Each node has a probability of P performed as a determined prior to become one cluster head, as is shown in Eq. (1).

$$T(n) = \begin{cases} \frac{P}{1-P[r \bmod (1/P)]} & \text{if } n \in G, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $T(n)$ represents the threshold value of node n , r is the current round index, and G is a set of unselected cluster heads of the nodes in the previous rounds. Nodes are evaluated on whether or not to be cluster heads first and the un-chosen nodes join to the nearest clusters.

In LEACH, the energy consumption of entire network is evenly distributed to each sensor node, which aims to reduce energy consumption and improve the network lifetime. The algorithm is simple, however, it has some deficiencies: First, it does not guarantee about even distribution of cluster heads over the network. Some very big clusters and very small clusters may exist in the network at the same time. Second, cluster head selection is unreasonable in heterogeneous networks where nodes have different energy. Third, in this protocol it is assumed that each cluster head transmits data to base station over a single hop, which may consume much energy.

Various other clustering algorithms have been proposed. In Ref. [4], each node communicates only with a close neighbor and takes turns transmitting to the base station. Ref. [5] elects cluster heads based on the average minimum reachability power. Ref. [6] considers the tradeoff of the energy expenditure between nodes to cluster heads and cluster heads to base station. Global knowledge of distance is required though. Such algorithm suffers from the energy hole problem, where energy consumption of sensors near the base station or on the critical paths is much faster than other nodes.

Ref. [7] follows LEACH to choose cluster heads using randomization, but clustering considers both energy and local distance. In Ref. [8] is applied as a node-weight heuristic algorithm with node's residual energy, number of nodes in the neighbor

partition and relative location under consideration. Ref. [9] determines suitable cluster sizes depending on the hop distance to the data sink to compensate for the requirement of high energy consumption. Similarly, Ref. [10] proposes a cluster allocation and routing algorithm based on node density and study on optimal density proportion for deploying sensor nodes. In this way, network lifetime can be prolonged. In such algorithms, however, too many clusters around the base station will produce a significant number of summary packets which results in heavy traffic load.

Appropriate cluster-head election is an essential consideration and nodes' location and connectivity have been primarily focused. Ref. [11] uses fuzzy logic technique considering two factors: neighbor nodes and remaining energy. Cluster heads elected in Ref. [12] are determined to have minimum composite distance of sensors to cluster head and cluster head to base station. In Ref. [13], the cluster-head selection depends on remaining energy level of sensor nodes for transmission. Ref. [14] provides the first trajectory based clustering technique for selecting the cluster heads and meanwhile extenuate the energy hole problem. Ref. [15] forms a cluster network with required coverage and connectivity and it avoids collisions and overhearing of data packets.

Density-based Clustering Protocol (DBCP) [16] is an improvement for LEACH on the basis of nodes' connectivity. A metric of nodes' relative density is introduced for cluster-head selection, as it is shown in Eq. (2).

$$T(n) = \begin{cases} \frac{P}{1-P[r \bmod (1/P)]} \left(1 + \frac{\text{Neighbor}(n)\text{-alive} - 1/P}{\text{Network_alive}}\right) & \text{if } n \in G, \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $1/P$ represents the average number of nodes in one cluster. By comparing to the alive neighboring nodes of n in certain round, the formula promotes that nodes in dense area have larger probability to become cluster head.

3. Our proposed DEGRA Algorithm

3.1. Relevant Models

We assume that the network is composed of N sensor nodes, denoted as: $\{s_1, s_2, \dots, s_N\}$ respectively. They are uniformly dispersed within a $M \times M$ square region. The nodes always have data to transmit to a base station, denoted as BS , which is often far from the sensing area. They continuously monitor the surrounding environment. We make the following assumptions:

- 1) All nodes are homogeneous and stationary after deployment.
- 2) Nodes can adjust their transmission power according to the relative distance to receiver
- 3) Links are symmetric. A node can compute the approximate distance to another node based on the received signal strength, once the transmitting power is given.

Figure 1 shows the scenario of a uniform dispersion of 100 sensor nodes in a $100 \times 100 m^2$ square region. Without loss of generality, here we assume that the base station is located at the coordination of $(-100, -100)$.

