

A Genetic Algorithm-Based Approach for Energy-Efficient Clustering of Wireless Sensor Networks

A. Zahmatkesh and M. H. Yaghmaee

Abstract—In this paper, we propose a Genetic Algorithm (GA) to optimize sensor nodes' energy consumption. Attention is paid to the clustering technique as an efficient way for reducing energy consumption of a sensor node as well as the cost of transmission. We use a multi-objective algorithm that generates an optimal number of sensor-clusters with cluster-heads and minimizes the cost of transmission. The components are thereafter used and the average fitness of the system is evaluated. The effectiveness of our algorithm is shown by simulation.

Index Terms—Sensor nodes, clustering, cluster-head, energy consumption; cost of transmission.

I. INTRODUCTION

Recently, Wireless Sensor Networks (WSNs) are becoming an essential part of many application environments that are used in military and civilians. A sensor network is composed of a large number of sensor nodes, which are densely deployed either inside the phenomenon or very close to it [1].

One of the most important features in WSNs belongs to the limited battery of sensor nodes. When battery-powered wireless sensor nodes are placed in a specific field, it is difficult to replace their batteries or supply additional energy. Furthermore, if one sensor node consumes completely its energy, part of the network may disconnect [2].

Data gathering is a common but critical operation in many applications of WSNs, while data aggregation and hierarchical mechanism are widely used techniques. Data aggregation can eliminate data redundancy and reduce communication load. Clustering mechanisms are effective means for managing such high population of nodes and can help reduce the nodes' energy consumption [3]. Some efforts in energy-efficient clustering are as follows.

Heinzelman *et al.* [2] have proposed a so-called Low Energy Adaptive Clustering Hierarchy (LEACH) as the first clustering protocol. In this protocol, data collection area is randomly divided into several clusters while each cluster has a cluster-head and some cluster members. Cluster members transmit data to their own cluster-heads, which aggregate and transmit data to the base station. Later, Heinzelman *et al.* [4] extended their work and represented a LEACH-C protocol.

In this clustering method, the base station makes sure that only nodes with enough energy are participating in.

A protocol named HEED has been proposed by Younis and Fahmy [5]. The protocol selects cluster-heads through $O(1)$ time iteration according to some metrics and adopts the multi-hop communication to reduce energy consumption.

The transmission cost in WSNs is regarded as the energy consumption of the nodes and when we can minimize this value, energy is saved. Fig. 1 shows a data delivery example in a cluster based network [1]. As shown, the calculation of transmission cost can be divided into two parts. The first part is the cluster members that transmit data to their cluster-heads. The second part, however, is the cluster-heads that transmit aggregated data to a sink node.

The Genetic Algorithm (GA) is a technique for randomized search and optimization and has been applied to a wide range of studies [6]-[7]. A basic operation flow of GA includes creating initial population, evaluating fitness, selection, crossover, mutation, updating optimal chromosome, and checking termination condition that are shown in Fig. 2.

A GA starts with a set of randomly generated possible solutions, called a population. Each individual solution, in the population, is known as a chromosome or an individual. Each chromosome may be represented as a simple string or an array of genes, which contain a part of solution. The values of genes are called alleles. The length of chromosomes in a population should be the same. A fitness function is provided to assign the fitness value for each individual.

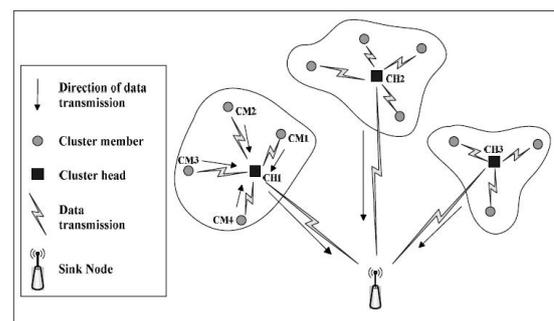


Fig. 1. Data delivery in a cluster based network.

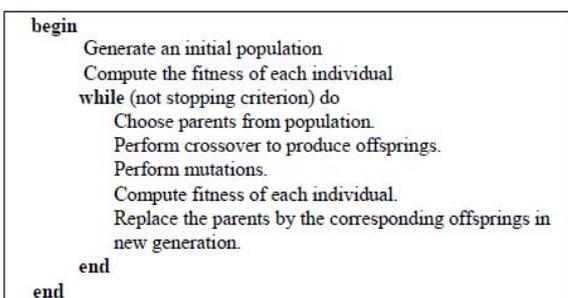


Fig. 2. Genetic Algorithm

Manuscript received January 2, 2012; revised February 29, 2012.

