

Genetic algorithm based protocols to select cluster heads and find multi-hop path in wireless sensor networks: review

Mohammed Al-Shalabi^{1,*}, Mohammed Anbar^{1,*}, and Tat-Chee Wan^{1,2}

¹National Advanced IPv6 Centre, Universiti Sains Malaysia, 11800 Penang, Malaysia

²School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia

Abstract. A wireless sensor network (WSN) is a modern technology in radio communication. A WSN comprises a number of sensor nodes that are randomly spread in a specific area for sensing and monitoring physical attributes that are difficult to monitor by humans, such as temperature, fire, and pressure. Many problems, including data transmission, power consumption and selecting cluster heads, may occur due to the nature of WSNs. Various protocols have been conducted to resolve these issues. Most of the proposed protocols are based on the Genetic Algorithm as an optimization technique to select the Cluster Heads (CHs) or to find a multi-hop path for sending the data from the CHs to the Base Station (BS). This paper presents a comprehensive study of the protocols for WSNs that are proposed to come up with these issues. This study emphasises on CHs selection protocols and multi-hop path finding protocols and their strengths and weaknesses. A new taxonomy is presented to discuss these protocols on the basis of different classes. A complete comparison of the main features and behaviors of the protocols is conducted. This study will give basic guidelines for the researchers those have a motivation to develop a new CHs selection protocol or a multi-hop path finding protocol.

1 Introduction

Data Transmission is the process of sending the data from the Cluster Heads (CHs) or the sensor nodes to the Base Station (BS) [1]. By contrast, the CH selection is the process of selecting the proper nodes to be CHs. There are many protocols have been proposed to deal with these issues in WSNs using different algorithms [1,2]. The evolutionary based algorithms play important role in solving such problems due to their efficiency in finding the optimal solutions. The most popular and familiar evolutionary algorithm is the Genetic Algorithm (GA) which is adapted in this research to solve the direct transmission problem [3,4].

This paper will review and compare between the protocols that have been conducted in the last few years. This review is realized as a clear understanding for the researchers of

* Corresponding author: mshalabi@nav6.my, anbar@nav6.usm.my

different CHs selection protocols' and multi-hop path protocols' vision, their strengths and weaknesses, and the used parameters in them.

The rest of the paper will be organized as following; Section II will clarify the Genetic Algorithm. Section III reviews the related protocols based on the new taxonomy. Section IV presents a comparison between the protocols. Finally, the paper is concluded in Section V.

2 Genetic algorithm

In this section, a background of Genetic Algorithm will be presented because it is considered as an optimization method used to find optimal solutions.

GA is an optimization algorithm that simulates a nature evolution, and it uses the same methods in the natural selection to find optimal solutions [5]. GA is used in optimization and search problems to find the best solution. The main operations in GA are selection, crossover, and mutation [6]. At first, initial population is randomly generated, and then a new population will be selected from the initial population depending on the fitness function and will contribute in the generation of the population [7]. The fitness function is very important in GA and depends on the problem, it is used to evaluate the quality of individuals and select the best of them to contribute in the next generation [8]. The main operators in GA are crossover and mutation, crossover operator is used to generate new children from the parents by combination them, and these children will be better than the two parents [9]. In mutation, random swap is applied on the parents in order to form new children. The algorithm stops when a specific condition is reached [10].

One of the GA operators is crossover which is applied on two selected individuals to produce new offspring in order to vary the programming of the chromosomes [9]. The individuals are selected depending on the fitness value. There are different types of crossover such as one-point crossover and two-point crossover which can be used to exchange portions; from these individuals; that are separated by a crossover point. The crossover point is randomly selected and then the crossover operation will be applied on the two selected individuals.

The mutation operation is applied on each gene of an individual with the probability of mutation rate. Mutation affects only a single chromosome, and the chromosome which is selected for mutation will have a randomly selected gene that is called mutation point. There are many ways for the mutation operation, flipping the value of a gene is one of the ways, the value of the gene that has a position equals to the mutation point will be changed from 0 to 1 or vice versa to give a new possible path.

In the selection process, some individuals are selected from the population in the current generation based on the values of the fitness function. The individuals with smaller fitness values are selected as the individuals of the population in the next generation. The algorithm stops if the number of generations equals a specific number which represents the maximum generations.

The main drawback in the WSNs is ignoring the residual energy of the nodes when selecting them as CHs. Moreover, WSNs consider the direct transmission of data between a CH and the BS, which consumes a lot of energy and negatively affects the lifetime of a network. For these reasons, there are many routing protocols that have been proposed to enhance the WSNs respecting to maximize the network's lifetime based on a proposed taxonomy, as will describe in the following section.

