Unequal clustering and scheduling in Wireless Sensor Network using Advance Genetic Algorithm

Ruwaieda M Yas 1, Soukaena H Hashem2

1 Information Institute For Postgraduate Studies, Iraqi Commission for Computer & Informatics, Iraq
2 Computer Science Department, University of Technology, Iraq

roueidam@yahoo.com

Abstract. Due to limited energy of the sensor nodes which forming the wireless sensor network (WSN), and almost deployed in harsh environment, therefore it is very important to minimize the consumption of the sensor node’s energy. For this reasons designing an energy efficient clustering and scheduling of sensor nodes considered the most efficient methods for extending the WSN’s lifetime. In this paper, we have proposed Genetic algorithm based methods for clustering and scheduling. The proposed methods have two stages; in the first stage GA is used for cluster formation where the chromosome is represented by using the sensor node’s position. While in the second stage GA is used for selecting a minimum number of nodes while maintaining the full coverage and the connectivity of the selected nodes. The simulation result shows that the proposed algorithm (AGA) is more efficient than the existent algorithms in terms of number of first node die per round, number of a live nodes, and energy consumption.

1. Introduction

In the ongoing years, wireless sensor network (WSN) have been popular for their applications that increased in the environment like, health care, disaster alarm system, agricultural, surveillance, etc. [1]. In such applications, a large number of wireless sensor nodes must deployed in a harsh environment to perform certain tasks [2]. These sensor nodes are small in size, limited in their sensing and transmission range, limited in their energy resource, memory, and cannot be replaced; for these reasons maximizing the WSN life time and minimizing energy consumption will be the main issues in WSN [3].

In order to deal with the WSN challenges, many energy efficient approaches have been proposed, which include energy efficient MAC layer protocol, clustering, routing, scheduling, etc. [4-6]. In cluster based protocol the sensor nodes are grouped to may clusters, with a predefined number of the clusters, where only the cluster head (CH) is responsible for communicating with the sink/base-station. It’s one of the efficient approach that used for energy-efficiency enhancement by minimizing and balancing the consumption of the sensor nodes energy [7]. In sensor node scheduling, a few number of the total deployed nodes are activated for a few rounds for monitoring the target region, while the remaining nodes will be in a sleep mode, this method will preserved the energy of the nodes and this will prolong the network life-time [8]. As mentioned before, the nodes has limited range in their sensing and transmitting, and the nodes can communicate only if they are within the same
transmission range, for this reason another problem arises which is how can ensure the full-coverage of the target region and the connectivity between the nodes and the sink (base-station) [8].

In order to solve the above problems, this paper proposed an energy efficient scheduling and clustering protocol which based on genetic algorithm (GA) for reducing the sensor nodes residual energy and increasing the network life-time.

The rest of this paper will organized as follows: section 2 present the related work. Section 3 shows an overview on GA. Section 4 describe the proposed work. Section 5 provide the result. Finally Section 6 concludes the paper and provide the future work.

