GAPSO-H: A Hybrid Approach Towards Optimizing the Cluster Based Routing in Wireless Sensor Network

Biswa Mohan Sahoo

Amity University, Greater Noida, India

E-mail: biswamohans@gmail.com

Hari Mohan Pandey

Department of Computer Science, Edge Hill University, Lancashire, United Kingdom

E-mail: Pandeyh@edgehill.ac.uk

Tarachand Amgoth

Indian Institute of Technology (Indian School of Mines), Dhanbad, India

E-mail: tarachand@iitism.ac.in

Abstract— Wireless Sensor Networks (WSNs) have left an indelible mark on the lives of all by aiding in various sectors such as agriculture, education, manufacturing, monitoring of the environment, etc. Nevertheless, because of the wireless existence, the sensor node batteries cannot be replaced when deployed in a remote or unattended area. Several researches are therefore documented to extend the node's survival time. While cluster-based routing has contributed significantly to address this issue, there is still room for improvement in the choice of the cluster head (CH) by integrating critical parameters. Furthermore, primarily the focus had been on either the selection of CH or the data transmission among the nodes. The meta-heuristic methods are the promising approach to acquire the optimal network performance. In this paper, the 'CH selection' and 'sink mobility-based data transmission', both are optimized through a hybrid approach that consider the genetic algorithm (GA) and particle swarm optimization (PSO) algorithm respectively for each task. The robust behavior of GA helps in the optimized the CH selection, whereas, PSO helps in finding the optimized route for sink mobility. It is observed through the simulation analysis and results statistics that the proposed GAPSO-H (GA and PSO based hybrid) method outperform the state-of-art algorithms at various levels of performance metrics.

Keywords— GA-based CH selection, PSO-based sink mobility, Wireless sensor network, clustering, Energy Consumption Rate (ECR), Sink mobility.

1. Introduction

The data dissemination from the network and sending it to the sink is one of the most significant roles a sensor node plays [1]. There are many purposes ranging from military to medical care, agricultural and atmospheric monitoring and so many more [2]. Essentially, the sensor nodes consist of low battery resources, restricted memory size range, and small-scale computational capabilities. Fundamentally, in hostile areas, these sensor nodes are meant to work, so the battery installed in them is irreplaceable. It introduces one of WSN's most important concerns, i.e. the energy-efficient use of the sensor nodes [3].

Making routing as energy efficient as possible is one of the keyways to preserve the power of the nodes [4]. The clustering helps to group the nodes of the sensor that are called clusters. The cluster has one Cluster Head (CH) with different cluster functions. CH gathers information from the members of the cluster and sends it to the sink. Clustering is beneficial in the network's energy balance and includes some of the following characteristics [5]: (a) the communication bandwidth is retained as only the CH is limited to the interaction of cluster nodes; (b) the energy of both the sensor nodes is saved in such a way that the CH aggregates the data collected, the sink only collects the significant data, thereby saving the nodes energy and (c) the network's scalability is greatly enhanced as CHs keep the local route configured for the other CHs.

The current research is focused on two rational aspects: (a) selection of CH and (b) sink mobility (it is an eminent solution to the hot-spot problem). CH selection is a significant concern of optimization that can be described as a NP-hard question. There are various parameters to consider when selecting CH: residual power, distance factor, node degree and much more [6]. The thing to recognize here is the trade-off between various parameters, for example we need high node energy but low separation from the sink. The optimization methodology is therefore necessary to pick CH by combining the important parameters into a designed fitness function [7][17]. When the network area is increasing, the next problem arises, and the clusters adopt multi-hop communications to reduce energy usage. The clusters closest to the sink are encumbered with heavy traffic information while doing so. Such a condition is called a problem at the hot spot [8]. Because of this, sink is ultimately disconnected from the network. To solve this problem, the mobility of sinks has proved promising [9]. This operates in the most affordable way that manages the load in the network and increases the lifespan of the network in the delivery of energy across the sink. It is worth noting that mobile sink carriers can be public transit which follows a predefined cyclic schedule to serve

2

a specific target area. Some of the applications where mobile sinks can be used effectively are smart houses, traffic management, pollution control and various safety measures [10].

It is studied that plethora of research has been carried out in handling the CH selection as well as in optimizing the data transmission among the sensor nodes while sink is on the move. However, a sprinkling of research is reported that focused on the both concerns at the same time with any such hybrid approach. The reason for employing hybrid approach is to get the striking beneficial characteristics from both optimization techniques so that the optimal network performance could be obtained. Furthermore, the reason for selecting GA (Genetic Algorithm) and PSO (Particle Swarm Optimization) for their specific task of selecting the CH and finding the optimized route for sink mobility, is supported by the significant existing research work in the domain [30].

Motivated by the aforementioned discussion, we propose a novel hybrid approach that consider the GA and PSO algorithm respectively for 'CH selection' and 'sink mobility. In particular, the key contributions of the proposed hybrid approach formulation are as follows:

a) We propose a GAPSO-H (GA and PSO based hybrid algorithm) clustering algorithm that utilizes GA for the selection of CH and optimizes the sink mobility to acquire the optimal network performance.

b) We showed CH selection optimization through GA considering five factors, namely, energy, distance, node degree, average energy and Energy Consumption Rate (ECR). These factors play a significant role during CH selection.

c) To pacify the concern of data transmission, the sink mobility is also optimized using PSO that considers the lowers energy CH and nearest CH from the moving sink and size of the cluster.

d) The performance validation of GAPSO-H clustering is done against the competitive algorithms to claim its superiority.

A broad paper outline is given as follows: **Section 2** gives the previous work; **Section 3** presents the motivation of adopting hybrid approach; **Section 4** outlines the proposed algorithm. Experimental results are given in **Section 5** and conclusions are provided in **Section 6**.

2. Related Work

Literature reveals that the clustering started from the LEACH [29] which had distributive approach to select CH and selected CH randomly. However, the clustering improved the network performance from the pre-existing algorithms. In this section, we have comprehensively discussed

the existing work for both GA and PSO along with the different protocols of heterogeneous WSN [12]. To highlight the research gaps, existing works are summarized in where Appendix A, Appendix B, and Appendix C (*Appendix A, B and C are attached as supplementary files*) respectively presents comparative study of GA based routing protocols and PSO based routing protocols.

2.1 CH Selection in Heterogeneous Protocols

The heterogeneous protocols consider the threshold concept for selecting the CH in the network. It initiated from the SEP protocols and subsequently various protocols are developed that considered different parameters to select CH. These protocols are discussed in Appendix A that exploits them for different parameters considered for the CH selection. Verma et al. [2] exploited heterogeneous routing protocols for multi sink environment using only three parameters. However, it was believed that when there are multiple parameters incorporated, they need to optimize for a common solution. It is discerned from Appendix A that these protocols are still deprived of the energy efficient parameters for example energy consumption rate; and some are deprived from the other parameters which are not considered in one or other protocols.

2.2 CH Selection using GA

GA is one of the most prominent optimization techniques that helps in selecting CH by converging to the fittest solution and rendering the most deserving node to be selected as the CH. In Appendix A various routing techniques that are based on GA are highlighted. Kuila et al. [6] utilized GA for managing the load balancing of the network. The proposed technique didn't consider the CH selection. Gupta and Jana [7] proposed GACR that included the distance and standard deviation factors in its fitness function aiming to acquire enhanced number of rounds before the first gateway is dead.

The other protocols GAHN [8], ASLPR [9], GADA-LEACH [10] are also discussed in Appendix A. These protocols were suffering from the scalability and the factor of node density was not taken into consideration for the CH selection in these protocols. Elhoseny et al. [11] presented DCH-GA that select CH based on six parameters as shown in Appendix A. It basically targeted the multi hop scenarios unlike to other protocols. However, it suffered from the hot-spot problem and excessive overheads.

Yuan et al. [12] proposed GASONeC that selected CH by considering the parameters of expected energy consumption, distance and the node density factor. However, it suffered as: "*increased in computational complexity and non-appropriate routing when the sink is located at quite far*". Moreover, no technique is defined to deal with long communication for large network. The discerning fact that can be drawn from the Appendix A can be stated as follows: (a) Though number of techniques have been proposed working towards CH selection while using GA, but none of the technique have been conserving energy optimally. There is need to select CH in a way that the multi hop communication is also taken into consideration along with the other factors. Therefore, the current work uses ECR in fitness function of its GA implementation. (b) The current work is inspired from the data scenario topology considered in the SEECP protocol in which the sink is placed inside the network.

2.3 Clustering using PSO

We have summarized the existing literatures in Appendix B where PSO was implemented as an optimization method for routing protocols for CH selection. In PSO, the numerous particles are initialized which define the solution. Furthermore, every solution is an array of number of the nodes in the cluster. These particles are checked for their fitness values and accordingly the personal best (pbest) value for the position of particle is computed. Thereafter, the global best (gbest) value is determined among all the particles. Particle's position and velocity get updated in each iteration of the PSO utilizing pbest and gbest. This process is repeated until the termination condition (or stopping criteria) achieved.

Latiff et al. [14] proposed PSO-MSB targeting different attributes, namely, network lifetime, data delivery and energy consumption are considered in which the distance between the node and sink were considered to determine the fitness value. Hu et al. [15] proposed immune orthogonal learning PSO algorithm (IOLPSOA). However, the long-haul transmission in this method consumes a lot of energy.

Sahoo et al. [25] proposed PSO-based energy efficient clustering and sink mobility (PSOECSM) algorithm targeting heterogeneous WSN for optimizing cluster head selection method as well as sink mobility are addressed using PSO that considers the lower energy CH and nearest CH from the moving sink and size of the cluster. PSO-UFC [26] proposed PSO-based method to derive the fitness function for unequal clustering as well as fault tolerant mechanism to stability the 'intra

and inter cluster energy consumption.

To have a deeper insight of various other meta-heuristic approaches, the comparative analysis is given in Appendix C. We have given the state-of-the-methods primarily from the year 2019 to highlight the attributes and consideration by the reported studies for the CH selection and sink mobility. All in all, the existing works have been summarized in Appendix B and Appendix C.

2.4 Differences between the existing and proposed work

It is imperative to highlight the differences between the proposed work against the existing work in the context of the attributes targeted by them and their key findings from their work. These differences are stated as follow.

a) Unlike the GAPSO-H many of the algorithms did not consider the energy efficient CH selection while implementing the PSO algorithms for the sink mobility.

b) GAPSO-H addresses the hot-spot problem whereas numerous algorithms did not target the hot-spot problem which could save the energy and enhance the network lifetime.

c) The computational complexity is another factor that makes the existing algorithms non-suitable for various real time applications. However, GAPSO-H is found to be computationally optimized.

d) The proposed work considers the five novel parameters which have not been considered altogether in any other method reported so far which makes it different and significant from the rest of works.

e) The comparative studies are presented in APPENDIX A, B, and C and after studying the research works in detail, the protocol GAPSO-H is incorporated with the left out significant parameters.

In one of the most recent work reported by Gharaei et al. [31] which defines the inter and intra cluster movement of mobile sink for clustering by using PSO. However, it suffered from the following drawbacks: (a) the corona-based model is used, and the mobile sinks used fermat spiral motion to collect data which tends to ignore the data traffic due to its spiral motion; and (b) the movement of sink is kept random inside the cluster which leads to increased energy consumption. Therefore, in consideration to the two sinks employed, single sink can be employed for cost-effective network. Furthermore, the PSO based optimization is employed to optimize the sink mobility.

