



An energy efficient weighted clustering algorithm in heterogeneous wireless sensor networks

Vivekanand Jha¹ · Rashika Sharma¹

Accepted: 3 March 2022 / Published online: 31 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Wireless sensor networks are most used to monitor remote environments. Multitudinous sensor nodes gather data in a self-governing manner, operating on an exhaustive source of energy or battery. Clustering process structures the network into a hierarchy wherein sensor nodes gather data passed to selected cluster head nodes which perform data processing, aggregation, and transfer it to a base station. Prolonging network lifetime and enhancing total data transmission to base station are major challenges in wireless sensor network and same is addressed in this work. A distributed energy-based epoch is used in this paper to determine node eligibility to become cluster head and a multi-parameter-weighted scalarization function is proposed to determine best cluster head candidates in order to manage dynamic and multi-characteristic node heterogeneity. The parameters used are distance to base station, expected cluster head lifetime, average cluster member node lifetime and maximum power consumed by a cluster member node. A novel weight computation strategy using analytical hierarchy process is introduced in this paper which enhances the optimality of scalarization function value. The proposed algorithm is distributed over two phases as network setup phase and clustering phase. The network setup phase computes the energy model and optimal number of cluster heads. The second phase proposes the cluster head selection process using weight-based scalarization and introduces the novel weight selection method. Finally, network operation enters the data transmission phase. The results show an enhancement in throughput at base station, with an increase of close to 30% along with an increase in the network lifetime of up to 20% as measured by last node death. The simulation results are produced in comparison with the considered base protocol of DEEC as well as other protocols using similar concepts for implementation. However, utilization of a two-step cluster heads selection process including unique node epochs for shortlisting and scalarization function-based node fitness, along with optimal weight selection procedure, has led the proposed model to give better results on simulation and analyzation than preexisting algorithms.

Extended author information available on the last page of the article

Keywords Wireless sensor network · Energy efficient · Clustering · Heterogeneous · Analytical hierarchy process

1 Introduction

Smart sensors have been widely manufactured since the onset of micro-electro-mechanical systems (MEMS) technology. These are tiny devices with extremely low power for processing and minimal computing resources. A sensor node consists of a memory, processor, a power supply, an actuator, etc. Wireless sensor networks (WSN) represent an infrastructure-less deployment of nodes spatially distributed in an ad hoc manner. These WSNs sense the environment, collect and process the data [1]. In order to send information across to a base station (BS), a radio module fulfills the wireless implementation. The principal source of energy in a sensor node is an exhaustive battery. This makes conservation of energy one of the central issues of wireless sensor networks. In the literature, certain techniques have been introduced to deal with this issue. Duty cycling has been mentioned as a method that uses topology control to select a minimum subset of nodes that ensure network connectivity to remain active and the rest go to sleep and save energy. Data-driven approaches minimize unneeded data collection and power consumption. The authors in [2, 3] have discussed more recent developments in battery alternatives based on energy harvesting including solar power, mechanical vibrations, temperature variations, electromagnetic fields, wind or water flow, etc.

Since WSNs are deployed in a wide variety of environments, each environment presents challenges in the network performance and the quality of services. MAC and routing protocols, security algorithms, scheduling mechanisms, CPU and operating systems together are used to construct solutions for these network issues [4]. Apart from the energy conservation, node failure in harsh environments [5], minimizing traffic load [6], determination of node positions [7], their programmability and reconfigurability and system security are also issues [8, 9]. The issue of efficient energy utilization has been addressed extensively in existing research.

Several “power-aware” routing protocols choose optimal routes for data transmission based on node energies along a route and show that “multi-hop” transmission over shorter links is more efficient than communication over longer ranges in wireless networks spread over a significantly large area [10]. Routing for data transmission can be in flat or in hierarchical WSN based on the topology. In a hierarchical network topology, it can be ensured that there is consistent energy dissipation, fairness of channel allocation, lower latency in multiple hops, designated data aggregation and data collision avoidance [11]. A hierarchy defines the network in a structured arrangement. Layers can be defined within the network- low energy layer for sensing the environment and high energy layer for data aggregation, processing, forwarding, etc. [12]. This translates to a two-tier approach of grouping nodes into clusters with a leader or cluster head (CH). The CH is responsible for data aggregation and forwarding to the BS while the remaining nodes sense the physical surroundings. CHs form the upper layer and member nodes occupy the lower layer, periodically sending data to their

respective CH. This points to the ensuing issue of selecting most suitable cluster heads from various available sensor nodes [13].

WSNs have sensor nodes with varying capabilities including different energy levels, computing and processing powers, sensing ranges, etc., which give rise to heterogeneity within the network. Compared with homogeneous implementations, the deployment, control of topology and managing energy efficiency are more complex in heterogeneous networks [14]. However, it is found that heterogeneity can be utilized constructively in various WSN clustering-based routing protocols. This is because giving it due consideration and adopting the right measures to evaluate nodes' characteristic values empowers the protocol to make highly informed decisions in assigning sensing and/or computing duties such as CH selection. This eventually leads to a higher efficiency and network lifetime. This is validated from the results of various clustering protocols which give evidence that accounting for heterogeneity plays a pivotal role in overall network performance. Such protocols prove more efficient and have a longer lifetime [15].

In this paper, it is proposed to utilize heterogeneity into cluster-based algorithms by combining heterogeneous characteristics of a node to formulate a single value. This gives an overall objective information about that node. This value will then be used by the protocol to assign roles to the node within the two-tier network hierarchy, i.e., either a member or CH node. In [16], it is depicted how node characteristics are often conflicting, meaning that for ideal results values may be minimum or maximum simultaneously (e.g., spent energy- minimum, and processing capacity- maximum). In [17], a mathematical method to form such combinations is given called the weighted sum or weighted scalarization technique. In this case, it is the sum of characteristic values pre-multiplied by weights.

Existing clustering algorithms have a prevalent issue of maximizing network lifetime as well as managing node characteristic heterogeneity. These issues are related as heterogeneity that eventually affects the network lifetime. This is because nodes with multiple highly varying characteristic values are not accounted for by the algorithm which is not tailored according to the node's characteristics. To manage node heterogeneity, it needs to be determined which parameters have a higher preference in order to select the best fit nodes such that they can perform high energy tasks within the network. After determination of the preference, optimal weight calculations for each parameter are required for the introduced weight-based clustering. To this end of achieving higher energy efficiency and network lifetime, an algorithm based on distributed energy and weighted scalarization has been proposed in this paper. Node characteristics contributing to heterogeneity like node degree, lifetime, distance from base station and maximum power consumed have been considered which would be pre-multiplied by calculated weights to give a single sum value. The weighted sum represents a fitness value which is used to select appropriate cluster head nodes. Further, weights are determined for each node characteristic by a weight determination technique that has been introduced in the following work and has been used to promote enhancement of the network lifetime.

