An Energy-efficient Clustering Algorithm in Wireless Sensor Networks with Multiple Sinks

Zhanyang Xu[1,2], Yue Yin[1], Jin Wang[1] and Jeong-Uk Kim[3]

[1]*School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China*

[2]*Institute of Information Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*

[3]*Department of Energy Grid, Sangmyung University, Seoul 110-743, Korea*

{zyxu, yinyue, wangjin}@nuist.edu.cn; jukim@smu.ac.kr

Abstract

In wireless sensor networks with single sink node, the energy consumption of sensors near the sink or on the critical paths is too fast besides other disadvantages. Therefore, we propose an Energy-efficient Multi-sink Clustering Algorithm (EMCA) for wireless sensor networks. In EMCA, residual energy plays a significant role in the procedure of selecting cluster heads. We also design an inter-cluster and intra-cluster routing algorithm aiming at minimizing the energy consumption. Inside a cluster, a multi-hop way is under consideration for a node to relay data instead of direct transmission. Simulation results show that our proposed algorithm consumes much less energy and owns longer network lifetime than the traditional routing algorithm LEACH.

Keywords: *wireless sensor networks; multi-sink; clustering; muti-hop*

1. Introduction

Wireless sensor networks (WSNs) [1] are wireless networks that consist of hundreds or thousands of randomly deployed static sensors. In a multi-hop and self-organized way, the sensors can monitor the physical environment such as temperature, sound and movement etc. They cooperatively sense, process and transmit data to provide people with direct, real and effective information. The physical world, information field and human society are therefore integrated to some extent.

WSNs are featured with large-scale deployment, dynamic topology, self-organization etc., which makes them data-centric and application-oriented. They can monitor wide areas while maintaining high precision. They are becoming increasingly useful in various critical applications, such as military, environmental monitoring, agricultural technology, industrial manufacturing and medical care.

Wireless transmission is one of the most important functions of the WSN. The relevant topology control and energy management algorithms are important and challenging research issues. The study of energy-efficient routing protocols or algorithms is one of the critical problems. The essence of the routing protocols is to find an optimal path that enables the efficient exchange of information between the source node and the sink node, and to ensure correct transmission of data along the path. Routing protocols can be divided into four categories: energy-based, query-based, geography-based and QoS-based routing protocols.

In simple routing scenarios, single-sink topology is often chosen. However, as the sensor nodes suffers from small communication range, limited battery, weak capability of computing, storage and data processing, etc., the single-sink infrastructure usually has some disadvantages: First, once the only sink node fails, the collected data can no longer be transmitted to the user, which will inevitably lead to the failure of the entire network. Second, the routing algorithms mainly focus on the shortest path or path with minimum hop number. The selected path may include nodes with less residual energy and large energy consumption for data transmission. Thus, the maximum network lifetime cannot be guaranteed. Third, for the nodes that are closer to the sink, they have more data to relay so that their energy consumption rate is much higher than the remote ones. They are under risk of becoming invalid. Such energy hole phenomenon will lead to the imbalance of energy consumption, which may seriously affect the lifetime of the entire network.

Compared with single-sink topology, multi-sink infrastructure allows sensor nodes to choose different sink nodes according to their own situation. Therefore, it can reduce the average transmission distance between sensor node and sink nodes. The deployment of multi-sink nodes has the advantages of dispersion, stability and robustness. The overall energy consumption is balanced. It can effectively solve the energy hole[2] problem and extend the network lifetime.

Clustering is an efficient routing method, where the entire network is divided into multiple clusters. Each cluster has one cluster head and it is responsible for data aggregation. Instead of direct communication with the sink, all the member nodes in one cluster send data to the cluster head. In this way, the traffic load can be reduced.

In this paper, we propose an Energy-efficient Multi-sink Clustering Algorithm (EMCA) for WSNs. We divide the sensing field into several equal clusters. Each cluster head collects data and sends it to one of the multiple sinks according to its own preference. We also propose an inter-cluster and intra-cluster routing algorithm. Therefore, it not only saves energy through clustering, but ensures that workload is dispersed so that the phenomenon of unbalanced energy consumption around one single sink can be alleviated. The network lifetime can also be prolonged.