3 Proposed taxonomy of GA-based protocols in wsns

In this paper, the focus is on the CHs selection and the multi-hop path finding as designated criteria to present the classification of the GA-based protocols in the WSNs. In Fig 1, the GA-based routing protocols are categorized into two categories: CHs selection-based and multi-hop path finding-based protocols.

Each main category is further classified into different classifications based on the main factor in the proposed approaches. The next subsections describe each main category in detail. Then, existing familiar protocols that belong to each main category are presented in detail.

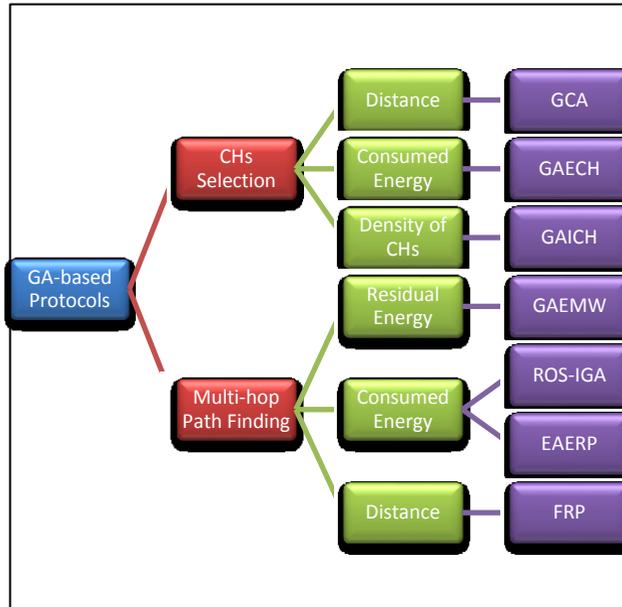


Fig. 1. Proposed taxonomy of LEACH-based routing protocols in WSNs

3.1 Multi-hop path finding protocols

This subsection presents the protocols that have been proposed to find the multi-hop path for transmitting the data from the CH to the BS based on the GA as will be described in the following paragraphs.

3.1.1 FRP (Fitness-base routing protocol):

Authors in [11] have been proposed a new protocol to improve the efficiency of data transmission by using GA. In order to find an efficient rout to transmit data from a sensor node to the BS, the fitness function uses parameters such as the distance from the sensor node to the BS, based on the fitness value, the optimal nodes are selected and then the data packet is routed in an optimized route to the BS [11]. The fitness value of a node is calculated as follows:

$$Fitness = dist(i, j) + dist(j, BS) \tag{1}$$

Where the distance between two sensor nodes or between any sensor node to the BS is calculated by Equation 1. The packet is transmitted from a sensor node to the BS through an efficient path, the intermediate nodes between the source node and the BS are selected depending on their fitness values, and these nodes route the packet to other nodes until reach to the BS [11].

In FRP protocol, the energy consumption in the transmission phase is reduced, and also the packet loss percentage is minimized by finding alternative nodes in case of node failure.

3.1.2 GAEMW (Genetic algorithm for energy-entropy based multipath routing in wsns):

In GAEMW protocol [12], authors proposed a GA for multipath routing in WSNs; the idea is to find the minimal node residual energy of each route in selecting multipath routing.

To initialize population, a routing path is encoded by a set of integers represent IDs, energy and other node's information. The first gene represents a source node and the last gene represents the BS, then several paths (chromosomes) will be created from the source to the destination. Fitness function is used to evaluate the quality of each chromosome (path) by finding the least cost path between the source and destination.

The fitness function of a chromosome C_i is defined as follows [12]:

$$f(C_i) = \frac{1}{G} \sum_{k=1}^{path} \left((E_{tr} + E_{rec}) \times P_{ki} \times |P_{ki}^{s,d}| - E_{avg} \right) \quad (2)$$

After that, the two operators in GA (crossover and mutation) are applied on the population and the next generation is created. This protocol provides efficient protocol for evaluating the route stability in WSNs.

3.1.3 ROS-IGA (Routing optimization strategy based on improved GA):

The literature that has been proposed in [13] describes an improved protocol using GA to find a suitable route between a sensor node and the BS, the new protocol called ROS-IGA. In this protocol, authors tried to solve the problem of generating invalid individuals; this problem negatively affects the efficiency of GA. In the ROS-IGA protocol, the locations of nodes and other information such as energy consumption between adjacent nodes, accumulated distance, and the remaining energy of nodes are considered in the fitness function to find suitable and practical path. The fitness function is maximization function, thus, the higher fitness value means more paths [13].