The rest of the paper is organized as: Sect. 2 describes literature review on existing research, Sect. 3 presents proposed work and Sect. 4 explains the simulation

results. Finally, Sect. 5 concludes the work presented in this paper and discusses future scope.

2 Literature review

The heterogeneity of a network depends on various differing sensor node attributes. However, the early cluster-based low energy adaptive clustering hierarchy (LEACH)-like protocols [18–21], assumed complete network homogeneity and CH selection procedure, did not accommodate the different node energy levels. It depended on a fixed probability. In stable election protocol (SEP) [22], however, the focus shifted toward heterogeneity which was solely based on two distinct energy levels- advanced nodes with higher and normal nodes with lower energies. Using two parameters, (i) higher energy nodes fraction (m) and (ii) extra energy factor (α), SEP optimized the network stability region, the clustering hierarchy process and CH selection. This protocol attempted to constraint the energy consumption in a balanced and equitable manner. Effectively, cluster head positions are held more often by the advanced nodes rather than by normal nodes. This ensured a better energy consumption distribution. More improvements in the connection methodology between normal nodes and BS were suggested by the authors in [23, 24]. However, multi-parameter node heterogeneity was not accounted for in these protocols and thus network lifetime was limited.

In an ideal case, nodes would have equal energy at each new round, and the probability of each node becoming CH in each round would be fixed, which was assumed by the above protocols. In terms of heterogeneity, the above protocols dealt with fixed energy levels. However, node energies vary continuously and cannot be confined into fixed slots and values; thus, it is important to handle individual node energies. This issue was handled in distributed energy-efficient clustering (DEEC) [25], which calculated a theoretical value of lifetime for the WSN. This gave a unique spending energy value to every node according to its current energy state in subsequent rounds. The value of remaining energy for a node was utilized by this protocol based on which a different value of epoch was assigned for each node's selection as CH in each round. Thus, DEEC selected a unique probability for each node. This solved the problem of nodes dying at variable times and extended the stability region. For every high energy value node, its probability to be CH was defined to be more than an ideal and fixed probability value.

In a shift toward addressing multiple differing node attributes contributing toward heterogeneity in addition to varying energy levels, lifetime sensitive weighted clustering algorithm (LTS-WCA) [26] was introduced. Moving a step further, the authors adapted weight-based clustering that utilized the weighted sum function method, in WSN. [27–29] show that this weighted clustering technique had been previously applied mainly on mobile ad hoc networks (MANET) [30], in algorithms which focused on selecting cluster heads in clustering protocols for MANET. LTS-WCA made modifications and enhancements to this method and introduced the extension of weight-based clustering to wireless sensor networks. In [31–33], the authors suggested similar weight-based approaches taken for the CH selection issue

with clustering protocols, by considering a variety of parameters and characteristics of heterogeneous nodes like node degree difference, distances between nodes and neighbors, device power, signal length, remaining energy, mobility, etc. They further highlighted how the scalarization method allowed incorporation of multiple heterogeneity parameters into the clustering process. However, the process of scalarization does not account for optimization of the function value calculated or suggests methods to compute weights used for each node characteristic.

In OWCA [34], an optimized weight-based clustering algorithm was proposed for mobile wireless sensor networks which incorporated parameters of sensors such as: transmission radius, degree difference, sum of all the distances between a single node and its neighbor nodes, speed of node, its individual cumulative time and energy to form the fitness function. This paper showed a step forward in the family of weight-based clustering algorithms, however the mechanism to select assigned weights was not explored, which had been allocated based on a general preference order.

In distributed weight-based energy-efficient hierarchical clustering (DWEHC) [35], the authors considered characteristics such as cluster range, distance between neighboring nodes and residual energy for fitness calculation. The hierarchy was established in multiple levels rather than two. The node self-evaluated its fitness and took its position in an appropriate hierarchy level. This paper showed a better performance than primary protocols such as hybrid energy-efficient distributed (HEED) [36]. Energy-based heterogeneity in WSN and multi-level hierarchical clustering with the introduction of weights was an important advancement and takeaway which were not taken up in preexisting protocols.

In [37], the authors reviewed and analyzed recent clustering algorithms. They highlighted the importance of using clustering which allows significant reduction of overhead, retaining small networks features in large and dense ones. Weight-based metric was recognized as a key performance heuristic to determine cluster heads in this paper. Control overhead was specified as a major parameter which greatly affects energy expenditure and greater control overhead was shown to result in higher loss of energy. An attempt has been made to minimize such an overhead in the following work, based on this conclusion.

The above literature showed that weight coefficients used in weighted calculations generally remain fixed and do not follow a selection methodology. In [38], authors tried to improve upon this issue and increase the optimality of the scalarization function value in a multi-parameter-based scenario to select CH nodes. They dealt with the problem of subjective assignment of coefficients by selecting these values through an iterative weight selection strategy based on comparisons of delay and maximum throughput given by the varying parameters. The weight selection process proposed in the following work presented by us extends this method to wireless sensor networks.

A detailed mathematical study linking scalarization in multi-parameter functions and their applications in WSN was shown in [39]. It highlighted the use of these functions in clustering protocols and the calculation of weights. The authors presented three main techniques to this end- (i) equal weights, (ii) rank order centroid weights and (iii) rank-sum weights. Other weights and score determination

techniques are highlighted in [40–43]. In [16], the authors highlighted the case of weight selection. It was suggested that the fitness value of the mathematical scalarization function formed by summing of weighted node parameter values is highly dependent on the process of weight selection. While few techniques exist to accomplish this goal, it highlighted the analytical hierarchy process (AHP) [42] as a good fit to be a powerful multi-criteria decision-making method for determination of relative weights. The following work incorporates this method to the proposed weight-based clustering algorithm to determine the function weights.

In [44], a weight-based algorithm was proposed for static wireless sensor networks that considered characteristics like remaining energy-based delay of node, average sensor node lifetime, CH lifetime, and maximum power consumed by a node. It, however, did not include a procedure for weights assignment. The algorithm proposed by us aims to further optimize this procedure by including node degree as an additional parameter, using distributed energy-based clustering and applying a weight determination scheme.

Consider a WSN for deployment that is spread in a square shaped area, where static sensor nodes have varying initial energy and are randomly deployed, with a central base station. Such a network is referred to in the subsequent context. The network is considered as heterogeneous based on varying node parameter values and energy levels. In clustering protocols based on the methodology followed by LEACH [18], SEP [22], etc., equitable node energy distribution is aimed for using a fixed and uniform CH rotation strategy. Such protocols show that equal energy nodes undergo dynamic, uneven drops in energies with progression in protocol rounds and data transmission. This occurs as some nodes that get selected as cluster heads drain more energy than others during data transmission due to an additional data aggregation and computation load. A node gets elected as cluster head based on a certain probability value on a rotation basis. It can only get elected once in a rotating epoch which is equal to a fixed number of rounds. Probability to become CH in such cases remains fixed for all nodes eligible, based on whether the node was not a CH in the previous epoch, the total number of nodes, the fixed number of CH in each round and the current round number.