The authors are with the Department of Computer Engineering, Ferdowsi University of Mashhad, Iran. (e-mail: zahmatkesh.afsane@stu-mail.um.ac.ir; e-mail: hyaghmae@ferdowsi.um.ac.ir).

This function is based on how close an individual is to the optimal solution. The higher the fitness value, the closer is the solution to the optimal solution. Two randomly selected chromosomes, known as parents, can exchange genetic information in a process called recombination or crossover. This produces two new chromosomes known as child or offspring. If both the parents share a particular pattern in their Chromosome, the same pattern will be carried over to the offspring. To obtain a good solution, mutation is often applied on randomly chosen chromosomes after the process of crossover. Mutation helps restore any lost genetic values when the population converges too fast. Once the processes of crossover and mutation have occurred in a population, the chromosomes for the next generation are selected [8].

GA can be effectively used in searching for optimal clusters. In the following, we explain an energy-efficient clustering technique using GA that has been developed so far.

Khanna et al. [9] have proposed a method to select a cluster-head that can minimize the maximum intra-cluster distance between itself and the cluster members. So, they optimize energy consumption of the network.

Islam et al. [10] have represented a GA that creates energy-efficient clusters for routing in wireless sensor networks. In order to evaluate the energy consumption, they used a radio model. Moreover, they assumed a fitness function with some fitness parameters such as direct distance to sink, cluster distance, transfer energy, and number of transmissions. Proposed algorithm maximizes the fitness function which is more energy efficient than a few existing cluster-based routing protocols.

The problem of energy-efficient clustering in WSNs has been concerned by Zhang et al. [11]. They proposed a novel clustering algorithm which improves the rate of data aggregation in the network. By ensuring that only nodes with a sufficient energy are selected as cluster-heads, they greatly decreased the consumed energy.

Seo et al. [12] considered the distance for data transmission between sensor nodes that evaluates the fitness of a network. In their study, the actual energy consumption of the sensor nodes was varied depending on defining cluster-heads and battery status of sensor nodes.

In all of these studies only limited factors have been concerned during the optimization process. In the present paper, we propose a multi-objective algorithm that generates an optimal number of sensor-clusters with cluster-heads and decreases the cost of transmission. Both of the objectives lead to energy saving in sensor nodes.

The rest of the paper is organized as followed. In section II, we explain the proposed multi-objective algorithm. Our simulation results are presented and discussed in section III. Finally, we conclude in section IV with a brief discussion on the future work.

II. OUR APPROACH

In this section, we explain our proposed GA algorithm. Diagram of our fitness function evaluation in our multi-objective algorithm for a parent is shown in Fig. 3. Clearly, two fitness functions are defined in this algorithm.

One of them evaluates total node fitness for each system of sensor networks that is going to be maximized here. The function generates an optimal number of sensor-clusters with cluster-heads and includes 3 factors such as cluster-head fitness. The other one evaluates the cost of transmission for each system of sensor networks that is going to be minimized here.

A. Node Selections Chromosome Representation

Each individual sensor node is represented by a 3-bit binary number called 'gene'. These three-bit genes which define the features of a node are represented as follows:

- 000 - Node inactive (powered off).
- 001 - Node (active) chosen as cluster-head.
- 100 - Node (active) chosen as sensor.

For example, if a system or parent includes 26 sensors, is represented by 78 bits, this string's system takes a form that is indicated in Fig. 4. We specify clusters in each string by using our clustering approach as define in Section II-D.

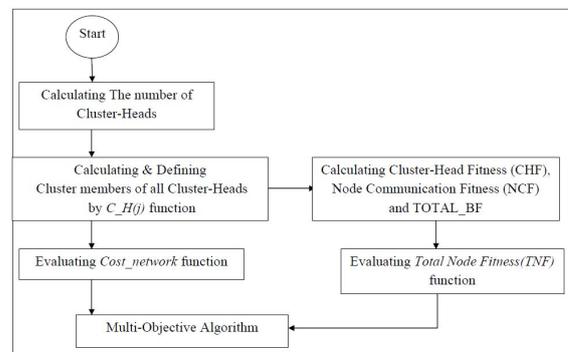


Fig. 3. Fitness Function Evaluation

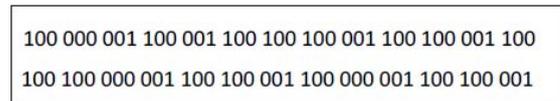


Fig. 4. A parent with 26 sensors.

B. Initial Population

Initial chromosomes' strings are seeded partially randomly using a random number generator and it includes 100 sensor nodes while each system is composed of 25 sensor nodes. The population is coded with gene structure as defined in Section II-A.