The crossover and mutation operations are improved in this protocol based on the neighboring information and the structure of the network. This improvement of crossover and mutation operations provides better routing and practical routes in WSNs.

3.1.4 EAERP (Energy-aware evolutionary routing protocol for dynamic clustering of wireless sensor network):

The idea of this protocol is to increase the stability period of the network and to increase the lifetime until the last node dies. Authors in [14] have used the concept of the centralized evolutionary routing protocol to optimize the CH selection process by the BS. The proposed fitness function includes the intercluster and the intracluster energies. The limitation in this protocol is the consideration of the single-hop transmission only without taking into account the multi-hop transmission.

3.2 Cluster heads selection protocols

This subsection presents the protocols that have been proposed to select the nodes to be CHs based on the GA. The following paragraphs describe each protocol in details with the formulated fitness functions and the considered parameters.

3.2.1 GAICH (*Genetic algorithm inspired clustering hierarchy*):

In this protocol, authors [15] have used the GA at the BS in order to select the suitable cluster heads and find the optimal number of them by applying a proposed fitness function which is based on the following parameters: density of cluster heads, distances from cluster heads to the BS, the centrality of cluster heads, and the residual energy of cluster heads. In this protocol, the authors proposed three levels of cluster heads percentage depend on the average energy of the network, and the population size depends on the number of candidate cluster heads. The limitation in this protocol is that the stability period is not achieved. In addition, this protocol doesn't expose the data transmission problem.

3.2.2 GAECH (*Genetic algorithm based energy efficient clustering hierarchy in wireless sensor networks*):

In this protocol, the authors [16] have enhanced the LEACH protocol by using the GA (run at the BS) with proposed fitness function in order to form balance clusters. The chromosomes in this protocol are presented using the binary presentation, and the single point crossover is used. The parameters which are used in the fitness function are the total energy consumption, the standard deviation among clusters in basis of energy consumption, the energy consumption in the cluster heads, and the cluster head distribution. The results showed that the performance of this protocol is better than other protocols. On the contrast, this protocol suffers from some drawbacks, such as it did not consider the distance when selecting the cluster heads, and it did not consider the routing process.

3.2.3 GCA (*A new robust genetic algorithm for dynamic cluster formation in wireless sensor networks*):

In order to increase the lifetime of the network, the authors in [17] have proposed a robust genetic clustering algorithm to balance the consumed energy within the clusters by controlling the number of them. The proposed fitness function in this protocol is presented in the following equation:

$$Fitness = W * CH + (1 - W) * D \tag{3}$$

where CH is the number of cluster heads, D is the total distance from cluster members to the CH in each cluster and W is a value based on the application. The authors of [17], based on this fitness function, tried to minimize the consumed energy by minimizing the number of CHs in the network and minimizing the total distance between the cluster members and the CH. The stability period of the network is not considered in this protocol.

4 Comparative analysis

Table 1 shows a comparison among the above mentioned protocols which proposed to select the CHs or to find the multi-hop path for transmitting the data from the CH to the BS using

different techniques based on the GA. In Table 1, some information such as the year of when the protocol is proposed, the modeling parameters, and the drawbacks and limitations of each protocol are highlighted.

Table 1. Comparison among GA-based Protocols

Protocol	Modeling Parameter	Drawbacks and Limitations
GCA (2007)	Number of CHs and distance to the BS.	-Does not consider the routing problem.
EAERP (2011)	Inter and Intra clusters energies to select the CHs.	-Consider just the single-hop transmission.
FRP (2014)	Distance from the sensor node to the BS	Fitness function just considers the distance to the BS.
GAEMW (2015)	Residual Energy of nodes	Fitness function concentrates only on the residual energy.
GAECH (2015)	The Total and standard deviation of energy, and CHs distribution.	-Does not consider the distance in selecting the CHs. -Does not consider the routing problem.
ROS-IGA (2016)	Locations of nodes, Energy consumption between adjacent nodes, Accumulated distance, Remaining energy of nodes	-Concentration on the invalid individuals, it is based on the locations of nodes which got during the deployment. -Does not consider the CH selection.
GAICH (2016)	Density of CHS, distance and residual energy.	-Stability period is not achieved. -Does not consider the routing problem.