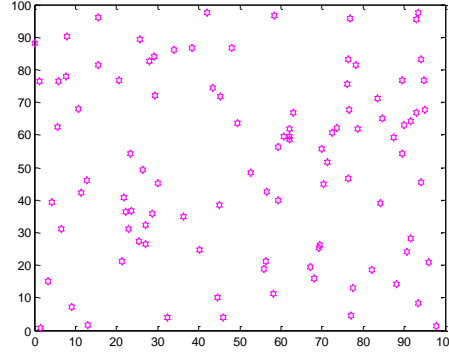


Figure 1. Network Model

We use similar energy model as Ref. [17]. Each sensor node will consume the following E_{Tx} amount of energy to transmit a l -bits packet over distance d , where the E_{elec} is the energy dissipated per bit to run the transmitter or receiver circuit, ε_{fs} and ε_{mp} represent the transmitter amplifier's efficiency and channel conditions:

$$E_{Tx}(l, d) = \begin{cases} lE_{elec} + l\varepsilon_{fs}d^2, & d < d_o \\ lE_{elec} + l\varepsilon_{mp}d^4, & d \geq d_o \end{cases} \quad (3)$$

To receive a packet, radio consumes energy

$$E_{Rx}(l) = lE_{elec} \quad (4)$$

Cluster heads aggregate n 1-bits packets received from its members into a single 1-bits fixed packet. The energy consumption is calculated as, where E_{DA} is the data aggregation cost of a bit per signal:

$$E_{aggregation}(n, l) = nlE_{DA} \quad (5)$$

3.2. Cluster Head Selection

We assume N sensor nodes in a $M \times M$ square region are divided into k clusters, with R representing the standard transmission radius for message exchange during the set-up stage of clusters. Thus we have:

$$M \times M = k\pi R^2 \Rightarrow R = \frac{M}{\sqrt{\pi k}} \quad (6)$$

We select cluster heads according to the density (denoted as den) of each node. Here, the metric of density represents the number of nodes located within a circle region for searching that takes the node itself as the center and R as the radius. The density of s_i can be calculated via searching the entire network as Eq. (7), where $d(s_i, s_x)$ represents the distance between s_i and another node s_x .

$$den(s_i) = \sum_{0 < x \leq N \ \& \ d(s_i, s_x) \leq R} 1 \quad (7)$$

Therefore, the selection procedure performs in k rounds. In each round, we sort all density values of possible nodes in descending order. Then choose the first one, namely the one with the largest den as the cluster head. If multiple nodes have the same den , we choose one randomly. If all nodes are included in one of the clusters, the selection ends.

However, we notice that as neighboring nodes of the determined cluster head often have similar density value which is large enough for disturbing selections in following rounds, we should exclude all nodes located in the previous searching circles. In this way, cluster heads are more evenly distributed than LEACH due to the consideration of density. Moreover, each cluster head tends to perform data fusion for more neighboring nodes, which saves energy.

As cluster heads consumes much more energy through data aggregation than normal nodes, after multiple rounds of DECA execution some cluster heads may drain out their residual energy and become invalid. It will largely reduce the network lifetime. Therefore, to solve such energy hole problem, we mention a metric η as the energy level to decide whether existing cluster head roles should be changed.

$$\eta = \frac{E_{residual}}{E_{initial}} \quad (8)$$

From the above formula, we can see η represents the proportion of the residual energy of certain node $E_{residual}$ in its initial energy $E_{initial}$. In our algorithm, we pre-determine a threshold number $\eta_{threshold}$ such as 10%. Once $\eta \leq \eta_{threshold}$, the cluster head is no longer responsible for data aggregation, instead it becomes a normal node. Another cluster head should be selected to compensate for the lost role with the metric density as the top priority.

3.3. Optimal Cluster Number

For certain network, if the cluster number is too small, many sensor nodes have to send data to the base station, which consumes much energy; On the contrary, if the cluster number is too large, thus clustering becomes unnecessary. Therefore, we try to find a relative optimal cluster number k . We assume the maximum distance of any inside node to its cluster head is relative small, thus the calculation of member nodes' energy consumption follows the free space channel model. With d_{toCH} representing the distance between the member node and its cluster head, the energy consumption is equal to:

$$E_{member} = lE_{elec} + l\epsilon_{fs}d_{toCH}^2 \quad (9)$$

Assuming that the nodes are uniformly distributed, with $\rho(x, y)$ represent the node distribution, it can be shown that:

$$E[d_{toCH}^2] = \iint (x^2 + y^2)\rho(x, y)dxdy = \frac{M^2}{2\pi k} \quad (10)$$

