Applied Soft Computing Journal 85 (2019) 105788

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Genetic Algorithm-based Optimized Cluster Head selection for single and multiple data sinks in Heterogeneous Wireless Sensor Network

Sandeep Verma [a,*], Neetu Sood [a], Ajay Kumar Sharma [b]

[a] Department of ECE, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar-144011, India
[b] National Institute of Technology, Delhi-110040, India

ARTICLE INFO

Article history:
Received 13 March 2018
Received in revised form 24 July 2019
Accepted 9 September 2019
Available online 23 September 2019

Keywords:
Genetic Algorithm
Multiple data sinks
Heterogeneous Wireless Sensor Network
GAOC
MS-GAOC
Clustering
CH selection

ABSTRACT

The constraints on the battery resources of sensor nodes have been the major stumbling block in achieving the network longevity and in exploring the potential of Wireless Sensor Network (WSN) to the maximum level. A plethora of research work has implemented multitudinous optimization techniques for the Cluster Head (CH) selection in homogenous WSN. However, for Heterogeneous WSN (HWSN), the CH selection is still left with a wide scope for further improvement for its exploitation capabilities. In this paper, Genetic Algorithm-based Optimized Clustering (GAOC) protocol is designed for optimized CH selection by integrating the parameters of residual energy, distance to the sink and node density in its formulated fitness function. Furthermore, to pact with the Hot-Spot problem, and to shorten the communicating distance from the nodes to the sink, Multiple data Sinks based GAOC (MS-GAOC) is proposed. The empirical investigations of MS-GAOC is carried out with protocols developed to operate with multiple data sinks so as to have fair comparative analysis. It is inferred from the simulation analysis that the GAOC and MS-GAOC outperform the state-of-the-art protocols on the benchmark of different performance metrics viz. stability period, network lifetime, number of dead nodes against rounds, throughput and network's remaining energy. The proposed protocols are expected to play a salient role in monitoring of hostile applications, i.e., forest fire detection, early detection of volcanic eruptions, etc.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

The emergence of wireless sensing-based technology has been one of the major advancements in the revolutionary growing technologies [1,2]. Wireless Sensor Network (WSN) is a network that connects various battery operated nodes. These nodes possess the capabilities to sense physical quantities viz. temperature, moisture, vibration, etc., handle data storage, compute and process signals, and undergo wireless communication [3]. These can be exploited to operate for various applications that can be categorized into four main genres stated as follows [4].

(a) Event-dependent (sensor nodes stay in sleep mode until an event is detected and upon happening of an event, it gets activated)
(b) Query-dependent, initiated from a sink (sensor nodes gets activated on query generated from a sink or else they remain in sleep mode)
(c) Periodic measurements (data dissemination is done at periodical intervals)
(d) Tracking based applications (combines the above three genres)

WSN has provided the unprecedented opportunities in various domains which range from military applications to agricultural monitoring and also play a salient role in various sectors viz. environmental, civil, industrial control, home automation, etc. [5].

However, the enormous potential of WSN for its maximum exploration, is limited by the power constraints on sensor nodes. Most of the researchers have focused on efficiently utilizing the energy of sensor nodes so as the network longevity can be achieved [6].

The traditional inefficient routing protocols i.e., Direct Transmission and Minimum Transmission Energy protocols failed to conserve the energy of nodes [7]. To address this issue, clustering was proposed in the first ever cluster-based hierarchical routing protocol i.e., Low Energy Adaptive Clustering Hierarchy (LEACH) [7]. It proved its worth against the aforementioned classical protocols for the sole aim of energy preservation. It basically operated for the homogeneous network in which all the nodes were having same capabilities in terms of energy, computation and coverage sources. Subsequently, many protocols based on LEACH architecture, were proposed which improved the network

* Corresponding author.
E-mail addresses: sandeepv.ec.13@nitj.ac.in (S. Verma), soodn@nitj.ac.in (N. Sood), director@nitdelhi.ac.in (A.K. Sharma).

https://doi.org/10.1016/j.asoc.2019.105788
1568-4946/© 2019 Elsevier B.V. All rights reserved.

longevity tremendously. Although, LEACH is homogeneous protocol, but ideally, homogeneity in sensor nodes does not exist due to manufacturing differences, different morphological factors, uneven physical terrain, etc. [8]. When the considered nodes are of different configuration in terms of aforementioned factors, the nodes are said to be heterogeneous nodes and hence the corresponding network framed is termed as a heterogeneous network. Among different genres of heterogeneity, energy heterogeneity is the most commonly exploited due to its prominent dependence on irreplaceable battery resources of nodes [9].

For a cluster-based routing in above-discussed modes of the network, a Cluster Head (CH) is selected in every cluster on the basis of some parameters. After data collection by the CH from the cluster nodes, data aggregation is performed to remove the data redundancy [10]. Consequently, CH suffers from more energy drainage as compared to the cluster member nodes.

The collected data from the cluster members is forwarded to the sink in a single or multi-hop communication and then it is forwarded to the user via internet sources as illustrated in Fig. 1 [11,12].

Many researchers have propounded their research strategies on optimizing the CH selection in the homogeneous network by exploiting some metaheuristics methods [13]. However, in Heterogeneous WSN (HWSN), the CH selection has been achieved through some tactical amendments in the threshold-based formula used for the nodes [10]. The optimization technique for CH selection aims to minimize its energy consumption. However, the optimized CH selection towards most energy efficient routing is a non-deterministic polynomial-time hard (NP-hard) problem. Nevertheless, the selection of CH can be optimized through some metaheuristics methods that incorporate some key factors for CH selection in the process of building up the fitness function [14]. While doing so, the metaheuristic method is applied on the grounds of its characteristics to converge it to an optimal solution.

So far, multitudinous optimization routing strategies have been developed to optimize the CH selection so as to ensure the reliability of the network undergoing data transmission [15]. Genetic Algorithm (GA) [16–18], PSO (Particle Swarm Optimization) [19–21] and Artificial Bee Colony (ABC) [22–24] are some of the prominent metaheuristic techniques taken into consideration for the optimized CH selection. Among the aforementioned optimization techniques, we employ a GA based CH selection algorithm, as it ensures a gradual improvement in the solution optimization due to its robust nature. Furthermore, it basically focuses on the energy saving of nodes.

It is to be noted that the one of the crucial energy devouring element in nodes, is the wireless radio that is responsible for communication to the other nodes and sink [25]. The transmission power in nodes decays in direct proportion to the distance within nodes or between nodes and sink by a 'distance factor squared' or higher order. In case of those applications, where Region of Interest (ROI) covers a large area, the long haul transmission from far placed nodes communicating in a single-hop lead to their energy depletion. Ultimately, the whole network stops functioning. Therefore, employing multi-hop communication in a network becomes inevitable that eventually leads to energy depletion of relaying nodes. After transmission of data packets over the time, when relaying nodes are completely exhausted of their energy, a no connection zone, i.e., Hot-Spot is created [26]. Although various algorithms are proposed in mitigating the Hot-Spot problem with a single sink intact, it is still a non-trivial problem.

To address this concern, we employ multiple sinks outside the ROI, making it possible for CH nodes to have single-hop communication to the nearest sink. Our main contribution can be listed as follows.

(a) *A novel approach for CH selection using GA in HWSN*:
In this paper, Genetic Algorithm-based Optimized Clustering protocol (GAOC) is proposed. The CH selection in GAOC is done by the metaheuristic GA that formulates its fitness function by integrating the parameters of residual energy, distance to the sink, and node density. To the best of our knowledge, it is the first ever consideration of genetic algorithm in HWSN with aforementioned parameters of CH selection.

(b) *Incorporating multiple data sinks outside ROI to mitigate Hot-Spot Problem:*
Multiple data sinks are employed outside ROI to alleviate the Hot-Spot problem hence, the proposed protocol is termed as Multiple Sink based-GAOC (MS-GAOC). It follows the same parameters for defining the CH selection as that of GAOC.

(c) *Performance evaluation of GAOC with state of art protocols and with MS-GAOC:*
The performance of GAOC is evaluated against the state-of-the-art protocols on the basis of performance metrics viz. stability period, network lifetime, number of dead nodes against rounds, throughput, and network's remaining energy. Furthermore, the performance evaluation of GAOC is done against MS-GAOC to introspect the essence of multiple data sinks employment outside the network. Lastly, for the fair comparative analysis, MS-GAOC is compared with the multiple data sinks based version of protocols taken into consideration for the comparison with GAOC.

The organization of rest of the manuscript is done as follows. Section 2 comprehensively discusses the related work and Section 3 presents the proposed system framework of GAOC and MS-GAOC. The result and discussions are presented in Section 4 followed by conclusion in Section 5. Then references are listed.

2. Related work

The recent unparalleled advancements in WSNs have fostered the researchers to design various routing protocols to pact with its most daunting constraint related to energy resources [27]. Among different routing techniques, cluster-based routing helps in distributing the load among sensor nodes in such a way that the energy consumption is balanced between them. The plethora of research has reported the cluster-based routing with homogeneous nodes having same stock of energy, connectivity, and computational and coverage capabilities [28]. However, homogeneity does not exist ideally, in fact, in consideration to the desideratum of crucial applications operating in harsh environment, some nodes with extra stock of energy will assist in acquiring network longevity [9]. One such application is the detection of forest fire that has been a serious concern over the years due to the heavy damage caused by the various wildfires happening all across the globe [29,30]. In the similar direction of fire detection, one of the attempts was reported that aimed to optimize the energy distribution of WSNs using harmony search algorithm in real time environment [31]. However, it has some serious drawbacks in various aspects that include inefficient CH selection, centralized approach, long haul transmission, etc.

Many authors have reported the retrospective review of the cluster-based routing protocols [32]. The concept of clustering was first ever introduced by the LEACH protocol [7]. The paramount concern of the cluster-based routing has been the formulation of certain parameters for the selection of CH.

The work done in the current paper focuses on the cluster-based routing in HWSN, so as the CH selection can be introspected based on the concept of the threshold value. Samargiks et al. [33]

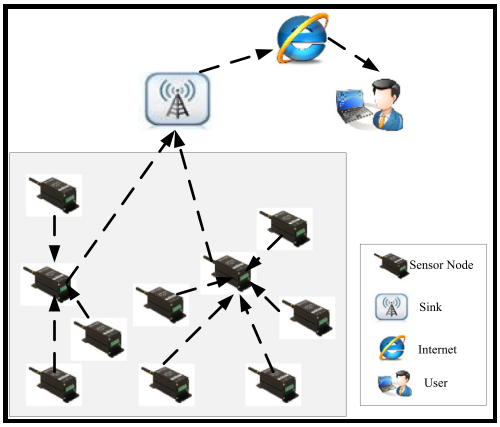

Fig. 1. Architecture of Wireless Sensor Network with cluster based routing.

reported the first heterogeneous protocol namely Stable Election Protocol (SEP) that operated at two energy levels of sensor nodes.

Some of the heterogeneous protocols focusing on CH selection are presented in Table 1, which will help in classifying these protocols in an appropriate genre.

The research gap corresponding to each protocol is discussed along with its characteristics like the number of heterogeneity level, number of sinks employed, hot-spot problem, mode of communication, reactive/ proactive protocol and fixed or variable number of CH.

The protocols SEP [33], DEEC [34] and DDEEC [35] operated at two energy levels but are not efficient enough for large area networks due to the single-hop communication employed for data transmission. CH selection was deprived of the distance factor in aforementioned protocols which led to the consumption of high magnitude of energy. DEEC suffered from the penalization of high energy nodes in a sense that the nodes are made CH, more frequently irrespective of their relative low energy at any moment of data transmission. DDEEC resolved this concern by introducing energy threshold for nodes selected as CH. EEHC [36] made use of three-level energy heterogeneity and selected CH on the basis of weighted election probability. However, it suffered from the penalization of high energy nodes which was further resolved by EDDEEC [37] following similar approach to that of DDEEC [35]. BEENISH [38] with four energy level of heterogeneity enhanced stability period by proposing energy efficient CH selection. Aforementioned protocols worked towards acquiring the ameliorated stability period and improving network lifetime. Paola et al. [39] proposed P-SEP that followed random selection approach for selecting CH rather than prioritizing the high energy nodes for same.

Above-discussed protocols worked for proactive applications, however, the protocols TSEP [40], DRESEP [41], SEECP [42] and TEDRP [43] were developed to work for reactive applications viz. forest fire detection, early detection of volcanic eruptions, etc. DRESEP, SEECP and TEDRP suffered from the Hot-Spot problem due to dual hop inter-cluster communication.

It is contemplated from the heterogeneous protocols integrating various parameters for CH selection that there is a need for some metaheuristic method that follows the inclusion of every significant parameter which can enhance network performance.

In consideration of the fact that CH selection is a dynamic process in which the role of CH is rotated on some predefined attributes, the optimization techniques play a significant role.

Furthermore, GA is one of the most popular optimization technique that is population-based and converges to the fittest solution [16]. While adopting GA, the selection of CH goes through various iterations while converging to an optimal solution.

The comprehensive review of GA based metaheuristic techniques is given in Table 2 to highlight the objectives and research gaps of evolutionary metaheuristics evolved in clustering. GA has been rigorously applied for the optimization of various cluster-based routing techniques as discussed in Table 2. These protocols are discussed further in Table 3 to highlight the GA process involved and the integration of fitness function incorporated for CH selection.

Khanna et al. [44] proposed GA-based optimization for multi-hop sensor network to acquire minimum energy consumption. As it involves multi-hop communication, Hot-Spot problem arises in the network. Hussain et al. [45] utilized GA to formulate energy efficient clusters for data collection. The random selection of CH and single-hop communication make the proposed protocol an energy inefficient for large area networks.

Bari et al. [46] utilized relay nodes as CHs for forwarding the data to the sink and inspected the performance of proposed GA based routing at different values of path loss component.

Norouzi et al. [47] investigated GA to obtain optimum states however, it suffered from long-haul transmission due to single-hop communication involved. Bhaskar [48] and Bayrakl et al. [49] proposed GA based GDSDA and GABEEC protocols, respectively. The former worked for the security aspects along with a reduction in transmission power, whereas, latter proposed a protocol to achieve network longevity. However, these protocols were observed to be non-scalable. Kuila et al. [50] utilized GA for bringing load balancing in the network. The proposed technique did not discuss about the clustering and CH selection in its approach. Kumar et al. [51] proposed GDFEC that utilized Fuzzy interference system to obtain total dissipated energy which was further optimized by GA.