According to Table 1, it is clearly seen that the parameters which used in the fitness function in order to optimize the data transmission problem in many protocols were not enough to find the best solution; which is the optimal multi-hop path between the CH and the BS; as the protocols that have been proposed in [11] and [12], which the authors just considered the distance from the sensor node to the BS in addition to the residual energy of nodes respectively. Otherwise, the authors [17] used GA to select the best CH without considering the data transmission path.

Other proposed protocols to select the CHs also did not consider suitable and enough parameters to select the proper nodes to be CHs. In general, gaps still exist in the proposed protocols in terms of the CH selection and the data transmission issues.

4 Conclusions

Several protocols have been discussed and proposed to solve the CHs selection and the data transmission problems in the hierarchical-based WSNs. In this paper, the common protocols are discussed and reviewed with considering the advantages and disadvantages for each protocol. The GA evolutionary algorithm was used to handle the data transmission between a CH and the BS by finding a multi-hop path between them based on some parameters such as the distance and the energy, or to select the CHs based on some parameters such as the number of CHs and the centrality of the CHs. Each of these protocols is analyzed in order to clarify the guidelines when the researchers want to propose mechanisms; these guidelines consider the gaps in the previous protocols.

References

1. M. Al-Shalabi, M. Anbar, T.C. Wan, A. Khasawneh, Variants of the Low-Energy Adaptive Clustering Hierarchy Protocol: Survey, Issues and Challenges. *Electronics* **7**, 136, (2018).
2. A. Abbasi, M. Younis, A survey on clustering algorithms for wireless sensor networks. *Computer Communications* **30**, pp. 2826-2841, (2007).
3. A. Norouzi, H. Zaim. Genetic Algorithm application in optimization of wireless sensor networks. *The Scientific World Journal*, (2014).
4. M. Kumar, M. Husian, N. Upreti, D. Gupta. Genetic Algorithm: Review and Application. *International Journal of Information Technology and Knowledge Management* **2**, pp. 451-454, (2010).
5. Q. Yang, X. Yang, Y. Wang, D. An, A strategy based on cooperative transmission for minimizing delivery delay in WSN. *International Journal of Distributed Sensor Networks* **2013**, (2013).
6. R. Zhou, M. Chen, G. Feng, H. Liu, S. He. Genetic clustering route algorithm in WSN. *Natural Computation* (**8**, pp. 4023-4026, 2010).
7. W. Liu, Y. Wu. Routing protocol based on genetic algorithm for energy harvesting-wireless sensor networks. *IET Wireless Sensor Systems* **3**, pp. 112-118, (2013).
8. S.N.Sivanandam, S. N. D. Genetic Algorithm Implementation Using Matlab. *Introduction to Genetic Algorithms*, pp. 453, (2008).
9. M. Warriar, A. Kumar. An Energy Efficient Approach for Routing in Wireless Sensor Networks. *Procedia Technology* **25**, pp. 520-527, (2016).
10. Y. Song, C. Gui, X. Lu, H. Chen, B. Sun. A Genetic Algorithm for Energy-Efficient Based Multipath Routing in Wireless Sensor Networks. *Wireless Personal Communications* **85**, pp. 2055, (2015).
11. A. Balamurugan. An Energy Efficient Fitness Based Routing Protocol In Wireless Sensor Networks. *Ictact Journal On Communication Technology* **5**, (2014).
12. Y. Song, C. Gui, X. Lu, H. Chen, B. Sun. A Genetic Algorithm for Energy-Efficient Based Multipath Routing in Wireless Sensor Networks. *Wireless Personal Communications* **85**, pp. 2055, (2015).
13. G. Yao, Z. Dong, W. Wen, Q. Ren. A Routing Optimization Strategy for Wireless Sensor Networks Based on Improved Genetic Algorithm. *Journal of Applied Science and Engineering* **19**, pp. 221-228, (2016).
14. E. A. Khalil and B. A. Attea. Energy-aware evolutionary routing protocol for dynamic clustering of wireless sensor networks. *Swarm and Evolutionary Computation* **1**, pp. 195-203, (2011).
15. Y. Lu, V. Wong. An energy-efficient multipath routing protocol for wireless sensor networks. *International Journal of Communication Systems* **23**, pp. 633-652, (2006).
16. B. Baranidharan, B. Santhi. GAECH: Genetic algorithm based energy efficient clustering hierarchy in wireless sensor networks. *J. Sensors* **2015**, (2015).
17. S. R. Mudundi, H. H. Ali. A new robust genetic algorithm for dynamic cluster formation in wireless sensor networks. *Wireless and Optical Communications*, (pp. 360-367, Montreal, Canada, June 2007).