2. Related work
In WSN there are many energy efficient clustering approaches have been proposed for maximizing the network life-time. LEACH (Low Energy Adaptive Clustering Hierarchy) was the first clustering approaches which was proposed by Heinzelman et al. [9]. In this approach the cluster-head (CH) is chosen based on generating random numbers between 0 to 1. The selected numbers are then compared with predetermined threshold (T). If the selected random number is less than (T) the node will become a CH. This disadvantage of this protocol is the single hop (one-hop) communication between CHs and the sink (Base-station). An improved version of LEACH which was called LEACH-C (Low Energy adaptive clustering hierarchy-Centralized) [10]. The CHs selection of this approach depends on addition parameters like nodes location and the remaining energy. These parameters lead select the better CHs, where the parameters are sent to the sink, then the sink calculates the average energy of the node, the nodes are selected as a candidate CH if they have energy more than the average energy. The main drawbacks of this approach is that the CH is randomly selected from the candidate CHs, so, this approach do not always guarantee that the higher residual energy become CHs. In [11] another improved version of LEACH proposed by Lindsey and Raghavendra which called (Power-efficient gathering in sensor information) (PEGASIS), in this routing protocol a good results can be achieved with regarding of balancing the energy and prolong the network-lifetime. But the main problems of this approach are the following: first part, the generating of long chain between the neighboring nodes, because the chain is depending on the closest neighbors. Second part, the early death of some nodes during the rotation mechanism of the cluster-head, particularly those are far from sink, as a result of long transmission distance. Third part, the communication overhead will be maximize due to the resection of the chain-head after each round. There would be many cluster based routing algorithm which developed for WSN depending on evolutionary algorithms. Chakraborty et al. in [12] proposed a new routing protocol GROUP which is form a chain for communication with the sink. In this approach each nodes is allowed for a non-periodical message transmission with the sink. This transmission is based on the remaining energy and the location of each sensor nodes, for this reasons the network life-time will be increased and the greedy chain will be avoided. In [13] Ataul et al. developed a new WSN routing protocol which is based on GA for the data transmission between the getaways in the two tier WSN. This approach used Roulette Wheel for the individual selection, and the fitness function is based on the network life-time in term of each rounds. A new routing approach which based on GA was proposed in [14] for improving the life time if the network, it is called GAR. This algorithm used the overall distance in each round for defining the fitness function. Gupta and Jana in [15] developed a new clustering and routing approach which based on GA. The clustering phase is depending on two factors, the remaining energy of the gateways, and the distance between the sensor nodes and their cluster head. While the routing phase is depending on the remaining energy of the gateways as well as make a trade of between the forwards’ number and the distance of the transmission.

However, the above algorithms are not take in consideration the scheduling of the sensor nodes in the region of interest with cluster formation, so the main contribution in this paper are as follows:
- Use a proposed GA for the cluster formation.
- Use the GA for solving the multi objective problem of the scheduling for providing the full coverage, maintaining the connectivity, and minimize the number of the active sensor nodes.
3. Overview of GA

The Genetic Algorithm is an Evolutionary Algorithm which used an Adaptive Strategy and a Global Optimization technique for finding a global optimum answers [16]. It was developed to simulate the principle of “the survival to the fittest” that found by “Darwinian evolution” and “Mendel’s theory of genetics”. The search in GA always starts with a set of solutions, which called population. Each solution in the population is represented as a chromosome [17]. The evolution usually starts from a population of randomly generated solutions, and is an iterative process, with the population in each iteration called a generation [18]. The next generation which contain the new chromosomes is created by using the reproduction operators (crossover and mutation). The chromosome’s suitability is defined by fitness function. The higher fitness value means the fitter chromosome for the next generation (i.e. has higher chance of surviving). The selection mechanism is depending on the fitness function for selecting the fittest chromosome that will survive for the next generation. This process will repeated until stopping condition is achieved. The stopping condition could be either “a predefined number of iterations of the executing algorithm or convergence during a predefined number of iterations” [16].

The main design issue that must take into consideration during GA development is the chromosome’s encoding representation. There are many encoding scheme for the chromosome like gray coding, real coding and the Hillis’s diploid binary encoding scheme, etc. but binary encoding is consider the first and popular in GA [17].

The basic steps in Genetic Algorithm (GA) is described as follow [19][20][22]:

a. Initialization: the first step for starting the search in GA is the initialization a population of solutions which is randomly generated for the problem. The size of this population is depending on the problem’s nature.

b. Selection: it is the reproduction operator which used to select the chromosomes that will survive to form the next generation’s population. The common selection schemes is listed below:
   i. Roulette-Wheel selection: it is called fitness based which is depend on the fitness function for selecting the individuals. The chromosome with a high fitness function has a higher chance to be selected for the next generation. The probability for selecting the chromosomes (x) for the next generation is denotes as
   \[ P(x) = \frac{f(x)}{\sum_{i=1}^{n} f(x)} \]  
   Where \( f(x) \) is the fitness value of the chromosome x.
   ii. Rank-based selection: in this type of selection, the individuals are sorted according to their fitness values. The probability for selecting the individual is depending on their positions in this sorted list.
   iii. Tournament based selection: the selection of the individuals will be randomly, then a tournament will carried out among these individuals for selecting the best for the next generation.

