

Data Gathering in Wireless Sensor Networks Using Intermediate Nodes

Abdullah I. Alhasanat¹, Ahmad Ali Alhasanat², Khitam M. Alatoun³, Aws AlQaisi⁴

¹Faculty of Engineering, Al-Hussein Bin Talal University, Ma'an, Jordan

²Faculty of Business Administration & Economics, Al-Hussein Bin Talal University, Ma'an, Jordan

³Faculty of Engineering, Al-Hussein Bin Talal University, Ma'an, Jordan

⁴Faculty of Engineering Technology, Al-Balqa' Applied University, Salt, Jordan

ABSTRACT

Energy consumption is an essential concern to Wireless Sensor Networks (WSNs). The major cause of the energy consumption in WSNs is due to the data aggregation. A data aggregation is a process of collecting data from sensor nodes and transmitting these data to the sink node or base station. An effective way to perform such a task is accomplished by using clustering. In clustering, nodes are grouped into clusters where a number of nodes, called cluster heads, are responsible for gathering data from other nodes, aggregate them and transmit them to the Base Station (BS).

In this paper we produce a new algorithm which focused on reducing the transmission path between sensor nodes and cluster heads. A proper utilization and reserving of the available power resources is achieved with this technique compared to the well-known LEACH_C algorithm.

KEYWORDS

WSN, BS, Clustering, Cluster head, Data aggregation.

1. INTRODUCTION

Wireless Sensor Networks (WSNs) are composed of large number of low power, small size and low cost sensor nodes. A sensor node is an electronic device with the capability of detecting physical conditions, computation and communication. Those sensor nodes can be scattered to perform a variety of applications such as wildlife monitoring, habitat monitoring, fire surveillance, etc. Figure 1 shows the basic structure of a WSN.

A general WSN is composed of sensor nodes, a base station (or sink), and the events being monitored [17].

- A sensor node typically consists of several parts including: a radio transceiver, a sensing unit, a microcontroller and power source usually a battery. The sensor nodes might vary in cost from few to hundreds of dollars depending on the functionality of

each sensor node. The constraints of cost and size of the sensor nodes led to constraints on its resources such as energy, communication and computation.

- A sink node is a resourceful node having unrestricted communication and computational capabilities in addition to energy source, it can be stationary or dynamic and act as an interface between the sensor network and management center.
- The event being monitored may be stationary or mobile, depending on the application of WSN. Mounting sensor nodes on wild animals for behaviour monitoring, where these animal move in an unexpected manner, is an example of mobile sensor nodes applications. On the other hand, sensor nodes may be deployed on stationary and known locations.

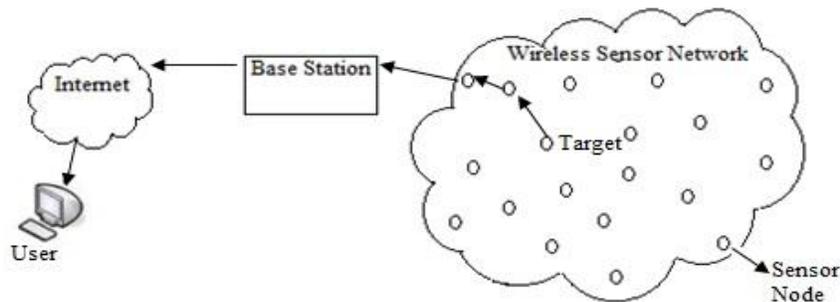


Figure (1) Basic structure of WSN

Each sensor node can communicate and exchange its data with other nodes and the base station. In this context, sensor nodes can use variable or fixed power for data transmission; as the distance between the source and destination nodes is increased, the required power is increased [17]. For instance, with single-hop communication, the transmission power should be sufficient to deliver data to the destination node. The direct communication method [18] is an example of dynamic power transmission power. However, for a static transmission power scheme, multi-hop communication is required to deliver data to distant nodes. The Minimum Transmission Energy algorithm (MTE) uses such type of transmission power [19].

