

A Tree Based Data Aggregation Scheme for Wireless Sensor Networks Using GA

Ali Norouzi¹, Faezeh Sadat Babamir², Zeynep Orman¹

¹Department of Computer Engineering, Istanbul University, Istanbul, Turkey

²Department of Computer Science, Shahid Beheshti University, Tehran, Iran

Email: norouzi@cscrs.itu.edu.tr, babamir@mail.sbu.ac.ir, ormanz@istanbul.edu.tr

Received May 9, 2012; revised June 18, 2012; accepted June 28, 2012

ABSTRACT

Energy is one of the most important items to determine the network lifetime due to low power energy nodes included in the network. Generally, data aggregation tree concept is used to find an energy efficient solution. However, even the best aggregation tree does not share the load of data packets to the transmitting nodes fairly while it is consuming the lowest possible energy of the network. Therefore, after some rounds, this problem causes to consume the whole energy of some heavily loaded nodes and hence results in with the death of the network. In this paper, by using the Genetic Algorithm (GA), we investigate the energy efficient data collecting spanning trees to find a suitable route which balances the data load throughout the network and thus balances the residual energy in the network in addition to consuming totally low power of the network. Using an algorithm which is able to balance the residual energy among the nodes can help the network to withstand more and consequently extend its own lifetime. In this work, we calculate all possible routes represented by the aggregation trees through the genetic algorithm. GA finds the optimum tree which is able to balance the data load and the energy in the network. Simulation results show that this balancing operation practically increases the network lifetime.

Keywords: Data Aggregation; Genetic Algorithm; Wireless Sensor Network; Routing; Energy Efficiency

1. Introduction

Nowadays, wireless sensor networks are appeared everywhere especially in ubiquitous and pervasive applications. These kinds of networks consist of a number of small, low power, certain communication range sensing nodes which cooperatively monitor the environment and transmit data through a route (a route is the collection of links edges in the network topology graph) between nodes from data sensor node to the Base Station (BS) or to the sink. This task is seriously energy consuming that leads the network out of service very soon due to the limited power of the nodes. Therefore energy efficient data dissemination routing-protocols are attended. The lifetime of the network is defined as the useful time where messages are exchanged toward to sink with minimum required active nodes. Intelligent optimum energy consumption protocols can extend the network lifetime and consequently withstand long transmission periods.

In this study, sensor nodes monitor the interested area to gather data and then remove the redundant ones in order to aggregate them according to the data aggregation spanning tree. Just then, they send packet data to the appropriate neighbors to conduct the packets toward to

BS through a route which consumes low energy. BS is a powerful station which all sensed data is gathered by sensor nodes and will be collected in here in order to process special goals. The proposed technique would be suitable for a homogeneous WSN with some spatial correlation to monitor the environment. Genetic Algorithm (GA) is classified as a heuristic function which can almost obtain an optimum solution by investigating the search space. In this paper, a GA based algorithm is used to create an efficient data aggregation tree in which any node has a value property. Consequently, any chromosome consists of some genes or indices that its index in gene indicates the corresponding parent node. In genetic algorithm generation section of the paper, we apply single point crossover to generate future population where corrupted individuals are distinguished by their corresponding value assigned by the selection process. The fitness function assigns a suitable value on the basis of residual energy, number of transmission and received data packets to individuals. Number of iterations and population size are determined according to the network size.

Following, we organize the paper with Section 2: Brief

description of the literature; Section 3: Problem discussion; Section 4: More detail about the usage of GA to create data aggregation tree; Section 5: Simulation results and finals; Section 6: Conclusion and future work.