This does not keep the energy distribution truly equitable due to the unequal energy dissipation, random deployment of nodes in a large area, a large variation in distances between nodes and varying cluster assignments. The network then seizes to remain homogenous, and heterogeneity in terms of various node parameters arises. Also, the length of the epoch remains predetermined and fixed. There is no change in the value of the epoch based on current energy or any other current attributes of the node. Hence, node heterogeneity is not accounted for by such protocols. On that account, robust algorithms that cater to heterogeneity among nodes by monitoring characteristic changes in energy, degree, delay, node lifetimes, power consumed, etc., can further maximize network lifetime [38].

Attributes like current node energy values, location, power, degree and proximity to base station are some examples of important node characteristics that are often considered while implementing various clustering protocols in heterogeneous networks [16, 34–39, 44]. These individual characteristics of a node have a direct impact on network energy dissipation during the execution of a protocol. Proximity

to base station as an example directly impacts lower transmission energy dissipation, and gradual node energy loss directly increases network lifetime. Thus, such attributes have high significance in determining the fitness of nodes, selection of cluster heads and in determining the optimality of network lifetime. They have been incorporated in this proposed paper for the scalarization method in weight-based clustering. This method accounts for current node attribute values by using a scalarization function [17]. The weighted sum method is used to express such a function:

$$\alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x) + \dots + \alpha_n f_n(x) = v \quad (1)$$

here $f_i(x)$ is an attribute of a particular node x like degree, mobility, delay value, etc. And the weights have the following relation:

$$\alpha_1 + \alpha_2 + \alpha_3 + \dots + \alpha_n = 1 \quad (2)$$

α_i is a constant and represents the weight allocated to $f_i(x)$ and signifies the extent of effect of $f_i(x)$ on the value of v . Above-mentioned attributes may need to be maximized or minimized individually in order to achieve optimality. To determine best cluster head candidates, the values generated by the function (v) are measured and the node that has minimum value is selected as cluster head.

In [16], it has been shown that the weights selected for attributes in the scalarization function primarily are pre-determined and fixed according to a generalized preference among the attributes, where higher weight means higher preference. The authors have concluded that weights have great significance in the optimality of the final function value. Also, typically the selector for weights is not aware of the exact values which would most appropriately generate a satisfactory optimal solution. Hence, a weight adjustment process is not known that can consistently change the solution and drive it toward optimality. This leads to preference order-based weight selection. This technique has been seen in various existing algorithms [27–35]. It also implies that it is difficult to formulate algorithms that are heuristic, which begin with a set of weights, and with iteration, generate weight vectors that give satisfactory optimality.

Thus, selection of values of weights in an objective function as a linear weighted sum is highly associated with the optimal value of that function. Since there is no a priori correlation between a weight vector and a solution vector, a selector generally resorts to subjective weights which results in poor objectivity to the linear weighted-sum method.

Existing scalarization function-based algorithms often rely on such a function alone to determine the best candidate as cluster head. Hence, data transmission in the protocol involves multiple rounds of calculating the function value for each node across the entire network. While this may be an efficient approach for small networks, generation of repetitive function values for all nodes with every round in considerably larger networks would not prove to be an efficient approach. This has been highlighted in [37] as one of the major issues in network control that generates overhead and leads to a substantial loss in energy of the network.

The work considers the issue of node heterogeneity which arises due to a disparity or differences in various node parameters such as memory, energy, bandwidth,

computing power and sensing range. It needs to be decided which node parameters are to be given a higher preference for the determination of nodes that are fit to perform high energy consumption duties within the network. After the preference order of parameters is decided, the assignment of weights that represent the preference order needs to be done for each parameter and a technique to perform this task is to be established. The above problem is addressed in the proposed work and the outcome is a distributed energy-based weighted clustering algorithm that uses mathematical scalarization. An attempt has been made to devise a weight selection strategy for the scalarization function of weighted clustering algorithm using proposed analytical hierarchy process (AHP). A distributed energy-efficient approach for conduction of protocol rounds has been suggested. The aim is to prolong the network lifetime and improve protocol performance in terms of throughput and efficiency in the following work.

3 Proposed solution

This paper aims to improve network lifetime, efficiency and throughput, while mitigating persisting issues in weight-based clustering algorithms. The cluster head selection process in the proposed algorithm introduces a weighted fitness function incorporating important node characteristics as a weighted sum. The focus is also on limiting repeated fitness function value evaluation for the entire set of nodes by adopting a distributed energy-based rotation strategy which restricts the eligibility of certain nodes to become cluster heads. This weakens excessive computation overhead and increases energy efficiency. Additionally, an AHP-based scheme for calculation of weights is introduced which optimizes the value of the weighted scalarization function. This further enhances the CH selection process as worthy cluster heads are chosen, leading to balanced energy dissipation and higher efficiency.

The algorithm is presented in two phases. The first phase presents the setup of the network including necessary computations of the energy model and optimal parameter values like node distances and number of protocol rounds required for data transmission. The second phase elaborates the cluster head selection process including distributed energy epoch-based CH eligibility and proposal of the weighted fitness function along with its weight determination strategy. This is followed by the assignment of remaining nodes to suitable clusters based on their transmission range and energy computations after data transfer. Lastly, an overall algorithm is presented that ties together the proposed method.

3.1 Phase 1: network setup

In this work it is assumed that N sensor nodes are deployed randomly and uniformly distributed across the network area assumed to be square shaped $M \times M$ region. The nodes have the ability to sense and transmit data from their surrounding which must reach the base station, placed at the center of the region.

The radio energy model for this paper has been considered same as stated in the literature [18, 45]. For any l -bits message at distance d , the transmission energy $E_{\text{Trans}}(l, d)$ and receiving energy $E_{\text{Rec}}(l)$ by radio model are considered as:

$$E_{\text{Trans}}(l, d) = l \times E_{\text{Elec}} + l \times e_{\text{fs}} \times d^2 \quad (3)$$

$$E_{\text{Rec}}(l) = l \times E_{\text{Elec}} \quad (4)$$

where E_{Elec} is energy dissipated per bit by the transmitter and $(e_{\text{fs}} \times d^2)$ is the energy required for amplifying the signal. Also, E_{Elec} and e_{fs} are constant. Energy value to a node x_i is assigned with a variable factor μ_i and is given by $(\mu_i + 1) \times E_o$ to build a two-level ($\mu_i = 1$) or multi-level heterogeneity ($\mu_i > 1$) that results into normal nodes ($\mu_i = 0$) and advanced nodes ($\mu_i \geq 1$). E_o is the initial energy for all normal nodes. A constant fraction of η among all the nodes is termed as the advanced nodes that have $(\mu_i + 1) \times E_o$ as initial energy, where $(0 < \eta < 1)$.