The limited energy resource, such as using non-rechargeable battery supplies to each sensor node, is one of the most crucial challenges in WSNs. Many routing algorithms have been proposed for WSNs. Most of the hierarchical routing algorithms proposed for WSNs concentrated mainly on prolonging the lifetime of the network by reducing the energy consumption [16]. Recent research proved that nodes clustering provided an effective approach for energy conservation in WSNs. In WSNs, data are collected from the deployed sensor nodes and sent to the sink node and then to the base station for analysis by the end user or application [1]. Sensor nodes suffer from limited-power sources and hence it is inefficient to convey their data directly to the sink node [7]. Instead, an appropriate data gathering algorithm is required to gather data from these nodes in an energy efficient way while maximizing the network lifetime ; the ultimate goal of all sensor networks.

In addition, data gathering would be more efficient with homogenous sensor networks. In this case, data aggregation is accomplished by collecting and aggregating data from a set of sensor nodes. The collected data is combined into a single data packet to be sent to the sink node. Certainly, this leads to minimize the number of transmissions by eliminating data redundancy and thus reduce the total power consumption in the network [3].

Clustering in WSNs is an important technique to reduce the energy consumption over these networks and thus prolonging it is network lifetimes. Many energy efficient protocols based on clustering and data aggregation have been studied [2][3].

In this paper, a new data gathering algorithm is proposed. The key idea behind this algorithm is to recursively divide the sensor network into four partitions symmetrical about a centroid node. Furthermore, a set of cluster heads in the middle of each partition are defined in order to aggregate data from cluster members and transmit these data to cluster heads in the next hierarchical level. This procedure continues until a prescribed number of sensor nodes in each partition are reached. At the end of this procedure, a set of partitions of almost equal number of nodes are produced.

The advantages of this algorithm are threefold. Firstly, equalizing the number of sensor nodes in each partition would greatly help to distribute the load among sensor nodes and therefore leads to proper utilization of the available power resources. Secondly, a set of cluster heads are assigned to each partition in each level. These nodes are selected as intermediate nodes in the cluster. This step is essential in order to prolong the network life time of cluster heads since these nodes usually consume their power more quickly compared to other normal nodes.

Moreover cluster heads do not need to send their data for long distances, as proposed in LEACH [4] where each cluster head transmits its data directly to the base station. In contrast, in our algorithm, cluster heads gather data from their cluster member nodes. Then, each cluster head computes the average and transmit it to the next cluster head in the hierarchical structure.

The remainder of this paper is organized as follows. An introduction to some related works in the literature is presented in Section 2. A proposed algorithm is presented in section 3. Result and discussion is section 4. The summary is drawn in section 5.

2. RELATED WORKS

Numerous clustering-based data aggregation protocols have been proposed recently. In this section we briefly discuss some of these algorithms. For example, in Low Energy Adaptive Clustering Hierarchy (LEACH) [4] algorithm, the deployed nodes group themselves into clusters for data aggregation.

In each cluster, a single node is elected to be a cluster head. Each cluster head aggregates data from its cluster members and sends this data directly to the base station. The cluster head eliminates redundant data and uses one of the aggregated functions to minimize the transmitted data to the sink node. LEACH protocol consists of two phases: setup phase and steady state phase. In the setup phase, the clusters are arranged and the cluster heads are elected. Each sensor node compares a random number r between 1 and 0 with a threshold H , H is given by

$$H = \frac{p}{1-p (r \bmod (1/p))} \quad (1)$$

Where p is the predetermined percent of cluster head in the network, r is the current round.

If the node was not a cluster head in the previous $1/p$ rounds, and $r > H$, then the node becomes a cluster head and broadcasts a message to all other nodes informing them that it is a cluster head, all non-cluster head nodes received a broadcast messages and determine to which cluster heads they belong based on the Received Signal Strength (RSS) of the received message.