The other GA based optimized protocols viz. GACR [52], GASONeC [53], DCH-GA [54], GAFTC [55], GADA-LEACH [56] and the one proposed by Pourzaferani et al. [57], are discussed in the context of their objectives and corresponding research gaps in Table 2.

Ragavan et al. [58] applied GA for dynamic topology deployment of sensor nodes and hence, it applied Tabu search technique to find the best optimal route.

Furthermore, it is observed from Table 3 that unlike other protocols, Khanna et al. [44], GASONeC, DCH-GA and GAHN have incorporated node density factor in their fitness function evaluation for the selection of CH. These protocols employ multi-hop communication consequently, suffers from Hot-spot problem.

It is inferred from the retrospective study that only sprinkling of research work has focused on the optimization in HWSN. Moreover, while going through the rigorous review of heterogeneous protocols, it is reported that almost all of the protocols have employed sink in the middle of the network for data dissemination that pertains to the non-hostile and attended applications.

However, for the applications (forest fire detection) where human intervention is not feasible, sink has to be placed outside the network and hence multi-hop communication has to be introduced to avoid long haul transmission. Consequently, the Hot-Spot problem arises [59,60].

In addition to that, in case of single sink failure, the whole network is disconnected and data is completely lost. To address this problem, the approach of introducing multiple data sinks works appropriately to minimize the communication hop between CH nodes and sink that enhances the reliability of the network.

It is observed that much of the research considering multiple data sinks, have been reported either aiming to find the optimal placements of sinks or for the sensor to sink binding by employing some optimization technique.

Table 1
Comparative study of state-of-the-art heterogeneous routing protocols.

Study reference	Name of protocols	Heterogeneity level	No. of sink	Reactive/ Proactive	Mode of communication	Hot-spot problem	No. of CH fixed	CH selection based on						Research gap
								Initial energy	Residual energy	Total energy	Average energy	Distance	Node density	
Smaragdakis et al. (2004) [33]	SEP	2	1	Pro-active	single-hop	No	×	✔	×	×	×	×	×	Not suitable for multi-level
Qing et al. (2006) [34]	DEEC	2	1	Pro-active	single-hop	No	×	×	✔	×	✔	×	×	Penalization of high energy nodes
Kumar et al. (2009) [36]	EEHC	3	1	Pro-active	single-hop	No	×	×	×	×	×	×	×	Energy of nodes is not considered for CH selection
Elbhiri et al. (2010) [35]	DDEEC	2	1	Pro-active	single-hop	No	×	×	✔	×	✔	×	×	Not suitable for multi-level heterogeneity
Javaid et al. (2013) [37]	EDDEEC	3	1	Pro-active	single-hop	No	×	×	✔	×	✔	×	×	Not suitable for multi-level heterogeneity
Qureshi et al. (2013) [38]	BEENISH	4	1	Reactive	single-hop	No	×	×	✔	×	✔	×	×	Penalization of high energy nodes
Kashaf et al. (2012) [40]	TSEP	3	1	Reactive	single-hop	No	×	×	×	×	×	×	×	Energy factor not included for CH selection
Kumar et al. (2015) [41]	DRESEP	3	1	Reactive	dual hop	Yes	×	×	✔	×	×	✔	×	Pre-fixed circular radius in random deployment scenario
Kumar et al. (2016) [42]	SEECP	3	1	Reactive	dual hop	Yes	✔	×	✔	×	✔	×	×	In-efficient selection of circular radius for dual hop comm.
Paola et al. (2017) [39]	P-SEP	2	1	Pro-active	single-hop	No	×	×	×	×	×	×	×	Selects CH randomly, No preference to advanced node
Kumar et al. (2017) [43]	TEDRP	3	1	Reactive	dual hop	Yes	×	×	✔	×	×	×	×	Hot-Spot Problem exists

The some of the protocols that employed multiple sinks in the network are QAZP [61], LP-KPS [62] and REBTAM [63]. LP-KPS was the 1st ever protocol to utilize multiple sinks for reducing the congestion problem in HWSN. It utilized two energy level of heterogeneous nodes to achieve the objective of load balancing in the network. However, it is not scalable enough to operate in large area networks and it also did not discuss the stability period of the network which is a quite crucial performance metric in HWSN.

In nutshell, following observations are formulated from the extensive literature work done in this paper.

(a) GA based optimized CH selection that employs node density factor along with energy and distance in its fitness function specifically for HWSN, is untouched so far in the literature work reported in this paper.
(b) To address the Hot-Spot problem and to pact with hostile applications, incorporation of multiple data sinks employing optimization technique for CH selection, may prove to be a promising approach. This imperative approach is still missing in the literature work reported till date.
(c) The performance evaluation of single sink and multiple data sinks in the same paper, for any proposed strategy, has never been reported till date. To the best of our knowledge, no research work has created multiple data sinks based version of existing protocols so as to have fair comparative analysis with proposed multiple data based technique.

The optimization heuristic based on GA is employed for the selection of CH in the scenario of single sink and multiple data sinks, is discussed as follows.

3. The proposed system framework of GAOC and MS-GAOC

This section explains the GAOC protocol by covering different steps of operation performed under GA. The basic overview and rationale behind choosing GA over other optimization techniques, are discussed as follows.

(a) **Overview and significance of GAs over other optimization techniques**
GAs are one of the most effective 'stochastic optimization search processes' which help in imitating the adaptive evolution procedure available in nature. GAs came into existence in the year 1970s, and since then it has been used actively in determining the solution for the optimization problems in various fields viz. computer networking, industrial engineering, machine learning, etc. [16]. GA works on the principle of Darwin's Theory, stating that best-adapted individuals of species will survive and others will get extinct. The most predominant significance of GA is in the problems of combinatorial and multi-objective optimization that are hard to determine.
Basically, GAs are those stochastic search engines that mimic the natural selection and biological evolution processes. Primarily, randomly selected individuals generated from the candidate solution space are contained by the population. Then these selected individuals are evolved through different successive generations via selection, crossover, and mutation processes. Thereafter, in order to improve the quality, fittest individuals are selected so as they can generate a new population of individuals for acquiring the objectives.
(b) **Rationale behind employing GAs over other optimization techniques**
In comparison with the conventional optimization methods [64,65], GAs are highly effective techniques. It is due to the fact that research in GAs is based on the parallel computations performed by the population of individuals rather than the operation of a single point. It helps in the improvement of chances of achieving the global optimum solution by avoiding the local stationary points. Furthermore, the role of fitness function used for evaluation rather than derivatives helps in extending the GAs to any kind of continuous or discrete optimization problems. It is reported that the best possible utilization of GAs is for the applications which suffer from difficulty to access domain knowledge and derivative information. In simple words, GAs reduces the complexity of the knowledge-based process by exploring only fitness function and the whole process is controlled by the fitness parameters taken into consideration.

3.1. GA operation of GAOC and MS-GAOC

GAOC and MS-GAOC are operated in energy heterogeneous WSN in which some of the nodes are equipped with extra source of energy. These high energy nodes are expected to play a paramount role as they will be collecting data from the other nodes in the network. These nodes ensure such task efficiently performed by using GA based algorithm.

The structure of a number of heterogeneous nodes and their respective energy compositions, is discussed as follows. In the proposed protocols, the total number of advanced, intermediate and normal nodes represented by N_{ADV}, N_{INT} and N_{NRM}, given by a set of Eqs. (1)–(9). In these equations, the proportions of a number of intermediate nodes and advanced nodes are represented by m_o and m, respectively.

$$\mathrm{N_{ADV}} = \mathrm{n} \times \mathrm{m} \tag{1}$$

$$\mathrm{N_{INT}} = \mathrm{n} \times m_o \tag{2}$$

$$\mathrm{N_{NRM}} = \mathrm{n} \times (1 - \mathrm{m} - m_o) \tag{3}$$

The advanced nodes α and intermediate nodes are β times higher in energy as compared to normal nodes. Energy fractions of intermediate and advanced nodes are denoted by β and α. In the Eqs. (4)–(7), E_{ADV}, E_{INT}, and E_{NRM} represent the energy of advanced, intermediate and normal nodes, respectively. The total energy of the network is represented by E_T given by Eq. (9) and is computed as follows.

$$\mathrm{E_{ADV}} = \mathrm{E_O} \times (1 + \alpha) \times \mathrm{n} \times \mathrm{m} \tag{4}$$

$$\mathrm{E_{INT}} = \mathrm{E_O} \times (1 + \beta) \times \mathrm{n} \times m_o \tag{5}$$

$$\mathrm{E_{NRM}} = \mathrm{E_O} \times (1 - \mathrm{m} - m_o) \times \mathrm{n} \tag{6}$$

$$\mathrm{E_T} = \mathrm{E_{ADV}} + \mathrm{E_{INT}} + \mathrm{E_{NRM}} \tag{7}$$

$$\mathrm{E_T} = \mathrm{E_O} \times (1 + \alpha) \times \mathrm{n} \times \mathrm{m} + \mathrm{E_O} \times (1 + \beta) \times \mathrm{n} \times m_o + \mathrm{E_O} \times (1 - \mathrm{m} - m_o) \times \mathrm{n} \tag{8}$$

$$\mathrm{E_T} = \mathrm{n} \times \mathrm{E_O} \times (1 + \beta \times m_o + \mathrm{m} \times \alpha) \tag{9}$$

The aforementioned equations give the laconic description of the heterogeneous structure of proposed protocols. These set of equations are used in building up network models for both protocols, discussed in further sections.

However, there is no such integrated mechanism provided by the classical optimization method as presented by GA which can ensure the alignment of roles according to nodes configuration. In addition, it might be the case that the random initialization and GA operations could introduce chromosomes that tend to violate the current sensor properties. So in the proposed work, heterogeneity is considered as constraints and hence there is a requirement of validation process before evaluation of chromosomes' fitness so as to ensure network integrity. In order to

Table 2
Comparative study of GA based routing protocols.

Study reference	Optimization technique	Objectives considered to improve	Research gap
khanna et al. (2006) [44]	GA	The optimal number of sensor clusters with Cluster-Heads, minimization of the power consumption	Performance evaluation is not done with the state-of-the-art protocols, Inefficient clustering
Hussain et al. (2007) [45]	GA	Network lifetime	Selection of CH is done randomly, inefficient inter-cluster communication
Bari et al. (2009) [46]	GA	Network lifetime at different path loss component, decreased execution time	Performance degrades at a higher value of path loss component, inefficient due to distance factor exclusion, long-haul transmission when sink place at the corner
Norouzi et al. (2011) [47]	GA	Network Lifetime, Reduction in energy consumption	Long-haul transmission makes routing inefficient, the inefficient fitness function formulation due to few parameters
Kuila et al. (2013) [50]	GA	Execution time, energy consumption, number of active sensor nodes, number of alive CHs, the rate of convergence	CH selection and clustering is not discussed
Bhaskar (2013) [48]	GA	Minimizes the energy consumption, ensures data security and reduces the transmission overhead	CH selection is done only on the factor of connectivity, network lifetime is not considered, does not focus on routing but security
Bayrakl et al. (2013) [49]	GA	Network Lifetime	Performance evaluation is done with LEACH not with the state-of-the-art protocols, Inefficient clustering
Kumar et al. (2015) [51]	GA, Mamdani Fuzzy Inference system	Network lifetime, Residual energy, packet delivery	Inefficient routing, distance parameter is not considered which enhances energy consumption, marginal improvement
Elhoseny et al. (2015) [54]	GA	Network lifetime, Residual energy	Non-scalable, long-haul transmission makes routing inefficient
Bhatia et al. (2016) [56]	GA	Network lifetime	Selection of relay nodes is not discussed, Inefficient routing, Hotspot problem, only network lifetime is considered for performance evaluation
Rajeswari et al. (2016) [55]	GA	Packet delivery ratio, throughput, an end to end delay, packet drop	Network lifetime is not discussed, marginal improvement, Hotspot problem, lack of results for performance evaluation
Pourzaferani et al. (2017) [57]	GA	Network Throughput, Delay, Network Lifetime, Average residual Energy	Hot-Spot problem, complex topology, too much overhead
Gupta et al. (2017) [52]	GA	First gateway node dead, reducing the distance and hop count	Tradeoff is observed b/w hop count and distance covered, Hot-spot problem, inefficient CH selection
Yuan et al. (2017) [53]	GA	Network longevity, scalability, dynamic network structure	Increased computational complexity, Inefficient routing due to long-haul transmission when sink placed at the boundary, No strategy defined for multi-hop communication for large network, No-clear demonstration of network is provided
Elhosen et al. (2017) [54]	GA	Dynamic clustering environment, average time, remaining energy of nodes	Hot-Spot problem, non-scalable, complex structure of the algorithm
Ragavan et al. (2018) [58]	GA, Tabu Search	QoS parameters; Packet Delivery, Reliability	Suffers from Hot-Spot problem, Network lifetime is not discussed, unstable network

pursue validation process, each gene plays an expected role for the corresponding sensor node, i.e., whether a sensor node will be a CH or a member node represented by bits '1' or '0', respectively. Validation process assists the initialization of population to be considered for further evaluation for the optimization.

3.1.1. Initialization

The initialization is executed after the validation process of chromosomes. There are following steps that are considered for the initialization.

(a) Firstly, the network parameters which are responsible for the network performance are initialized. It basically includes dimensions of network area taken into consideration, number of nodes, location, number of sinks used, etc.
(b) Thereafter, the energy parameters are initialized which comprises the energy for operating 'transmitter and receiver of a node circuitry' and energy consumed in 'data aggregation and amplification' of data packets.

After initialization, fitness function is formulated which is discussed as follows.

Table 3
Relative comparison of GA based CH selection in different routing protocols.