In this phase, the values of optimal number of CHs and maximum rounds for transmission are computed in this paper as per given in the literature [25]. The value of optimal number of cluster heads (k_{opt}) in a round is given by:

$$k_{\text{opt}} = \frac{\sqrt{N}}{\sqrt{2\pi}} \times \sqrt{\frac{e_{\text{fs}}}{e_{\text{mp}}}} \times \frac{M}{d_{\text{BS}}^2} \quad (5)$$

where d_{BS} is the Euclidean distance of a node from its base station, the value of N is the total number of nodes that are deployed and M is the dimension of deployed region. The neighborhood nodes are determined based on communication range and the sensor nodes senses as per their sensing range. In this work an assumption is made that the communication range is on much higher side than the sensing range.

3.2 Phase-2: cluster head (CH) selection and data transmission

The phase begins with determination of nodes' eligibility to qualify as CH. This is done using a node's unique probability value given by an epoch that is based on optimum dissipation energy assigned to that node. This section then introduces the devised weighted fitness function, which is the sum of weighted node parameters, used as the key decision criteria for selection of cluster heads. Following this, a new technique for selection of weights in weight-based clustering is introduced that follows the analytical hierarchy process.

Eligibility of an alive node x_i to become CH can be given by a unique probability $P^i(r)$. The value of $P^i(r)$ has been derived as per the derivation provided in the literature [25]. Let $E_{\text{Res}}^i(r)$ for a node x_i represents its residual energy in a round r and the average energy of network that consists of N nodes in a single round is given by $E_{\text{Avg}}^N(r)$. In an ideal scenario, nodes would have equal energy in an epoch and average probability of P_{Avg} to become CH. Hence, the total number of possible CH is $(P_{\text{Avg}} \times N)$ in every round and all nodes would eventually die at about the same time. Since the considered conditions are not ideal and node energies are highly variable,

the unique probability of a node x_i to become CH is given by its residual energy relative to average network energy. Hence, the probability is given as

$$P^i(r) = P_{Avg} \times \frac{E_{Res(r)}^i}{E_{Avg(r)}^N} \tag{6}$$

Since a node x_i belonging to a fraction η of total nodes (N) contains μ_i times more initial energy, P_{Avg} cannot be directly considered as constant for all nodes. The value of μ_i can be either fixed for a two-level heterogeneity or belong to a range for multi-level heterogeneity where $\mu_i > 0$. For each node, P_{Avg} is replaced by P_{Avg}^i , which is given by a weighted probability based on the node's initial energy as given below.

$$P_{Avg}^i = P_{Avg} \times \frac{N(1 + \mu_i)}{N + \sum_{i=1}^N \mu_i} \tag{7}$$

Hence, $P^i(r)$ can be written as

$$P^i(r) = P_{Avg}^i \times \frac{E_{Res(r)}^i}{E_{Avg(r)}^N} = P_{Avg} \times \frac{N(1 + \mu_i)}{N + \sum_{i=1}^N \mu_i} \times \frac{E_{Res(r)}^i}{E_{Avg(r)}^N} \tag{8}$$

The probability for a node to be CH is given by Eq. (8) and now it is the turn to fix the eligibility of CH. The eligibility is fixed here based on proposed weighted fitness function. The proposed weighted scalarization sum is designed by using following parameters like distance from base station, lifetime of CH, node degree, power transmission and average node lifetime. These are combined to give a parametric fitness function that computes to a novel value used to determine which nodes are best fit to become CH. The proposed fitness function value (F_{Value}) is given as:

$$F_{Value} = (\alpha \times D_{BS}) - (\beta \times CH_{Life}) - (\gamma \times M_{AvgLife}) + (\delta \times P_{Max}) \tag{9}$$

where D_{BS} is Euclidean distance between the considered node and the base station, CH_{Life} is the expected life of the considered node if chosen to be CH, $M_{AvgLife}$ is average lifetime of the member nodes of the cluster and P_{max} is maximum power consumption by a member node of the cluster. For a good candidate of cluster head, the F_{Value} should be minimum as far as possible. The selected node characteristic of D_{BS} which is Euclidean distance between the considered node and the base station should have a low value for minimum energy dissipation according to the existing literature and hence more optimal. Similarly, CH_{Life} which is the expected life of the considered node if chosen to be CH should have a higher value for minimum dissipation of energy and thus taken as negative within the scalarization function. Also, $M_{AvgLife}$ is average lifetime of the member nodes of the cluster, also indicating having a higher value for optimality, and is also taken as negative within the function. P_{max} is maximum power consumption by a member node of the cluster, which is directly proportional to energy dissipation; thus, it must be minimum to be more optimal. Thus overall, the construction of the scalarization function has been done in a manner such that a lower value indicates higher optimality.

Thus, the values of D_{BS} and P_{Max} are minimized and the values of CH_{life} and $M_{AvgLife}$ are maximized in order to achieve minimum F_{Value} . Hence, in Eq. (9) the term represented by CH_{life} and $M_{AvgLife}$ is kept negative intentionally.

Here, α, β, γ and δ are constants and are respective weights assigned to each of the parameters. Also, their relation is derived as

$$\alpha + \beta + \gamma + \delta = 1 \quad (10)$$

The cluster head lifetime (CH_{Life}) is calculated as the ratio of residual energy of probable cluster head and transmission energy required for data transfer to the base station:

$$CH_{Life} = \frac{E_{Res(CH)}}{E_{Trans(CH-BS)}} \quad (11)$$

here $E_{Res(CH)}$ is the residual energy of the probable cluster head and $E_{Trans(CH-BS)}$ is the energy required to transmit the unit data from probable cluster head to base station.

Here, for each node that is eligible to be cluster head, a set of its probable cluster member nodes is found, based on the nodes' transmission range. This gives a fair chance at predicting an energy expenditure value of the cluster formed by the eligible cluster head, using its members average life.

The average lifetime ($M_{AvgLife}$) of member nodes within a cluster is the ratio of summation of individual cluster member lifetime and total number of member nodes. If CH_{Mem} is the total number of member nodes within any i^{th} cluster, the average lifetime of member nodes for that cluster can be expressed by

$$M_{AvgLife} = \frac{\sum_{i=1}^{CH_{Mem}} N_{Life}^i}{CH_{Mem}} \quad (12)$$

where the value of lifetime of any i^{th} member node of a cluster is derived as,

$$N_{Life}^i = \frac{E_{Res}^i}{E_{Trans(i-CH)}} \quad (13)$$

Here E_{Res}^i is the residual energy of the i^{th} member node and $E_{Trans(i-CH)}$ is the transmission energy required for i^{th} member node to communicate unit data with cluster head.