C. Node Distribution and Battery Status

Nodes are placed at random positions in a 30×30 space. Each of the nodes picks up a random coordinate between (0, 0) and (30, 30) that assigns a random battery capacity between 0 and 15.

D. Our Clustering Approach

We can recognize each kind of sensor nodes by representing bits. We consider energy level of each cluster-heads ignore those with insufficient energy for transmitting data.

In clustering algorithms, each cluster has a cluster-head while cluster members are represented by inactive/active sensors. To distribute sensor nodes between the cluster-heads

fairly, we define a ρ parameter as

$$\rho = \text{Totalsensornodes} / \text{Totalcluster-heads} \quad (1)$$

Obviously, ρ is the average number of nodes per cluster in a system.

Additionally, we generate costs or power of transmission between sensor nodes and cluster-heads. The cost is based on their distance and is expressed by a PL function as [9].

$$PL(d) = PL_0 + 10\mu \log_{10}(d/d_0) + S \quad (2)$$

Here, d is the distance between the sensors. d_0 is a reference distance typically chosen as 1m for sensor networks. PL_0 is the path loss at d_0 . μ is the path loss exponent that is in the range of 2 ~ 4 and S is a zero mean Gaussian random variable. This variable gives the deviation in path loss from its average value.

Then, by introducing a function that depends on some factors, the number of members and sensor nodes as members of every cluster are chosen. These factors are the costs of transmission between sensor nodes and each cluster-heads, the value of ρ , and energy level of each cluster-heads. The function is expressed as

$$PL(d) = F(PL(jk), \rho, \text{Energy of each cluster-heads}) \quad (3)$$

Here, j and k are the cluster-head number and the number of sensor nodes, respectively. $C_H(j)$ represents the number of members and sensor nodes as members of cluster-head(j).

E. Node Selection

The node selection evaluates the fitness of the network or the performance of the system. The fitness parameters affect the performance of the system and therefore we try to consider the most important factors in this function. We use the following fitness parameters.

1) Cluster-Head Fitness (CHF):

We've already explained the distribution of the sensor nodes between the cluster-heads in a fair manner. This prevents cluster-head overloading. CHF defines the fitness based on the uniformity of the sensor nodes and cluster-heads:

$$CHF = 1 - \min(1, \sum_N (\rho_N - \rho) / \rho) / NH \quad (4)$$

where N is the cluster-head number. NH is the number of cluster-heads in this system. ρ_N is the number of nodes attached to this cluster-head. We've used ρ parameter to select the cluster members for each cluster.

2) Node Communication Fitness (NCF):

As discussed previously, a sensor node consumes energy to communicate with other sensor nodes as well as the cluster-heads. Based on the PL function, we define p_{ij} that represent the needed power to communicate between nodes i and j . Then, the NCF function is obtained as

$$NCF = 1 - \min(\sum_i \sum_j (\max(0, p_{ij} - p_i) / p_i) / n) \quad (5)$$

p_i represents energy threshold and n is the number of sensor-nodes in this system.

3) Battery Status Fitness (BF):

As already mentioned, we generate battery status for every

sensors randomly. Whenever a sensor node communicates with other nodes (whether sensor nodes or cluster-head), we evaluate its new battery status. For a cluster-head, depending on its cluster members and its cost of transmission to the sink, we calculate its battery capacity. For the other sensor nodes that depends on the cost of transmit to its cluster-head its battery capacity is decreased. Whenever a sensor node consumes all its battery capacity, we consider it as a node inactive. Also, we define a new parameter as named TOTAL_BF. It goes back to the total battery of all sensor nodes in a network.

4) Total Node Fitness (TNF):

TNF is the final fitness that is evaluated in the GA algorithm for an appropriate node assignment. It is described by

$$TNF = \alpha_1 CHF + \alpha_2 NCF + \alpha_3 \text{Total_BF} \quad (6)$$

where $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and α_i depends upon the relative significance of the components.

F. Cost Network Function

The other objective that we want to optimize is the transmission cost. As already mentioned, in sensor networks, the transmission cost is divided into two parts: the cost to communicate between sensor nodes and cluster-heads and the cost to transmit from the cluster-heads to the sink. We express them as dCH and dSN, respectively. Cost_network function is going to be minimized.

$$\text{Cost_network} = \sum_i \sum_j d_{CH(i,j)}^2 + \sum_i d_{SN(i)}^2 \quad (7)$$

where i is the number of cluster-heads and j is the number of cluster members managed by their cluster-head numbered i . Therefore, it is important to find a proper number of cluster heads for organizing low-cost clusters. TNF and its factors help reach this aim. The cost is used in the selection phase and in the evaluation of an optimal chromosome at the current generation.