c. Crossover: is the process of combining (mating) the chromosomes to produce a new chromosomes. The types of crossover are:
   i. One-Point Crossover: is select the bit location is randomly which need to change.
   ii. Two-point Crossover: is select 2- position and the bit between 2-positions are change only.

d. Mutation: is the process of selecting one or more genes in the chromosome randomly and alter them for producing best chromosomes for the next generation. This process is used to increase the population diversity to overcome the effect of the local optimum trap in GA search. Figure 1 [15] and algorithm 1 show the GA process [16][23][24].

e. Replacement: possibly defined as a process that exchange part of the population with the newly generated individual for next generation search.

Genetic Algorithm can be used clusters building in wireless network sensor. Routing based on GA will create a tree that collect all sensor nodes, and it will extend the lifetime of the network [21][25].
4. The proposed work

4.1. GA - Clustering
The sensor nodes (N) are deployed in the 2-D space, so each node (x) can be represented by (x, y) dimension where i = 1, 2, 3, ..., n. The clustering process partitions the deployed nodes to number of clusters (K), i.e. Cj where j = 1, 2, ..., k depending on some similarity metric. The main goal here is to minimize the consuming energy of the sensor nodes.

The steps of the GA-based clustering can be described as follows:

1. The chromosome representation
Each chromosome represented as an integer numbers which represent the center of the clusters. The length of the chromosome is the number of the selected clusters. Each gene in the chromosome represent the cluster center.
Example 1: let K = 4, i.e. the number of the clusters is four i.e. CH = C1 C2 C3 C4, and the chromosome will be as follows:
(50,60) (31,22) (15,29) (32,40).
(50,60) is the first cluster’s center where 50 represent the value of x-axis the 2-D area while 60 represent the value of y-axis in 2-D the area , (31,22) is the second cluster’s center, (15,29) is the third cluster’s center, (32,40) is the fourth cluster’s center.

2. Initialization of the population
Each chromosome are initialized randomly, the K cluster centers (Z1, Z2, ..., Zk) are selected randomly from the deployed nodes (X1, X2, ..., Xn). This operation is repeated for number of population (P).

3. The Cluster formation
The cluster is created according to the cluster centers that selected randomly in the chromosome. Where each nodes (Xi) are assigned to one cluster (Ci) in the chromosome if and only if
\[ \|Xi - Zj\| < \|Xi - Zp\|, \quad i = 1, 2, ..., n, j \neq p = 1, 2, ..., k \text{ but } j \neq p \]
Where Xi is the location of the nodes, Zj is the center of the clusters in the chromosome.
After the cluster formation, each cluster center in the chromosome are replaced by the mean of nodes in this cluster. The new cluster center is computed by using the following equation:
\[ Zj^* = \frac{1}{n_j} \sum X_j \]
(2)
Where \( X_j \in C_i \), i = 1, 2, ..., K, and \( n_j \) is the number of nodes that belong to the \( C_i \) cluster.

4. The Fitness computation
The fitness function f is computed as follows:
\[
\begin{align*}
(\text{max}) f &= 1/m \\
m &= \sum_{i=0}^{K} m_i \\
m_i &= \sum_{X_i \in C_i} ||X_i - Z_j||
\end{align*}
\]

5. GA-operation

Selection-This paper use Roulette Wheel selection, where the chromosome with the large fitness value will survive for the next generation.

Crossover-This paper use one point crossover with a predefined probability \(P_c\).