The rest of the paper is organized as follows. Section 2 introduces some related work of clustering algorithms and multi-sink applications. In Section 3 we first present relevant network and energy model. Then we show the clustering-related method of our EMCA and describe the details of its inter-cluster and intra-cluster routing protocols. Simulation evaluation and performance comparison are given in Section 4 and Section 5 concludes this paper.

2. Related Work

Many energy-efficient routing protocols have been proposed based on the clustering structure. In clustering topology, clusters are formed so that cluster head can fuse data from ordinary sensors in the same cluster. Therefore, energy can be saved and longer network lifetime can be achieved.

LEACH [3] is a classical clustering algorithm. In a periodical way, it randomly chooses the cluster heads. PEGASIS [4] is an improvement over LEACH. It's a chain-based protocol. Each node communicates only with a close neighbor and takes turns transmitting to the sink. HEED [5] also improve LEACH. Cluster heads are decided based on the average minimum reachability power. In TTDD [6], a grid structure is maintained. It provides scalable and efficient data delivery.

To solve the energy hole problem, some scholars choose the unequal clustering algorithms. For the clusters, the closer they are to the sink, the smaller size they are formed. Therefore, it saves energy for the inter-cluster communications. MRPSUC [7] is a novel energy-aware clustering protocol. It not only preserves energy by clustering, but takes various measures (such as considering the residual energy) to balance the energy of each node. PEBECS [8] divides the entire WSN into several equal partitions reasonably and groups the nodes into clusters of unequal sizes. A node-weight heuristic algorithm is applied, considering the node's residual energy, the number of nodes in the neighbor partition and the relative location. More balanced load can be ensured. EC [9] also determines suitable cluster sizes depending on the hop distance to the data sink. It achieves approximate equalization of node lifetimes and reduced energy consumption levels. However, too many clusters around the sink will produce a significant number of summary packets. In result, it will cause heavy traffic load.

The traditional single-sink infrastructure has several disadvantages, such as sudden invalidation of the sink node. It will affect the lifetime of the entire network. To solve the problem, the research of multi-sink topology is deployed.

MRMS [10] is a multi-sink routing mechanism, which includes three parts: topology discovery, clustering maintenance and path reselection. Simulations show that it effectively extends the network lifetime, but the iteration can cause large overheads. In HLBR [11], both of the balances of inter-sink and locally inter-sensor are considered. It's based on the geographical location of each node. It successfully solves the problem that the energy of sensor nodes in specific area depletes much faster than the others. DispersiveCast [12] is another scheme. It studies data transmission between just two sinks. Sensor nodes choose sink randomly until any sink detects more data than the threshold, then start to send data to both sinks with certain probabilities. Voronoi scoping [13] controls the query dissemination sent by the sink. Instead of overall flooding, each query is sent within the Voronoi cluster, which minimizes the traffic overhead. In this method, nodes will not be affected by the scale of the network, the number of the sinks or the location. In PBR [14], a priority is calculated. This method considers both the energy level and the energy cost of the routing path, so the energy consumption is more balanced and the network lifetime is more prolonged. However, the reasonable deployment of the sinks is not mentioned. Similar to the PBR, Z. B. Wu etal. [15] proposed an energy level-based routing algorithm. it chooses the path with the highest energy level to send data to the sink. Traditional TopDisc [16] is applied to form clusters, so the network lacks flexibility to some extent.

3. Our Proposed EMCA Algorithm

3.1. Relavant Model

3.1.1. Network Model: We assume that the network is composed of N sensor nodes, denoted as: $\{S_1, S_2, ..., S_N\}$ respectively. They are uniformly dispersed within a rectangle field and continuously monitor their surrounding environment. There are k sink nodes (or Base Station, BS) have been deployed with random location, represented as $\{BS_1, BS_2, ..., BS_k\}$, as is shown in Figure 1. We make the following assumptions:

（1）All nodes are homogeneous and stationary after deployment.

（2）The multiple sink nodes are pre-located within the sensing field randomly.