In [8], the author proposed LEACH Centralized (LEACH-C). This protocol is identical to the LEACH protocol but the Base Station (BS) organizes the clusters. In the setup phase, each sensor node sends information about its location and its residential energy to the BS, where BS organizes the clusters and defines cluster head. BS then sends a broadcast message with the IDentification (ID) of cluster head for each sensor node.

Distance-based Clustering Routing Protocol in Wireless Sensor Networks algorithm [6] proposed a different approach to picks cluster heads based on distance. In this algorithm, non-cluster head nodes defined the cluster head which is closest to the center point between the node itself and the sink node. Each round in this method is consists of two phases; the setup and steady phases. The setup phase defines clusters and cluster heads as proposed in LEACH [4], and each node selects its cluster according to the distance.

In the Power-Efficient Gathering in Sensor Information Systems algorithm (PEGASIS) [5], each node sends data to its nearest neighbor and only a single node transmits the aggregated data to the sink node. The chain of nodes is formed using a greedy algorithm that collects data in each round, the chain leader aggregates data to transmit to the sink node.

Ying Liang and Hongwei Gao [11] proposed a clustering algorithm referred to as OCABTR. In this algorithm, the authors suggested a strategy for forming clusters using a genetic algorithm, and selecting the cluster-head. Each round in OCABTR is composed of setup phase, which organizes adjacent nodes into clusters then elect the cluster heads, and steady-state phase when each node sends its data to a cluster-head. The cluster-head aggregate the received data and transmit them to the BS.

This technique expends intra-cluster communication cost for electing cluster heads but select them based-on residual energy.

In addition, authors in [12] proposed an approach called unequal clustering size (UCS) where the sensor network is organized into different size clusters. The cluster heads are determined a priori and located around the BS.

The nodes around the cluster head form the cluster. The number of nodes in each cluster depends on the number of node in the next cluster. This algorithm assumes that the BS is located in the middle of the network.

Another clustering algorithm, known as Energy-Aware Routing Protocol (EAP), was introduced [13]. In this protocol a novel scheme for inter-cluster communication is proposed and used new parameters for selecting cluster heads. Each node has a table of the remaining energy of all neighboring nodes within its cluster area. This table helps each node to compute the average remaining energy of its neighbors. Any node whose remaining energy is higher than the average value will be assigned higher probability to become a cluster head.

In EAP, each round begins with constructing a routing tree of cluster heads followed by transmitting data from nodes to cluster heads, where the data are aggregated and sent to the BS. EAP uses a computation model to determine routing tree.

In [14], a clustering and leveling algorithm, called Energy Efficient Threshold Sensitive Hierarchical Routing Algorithm for Cognitive Wireless Sensor Networks (ETSHRA), was proposed. ETSHRA is composed of four phases. In the first phase, which is called leveling, nodes are divided into logical levels based on the power level of the received signal from BS. After leveling, clustering phase is initiated, first cluster heads are selected randomly, and then the nodes start to belong to a cluster head based on the level which they locate.

Once the leveling and clustering phases are completed, a chain from cluster heads to BS is established. At the last phase, the soft and hard thresholds are used to allow sensors to transmit their data.

The authors in [15] developed a model to form clustering in sensor area. The clusters formulated in this method are heterogeneous-sized clusters, where the largest clusters are those located farther from the sink. A greedy algorithm was used to choose cluster heads. Firstly, the nodes with the highest energy needed to reach the sink are marked, and then every node computes the gain achieved by being a cluster head. Those with the highest gain are selected to act as cluster heads.

3. THE PROPOSED ALGORITHM

In this work, we assume a homogenous WSN consists of N sensor nodes that are uniformly and randomly distributed within an area of $L \times W$ meter square. The sink node is located at $(L + 10, W/2)$. The problem considered in this paper is to gather data generated from N sensor nodes every μ seconds. Cluster-heads are defined to receive data from all member nodes of their clusters and transmit the aggregated data to the sink node directly or through other cluster heads. The network is assumed to be homogenous in that sensor nodes are required to sense identical type of information. The goal of this algorithm is to present a strategy for defining intermediate cluster heads to minimize the distance between the cluster heads and their member node, so that the total energy consumed in the WSN is reduced.