3. Motivation

GA is one of the pertinent stochastic optimization processes that helps in adapting the evolution procedure available in nature. In GA, the individuals are selected randomly and are contained in the population. Thereafter, through different steps the individuals are evolved viz., selection, crossover and mutation processes. After selecting the fittest individuals, the new population of individuals are generated. The reason behind selecting GAs over the other metaheuristics algorithms is the robust behavior and its characteristics to perform with parallel computations rather operating with the single point. It not only helps in acquiring the global optimal solution by avoiding the local optimal point. The applications which are derived from the derivative information, there GAs can be used to reduce the complexity by handing over the control to the fitness function. Hence, GA was applied across domains to deal with combinatorial and multi-objective optimization problems [32] [44] [45] [46] [47] [48].

It is seen that the performance of classical optimization degrades as the size of the network is increased. PSO has demonstrated the ability to resolve this type of problems. It was implemented to solve various problems related to manufacturing [27][33]. PSO has ease of implementation, better quality solution and capability to escape from local optimal point. Furthermore, it has the characteristics of quick convergence as compared to other nature inspired algorithms.

It is seen that GA gets stuck into its local optima, therefore solely dependence on GA techniques for optimizing the clustering problem may bring delay in acquiring the solution. Therefore, to avoid the delay in the data reception at the user end, the early convergence is required. Furthermore, if the PSO based technique is dealt solely, the best solution is not saved in a way which is facilitated in GA based technique where elite solutions are saved. However, if the hybrid approach is used, it may lead to the early convergence for getting the optimized solution and best solution is also preserved.

4. Proposed System

This section shade light on the proposed GAPSO-H system framework.

4.1 Heterogeneous Model of GAPSO-H

GAPSO-H utilizes the heterogeneous model for its operation. The nodes are energy heterogeneous and three level of energy heterogeneity is employed. The number of normal,

7

advanced and super nodes used in the network are represented as N_{NORM}, N_{ADVN} and N_{SUP} as given eq. (1-9). The quantity of these high-energy nodes i.e., advanced and super nodes are represented by p_o and p, respectively.

$$N_{SUP} = n \times Ʊ \tag{1}$$

$$N_{ADVN} = n \times ɵ \tag{2}$$

$$N_{NORM} = n \times (1 - Ʊ - ɵ) \tag{3}$$

The advanced and super nodes are ϖ and ϕ times more in energy as compared to normal nodes. The symbols ϖ and ϕ represent the energy fractions of advanced and super nodes, respectively. The computation of total energy of the network represented by E_T is done through eq. (4-9). E_{SUP}, E_{ADVN}, and E_{NORM} represent the energy of super, advanced, and normal nodes, respectively.

$$E_{SUP} = E_O \times (1 + \varpi) \times n \times Ʊ \tag{4}$$

$$E_{ADVN} = E_O \times (1 + \phi) \times n \times ɵ \tag{5}$$

$$E_{NORM} = E_O \times (1 - Ʊ - ɵ) \times n \tag{6}$$

$$E_T = E_{ADVN} + E_{SUP} + E_{NORM} \tag{7}$$

$$E_T = E_O \times (1 + \phi) \times n \times Ʊ + E_O \times (1 + \varpi) \times n \times ɵ + E_O \times (1 - Ʊ - ɵ) \times n \tag{8}$$

$$E_T = n \times E_O \times (1 + \varpi \times ɵ + Ʊ \times \phi) \tag{9}$$

The E_T computed above is used further in CH selection while integrating fitness function in the following subsection.

4.2 The overall operational structure of GAPSO-H

The overall operation of the GAPSO-H is discussed in the following steps.

Step 1: The heterogeneous nodes are deployment randomly in the network and the sink is located initially at the one corner of the network.

Step 2: Once the network is established, second step involves the clustering process in which the *CH selection* is accomplished through applying GA. The CH selection is done in set up phase and the cluster members send the selected data to the CH in the steady state phase. This process of CH selection is explained in the proceeding *Subsection 4.3*.

Step 3: The next task is moving the sink by implying the PSO strategy. It will help the sink to decide where it should be moving to collect data. The optimized movement of the sink is explained in detail in *Subsection 4.4*.

Step 4: The nodes consume their energy with respect to the radio energy model as discussed in *Subsection 4.5*.

Step 5: Finally, the algorithm is implemented in the MATLAB and the simulation results acquired are discussed in detail in **Section 5**.

4.3 GAPSO-H for CH selection using GA

The protocol GAPSO-H, we consider the validation process to work with and for this let assume the nodes as a collection of chromosomes, which will be re-calculated the concerning shape of bit frame. For instance, when the bit is '1' or '0', the node is indicated as a CH or as a member node respectively. Validation process helps in the initialization identifying the eligible nodes that are considered further during implementation.

4.3.1   Initialization

The certain chromosomes are initialized based on their desirable characteristics. It is analogous to the network parameters comprising network area, number of nodes, sink position in the network and the value of transmitting and receiving energy encountered while data transmission.

4.3.2   Fitness Function

Fitness function deals with various fitness parameters that decide for the fitness of current individual. Here, the fitness function comprises five (residual energy of a node, average energy of a node, distance between sink and node, number of neighbors surrounded by a node, and energy consumption rate (ECR)) fitness parameters (FPs) that are employed to determine the function value. The FP is computed for its current value depending upon various factors. It is to be noted that more significant the parameter is; more optimized value will be acquired. Here, the fitness parameters aim to reduce the energy consumption and rendering the network longevity to the network.

a)   The Residual Energy of a Node (FP$_{1st}$)

Residual energy of a node is one of the most prominent factors while selecting the CH after each round. The reason behind selecting this factor is the rotation of CH based on the residual energy of the node. Rotation of CH is required to bring energy balancing in the network. As the network considered is heterogeneous in nature, the nodes with maximum energy is favored to be selected as CH.

9

$$FP_{1st} = 1 \Big/ \sum_{i=1}^{N} \left(\frac{E_{R(i)}}{E_T} \right) \qquad (10)$$

Eq. (10) is used to determine the residual energy of a node. It is summation of ratio of residual energy of ith node represented by $E_{R(i)}$ and total energy denoted by E_T . The total number of nodes is denoted by *N*. The following phenomenon "*Lower the value of FP_{1st} of a node, lower will be the chances of selecting it as CH*" is used.

b) Distance between Sink and Node (FP_{2nd})

In this aspects, consumption of energy of a node to decide the distance in compliance with the communication among themselves or with the sink. The distance from sink to node remain directly proportion to amount of this energy utilization along with a node. Consequently, networking strategy towards CH selection looks after of the parameter accordingly in which the median distance among the nodes along with sink may be optimized. FP_{2nd} instead of framing the objective function instead of which is CH selection accords along with distance component is presented in accordance with the equation (11).

$$FP_{2nd} = \sum_{i=1}^{N} \left(\frac{D_{(N(i)-S)}}{D_{AVG(N(i)-S)}} \right) \qquad (11)$$

In above equation (11), Third fitness parameters (FP_{2nd}) evaluates the summation of distance expense acquired for each ith node. where *N* is the total number of nodes of the networks. Whereas the Euclidean distance from sink to i^{th} node is denoted by $D_{N(i)-S}$ however, $D_{AVG(N(i)-S)}$ signifies an avg. distance at the center of ith node and sink.

c) The Average Energy of a Node (FP_{3rd})

Average energy nodes are preferred for the selection of CH. This node is measured since the network is positioned by the energy heterogeneous nodes. It is then super nodes sustain for longer extent as related to the advanced nodes. Likewise, advanced nodes are favored over the normal nodes. Node has an average energy as FP_{3rd} and its standardized to consume value between 0 and 1. $E_{(i)}$ denoted the energy of its node, as well as N denoted the total number of nodes in the network of eq. (12).

$$FP_{3rd} = \frac{1}{N} \sum_{i=1}^{N} E_{(i)} \qquad (12)$$

d) Number of Neighbors Surrounded by a Node (FP_{4th})

Once the network expanded in the larger area, the intra cluster communication enhances a controlling entity. The selection of a node as CH if made independent of number of neighboring nodes to that node, it will result in selecting node as CH which is far located with respect to other nodes. Consequently, the CH node will consume more energy in collecting data form the other nodes in a cluster. Therefore, to avoid such selection, the number of neighboring nodes is taken into consideration. Therefore, FP_{4th} deals with number of neighboring nodes and is defined by the following eq. (13).

$$FP_{4th} = \left(\frac{\sum_{i=1,j=1}^{N_{CL}} D_{(N(i)-N(j))}}{N_{CL}} \right) \tag{13}$$

In above eq. (13), FP_{4th} is evaluate the ratio summation of $D_{(N(i)-N(j))}$ and N_{CL} . Distance among i^{th} node and j^{th} node in a network cluster is denoted by $D_{(N(i)-N(j))}$. Whereas the total number of nodes in the cluster is denoted by N_{CL}. Therefore, FP_{4th} must be reduced to do it an energy effective CH selection.

e) Energy Consumption Rate (ECR) ($FP_{5th}(ECR)$)

The amount of energy consumption by a node is decided by substantial factor and converts the projecting concern used for selection of CH. After first round it's the difference among the initial energy of the node and the remaining energy. The energy of the node in the prior round turn into its initial energy in place of the number of rounds were proceeded. Hence, ECR is figured and linked by the threshold average value of ECR. The node becomes suitable to become CH if the figured value is originated to be lesser than threshold average rate, it will not be eligible for that round to be CH. FP_5th (ECR) is agreed by equation (14).

$$FP_{5th}(ECR) = \sum_{i=1}^{N} \left(\left(E_{p(i)} - E_{RC(i)} \right) \middle/ \left(E_{p(i)} \right) \right) \tag{14}$$

As the above equation (14), $E_{RC(i)}$ represents the energy consumed in the existing round by i^{th} node besides $E_{p(i)}$ denotes the energy value of the ith node in the preceding round.

Each node of the network analyzes the threshold average value of energy consumption rate (ECR), that node is consider for cluster with lowest energy. It is not counted into consideration for the CH selection, if any one node consumes energy with a lot of magnitude and gets lesser in worth as relate to the lowest energy node.

11

f) Fitness Function for the network

The fitness function of the network is the integration of different FPs (Eq. (10-14)) are integrated altogether in a single expression in eq. (15).

$$F = \frac{1}{\varphi \times FP_{1st} + \delta \times FP_{2nd} + \gamma \times FP_{3rd} + \alpha \times FP_{4th} + \sigma \times FP_{5th}} \tag{15}$$

The fitness function characterized by F in equation (15) must be decreased to create the network efficiency at the best possible cost. In eq. (15), consequent fitness parameters are increased with weight coefficients i.e., φ, δ, γ, α, and σ. Such elements are uniformly weighted as given as follows in equation (16).

These factors are evenly weighted such that it follows eq. (16).

$$\varphi + \delta + \gamma + \alpha + \sigma = 1 \tag{16}$$

Therefore, the main objective function defined for the GA is given by eq. (15) and the PSO based operations are applied to minimize this function for network lifetime and stability period enhancement. *It is imperative to mention that though we have proposed multi-objective PSO-based algorithm. However, while formulating a fitness function that considers its different fitness parameters the multi-objective function is transformed into single objective function which is defined by eq. (16).*

4.3.3 Selection

The ranking selection method is used in this work for selecting the fittest chromosome. This method is employed in a way that a rank table is generated for every fittest chromosome. After creating table sorting is done to extract out the best chromosome. Elitism strategy is employed to retain the best fittest chromosome i.e., fittest node is selected as CH and kept is until it remains fitter than the rest chromosomes.

4.3.4 Crossover

Crossover is operated on two parents and two different offspring are generated. Single point crossover is considered in this work in which the bits for a particular chromosome are exchanged with the other. The crossover decides for the CH rotation in the network.

4.3.5 Mutation

A bit flip mutation is adopted where a bit value of bit stream representing the individual chromosome is changed to change the characteristics of offspring generated from crossover operation.