Energy dissipated in the cluster head during a round includes the energy consumption of data reception, aggregation and transmission. It is given by Eq. (11) where d_{toBS} represents its distance to the base station, and as the base station is often far from the sensing area, we assume the formula follows multi-path fading model.

$$E_{CH} = \left(\frac{N}{k} - 1\right)lE_{elec} + lE_{DA} + (lE_{elec} + l\epsilon_{mp}d_{toBS}^4) \quad (11)$$

The total energy consumption of the entire network is:

$$E_{total} = k(E_{CH} + \left(\frac{N}{k} - 1\right)E_{member}) \approx E_{CH} + \frac{N}{k}E_{member} = l(2NE_{elec} + \frac{M^2}{2k\pi}N\epsilon_{fs} + E_{DA}N + k\epsilon_{mp}d_{toBS}^4) \quad (12)$$

By differentiating E_{total} respect to k and equating to zero, the relative optimal number of constructed clusters k_{opt} can be found as:

$$\frac{dE_{total}(k)}{dk} = l(\epsilon_{mp}d_{toBS}^4 - \epsilon_{fs} \frac{M^2}{2\pi k^2}N) \Rightarrow k_{opt} = \frac{M}{d_{toBS}^2} \sqrt{\frac{N\epsilon_{fs}}{2\pi\epsilon_{mp}}} \quad (13)$$

Eq. (6) shows that the message transmission radius R is related to the number of clusters k . By integrating Eq. (6) and Eq. (13), we can find certain value of a standard transmission radius R that can be adjusted for the optimal number k :

$$R = \frac{d_{toBS}^2 \sqrt{2\epsilon_{mp}}}{\sqrt{N\epsilon_{fs}}} \quad (14)$$

3.4. Routing Procedure

In our algorithm, each sensor node sends data to its cluster head directly within one hop. The corresponding cluster head should be determined by the distance between the node and the cluster head according to Eq. (3) in the energy model.

We compare the distances from the node to each cluster head and choose the shortest one for communication. In this way, the node will find the optimal cluster head with the least energy consumption. The intra-routing algorithm can be formulated as to find:

$$\text{Min}_k(d(s_x, CH_k)), \quad x=1,2,\dots,N \ \& \ s_x \neq CH_k \quad (15)$$

In LEACH, cluster heads send data to the base station directly within one hop. There is high chance that it consumes large energy due to the remote location of some cluster head. In our DECA, we perform inter-routing in a multi-hop way.

Suppose cluster head CH_i chooses another CH_j as its relay node and let CH_j communicate directly with the base station BS . In order to deliver a l -length packet to BS via CH_j , the energy consumed of CH_i is calculated as Eq. (16) where ϵ and α vary in different situations according to the energy model.

$$E_{CH_i} = E_{Tx}(l, d(CH_i, CH_j)) + E_{Rx}(l) + E_{Tx}(l, d(CH_j, BS)) = 3lE_{elec} + \epsilon d^\alpha(CH_i, CH_j) + \epsilon d^\alpha(CH_j, BS) \quad (16)$$

For each cluster head CH_i , we choose an optimal relay cluster head which maintains the least energy consumption E_{CH_i} . We compare it with the direct communication cost to BS , and determine the optimal inter-routing according to the smaller energy dissipation.

4. Performance Evaluation

4.1. Simulation Environment

We evaluate the performance of the DECA via simulations in Matlab. The simulation environment is set up with the parameters listed in Table 1.

Table 1. Network Parameters

Parameter Name	Value
Network Scale	$100 \times 100 m^2$
Number of the sensor nodes	100
Length of the packet (l)	4000bits
Initial energy of the sensor nodes ($E_{initial}$)	0.25J
Energy consumption on circuit (E_{elec})	50nJ/bit
Channel parameter in free-space model (ϵ_{fs})	$10 pJ / bit / m^2$
Channel parameter in multi-path model (ϵ_{mp})	$0.0013 pJ / bit / m^4$
Channel parameter for data aggregation (ϵ_{DA})	5pJ/bit/signal
Pre-determined energy level of cluster heads ($\eta_{threshold}$)	10%

