An Energy-Efficient Clustering Algorithm for Large Scale Wireless Sensor Networks

Maryam Soleimani, Amirali Sharifian, Ali Fanian
Department of Electrical and Computer Engineering,
Isfahan University of Technology, Iran
Maryam.soleimani@ec.iut.ac.ir, a.sharifian@ec.iut.ac.ir, a.fanian@cc.iut.ac.ir

***Abstract*— Wireless sensor networks (WSNs) consist of a large number of sensor nodes with limited energy resources. Collecting and transmitting sensed information in an efficient way is one of the challenges in these networks. The clustering algorithm is a solution to reduce energy consumption. It can be helpful to the scalability and network life time. However, the problem of unbalanced energy dissipation is an important issue in cluster based WSNs. In this paper, a new clustering algorithm, named PDKC, is proposed for wireless sensor networks based on node deployment knowledge. However, in PDKC, sensor node location is modelled by Gaussian probability distribution function instead of using GPSs or any other location-aware devices. In the proposed method, cluster heads are selected based on node deployment information, residual energy, node degree and their distance from the base station. The Simulation results indicate that PDKC algorithm prolongs network lifetime, improves the network coverage and balance energy dissipation in comparison to other works.**

Keywords*— *Wireless sensor networks, clustering, energy efficient, coverage, data aggregation.

I. INTRODUCTION

Recent developments in wireless communications have provided low cost and small size sensor nodes [1]. A wireless sensor network consists of thousands of limited resource sensor nodes, used to collect information from the surrounding environment. These sensor nodes are able to communicate with each other and also with base station (BS) [2]. In other words, each sensor node can sense, process and transfer data to the BS.

Due to some constraints on the size and the cost of sensor nodes, they usually are equipped with limited and non-rechargeable batteries; therefore, reducing the amount of energy dissipation is an important challenge in the network lifetime. To reduce energy consumption, unnecessary information should not be sent to the BS. It has been proved that clustering is an effective scheme in reducing energy consumption and increasing the scalability of the WSNs. In this method the network is divided into some separate clusters. Each cluster has a cluster head (CH) and each sensor node belongs to a cluster to deliver the sensed data to its cluster head. Finally, each CH aggregates the received data and sends it to the BS. Since some sensor nodes are located around a special event similar data would be sent to the CH, aggregation mechanisms in the CH can reduce the amount of transmitted data to the BS. Hence, the energy consumption reduces significantly in the sensor nodes. Furthermore, [3] network coverage is another challenge especially in large scale WSNs such as military applications [4]. In such networks, it is important to cover the whole network area while decreasing the overlap between clusters [5]. In other words, decreasing the overlapping areas has a significant influence on the reduction of the energy dissipation.

In the clustering algorithms, there are two kinds of communications, intra-cluster and inter-cluster communications. In intra-cluster, since the distance between CH and sensors are too short, the commutation is established by one or at most two hops. However, in inter-cluster communication the distance between CHs and BS may grow extremely. In previous researches [6], it has been shown that multi-hop communication between a data source and a data sink is usually more energy efficient than direct transmission. The main problem of this approach is that the sensors closer to the BS suffer from heavy relay traffic. Therefore, they die much sooner than other sensors, thus sense coverage of the network decreases.

In the current paper, a clustering algorithm, namely an efficient Power consumption based on Deployment Knowledge for Cluster based WSNs (PDKC), is proposed in which both network coverage and energy consumption are considered as main factors. PDKC algorithm is based on deployment knowledge that improves the sensor node energy consumption and the network coverage simultaneously. The proposed algorithm, however, benefits from sensor nodes deployment knowledge; sensor nodes do not require any GPS or other location aware devices. Additionally, they need not to be synchronized with each other; accordingly, their expenses and energy consumption are decreased [7]. In the proposed algorithm, selecting a CH is done according to some parameters including approximate sensor node position, residual energy, and sensor node degrees. In addition, in PDKC, the transmission range of each CH varies according to the distance between the CH and BS. Energy consumption, in this case, in CHs is almost balanced in the network.