（3）Nodes can adjust their transmission power according to the relative distance to its receiver.
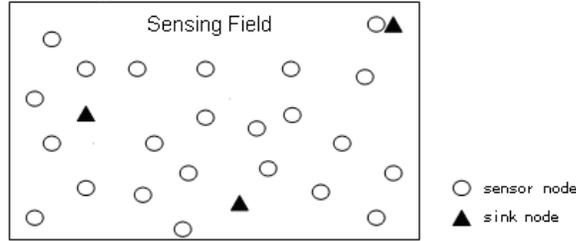
（4）Links are symmetric.

Figure 1. Network Model

3.1.2. Energy Model: We use similar energy model in [17]. Based on the distance between transmitter and receiver, a free space (d^2 power loss) or multi-path fading (d^4 power loss) channel models are used.

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:

$$E_{Tx}(l,d) = \begin{cases} lE_{elec} + l\varepsilon_{fs}d^2, d < d_o \\ lE_{elec} + l\varepsilon_{mp}d^4, d \geq d_o \end{cases} \tag{1}$$

To receive a packet, radio consumes energy:

$$E_{Rx}(l) = lE_{elec} \tag{2}$$

3.2. Cluster Formation and Cluster Head Selection

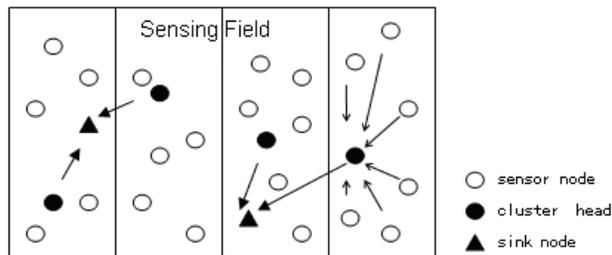We divide the entire network into *n* equal clusters, as is shown in Figure 2.

Figure 2. Cluster Formation

Each cluster has only one cluster head. The other nodes in the same cluster send data to the cluster head. Then cluster head send aggregated data to its relevant sink. The clustering method proposed in our EMCA has various advantages. First and foremost, data aggregation reduces traffic load. Second, the cluster heads locate in a more uniform way comparing to the probabilistic deployment in LEACH. It is more suitable for the large-scale deployed networks. Last but not the least, it can prolong the network lifetime, as a majority of nodes close the communication module for relatively long time.

After equally dividing the sensing field into several equal areas, we will next choose each cluster head. As the network is considered to be heterogeneous, we determine each cluster head by its residual energy.

When the selection begins, we first motivate the sensor node that is located in the center of each cluster like S_i. It is regarded as the cluster head candidate. It broadcasts one message within a neighborhood of radius **R**. This message aims to motivate other nodes for the competition of the cluster head. It contains the node's id and its residual energy. Only the nodes within the transmission range can receive the message and become active, while the outside nodes remain idle. If any node S_j has larger residual energy than S_i, it becomes the new cluster head candidate and broadcasts new message with its own information to the others. If S_j has equal residual energy with S_i, compare the ID. The node with a smaller ID wins. If S_j has smaller residual energy than S_i, it still broadcasts the message of S_i. As soon as the comparison is done, the un-chosen node becomes idle again. All nodes in the cluster should be compared only once. In this way, the node with the largest residual energy is chosen as the cluster head.

The cluster-selection algorithm can be formulated as to find $Max(E_{residual})$

3.3. Routing Procedure

3.3.1. Inter-cluster Routing: After data fusion, the cluster heads should send aggregated data to sink nodes. We can make good use of the multi-sink topology. In our algorithm, each cluster head selects one optimal sink respectively. The minimization of energy consumption is our top concern.

For any cluster head CH_n, the energy consumption to sink BS_k is represented as $E(CH_n, BS_k)$. Its calculation follows the energy model.

$$E(CH_n,BS_k) = \begin{cases} lE_{elec}+l\varepsilon_{fs}d(CH_n,BS_k)^2, d(CH_n,BS_k)<d_o \\ lE_{elec}+l\varepsilon_{mp}d(CH_n,BS_k)^4, d(CH_n,BS_k)\geq d_o \end{cases} \tag{3}$$

Since the cluster heads send data to the sink nodes directly within one hop, from the formula we can see that the smaller $d(CH_n, BS_k)$ is, the smaller $E(CH_n, BS_k)$ becomes. Therefore, we only need to compare the distances from each cluster head to different sinks and choose the shortest one. In this way, the cluster head will find the optimal sink with the least energy consumption.