Study reference	Optimization technique	Type of selection and crossover	Name of technique	homo/ hetero/mode	Mode of communication	Hot-spot problem	Fitness Function Includes							
							Initial energy	Residual energy	Total energy	Average energy	Distance	Total no. of nodes	Node density	Total dissipated energy
Khanna et al. (2006) [44]	GA	Many point	–	homo	Multi-hop	✔	×	✔	×	×	×	✔	✔	×
Hussain et al. (2007) [45]	GA	Tournament, Single point	–	homo	Single-hop	×	×	×	×	×	✔	×	×	✔
Bari et al. (2009)(46)	GA	Roulette wheel, random point (1,2,3)	GA based algorithm	homo	Multi-Hop	✔	✔	×	×	×	×	×	×	✔
Norouzi et al. (2011) [47]	GA	Roulette wheel, Single point	–	homo	Single-hop	×	×	×	×	×	×	×	×	✔
Kuila et al. (2013) [50]	GA	Tournament selection, Single point	–	homo	Single-hop	×	×	×	×	×	×	✔	×	×
Bhaskar (2013) [48]	GA	Roulette wheel, Single point	GDSDA	homo	Single-hop	×	×	✔	×	×	✔	✔	×	×
Bayrakl et al. (2013) [49]	GA	Roulette wheel, Single point	GABEEC	homo	Single-hop	×	×	✔	×	✔	×	×	×	×
Kumar et al. (2015) [51]	GA, Mamdani Fuzzy Inference system	–	GDFEC	hetero	Multi-hop	✔	✔	✔	×	×	×	×	×	✔
Elhoseny et al. (2015) [54]	GA	Tournament, ___	GAHN	hetero	Single-hop	×	✔	✔	×	×	✔	×	✔	✔
Rajeswari et al. (2016) [55]	GA	Roulette wheel, Two-point	GAFTC	homo	Multi-Hop	✔	✔	✔	×	×	✔	×	×	×
Bhatia et al. (2016) [56]	GA	Roulette wheel, Single point	GADA-LEACH	homo	Multi-Hop	✔	×	✔	×	×	✔	×	×	×
Pourzaferani et al. (2017) [57]	GA	Roulette wheel, Single point	NA	homo	Multi-hop	✔	×	✔	✔	✔	✔	✔	×	✔
Gupta et al. (2017) [52]	GA	Roulette wheel, Single point	GACR	homo	Multi-hop	✔	×	✔	×	×	✔	×	×	×
Yuan et al. (2017) [53]	GA	–	GASONeC	homo	Single	×	✔	✔	×	×	✔	×	✔	✔
Elhosen et al. (2017) [54]	GA	Roulette wheel, Single point	DCH-GA	homo	Single and Multi-Hop	✔	×	✔	×	×	✔	×	✔	×
Ragavan et al. (2018) [58]	GA, Tabu Search	–	–	homo	Multi-hop	✔	✔	×	✔	×	✔	×	×	✔

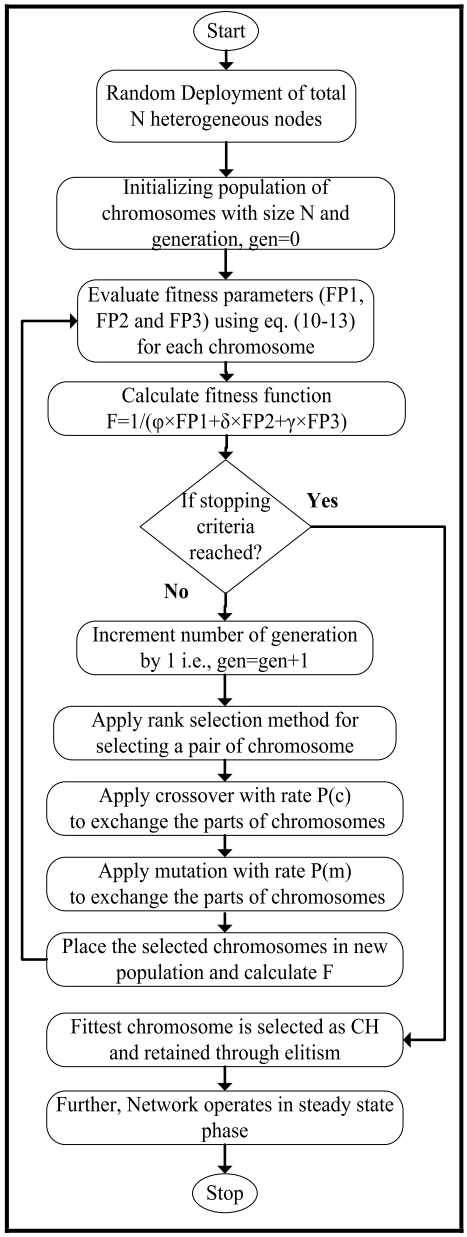

Fig. 2. Flow chart for GA operation in GAOC.

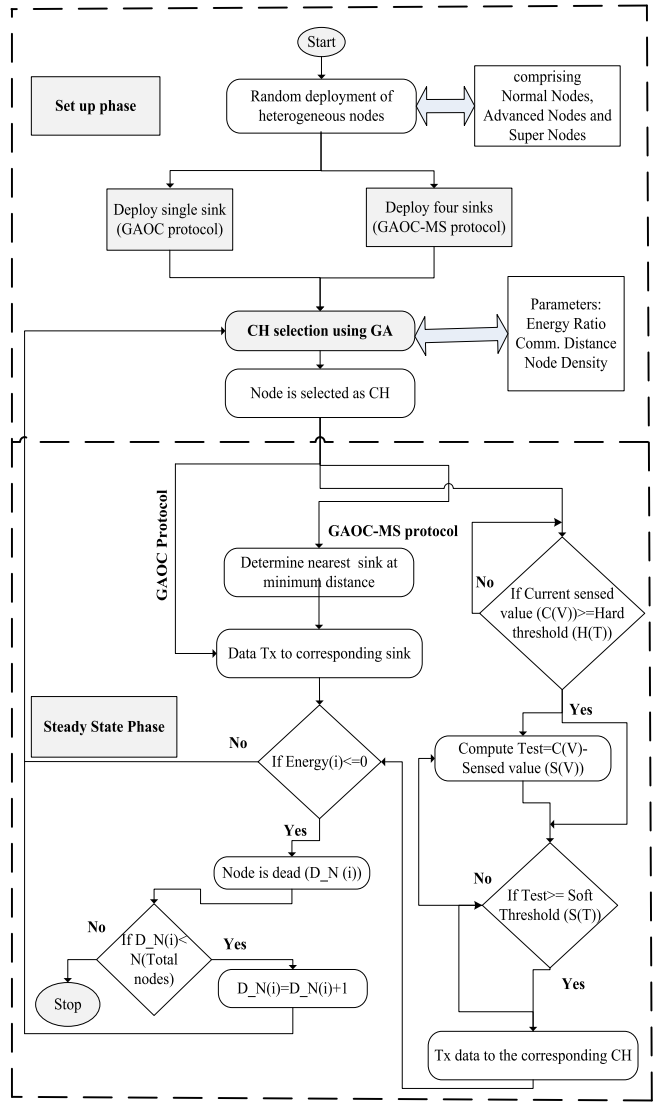

Fig. 3. Flow chart for working process of GAOC and GAOC-MS.

3.1.2. Fitness function

One of the critical issue to a successful GA performance, is the building up of fitness function. It is due to the fact that it forms the basis for the selection process and subsequently, it facilitates improvements. Improvements are defined in two perspectives; firstly, from the problem-solving perspective in which the task to be solved is represented by the proposed strategy by GA and secondly, from the technical aspects in which the assignments of quality measures to the individual solutions are done.

The fitness function is essentially a cost function which is basically mathematically expressed and developed for any attribute which is to be optimized. The weak individuals considered for fitness evaluation are eliminated from the population and fittest individuals are identified. Therefore, a chromosome is said to be fittest if it converges to an optimal point by bringing its fitness evaluation closer to that point. Fitness function integrates some essential fitness parameters which are discussed as follows.

3.1.3. Fitness parameters (FPs)

The fate of any individual in GA is decided by its fitness value that is further used in designing the fitness function. The FPs are directed to minimize the energy consumption and maximize the stability period of the network. Three FPs are considered in the selection of CH in the network which are reported as follows.

(a) The energy factor of a node

Due to the specific responsibility of collecting, aggregating and then forwarding the data, the CH results in consuming more energy as compared to the other sensor nodes. After every round, CH depletes its energy faster than the normal nodes. Therefore, the periodical rotation of CHs on the basis of residual energy becomes essential. So the factor of residual energy of a node, becomes the most prominent entity to be included in formulating fitness function.

The First Fitness Parameter (FP1) for designing the fitness function is given by following Eq. (10).

$$\mathrm{FP1} = \sum_{i=1}^{N} \left(\frac{\mathrm{E_{R(i)}}}{\mathrm{E_{Max}}\,(i)} + \frac{\mathrm{E_{TH}}}{\mathrm{E_T}} \right) \tag{10}$$

It shows the ratio of residual energy of ith node represented by E_R to the maximum initial energy of a node represented by

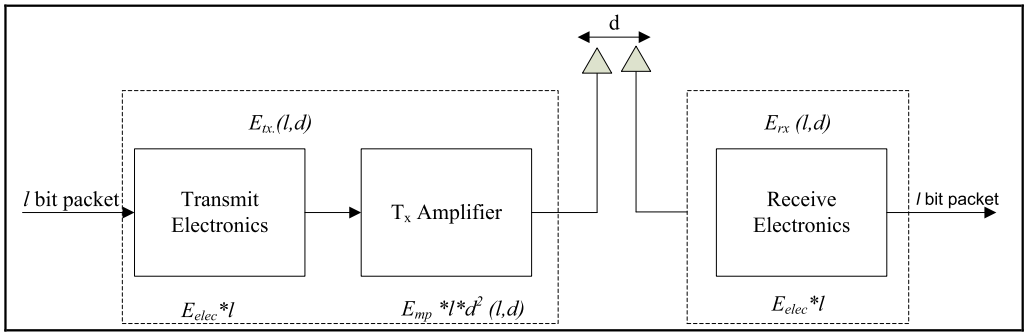

Fig. 4. Radio Energy Dissipation Model.

E_{Max}. It is to be noted that for each types of heterogeneous nodes, the value of E_{Max} will be different. The value of i used in the summation, ranges from 1 to N (total number of nodes in the network).

The energy threshold value is given by E_{TH}. The value of E_T is computed from the Eq. (9) and E_{TH} is the energy threshold that is incorporated to avoid the penalization of high energy nodes i.e., the frequent selection of high energy nodes as CH. The value of E_{TH} is calculated through number of successive simulations. So, *FP1* is the summation of the energy ratios as contemplated from Eq. (10). It is evident that higher value of *FP1* favors the selection of a node to be CH.

(b) Distance between node and sink

The nodes consume their energy when they communicate with other nodes or sink placed at some distance. More the distance a node covers to communicate, higher energy will be consumed and vice versa. Therefore, the Second Fitness Parameter (FP2) for designing the fitness function for the CH selection pacts with distance factor and is given by Eq. (11).

$$\mathrm{FP2} = \sum_{i=1}^{N}\left(\frac{\mathrm{D_{N-S(i)}}}{\mathrm{D_{F(N-S)}}} + \frac{1}{\mathrm{D_{AVG(N-S)}}}\right) \tag{11}$$

$$\mathrm{D_{AVG(N-S)}} = \frac{\sum_{i=1}^{N}\mathrm{D_{N-S(i)}}}{N} \tag{12}$$

FP2 calculates the summation of distance cost incurred for each ith node where i ranges from 1 to N (total number of nodes in the network). In Eq. (11), $D_{N-S(i)}$ represents the Euclidean distance of a ith node from the sink whereas, $D_{F(N-S)}$ represents the Euclidean Distance between farthest node and the sink. The average Euclidean distance of all the nodes from the sink is denoted by $D_{AVG(N-S)}$ and is given by Eq. (12). $D_{AVG(N-S)}$ is incorporated to favor the selection of CH nearby to the sink by fostering the CH selection of nodes with distance less than or equal to $D_{AVG(N-S)}$. It is observed that more the value of *FP2*, more will be the chances of a node to be selected as CH as it ensures the least possible communicating distance from the sink.

(c) Node density

The intra-cluster communication plays a significant role in efficiently utilizing the energy of nodes of cluster. The higher number of nodes in the vicinity of CH helps in preserving the energy that otherwise would have consumed gigantically. Therefore, a surrogate node is selected as CH that is surrounded by more number of nodes compared to any other node in the cluster, in other words, having higher node density. Therefore, Third Fitness Parameter (FP3) pacts with node density and is defined by the following Eq. (13).

$$\mathrm{FP3} = \left(\frac{\sum_{i=1}^{\mathrm{N_C}}\mathrm{D_{(N-NN)(i)}}}{\mathrm{N_C}} \times \frac{1}{\mathrm{D_{(N-FN)(i)}}}\right) \tag{13}$$

In above Eq. (13), the Euclidean average distance between ith node and the farthest neighboring node and is given by $D_{(N-FN)(i)}$. Also, $D_{(N-NN)(i)}$ denotes the distance between ith node and the neighboring node. It is to be noted that i ranges from 1 to N_C i.e. total number of nodes in a cluster. The whole term i.e. $\frac{\sum_{i=1}^{\mathrm{N_C}}\mathrm{D_{(N-NN)(i)}}}{\mathrm{N_C}}$ denotes the average distance of all nodes from their neighboring nodes. It is to be observed from the Eq. (13), D_{N-NN} should be smaller so as the selection of that node could be done which has least distance from the neighboring nodes. Consequently, the value of *FP3* must be maximized for efficient selection of CH that minimizes the distance factor and hence the energy consumption of nodes.

3.1.4. Fitness function for the network

In consideration to the above-discussed fitness parameters, the fitness function can be formulated by integrating all weighted fitness parameters in a single expression. The fitness function holds the control over the measure and performance of network design for the proposed work and is given by Eq. (14).

$$\mathrm{F} = \frac{1}{\varphi \times \mathrm{FP1} + \delta \times \mathrm{FP2} + \gamma \times \mathrm{FP3}} \tag{14}$$

In Eq. (14), φ, δ and γ are the weight coefficients where the attribute of residual energy is represented by φ, distance factor from a node to the sink by δ and the distance from a node to neighboring nodes is represented by γ. In the proposed work, these co-efficient are approximately evenly weighted (i.e., $\varphi = 0.3$, $\delta = 0.3$ and $\gamma = 0.35$) so as to emphasize the significant contribution of each factor. The proposed work finds its applicability for large area network that is the reason the value of $\gamma = 0.35$ is kept a bit higher than the values of other coefficients. Consequently, node density becomes a crucial concern in the proposed work. It is to be noted that the summation of all the coefficients is equal to unity as given in Eq. (15).

$$\varphi + \delta + \gamma = 1 \tag{15}$$

GAs aims to minimize F over the evolutionary processes to acquire the optimal network performance.