4.2. Simulation Results

Figure 2 shows the distribution of 5 cluster heads in LEACH. As it adopts randomization in the selection procedure, there is high chance that some cluster heads locate relatively close to each other. Thus it results in heavy traffic load for remote nodes to transmit data to any cluster head. Differently, with nodes' density, namely connectivity under consideration in our DECA, cluster heads are distributed much more evenly as is shown in Figure 3

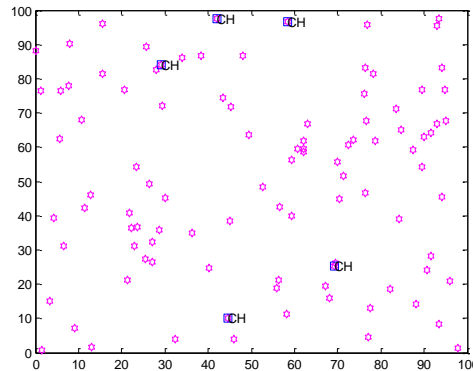


Figure 2. Location of Cluster Heads in LEACH

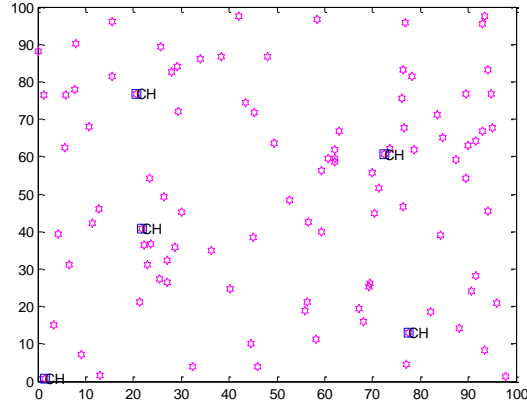


Figure 3. Location of Cluster Heads in DECA

With proper message transmission radius R , we have assumed that a relative optimal cluster head number k can be found to ensure good cluster coverage. Figure 4 illustrates the energy consumption of our DECA with different numbers of clusters in various network scales. For example, “100*100” represents a $100 \times 100m^2$ wireless sensor network). As the inter-cluster routing is performed in a multi-hop way, we can see that total energy consumption decreases while the number of clusters increases. However, the decreasing rate of the energy consumption becomes relatively small after certain clusters are formed under all circumstances. It is because at that time mostly nodes are included in one of the existing clusters with relative small transmission distance to the cluster head. Forming more clusters can hardly further reduce energy consumption. In Figure 4, we can ensure five clusters do have good performance.

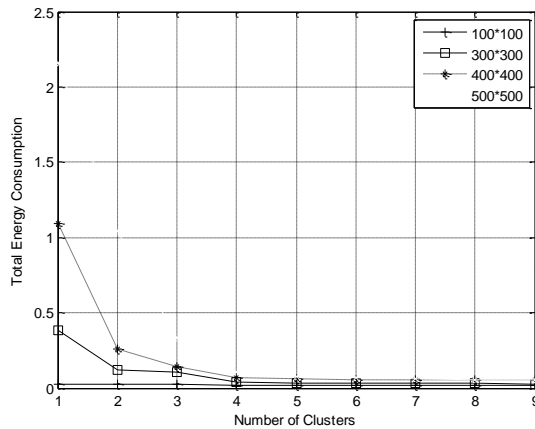


Figure 4. Energy Consumed in DECA with Various Cluster Numbers

We compare the total energy consumption of LEACH, DBCP and our DECA, as is shown in Figure 5 where the network is set as $100 \times 100m^2$. During 20 rounds, the energy consumption of LEACH and DBCP are quite similar. In comparison, DECA remains much better performance with less energy consumption than both LEACH and DBCP. This is because the clusters are evenly distributed and the multi-hop inter-cluster routing also saves much energy.