The inter-cluster algorithm can be formulated as to find $\underset{k}{Min}(d(CH_n,BS_k))$.

3.3.2. Intra-cluster Routing: Moreover, in many clustering algorithms such as LEACH, sensor nodes in the same cluster send data directly to the cluster head. Due to the fact of their various locations, some sensor nodes may consume relative large amount of energy due to long-distance transmission. Therefore we set a multi-hop routing protocol.

For any member node S_i in one cluster, the energy consumption it costs to send data directly to its cluster head CH_{S_i} is represented as $E_1(S_i, CH_{S_i})$.

$$E_1(S_i, CH_{S_i}) = \begin{cases} lE_{elec} + l\varepsilon_{fs} d(S_i, CH_{S_i})^2, d(S_i, CH_{S_i}) < d_o \\ lE_{elec} + l\varepsilon_{mp} d(S_i, CH_{S_i})^4, d(S_i, CH_{S_i}) \geq d_o \end{cases} \qquad (4)$$

In the mean time, it is possible that S_i tries to find another sensor node S_j to relay data to save energy by avoiding directly communication with CH_{S_i}, as it is shown in Figure 3.
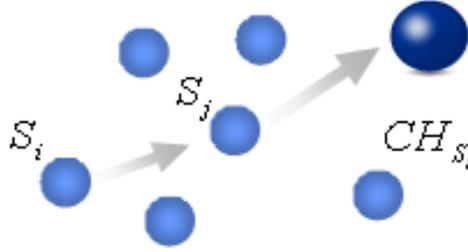

Figure 3. Possible relaying scenario

To deliver a *l*-length packet to the cluster head, the energy consumption $E_2(S_i, S_j, CH_{S_i})$ is calculated as formula (5) and the optimal relay node is determined based on the smallest value of $E_2(S_i, S_j, CH_{S_i})$ where ε and α vary in different situations according to the energy model.

$$\begin{aligned} E_2(S_i, S_j, CH_{S_i}) &= E_{Tx}(l, d(S_i, S_j)) + E_{Rx}(l) + E_{Tx}(l, d(S_j, CH_{S_i})) \\ &= l(E_{elec} + \varepsilon d^\alpha(S_i, S_j)) + lE_{elec} + l(E_{elec} + \varepsilon d^\alpha(S_j, CH_{S_i})) \qquad (5) \\ &= 3lE_{elec} + \varepsilon d^\alpha(S_i, S_j) + \varepsilon d^\alpha(S_j, CH_{S_i}) \end{aligned}$$

Each S_i chooses S_j with the smallest value of $E_2(S_i, S_j, CH_{S_i})$ as the relay node if necessary.

$$E_2(S_i, CH_{S_i}) = Min(E_2(S_i, S_j, CH_{S_i})) \qquad (6)$$

Compare formula 4 and formula 6, and the smaller one is chosen.

$$E(S_i, CH_{S_i})$$
$$= Min(E_1(S_i, CH_{S_i}), E_2(S_i, CH_{S_i})) \quad\quad (7)$$
$$= Min((lE_{elec} + \varepsilon d^\alpha(S_i, CH_{S_i})), Min(3lE_{elec} + \varepsilon d^\alpha(S_i, S_j) + \varepsilon d^\alpha(S_j, CH_{S_i})))$$

In our algorithm, however, the sink nodes are randomly located. Therefore, some nodes may consume less energy through sending data directly to the sink rather than to its cluster head. So it is necessary to compare $E(S_i, CH_{S_i})$ and $E(S_i, BS_k)$ and decide the final route.

For simplicity, the intra-cluster algorithm can be formulated as to find $Min(E(S_i, CH_{S_i}), E(S_i, BS_k))$

4. Performance Evaluation

4.1. Simulation Environment

We evaluate the performance of the EMCA via simulations in Matlab. It is assumed that all the senor nodes and the sink nodes are uniformly deployed in a square sensing area. For example, Figure 4 shows the scenario of a uniform dispersion of 100 sensor nodes in a $500*500\,m^2$ square region.
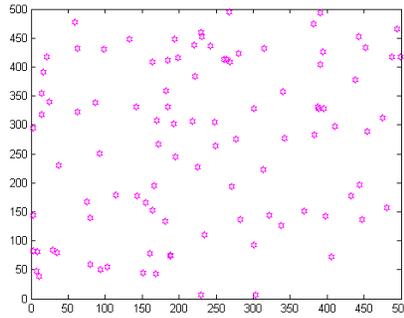

Figure 4. Example of a 500*500 m^2 Network

Some other parameters for setting up the simulation environment are listed in Table 1.