3.1.5. Selection

In order to enhance the average quality of a population at current generation, the selection process is carried out. In this work, the ranking selection along with the elitist strategy is adopted as compared to the roulette wheel selection due to the following reasons.

(a) Roulette-wheel selection considers the chromosomes by considering their cost values. However, in the case, if the manipulated cost is not too high; this type of selection becomes difficult to process as it is dependent on chromosome quality.

Table 4
A single-point crossover.

	Chromosomes	Bits representation
Before cross over	First	11001|01100000110
	Second	10010|**10110001110**
After cross over	Offspring	
	First	11001 | **10110001110**
	Second	10010 | 01100000110

Table 5
Example of mutation.

	Chromosomes	Bits representation
Before mutation	First	1100110**1**10001110
	Second	1001001100000110
After mutation	Offspring	
	First	1100110**0**10001110
	Second	1001001100000110

(b) Ranking selection assigns the rank to the chromosome based on their costs. Due to the selection based on the rank, this method has a higher probability of selecting the chromosomes better than the roulette-wheel selection.

Subsequently, the Elitist strategy is employed so as to continue with the best-selected parent chromosome for the next iteration. In sensor network topology, the aforementioned selection process of the chromosomes is taken into consideration in a following way. Based on the ranking; the eligible nodes to be selected as CH are listed and thereafter

Elitism is applied by keeping the CH same for the next round until it has a rich stock of energy. Consequently, numbers of overheads are decreased.

3.1.6. Crossover

It is a process in which the recombination of component material takes place due to the process of mating. The output from the crossover is directly dependent on the selection process performed on the chromosomes. Crossover operator is a binary genetic operator which performs on two parents. Though, there are various types of crossover which are applied according to the applications but in our work, we employ single point crossover, in it the chromosomes are exchanged after any 1-point as shown in Table 4.

During crossover operation, bit stream (from a point of crossover) of one of the parent chromosomes are exchanged with the other parent chromosome. After applying 1-point crossover, the two new bit stream generated correspond to two different chromosomes, termed as offspring as illustrated in Table 4.

In network topology, the crossover rate is basically analogous to the probability which decides whether there will be any periodical rotation of CHs or not. In the proposed strategy, the crossover rate i.e., $Pc = 0.6$ is employed.

3.1.7. Mutation

While the iteration is in process, after applying the crossover, the new generation of chromosomes will be bound to have the characteristics of the parental chromosomes. So the evolutionary optimization gets stuck among the local solutions. To resolve this issue, mutation is applied, which ensures the introduction of a new sequence of genes into a chromosome. Similar to the crossover, the mutation rate decides how often mutation will be applied. It is either per-bit or a per-chromosome basis, if the mutation rate is 0.0001 for per bit basis, it indicates that there is 0.01 percent chance for each bit in chromosome being mutated. Similarly, for 0.001 mutation rate for per bit chromosome ensures the 0.1% chance of each chromosome for its mutation.

In this paper, the mutation is applied on a per bit basis which affects a single chromosome as shown in Table 5. The eight bit of offspring 1 is mutated or changed from 0 to 1, however, the offspring 2 remains same due to the very low mutation rate.

In a network topology, the mutation process search for the best chromosomes by the transformation of CHs to cluster member and similarly from cluster members to the CHs. However, the likelihood of transformation from CHs to the individual is kept higher so as to put a control over the increase in the number of CHs. In the proposed strategy, the mutation rate i.e., Pm =0.006 is employed.

3.1.8. Termination

When the above-discussed GA operations are executed, the termination occurs after the predefined number of generations. During the iterations, the fitness value is compared with the one obtained from the previous iteration. So the chromosomes are updated accordingly.

Subsequently, the best chromosome or the best nodes are selected as CHs. It is reported that the maximum value of fitness function ensures the optimized CH selection with high residual energy and least communicating distance from the sink.

The whole process of GA applied in the proposed work is presented in algorithm 1 which is discussed as follows.

The description to the Algorithm 1 is given as follows.

Step 1: Initialization of GA parameters viz. number of chromosomes (N), *Population_size,* crossover rate (Pc), mutation rate (Pm), and generation number (gen) is done. In addition to these, weight coefficients used for remaining energy (φ), distance factor (δ), and node density (γ) are initialized.

Step 2: After performing validation and initialization, set of chromosomes are selected for further evaluation in GA operation.

Step 3: For selected chromosomes, the fitness parameters are computed which are then integrated in the fitness function given by Eq. (14).

Step 4: The process starts iteratively till the stopping criteria is reached.

Step 5: Count to the generation is incremented by 1 (gen = gen+1)

Step 6: The selection of fittest chromosome i.e., fit_chromo (*ka, kb)* is done by using rank selection method.

Step 7: 1-point crossover (C) is applied for the generation of new individuals i.e., $(\widetilde{ka}, \widetilde{kb})$ by combination of parents with pre-fixed crossover rate (Pc) i.e., $C(ka, kb|Pc) \Rightarrow \widetilde{ka}, \widetilde{kb}$.

Step 8: The new chromosomes generated after the cross over operation $(\widetilde{ka}, \widetilde{kb})$ is applied over the mutation operator with mutation rate Pm to generate $(\widetilde{ka}, \widetilde{kb})$, i.e., $M(ka|Pm) \Rightarrow \widetilde{ka}, M(kb|Pm) \Rightarrow \widetilde{kb}$

Step 9: Update the initial population by corresponding new generation and retain the fittest one through elitism strategy

Step 10: Repeat steps 3–10.

GA operates in steps explained in Algorithm 1. It is illustrated in the flow chart shown in Fig. 2. The heterogeneous sensor nodes are deployed randomly which is done with the set of selected chromosomes after validation process. Then fitness function is defined and the fittest chromosomes are selected which are further operated for cross over and mutation process. Set of chromosomes hence selected, correspond to the CHs selection. The process keep on iterating till the stopping criteria is reached. Thereafter, GAOC follows steady state phase and process is halted on complete energy exhaustion of network.

Algorithm 1 The Proposed Algorithm for CH Selection.

1. Initialize N, Pc, Pm, φ, δ,γ,and $gen=0$.
   // N is number of nodes or chromosomes, Pc is crossover rate,
   //Pm is mutation rate, gen is generation number, φ, δ and γ are the weight coefficient used for residual
   //energy, distance factor, and node density, respectively.
   //FP1,FP2, and FP3 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 *FP1, FP2 and FP3* according to eq. (10-13) for each ith chromosome
6. Fitness_Function

$$F(j) = \frac{1}{\varphi \times FP1 + \delta \times FP2 + \gamma \times FP3}$$

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
   $C(ka, kb|Pc) \Rightarrow \tilde{k}a, \tilde{k}b$.
   iv. Perform mutation on *ka* and *kb* with mutation rate*Pm*
   $M(ka|Pm) \Rightarrow \tilde{k}a$,
   $M(kb|Pm) \Rightarrow \tilde{k}b$
   v. chromosomes[j]= {$\tilde{k}a$, $\tilde{k}b$} *// update the initial population by corresponding new generation and retain the fittest one through elitism strategy*
   vi. repeat steps (5-7)
8. **end while**
9. **end for**

3.2. The working process of GAOC and MS-GAOC

The flow chart for the working process of GAOC and MS-GAOC is illustrated in Fig. 3. The protocols GAOC and MS-GAOC operates under two phases viz. setup phase and steady state phase.

3.2.1. Set up phase

The network formation and CH selection is processed in set up phase. It is discussed in following steps.

Step 1. The whole process is initiated from a random deployment of three energy level heterogeneous nodes which basically comprise of normal nodes, intermediate nodes, and advanced nodes. A single data sink is employed in the middle of the network in case of GAOC protocol whereas in case of MS-GAOC protocol, multiple data sinks equidistant to each other, are placed outside the network.

Step 2. After network formation, clustering is followed, in which selection of CH is done using GA. The parameters for CH selection used in the fitness function formulation includes, residual energy, distance factor, and node density.

3.2.2. Steady state phase

After selection of CH, the data transmission is initiated in steady state phase and the steps followed are discussed below.

Step 1. The moment CH is selected; it disseminates data from the cluster member nodes based on threshold approach. The nodes transmit data based on following conditions and checked for each cluster member node if current sensed value (C(V)) is more than pre-defined hard threshold (H(T)). Else, no transmission is preceded. However, in otherwise case, it is further checked if C(V) is greater than H(T) for any node, then further a test value is computed by subtracting new sensed value S(V) to the previously sensed value (C(V)). Otherwise, if test value is found to be greater than pre-defined soft threshold (S(T)) value, then data is transmitted to the corresponding CH else transmission is awaited till it satisfies the threshold conditions.

Step 2. When data is received by the CH, it aggregates the data and forwards it to the corresponding sink in case of GAOC protocol whereas, in case of MS-GAOC, it determines nearest located sink by computing minimum Euclidean distance between nodes and multiple data sinks.

Step 3. Energy of each node is checked if it is equal to or less than zero, if it happens so, node is said to be dead node and is represented by *D_N(i)* where value of *i* ranges from 1 to *N* (total number of nodes).

Step 4. Furthermore, it is checked for *i*th node if it is less than total number of nodes, if it is so, *D_N* is incremented by 1 and again CH selection is preceded for the next node. However, if all nodes are dead, the network stops functioning.

3.3. System energy model and network framework assumptions of GAOC and MS-GAOC

The structure of a number of heterogeneous nodes and their respective energy compositions, are discussed as follows. In the proposed protocols, the total number of advanced, intermediate nodes and normal nodes represented by N_{ADV}, N_{INT} and N_{NRM} are given by a set of Eqs. (1)–(3). In these equations, the proportions of a number of intermediate and advanced nodes are represented by m and m_o, respectively. These set of equations are used in building up network models for both protocols as discussed in further sections.

3.3.1. Analysis of energy model for GAOC and MS-GAOC

In this paper, the fundamental radio energy consumption model is used as shown in Fig. 4. It decides for the amount of depletion of energy for all the nodes that communicate in the

network. The energy consumed while the network is in operation is demonstrated by the set of Eqs. (16)–(20). The computational energy is considered to be negligible. So, while formulating energy model, the energy depleted during communication is taken into consideration. The Eqs. (16)–(17) describe the consumption of energy while transmission at a different value of distance. The energy consumed in the transmission of *l*-bit data at the distance '*d*', is represented by E_{tx} *(l,d)* and given as follows.

$$E_{tx}(l, d) = l \times E_{elc} + l \times E_{efs} \times d^2 \text{ for } d \le d_o \tag{16}$$

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

In above Eqs. (16)–(17), '*d*' represents the distance between the source and destination nodes or between nodes and sink. E_{elc} represents the energy consumed for switching on and off the transmitter and receiver circuitry. The threshold distance is represented by 'd_o' and is expressed as in Eq. (18).

$$d_o = \sqrt{\frac{E_{efs}}{E_{amp}}} \tag{18}$$

The characteristics of transmitter amplifiers are given by E_{efs} and E_{amp} where E_{efs} is for free space energy model (power loss d^2) and E_{amp} accounts to energy consumed for multi path energy model (power loss d^4).

The energy consumed while receiving the data per bit is given by Eq. (19).

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

The CH which performs the data aggregation consumes energy as given by the Eq. (20).

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

The energy consumed in the reception of *l*-bit data, is represented by E_{rx}. E_{da} is the energy consumed in the data aggregation of 1-bit data. Moreover, $E_{dx}(l)$ is the energy expenditure during data aggregation of received *l*-bit data of *p* number of data packets. The energy consumption by the network follows the equations discussed in radio energy mode.

3.3.2. The network assumptions for framework of GAOC

While developing GAOC, the assumptions are made about the characteristics for some of the sensor nodes listed as following.

(a) The network is static in the context that all the nodes including sink, are stationary in the monitoring area. They are capable enough for data forwarding and data receiving from the other nodes and sink, within their sensing range.
(b) The deployed nodes are heterogeneous in energy parameter i.e. some of the nodes are enriched with additional energy resources. These nodes are intermediate nodes and advanced nodes. So, three types of nodes normal; intermediate and advanced nodes with three different levels of energy are taken into consideration.
(c) Nodes possess some unique identifiers (ID) which are allocated to them.
(d) The sink is considered to be having an infinite source of energy.
(e) Once the nodes are exhausted or depleted of their energies, they cannot be recharged again.
(f) These sensor nodes are not equipped with any GPS-enable antenna on their circuitry i.e., these nodes are location unaware.
(g) The factors viz. signal fading, interference and signal loss resulted due to various environmental factors and physical obstacles are not taken into consideration.
(h) The monitoring/target area taken for the consideration is assumed to be a square area given by $A = M \times M$ in m^2(meter square).
(i) Whenever the sensor nodes are deployed randomly in network field, sink broadcasts *HELLO* message which is acknowledged by each node and *HELLO_neighbor* message is spreaded by each node in its neighborhood, i.e., up to its communication range. With the help of these messages, the distance between the sink and neighboring nodes is computed by analyzing the RSSI (Received Signal Strength Indicator) value.

3.3.3. The network assumptions for framework of MS-GAOC

The system framework of MS-GAOC has some network assumptions and operational characteristics that are similar and different in the following way.

(a) During set up phase of clustering in MS-GAOC, four data sinks are employed outside the network around each periphery unlike the single sink in case of GAOC as illustrated in Fig. 5. All sinks are identical and placed equidistant to each other and they are assumed to have no constraints on the energy resources, computational and coverage capabilities.
(b) Rest of the network assumptions taken into consideration in MS-GAOC stays same as that of GAOC. The same number of nodes with same energy stock, are deployed in a network area of similar dimensions as that of GAOC. The radio energy consumption model stays same in MS-GAOC as used in GAOC.
(c) The operational behavior of MS-GAOC is same with GAOC in a sense that the CH selection in case of MS-GAOC follows the same GA operation as adopted by GAOC. The fitness function also stays same in the case of MS-GAOC as used in GAOC.
(d) However, the operation of MS-GAOC differs in a way the data transmission is carried out from the CHs to the sink. The nearest sink is selected by the CH for the data transmission as illustrated in Fig. 6.

In nutshell, it can be said that MS-GAOC is an extended version of GAOC, explored with multiple data sinks placed outside the network. The rationale behind extending the work from GAOC to MS-GAOC i.e. essence of employing multiple sinks and is comprehensively discussed in the later stage of Section 2.