Our algorithm is divided into two phases; setup phase and steady state phase. In the setup phase a recursive algorithm is used to define clusters and cluster heads, which remain fixed over the network lifetime. The setup phase is occurred once over the network lifetime, as a result of this phase, the network is grouped into clusters, where each cluster defines its cluster heads. During steady state phase, data transmission continues until the network lifetime is over.

With this algorithm, the network is firstly partitioned into four areas of almost equal numbers of nodes. Then, each area is further partitioned into four sub regions. This recurs until the constraint $N < N_{th}$ is matched, where N is the number of nodes in the partition and N_{th} is the maximum number of nodes allowed in each partition. There is no optimal value of such number.

The objective of this algorithm is to select a cluster head that is close to all other nodes in the cluster, which then minimize the transmission power needed to transmit data from a cluster member to a cluster head. Therefore, the cluster heads are defined as the intermediate nodes in the

cluster. These nodes can be described as the nodes that have the median of the sum of the pairwise Received Signal Strength (RSS) values r with other nodes in the cluster. Hence, the cluster head c_l^k for cluster l in the k_{th} level is computed as

$$c_l^k = \arg \operatorname{median}_{i=1 \dots N} \left\{ \sum_{j=1}^N r_{ij} \right\} \quad (2)$$

where r_{ij} is the RSS value between node i and j , for $j = 1 \dots N$ and N is the number of cluster's members. For each cluster a number of cluster heads are assigned so that they can share the function of gathering and sending cluster's data in a round robin fashion.

Figure 2 shows the hierarchical clustering scheme used with the proposed algorithm. The circle represents a network partition.

Figure 3 shows the flow chart of the proposed algorithm through the setup phase. During the steady state phase, each node maintains a table of cluster heads. In each round r the node select its cluster head h from number of cluster heads n as:

$$h = r \bmod n \quad (3)$$

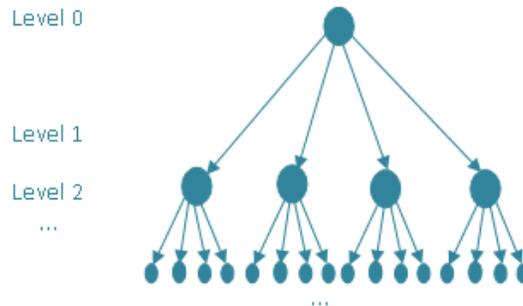


Figure (2) The hierarchical clustering scheme used with the proposed algorithm.

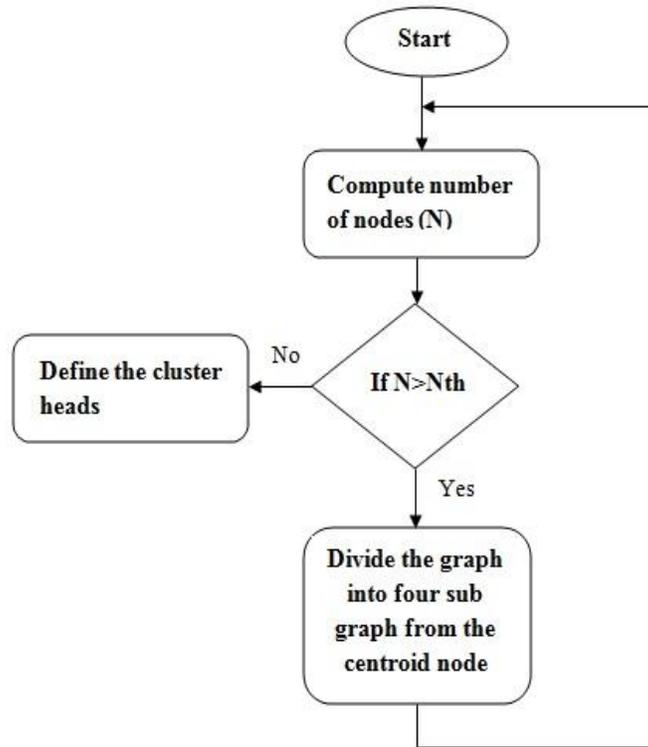


Figure (3) The flow chart of the algorithm

The algorithm can be briefly described by the following steps:

Input (Network graph topology (G), Threshold number of nodes (N_{th}), Sink node (S))

Output (network cluster, cluster heads)

1. *Compute number of nodes in graph (N).*
2. *If N is greater than N_{th} , then partition the graphs into four sub graphs from the centroid node and return to step 1.*
3. *If N is less than N_{th} , then define the cluster heads. Each node in the area maintains a table of the cluster heads.*

4.RESULTS AND DISCUSSION

The performance of the proposed algorithm is evaluated using 200 independent simulation runs through Matlab. This algorithm is compared to the LEACH-C algorithm. The performance of both algorithms was assessed in terms of total energy dissipation under different network diameter, network lifetime of the network and number of dead nodes over the simulation time.

4.1 Simulation Scenario

For our experiments, 200 sensor nodes are uniformly and randomly distributed within an area of 200 x 200 meter square, the sink is fixed outside the network area at (210,100)m. Each sensor node has 1 Joule initial energy, the size of data message (k) is set to 100 Kbits.

4.2 Radio Model

We assume simple radio model, where the energy dissipated to run the receiver's and transmitter's circuitry is set to $E_{elec}= 50\text{nJ/bit}$, and the transmitter's amplifier $\epsilon_{amp}=10$ pJ/bit/m². The cost to transmit a message depends on the distance between the transmitter and receiver [9], both the free space (d^2 power loss) and the multipath fading (d^4 power loss) channel models were used, depending on the distance between the transmitter and receiver [9], if the distance is less than a threshold (d_0), the free space (d^2) model is used; otherwise, the multipath (d^4) model is used.

The distance threshold is set to 44 meters, which represents a typical communication range for IEEE std. 802.15.4 (ZigBee) nodes at environments with moderate number of obstructions [10]. Thus, the transmission cost considered in this paper is given as

$$E_T(k, d) = \begin{cases} E_{elec} * k + \epsilon_{amp} * k * d^2, & \text{if } d \leq d_0 \\ E_{elec} * k + \epsilon_{amp} * k * d^4, & \text{if } d > d_0 \end{cases} \quad (4)$$

and the receiving cost is given as

$$E_R(k) = E_{elec} * k \quad (5)$$

With this algorithm there is no fixed number of clusters; it depends on the number of sensor nodes in the area. In our experiments, we used a threshold value N_{th} of 50 nodes. Therefore, the number of clusters varies as a function of number of sensor nodes; at the end of the setup phase each cluster will contain almost the same number of nodes around N_{th} .

The percentage of cluster heads is set to 5% of the cluster size. In each communication round one cluster head is selected through round robin basis in order to gather, and transmit the cluster data. Our algorithm and LEACH_C use the same constants (E_{elec} , ϵ_{amp} , and k) for calculating energy costs, therefore our algorithm achieves its energy savings by minimizing d (the distance between sensor nodes and cluster head nodes) and the number of transmissions and receives for each node by reducing multi-hop communication.

4.3 Energy dissipation under different network diameters

It is important to examine the performance of our algorithm and LEACH-C algorithm for different network sizes. Figure 4 shows the energy dissipated at the area of sensor fields is varied from 100 to 550 meter square.

As notice from this figure that our algorithm consumed less energy than LEACH-C at different network diameters. It is also shown that further improvement of consumed energy is achieved with our algorithm as the diameter of the network increases. For a small network diameter both algorithms exhibited an identical performance. In fact, when the network diameter is increased, transmission path between sensor nodes and cluster heads is also increased and so higher transmission power is needed. In addition, in the presence of large inter-sensor distances the shadowing channel model described by equation (4) is used, which greatly affected the energy dissipated of the network.