Mutation-Each chromosome is mutated based on a fixed probability \(P_m\), the mutation is done for the selected gene by using the following equation:
\[
\nu = v \pm (\delta + \varepsilon) \cdot v
\]
Where \(v\) is the value of the gene, \(\delta\) is a random number between \([0,1]\), \(\varepsilon\) is a number \(0 < \varepsilon < 1\).

6. Termination condition

These operation will terminate after a maximum number of iterations. The best chromosome in the last generation represent the best cluster centers locations.

Algorithm (1): Genetic Algorithm based clustering

Input: (1) \(N\): number of deployed nodes \(X\).
(2) \(P_{\text{size}}\): number of chromosomes in the initial population.
(2) \(K\): number of genes in the chromosomes, i.e. number of clusters \(C\).

Output: clusters formation with best location of the cluster center

1. \(POP = \emptyset\)
2. \(Chrom = \{ C_i | \forall i, 1 \leq i \leq K \}\)
3. While (the stopping condition is not reach) do
4. For \(j = 1\) to \(P_{\text{size}}\)
5. For \(i = 1\) to \(K\)
6. For \(r = 1\) to \(N\)
7. Choose \(X_r\) randomly from \(N\).
8. Assign nodes \(X_r\) to \(C_i\)
9. End For
10. Replace the cluster center with the new cluster center.
11. Evaluate each chromosome according to fitness function.
12. Apply GA operator.
13. End For
14. \(POP = POP \cup Chrom\)
15. End For
16. Return \(POP\)
17. End

4.2. GA based scheduling

Some text. After cluster formation, sensor nodes in each clusters will be scheduled so that subset of these nodes will be activated while the remaining will be inactive. The main purpose here is to select minimum number of nodes that sense the target area with maintaining full coverage and connectivity among the active nodes. The complete processes of this phase are repeated for each cluster in the WSN.

GA- based scheduling can be explained in the following steps:

1. Chromosome representation

The chromosome represented as a string of binary, and the chromosome’s length \((N)\) represent the number sensor nodes \((s)\) in one cluster in the WSN.
CH = \{g_i | g_i \in \{0, 1\}, \forall i, 1 \leq i \leq N\}. The value of g_i represent whether the selected \(S_i\) is active or sleep in the current round.

Where:

\[
g_i = \begin{cases} 
1, & \text{if } s_i \text{ is active node} \\
0, & \text{otherwise}
\end{cases}
\]  

(7)

2. Population initialization

The creation of the population will be done by generating the chromosomes in random form. Here, a note must be mentioned that each chromosome in the population represent a complete cluster. The generated number is either 0 or 1.

3. Calculation of Fitness Function

The evaluation of the chromosomes is depending on the fitness function. The main objective of this work is to activate minimum number of sensor nodes with preserving on the full coverage of the region of interest, and connectivity between the selected these selected nodes. The selected nodes must have sufficient energy to complete there mission. So to achieve these objectives, the fitness function will be formulated as follows:

\[
\text{Fitness (max)} = \left\{ W_1 \times \left( 1.0 - \frac{1}{N} \sum_{i=1}^{N} g_i \right) + W_2 \times \frac{1}{K} \sum_{i=1}^{K} y_{cost}(S_i) + W_3 \times \frac{E_{MIN}}{E_{MAX}} \right\}
\]  

(8)

Where:

- \(N\) number of nodes in the cluster.
- \(W_1\) Are the weight value, \(0 \leq W_i \leq 1, \forall i, 1 \leq i \leq 4\).
- \(E_{MAX}\) is the maximum energy of the sensor nodes.
- \(E_{MIN}\) is the current minimum energy.
- The objective of \(\frac{1}{N} \sum_{i=1}^{N} g_i\) is to select minimum number of sensor nodes, \(g_i\) is the number of ones in the chromosome.
- The objective of \(\frac{1}{K} \sum_{i=1}^{K} y_{cost}(S_i)\) is for ensuring full coverage of the area, \(K\) is the number of nodes in one cluster, and \(R_{sen}\) is a sensing range.

\[
y_{cost}(S_i) = \begin{cases} 
+1, & \text{if } (\xi_{cov}(S_i) = \{S_i\} | g_i = 1 \& \text{dis}(S_i, S_j) \leq R_{sen}, i, j = 1, 2, ..., K, i \neq j) \\
-1, & \text{Otherwise}
\end{cases}
\]  

(9)

4. GA operators

Selection - The chromosomes with the higher fitness values are selected for the new generation. In this paper tournament selection is used.