2. Literature

During 2000's, many researchers have been devoted to finding efficient algorithms for WSNs. For instance, Kalkanis *et al.* [1] propose MLDA (Maximum Lifetime Data gathering Algorithm) which finds edge capacities that flow maximum transmission by running a linear program. This algorithm is able to maximize the lifetime of the network with certain locations of each node and the base station. One year later, Dasgupta *et al.* [2] extended MLDA by applying a cluster based heuristic algorithm called CMLDA where nodes are grouped into several clusters with a pre-defined size. The energy summation of cluster member nodes is their cluster's energy. The distance between clusters is computed by the maximum distance between every pair of nodes of two clusters. Based on this cluster information, MLDA is applied. In 2007, Hussain *et al.* [3] proposed EESR which is implemented on a multi-hop network. In order to make the energy consumption efficient, they also utilized a spanning tree which is the collection of routing trees. In EESR, the lowest energy node is selected to calculate its energy edge cost [4]. EESR chooses the lowest energy node and considers its lowest and highest edge cost links which receive and forward data packets to the neighbor or BS respectively. If selected node leads to make a cycle in the route, EESR chooses the second one in priority such as second highest edge cost link, to prevent node overloading. This technique balances the packet data load in the network and also the BS broadcasts the schedule or the individual tree to all nodes. Tan *et al.* [4] studied two spanning trees where the aggregation scheme and data gathering have been applied to extend the network lifetime. In this paper, generally, two methods are considered to manage the power among the nodes. The first method is the power aware version (PEDAPPA) which has an attempt to extend the lifetime by balancing the energy consumption among the nodes, unlike in the second method, PEDAP, the non-power aware version minimizes the total energy that is consumed by the system in each data gathering round [5]. This method extends the lifetime of the last node. The edge cost is calculated in different ways in each method. In PEDAP, the edge cost is the same summation of energy mounts for transmission and receiving while PEDAPPA divides PEDAP edge cost with transmitter residual energy that results asymmetric communication costs. A node with higher cost is considered later in the tree as it has few incomings. After determining of the edge costs, Prime's

minimum spanning tree is rooted at the BS, will be formed for routing the packets. This calculation is computed per every 100-iteration. Being active for all nodes and awareness of node locations at the BS are also considered as assumptions.

Jin *et al.* [6] utilized the GA to fulfill the energy consumption reduction. This algorithm gets a prime number of pre-defined independent clusters and then biases them toward to an optimal solution with minimum communication distance by the iterations of generation. They come to the conclusion that the number of cluster heads is reduced about 10% of the total number of nodes. They also show that the cluster based methods decrease 80% of the communication distance when compared with the direct transmission distance. In 2005, Ferentinos *et al.* [7] improved the proposed Jin *et al.* [6] algorithm with an extended fitness parameter. They investigated the energy consumption optimization and the uniformity of the measurement point by using a fitness function that involves the status of the sensor nodes, network clustering with the suitable cluster heads, and also the selection between two signal ranges from normal sensor nodes. In the next section we're going to describe the problem, the primary conditions of the nodes and the environment and the assumptions which will be considered for solving the problem.

3. Problem Statement

Assume a Euclidian square field with length l and is subdivided into equal area so that any interested sub area can be monitored by sensors which are located at the vicinal intersection lines (See **Figure 1**). A route consists of some nodes from the data monitor node to the BS which has not any cycle at all. A cycle leads to overloading some node(s) by many messages and shows an incorrect routing algorithm. This trick is intended by many researchers as a grid based wireless sensor network layout.

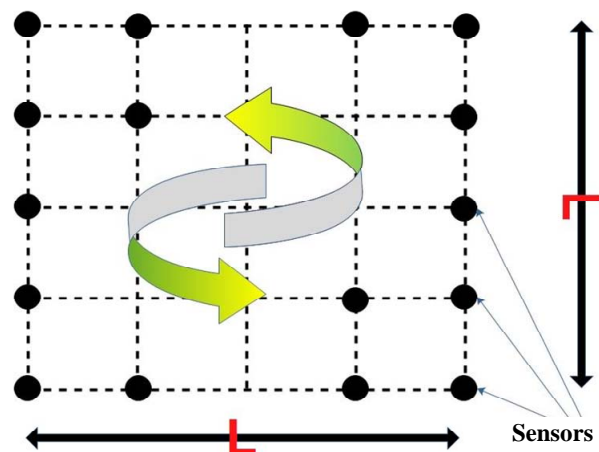


Figure 1. A sample wireless sensor network.