Despite this, dividing the network into smaller cluster sizes as in our algorithm and select the intermediate nodes to play as cluster heads to minimize the total amount of the dissipated energy.

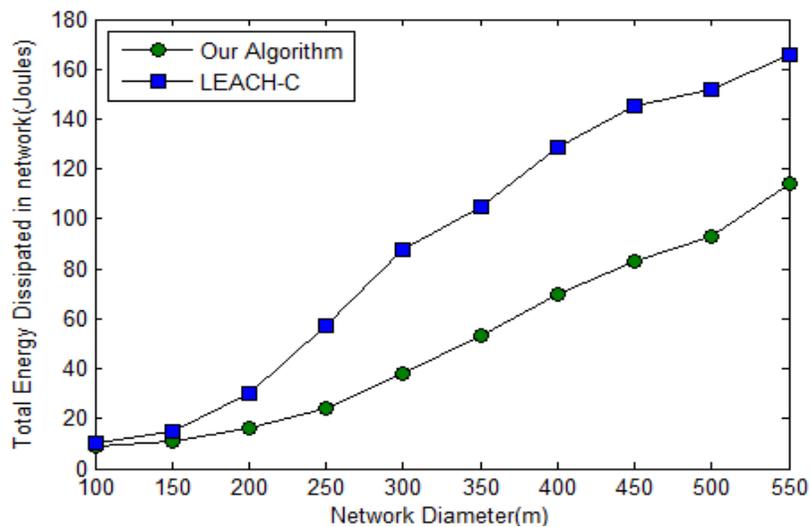


Figure (4) Total energy dissipation of LEACH-C and our algorithms at different network sizes.

4.4 Number of dead nodes

To evaluate the total number of dead nodes over the simulation time, we used a network with 200 nodes and saved the number of dead nodes at each rounds, as seen in Figure 5.

The results of the two algorithms showed that the death of the first set of nodes appeared around 2000 rounds. After that, the nodes for LEACH-C sharply die, in contrast to the proposed algorithm which showed gradual death of sensor nodes.

This indicates that the dissipation energy in the proposed algorithm is distributed over a longer time period in comparison to LEACH-C; allowing sensor nodes in this algorithm to live longer.

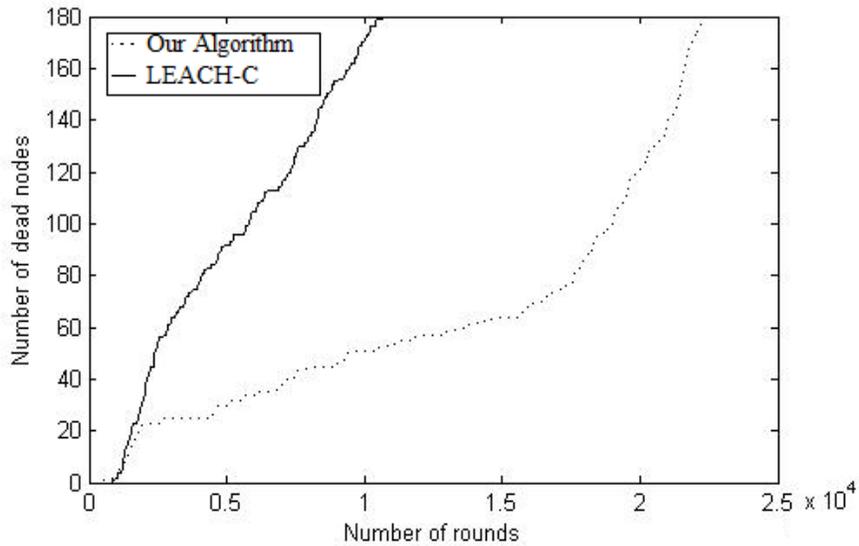


Figure (5) Total number of dead nodes over the simulation time.