Crossover – two point crossover is applied for producing the new chromosomes.

Mutation – in this step the gene is randomly selected and converted from (0 to 1) or from (1 to 0).

5. Stopping condition

The algorithm is stopped after number of iterations.

5. Simulation results and analysis

The proposed algorithm was tested using MATLAB R2019b with the following parameters, shown in table 1.

| Table 1. Simulation parameters. |
|---------------------------------|
| Parameters                      | Value          |
| Area                            | \(100 \times 100\) m² |
| Number of deployed nodes        | 100            |
| Initial energy of sensor nodes  | 0.3 J          |
| Location of the Sink            | (50,100)       |
| Maximum communication range     | 100            |
The performance of the proposed algorithm is evaluated in terms of network energy consumption, network-lifetime, No. of a live nodes, and first node die. The results of the proposed algorithm is compared with the LEACH, GAR. Figure 2 shows the comparison results in term of Network-lifetime, the lifetime of the network can be defined as the number of nodes that still alive during the communication rounds of the network. The result shows that the proposed algorithm outperform other algorithms that compared with them. This is due to selection of minimum number of nodes with the same sensing range to be active in each cluster during the scheduling.

In figure 3 it is clear that the proposed algorithm extends the death of its first node compared with the other two protocols. Where we noticed that the proposed algorithm is better in term of efficiency than GAR and LEACH due to our algorithm that used for cluster formation. It can observed from figure 4 that the proposed algorithm reduced the energy consumption more than the compared algorithms. This is because the merge of clustering with the scheduling algorithms.
6. Conclusion

In this paper, two algorithms which based on GA were proposed, one for the energy efficient cluster and the other for sensor nodes scheduling. These two algorithms were combined in order to minimize the consumption of the energy as well as maximize the network life-time. The first algorithm the GA was used for selecting best cluster head (CH) which ensure single-hop communication between the nodes and their corresponding CH. The second algorithm proposed an enhance GA was used by considering a trade-off between the conflicting objectives which include the selection of less number of the nodes, ensuring the full coverage of the target area, and the remaining energy of the chosen sensor nodes. The experimental results shows that the proposed algorithms provide better performance than other algorithms in terms of network life-time, a number of selected sensor nodes in each cluster, energy-consumption, and number of rounds until first node die.

References

[1] Sajid H, Abdul W Matin and Obidul I 2007 Las Vegas Genetic Algorithm for Energy Efficient Clusters in Wireless Sensor Networks Fourth International Conference on Information Technology (ITNG’07) pp 147-154.

[2] Ali P, Habib R and S Hamed J 2011 An optimal energy-efficient clustering method in wireless sensor networks using multi-objective genetic algorithm INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS Int. J. Commun. Syst. 26 14–126

[3] Jin W, Jiayi C, Sai J, and Jong H Park 2017 Energy-efficient cluster-based dynamic routes adjustment approach for wireless sensor networks with mobile sinks The Journal of Supercomputing 73(7), 3277–3290.

[4] Yigitel M A, Incel O D, and Ersoy C 2011 QoS-aware MAC protocols for wireless sensor networks: A survey Computer Networks The International Journal of Computer and Telecommunications Networking 55(8).

[5] Kuila P and Jana P K 2015 Heap and parameter-based load balanced clustering algorithms for wireless sensor networks International Journal of Communication Networks and Distributed Systems 14(4) 413–432.

[6] Kuila P and Jana P K 2014 A novel differential evolution based clustering algorithm for wireless sensor networks Applied Soft Computing 25 414–425.