At first, we suppose that every node has a primary pre-determined energy for receiving the multi data packets in initializing state of the network and a fix communication radio rang. The whole nodes are able to monitor the interested area in addition to transmitting children packets to the neighbors as well as sending single ones to the corresponding parents. This task will periodically continue as long as possible. In our algorithm, firstly, all nodes send a sample certain packet to the BS once they are ready. When a packet should be transmitted through a route, the BS finds the monitor node and its property by using a routing table which has all the properties and current energy of all nodes and then feeds GA with this information. GA generates all possible routes and finds an optimum one. Just then BS schedules all nodes in the selected route and sends them their schedule. After that, GA updates the routing table by applying the energy reduction to all nodes. In fact, GA minimum spanning tree is the same aggregation tree which is rooted at the environment-monitor node that is formed to investigate the best edges toward to BS to balance the load of the data packets. In this paper, the network lifetime is credited until the minimum possible required nodes needed to transmit a data packet are active.

4. Genetic Algorithm

GA is commonly used in applications where the search space is huge and the precise results are not very important [8,9]. As GA is relatively computation intensive, it is executed only at the base station. It is categorized as a global search heuristic algorithm in which optimal solution is estimated by generating different individuals. This algorithm is made up of some procedures such as fitness and selection functions. In this study we focus on the fitness function which assigns a weight to the individuals according to their quality. High quality individuals have much more chance to survive and reproduce, however, the low quality ones will be removed during the passes of generation. In this study, GA is applied to obtain a balanced and energy efficient spanning tree. Every chromosome represents a tree where gene indicates a node and contains a value that points to the corresponding parent. Using the standard GA, the optimum minimum spanning tree will be obtained.

4.1. Gene and Chromosome

A chromosome is the collection of genes or nodes which have fixed length according to the number of former nodes. Every gene contains a value which provides the parent ID.

Figure 2 shows a chromosome where indexes indicate the children and contain a value or gene that shows the

corresponding parent. For example the child ID 0 has no parent. It means that it is the root and the data sensor node. ID 1 is the parent of child 3 and so on. Although the gene value is denoted as decimal, it can be calculated as binary. For instance: null-000-101-001-010-001. However, the genes are atomic and in the GA operations, chromosomes are considered as individuals.

GA forms all possible routes represented by aggregation trees and then calculates the corresponding fitness value to find an optimum one.

4.2. Crossover

The main step to produce a new generation is called as the crossover or reproduction process. In fact, it is a simulation of the sexual reproductive process that the inheritance characteristics are naturally transferred into the new population. To generate a new offspring, crossover selects a pair of individuals as parents from the collection that is formed by the selection process for breeding. This process will continue until the certain size of this new population is obtained. In general, there are various crossover operations which have been developed for different aims. The simplest method is the single-point in which a random point is chosen whereby two parents exchange their characteristics. Figure 3 shows an example of mating two chromosomes in the single point way.

4.3. Fitness Function

Fitness function is a procedure which scores any chromosome. This value helps us to compare the whole ones to each other for survival or death.

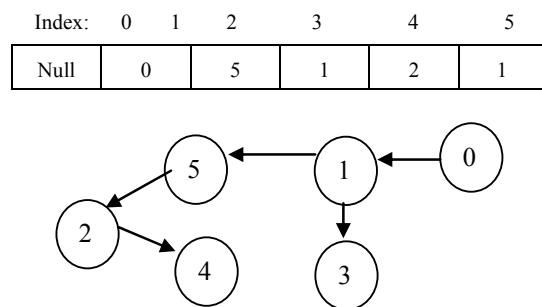


Figure 2. Chromosome and corresponding tree example.

	First	Second
Parents	101101'01101101	011110'10001011
Offspring	101101'10001011	011110'01101101

Figure 3. Single point method at random point 6.