4.5 Network lifetime with different death node percentage

The network lifetime is described as the amount of time elapsed during which the network is functioning properly. One important factor that affects the network from functioning well is a number of dead nodes. Determining the number of dead nodes for which the network is considered failed is not an easy task since it mostly depends on the network type, the applications itself, and the protocol being used.

In our simulations, we evaluate the lifetime of the sensor network for several percentages of dead nodes. Figure 6 shows the network lifetime until 10%, 20%, 30%, 40%, 50% and 60% of nodes die. 200 nodes were randomly distributed in 200x200 meter square.

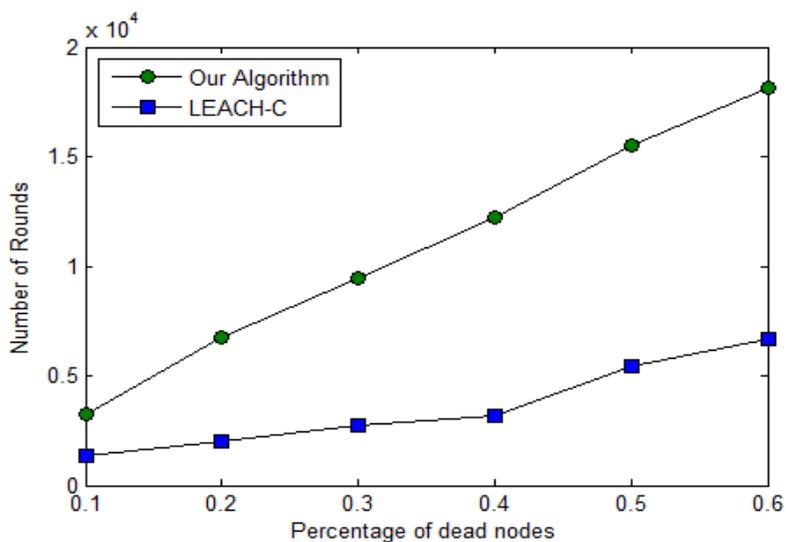


Figure (6) Network lifetime versus percentage of dead nodes

As can be shown from this figure, when the percentage of dead nodes increased, it allowed the sensor nodes to live longer, and this is true for the two algorithms. More interestingly, the proposed algorithm showed significant savings in the network lifetime, which outperformed the LEACH-C over the entire range of the percentage dead nodes. This result emphasized the benefits of using our algorithm for any percentage of dead nodes.

5. CONCLUSION

In this paper, we proposed a new energy efficient data aggregation protocol in wireless sensor networks. The key idea behind this algorithm is to recursively divide the sensor network into four partitions symmetrical about a centroid node. Furthermore, a set of cluster heads in the middle of each partition are defined in order to aggregate data from cluster members and transmit these data to cluster heads in the next hierarchical level. The new algorithm adopts the concept of hierarchical clustering which prevents cluster heads from sending their data for long distances and thus the energy consumption of the sensor nodes is significantly improved.

This algorithm focused on avoiding the overhead of dynamic clustering, reducing the transmission path between sensor nodes and cluster head nodes, and minimizing the direct communication between the sink node and cluster heads. Simulation results showed that the proposed algorithm achieved better performance in comparison with the LEACH-C algorithm in terms of energy consumption, network lifetime, and number of dead nodes.

As a future work, the intermediate nodes in each cluster can serve as cache points in WSN with mobile elements, in an attempt to reduce the contact time of mobile sink, and then reduce the latency of the network. Also, it would be useful if the residual energy is taken into account for cluster heads selection in each transmission round.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers of the paper.