4.3.6 Termination

The aforementioned steps are executed in an iteration till the satisfying conditions are met. The chromosome updating keeps occurring till the fittest value is acquired. After each iteration the fitness value is retained and compared with the previous value and the fittest among is retained. The fittest chromosome corresponds to the node with the maximum fitness value acquired after the iteration. Algorithm 1 presents the overall process of the GA implemented for CH selection.

Algorithm 1. The Proposed Algorithm for CH Selection.

1. Initialize N, Pc, Pm, φ, δ, γ, α, σ, and $gen=0$.

   // N is number of nodes or chromosomes, Pc is crossover rate,

   //Pm is mutation rate, gen is generation number,

   // FP_{1st}, FP_{2nd}, FP_{3rd}, FP_{4th}, FP_{5th} are the fitness

   parameters

2. **for** j=1:*Population_size* **do**

            chromosomes[j]= getChromo{set of N}

3. **end for**

4. **for** j= 1*: length(chromosomes)* **do**

5. Fitness_value:= getFitness (chromosomes)

   Calculate FP_{1st}, FP_{2nd}, FP_{3rd}, FP_{4th}, FP_{5th}, according to eq. (10-13) for each ith chromosome

6. Fitness_Function

$$F = \frac{1}{\varphi \times FP_{1st} + \delta \times FP_{2nd} + \gamma \times FP_{3rd} + \alpha \times FP_{4th} + \sigma \times FP_{5th}} \qquad (15)$$

7. **while** stopping_criteria_reached **do**

      i. gen= gen+1

      ii. fit_chromo [j]=rank_selection (j, F)

      iii. Cross_select_chromo=cross_operation{fit_chromo [j]} with crossover rate Pc

$C(ka, kb/Pc) \Rightarrow k\tilde{}a, k\tilde{}b.$

Perform mutation on *ka* and *kb* with mutation rate*Pm*

$M(ka/Pm) \Rightarrow \tilde{}ka,$

$M(kb/Pm) \Rightarrow \tilde{}kb$

    iv. chromosomes[j]= {*˜ka, ˜kb}*

    v. repeat steps (5-7)

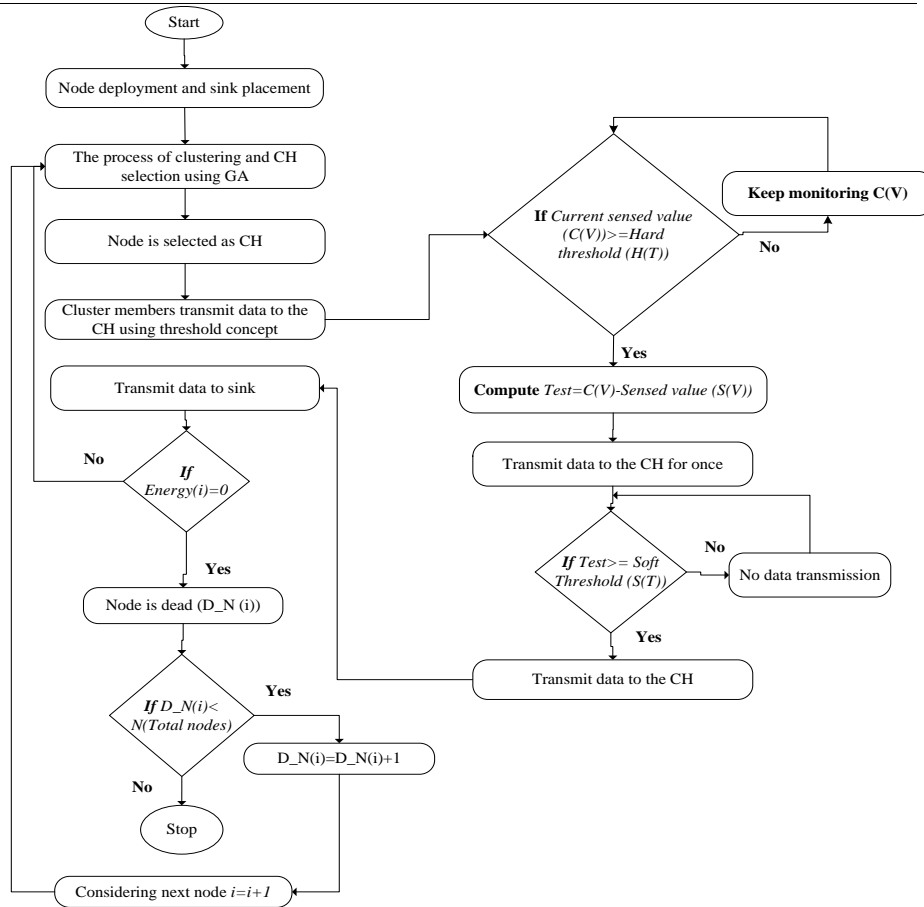**8.**   **end while**

9. **end for**

Fig. 1. Flow chart for working process of GA in GAPSO-H.

14

In 1st step, the parameters are initialized that covers the size of population, crossover rate, generation number, mutation rate, and chromosomes total number. In 2nd step, selection of some chromosomes is done. In 3rd step, the fitness parameters are determined for the chromosomes selected in Step 2. These parameters are then taken into consideration for determining the fitness function value eq. (15). In 4th step, the process keeps running till the stopping criteria is acquired. In 5th step, generation counter is incremented by 1 (gen=gen+1). In 6th step, the fittest chromosome i.e., fit_chromo (*ka, kb)* is selected through the rank selection method. To generate new individual, 1-point cut crossover (C) is applied i.e., *(k˜a, k˜b)* by combining parents with crossover rate (Pc) i.e., *C (ka, kb/Pc) ⇒ k˜a, k˜b.* As a result of crossover, the new chromosomes are produced which are given by *(k˜a, k˜b).* They are further exploited by the mutation operation with mutation rate *(Pm)* to generate (˜ka, ˜kb), i.e., *M(ka/Pm) ⇒˜ka, M(kb/Pm) ⇒˜kb.* Initial population is updated by new generation and among them fittest individual is retained through elitism process. Repeat steps 4-8.

Figure 1 demonstrate the working of GAPSO-H for CH selection in heterogenous network. Overall working is divided into 2-phases: (a) setup phase and (b) steady state phase. The network that has been designed according to the following phases. First of all, the node with energy heterogeneity like advance, intermediate and normal level of node are deployed randomly in the network area which is targeted and then for data transmission from cluster to the user through internet the sink will be put in anywhere in the midway of communications network. After node deployment, the clustering of nodes is performed and selection of CH for each cluster is done by implementing Algorithm 1.

The inter and intra cluster and among the CH and sink communication is performed in steady state phase. As the GAPSO-H is a reactive protocol so it works on the concept of hard and soft threshold as adopted in TSEP protocol [2]. Transfer data depends on this hard and soft threshold. Suppose if the current sensed value represented by (C(V)) is more than hard threshold (H(T)) then data transmission will be done from the node to the CH i.e. intra-cluster data transmission will be happened. Additionally, if the change among the present and formerly sensed value is greater than the preset soft threshold then just the next round communication will be performed. Otherwise, the next round data transmission will be placed on hold until the necessary conditions have been met. Furthermore, at the time of data transmission if the node is dead entailing the totally energy of node exhausted, additional stand towards the dead nodes that are through incrementing 1.

15

However, data aggregation remains carried out on the CH when the data transfer has been completed in accordance with the CH and subsequently an aggregated useful data will be forwarded to the sink from CH. This process remains identically repeatedly until all nodes remain became exhausted and the instant finally nodes remain dead, a single network which is dead.

4.4 GAPSO-H for PSO-based sink mobility

The pseudocode of the PSO based algorithm for sink mobility is given in Algorithm 2.

Algorithm 2. PSO based routing proposed work

---

1. **Input:** Initialize gateways $\xi = \{g1, g2, \ldots, gm\}$,

2. *Next_Hop_G(gi),* $\forall i,\ 1 \leq i \leq m$ and *NP* size.

3. **Output:** Route *R:* $\xi \rightarrow \{\xi + gm+1\}$

4. **Step one:**

5. Set number of particles *Pi,* $\forall i,\ 1 \leq i \leq NP.$

6. **Step two:**

7. **for** $i = 1$ to *NP* **do**

8. Determine *Fitness_f (Pi)* /*Using Eq. (15)*/

9. *Pbesti = Pi*

10. **end for**

11. **Step three:**

12. *Gbest = {Pbestk |Fitness_f (Pbestk)*

13. *min (Fitness_f(Pbesti),* $\forall i,\ 1 \leq i \leq NP)$*}*

14. **Step four:**

15. **while** (stopping criteria meet) **do**

16. **for** $i = 1$ to *NP* **do**

17. Velocity and position of *Pi* are updated.

18. Determination of *Fitness_f (Pi)*

19. **if** *Fitness_f(Pi) < Fitness_f (Pbesti)* **then**

20. *Pbesti = Pi*

21. **end if**

22. **if** *Fitness_f(Pi) < Fitness_f(Gbest)* **then**

23. *Gbest = Pbesti*

24. end if

25. end for

26. end while

27. Step five:

28. Determination *Next_Hop(gi),* $\forall i,\ 1 \le i \le NP,$ (i.e., route *R*) using

29. *Gbest*.

30. **Step six: Terminate**

In 1st step, the number of particles are initialized. The other required network parameters are also initialized. In 2nd step, selection of some particles is done based on fitness function defined in eq. (15). In 3rd step, the fitness parameters are determined to select the CH and thereby the optimized sink mobility can be taken place. The fitness function is computed using eq. (15). In 4th step, the process keeps running till the stopping criteria is acquired. In 5th step, the positions and velocity of the particles are updated based on their fitness values. In 6th step, the optimized position helps in obtaining the next move of the sink. In 7th step, global best and local best are computed using PSO computations. In 8th step, repeat steps 3-8.

Lemma 1: *GAPSO-H terminates in fixed iterations* $It_r = O(1)$ *and renders the optimized CH selection and covers every significant aspect*

Proof: *GAPSO-H* executes it the steady-state phase after completing set-up phase where the data is collected and aggregated to the CH. Additionally, the inevitable data has been transmitted to the sink and this process remain goes on till the network became dead i.e. all of the nodes of that network became exhausted. When nodes start to transmit data to CH, they stat deplete their energies simultaneously and it depends upon the hard and soft threshold. The next round of data transmission cannot be executed until the node completely drains its energy or became a dead node. This process remains identically processed until all nodes are exhausted. Therefore, if the number of iterations increments then technically the number of nodes become lesser. That implies, the iterations and the number of dead nodes is inversely proportional to each other. Additionally, so the quantity of dead nodes remains identical and endless procedure of *GAPSO-H* ceases all through predetermined iteration.