Table 1. Network Parameters

Parameter Name	Value
Number of the sensor nodes (N)	100
Length of the packet (l)	6bit
Initial energy of the sensor nodes (E_{init})	0-2J
Energy consumption on circuit (E_{elec})	50nJ/bit
Channel parameter in free-space model (ε_{fs})	10pJ/bit/ m^2
Channel parameter in multi-path model (ε_{mp})	0.0013pJ/bit/ m^4

4.2. Simulation Results

Figure 5 illustrates the energy consumption of our EMCA with different numbers of the sink nodes in a 500*500 m^2 network of n equal clusters (n =3,4,5,6). From Figure 5, we can see that the total energy consumption decreases while the number of the sinks increases under all circumstances.
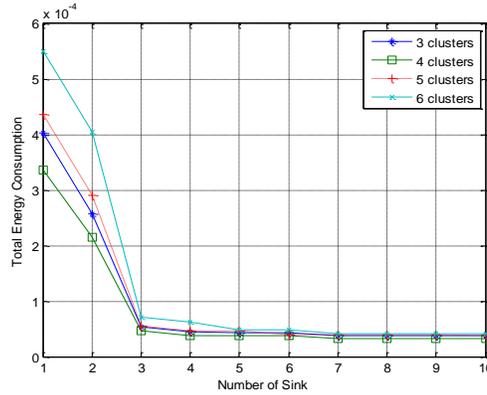

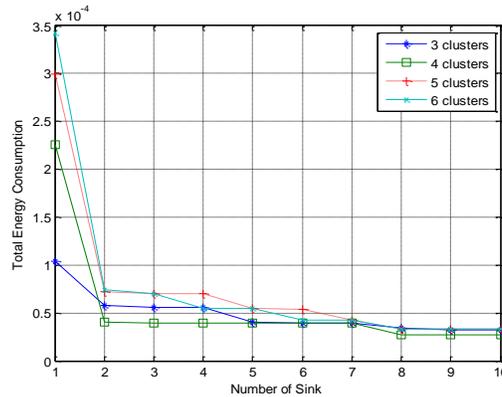**Figure 5. Total Energy Consumption in a 500*500 m^2 Network**

Figure 6. Total Energy Consumption in a 400*400 m^2 Network

Moreover, from Figure 5 we can conclude that once three sinks have been deployed in this 500*500 m^2 area, the decreasing rate of the energy consumption becomes relatively small even if more sink nodes are added later. We can regard the later deployment unnecessary to some extent. Similarly, Figure 6 shows the case with the 400*400 m^2 network, where two sinks performs well enough for the energy consumption. Without loss of generality, we can conclude that the optimal number of the sink nodes can be found in certain scale of the network.

Even though more sinks can save more energy, the number of the sinks should be limited as one sink node cost C_{sink} much more than that of one sensor node C_{node} . The total cost is in proportion with the number of the sinks. Since the energy consumption $\boldsymbol{E}$

is in contrast with the number of the sinks, for certain scale of the network, where there are N sensor nodes and k sink nodes, we can take a value of $E*C$, where C is defined as $1 + \dfrac{C_{\sin k}}{N * C_{node}} k$, deduced by formula 8, as an evaluation metric for determining the optimal k.

$$\frac{1}{C} = \frac{N * C_{node}}{N * C_{node} + k * C_{\sin k}} \quad => \quad 1 + \frac{C_{\sin k}}{N * C_{node}} k \tag{8}$$