978-1-4673-5634-3/13/$31.00 ©2013 IEEE

The rest of the paper is organized as follows. Section 2 summarizes previous and related works. Section 3 outlines the assumptions for the energy consumption model and sensor node distribution model. The proposed method, PDKC, is described in Section 4. Section 5 discusses the effectiveness of PDKC algorithm via simulations and compares PDKC to other clustering techniques. Finally, the conclusion is presented in the last section.

II. RELATED WORKS

The first proposed clustering algorithm for WSN was LEACH [8]. The clustering operation in this method is done in two phases. In the first phase, named setup, some sensor nodes are randomly chosen as CHs in each round. In the latter, named steady-state, data is transmitted from nodes to the CH and CH sends the aggregated data to the BS. PEGASIS [9] is an extension of LEACH, in which all nodes are grouped into a chain and each node sends its data to the nearest neighbor. Only one sensor node is selected from that chain to transmit the aggregated data to the base station. Compared with LEACH, this algorithm decreases energy consumption of nodes significantly, but the data delay is much extensive. Therefore, the use of this protocol in large scale networks seems impossible [8]. Some other algorithms try to balance energy consumption of the nodes by using equal clusters in the network. In [10], the network area is divided into a set of equal squares and then a GPS-free scheme is used to get nodes deployment information. Nodes belonging to a same cell make a cluster and a new CH is selected in each round.

USC [11] was the first algorithm proposed in order to balance energy consumption of the sensor nodes by forming unequal clusters. However, in the latter research, it was proved that this algorithm seems impractical to be used in the WSNs. Because the CHs are performed by some super nodes all the time and are deployed at pre-determined locations [12]. A clustering algorithm based on the highest degree is proposed in [13]. In this protocol, the node that has the most neighbors would be selected as CH. The presented results indicate that the rate of CH changes and the throughput is low. Another clustering algorithm, namely EEUC, is proposed in [6]. In this algorithm, the network is divided into some unequal clusters based on sensor nodes residual energy and their distance from the BS. The main disadvantage of this algorithm is that the algorithm efficiency depends on the primary distribution of nodes and CHs. Moreover, in EEUC, distance from the BS is considered as a major factor regardless of sensor nodes density. This issue would cause some difficulty in network coverage.

III. PROPOSED DEPLOYMENT-BASED CLUSTERING ALGORITHM

As mentioned in the previous section, clustering is one of the important methods for reducing power consumption in sensor nodes sending data to the BS. Different methods for cluster-based WSNs have been proposed. Some of these methods make use of additional equipment, such as GPS or other location aware mechanisms. Using such equipment would increase the cost and the energy consumption in the sensor nodes. Besides, in all other proposed algorithms, the network coverage performance is low and they are considered to cover a small area, yet in most of applications the network area is so small [2]. This paper proposes a clustering algorithm, namely PDKC, for large scale network. While, in PDKC, the clustering algorithm requires sensor node deployment information, the sensor node doesn't need to have any additional equipment such as GPS. Firstly, in order to explain the proposed algorithm, the network model is presented, and then the clustering algorithm is discussed.

A. Network Model

In order to receive regional information, in most of applications, many sensor nodes are distributed in a wide area [6, 7]. In WSNs sensor nodes are usually distributed by an airplane. If the distribution of the sensor nodes is done uniformly, sensor nodes can be located in each area with the same probability. If so, there isn't any deployment knowledge available about the sensor node resident point. In the literature, it has been proved that having deployment knowledge invigorates algorithm performance. In the proposed protocol, since no location aware device is used in sensor nodes, deployment knowledge is modelled using a Gaussian probability distribution function (pdf) [14]. On the other hand, former proposed clustering algorithms use a uniform pdf.

In the proposed scheme, the network area is divided into non-overlapping hexagonal cells, and the sensor nodes are allocated in groups into these cells. The centre of a cell is defined as the deployment point of the sensor nodes allocated to that specific cell. Fig. 1, shows the division of the network into hexagonal cells. The distance between adjacent cell deployment points is L. Each cell in the proposed scheme has a pair of credentials (i, j) which are the cell position. Using two-dimensional Cartesian coordinates and assuming that the deployment point of cell Ci, j is (xi, yi); [14] the pdf of the sensor node resident points can be formulated as:

$$f_k^{i,j}(x,y|k \in C_{i,j}) = f(x-x_i, y-y_i) = \frac{1}{2\pi\sigma^2} e^{-[(x-x_i)^2+(y-y_i)^2]/2\sigma^2} \quad (1)$$

$$f(x,y) = \frac{1}{2\pi\sigma^2} e^{-[x^2+y^2]/2\sigma^2} \quad (2)$$

Assuming identical pdfs for every group of sensor nodes $f_k(x, y \mid k \in C(i,j))$ can be used instead of $f_k^{ij}(x, y | k \in C(i,j))$.