For CH selection residual energy and initial energy is more favorable because they are implanted with more energy. Therefore, the main objective of *GAPSO-H* integrates several key

factors which consider both favorable energy which considers as the available stock of energy for this process as well. Additionally, distance between the nodes and the sink is another major factor for energy consumption. This distance factor and energy consumption is inversely proportional to each other. Therefore, to alleviate the energy consumption due to this distance function between the nodes and sink node the same is also conducted for the consideration.
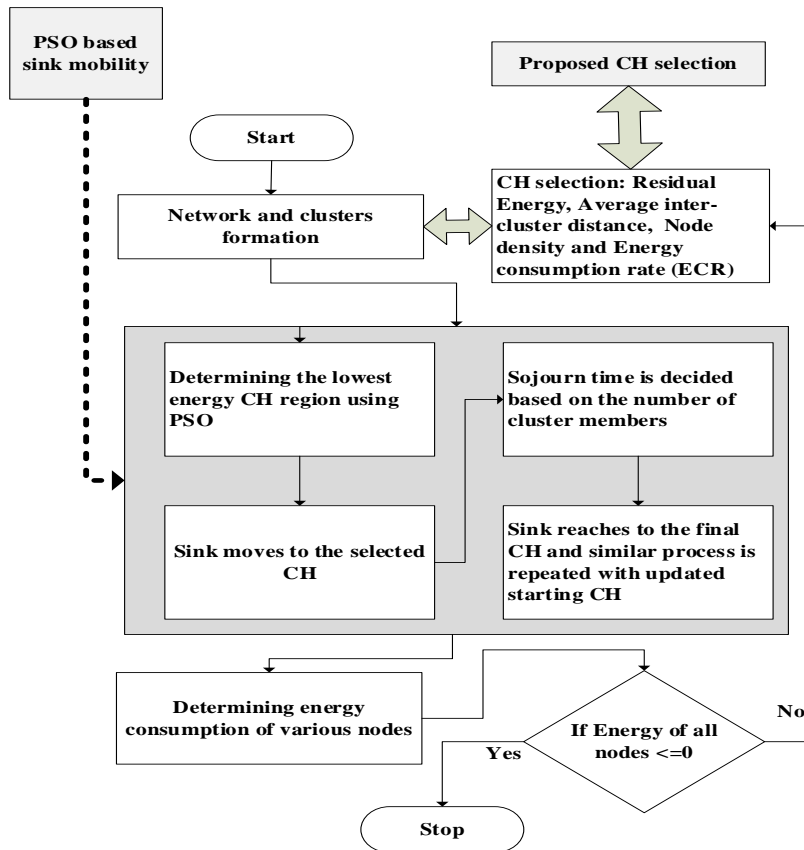

Fig. 2. The design methodology for the proposed work.

The methodology for the sink mobility is illustrated in Fig. 2. With the PSO optimization technique, sink mobility is implemented as the lowest energy CH regions are primarily targeted at saving the energy of those CHs that are about to die. This can be determined through the fitness function.

4.5 Proposed System and Network framework of GAPSO-H

A heterogeneous network with three level energy of heterogeneity model is already discussed in Section 4. Nodes suffer from energy consumption when they are involved in data transmission. The radio energy model that decides the amount of energy consumption is discussed here.

4.5.1 Radio Energy Model for GAPSO-H

Standard for the sensor radio energy paradigm which is responsible for energy consumption of the nodes is proposed in the system structure of GAPSO-H is shown in Fig. 3. The data transceiver for energy consumption are elaborates in equations (16-20). The amount of energy consumption is dependent on the distance between nodes. The energy consumption for transferring the z bit data at the distance which is represented by 'd', is given by E_{tx} (z, d) as shown in eq. (16 – 17).
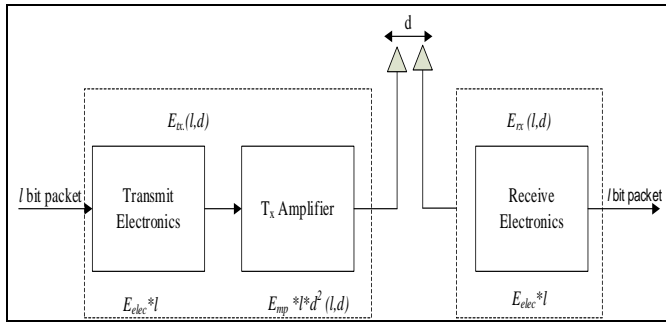

Fig. 3. Radio Energy Dissipation Model.

$$E_{tx}(z, d) = z \times E_{elc} + z \times E_{efs} \times d^2 \text{ for d} \leq d_o \quad (16)$$

$$E_{tx}(z, d) = z \times E_{elc} + z \times E_{amp} \times d^4 \text{ for d} > d_o \quad (17)$$

In eq. (16-17) The symbol 'd' accounts for the distance between the different entities which may be source node or destination node or sink too. E_{elc} denotes the energy consumed for activating the circuits of transceiver. The threshold distance is given by 'd_o' and is expressed by eq. (18).

$$d_o = \sqrt{\frac{E_{efs}}{E_{amp}}} \quad (18)$$

E_{efs} is for free space energy model given in power loss i.e., d^2 and E_{amp} denotes the energy consumption for multi path energy model given in power loss d^4. In reception of single bit, the energy consumption is given by eq. (19).

$$E_{rx}(z) = z \times E_{elec} \tag{19}$$

The energy consumption by the CH for performing data aggregation is given by the eq. (20).

$$E_{dx}(l) = x \times z \times E_{da} \tag{20}$$

The energy exhausted in receiving the z bit data, is given by E_{rx}. E_{da} is the energy exhausted in the data aggregation of 1-bit data. Moreover, $E_{dx}(z)$ is the energy spent during process of data aggregation of receiving z-bit data of *finite* number of packets of data.

4.5.2 Network model assumptions for GAPSO-H

The characteristics of sensor node have left some effects on the network framing. Therefore, to structure the framework for GAPSO-H with sensor nodes we need to look after to their characteristics as follows.

a) The network is immobile that indicates the nodes are unmoving by nature and sink is also kept stationary in the whole network run.

b) There are 3-level of energy heterogeneity nodes within the heterogeneous system specifically, super node, advanced node, as well as normal nodes. Between these nodes super nodes come about having maximum energy and even the smallest is initiated with normal nodes.

c) The energy consumption is constrained for the nodes whereas for the sink there has no such limitation. Consequently, after completing a certain round of data transmission to the CH the nodes became exhausted. Therefore, having no such restriction on energy consumption for sink become unlike for nodes.

d) Nodes used are location oblivious i.e., no GPS is connected on the node's circuitry.

e) Additional aspects causing signal attenuation are not taken into consideration.

f) The network area is assumed to be square shaped.

g) The signal intensity calculated from Received Signal Strength Indicator (RSSI) is considering as per the Euclidean distance between nodes.

5. Numerical Experiments

Extensive computer simulations have been conducted to analyze the performance of the proposed GAPSO-H. All simulations have been performed on the MATLAB 2016, with system configuration of 4GB RAM, 1 TB HD, Intel i5 processor with working with at 2.60 GHz and Window 10.

Table 1. Simulation parameters.

Model of Networks and GAPSO-H Parameters	Values
Size of Network Area	$100 \times 100 \text{ m}^2$, $500 \times 500 \text{ m}^2$
Number of Nodes (N)	100, 200
GAPSO-H required data sink	1
Initial energy (E_o)	0.5 (in Joules)
Heterogeneous levels of nodes	3-level, normal, intermediate and advanced nodes
Intermediate and advanced nodes in Energy fractions	$\beta=1, \alpha=2$
Required intermediate nodes (m) and advanced nodes fraction (m_o)	$m=0.1$, $m_o=0.2$
Energy consumed by transmitter and also by receiver E_{elc}.	50nJ/bit
Threshold distance (d_o)	87m
Amount of Amplification energy for smaller and larger distance $d \leq d_o$ (E_{efs}) And $d > d_o$ (E_{mp})	10pJ/bit/m^2 and 0.0013pJ/bit/m^4
Data aggregation (E_{da})	5nJ/bit/signal
Size of data packet	2000bits
Size of population (P)	30
Crossover rate (P_c)	0.6
Crossover types	Single Point
Mutation rate (P_m)	0.006
Method of selection	Rank selection method
Total chromosomes	30
Generations number	30
Simulation runs	20
Total Particles	30
Acceleration factors (c_1, c_2)	2.0
Initial velocity	0.0
Initial position	9.6, 13.6

5.1 Control Parameters Tuning

The network, radio energy model and simulation parameters are given in Table 1. Here, we have considered two cases, in first case, 100 nodes are deployed in 100 X 100 square meter area, and in the second case, the 200 nodes are deployed in 500 X 500 square meter area. The nodes are of three level of energy heterogeneity viz. normal, advanced and super nodes. An extensive control parameter tuning is done by Taguchi signal to noise ratio (SNR) method along with orthogonal

matrix as done in [28]. Taguchi SNR is a log function of the desired output that serves as an objective function as shown in eq. (21).

$$SNR_i = -10 \log \left(\sum_{s=1}^{T_s} \frac{y_s^2}{T_i} \right)$$ (21)

Where i, s, T_i and y_s respectively represent experiment number, trial number, total number of trial for the experiment and number of iterations performed in each trial to get a solution. For control parameter tuning 3-levels were identified for size of population (P) = [20, 30, 60], crossover rate (P_c) = [0.2, 0.4, 0.6], mutation rate (P_m) = [0.002, 0.004, 0.006] and total chromosomes = [10, 30, 50]. The following setting produced the best results [P, P_c, P_m, total chromosomes] = [30, 0.6, 0.006, 30] for the proposed GAPSO-H.

The performance of the proposed GAPSO-H is compared against the state-of-the-art algorithm such as PSOECSM [25], PSO-UFC [26], GADA-LEACH [10], PSOBS [24], GABEEC [30] and DCH-GA [11] (See Section 5.3). For a fair comparison the same control parameter values (as used for the GAPSO-H) have been used for other GA based algorithms implemented in this paper. The control parameter setting has been done in the following manner: "*population size (P) lies between 20 to 40 with step size 10, crossover rate (Pc) lies between 0.5 to 0.7 with step size 0.1, mutation rate (Pm) lies between 0.005 to 0.007 with step size 0.001 and total chromosomes lies between 20 to 40 with step size 10*". On the other hand, parameter tuning for the PSO based algorithm has also been performed where total particles = 30, acceleration factors (c1, c2) = 2.0, initial velocity = 0.0, and initial position = [9.6, 13.6] produced the best results. The control parameter values after tuning process is presented in Table 1.

It is noted that the nodes are classified into different genres and defined unique id's based on their initial energy stock. However, with the gradual progression of data transmission the energy of the nodes decreases. A moment will come when the nodes from one category might switch from one genre to other. However, the fact is, the rules designated for any genres of nodes stays same throughout the functioning of network. For an example, if a node a designated as super node gets to the energy to that of normal nodes, it will still have the same CH selection criteria as that of super nodes.

Commented [HP1]: Comment-1: Have authors explained how competing methods were tuned on the selected problems.

Response: Many thanks for pointing this to us. We performed extensive parameters tuning for each algorithm. Design of experiment (DOE) method is used to develop a robust experimental environment. Taguchi signal to noise ratio (SNR) method along with orthogonal matrix is utilized for tuning the control parameters for each algorithm.

22

5.2 Performance Metrics

Five performance metrics are discussed below have been considered to observe the performance of the proposed GAPSO-H.

a) **Stability Period:** It can be defined as the number of rounds covered until any node is completely exhausted in its energy stock. Hence that node is said to be dead node. This parameter can play a significant role in those applications where the leakage of information cannot be accepted. The applications with such narrow acceptance for loss of information values stability period to the maximum level. With higher value of stability period, the higher reliability is ensured.

b) **Network Longevity**: It can be defined as the rounds covered when there is no node left alive to communicate in the network. It is crucial for those applications where the monitoring of the network is a continuous process. Applications like agricultural, flood detection, etc. are some of examples where the network longevity is highly significant.

c) **Number of dead nodes**: To observe the behavior of protocol when it undergoes the energy consumption while data transmission, the pace is observed at which the nodes keep depleting their energies until they are fully dead.

d) **Throughput:** It can be defined as the number of data packets that follow successful transmission is termed as throughput. It helps in explaining the QoS status for each node. The stability period and network longevity have no value if the QoS is poor for any routing protocol.

e) **Network's remaining energy**: The remaining energy of the network is observed at various number of rounds. It shows the load balancing of the network. Network energy is basically the sum of energy of all nodes deployed in the network.