After the extensive investigation of the proposed methodology, the following Lemmas are defined that highlights the significance of proposed protocols.

Lemma 1. *GAOC protocol is distributed and scalable in nature and CH selection is optimized through rigorous genetic process.*

Proof. The notable characteristics of GAOC are discussed as follows.

(a) The fitness parameters considered for the formulation of fitness function includes residual energy of the node (which is updated after each round, among the nodes), distance factor that is determined through the Euclidean distance computations. This distance factor is basically based on the Received Signal Strength Indicator (RSSI) value of a received signal. Furthermore, the node density parameter is taken into consideration that ensures the least communicating distance for the cluster nodes to the CH. These factors are computed through the distributed approach by exchanging the messages among nodes and sink.

(b) Due to its distributed approach, the CH can be extended to the thousands of nodes and area dimensions can also be enhanced significantly with no compromise to the network performance.
(c) The aforementioned parameters used in fitness function of GAOC are iterated through different steps of GA which brings out the optimal selection of CH as explained in previous sections.

Lemma 2. *MS-GAOC protocol is distributed, scalable, mitigates Hot-Spot problem and best suited for unattended/ hostile applications.*

Proof. The some of the prominent characteristics of MS-GAOC are discussed below.

(a) The selection of CH takes place through the optimized process of GA that runs on the grounds of different fitness parameters. The set up phase and steady state phase of GAOC are operated by exchanging the *declaration* and *connection_establishment* messages.
(b) Due to employment of multiple data sinks, the effective communicating distance between nodes and sink is reduced comprehensively. So the network area can be extended to a level up to which the nodes are under communication range to each other.
(c) For those applications, where sink placement has to be done outside the network, MS-GAOC has a huge role to play. Inevitable placement of single data sink outside the network, will lead to desideratum of multi hop communication. Therefore, the approach of multiple data sinks i.e., given by MS-GAOC, will be mitigating the Hot-Spot problem by avoiding dual hop communication and employing single hop. The network area can be extended to enhanced dimensions without compromising its performance.

Lemma 3. *The protocols GAOC and MS-GAOC perform while maintaining Quality of Service (QoS) parameters viz. throughput and delay to improve network performance*

Proof. The supporting reasons to verify the performance of proposed protocols on the benchmark of QoS are discussed as follows.

(a) As a result of multiple data sinks employment, the throughput is successfully enhanced due to reduced communicating distance between nodes and sink. Moreover, due to the congestion avoidance around the sink, the data packets are successfully transmitted to the respective sink.
(b) When employed for the large area network, the multi hop communication employed at larger distance between nodes or between nodes and sink, will always end up in consuming more time units as compared to the single hop at shorter distance. Employing multiple data sinks around the network results in single hop communication that too at a shorter distance. Hence, the delay in reception of packets at the end of data sink will be shortened making it much suitable for hostile applications where the response time for rescuing operation against any calamities is very low.

The performance evaluation of GAOC and MS-GAOC are discussed in the following section.

4. Results and discussions

The performance comparison of GAOC is done with the state of the art protocols based on genetic algorithms having their distinct fitness functions based on different fitness parameters. These state-of-the-art protocols GADA-LEACH, and DCH-GA are selected on the basis of literature study carried out comprehensively as discussed in Table 3. Moreover, GAOC protocol is also compared with the optimized but reactive protocol i.e. TEDRP as it has outperformed various routing strategies due to its energy efficient CH selection as discussed in Table 1. To highlight the essence of the multiple data sinks, MS-GAOC is compared with GAOC protocol and further for the fair comparative analysis, MS-GAOC is also compared with the MS-DCHGA, MS-GADA and MS-TEDRP protocols (developed in this work). The performance metrics used for the performance comparison of proposed protocols with the aforementioned protocols are discussed as follows.

4.1. Performance metrics

There are some crucial performance metrics which are utilized to inspect the performance of GAOC and MS-GAOC against the other protocols. These are discussed as follows.

(a) **Stability Period**
The network stability depends upon this factor as it ensures the reliable data dissemination from the network. The number of rounds covered till the first node of any type i.e. normal, intermediate and advanced nodes, depletes its full stock of energy, is termed as stability period. In some applications, where the loss of even little information can have a high magnitude of repercussions on the network performance, the stability period becomes a prominent performance metric to be considered for performance evaluation. The higher stability period ensures the higher reliability of proposed routing protocols operating for any network.
(b) **Network Longevity**
The large area applications, for which the data dissemination is a result of continuous monitoring, network longevity has a crucial role to play. Network longevity can be defined as the number of rounds covered till all nodes are exhausted of energy while pursuing data transmission.
(c) **Number of dead nodes against rounds**
The network performance is also evaluated on the factor which gives the status of a number of dead nodes with the passage of rounds. When a number of rounds are covered against the number of alive nodes, the network performance is said to be upgraded.
(d) **Throughput**
The number of data packets successfully transmitted to sink is termed as throughput. This is one of the recurring parameter for fostering the QoS so as to ensure the reliability of the network. The sole consideration of network longevity does not help in acquiring the best of network performance. Therefore, QoS parameter upgrades the network performance and enhances the credibility of proposed routing strategy.
(e) **Network's remaining energy**
As the data transmission proceeds, the network's total energy i.e. sum of energy of all nodes reduces gradually due to the energy consumption by the nodes while communicating with the other nodes or with the sink. This metric helps in exposing the status of total energy of nodes after each round. When the routing strategy is an efficient one, the network preserves its energy and covers more number of rounds thereby resulting in acquiring network longevity.

Table 6
Simulation parameters.

Network model and GA parameters	Values
Network Area Size	100 ×100 m^2, 500 × 500 m^2
Number of Nodes (N)	100, 200
Number of data sinks for GAOC and MS-GAOC	1 and 4
Initial energy of nodes (in Joules) (E_o)	0.5
Energy heterogeneity Node Type	3-level; normal, intermediate and advanced nodes
Energy fraction of intermediate nodes (β) and advanced nodes(α)	$\beta = 1$, $\alpha = 2$
Number of intermediate nodes (m) and advanced nodes fraction (m_o)	$m = 0.1$, $m_o = 0.2$
Energy required for running transmitter and receiver E_{elc}	50 nJ/bit
Threshold distance (d_o)	87 m
Amplification energy required for smaller distance $d \leq d_o$ (E_{efs})	10 pJ/bit/m^2
Amplification energy required for larger distance $d > d_o$ (E_{mp})	0.0013 pJ/bit/m^4
Energy consumption incurred while data aggregation (E_{da})	5 nJ/bit/signal
Data packet size	2000 bits
Population Size (P)	30
Crossover rate (P_c)	0.6
Mutation rate (P_m)	0.006
Type of crossover	Single Point
Selection method	Rank selection method
Number of total chromosomes	30
Number of Generations	30
Number of Simulation run	20
Confidence interval	95%

4.2. Simulation settings

This section presents the simulation environment utilized for simulating the GAOC and MS-GAOC protocols. The simulation software MATLAB version 2016 running on Windows 10, with an Intel Core i3 CPU 540 with processor operating at 3.07 GHz and RAM of size 4 GB, is employed for simulation purpose. The network area of 100 ×100 m^2 is simulated and 100 energy heterogeneous nodes are randomly deployed in it. The sink is placed in the middle of the network in the case of GAOC whereas four sinks outside the network are placed in the case of MS-GAOC. These four sinks are placed at (50, -10), (-110, 50), (50, 110) and (-10, 50) coordinates of the network area .

The simulation parameters considered for GAOC and MS-GAOC are listed in Table 6. The initial energy of normal nodes is 0.5 Joules and the intermediate nodes are two fold and advanced nodes are three fold to the energy of normal nodes. The number of intermediate nodes and advanced nodes have number fractions i.e. $m = 0.1$, $m_o = 0.2$ and energy fractions i.e. $\alpha = 2$ and $\beta = 1$, respectively.

The radio energy parameters are kept same as that used in other protocols taken for comparative analysis. The set of 30 chromosomes are used for GA operation. The single point cross over with cross over rate ($P_c = 0.6$) is used. The mutation rate is (P_m =0.006).

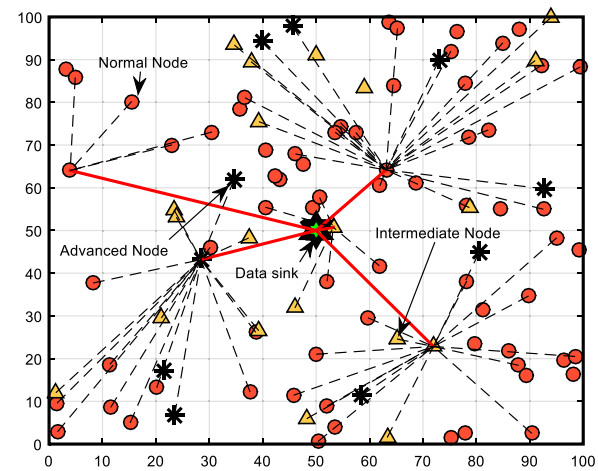

Fig. 5. Data transmission scenario of GAOC.

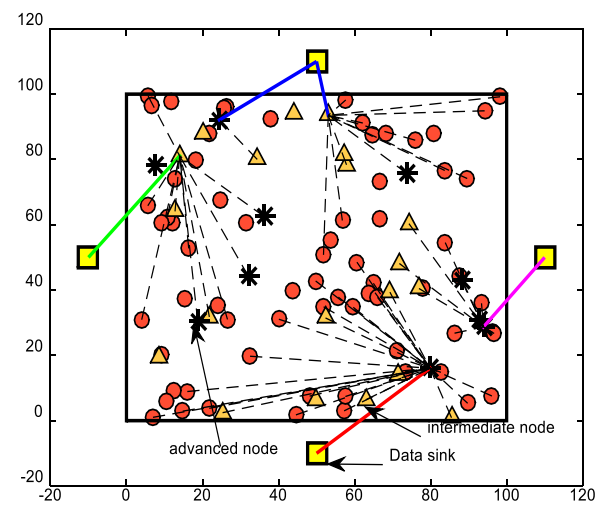

Fig. 6. Data transmission scenario of MS-GAOC.

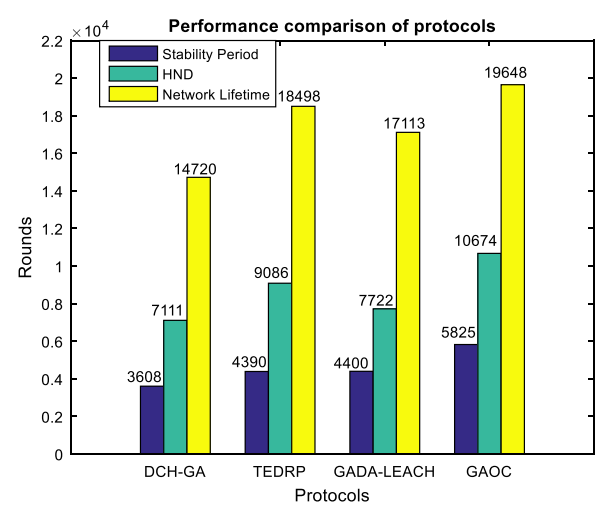

Fig. 7. Stability Period, HND and Network lifetime comparison of GAOC with other protocols.

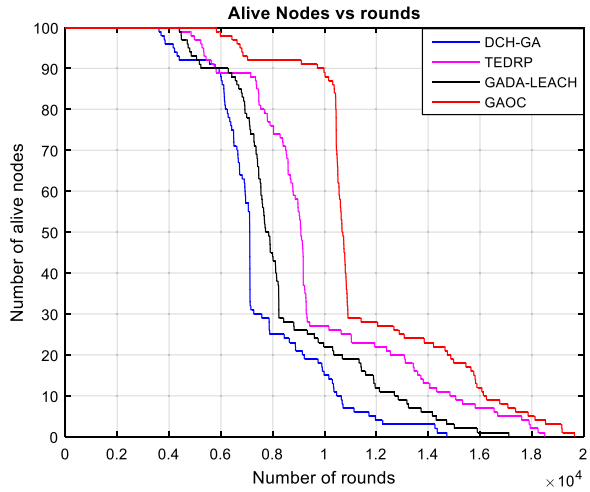

Fig. 8. Comparison of alive nodes vs rounds of GAOC with other protocols.

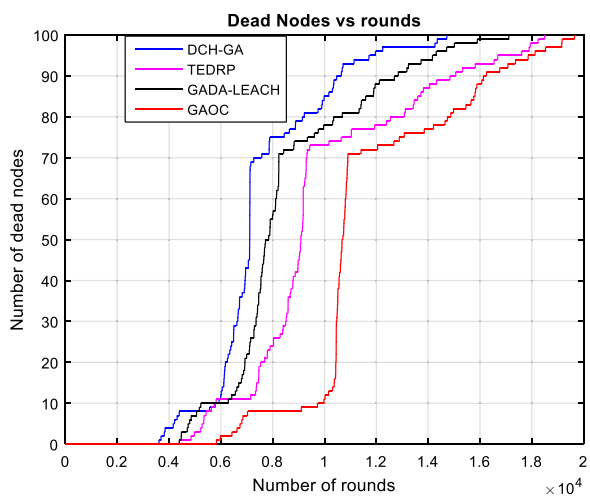

Fig. 9. Comparison of dead nodes vs rounds of GAOC with other protocols.

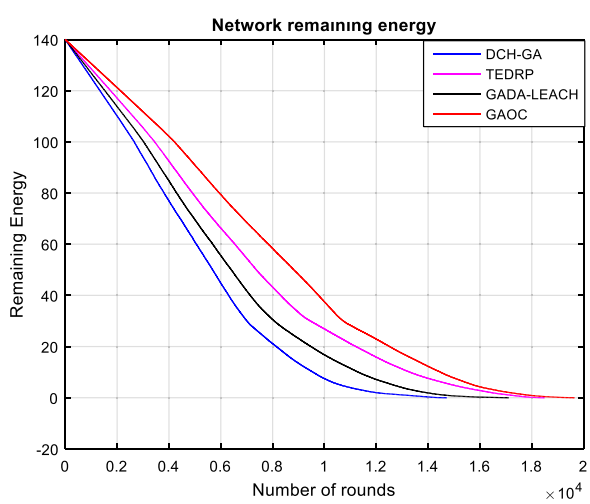

Fig. 10. Comparison of Network's remaining energy of GAOC with other protocols.