[7] Niayesh G, Kamarlunizam A Bakar, Siti Z M Hashim and Ali Pourasl 2018 Energy-Efficient Intra-Cluster Routing Algorithm to Enhance the Coverage Time of Wireless Sensor Networks sensors 17(8) 1858.
[8] Subash H and Pratyay K 2019 Coverage and connectivity aware energy efficient scheduling in target based wireless sensor networks: an improved genetic algorithm based approach *Wireless Networks* **25**(4).

[9] Wendi R Heinzelman, Anantha C and Hari B 2000 USA Energy-efficient communication protocol for wireless microsensor networks *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences* vol 2 pp 10.

[10] Heinzelman, W B, Chandrakasan, A P, & Balakrishnan H 2002 An application-specific protocol architecture for wireless microsensor networks *IEEE Transactions on Wireless Communications* **1**(4) 660–670.

[11] Stephanie L and Cauligi S Raghavendra 2002 PEGASIS: power efficient gathering in sensor information systems *IEEE Aerospace Conference* pp 3-3.

[12] Ayon C, Subir K Mitra and Mrinal K Naskar 2018 A genetic algorithm inspired routing protocol for wireless sensor networks *Journal of Advance Computational Research* Vol **3** No 1.

[13] Ataul B, Shamsul W, Jaekel, Jaekel A and Subir B 2009 A genetic algorithm based approach for energy efficient routing in two-tiered sensor networks *Ad Hoc Networks* **7**(4) 665–676.

[14] Suneet K Gupta, Pratyay K and Prasanta K Jana 2013 GAR: An energy efficient GA-based routing for wireless sensor networks *International conference on distributed computing and internet technologies* vol **7753** pp 267–277.

[15] Suneet K Gupta and Prasanta K Jana 2015 Energy Efficient Clustering and Routing Algorithms for Wireless Sensor Networks: GA Based Approach *Wireless Personal Communications* vol **83** Issue 3 pp 2403–2423.

[16] Ali P, Habib R Mashhadi and S Hamed J 2013 An optimal energy-efficient clustering method in wireless sensor networks using multi-objective genetic algorithm *INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS* **26**(1) 114-126

[17] Jagdish C Bansal, Pramod K Singh and Nikhil R Pal 2019 *Evolutionary and Swarm Intelligence Algorithms* vol 779 ed J Kacprzyk, (Poland) Springer International Publishing AG, part of Springer Nature.

[18] Jason B 2011 *Clever Algorithms: Nature-Inspired Programming Recipes* 1st Edition (Australia), lulu.com.

[19] David E Goldberg 1989 *Genetic Algorithms in Search, Optimisation, and Machine Learning* Addison-Wesley.

[20] Melanie M 1996 *An Introduction to Genetic Algorithms* MIT Press Cambridge MA.

[21] Wenjing G and Wei Z 2014 Review: A survey on intelligent routing protocols in wireless sensor networks *Journal of Network and Computer Applications* Vol **38**

[22] Ruwaida M Yas 2017 Permuting Convergence Overcoming of Genetic Algorithm Using Arnold Cat Map *International Journal of Science and Research* Vol **6** Issue 5.

[23] Ahmed T Sadiq and Amaal G Hammed 2010 BSA: A Hybrid Bees’ Simulated Annealing Algorithm To Solve Optimization & NP-Complete Problems *International Journal of Engineering and Technology* **28**(2) 271-281

[24] Ahmed T Sadiq and Amaal G Hammed 2012 Exploration-Balanced Bees Algorithms to Solve Optimization and NP-Complete Problems *International Journal of Research and Reviews in Soft and Intelligent Computing* Vol **2** No. 1 pp 108-113.

[25] Amaal G Hammed and Enas M 2016 A Proposal for Escaping Local Optima in C4.5 Decision Tree by Using Explorative Search Space Guiding Through Random Search technique *Al-Ma'mon College Journal* Issue **28** pp 341-353