5.3 Comparison with the state-of-the-art methods

The performance of the proposed GAPSO-H is tested against the state-of-the-art methods such as PSOECSM [25], PSO-UFC [26], GADA-LEACH [10], PSOBS [24], GABEEC [30] and DCH-GA [11] protocols to validate its promising performance. The aforesaid protocols are selected for comparison with GAPSO-H due to the hybrid nature of the proposed protocol. While comparing with GA and PSO based algorithms, it is ensured that the fair comparison is performed with the proposed protocol.

5.4 Results and Analysis

Results and analysis are presented by considering all 5-performance metrics.

5.4.1 Stability Period

Fig. 4 reveals that the GAPSO-H's stability period is better than other algorithms. It is seen that in GAPSO-H, the first node was dead after 6387 rounds whereas in the case of DCH-GA, GABEEC, PSOBS, GADA-LEACH, PSO-UFC and PSOECSM it is just 3700, 4404, 4578, 4424, 5556, and 6052 respectively.

It is comprehended that GAPSO-H ameliorates stability period by 72.6%, 44.8%, 39.5%, 44.3%, 14.9% and 5.5% as compared to DCH-GA, GABEEC, PSOBS, GADA-LEACH, PSO-UFC and PSOECSM respectively. Moreover, as shown in Fig. 5, which considers the larger network area, the network lifetime of GAPSO-H is enhanced as compared to the other protocols. The network lifetime of GAPSO-H is 3640 rounds whereas the PSOECSM, PSO-UFC, GADA-LEACH, PSOBS, GABEEC and DCH-GA covers 3188, 2805, 2615, 2276, 2002, and 1894 rounds, respectively until all nodes are operational.
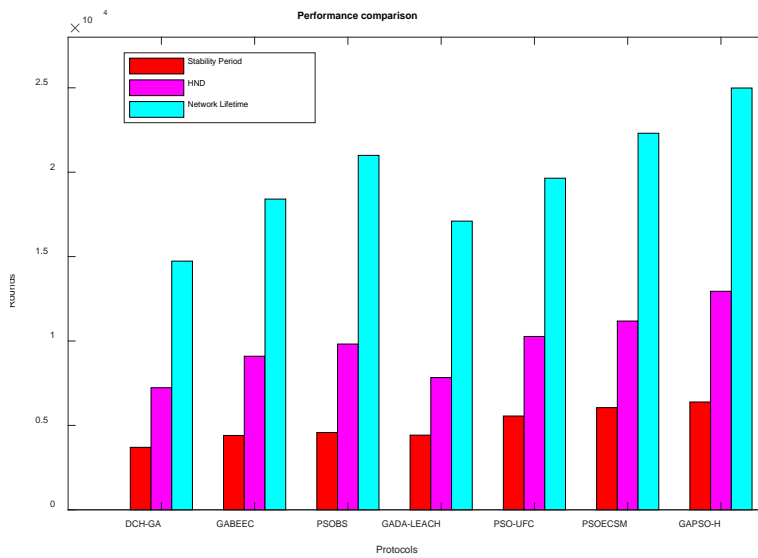

Fig. 4. Overall performance comparison of GAPSO-H with other protocols.
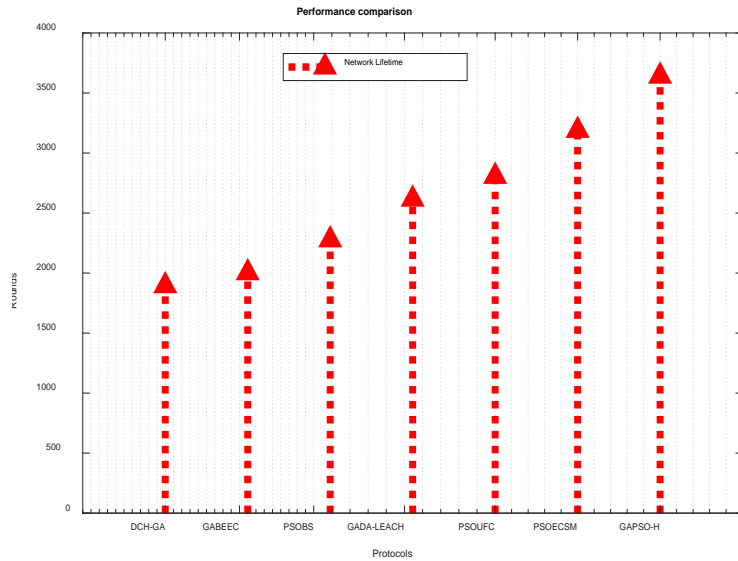
Fig. 5. Performance comparison of GAPSO-H with other protocols for enhanced network area.

The dominant cause behind such improvement is the incorporation of FPs (Eq. (10-14) that ensure the energy preservation while the data transmission is in process. The distance among the nodes and sink and within the nodes is effectively decreased. It can be comprehended that GAPSO-H outperform the state-of-art protocols in the context of stability period.

5.4.2 Network Lifetime

Two-fold (alive nodes vs rounds and dead nodes vs rounds) analysis have been performed to analyze the network lifetime. Fig. 6-7 and 8-9 respectively present network lifetime analysis for alive nodes vs rounds and dead nodes vs round. We noticed that the network lifetime for GAPSO-H is achieved at 24988 rounds whereas DCH-GA, GABEEC, PSOBS and GADA-LEACH, PSO-UFC and PSOECSM it is observed at 14732, 18408, 20998, 17102, 19645 and 22308 rounds respectively. This analysis reveals that 10256, 6580, 3990, 7886, 5343 and 2680 more rounds are covered by GAPSO-H as compared to state-of-the-art methods. It is observed from the Fig. 7, there is slight increase in the performance when the network area is enhanced.

The improvement in network lifetime is observed due to the ECR factors are integrated in fitness function oppressed for GAPSO-H. Moreover, through high number of neighboring nodes, the

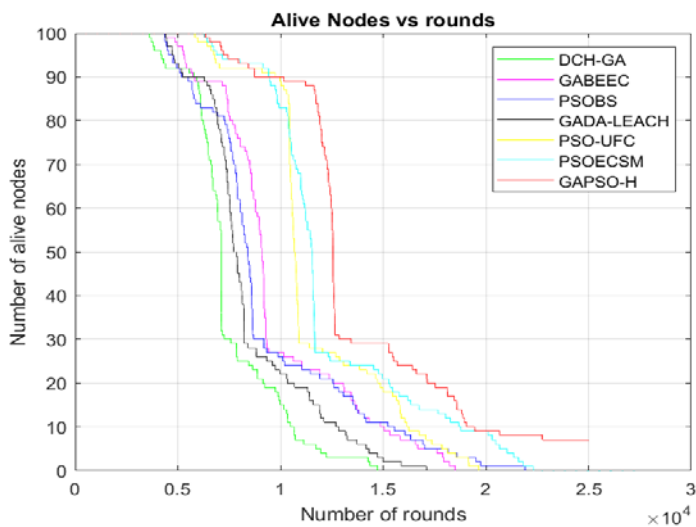average distance between CH and node is decreased comprehensively.

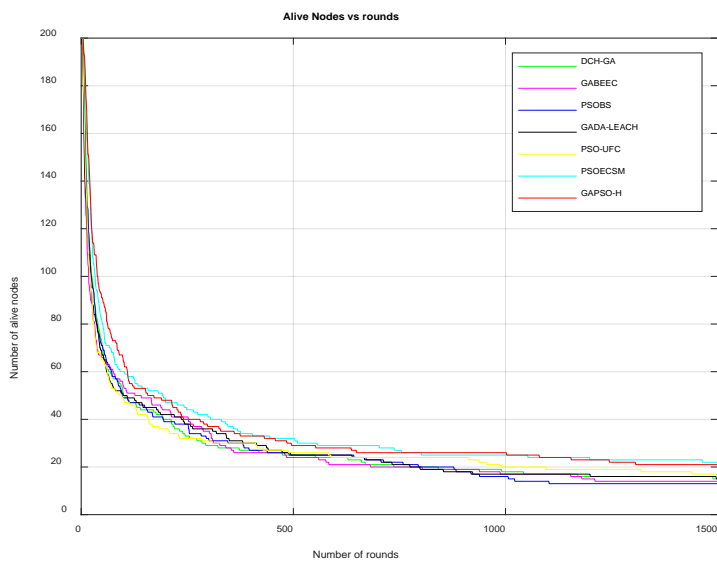**Fig. 6**. Comparison of alive nodes vs rounds of GAPSO-H with other protocols.

Fig. 7. Comparison of alive nodes vs rounds of GAPSO-H with other protocols for enhanced network area.
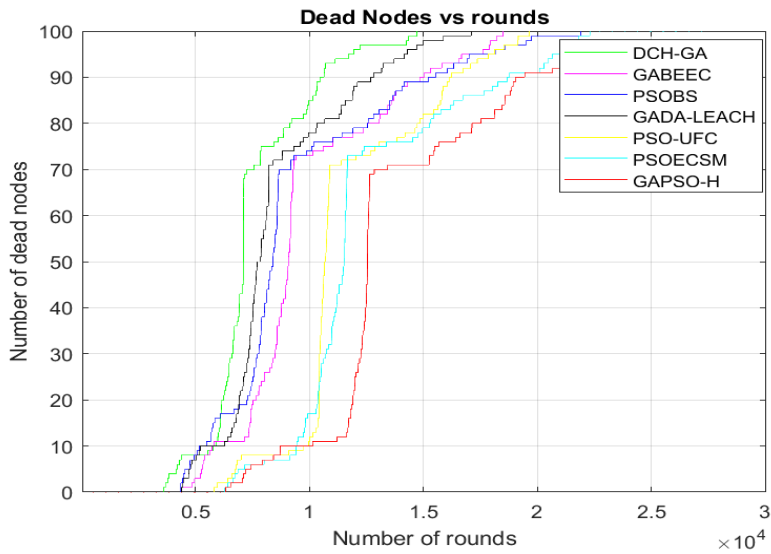
Fig. 8. Comparison of dead nodes vs rounds of GAPSO-H with other protocols.

5.4.3 Number of dead nodes against rounds

It is seen that GAPSO-H as compared with other protocols, the smaller number of rounds with respect to number of dead nodes. In Fig. 8, shows that, the First Node Dead (FND) for GAPSO-H is 6387 rounds which is 6052, 5556, 4424, 4578, 4404 and 3700 rounds in protocols PSOECSM, PSO-UFC, GADA-LEACH, PSOBS, GABEEC and DCH-GA respectively.
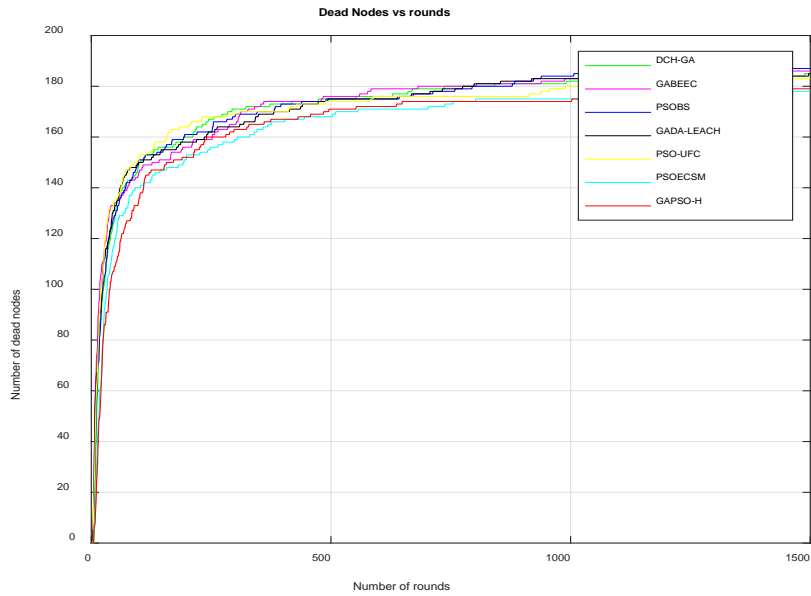
Fig. 9. Comparison of dead nodes vs rounds of GAPSO-H with other protocols for enhanced network area.