It is observed that larger mutation rate may results in losing some good solutions and the small cross over rate may result in hinging the convergence. Increase in the number of generations may lead to better solutions but it enhances the time required to obtain the solution which affects its efficiency and hence, it might not be acceptable for some applications. The rank selection method is considered for selection of fittest chromosomes.

The Confidence Interval for the simulation run is 95%. The data transmission scenario for GAOC and MS-GAOC are illustrated in Figs. 5 and 6, respectively. In GAOC sink is placed at the middle whereas in MS-GAOC four sinks are placed.

4.3. Simulation analysis

The simulation analysis of GAOC and MS-GAOC, is done on the grounds of performance metrics discussed in Section 4.1. The working process of GAOC is based on the heterogeneous network framework discussed in Section 3.1 and is illustrated in the flowchart shown in Fig. 3. Firstly, the simulation results of GAOC are discussed thereafter, the simulation analysis of MS-GAOC is discussed against the multiple data sinks form of other protocols.

4.3.1. Simulation results of GAOC protocol

It is to be noted that simulation analysis follows the performance evaluation of GAOC follows the performance metrics defined in Section 4.1. The comparative analysis of the performance metrics for GAOC against the other protocols is discussed as follows.

(a) **Stability Period:**
In GAOC, the first node depletes its energy after 5825 rounds which was 4400, 4390 and 3608 rounds in the case of GADA-LEACH, TEDRP and DCH-GA, respectively as illustrated in Fig. 7. It is evident that GAOC contributes to 32.39%, 32.69% and 61.45% gigantic amelioration in stability period in comparison to the protocols GADA-LEACH, TEDRP and DCH-GA, respectively.
The prominent reason behind such amelioration in stability period is the inclusion of energy efficient fitness parameters formulating fitness function. Incorporation of distance and energy factor for the selection of CH reduces the uneven and abrupt energy consumption and further helps in energy preservation. In addition to that, node density factor ensures the minimal intra-cluster distance among the nodes and sink.
Unlike to the protocols GADA-LEACH, TEDRP and DCH-GA that are taken into consideration for comparative analysis, GAOC ensures optimized energy efficient CH selection that enhances stability period.

(b) **Network Lifetime**
It is observed from Figs. 7 and 8 that GAOC covers 19648 rounds whereas GADA-LEACH, TEDRP and DCH-GA covers 17113, 18498 and 14720 rounds, respectively, before complete energy exhaustion of all the nodes in the network. GAOC accounts 2535 rounds more as compared to the GADA-LEACH protocol hence contributing in 14.81% enhancement in the network lifetime. Moreover, GAOC ameliorates network lifetime by 6.22 and 33.48% in comparison to TEDRP and DCH-GA protocols, respectively.
Such amelioration is reported due to the energy efficient fitness parameters taken into consideration for the selection of CH. The node density factor reduces the communication cost for the sensor nodes in a cluster. It is due to the reason that node density factor favors the CH selection of a node which is surrounded by more neighboring nodes. Consequently, overall network energy is preserved leading to improved network lifetime.

(c) **Number of dead nodes against the rounds**
The performance comparison of GAOC is illustrated in Fig. 9 which investigates the different percentage of dead nodes of the network with respect to the number of rounds. It is evident from Fig. 9 that First Node Dead (FND) for GAOC is 5825 rounds which is 4400, 4390 and 3608 rounds in case of GADA-LEACH, TEDRP and DCH-GA, respectively. Half Nodes Dead (HND) for GAOC is 10674 whereas it is just 7722, 9086 and 7111 rounds in case of GADA-LEACH, TEDRP and DCH-GA protocols, respectively. The improvement in Last node dead (LND) is also observed in GAOC as it covers 19648 rounds whereas GADA-LEACH, TEDRP and DCH-GA covers 17113, 18498 and 14720 rounds, respectively. It can be contemplated from the above analysis that GAOC covers more number of rounds at different stages of dead nodes. It is ascribed to the reduced energy expenditure incurred by the optimized CH selected under GA operation.

(d) **Network's remaining energy:**
The rate of network's energy expenditure is measured by this metric. When the data transmission proceeds, the network's energy reduces. GAOC performs better as compared to GADA-LEACH, TEDRP and DCH-GA protocols, respectively in a way that it covers a greater number of rounds while the data transmission is in progress as illustrated in Fig. 10. GAOC covers maximum number of rounds but DCH-GA covers least number of rounds during the network operation. It is due to the reason that energy efficient selection of CH consumes a very little stock of energy among all nodes. Moreover, the intra cluster communication also preserves energy of nodes in the most efficient way.

(e) **Throughput/Number of data packets sent to sink:**
In case of GAOC, as illustrated in Fig. 11, the throughput is enhanced comprehensively as it successfully transmits 698600 data packets whereas GADA-LEACH, TEDRP and DCH-GA transmit 506203, 567684 and 443842 data packets, respectively. It is evident from the throughput comparative analysis, GAOC improves throughput by 57.4%, 23% and 38% as compared to GADA-LEACH, TEDRP and DCH-GA protocols, respectively. Such improvement is observed due to optimized CH selection in the network which further helps in acquiring network longevity. Consequently, nodes transmit data packets for a longer duration thereby enhancing throughput significantly.
The following subsection covers the simulation analysis of MS-GAOC protocol.

4.3.2. Simulation analysis of ms-gaoc protocol

In this Section, for the fair comparative analysis of MS-GAOC with the state of the art protocols GADA-LEACH, TEDRP and DCH-GA, the multiple data sink based version of aforementioned protocols is generated. The simulation settings to design MS-DCHGA, MS-GADA and MS-TEDRP protocols are kept same as that of used in MS-GAOC. Only difference in the designing of multiple sinks based version protocols with the single sink based version (GAOC, GADA-LEACH, TEDRP and DCH-GA), is the employment of four sinks outside the network with the same network dimensions. Performance investigation of MS-GAOC is carried out against the other protocols by considering the previously explored performance metrics (discussed in Section 4.1).

(a) **Stability Period**
It is illustrated from Fig. 12 that MS-GAOC has a stability period of 6515 rounds whereas MS-GADA, MS-TEDRP and MS-DCH has a stability period of 5334, 4791, 3624 rounds, respectively. It is observed that MS-GAOC contributes to 22.14, 35.98 and 79.77% gigantic amelioration in stability period in comparison to the protocols MS-GADA, MS-TEDRP and MS-DCHGA, respectively. The comparison

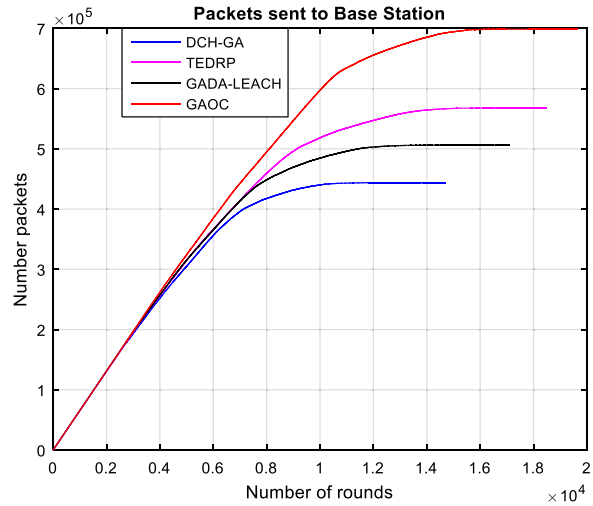

Fig. 11. Comparison of throughput of GAOC with other protocols.

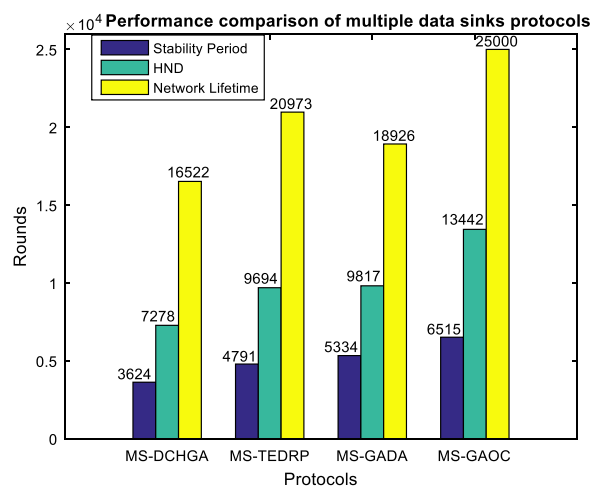

Fig. 12. Stability Period, HND and Network lifetime comparison of MS-GAOC with other protocols.

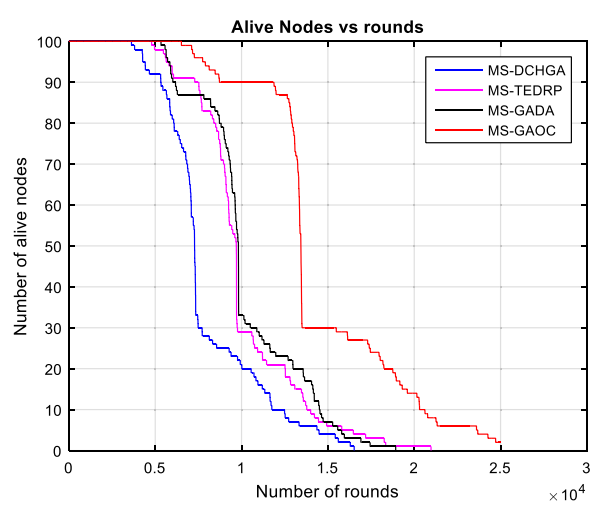

Fig. 13. Comparison of alive nodes vs rounds of MS-GAOC with other protocols.

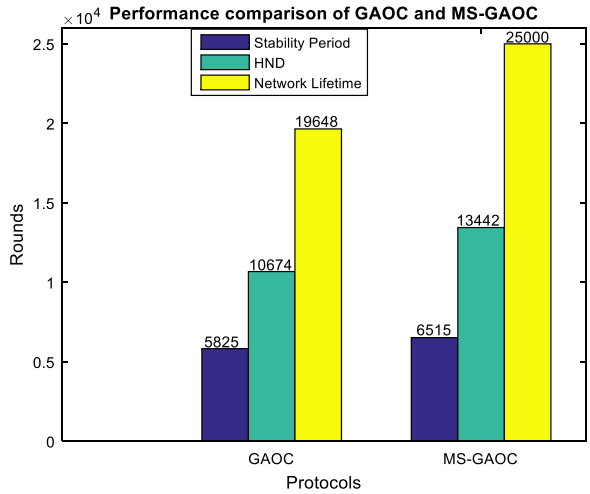

Fig. 14. Stability Period, HND and Network lifetime comparison of MS-GAOC with GAOC.

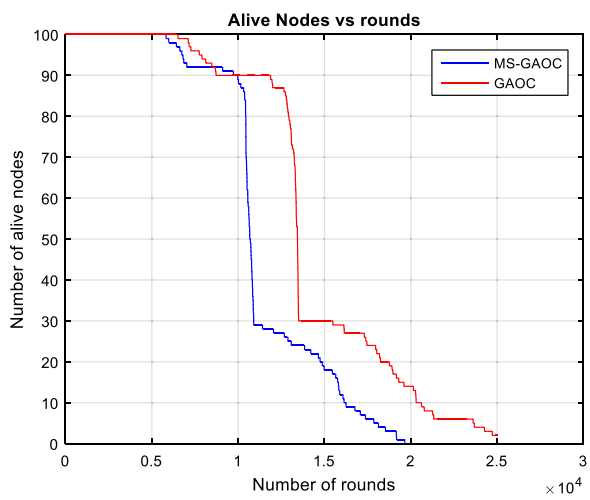

Fig. 15. Comparison of alive nodes vs rounds of MS-GAOC with GAOC.

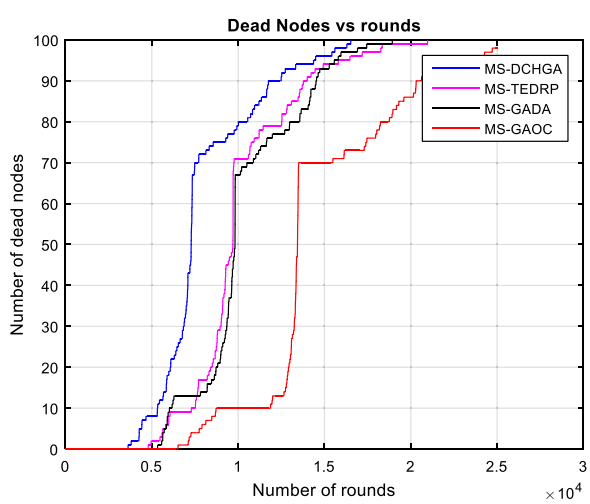

Fig. 16. Comparison of dead nodes vs rounds of MS-GAOC with other protocols.

of the number of alive nodes over the rounds is done for MS-GAOC against MS-GADA, MS-TEDRP, and MS-DCHGA

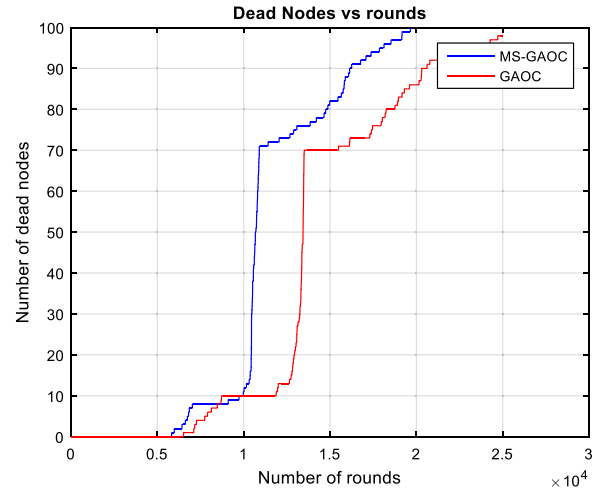

Fig. 17. Comparison of dead nodes vs rounds of MS-GAOC with GAOC.

protocols as shown in Fig. 13. It is observed that in MS-GAOC more number of rounds are covered corresponding to the different number of alive nodes during the network run. This improvement is observed due to the energy efficient CH selection incorporating node density factor along with energy and distance. It is due to the fact that the node density factor helps in shortening the communicating distance between nodes and corresponding CH in a cluster. Moreover, as illustrated in Fig. 14, MS-GAOC improves stability period of GAOC by 11.85% which is accounted to the employment of multiple data sinks due to which the geometric distance between the sink and CH is reduced comprehensively. The comparison of the number of alive nodes over the rounds is also done for MS-GAOC against GAOC as shown in Fig. 15, MS-GAOC covers a greater number of rounds for different number of alive nodes left in the network.