G. Selection

Selection of individuals is carried out using the Roulette-Wheel selection method. In this method the probability of being selected increases with the fitness value of the individual chromosome. In addition, in order to assure that the best individual of each generation was not destroyed by the crossover and mutation operators during the evolution process, elitism was included in the algorithm. It means that the current best individual at each generation of the algorithm always survived to the next generation.

H. Crossover

The types of crossover and mutation are of major importance to the performance of the GA optimization. To produce new generation from the selected parents, we choose multiple cross-over points. Crossover is also applied with some specific probabilities. These probabilities are tuned after proper experimentations.

I. Mutation

The mutation is an exploration process which randomly transforms genes to overcome the limitation of the crossover. In this paper, this operation enables the search for optimal

chromosome by transforming a cluster-head to a cluster member and a cluster member and a cluster-head, with a small probability. In this paper, the probability of transforming from cluster member to cluster-head is set greater than that of the opposite case for preventing abnormal increase of cluster-heads. After executing the crossover and mutation, clusters should be reconstituted since the cluster-heads' positions could have been shifted.

III. SIMULATION RESULTS

In this section, we present the simulation results as the performance evaluation of our proposed algorithm. The performance of GA is greatly affected by a number of factors, such as the population size, the probability of mutation and crossover, and the method of replacement. We used MATLAB for our simulations. We have run a number of experiments with different values of these parameters to determine the optimal set for our network size. Tables 1 and 2 indicate the WSN parameters and final GA parameters, respectively.

At the first, GA is run for TNF function. As seen in Fig. 5, CHF and NCF increase with the number of generations for 100 nodes and they are converged to 1. It shows optimal number of cluster-heads are generated and the distribution cluster members between cluster-heads is uniform. Fig. 6 indicates that total battery of the sensor network is maximized. For these three parameters, the experiments are simulated with the number of generations for 100 nodes.

Experimental setup in Fig. 7 consists of 100, 200, 400 and 600 nodes. In this simulation, the total node fitness is evaluated. As seen in this figure, convergence points depend upon the number of nodes being optimized. As already discussed, by increasing TNF increases, appropriate nodes are assigned and the nodes' energy consumption is reduced.

TABLE I: WSN PARAMATERS

Parameter	Value
Node distribution	30*30 m ²
Initial population size	4
Bit for representing every sensor nodes	3
Battery capacity	(0,15)
Location of sink node	(0,0)
Number of sensor nodes	100, 200, 400, 600

TABLE II: GA PARAMATERS

Parameter	Value
Mutation rate	0.004
Crossover rate	0.7
Max_Generations	1000

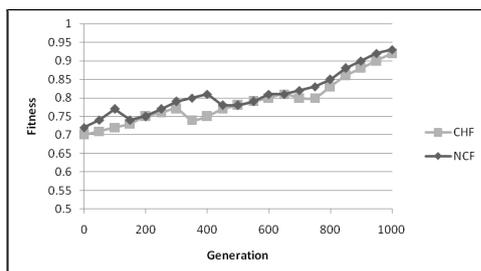


Fig. 5. Fitness chart for CHF/NCF for 100 nodes.

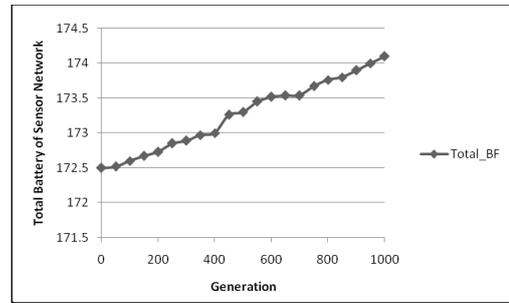


Fig. 6. Total battery of sensor network for 100 nodes.

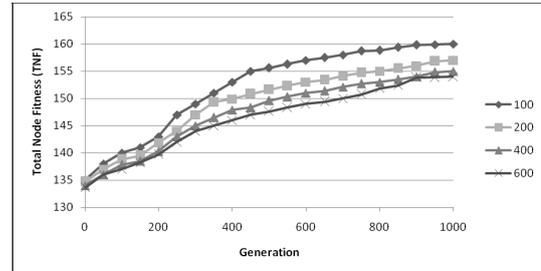


Fig. 7. TNF chart for 100, 200, 400, and 600 nodes.