As in [14-16], in the proposed scheme it is assumed that the routing protocol delivers the transmitted data to the correct destinations.

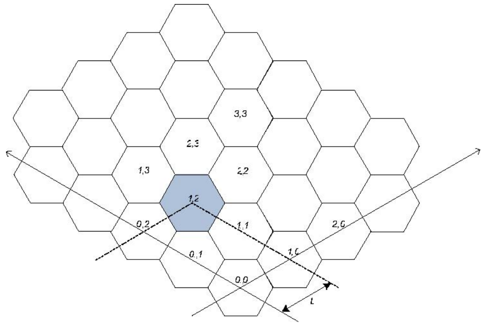

Fig. 1. A two-dimensional WSN with adjacent deployment point distance L.

With a Gaussian distribution, the distance between the resident and deployment points is less than 3σ with the probability of 0.9987, where σ is the standard deviation [15]. In addition, if the distance between deployment points in two cells is more than 6σ, there is little chance that two sensor nodes, each belonging to one of these cells, be adjacent [15], [16].

As mentioned, the distribution of sensor nodes in each cell is modelled by a Gaussian probability distribution around the deployment point. Although using Gaussian distribution would allow having information about the resident point of the sensor nodes, yet it is preferable to have uniform distribution over the whole network.

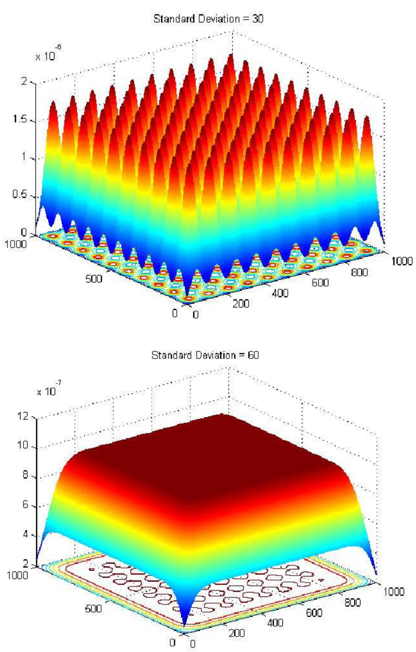

Fig. 2. Deployment distribution of the sensor nodes in the network for different values of σ.

In other words, in the proposed algorithm not only is the deployment knowledge considered, but also the uniform density of the sensor nodes in the network area becomes very important. However, by choosing an appropriate amount for standard deviation in Gaussian distribution this goal becomes achievable.

In the Fig. 2 the distribution of sensor nodes are illustrated for different amounts of standard deviation. As depicted in this figure, for standard deviation of 60, the overall distribution is very close to the uniform distribution.

B. Energy Consumption Model

In this paper, the energy consumption model is based on the proposed model in DEECIC algorithm [9]. In this model, the multipath fading channel is used when the distance between the sender and the receiver is more than threshold d_0; otherwise the free space model is used. Therefore, in order to transmit an L-bit message with the distance d there would be:

$$\varepsilon_{Tx}(L,d) = \begin{cases} L \times E_{elec} + L \times \varepsilon_{fs} \times d^2 & d < d_0 \\ L \times \varepsilon_{elec} + L \times \varepsilon_{mp} \times d^4 & d \geq d_0 \end{cases} \tag{3}$$

Where E_{elec} is the energy dissipated per bit to run the transmitter or the receiver circuit, ε_{fs} or ε_{mp} is the energy dissipated per bit to run the transmit amplifier. To receive this message, the required amount of energy would be:

$$E_{Rx}(L) = L \times E_{elec} \tag{4}$$

E_{DA} Is amount of energy for data aggregation, thus the energy dissipated for aggregating m-bit data is as follow:

$$E_A(m,L) = m \times L \times E_{DA} \tag{5}$$

The cluster head can always aggregate the data gathered from its members into a single length-fixed packet.