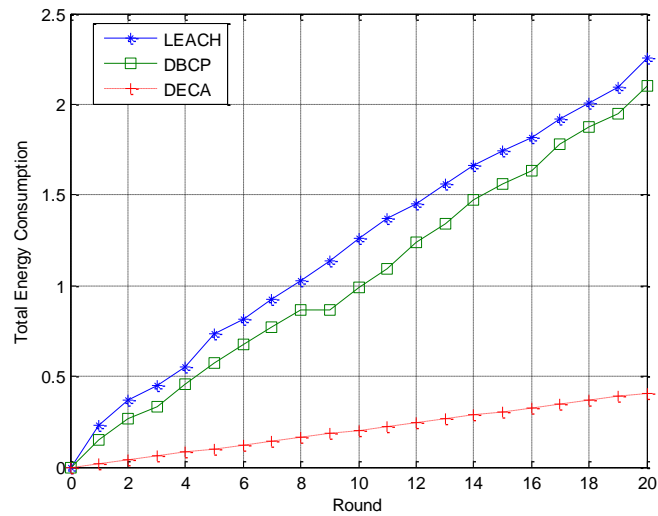


Figure 5. Comparison of Energy Consumption

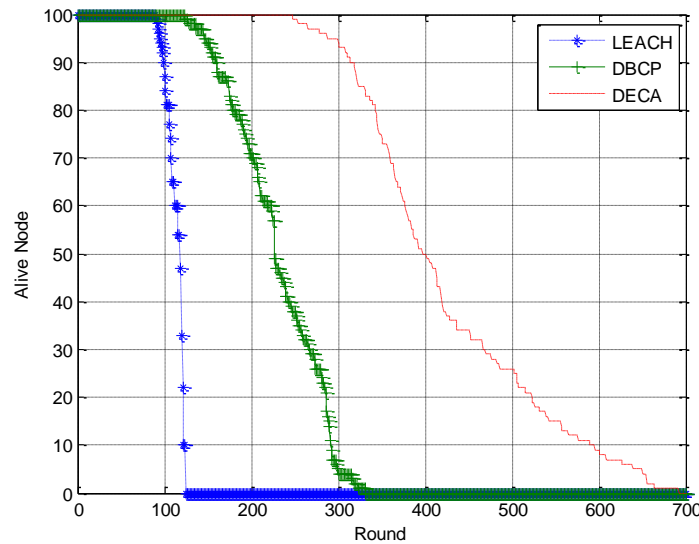


Figure 6. Comparison of Network Lifetime

Moreover, we compare the network lifetime of LEACH, DBCP and our DECA, as is shown in Figure 6 where the network is set as $100 \times 100 m^2$ and there are 10 clusters. For LEACH, the first node that becomes invalid appears in 94th round; DBCP has the first inactive node in 124th round; DECA shows the best performance as the first node is found in 248th round, which is twice larger than the DBCP situation. It is due to the changes of cluster head roles considering its residual energy proportion as well as the energy-efficiency via multi-hop routing technique. Both improvements prolong the network lifetime.

5. Conclusions

For wireless sensor networks, clustering is one of the most popular routing methodologies that can effectively manage network energy consumption via data aggregation. We propose a Density-based Energy-efficient Clustering Algorithm (DECA) for WSNs. In our algorithm, cluster heads are elected according to the density, so that clusters are evenly distributed. An intra-cluster routing algorithm and a multi-hop inter-cluster routing algorithm are proposed. An optimal number of clusters can be found. Simulations show that the energy consumption in DECA is largely reduced and the network lifetime becomes much longer with comparison to some existed algorithms.

Acknowledgements

This work was supported by Jiangsu Province Universities Natural Science Research Program (NO.11KJB510010) and Jiangsu Province Research and Innovation Project for College Graduates (NO.CXZZ12_0515). This work was also supported by the Industrial Strategic Technology Development Program (10041740) funded by the Ministry of Knowledge Economy (MKE) Korea, and by the Natural Science Foundation of Jiangsu Province (No. BK2012461).