To propose an appropriate fitness function as a part of the genetic algorithm, we inspired from a formula [10] which is then improved as the Nakamura formula [9], where A is the given monitoring area, S is the set of sensor nodes, D is the set of demanded points, A^d is the set of sensors monitoring the demanded areas, AL is the turning energy on in low energy level mod, AH is the turning energy on in high energy level mod which leads to send the packet load to the just BS, EC is the edge cost of a node in 3 states A , B and C and finally S is the set of edges collection.

The model can be formulated as:

$$G(i) = \sum_{i \in S} (AE_i + EC_i)$$

$$\forall i \in D \& \forall ii \in A^d, AE \in \{AL, AH\} \quad (1)$$

Mentioned formula calculates all possible paths as well as the state of the nodes which is needed to establish a node in low/high level energy mod. We improve our fitness function by considering 3 states at least for all nodes to calculate the almost exact fitness. It can be extended for nodes with high technology.

$$EC = \frac{4n_A + 2n_B + n_C}{\sum_{i \in S} n_i} \quad (2)$$

$$M \in \{A, B, C\}$$

EC is numerically measured and depends on the mode of the sensor network. Obviously, since the sensor node operating in mode A has a high communication range, it consumes the most energy and consequently, mode B has an intermediate communication range as well as mode C consumes the lowest power. Following, we assume that mode A use 4 times more power than C and also B is 2 times more than mode C . EC is given by the formula (2).

Below, we propose a fitness function where N is the number of nodes and setup energy is considered for calculating the electrical power. $E_{children}$ is the required energy to sending data packet received from children

$$E(i)_{Total} = E_{Monitor} + G(i)$$

$$F(i) = E(i)_{Total} / N$$

Above, we try to compute the average energy mount with division to the number of nodes. Each individual is assessed by the selection function where the better fitness value has the more chance to participate in the next generation.

5. Simulation and Evaluation

In order to implement the algorithm, we utilize a network simulator to assess our proposed algorithm which has two aspects. First one is the implementation of the ge-

netic algorithm based portion, using the Java editor. In this case, we installed Java Genetic Algorithm package (JPAC) to test which have already done by some researchers. After, we utilized OMNET++ to trace the different routes from sensor node to the BS in some virtual environments such as **Figures 4-6**.

The simulation parameters for the experienced sensor network are: 1) The network dimension is $100 \times 100 \text{ m}^2$; 2) Initial energy of each node is 0.8 J; 3) Random nodes placement is considered; 4) Each tree is used just for 15 periods; 5) Each scenario is simulated for 5 times that average one is reported; 6) BS is situated at the center of resource; 7) Wireless sensor network simulator, OMNET++ and JPAC.

5.1. Genetic Algorithm Parameter

The simulation for our heuristic algorithm is as follows: 1) Population size is 500 equal to network nodes; 2) Number of generation is 200; 3) Mutation and crossover rate is 0.7.

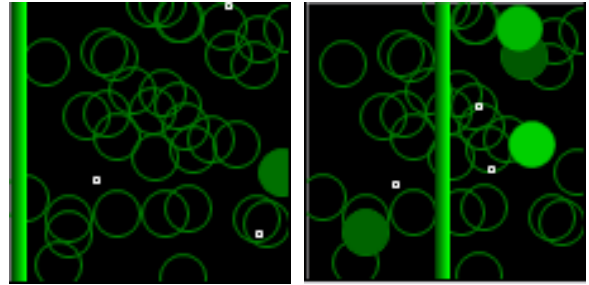


Figure 4. Network status in early setup.

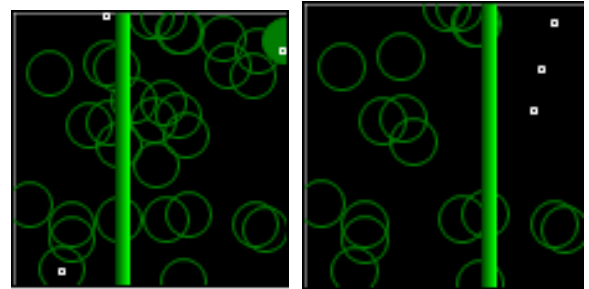


Figure 5. Failing some nodes after several periods.