C. PDKC Algorithm Description

The proposed clustering algorithm considers the energy consumption besides the network coverage. In the previous works just one of these factors was subjected. If both parameters are considered, the clustering algorithm would be more realistic. The aim of PDKC is to cover the whole network by having the constant number of cluster heads, and then having executed the algorithm, each node becomes a member of only one cluster. According to the sensor node distribution model in the network, this approach will result in an appropriate coverage in the whole network. In the proposed algorithm each node sends its data with at most 2 hops to its corresponding cluster head. Therefore, sensor nodes that cannot communicate to the CH directly can profit from an intermediate sensor node to communicate indirectly with its CH. The transmitted data to a cluster head consists of the node identity and the received information from the environment. Hence, the CH can recognize the source of the received message either directly or indirectly. Consequently, in the proposed algorithm, without imposing significant overhead to the nodes, the network area becomes expandable. In PDKC, each sensor node has a threedimensional identity, namely that of (i,j,k). While the first two dimensions, (i,j), specify the cell of the sensor node the last one is the sensor node identity in the cell.

In each round a new cluster head is selected. In order to select the appropriate new cluster head, sensor nodes belonging to a cell must send some information to the current cluster head. Since, in PDKC, the number of sensor node degree is a criterion of CH selection, the nodes located in the denser parts of the network would be more likely to be selected as cluster head. In the proposed algorithm each sensor node can have 3 statuses; head of a cluster, member of a cluster, and/or non-member of a cluster. In the following subsections, more detail of PDKC algorithm is presented.

1) Cluster Formation

In the first step of the algorithm, after deployment of the sensor nodes to the network, BS broadcasts a "Hello" message to the entire network in order to estimate the distance of each node from the BS. This message contains the primary energy that the BS needs to send the message to all deployed sensor nodes. Having received the message, each node can estimate its distance from the BS by comparing the amount of the received signal energy to the primary energy level received in the message. Suppose during the execution of the algorithm all sensor nodes are awake in the network. Moreover, at the beginning of each round, every node broadcasts an "update" message to the network to indicate its existence to the neighbors. The TTL of the broadcasted message is one. Therefore, sensor nodes that have received the update message do not relay it. The reason behind sending this message is that each node computes its degree. Therefore, the nodes' degree is calculated as follows [9]:

$$NodeDegree\ S_i = count\ s_j\ dist\ s_i < R_c, s_j \in \mathbb{S}, s_i \neq s_i \quad (6)$$

Where S_j is a neighbour of S_i in the wireless transmission range of S_i (R_c).

After sending "Update" messages at the beginning of each CH selecting round, each node sends a unicast message, called feature message, including its identity, residual energy, and degree to its cluster head or to the virtual cluster head which will be described in the following. Then each CH will select two sensors nodes with the highest degree based on received feature messages. Afterward, one of them with the highest level of residual energy is selected as a new cluster head. This decision is announced to the new cluster head by a unicast message. Then, the new cluster head broadcasts a message to the network with TTL equal to one, called advertise message. In this message, the new CH advertises himself to its neighbors. Also, the residual energy of the CH and the estimated distance between the node and the BS is announced to the sensor nodes by the advertise message.

Therefore, each node that receives the advertise message can decide which CH is suitable. The selection of an appropriate cluster head is done based on the residual energy and the distance of the CH to the BS. Therefore, each sensor node selects two CHs with the highest residual energy. Then the sensor node selects the one that is the closest to the BS as its cluster head.

Each sensor node that selects its cluster head broadcasts a message with TTL equal to one to inform its neighbors about the new cluster head. Therefore, some sensor nodes that have not received any advertise message get informed about the nearest cluster head. Hence, the sensor node with at most two hop distances from a cluster head can select it as CH. By this operation, almost all sensor nodes can select at least one cluster head and the probability of uncovered area will decrease extremely. The selection of cluster head with two hop distance is similar to one hop distance. It is mentioned that if two cluster heads have equal estimated distance from the BS, one of them is selected randomly.