For achieving energy-efficiency of our algorithm, we aim to minimizing E while maximizing $1/C$, namely the minimization of $E*C$ in general. Figure 7 displays thw evaluation of the factor $E*C$. It shows that in a $500*500\, m^2$ wireless network, 3-8 sinks provide relatively small value of $E*C$, which is preferable. In Figure 7, the metrics "1:5", "1:10", "1:20", "1:50" represent the cost rate of one sensor node and one sink $\dfrac{C_{node}}{C_{\sin k}}$. For example, "1:10" means one normal node cost 10 times larger than the cost of a sink. Moreover, as the cost rate becomes larger, 3 turns out to be the optimal number of the sink.
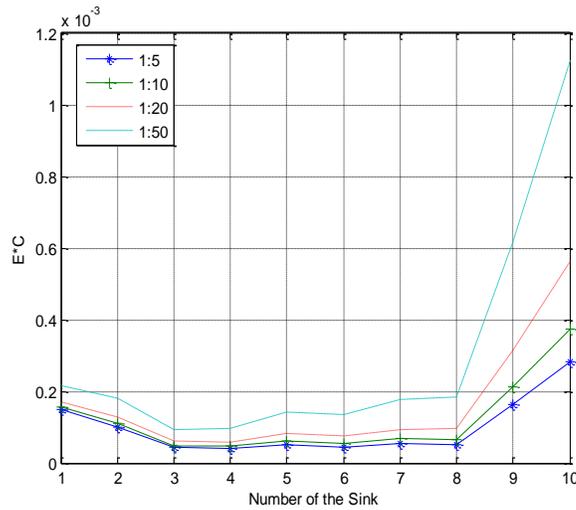

Figure 7. Optimal Number of the Sink

We compare the total energy consumption of our EMCA and LEACH algorithm, as is shown in Figure 8. The network is set as $500*500\, m^2$. In 20 rounds, EMCA remains much better performance with less energy consumption than LEACH algorithm. This is mainly because of the clustering method that implements data fusion to reduce the transmission cost along the path. Multi-hop also saves energy inside each cluster
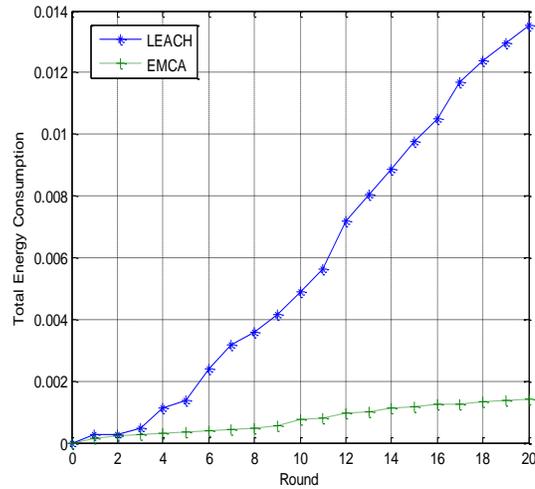
Figure 8. Total Energy Consumption in of EMCA Compared with LEACH

Moreover, we compare the network lifetime of our EMCA and LEACH algorithm under certain same scenario where the network is set as $500*500\,m^2$ and equipped with 3 sinks. The result is shown in Figure 9. For LEACH, the first node that becomes invalid appears in 390th round, while EMCA has the first inactive node in 503rd round. It is due to the changes of cluster head roles considering nodes' residual energy, as well as the main focus on minimization of energy consumption in EMCA that efficiently prolongs the network lifetime.
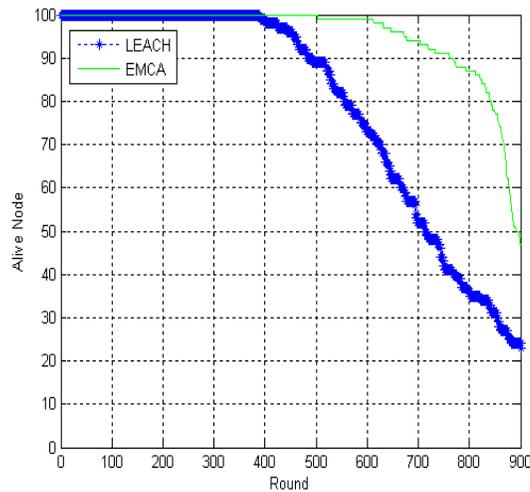

Figure 9. Network Lifetime of EMCA Compared with LEACH

5. Conclusions

The multi-sink deployment helps solve the energy hole problem. We propose an Energy-efficient Multi-sink Clustering Algorithm (EMCA) for WSNs. The inter-cluster and intra-cluster routing algorithm is explained in details. Moreover, we deduce the deployment of the sink nodes with an optimal multiple sink number through experiments. Simulations show that the energy consumption is largely reduced than LEACH algorithm.