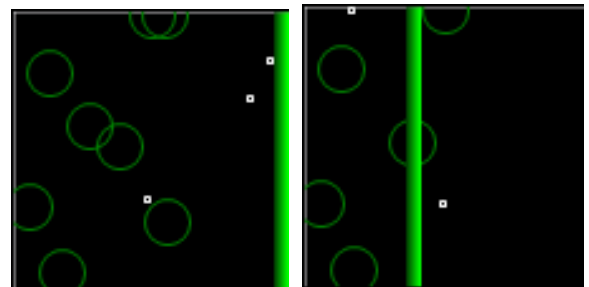


Figure 6. Network status in later lifetime.

Figure 4 represents the network under experience in early setup where little white point is the data packet while they are transmitting among nodes and green circles are active nodes at that moment. The light ones are transmitting while dark ones transmitted at this near time.

Hereinabove shows the network with fewer active nodes but it is not out of order as the required paths for transmitting are credited.

At last, **Figure 6** indicates our experienced network at its later lifetime. According to this figure, just a few nodes are valid for transmitting children and monitored data. In this status, the network is being to be out of order. As a result, these figures indicate that using proposed algorithm, the network has most withstanding to transmit data packet, as minimum spanning tree intelligently selects an affordable routing leading to extend lifetime.

Figures 7 and 8 compare proposed algorithm with LEACH protocol of network energy and lifetime which is considered in 200 periods of time (years). In **Figure 7**, unified packet load and consuming of energy by finding optimum rout toward BS. **Figure 8** represents time of removing first node because of energy status or briefly death time of first node is postponed rather than LEACH protocol. Also the network can be in order while minimum number of nodes is alive. Generally, due to using routing algorithm fitness function considers energy status of nodes in three states of low, middle and high rang communication specialized for the BS, final individuals provide rout which almost uniformly consume energy. This phenomenon significantly extends lifetime of network.

6. Conclusion

Presented paper utilizes genetic algorithm to find routs which balance energy and data load in network. It was attracted many researchers. Step by step property can

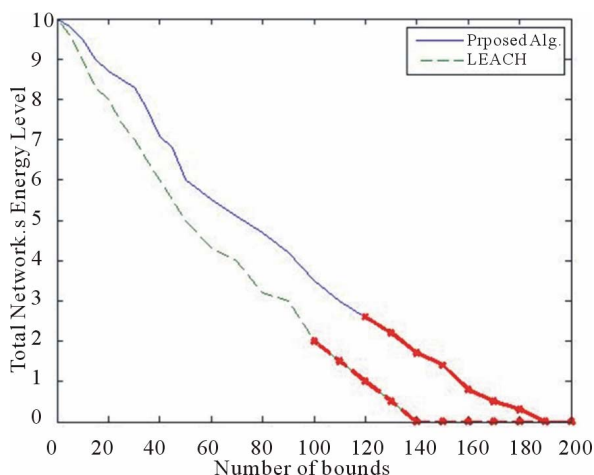


Figure 7. Energy consumption rate in lifetime of virtual environments.

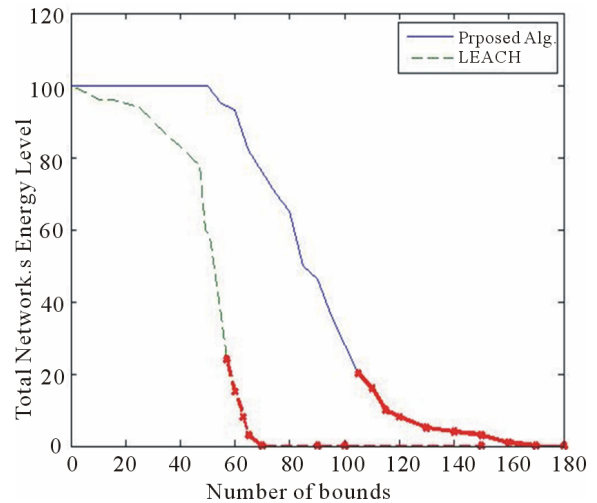


Figure 8. Comparison of coverage in two methods.