As mention before, at the begging of each round, the feature message is sent to the cluster head or virtual cluster head. In all rounds except for the first one, each sensor node knows its cluster head. In the first round, no cluster head is introduced to the network. Therefore, in this round a special sensor node in each cell is considered as a cluster head for that cell. Assume a sensor node with identity (i, j, 0) is considered as a virtual cluster head for cell (i, j). Consequently, this sensor node is responsible for the new cluster head selection when sensor nodes are deployed to the network.

2) Algorithm Specification

According to network distribution model, nodes distribution models, and clustering method the proposed algorithm enjoys the following specifications:

- In PDKC, the clustering procedure is distributed thoroughly, and each sensor node sends some information to the neighboring sensor nodes. Since the algorithm is distributed, it does not cope with some challenges in the centralized algorithm.
- PDKC algorithm does not need the initial time synchronization. As mentioned before, in the proposed algorithm, each sensor node can make its own decisions based on the received information independently from other sensor nodes.
- PDKC algorithm is scalable regardless of network size and the number of deployed sensor nodes in the network.

IV. Performance Analysis

In this section, we analyse the performance of PDKC algorithms using MATLAB which allows realistic modelling of sensor nodes. In our simulation the BS is located in a fixed point outside the network area. The simulation parameters are presented in Table I:

According to the simulation parameters, the performance of PDKC, LEACH, EEUC and EEBCDA algorithms are compared. Theses algorithms are evaluated base on the following criteria:

- End to end delay

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Network size	$(1000 \times 1000)\text{m}^2$
Node Number	10000
BS position	500*1250
Initial energy	5J
Deviation of Gaussian(δ)	50m
Node Number in each cell	100
Wireless transmission	40m

- Average cluster size
- Number of cluster head in each round
- Energy consumption in each round
- The number of alive sensor nodes in each round.

A. *End-to-End Delay*

At first, the evaluation of end to end delay for the proposed algorithm is discussed. The delay is the average number of hops between a member of a cluster and cluster head for a connection. The result is as shown in the Fig 3.

As mentioned before, in the proposed algorithm the number of hops between sensor nodes and their cluster head can reach to two hops. The two hops connection is suitable for some sensor nodes which cannot connect to the cluster head directly. Therefore, they use an intermediate sensor node which is one hop from cluster head to communicate with. As it is inferred from the figure, the most connections between sensor nodes and cluster head are of one hop type. The percentage of two hops connection is about 5 to 10.

B. *Cluster Size and the Number of Cluster Heads*

In order to evaluate the mean cluster size in PDKC algorithm the following equation is used:

$$R_r = \frac{R_{alive}}{count(CH)} \qquad (7)$$

In Equation 7, R_{alive} is the number of alive sensor nodes in the new round of cluster head selection. The average

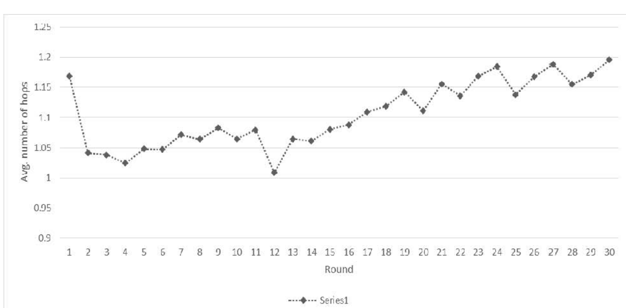

Fig. 3. End-to-End delay

cluster size for considered algorithm is depicted in Fig 4. As it is inferred from the figure, the cluster size for PDKC algorithm is significantly more than other algorithms. The number of cluster in the network has an important effect on the clustering performance. If the number is low, some sensor nodes may be located far from the cluster heads. Therefore, the communication between sensor nodes and their cluster head may impose more overhead. If the number of clusters is greatly increased accordingly the overlap regions between them increases as well. As it is clearly inferred, both cases seriously impact the performance. However, in the proposed protocol, the number of clusters is fixed to the number of cells. Therefore, not only is the average of hop count closer to one, but also the overlap between clusters is very low. The number of cluster head for LEACH and EECU are depicted in Fig. 4 for different rounds. Since, the number of cluster head is fixed in PDKC algorithm, it is not depicted in the Fig. 5.