The maximum power consumption by any i^{th} member node within any cluster is represented by P_{Max} and it is the maximum transmission energy required to communicate with cluster head. Hence, this value of maximum power consumption is represented by

$$P_{Max} = \text{Max}(E_{Trans(i-CH)}), \forall i \quad (14)$$

After formulation of the fitness function, a new technique is proposed here to compute the weights for the function, i.e., values of α, β, γ and δ . The

computation is built on top of the analytical hierarchy process, which spans over various stages of the weight determination procedure.

AHP presented by the researchers in [40] outlines four well-defined stages in the process of determination of parametric weights. It describes the first stage as the creation of hierarchical levels with a top-level central goal followed by another level of important decision-making criteria (attributes) for that goal. These are supported further by distinct parameters or entities in the last level that complete the hierarchy. In the next two stages, priorities among multiple entities are established by comparing and studying their individual effects on objective outcomes from the system on which this process is applied. In the last stage, weight assignment to the parametric entities is done.

The presented technique for weight selection in this paper considers these stages as a foundation and introduces a process of choosing weights for weight-based scalarization functions in clustering algorithms. The four stages of AHP in this paper include hierarchical structuring, grouping and comparing entities, priorities setting and weight assignment.

Beginning with the step of hierarchical structuring, the goal established is selection of cluster head. A detailed hierarchical structure that elaborates on various contributing criteria and node parameters that affect or cause heterogeneity within the network is shown in Fig. 1(a). It is a comprehensive representation for a multi-parameter model. For the proposed work, we have considered the hierarchical structure as shown in Fig. 1(b). Consideration of all parameters would be a more complex implementation although it would be the ideal and more accurate model. However, for the clarity of our proposed work, Fig. 1(b) is considered. Proximity to base station and the energy of clusters within network have been selected as the two major criteria to determine node's eligibility to be cluster head. Four entities (A , B , C and D) which are same as the parameters selected for the fitness function, fall under these criteria to complete the proposed hierarchical structuring.

Next, the hierarchy is established, a priority order for entities on the last level needs to be set under the criteria they belong to, based on a network performance metric. This priority reflects which entity is potentially more likely to influence the selection of a desired cluster head based on its effect on the performance metric. The considered network performance metric is the network lifetime that is compared under independent effect of each entity parameter.

For priority building within the two criteria of *Cluster Energy* and *Base Station Proximity*, the *Cluster Energy* criteria with three parameters is considered first and a priority order is set for these three parameters. The sole parameter under *Base Station Proximity* is then considered and given a position among the priority order previously set under *Cluster Energy*.

Studying the *Cluster Energy* criterion, three entity parameters B , C and D give six distinct possible priority orders like: (B, C, D) , (B, D, C) , (C, B, D) , (C, D, B) , (D, B, C) and (D, C, B) . The aim is to find the best suited order influencing *Optimal Cluster Head Selection*. In order to judge the effect of these parameters on network lifetime, the study of the existing protocols like LEACH [18], SEP[22] and DEEC[25] has been conducted through the simulation. The obtained data for energy dissipation patterns in clusters by executing various number of simulation iterations

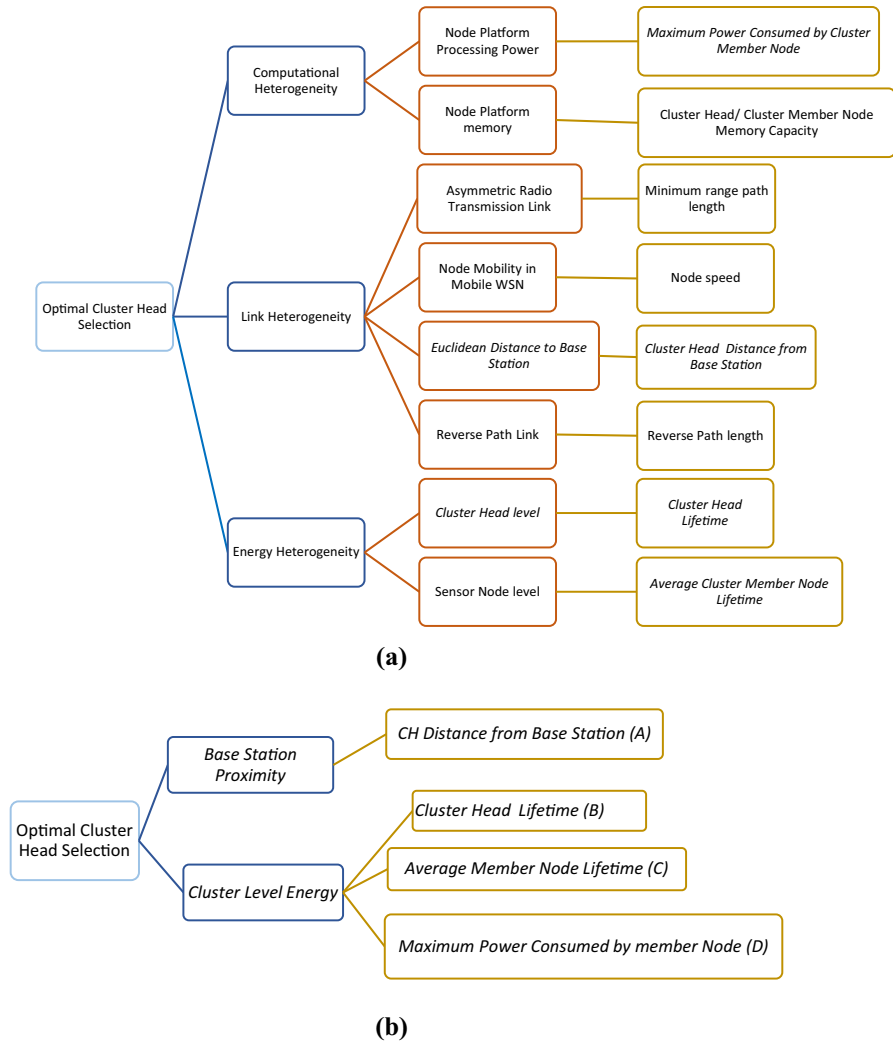


Fig. 1 Proposed hierarchy structure **(a)** comprehensive model **(b)** model used in this work

with varying initial energy level have been studied. The network lifetime performances have been studied here in association with cluster energy dissipation patterns and conclusions are drawn for the priorities.