References

- [1] I. Akyildiz and W. Su, Y. Sankarasubramaniam and E. Cayirci, "A Survey on Sensor Networks", *IEEE Commun. Mag.*, vol. 8, (2002), pp. 102.
- [2] T. J. Kwon and M. Gerla, "Clustering With Power Control", *Proceedings of IEEE Military Communications Conference*, (1999) October 31–November 3; Atlantic City, New Jersey.
- [3] W. Heinzelman and A. Chandrakasan and H. Balakrishnan, "Energy-efficient Routing Protocols for Wireless Microsensor Networks", *Proceedings of the 33rd Annual Hawaii International Conference of System Sciences*, (2000) January 4-7; Maui, Hawaii.
- [4] S. Lindsey and C. S. Raghavendra, (Eds.), "PEGASIS: Power-efficient Gathering in Sensor Information Systems", *Parallel and Distributed Systems*, vol. 9, (2002), pp. 924.
- [5] O. Younis and S. Fahmy, (Eds.), "HEED: A Hybrid, Energy-efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks", *Mobile Comput.*, vol. 3, (2004), pp. 366.
- [6] M. Ye and C. Li, G. Chen and J. Wu, "EECS: An Energy Efficient Cluster Scheme in Wireless Sensor Networks", *Proceedings of IEEE International Workshop on Strategies for Energy Efficiency in Ad Hoc and Sensor Networks*, (2005) April 7–9, Phoenix, Arizona.
- [7] J. M. Kim and S. H. Park and Y. J. Han, "CHEF: Cluster Head Election Mechanism Using Fuzzy Logic in Wireless Sensor Networks", *Advanced Communication Technology*, vol. 1, (2008), pp. 654.
- [8] Y. Wang and T. L. X. Yang, D. Zhang, "An Energy Efficient and Balance Hierarchical Unequal Clustering Algorithm for Large Scale Sensor Networks", *Information Technology Journal*, vol. 8, (2009), pp. 28.
- [9] D. Wei and Y. C. Jin and S. Vural, "An Energy-efficient Clustering Solution for Wireless Sensor Networks", *IEEE transactions on wireless communications*, vol. 10, (2011), pp. 3973.
- [10] B. S. Lee and H. W. Lin and W. Tarng, "A Cluster Allocation and Routing Algorithm Based on Node Density for Extending the Lifetime of Wireless Sensor Networks", *Proceedings of the 26th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, (2012) March 26-29; Fukuoka, Japan.
- [11] Y. Hu and X. R. Shen, Z. H. Kang, "Energy-efficient Cluster Head Selection in Clustering Routing for Wireless Sensor Networks", *Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing* (2009) September 24-26; Beijing, China.
- [12] X. X. Zhang and M. Zhang, Z. C. Zhang, "An Improved WSNs Clustering Routing Algorithm and Its Performance", *Sciencepaper Online*, vol. 2, (2010), pp. 96.

- [13] M. C. M. Thein and T. Thein, "An Energy Efficient Cluster-head Selection for Wireless Sensor Networks", Proceedings of International Conference on Intelligent Systems, Modelling and Simulation (ISMS), (2010) January 27-29; Liverpool, United Kingdom.
- [14] H. Munaga and J. V. R. Murthy and N. B. Venkates-warlu, "A Novel Trajectory Clustering Technique for Selecting Cluster Heads in Wireless Sensor Networks", International Journal of Recent Trends in Engineering, vol. 1, no. 1, (2009).
- [15] N. Xu, A. P. Huang, T. W. Hou, H. H. Chen, "Coverage and Connectivity Guaranteed Topology Control Algorithm for Cluster-based Wireless Sensor Networks", Wireless Communications and Mobile Computing, vol. 12, (2012), pp. 23.
- [16] J. F. Qiao and S. Y. Liu, X. Y. Cao, "Density-based Clustering Protocol for Wireless Sensor Networks", Computer Science, vol. 12, (2009), pp. 46.
- [17] T. Rappaport, "Wireless Communications: Principles & Practice", Englewood Cliffs, Prentice-Hall, (1996).

Authors



Zhanyang Xu obtained his Masters in Computer Engineering from the School of Computer Science and Technology, China University of Mining and Technology(CUMT) in 2004. He graduated as the top student in the Nanjing University of Posts and Telecommunications(NJUPT) where would obtain his PhD in 2012. At the same time, he is serving as a full time faculty in the School of Computer & Software, Nanjing University of Information Science & Technology(NUIST). His research interest includes wireless communication network, wireless sensor networks and Internet of things.



Yue Yin received the Bachelor degree in Applied Computing from Nanjing University of Information and Science Technology in 2011. She is currently pursuing a Master degree in Technology of Computer Application in the former institution. Her research interests include wireless sensor networks and Internet of things, specifically in fields of algorithm improvement and data replication.



Jin Wang Dr. Jin Wang received the B.S. and M.S. degree in the Electronical Engineering from Nanjing University of Posts and Telecommunications, China in 2002 and 2005, respectively. He received Ph.D. degree in the Ubiquitous Computing laboratory from the Computer Engineering Department of Kyung Hee University Korea in 2010. Now, he is a professor in the Computer and Software Institute, Nanjing University of Information Science and technology. His research interests mainly include routing protocol and algorithm design, performance evaluation and optimization for wireless ad hoc and sensor networks. He is a member of the IEEE and ACM.