As far as Half Nodes Dead (HND) for GAPSO-H is concerned, it is 12946 whereas it is just 11182, 10268, 7834, 9820, 9098 and 7232 rounds in case of PSOECSM, PSO-UFC, GADA-LEACH, PSOBS, GABEEC and DCH-GA protocols, respectively. Furthermore, the Last node dead (LND) i.e., also termed as network lifetime is also found to be improved, is also observed in GAPSO-H as it covers 24988 rounds whereas PSOECSM, PSO-UFC, GADA-LEACH, PSOBS, GABEEC and DCH-GA covers 22308, 19645, 17102, 20988, 18408 and 14732 rounds, respectively. Further, as shown in Fig. 9, the protocol GAPSO-H covers more rounds as stated above, when the network is expanded.

Such improvement is reported due to the fact that when the CH selection is optimized under different attributes, more energy preservation is acquired in GAPSO-H as compared to other protocols, respectively.
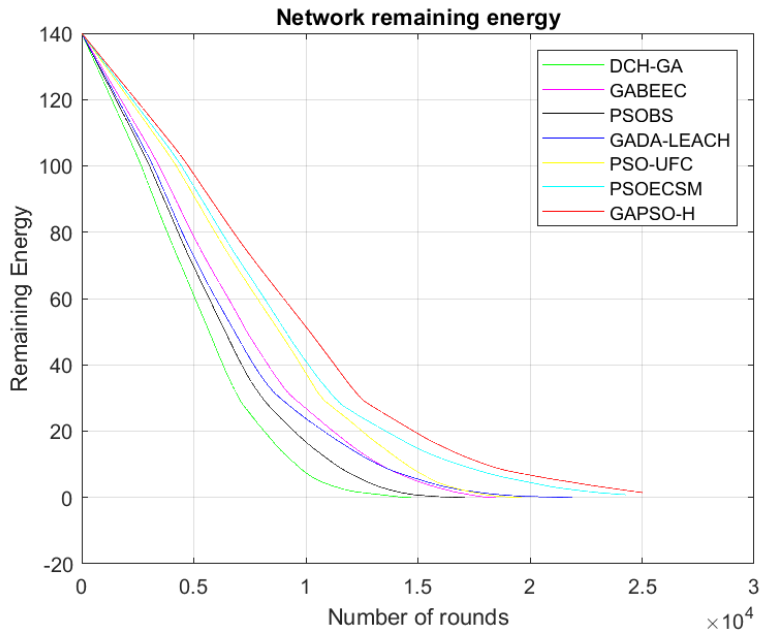
Fig. 10. Comparative analysis of Network's remaining energy of GAPSO-H with other protocols.

5.4.4 Network's remaining energy

It is seen that in GAPSO-H protocol, reducing of networks energy as per the data transmission is assumed. It is quite essential to observe the behaviour of network's remaining energy with increase in number of rounds. GAPSO-H performs better as compared to DCH-GA, GABEEC, PSOBS, GADA-LEACH, PSO-UFC and PSOECSM protocols, respectively such that the number of rounds covered are more in the process of data transmission as given in Fig. 10.
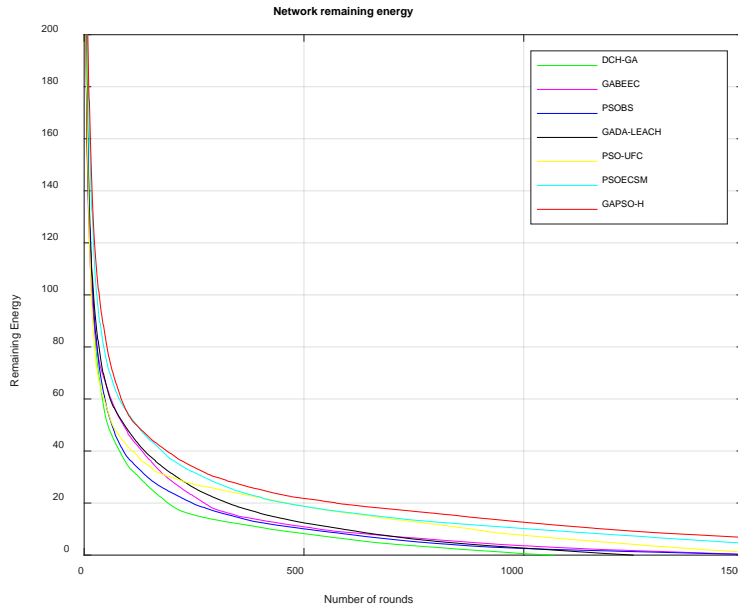
Fig. 11. Comparative analysis of Network's remaining energy of GAPSO-H with other protocols for enhanced network area

As shown in Fig. 11, the network's remaining energy of the proposed protocol is still better than the other protocols even when the network area is expanded. The energy of a node is preserved in each round due to the minimum energy consumption resulted due to the energy efficient dual hop communication.

5.4.5 Throughput

It is seen that GAPSO-H, as given in Fig. 12, the throughput is observed to be incremented as the successful transmission of 792952 data packets were done. However, PSOECSM, PSO-UFC, GADA-LEACH, GABEEC and DCH-GA sent 637784, 698424, 506248, 486731, 567692 and 443858 packets of data, respectively. When the comparative analysis is performed, it is found that GAPSO-H improves throughput by 24.3%, 13.5%, 56.6%, 63%, 39.6% and 78.6% percent in comparison to PSOECSM, PSO-UFC, GADA-LEACH, PSOBS GABEEC and DCH-GA protocols, respectively. Further, it is seen from the Fig. 13, the protocol sends more data packets to the sink even when the network area is enhanced.
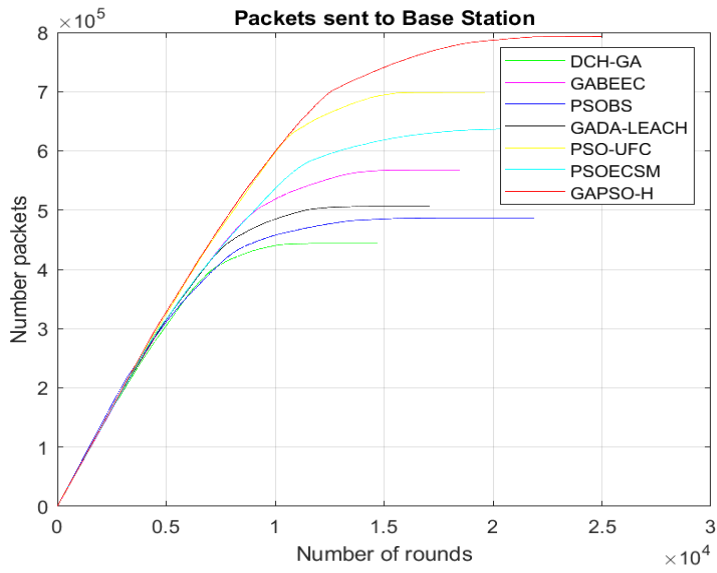
Fig. 12. Throughput comparison of GAPSO-H with other protocols.
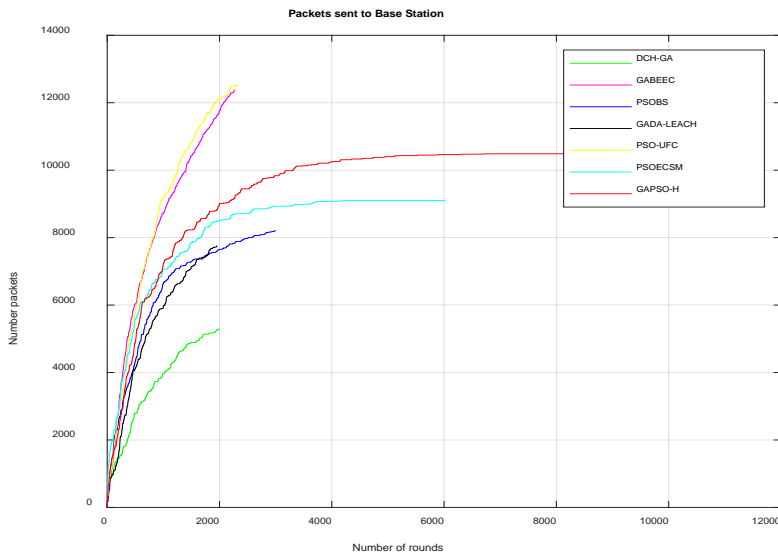

Fig. 13. Throughput comparison of GAPSO-H with other protocols for enhanced network area.

Throughput is enhanced gigantically due to the fact that with the selection of optimized CH, the data packets are forwarded in the most optimal way leading to the reduction in packet loss during transmission.

5.5 Summary

In nutshell, the summary of computer simulation is reported in Table 2. The comparative analysis indicates that GAPSO-H has outperformed other state-of-the-art algorithms with respect to different performance metrics.

Table 2. Comparative analysis of GAPSO-H with other protocols for different metrics.

Value of advanced fractions and quantity fractions of node $m=0.1, m_o=0.2, \beta=1, \alpha=2$					
References	Protocols	Total Energy of Network (Joules)	Stability Period (rounds)	Half Node Dead (rounds)	Network Lifetime (rounds)	Throughput (packets)
Elhoseny et al. [11]	DCH-GA	140	3700	7232	14732	443858
Bayrakh et al. [30]	GABEEC	140	4404	9098	18408	567692
Tabibi et al. [24]	PSOBS	140	4578	9820	20998	486731
Bhatia et al. [10]	GADA-LEACH	140	4424	7834	17102	506248
Kaur et al. [26]	PSO-UFC	140	5556	10268	19645	698424
Sahoo et al. [25]	PSOECSM	140	6052	11182	22308	637784
Proposed	GAPSO-H	140	6387	12946	24988	792952

The percentage improvement by the GAPSO-H in the context of stability period, HND, Last Node dead which is 'Network Lifetime' and Throughput, are given in Table 3.

Table 3. Improvement analysis by GAPSO-H compared to other protocols.

Improvement in Percentage (%) by GAPSO-H protocol					
References	Protocols	Stability Period	HND	LND or Network Lifetime	Throughput
Elhoseny et al. [11]	DCH-GA	72.6	79	69.6	78.6
Bayrakh et al. [30]	GABEEC	44.8	42.2	35.7	39.6
Tabibi et al. [24]	PSOBS	39.5	31.8	19	63
Bhatia et al. [10]	GADA- LEACH	44.3	65.2	46.1	56.6
Kaur et al. [26]	PSO-UFC	14.9	26.1	27.1	13.5
Sahoo et al. [25]	PSOECSM	5.5	15.7	12	24.3

5.6 Statistical analysis

Rigorous statistical tests have been conducted to determine the performance significance (95% level of confidence interval). Samples for remaining energy has been selected for each algorithm.

Table 4 presents the descriptive analysis with respect to reaming energy for each algorithm. It can be seen that the proposed GAPSO-H has achieved the highest mean (= 1.753).

Table 4. Descriptive analysis with respect to remaining energy for each algorithm.