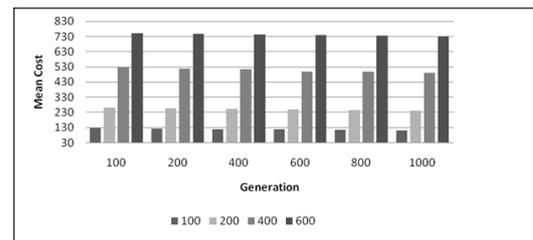


Fig. 8. Mean costs of the proposed algorithm for 100, 200, 400 and 600 nodes.

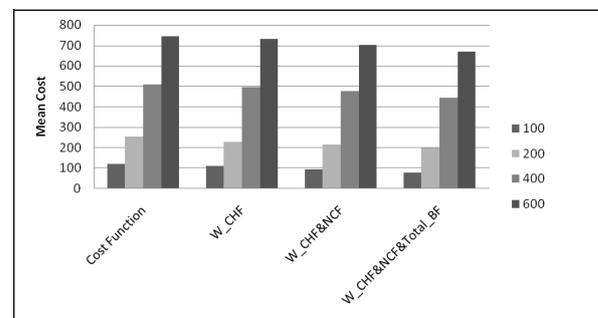


Fig. 9. Total fitness chart for 100, 200, 400, and 600 nodes.

The second objective for energy-efficient clustering is to minimize the cost of transmission. This objective without the first objective with GA is run. For evaluating the performance of proposed algorithm in various node densities, simulation is performed for 100, 200, 400 and 600 nodes. Fig. 8 shows the mean cost of the proposed algorithm. Clearly, the cost of transmission with using the Cost_network function is decreased.

Fig. 9 shows that the first objective and its factors are selected appropriately. In this case, the multi-objective algorithm is run. The result of Cost_network function is compared with some conditions as follows:

First objective with CHF (W_CHF)

First objective with CHF and NCF (W_CHF&NCF)

First objective with CHF, NCF and Total_BF (W_CHF&NCF& Total_BF)

Fig. 9 represents the cost of transmission is minimized with adding the factors in the first objective. Evidently, the cost decreases with growing the number of generations up. Moreover, it is demonstrated that with increase in the nodes density, the performance of the system improves. This can be attributed to the fact that as the number of nodes increases, it becomes less likely to find a better population by luck.

IV. CONCLUSION

In this paper, we proposed a multi-objective Genetic Algorithm to search energy-efficient clusters for sensor networks. Total energy consumption is closely concerned with the number of cluster-heads and their position. Therefore, it is important to find out an energy-efficient clustering method. So, as the first objective, we tried to generate an optimal number of cluster-heads and optimize the number of cluster members of each cluster-heads. In regard to the second objective, the distance for data transmission between sensor nodes was considered to evaluate the fitness of a network. By this function, we minimized the cost of transmission in our network. Finally, the experimental results showed our approach can be concerned as an effective energy-efficient clustering method. It was demonstrated that our algorithm generates optimal clusters and minimizes the cost of transmission.

As a part of future research, we will continue to work on improving the routing mechanisms in clustering. If this factor is considered in our algorithm, more practical and energy-efficient clusters can be achieved.

REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, March 2002.
- [2] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. of the 33rd Annual Hawaii International Conference on System Sciences*, pp. 3005-3014, January 2000.
- [3] A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, vol. 30, pp. 2826-2841, October 2007.
- [4] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. on Wireless Communications*, 2002, 1(4), pp. 660-670.
- [5] O Younis and S Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, pp. 366-379, 2004.
- [6] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison Wesley, Reading, MA, 1989.
- [7] K. Sastry, D. Goldberg, and G. Kendall, Genetic Algorithms. Chapter 4 of *Introductory Tutorials in Optimization and Decision Support Techniques*, (eds. E. Burke and G. Kendall), pp. 97–125, Kluwer, 2005.
- [8] M. Mitchell, *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, 1996.
- [9] R. Khanna, H. Liu, and H. H. Chen, "Self-organization of sensor networks using genetic algorithm," *IEEE ICC '06*, pp.3377-3382, June 2006.
- [10] O. Islam, S. Hussain, and H. Zhang, "Genetic algorithm for energy efficient clusters in wireless sensor networks," *IEEE ITNG '07*, pp. 147 – 154, 2007.
- [11] J. Zhang, Y. Lin, C. Zhou, and J. Ouyang, "Optimal model for energy-efficient clustering in wireless sensor networks using global simulated annealing genetic algorithm," *IEEE IITAW '08*, pp. 656 – 660, 2008.
- [12] H. Seo, S. Oh, and C. Lee, "Evolutionary genetic algorithm for efficient clustering of wireless sensor networks," *IEEE CCNC 2009*, pp. 1 – 5, 2009.