The results confirm that a high power-consuming cluster member node quickly drains energy due to large range transmission. Within few such cluster assignments, such a node’s death is imminent. This rapid energy depletion factor largely deems the corresponding cluster head as unfit and establishes stark correlation with network lifetime and rapid energy depletion. Next, data confirm that nodes with higher values for residual node energy have direct influence on optimal selection as cluster head, as it performs high energy-consuming aggregation

Table 1 Ranked entities

Rank	1	2	3	4
Entities	A	D	B	C

Table 2 Parameter comparisons

Comparisons	Scores			
	S(A)	S(D)	S(B)	S(C)
(AwithD)	10	9	N/A	N/A
(DwithB)	N/A	10	9.3	N/A
(Bwith C)	N/A	N/A	10	8

and transmission to base station duties. Thus, it plays a key role in equitable energy dissipation and hence in increasing network lifetime. Lastly, while carrying out study, the values of average lifetime of member nodes for clusters have been tabbed. The study confirms that the simulations that have clusters having members with higher lifetime as an overall average generally give better network lifetime performance. The conclusion drawn here from the study undertaken is that the values of network lifetime when *Maximum Power Consumed by Member Node (D)* is considered precedes that when *Cluster Head Lifetime (B)* is considered which in turn precedes the network lifetime when *Average Member Node Lifetime (C)* is considered. Hence, the order established is (D, B, C), i.e., D followed by B followed by C in order to achieve maximum lifetime by the network under *Cluster Energy* criteria.

Next, a priority order of all the entities is to be determined. Since distance from base station is the sole entity under *Base Station Proximity* criteria, there is no priority order required here.

There exists a straightforward connection between *CH Distance from Base Station (A)* and network lifetime. Whenever the former is consistently lower across cluster heads, nodes die out slowly and it improves the lifetime of the overall network performance. In order to put *CH Distance from Base Station (A)* in the final priority order this preceded all the previously computed values under *Cluster Energy* criteria. Hence, the final priority established is (A, D, B, C).

After the determination of priority order among all entity parameters, the next step is assignment of weights. For this, a separate performance metric is considered to give relative scores to all entity parameters. As a start, all entities are placed in rank order table according to their set priorities which are shown in Table 1.

The rank order set in Table 1 is now used for pair-wise comparisons between entity parameters placed in consecutive ranks to give them relative scores. The score assignment to each parameter is done based on the calculated value of total data throughput at base station when CH selection takes place under independent effect of that entity parameter. These comparisons are shown in Table 2, wherein for four entity parameters three consecutive pair-wise comparisons are made. For each pair-wise comparison, the score assignment under the remaining entity parameters is

given as “N/A” for not applicable. Let the throughput given under the entity parameter having the higher rank is T_h , which is normalized to a value of 10.

Let the throughput given under the lower ranked entity is T_l , then the score of this parameter is the normalized value relative to 10 as $(T_l/T_h) \times 10$. The results from studies conducted during determination of priority order clearly indicates that while network lifetime increases with higher priority, greater lifetime induces consistent enhanced network throughput as well. This is directly attributed to longer stable period of the network during data transmission and efficient energy utilization which causes nodes to survive longer and hence participate in more rounds of transmission. Thus, the value for T_h comes out to be greater than T_l .

As an example, comparison in (AwithD), if throughput under entity A is $T_h = 150$, which is normalized to give a score as $S(A) = 10$ and throughput under entity D is $T_l = 135$, then the score assigned to entity D relative to entity A is $(135/150) \times 10 = 9$. After this, subsequent comparisons occur in a similar manner until the last entity. The comparisons give each entity a relative score. Next, final score ratios for these parameters is calculated and are shown in Table 3.

Let score of the preceding parameter relative to highest priority parameter be S_H and score of current parameters be S_C . Then, the value of score ratio is given by formula $(S_C/10) \times S_H$.

This score ratio for the highest ranked parameter A remains same as the normalized throughput score value taken in previous step, i.e., 10, whereas all parameters that follow it in rank get a score relative to this normalized value. The score ratio value for D is 9, same as its score in the previous step, since this was already calculated relative to A's score of 10. However, for the next two parameters B and C, since scores were not directly compared with A, but with preceding parameters, their scores are relative to A only in a cascading manner. Final score ratios for these parameters are calculated further. In Table 2, for B, the score $S(B)$ is 9.3 relative to D and D has a score $S(D)$ of 9, relative to A. In Table 3 score ratio for B relative to A is given by $(9.3/10) \times 9 = 8.37$. Similarly, for C the value is computed as $(8/10) \times 8.37 = 7$.

Next, the considered step is to generate score-bounded values which are shown in Table 3. To reduce any anomalies within the process for weight calculations, a background study through basic implementations has been done to understand an optimum window of difference in score ratios between the highest and lowest priority entity parameters. A considerably large difference would mean greater emphasis or excessive dependency upon a particular parameter and a low difference would mean disregarding the priority order and its positive impact on network performance. It

Table 3 Weights computation

Objective	Parameter			
	A	D	B	C
Score Ratio	10	9	8.37	7
Score-bounded Values	10	9.33	8.91	8
Weight	$\alpha = 0.276$	$\delta = 0.25$	$\beta = 0.24$	$\gamma = 0.22$

is concluded that the former case of excessive dependency turns up above a 20% difference value and protocol results show inconsistency, start to vary with network model and computation method of high priority parameter. Similarly, a lower difference wherein the priority effects start to wear off are observed below the 20% mark. Therefore, for the purpose of simulation of proposed protocol, a cap has been set on score ratio difference at 20%. This also ensures reducing any anomalies that might have arisen within the weight calculation process. Score bounds have been set on the highest and lowest entity parameters, such that these have an optimum 20% difference. Thus, the lower end on score ratio which was 7 has been incremented by 1 and set to 8 to reduce the range between highest and lowest score ratios. Once the bounds are set as score-bounded values for A and C , new score-bounded values for D and B need to be generated. Let initial score ratio range be R_i and the score ratio for any parameter (X) be s_i . Let new score ratio range, i.e., difference set between highest and lowest score ratio be R_b and highest score ratio be S_H . Then, new score-bounded value of the parameter (X) is given by

$$S(X) = \left(\frac{S_H - s_i}{R_i} \right) \times R_b \quad (15)$$

For example, when the score ratio range is initially $R_i = (10 - 7) = 3$, D has a score ratio $s_i = 9$, which is at a distinct interval of 1 from highest score ratio of $S_H = 10$ and at an interval of 2 from lowest of 7. New intervals for when the score ratios are bounded as $R_b = (10 - 8) = 2$ are now evaluated. The new score bounded value for D would be at an interval of $[(1/3) \times 2]$ from 10 and $[(2/3) \times 2]$ from 8. This value comes to be 9.33. Similarly, for B , the score ratio is 8.37 when range is 3 and intervals are 1.63 and 1.37 from highest and lowest score ratios respectively. The new score bounded value for B would be thus at $[(1.63/3) \times 2]$ from 10 and $[(1.37/3) \times 2]$ from 8. This value comes to be 8.91. Thus, score-bounded values have also been computed for all entity parameters in Table 3.

In the last step, weight values are finally calculated based on above score-bounded values. The sum of these score-bounded values is calculated, and final weights are determined by calculating the percentage contribution of each entity parameter to this sum. As an example, the sum is 36.24 and score-bounded value for D is 9.33 then weight for D is the ratio of 9.33 and 36.24, which is 0.25. Similarly, the weights for all other parameters are also determined in Table 3. Thus, the set of weights calculated as part of the simulation give, $\alpha = 0.276$, $\beta = 0.24$, $\gamma = 0.22$, and $\delta = 0.25$.