References	Algorithms	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Elhoseny et al. [11]	DCH-GA	30	-.013	.005	.001	-.016	-.011	-.0204	-.006
Sahoo et al. [25]	PSOECSM	30	1.480	.025	.004	1.470	1.489	1.400	1.501
Kaur et al. [26]	PSO-UFC	30	-.008	.002	.000	-.008	-.007	-.011	-.004
Bhatia et al. [10]	GADA-LEACH	30	-.009	.002	.000	-.0105	-.008	-.013	-.005
Proposed	**GAPSO-H**	**30**	**1.753**	**.415**	**.075**	**1.598**	**1.908**	**1.127**	**2.033**
Tabibi et al. [24]	PSOBS	30	.071	.057	.010	.0497	.092	-.011	.140
Bayrakh et al. [30]	GABEEC	30	-.012	.001	.000	-.0127	-.011	-.015	-.009
	Total	210	.465	.750	.051	.3638	.567	-.020	2.033

Table 5. Result of ANOVA test with respect to remaining energy for each algorithm.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	**112.555**	**6**	**18.759**	**742.729**	**0.000**
Within Groups	5.127	203	.025		
Total	117.682	209			

Table 5 presents ANOVA test results where p-value (Sig.) is less than 0.05 (0.000 < 0.05). Therefore, null hypothesis (mean value of all algorithms are equal) can be rejected. However, this result does not reveal that which algorithm is showing the best, average or worst performance. Hence, multiple comparison (paired t-test/Posthoc test) test has been conducted. Table 6 shows the results of multiple comparison test. It can be seen that the performance of the GAPSO-H is significantly (Sig. value in 5th Column is 0.000 < 0.05 for each pair of algorithms of 6th row) better than the other six algorithms.

Table 6. Result of multiple comparison t-test (Posthoc test).

References	(I) Algorithms	(J) Algorithms	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Elhoseny et al. [11]	DCH-GA	PSOECSM	-1.4939299*	.0410342	.000	-1.616136	-1.371724
		PSO-UFC	-.0056643	.0410342	1.000	-.127870	.116542
		GADA-LEACH	-.0042423	.0410342	1.000	-.126448	.117964
		GAPSO-H	-1.7674309*	.0410342	.000	-1.889637	-1.645225
		PSOBS	-.0851974	.0410342	.371	-.207403	.037008
		GABEEC	-.0017827	.0410342	1.000	-.123989	.120423
Sahoo et al. [25]	PSOECSM	DCH-GA	1.4939299*	.0410342	.000	1.371724	1.616136
		PSO-UFC	1.4882656*	.0410342	.000	1.366060	1.610472
		GADA-LEACH	1.4896876*	.0410342	.000	1.367482	1.611893
		GAPSO-H	-.2735010*	.0410342	.000	-.395707	-.151295
		PSOBS	1.4087325*	.0410342	.000	1.286527	1.530938
		GABEEC	1.4921472*	.0410342	.000	1.369941	1.614353
	PSO-UFC	DCH-GA	.0056643	.0410342	1.000	-.116542	.127870

Kaur et al. [26]		PSOECSM	-1.4882656*	.0410342	.000	-1.610472	-1.366060
		GADA-LEACH	.0014219	.0410342	1.000	-.120784	.123628
		GAPSO-H	-1.7617667*	.0410342	.000	-1.883973	-1.639561
		PSOBS	-.0795331	.0410342	.458	-.201739	.042673
		GABEEC	.0038816	.0410342	1.000	-.118324	.126087
Bhatia et al. [10]	GADA-LEACH	DCH-GA	.0042423	.0410342	1.000	-.117964	.126448
		PSOECSM	-1.4896876*	.0410342	.000	-1.611893	-1.367482
		PSO-UFC	-.0014219	.0410342	1.000	-.123628	.120784
		GAPSO-H	-1.7631886*	.0410342	.000	-1.885394	-1.640983
		PSOBS	-.0809551	.0410342	.435	-.203161	.041251
		GABEEC	.0024596	.0410342	1.000	-.119746	.124666
Proposed	**GAPSO-H**	**DCH-GA**	**1.7674309***	**.0410342**	**.000**	**1.645225**	**1.889637**
		PSOECSM	**.2735010***	**.0410342**	**.000**	**.151295**	**.395707**
		PSO-UFC	**1.7617667***	**.0410342**	**.000**	**1.639561**	**1.883973**
		GADA-LEACH	**1.7631886***	**.0410342**	**.000**	**1.640983**	**1.885394**
		PSOBS	**1.6822335***	**.0410342**	**.000**	**1.560028**	**1.804439**
		GABEEC	**1.7656482***	**.0410342**	**.000**	**1.643442**	**1.887854**
Tabibi et al. [24]	PSOBS	DCH-GA	.0851974	.0410342	.371	-.037008	.207403
		PSOECSM	-1.4087325*	.0410342	.000	-1.530938	-1.286527
		PSO-UFC	.0795331	.0410342	.458	-.042673	.201739
		GADA-LEACH	.0809551	.0410342	.435	-.041251	.203161
		GAPSO-H	-1.6822335*	.0410342	.000	-1.804439	-1.560028
		GABEEC	.0834147	.0410342	.397	-.038791	.205621
Bayrakh et al. [30]	GABEEC	DCH-GA	.0017827	.0410342	1.000	-.120423	.123989
		PSOECSM	-1.4921472*	.0410342	.000	-1.614353	-1.369941
		PSO-UFC	-.0038816	.0410342	1.000	-.126087	.118324
		GADA-LEACH	-.0024596	.0410342	1.000	-.124666	.119746
		GAPSO-H	-1.7656482*	.0410342	.000	-1.887854	-1.643442
		PSOBS	-.0834147	.0410342	.397	-.205621	.038791

*. The mean difference is significant at the 0.05 level.

Table 7. Homogeneous subset test.

References	Algorithms	N	Subset for alpha = 0.05		
			1	2	3
Elhoseny et al. [11]	DCH-GA	30	-.013		
Bayrakh et al. [30]	GABEEC	30	-.012		
Bhatia et al. [10]	GADA-LEACH	30	-.009		
Kaur et al. [26]	PSO-UFC	30	-.008		
Tabibi et al. [24]	PSOBS	30	.071		
Sahoo et al. [25]	PSOECSM	30		1.480	
Proposed	**GAPSO-H**	**30**			**1.753**
	Sig.		.371	1.000	1.000

Means for groups in homogeneous subsets are displayed.
a. Uses Harmonic Mean Sample Size = 30.000.

Results of homogeneous subset test is depicted in Table 7. This result reveals that the proposed GAPSO-H has outperformed other algorithms, showed mean value = 10.753 at 1.000 level of significance and fall in group-3. It can be noted that PSOECSM showed average performance and fall in group-2 at 1.000 level of significance. Remaining algorithms showed similar performance; hence, they fall in the same group (i.e. group 1). This result also indicates that DCH-GA and GABEEC showed the worst overall performance.
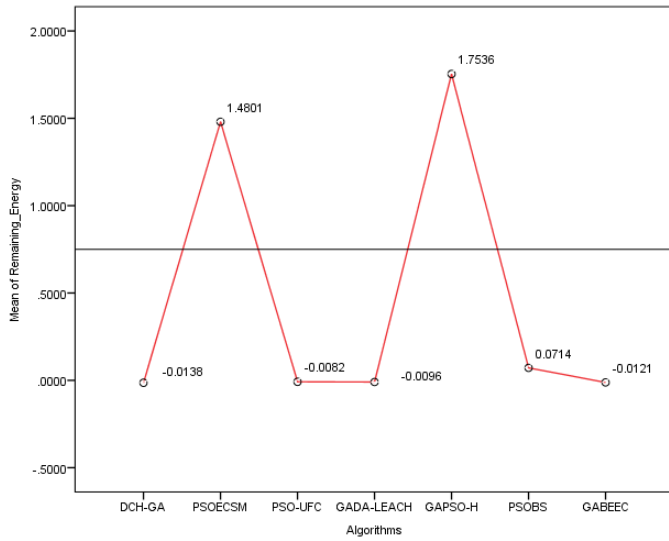
Fig. 14. Estimated marginal mean plot for remaining energy with respect to each algorithm (sample size (n) = 30)
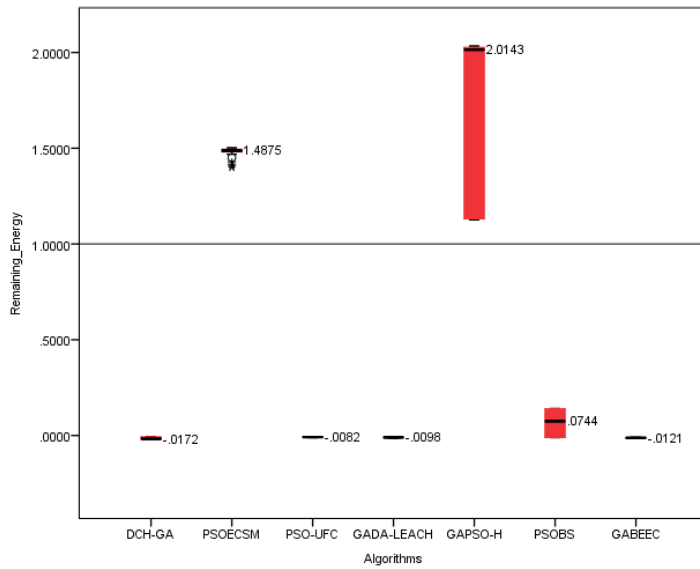

Fig. 15. Box plot for remaining energy with respect to each algorithm (sample size (n) = 30).

Figure 14 displays mean plot for remaining energy with respect to each algorithm. The X-axis shows the algorithms whilst Y-axis presents the mean of the remaining energy. The data label is added for each algorithm. It can be seen that the proposed GAPSO-H outperformed all the other algorithms. Figure 15 presents the box plot for each algorithm. It can be seen that the GAPSO-H showed the best performance with median value = 2.014.

Table 8. Tukey's Hinges percentile distribution for remaining energy with respect to each algorithm.

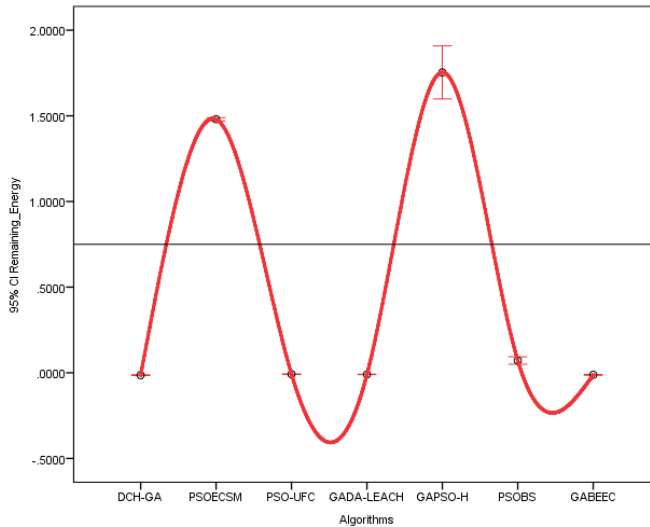| References	Algorithms	Percentiles						
		5	10	25	50	75	90	95
Elhoseny et al. [11]	DCH-GA			-.018	-.017	-.007		
Sahoo et al. [25]	PSOECSM			1.480	1.487	1.494		
Kaur et al. [26]	PSO-UFC			-.009	-.008	-.006		
Bhatia et al. [10]	GADA-LEACH			-.011	-.009	-.007		
Proposed	**GAPSO-H**			**1.131**	**2.014**	**2.027**		
Tabibi et al. [24]	PSOBS			-.010	.074	.139		
Bayrakh et al. [30]	GABEEC			-.013	-.012	-.0104		

Fig. 16. Error bar plot for remaining energy with respect to each algorithm (sample size (n) = 30).