Acknowledgements

References

[1] I. Akyildiz and W. Su, Y. Sankarasubramaniam and E. Cayirci, Editors, IEEE Commun. Mag., vol. 8, **(2002)**, pp. 102.
[2] C. Xu and L. Cao, G. A. Zhang and J. Y. Gu, Editors, Application Research of Computers, vol. 3, **(2010)**, pp. 816.
[3] W. Heinzelman and A. Chandrakasan and H. Balakrishnan, Editors, "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, Editors, Parallel and Distributed Systems, vol. 9, **(2002)**, pp. 924.
[5] O. Younis and S. Fahmy, Editors, Mobile Comput., vol. 3, **(2004)**, pp. 366.
[6] F. Ye and H. Luo, J. Cheng, S. Lu, L. Zhang, Editors, "A Two-tier Data Dissemination Model for Large-scale Wireless Sensor Networks", Proceedings of the 8th Annual International Conference on Mobile Computing and Networking, **(2002)** September 23-28; Atlanta, Georgia, USA.
[7] B. Gong and L. Li, S. Wang, X. Zhou, Editors, "Multihop Routing Protocol with Unequal Clustering for Wireless Sensor Networks", Proceedings of ISECS International Colloquium on Computing, Communication, Control, and Management, **(2008)** August 3-5; Guangdong, China.
[8] Y. Wang and T. L. X. Y. D. Zhang, Editors, Information Technology, vol. 1, **(2009)**, pp. 28.
[9] D. Wei, Y. C. Jin, S. Vural, K. Moessner and R. Tafazolli, Editors, Wireless Communications, vol. 10, **(2011)**, pp. 3973.
[10] Y. Q. Chen and C. S. Han, Editors, "Energy Efficient Multipath Routing in Large Scale Sensor Networks with Multiple Sink Nodes", Proceedings of the 6th International Workshop on Advanced Parallel Processing Techniques, **(2005)** October 27-28; Hong Kong, China.
[11] F. R. Li and H. Bao and H. Q. Huang, Editors, Computer Research and Development, vol. 47, **(2010)**, pp. 139.
[12] T. Suzuki, M. Bandai and T. Watanabe, Editors, "DispersiveCast: Dispersive Packets Transmission to Multiple Sinks for Energy Saving in Sensor Networks", Proceedings of the 17th Annual IEEE International Symposium on Personal,Indoor and Mobile Radio Communications, **(2006)** September 11-14; Helsinki, Finland.
[13] H. D. Ferriere, D. Estrin and T. Stathopoulos, Editors, "Efficient and Practical Query Scoping in Sensor Networks", Proceedings of IEEE International Conference on Mobile Ad hoc and Sensor Systems, **(2004)** October 25-27; Pittsburgh, USA.
[14] M. Meng, H. Xu and X. L. Wu, Editors, "PBR: Priority Based Routing in Multi-sink Sensor Networks", Proceedings of the International Conference on Wireless Networks, **(2007)** June 25-28; Las Vegas, USA.
[15] Z. B. Wu, X. B. Fan and H. Chen, Editors, Journal of Computer Research and Development, vol. 1, **(2008)**, pp. 1282.
[16] B. Deb, S. Bhatnagar and B. Nath, Technical Report of Rutgers University DCS-TR-441, **(2001)**.
[17] T. Rappaport, Editor, "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 routing algorithms and data replication of wireless sensor networks.

Jin Wang is a professor in the Computer and Software Institute, Nanjing University of Information Science and technology. He 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. 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.

Jeong-Uk Kim received his B.S. degree in Control and Instrumentation Engineering from Seoul National University in 1987, M.S. and Ph.D. degrees **in** Electrical Engineering from Korea Advanced Institute of Science and Technology in 1989, and 1993, respectively. He is a professor in SangMyung University in Seoul. His research interests include smart grid demand response, building automation system, and renewable energy.