(b) **Network lifetime**

It is reported that network lifetime of MS-GAOC is improved by 32.09, 19.2 and 51.231% as compared to the MS-GADA, MS-TEDRP and MS-DCH protocols, respectively. The total number of rounds covered by MS-GAOC is 25000 till the complete energy exhaustion of all nodes. Whereas, it is 18926, 20973 and 16522 rounds in case of MS-GADA, MS-TEDRP and MS-DCHGA protocols, respectively as shown in Fig. 12. The enhancement in network lifetime is attributed to the fact that genetic operation selects the fittest chromosomes that are analogous to the selected CHs in the clusters. The inclusion of energy efficient fitness parameters makes GAOC outperforming the other protocols. Furthermore, the comparative analysis of network lifetime of MS-GAOC is done with GAOC. It is evident from Fig. 15 that MS-GAOC covers 25000 rounds against the 19648 rounds of GAOC. The reason behind amelioration in the network lifetime is the reduction in the energy consumption due to nearly located data sinks. Due to the single hop and reduced distance communication, Hot-Spot problem is mitigated that helps in the energy balancing in the network. Due to this energy balancing, the network acquires network longevity.

(c) **Number of dead nodes against rounds**

The performance comparison of MS-GAOC is illustrated in Fig. 16 in which MS-GAOC covers more number of rounds at different stages of dead nodes. It is evident from Fig. 12 that First Node Dead (FND) for MS-GAOC is 6515 rounds

which is 5334, 4791 and 3624 rounds in case of MS-GADA, MS-TEDRP and MS-DCHGA protocols, respectively. Half Nodes Dead (HND) for MS-GAOC is 13442 whereas it is just 9817, 9694 and 7278 rounds in case of MS-GADA, MS-TEDRP and MS-DCHGA protocols, respectively.

The improvement in Last node dead (LND) is also observed in MS-GAOC as it covers 25000 rounds whereas MS-GADA, MS-TEDRP and MS-DCHGA covers 18926, 20973 and 16522 rounds, respectively. It can be contemplated from Fig. 16 and above analysis that MS-GAOC covers more number of rounds at different stages of dead nodes which is due to the reduced energy expenditure as a result of energy preservation in the CH selection as well as in the intra-cluster communication.

Furthermore, the comparative analysis of dead nodes vs rounds is done for MS-GAOC and GAOC as shown in Fig. 17. It is found that MS-GAOC improves the network lifetime by the gigantic improvement of 27.24%. It is again due to the reduced energy expenditure incurred due to employment of multiple data sinks outside the network.

(d) **Network's remaining energy:**
As the network operates in MS-GAOC, it covers more number of rounds as discussed above. As illustrated in Fig. 18, the network's energy of MS-GAOC covers more number of rounds as compared to MS-GADA, MS-TEDRP and, MS-DCHGA protocols, respectively. It is due to the reason that the energy of nodes is preserved due to the optimized selection of CH which helps in energy preservation. It is basically due to the node density factor considered for CH selection that reduces the communicative distance among the CH and cluster member nodes. Furthermore, in a Fig. 19 the comparative analysis of MS-GAOC is done against GAOC. MS-GAOC perform better than GAOC due to the incorporation of multiple data sinks which helps in abating the network's remaining energy at a very low rate as compared to the networks of other protocols. With multiple data sinks employment, the residual energy is saved by the much larger amount due to the reduction in the number of communication hops. Consequently, it leads to the reduction in the effective communication distance from the CH to the respective sink.

(e) **Throughput/Number of data packets sent to sink:**
The throughput in MS-GAOC improves by high magnitude i.e., by 36.58, 43.46 and 92% as compared to MS-GADA, MS-TEDRP and MS-DCHGA protocols, respectively. The total number of packets transmitted to the respective sink in case of MS-GAOC is 867047 whereas it is 634813, 604364, 451133 packets in case of MS-GADA, MS-TEDRP and MS-DCHGA protocols, respectively as illustrated in Fig. 20. The improvement in throughput is obtained due to the network longevity achieved by MS-GAOC which helps in dissemination and hence transmission of more number of packets in the network. The fitness parameters used in the selection of CH play a major role in enhancing throughput. In addition to this, MS-GAOC also outperform GAOC in throughput metric. It is observed that MS-GAOC transmits 168447 packets more than that sent by GAOC i.e. improving throughput by 24.11% as illustrated in Fig. 21. It is due to the reduction in the average communicating distance of CH nodes from the sink that helps in fostering the data delivery at the sink.

4.4. Summarized analysis of GAOC and MS-GAOC protocols

It can be encapsulated from the results obtained by simulation analysis of GAOC and MS-GAOC that by incorporating multiple data sinks in the network; it not only enhances the reliability but also stability period in the network. The comparative analysis of proposed protocol is reported in Table 7. Furthermore, Tables 8 and 9 demonstrate the percentage improvement by GAOC and MS-GAOC to the protocols (DCH-GA, TEDRP and GADA-LEACH) and GAOC respectively.

It is to be noted that the performance improvement for the MSO-GAOC is observed due to the proposed CH selection not solely due to the multiple data sinks. The enhanced stability period achieved by the MSO-GAOC as compared to MS-GADA, MS-TEDRP and MS-DCHGA as discussed in Table 9 supports the aforementioned reason for its performance improvement. Load balancing is also achieved with MS-GAOC as compared to the other protocols in the single sink as well as in multiple data sinks scenarios thereby making it favorable to explore for some apposite time critical unattended applications.

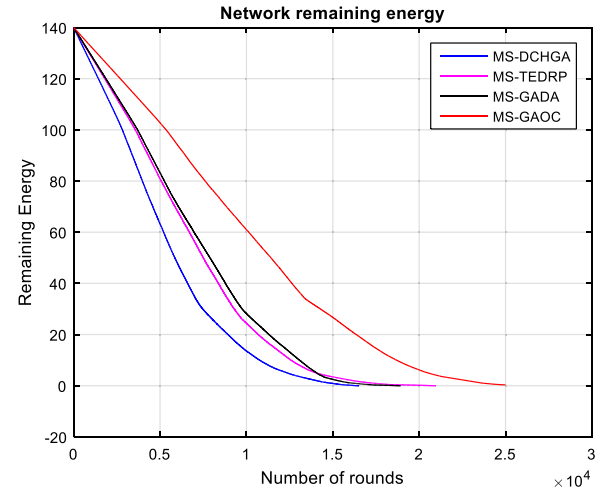

Fig. 18. Comparison of Network's remaining energy of MS-GAOC with other protocols.

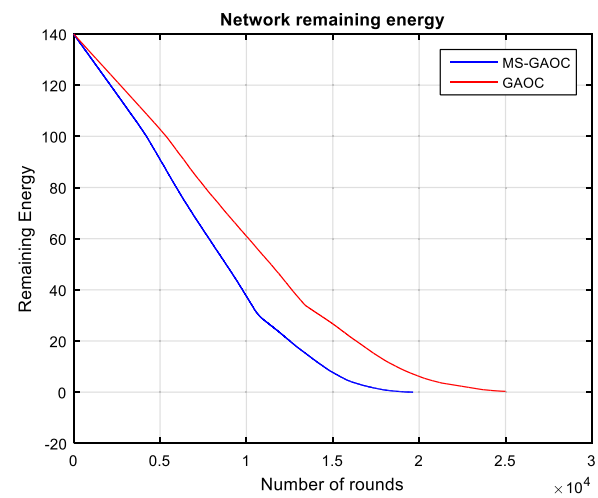

Fig. 19. Comparison of Network's remaining energy of MS-GAOC with GAOC.

5. Conclusion

The optimized selection of Cluster Head (CH) is of paramount concerns for acquiring energy efficiency in WSN. To acquire that, two novel reactive routing protocols, i.e., Genetic Algorithm-based Optimized Clustering protocol (GAOC) and Multiple data

Table 7
Comparative analysis of GAOC and MS-GAOC with other protocols for different metrics.

Value of advanced fractions and quantity fractions of node $m = 0.1$, $m_o = 0.2$, $\beta = 1$, $\alpha = 2$						
Protocols	No. of data sinks	Total energy of network (Joules)	Stability period (rounds)	Half node dead (rounds)	Network lifetime (rounds)	Throughput (packets)
DCH-GA	1	140	3608	7111	14720	443842
TEDRP	1	140	4390	9086	18498	567684
GADA-LEACH	1	140	4400	7722	17113	506203
GAOC	**1**	**140**	**5825**	**10674**	**19648**	**698600**
MS-DCHGA	4	140	3624	7278	16522	451133
MS-TEDRP	4	140	4791	9694	20973	604364
MS-GADA	4	140	5334	9817	18926	634813
MS-GAOC	**4**	**140**	**6515**	**13442**	**25000**	**867047**

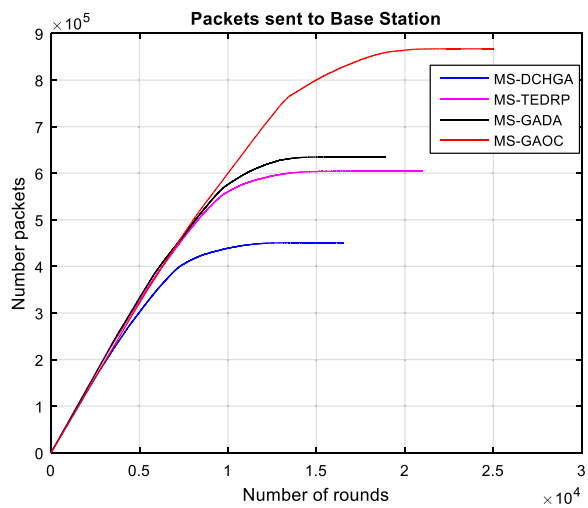

Fig. 20. Comparison of throughput of MS-GAOC with other protocols.

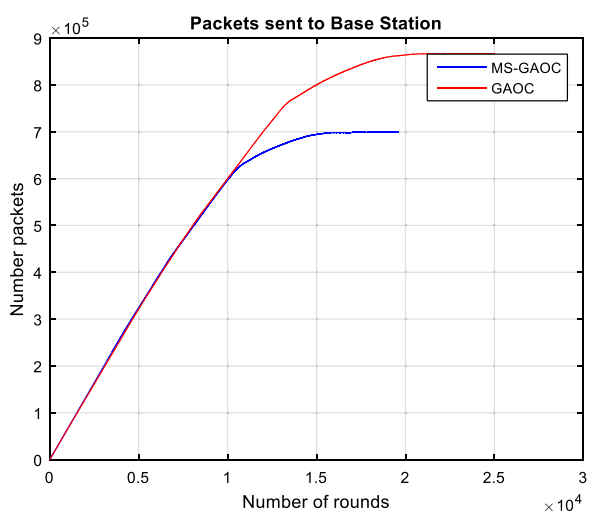

Fig. 21. Comparison of throughput of MS-GAOC with GAOC.

Sink based GAOC (MS-GAOC) for Heterogeneous WSN are proposed. The parameters of residual energy, distance to the sink and node density are employed in formulating the fitness function for both protocols. The node density factor has contributed to energy preservation for the intra-cluster communication. The extensive description of chromosome representation and framing of fitness function is discussed. In MS-GAOC, the employment of multiple

Table 8
Comparative percentage improvement by GAOC to other protocols.

Percentage (%) Improvement by GAOC protocol				
Protocols	Stability period	Half node dead	Network lifetime	Throughput
DCH-GA	61.45	50.11	33.48	57.4
TEDRP	32.69	17.48	6.22	23
GADA- LEACH	32.39	38.23	14.81	38

Table 9
Comparative percentage improvement by MS-GAOC to other protocols.

Percentage (%) Improvement by MS-GAOC protocol				
Protocols	Stability period	Half node dead	Network lifetime	Throughput
GAOC	11.85	25.93	27.24	24.11
MS-DCHGA	79.77	84.69	51.31	92
MS-TEDRP	35.98	38.66	19.2	43.46
MS-GADA	22.14	36.93	32.09	36.58

data sinks has helped in mitigating the hot-spot problem, i.e., premature death of network lifetime in a large network area where multi-hop communication is inevitable.

The simulation and hence comparative analysis of GAOC and MS-GAOC with GADA-LEACH, TEDRP and DCH-GA protocols has been performed on benchmark of different performance metrics. For fair comparative analysis with MS-GAOC, multiple data sinks based version viz. MS-GADA, MS-TEDRP and MS-DCHGA are developed.

The extensive simulations showed that GAOC has improved stability period by 32.39, 32.69, and 61.45% and network lifetime by 14.81, 6.22, and 33.48% as compared to the GADA-LEACH, TEDRP and DCH-GA protocols, respectively. Such amelioration is acquired due to the energy efficient fitness parameters taken into consideration.

Furthermore, it is observed that MS-GAOC has ameliorated stability period by 11.85, 22.14, 35.98 and 79.77% and network lifetime by 27.24, 32.09, 19.2, and 51.34 as compared to the GAOC, MS-GADA, MS-TEDRP and MS-DCHGA protocols, respectively. Throughput is also improved substantially by MS-GAOC as compared to aforementioned protocols. Such amelioration is obtained not only due to the shortening of the effective communicating distance between nodes and the respective sink that but also due to energy efficient CH selection.