The percentile distribution for the remaining energy is presented in Table 8. It shows 1st (25% percentile), 2nd (50% percentile, also known as the median) and 3rd (75% percentile) quartile. One can see that median value (50% percentile) = 2.014 for the GAPSO-H is the best with interquartile range 0.896. Figure 16 depicts the error bar chart where interpolation line has been

drawn to highlight the difference in the algorithm's performance. Figure 16 reveals that the performance of the GAPSO-H is the best whilst PSOECSM showed fairly good performance. It is noticed that the remaining algorithm either showed average or the worst performance.

6. Conclusion

In this paper, a GAPSO-H is presented for the CH selection and to optimize the sink mobility to acquire an optimal network performance. We implemented a GA which considers various factors such as residual energy, length, node level, average energy and energy consumption rate (ECR) during CH selection. These factors have been used as fitness parameters for determining the fitness function values. We observed that data traffic relaying in multi-hop communication have contributed to a hot-spot problem. Hence, a sink mobility has been implemented which was optimized through the PSO by considering three key factors: (a) the discovery of lowest energy CH; (b) distance to CH; and (c) cluster length. Extensive computer simulations have been conducted to show the effectiveness of the proposed GAPSO-H. We developed a robust experimental environment by tuning the control parameters for both GA and PSO. Five performance metrics have been considered to measure the performance. The performance of the proposed GAPSO-H has been tested against the state-of-the-arts methods, namely DCH-GA, GABEEC, PSOBS, GADA-LEACH, PSO-UFC and PSOECSM. The simulation results have been summarized and presented in Table 2 and Table 3. This result reveals that the proposed GAPSO-H has improved the stability period significantly. We noted that on various other performance metrics GAPSO-H has outperformed the other state-of-the-arts algorithms. Rigorous statistical tests have been performed and results have been presented. The results statistic has indicated that the performance of the proposed GAPSO-H is significantly better that the other six algorithms at 95% level of confidence interval. The current results are encouraging but it has wide scope of extension in the near future. The alternative to sink flexibility will be pursued as an immediate future work in order to make the network cost-effective and at the same time reduce the end-to-end delay. Furthermore, it will be interesting to examine the multiple sink mobility scenarios in the network through this hybrid optimized approach.

References

[1]. Abbasi AA, Younis M (2007) A survey on clustering algorithms for wireless sensor networks. Computer communications 30:2826–2841

[2]. Verma S, Sood N, Sharma AK (2019) Genetic Algorithm-based Optimized Cluster Head selection for single and multiple data sinks in Heterogeneous Wireless Sensor Network. Applied Soft Computing 105788

[3]. Hussain S, Matin AW, Islam O (2007) Genetic algorithm for hierarchical wireless sensor networks. JNW 2:87–97

[4]. Liu J-L, Ravishankar CV (2011) LEACH-GA: Genetic algorithm-based energy-efficient adaptive clustering protocol for wireless sensor networks. International Journal of Machine Learning and Computing 1:79

[5]. Singh VK, Sharma V (2013) A multiobjective coverage and connectivity strategy for improving the performance of wireless sensor networks. Int J Adv Res Computer Sci Software Eng (IJARCSSE) 3:

[6]. Kuila P, Gupta SK, Jana PK (2013) A novel evolutionary approach for load balanced clustering problem for wireless sensor networks. Swarm and Evolutionary Computation 12:48–56

[7]. Gupta SK, Jana PK (2015) Energy efficient clustering and routing algorithms for wireless sensor networks: GA based approach. Wireless Personal Communications 83:2403–2423

[8]. Elhoseny M, Yuan X, Yu Z, et al (2015) Balancing energy consumption in heterogeneous wireless sensor networks using genetic algorithm. IEEE Communications Letters 19:2194–2197

[9]. Shokouhifar M, Jalali A (2015) A new evolutionary based application specific routing protocol for clustered wireless sensor networks. AEU-International Journal of Electronics and Communications 69:432–441

[10]. Bhatia T, Kansal S, Goel S, Verma AK (2016) A genetic algorithm-based distance-aware routing protocol for wireless sensor networks. Computers & Electrical Engineering 56:441–455

[11]. Elhoseny M, Farouk A, Zhou N, et al (2017) Dynamic Multi-hop Clustering in a Wireless Sensor Network: Performance Improvement. Wireless Personal Communications 1–21

[12]. Yuan X, Elhoseny M, El-Minir HK, Riad AM (2017) A genetic algorithm-based, dynamic clustering method towards improved wsn longevity. Journal of Network and Systems Management 25:21–46

38

[13]. Hamidouche R, Aliouat Z, Gueroui AM (2018) Genetic Algorithm for Improving the Lifetime and QoS of Wireless Sensor Networks. Wireless Personal Communications 101:2313–2348

[14]. Latiff NAA, Latiff NMA, Ahmad RB (2011) Prolonging lifetime of wireless sensor networks with mobile base station using particle swarm optimization. In: 2011 Fourth International Conference on Modeling, Simulation and Applied Optimization. IEEE, pp 1–6

[15]. Hu Y, Ding Y, Hao K, et al (2014) An immune orthogonal learning particle swarm optimisation algorithm for routing recovery of wireless sensor networks with mobile sink. International Journal of Systems Science 45:337–350

[16]. Singh B, Lobiyal DK (2012) A novel energy-aware cluster head selection based on particle swarm optimization for wireless sensor networks. Human-Centric Computing and Information Sciences 2:13

[17]. Kuila P, Jana PK (2014) Energy efficient clustering and routing algorithms for wireless sensor networks: Particle swarm optimization approach. Engineering Applications of Artificial Intelligence 33:127–140

[18]. RejinaParvin J, Vasanthanayaki C (2015) Particle swarm optimization-based clustering by preventing residual nodes in wireless sensor networks. IEEE sensors journal 15:4264–4274

[19]. Ni Q, Pan Q, Du H, et al (2015) A novel cluster head selection algorithm based on fuzzy clustering and particle swarm optimization. IEEE/ACM Transactions on Computational Biology and Bioinformatics 14:76–84

[20]. Hu Y-F, Ding Y-S, Ren L-H, et al (2015) An endocrine cooperative particle swarm optimization algorithm for routing recovery problem of wireless sensor networks with multiple mobile sinks. Information Sciences 300:100–113

[21]. Rao PS, Jana PK, Banka H (2017) A particle swarm optimization-based energy efficient cluster head selection algorithm for wireless sensor networks. Wireless networks 23:2005–2020

[22]. Azharuddin M, Jana PK (2017) PSO-based approach for energy-efficient and energy-balanced routing and clustering in wireless sensor networks. Soft Computing 21:6825–6839

[23]. Edla DR, Kongara MC, Cheruku R (2019) A PSO Based Routing with Novel Fitness Function for Improving Lifetime of WSNs. Wireless Personal Communications 104:73–89

[24]. Tabibi S, Ghaffari A (2019) Energy-Efficient Routing Mechanism for Mobile Sink in Wireless Sensor Networks Using Particle Swarm Optimization Algorithm. Wireless Personal Communications 104:199–216

[25]. Sahoo, B. M., Amgoth, T., & Pandey, H. M. (2020). Particle Swarm Optimization Based Energy Efficient Clustering and Sink Mobility in Heterogeneous Wireless Sensor Network. Ad Hoc Networks, 102237

[26]. Kaur, T., & Kumar, D. (2018). Particle swarm optimization-based unequal and fault tolerant clustering protocol for wireless sensor networks. IEEE Sensors Journal, 18(11), 4614-4622

[27]. Lynn, N., & Suganthan, P. N. (2015). Heterogeneous comprehensive learning particle swarm optimization with enhanced exploration and exploitation. Swarm and Evolutionary Computation, 24, 11-24

[28]. Pandey, H. M., Chaudhary, A., Mehrotra, D. & Kendall, G. (2016). Maintaining regularity and generalization in data using the minimum description length principle and genetic algorithm: Case of grammatical inference. Swarm and Evolutionary Computation, 31, 11-23.

[29]. Heinzelman WB, Chandrakasan AP, Balakrishnan H (2002). An application-specific protocol architecture for wireless micro sensor networks. IEEE Transactions on wireless communications 1:660–670

[30]. Bayrakh, S., & Erdogan, S.Z. (2012). Genetic algorithm-based energy efficient clusters (gabeec) in wireless sensor networks. Procedia Computer Science, 10, 247-25

[31]. Gharaei N, Bakar KA, Hashim SZM, Pourasl AH (2019) Inter-and intra-cluster movement of mobile sink algorithms for cluster-based networks to enhance the network lifetime. Ad Hoc Networks 85:60–70

[32]. Goldberg DE, Holland JH (1988) Genetic algorithms and machine learning. Machine learning 3:95–99

[33]. Kennedy J (2010) Particle swarm optimization. Encyclopedia of machine learning 760–766

[34]. Pitchaimanickam, B., & Murugaboopathi, G. (2019). A hybrid firefly algorithm with particle swarm optimization for energy efficient optimal cluster head selection in wireless sensor networks. *Neural Computing and Applications*, 1-15.

[35]. Vijayalakshmi, K., & Anandan, P. (2019). A multi objective Tabu particle swarm optimization for effective cluster head selection in WSN. *Cluster computing*, *22*(5), 12275-12282.

[36]. Chandirasekaran, D., & Jayabarathi, T. (2019). Cat swarm algorithm in wireless sensor networks for optimized cluster head selection: a real time approach. *Cluster Computing*, *22*(5), 11351-11361.

[37]. John, J., & Rodrigues, P. (2019). MOTCO: Multi-objective Taylor Crow Optimization Algorithm for Cluster Head Selection in Energy Aware Wireless Sensor Network. *Mobile Networks and Applications*, *24*(5), 1509-1525.

[38]. Alghamdi, T. A. (2020). Energy efficient protocol in wireless sensor network: optimized cluster head selection model. *Telecommunication Systems*, 1-15.

[39]. Gupta, G.P., Saha, B. (2020). Load balanced clustering scheme using hybrid metaheuristic technique for mobile sink based wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing.

[40]. Krishnan, M., Yun, S., & Jung, Y. M. (2019). Dynamic clustering approach with ACO-based mobile sink for data collection in WSNs. *Wireless Networks*, *25*(8), 4859-4871.

[41]. Vijayashree, R., & Suresh Ghana Dhas, C. (2019). Energy efficient data collection with multiple mobile sinks using artificial bee colony algorithm in large-scale WSN. *Automatika*, *60*(5), 555-563.

[42]. Zhang, H., Li, Z., Shu, W. (2019). Ant colony optimization algorithm based on mobile sink data collection in industrial wireless sensor networks. *J Wireless Com Network*, 152, 1–10.

[43]. Wang, H., Li, K., and Pedrycz, W. (2020). An Elite Hybrid Metaheuristic Optimization Algorithm for Maximizing Wireless Sensor Networks Lifetime with a Sink node. *IEEE Sensors Journal*.

[44]. Pandey, H. M., Chaudhary, A., & Mehrotra, D. (2014). A comparative review of approaches to prevent premature convergence in GA. Applied Soft Computing, 24, 1047-1077.

[45]. Pandey, H. M., Chaudhary, A., & Mehrotra, D. (2016). Grammar induction using bit masking oriented genetic algorithm and comparative analysis. Applied Soft Computing, 38, 453-468.

[46]. Pandey, H. (2017, September). Performance Review of Harmony Search, Differential Evolution and Particle Swarm Optimization. In IOP Conference Series: Materials Science and Engineering (Vol. 225). IOP Publishing.

[47]. Pandey, H. M. (2016). Performance evaluation of selection methods of genetic algorithm and network security concerns. Physics Procedia, 78, 13-18.

[48]. Dhiman, G., Soni, M., Pandey, H. M., Slowik, A., & Kaur, H. (2020). A novel hybrid hypervolume indicator and reference vector adaptation strategies based evolutionary algorithm for many-objective optimization. Engineering with Computers, 1-19.