C. *Energy Consumption in Sensor Nodes*

At the begging of simulation, all sensor nodes have identical energy. When a node runs out of energy, it is considered as a dead one. The number of live sensor nodes is shown in the Fig. 6 for different rounds. The number of surviving nodes is much greater in PDKC algorithm, because not only considers it the remaining energy of the node, but also it tries to choose sensor nodes as CH in denser areas. Therefore, the cluster head can communicate with more sensor nodes directly. This method of choosing a CH can balance energy consumption of the cluster head in such a wide network with numerous sensor nodes.

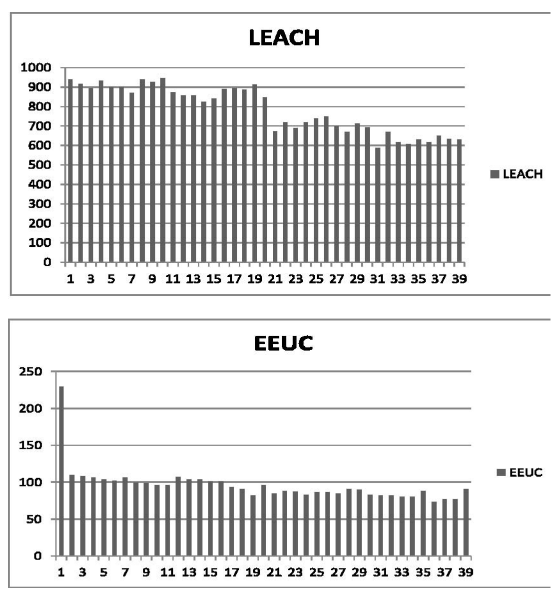

Fig. 4. The number of clusters in each round

Fig. 7 shows average energy consumption for one CH in a round. In the simulation, if the energy of a CH runs out or even gets negative, the algorithms continue until the round is ended. Considering this fact, in EEUC and EEBSDA

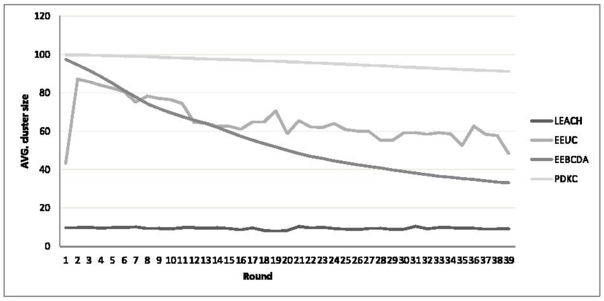

Fig. 5. Average size of clusters

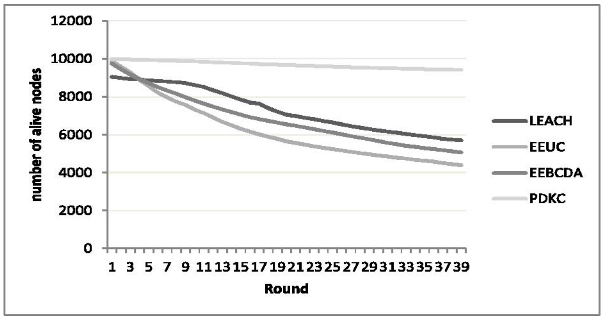

Fig. 6. The number of alive nodes

algorithms, the average energy consumption for each CH in the first rounds outnumbers the initial energy level. According to TABLE I, the initial energy of a sensor node is considered 5 J. As it is shown in Fig. 7 CHs in PDKC and LEACH consume more energy than the initial value. Therefore, EEUC and EEBSDA algorithms are not appropriate for wide network areas with a large number of nodes.