GAOC finds its utility in attended and small area based network whereas MS-GAOC can be employed for hostile reactive large area applications viz. forest fire detection, early detection of volcanic eruptions, etc. In future, this work can be extended for moving sink scenario for achieving better performance.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a survey, Comput. Netw. 38 (2002) 393–422.
[2] G.J. Pottie, Wireless sensor networks, in: Inf. Theory Workshop, IEEE, 1998, pp. 139–140.
[3] M.A.M. Vieira, C.N. Coelho, D.C. Da Silva, J.M. da Mata, Survey on wireless sensor network devices, in: Proc. of IEEE Conf. on Emerg. Technol. Fact. Autom., ETFA03, 2003, pp. 537–544.
[4] R.E. Mohemed, A.I. Saleh, M. Abdelrazzak, A.S. Samra, Energy-efficient routing protocols for solving energy hole problem in wireless sensor networks, Comput. Netw. 114 (2017) 51–66.
[5] T. Arampatzis, J. Lygeros, S. Manesis, A survey of applications of wireless sensors and wireless sensor networks, in: Proc. of Int. Symp. Mediterrean Conf. on Control and Autom., IEEE, 2005, pp. 719–724.
[6] A.Z. Abbasi, N. Islam, Z.A. Shaikh, A review of wireless sensors and networks' applications in agriculture, Comput. Stand. Interfaces 36 (2014) 263–270.
[7] W.R. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in: Proc. of 33rd Annu. Int. Conf. on Syst. Sci. Hawaii, IEEE, 2000, p. 10.
[8] V. Mhatre, C. Rosenberg, Homogeneous vs heterogeneous clustered sensor networks: a comparative study, in: Int. Conf. on Commun., IEEE, 2004, pp. 3646–3651.
[9] M. Yarvis, N. Kushalnagar, H. Singh, A. Rangarajan, Y. Liu, S. Singh, Exploiting heterogeneity in sensor networks, in: Proc. of 24th Annu. Jt. Conf. on Comput. Commun. Soc., INFOCOM 2005, IEEE, 2005, pp. 878–890.
[10] S. Tanwar, N. Kumar, J.J. Rodrigues, A systematic review on heterogeneous routing protocols for wireless sensor network, J. Netw. Comput. Appl. 53 (2015) 39–56.
[11] J. Huang, D. Ruan, Y. Hong, Z. Zhao, H. Zheng, IMHRP: Improved multi-hop routing protocol for wireless sensor networks, in: J. Phys. Conf. Ser, IOP Publishing, 2017, p. 012054.
[12] S. Kumar, P. Ranjan, R. Ramaswami, M.R. Tripathy, Resource efficient clustering and next hop knowledge based routing in multiple heterogeneous wireless sensor networks, Int. J. Grid High Perform. Comput. 9 (2017) 1–20.
[13] A.A. Bara'a, E.A. Khalil, A new evolutionary based routing protocol for clustered heterogeneous wireless sensor networks, Appl. Soft Comput. 12 (2012) 1950–1957.
[14] C.-W. Tsai, T.-P. Hong, G.-N. Shiu, Metaheuristics for the lifetime of WSN: A review, IEEE Sens. J. 16 (2016) 2812–2831.
[15] B.P. Deosarkar, N.S. Yadav, R.P. Yadav, Clusterhead selection in clustering algorithms for wireless sensor networks: A survey, in: Int. Conf. on Comput. Commun. Netw., ICCCn, IEEE, 2008, pp. 1–8.
[16] D.E. Goldberg, Genetic Algorithms in Search Optimization and Machine Learning, Addison-wesley Reading, 1989.
[17] A. Konak, D.W. Coit, A.E. Smith, Multi-objective optimization using genetic algorithms: A tutorial, Reliab. Eng. Syst. Saf. 91 (2006) 992–1007.
[18] H. Tamaki, H. Kita, S. Kobayashi, Multi-objective optimization by genetic algorithms: A review, in: Proc. of Int. Conf. on Evol. Comput., IEEE, 1996, pp. 517–522.
[19] R.S. Elhabyan, M.C. Yagoub, Two-tier particle swarm optimization protocol for clustering and routing in wireless sensor network, J. Netw. Comput. Appl. 52 (2015) 116–128.
[20] B. Singh, D.K. Lobiyal, A novel energy-aware cluster head selection based on particle swarm optimization for wireless sensor networks, Hum.-Centric Comput. Inf. Sci. 2 (2012) 1–18.
[21] R.V. Kulkarni, G.K. Venayagamoorthy, Particle swarm optimization in wireless-sensor networks: A brief survey, IEEE Trans. Syst. Man Cybern. C 41 (2011) 262–267.
[22] D. Karaboga, B. Gorkemli, C. Ozturk, N. Karaboga, A comprehensive survey: artificial bee colony (ABC) algorithm and applications, Artif. Intell. Rev. 42 (2014) 21–57.
[23] D. Karaboga, S. Okdem, C. Ozturk, Cluster based wireless sensor network routing using artificial bee colony algorithm, Wirel. Netw. 18 (2012) 847–860.
[24] R.V. Kulkarni, A. Forster, G.K. Venayagamoorthy, Computational intelligence in wireless sensor networks: A survey, IEEE Commun. Surv. Tutor. 13 (2011) 68–96.
[25] T.S. Rappaport, Wireless Communications: Principles and Practice, Prentice Hall PTR, New Jersey, 1996.
[26] S. Jannu, P.K. Jana, A grid based clustering and routing algorithm for solving hot spot problem in wireless sensor networks, Wirel. Netw. 22 (2016) 1901–1916.
[27] C. Li-jun, C. Dao-xu, X. Li, C. Jian-nong, Evolution of wireless sensor network, in: Wirel. Commun. Netw. Conf., WCNC 2007, IEEE, 2007, pp. 3003–3007.
[28] A.A. Abbasi, M. Younis, A survey on clustering algorithms for wireless sensor networks, Comput. Commun. 30 (2007) 2826–2841.
[29] L. Yu, N. Wang, X. Meng, Real-time forest fire detection with wireless sensor networks, in: Proc. of Int. Conf. on Wirel. Commun. Netw. and Mob. Comput., IEEE, 2005, pp. 1214–1217.
[30] List of wildfires, Wikipedia, 2018, https://en.wikipedia.org/w/index.php?title=List_of_wildfires&oldid=819470775, (Accessed 9 January 2018).
[31] D.C. Hoang, P. Yadav, R. Kumar, S.K. Panda, Real-time implementation of a harmony search algorithm-based clustering protocol for energy-efficient wireless sensor networks, IEEE Trans. Ind. Inform. 10 (2013) 774–783.
[32] S. Tyagi, N. Kumar, A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks, J. Netw. Comput. Appl. 36 (2013) 623–645.
[33] G. Smaragdakis, I. Matta, A. Bestavros, SEP: A Stable Election Protocol for Clustered Heterogeneous Wireless Sensor Networks, Boston University Computer Science Department, 2004.
[34] L. Qing, Q. Zhu, M. Wang, Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks, Comput. Commun. 29 (2006) 2230–2237.
[35] B. Elbhiri, R. Saadane, D. Aboutajdine, et al., Developed distributed energy-efficient clustering (DDEEC) for heterogeneous wireless sensor networks, in: 5th Int. Symp. on I/V Commun. and Mob. Netw., ISVC 2010, IEEE, 2010, pp. 1–4.
[36] D. Kumar, T.C. Aseri, R.B. Patel, EEHC: Energy efficient heterogeneous clustered scheme for wireless sensor networks, Comput. Commun. 32 (2009) 662–667.
[37] N. Javaid, T.N. Qureshi, A.H. Khan, A. Iqbal, E. Akhtar, M. Ishfaq, ED-DEEC: Enhanced developed distributed energy-efficient clustering for heterogeneous wireless sensor networks, Procedia Comput. Sci. 19 (2013) 914–919.
[38] T.N. Qureshi, N. Javaid, A.H. Khan, A. Iqbal, E. Akhtar, M. Ishfaq, BEENISH: Balanced energy efficient network integrated super heterogeneous protocol for wireless sensor networks, Procedia Comput. Sci. 19 (2013) 920–925.
[39] P.G.V. Naranjo, M. Shojafar, H. Mostafaei, Z. Pooranian, E. Baccarelli, P-SEP: a prolong stable election routing algorithm for energy-limited heterogeneous fog-supported wireless sensor networks, J. Supercomput. 73 (2017) 733–755.
[40] A. Kashaf, N. Javaid, Z.A. Khan, I.A. Khan, TSEP: Threshold-sensitive stable election protocol for WSNs, in: 10th Int. Conf. on Front. Inf. Technol., FIT 2012, IEEE, 2012, pp. 164–168.
[41] N. Mittal, U. Singh, Distance-based residual energy-efficient stable election protocol for WSNs, Arab. J. Sci. Eng. 40 (2015) 1637–1646.
[42] N. Mittal, U. Singh, B.S. Sohi, A stable energy efficient clustering protocol for wireless sensor networks, Wirel. Netw. 23 (2017) 1809–1821.
[43] N. Mittal, U. Singh, B.S. Sohi, A novel energy efficient stable clustering approach for wireless sensor networks, Wirel. Pers. Commun. 95 (2017) 2947–2971.
[44] R. Khanna, H. Liu, H.-H. Chen, Self-organisation of sensor networks using genetic algorithms, Int. J. Sens. Netw. 1 (2006) 241–252.
[45] S. Hussain, A.W. Matin, O. Islam, Genetic algorithm for hierarchical wireless sensor networks, J. Netw. 2 (2007) 87–97.
[46] A. Bari, S. Wazed, A. Jaekel, S. Bandyopadhyay, A genetic algorithm based approach for energy efficient routing in two-tiered sensor networks, Ad Hoc Netw. 7 (2009) 665–676.
[47] A. Norouzi, F.S. Babamir, A.H. Zaim, A new clustering protocol for wireless sensor networks using genetic algorithm approach, Wirel. Sens. Netw. 3 (2011) 362.
[48] L. Bhasker, Genetically derived secure cluster-based data aggregation in wireless sensor networks, IET Inf. Secur. 8 (2014) 1–7.
[49] S. Bayraklı, S.Z. Erdogan, Genetic algorithm based energy efficient clusters (gabeec) in wireless sensor networks, Procedia Comput. Sci. 10 (2012) 247–254.
[50] P. Kuila, S.K. Gupta, P.K. Jana, A novel evolutionary approach for load balanced clustering problem for wireless sensor networks, Swarm Evol. Comput. 12 (2013) 48–56.

[51] S.S. Kumar, S. Vishwas, GDFEC protocol for heterogeneous wireless sensor network, in: Comput. Intell. in Data Min., vol. 1, Springer, 2015, pp. 345–354.
[52] S.K. Gupta, P.K. Jana, Energy efficient clustering and routing algorithms for wireless sensor networks: GA based approach, Wirel. Pers. Commun. 83 (2015) 2403–2423.
[53] X. Yuan, M. Elhoseny, H.K. El-Minir, A.M. Riad, A genetic algorithm-based dynamic clustering method towards improved wsn longevity, J. Netw. Syst. Manag. 25 (2017) 21–46.
[54] M. Elhoseny, A. Farouk, N. Zhou, M.-M. Wang, S. Abdalla, J. Batle, Dynamic multi-hop clustering in a wireless sensor network: Performance improvement, Wirel. Pers. Commun. (2017) 1–21.
[55] K. Rajeswari, S. Neduncheliyan, Genetic algorithm based fault tolerant clustering in wireless sensor network, IET Commun. 11 (2017) 1927–1932.
[56] T. Bhatia, S. Kansal, S. Goel, A.K. Verma, A genetic algorithm based distance-aware routing protocol for wireless sensor networks, Comput. Electr. Eng. 56 (2016) 441–455.
[57] S. Dehghani, B. Barekatain, M. Pourzaferani, An enhanced energy-aware cluster-based routing algorithm in wireless sensor networks, Wirel. Pers. Commun. 98 (2018) 1605–1635.
[58] P.S. Ragavan, K. Ramasamy, Optiized routing in wireless sensor networks by establishing dynaic topologies based on genetic algorith, Clust. Comput. (2018) 1–7.
[59] S. Jannu, S. Dara, K.K. Kumar, S. Bandari, Efficient algorithms for hotspot problem in wireless sensor networks: Gravitational search algorithm, in: Int. Symp. Intell. Syst. Technol. Appl., Springer, 2017, pp. 41–53.
[60] K. Sundaran, V. Ganapathy, P. Sudhakara, Fuzzy logic based unequal clustering in wireless sensor network for minimizing energy consumption, in: 2nd Int. Conf. on Comput. Commun. Technol., ICCCT 2017, IEEE, 2017, pp. 304–309.
[61] S. Yasotha, V. Gopalakrishnan, M. Mohankumar, Multi-sink optimal repositioning for energy and power optimization in wireless sensor networks, Wirel. Pers. Commun. 87 (2016) 335–348.
[62] M. Masdari, F. Naghiloo, Fuzzy logic-based sink selection and load balancing in multi-sink wireless sensor networks, Wirel. Pers. Commun. 97 (2017) 2713–2739.
[63] F.H. El-Fouly, R.A. Ramadan, M.I. Mahmoud, M.I. Dessouky, REBTAM: reliable energy balance traffic aware data reporting algorithm for object tracking in multi-sink wireless sensor networks, Wirel. Netw. (2016) 1–19.
[64] B. Zeng, Y. Dong, An improved harmony search based energy-efficient routing algorithm for wireless sensor networks, Appl. Soft Comput. 41 (2016) 135–147.
[65] Adnan, Md Akhtaruzzaman, et al., Bio-mimic optimization strategies in wireless sensor networks: A survey, Sensors 1 (2013) 299–345.

Sandeep Verma is currently a Research Scholar in ECE Department in Dr BR Ambedkar National Institute of Technology, Jalandhar. He completed his M.E. in Electronics and Communication from NITTTR Chandigarh in 2013. He received his B.Tech degree in Electronics and Communication from LCET Katani Kalan, Ludhiana. Hi research interest includes energy efficient Wireless Sensor Networks and IoT based architectures.

Dr. Neetu Sood received the B.Tech. degree (with Honour) in Electronics and Communication from Sant Longowal Institute of Engineering and Technology, India 2000. In year Jan. 2002, she completed her M.E.Degree (With Honour) in Electronics and Communication Engineering from Thapar University, Patiala, India. She completed her PhD in ECE from Dr. B R Ambedkar National Institute of Technology, Jalandhar. Currently, she is working as an Assistant Professor at Dr. B R Ambedkar National Institute of Technology, Jalandhar. Her research interest includes wireless sensor network, simulation of wireless systems based on OFDM and simulation of fading channels.

Dr. Ajay K Sharma received his BE in Electronics and Electrical Communication Engineering from Punjab University Chandigarh, India in 1986, MS in Electronics and Control from Birla Institute of Technology (BITS), Pilani in the year 1994 and PhD in Electronics Communication and Computer Engineering in the year 1999. Currently, He is working as Director at NIT Delhi and also has a charge as a Director at NIT Hamirpur. He has published 272 research papers in the International/National Journals/Conferences and 12 books. He has supervised more than 18 Ph.D. and 48 M.Tech thesis. He is technical reviewer of reputed international journals.