After the determination of weighted scalarization function value and computation of parametric weights, the F_{value} for all eligible nodes eligible to be cluster heads is computed. An optimal number of cluster heads (k_{opt}) from those having the minimum value or least cost according to F_{value} are selected. The next step is to find appropriate cluster member nodes which along with the cluster heads would complete the clusters within the network for a single round of data transmission. Each non-cluster head node is assigned as cluster member in the cluster of the closest existing cluster head that lies within its transmission range. After all cluster assignments are complete, the relay of data from the members to cluster heads and from cluster heads to the base station is done.

The cluster head energy dissipation is the sum of energy required for transmission to base station, energy required to receive data from cluster members and data aggregation energy, while energy dissipation of a cluster member node is that required for data transmission to cluster head. According to this, energy values are subtracted from the appropriate nodes, i.e., cluster head or member node's current energy value for the current round of transmission. The network setup phase of the overall algorithm is shown in Algorithm 1.

Algorithm Network Setup ()
<pre> { Set total nodes as N; Distribute N nodes over network region M and set their coordinates as (X_n, Y_n); Initialize node energy as $((1 + \mu_i) \times E_0)$; Set constants E_{elec}, e_{fs} for Equations 3, 4; Set optimal clusters k_{opt} as in Equation 5; Calculate degree of node x_i as array of neighbourhood nodes for each x_i; Set maximum protocol rounds as r_{max}; } </pre>

Algorithm 1: Algorithm to initialise the network

Next, the network moves into the cluster head selection phase using weighted scalarization function with weight assignment technique. The algorithm for this phase is shown in Algorithm 2.

Algorithm Cluster Head Selection ()
<pre> { Do { For each node (x_i) { If ($E_{res} > 0$) then do { Calculate $P^l(r)$ for CH eligibility as in Equation 8; Compute D_{BS}, CH_{Life}, $M_{AvgLife}$ and P_{Max} as in Equations 11, 12 and 14; Set weights as shown in Tables 1, 2 and 3; Calculate F_{value} as per Equation 9; } } Select k_{opt} nodes as CH from all nodes arranged in increasing order of F_{value}; } While (round $\leq r_{max}$ OR nodes dead) } } </pre>

Algorithm 2: Algorithm for selection of cluster heads in each round of the network

Finally, the associations of cluster head with member nodes are performed and the data transmission phase in the network take place as shown in Algorithm 3.

```

Algorithm Data Transmission ( )
{
Do
{
For each non-CH node ( $x_i$ )
{ If ( $E_{Res} > 0$ ) then do {
 $\forall$  selected CH, compute Euclidean distance  $D_{x_i-CH}$ ;
Assign ( $x_i$ ) to CH having least  $D_{x_i-CH}$  value;}}
For each CH node
{ Compute  $E_{Trans}$  for transmission to BS;
Compute  $E_{DA}$  for data aggregation;
Compute  $E_{Rec}$  for reception from cluster members;
 $E_{Res} = E_{Res} - (E_{Trans} + E_{DA} + E_{Rec});$ 
For each non-CH/cluster member
{Compute  $E_{Trans}$  for transmission to CH;
 $E_{Res} = E_{Res} - E_{Trans};$  }
} While (round  $\leq r_{max}$  OR nodes dead )
}
}
    
```

Algorithm 3: Algorithm for estimation of energy dissipation by the network in each round

These rounds exhaust either when the initially set maximum number of rounds is complete or when the nodes of the network all exhaust, whichever occurs first. Figure 2 presents the overall flow of the proposed work covering both the execution phases.

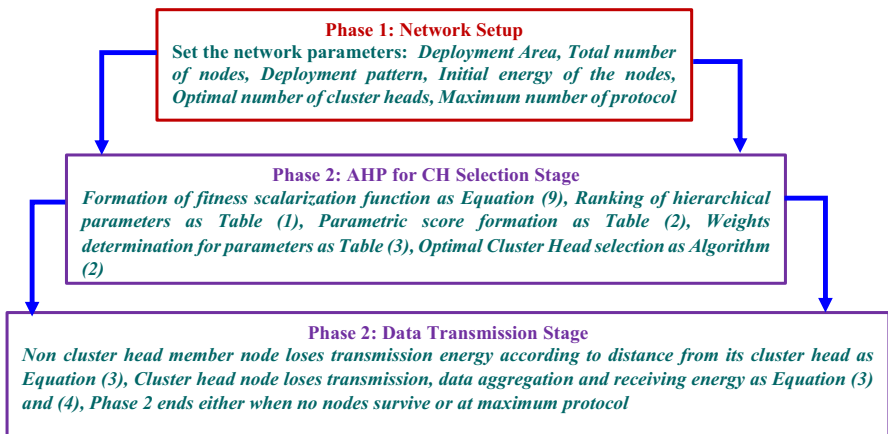


Fig. 2 Flow of the proposed work

4 Network simulation and results

The performance of the algorithm proposed in the previous section has been simulated and evaluated using MATLAB software. The nodes in the network have been deployed randomly in a square field. The base station is positioned at the geometric center of the network field.

The effectiveness and overall performance of the proposed algorithm is analyzed by the performance metrics of network lifetime, net data received at base station and system energy dissipation.

The results of the proposed algorithm in terms of network lifetime, data transmitted to BS and cumulative system energy are then compared with preexisting algorithms such as LEACH [18], DEEC [25] and multi-parameter-based heterogeneous clustering algorithm [44]. These performance comparisons have been represented using plots.

The network parameters that have been used during the simulation and their values are listed in Table 4. These parameters have been defined within the implementation in the setup phase. This part is represented in Algorithm 1: Algorithm to initialize the network in Sect. 4.2. Further the simulation of cluster head selection and data transmission phases has been implemented as per algorithms in Algorithm 2: Algorithm for selection of cluster heads in each round of the network and Algorithm 3: Algorithm for estimation of energy dissipation by the network in each round. The simulation maintains the values of specific node parameters and characteristics for each node within the network within an array, which gets updated with each round of execution of the protocol and its phases. The data for node parameters are thus obtained and are used to plot the results in a graphical representation.