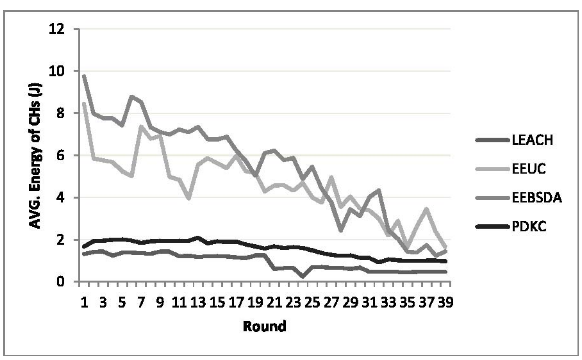

Fig. 7. Average amount of energy dissipated by CHs in each round for different algorithms

V. Conclusions and Future Work

In this paper a distributed energy efficient clustering algorithm, namely PDKC, based on deployment knowledge was proposed. The deployment knowledge was modelled based on Gaussian probability distribution function instead of using any location aware device such as GPS. In the proposed algorithm, the network has been divided into hexagonal cells. The cluster head was selected based on the residual energy, node degree and deployment knowledge. Simulation results indicate that PDKC algorithm improves network coverage and prolongs network lifetime in comparison with LEACH, EEUC and EEBCDA especially for large-scale-networks.

References

[1] A. Kemal and Y. Mohamed, "A survey on routing protocols for wireless sensor networks," Department of Computer Science and Electrical Engineering, University of Maryland, USA, 2003.

[2] C. Intanagonwiwat, R. Govindan, D. Estrin and J. Heidemann,"Directed Diffusion for Wireless Sensor Networking ," *IEEE Trans. Networking*, vol. 11, no. 1, 2003.

[3] M. Younis, M. Youssef and K. Arisha, "Energy-aware management in cluster-based sensor networks," *Computer Networks 43 (5) (2003)* 649–668.

[4] M.T. Thai, F. Wang and D. Du, "Coverage problems in wireless sensor networks design and analysis," *International Journal of Sensor Networks 3 (3) (2008)* 191–200.

[5] M. Cardei and J. Wu, "Energy-efficient coverage problems in wireless ad hoc sensor networks", *Computer Communications Journal 29 (4) (2006)* 413–420.

[6] M. Ye, C. Li, G. Chen and J. Wu, "An energy efficient clustering scheme in wireless sensor networks," Ad Hoc & Sensor Wireless Networks, 2006, 3, p. 99–119.

[7] Z. Liu, Q. Zheng, L. Xue and X. Guan, "A distributed energy-efficient clustering algorithm with improved coverage in wireless sensor networks," In *Future Generation Computer Systems Journal*, Vol. 28, p.780-790, 2011.

[8] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-efficient communication protocol for wireless micro-sensor networks". In Proc of *IEEE Conf on System Sciences*, 2000, p. 3005–3014.

[9] A.P. Chandrakasan, A.C. Smith and W.B. Heinzelman, "An application specific protocol architecture for wireless microsensor networks", *IEEE Transaction on Wireless Communications 1(4) (2004)* 660–669.

[10] J.N. Al-Karaki, R. Ul-Mustafa and A.E. Kamal. "Data aggregation in wireless sensor networks-exact and approximate algorithms," In *Proc of Workshop on High Performance Switching and Routing, 2004*, p. 241–245.

[11] J. Yuea , W. Zhang, W. Xiao, D. Tang and J. Tang. "Energy Efficient and Balanced Cluster-Based Data Aggregation Algorithm for Wireless Sensor Networks". Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology Changsha-410073, China.

[12] S. Soro, W.B. Heinzelman. "Prolonging the lifetime of wireless sensor networks via unequal clustering," In *Proc of IEEE Parallel and Distributed Processing Symposium, 2005*, p. 1–8.

[13] M. Gerla and J.T.-C. Tsai, "Multicluster, mobile, multimedia radio network," *Wireless Networks*, vol.1, no. 3, 1995, pp. 255-265.

[14] A. Fanian, M. Berenjkoub, H. Saidi, T.A. Gulliver, "A High Performance and Intrinsically Secure Key Establishment Protocol for Wireless Sensor Networks", *Computer Networks Vol. 55, Issue 8, 1 June 2011*, Pages 1849–1863.

[15] W. Du, J. Deng, Y.S. Han, S. Chen, and P.K. Varshney, "A Key Management Scheme for Wireless Sensor Networks Using Deployment Knowledge," in *Proc. IEEE Conf. on Computer Communications.* (INFOCOM), pp. 586-597, Mar. 2004.

[16] M. Ramkumar and N. Memon, "An Efficient Random Key Predistribution Scheme," in *Proc. IEEE Global Telecommunication. Conference* (GLOBECOM), pp. 2218-2223, Dec. 2004.