REFERENCES

- [1] Jasmine Norman, J. Paulraj Joseph and P. PrapoornaRoja, "A Faster Routing Scheme for Stationary Wireless Sensor Networks - A Hybrid Approach", *International Journal of Ad Hoc, Sensor & Ubiquitous Computing*, Issn: 09762205, EIssn: 09761764, Volume: 1, Issue: 1, pages: 1-10, 2010.
- [2] Ramesh Rajagopalan and Pramod K. Varshney; "Data aggregation techniques in sensor networks: A survey", *Communications Surveys and Tutorials*, IEEE Issn: 1553-877X, Volume:8, Issue: 4, pages: 48 – 63, 2006.
- [3] KiranMaraiya, Kamal Kant and Nitin Gupta; "Wireless Sensor Network: A review on Data Aggregation", *International Journal of Scientific & Research*, Volume 2, Issue 2, 2011.
- [4] Wendi RabinerHeinzelman et al. "Energy-Efficient Communication Protocol for Wireless Microsensor Networks", in *Proceeding of the 33rd Hawaii International Conference on System Sciences*, January 2000, pages: 1-10.
- [5] S. Lindsey, C. Raghavendra, and K.M. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics," *IEEE Trans. Parallel and Distributed Systems*, Volume 13, no. 9, September 2002, pages: 924-935.

- [6] WANG Jun, Zhang, Junyuan and Zhengkun "A Distance-based Clustering Routing Protocol in Wireless Sensor Networks", Communication Technology (ICCT), 2010, 12th IEEE International Conference, Nov. 2010.
- [7] X. Lin and I. Stojmenovic. "Power-Aware Routing in Ad Hoc Wireless Networks", In SITE, University of Ottawa, TR-98- 11, Dec. 1998.
- [8] Heinzelman W. B., Chandrakasan A. P., Balakrishnan H., "An application specific protocol architecture for wireless microsensor networks," *IEEE Trans on Wireless Communications*, Volume 1, No. 4, 2002, pages: 660-670.
- [9] T. Rappaport, *Wireless Communications: Principles & Practice*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [10] "PIC microcontroller with 2.4GHz IEEE 802.15.4 transceiver and ZigBee stack," FlexiPanel Ltd, 2008.
- [11] Ying Liang and Hongwei Gao, "An energy-efficient clustering algorithm for data gathering and aggregation in sensor networks," *Industrial Electronics and Applications*, 2009. ICIEA 2009. 4th IEEE Conference on 25-27 May2009.
- [12] Soro, Heinzelman, W.B., "Prolonging the Lifetime of Wireless Sensor Networks via Unequal Clustering", 19th IEEE International Parallel and Distributed Processing Symposium, 2005.
- [13] Ming Liu, Jiannong Cao, Guihai Chen and Xiaomin Wang, "An Energy-Aware Routing Protocol in Wireless Sensor Networks," *Sensors*, ISSN 1424-8220, 2009.
- [14] Shahansha Quadri and Garimella R. Murthy, "ETSHRA: Energy Efficient Threshold Sensitive Hierarchical Routing Algorithm for Cognitive Wireless Sensor Networks", *International Journal of Information and Electronics Engineering*, Volume 2, No. 3, 2012.
- [15] Ali Dabirmoghaddam, Majid Ghaderi and Carey Williamson," Cluster-Based Correlated Data Gathering in Wireless Sensor Networks," *Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS)*, 2010 IEEE International Symposium on.
- [16] Chaurasiya, S.K., Pal, T., Bit, S.D., "An Enhanced Energy-Efficient Protocol with Static Clustering for WSN". *International Conference on Information Networking (ICOIN)*, 2011 Kuala Lumpur, Malaysia, pages: 58 – 63
- [17] Jamal N. Al-Karaki Ahmed E. Kamal," Routing Techniques in Wireless Sensor Networks: A Survey," *Wireless Communications*, IEEE Publication, Volume: 11, Issue: 6, pages: 6-28, 2004.
- [18] X. Lin and I. Stojmenovic. "Power-Aware Routing in Ad Hoc Wireless Networks", In SITE, University of Ottawa, TR-98- 11, Dec. 1998.
- [19] T. Meng and R. Volkan. "Distributed Network Protocols for Wireless Communication", In Proc. IEEE ISCAS, May 1998.