conduct the sensor nodes to the optimum directions, considering amount of remained energy leading to maximum environment power consumption metrics simultaneously. In this work, considering some priorities such as remained energy and packet size, we formed all routes which optimum one will be selected by GA. Then BS schedules all nodes throughout the selected rout and sends them their schedule. The simulation confirms that proposed fitness function fulfills these objectives compare to other works [11]. In future, the fitness function and other GA parameters can be improved to present more efficient algorithm. Furthermore, we would like to investigate tree which can efficiently adapt with network structure.

REFERENCES

- [1] K. Kalpakis, K. Dasgupta and P. Namjoshi, "Maximum Lifetime Data Gathering and Aggregation in Wireless Sensor Networks," *IEEE International Conference on Networking*, Singapore, 27-30 August 2002, pp. 685-696.
- [2] K. Dasgupta, K. Kalpakis and P. Namjoshi, "An Efficient Clustering-Based Heuristic for Data Gathering and Aggregation in Sensor Networks," *IEEE Wireless Communications and Networking Conference*, New Orleans, 20 March 2003, pp. 1948-1953.
- [3] S. Hussain and O. Islam, "An Energy Efficient Spanning Tree Based Multi-Hop Routing in Wireless Sensor Network," *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, Hong Kong, 11-15 March 2007, pp. 4383-4388. [doi:10.1109/WCNC.2007.799](https://doi.org/10.1109/WCNC.2007.799)
- [4] H. O. Tan and I. Körpeoğlu, "Power Efficient Data Gathering and Aggregation in Wireless Sensor Networks," *SIGMOD Record*, Vol. 32, No. 4, 2003, pp. 66-71. [doi:10.1145/959060.959072](https://doi.org/10.1145/959060.959072)
- [5] O. Islam and S. Hussain, "Genetic Algorithm for Data Aggregation Trees in Wireless Sensor Networks," *3rd*

- International Conference on Intelligent Environments*, Ulm, 24-25 September 2007, pp. 312-316.
[doi:10.1049/cp:20070386](https://doi.org/10.1049/cp:20070386)
- [6] S. Jin, M. Zhou and A. S. Wu, "Sensor Network Optimization Using a Genetic Algorithm," *Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics*, Orlando, 30 March-2 April 2003, pp. 109-116.
- [7] K. P. Ferentinos, T. A. Tsiligiridis and K. G. Arvanitis, "Energy Optimization of Wireless Sensor Networks for Environmental Measurements," *Proceedings of the International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA)*, Giardini Naxos, 20-22 July 2005, pp. 250-255.
[doi:10.1109/CIMSA.2005.1522872](https://doi.org/10.1109/CIMSA.2005.1522872)
- [8] F. G. Nakamura, "Planning to Control Dynamic Coverage and Connectivity in Wireless Sensor Networks Planas," Master's Thesis, Federal University of Minas Gerais, Belo Horizonte, 2003.
- [9] D. Goldberg, B. Karp, Y. Ke, S. Nath and S. Seshan, "Genetic Algorithm in Search, Optimization, and Machine Learning," Addison-Wesley, Boston, 1989.
- [10] F. P. Quintão, F. G. Nakamura and G. R. Mateus, "A Hybrid Approach to Solve the Coverage and Connectivity Problem in Wireless Sensor Networks," *4th European Workshop on Meta-Heuristics: Design and Evaluation of Advanced Hybrid Meta-Heuristics*, Nottingham, 3-4 November 2004, pp. 1-5.
- [11] A. P. Bhondekar, R. Vig, M. L. Singla, C. Ghanshyam and P. Kapur, "Genetic Algorithm Based Node Placement Methodology for Wireless Sensor Network," *Proceedings of the International Multiconference of Engineers and Computer Scientists*, Hong Kong, 18-20 March 2009, pp. 106-112.