The simulation has been carried out in two distinct energy-based heterogeneous environments. The energy of a node x_i is given by $(\mu_i + 1) \times E_o$. The first case is of a two-level heterogeneity where nodes have been classified as normal and advanced nodes, each on a distinct energy level. The normal nodes have energy E_o as $\mu_i = 0$ and advanced nodes have energy $2 \times E_o$ as $\mu_i = 1$. The second case is of a multi-level heterogeneity where nodes belong to varying energy levels. Normal nodes

Table 4 Network simulation parameters

Parameter name	Values
Initial energy of sensor nodes (E_o)	0.25 J
Heterogeneity factor (μ_i)	0—normal nodes 1—two-level advanced node $1 < \mu_i < 2$ - multi-level advanced node
Number of nodes in Network	100
Network area	$(100 \times 100)m^2$
Rounds (r_{max})	100–1000
E_{Elec}	50 n J/bit
Free Space loss (e_{fs})	10 p J/bit/ m^2

have initial energy E_o as $\mu_i = 0$ and advanced nodes have varying energy values as $1 < \mu_i < 2$.

4.1 Case-1: two level heterogeneity

The two-level heterogeneity simulation has been done by setting up the same node energy model for all the protocol implementations. LEACH does not take into consideration the varying energy levels that come up later during transmission and keeps a fixed probability value for the selection of cluster head. Although conventionally starting with equal energy nodes, however for the purpose of standardized comparison, the initial energy two-level heterogeneity has been setup for its simulation. This does not affect the working of the protocol itself. The results for Case-1 are shown in Figs. 3, 4, 5 and 6.

Figure 3 shows the network lifetime performance for mentioned protocol simulations. The cluster head selection process followed by each of the simulations plays the most significant role in determining network lifetime performance. LEACH has a steep decline in node population which is the earliest among all. Heterogeneous clustering algorithm has a slow and steady decline until it exhausts around Round = 1000. It considers multi-parameter heterogeneity for selection of cluster heads over a fixed probability-based process followed by LEACH which accounts for a more stable decline in node population. Next, DEEC uses individual node epochs for cluster head selection based on energy which controls a consistent energy dissipation pattern and thus performs better.

The result shows a significant improvement in prolonging network lifetime for the proposed algorithm and efficient dissipation of energy in terms of a longer stability period. This is because proposed cluster head selection process uses a unique individual probability value for each node in order to shortlist eligibility for cluster

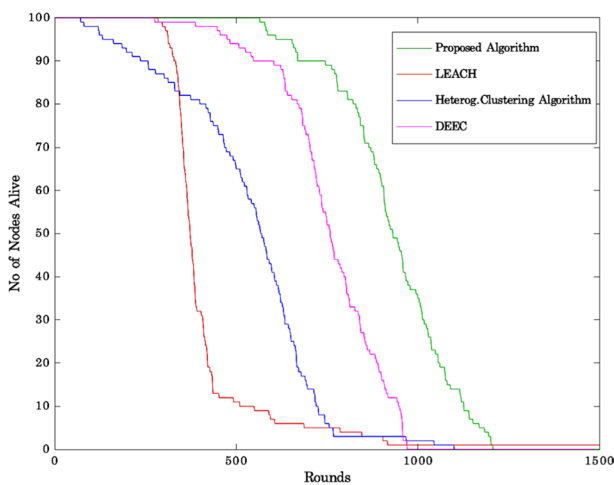


Fig. 3 Comparison of network lifetime

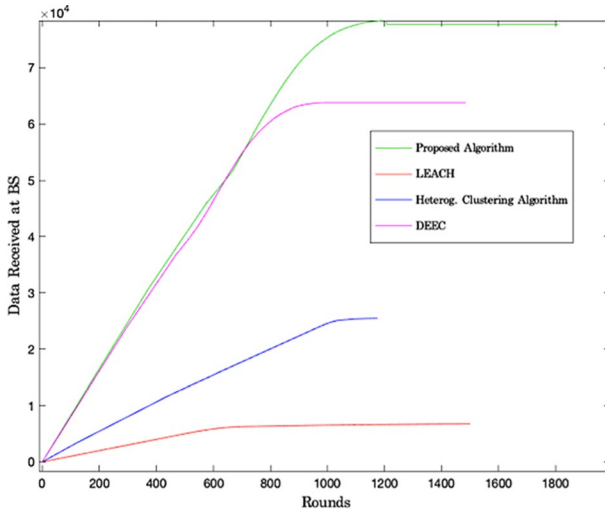


Fig. 4 Comparison of data received at BS

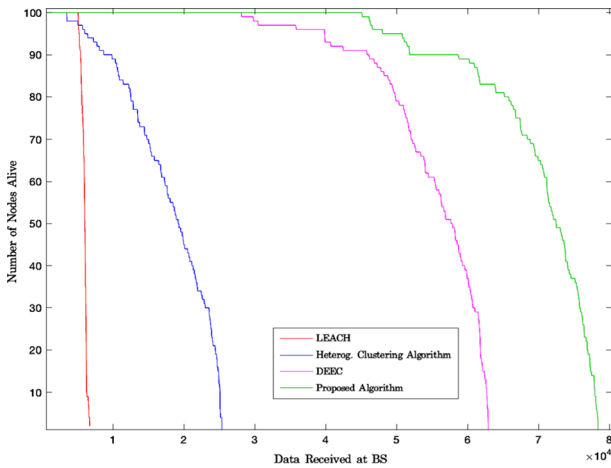


Fig. 5 Comparison of lifetime and data

heads after which a multi-parameter-weighted scalarization function is used to finalize the selection based on function value.

This ensures a more optimal two-level cluster head selection process. As a direct improvement in implementing the scalarization function, a novel weight selection process is also introduced and utilized which increases certainty of optimality in the selection process.

Figure 4 gives the data received at base station in case of mentioned protocols. LEACH which has the quickest node death and decline has limited its net data at base station, lower than the rest of the protocols. It has stable increase in

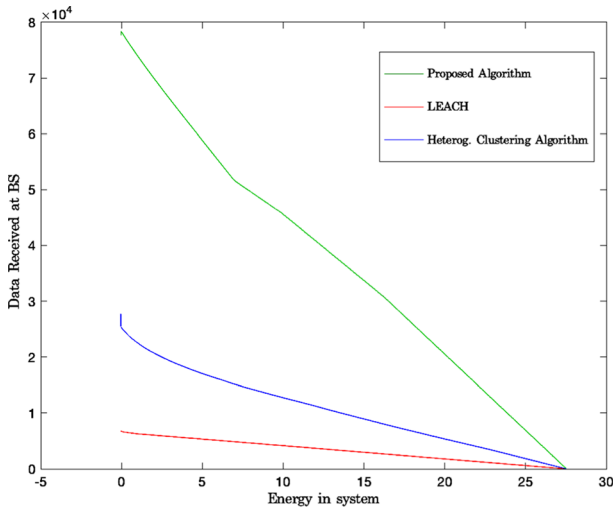


Fig. 6 Comparison of data and system energy

data transfer till Round = 600 around which it has minimal nodes left alive and hence data at base station become static and constant. Heterogeneous clustering algorithm has steady rise and higher net-data transfer until Round = 1000, which is around when the network energy exhausts and data stagnates as a constant. DEEC and proposed algorithm have similar rate of data transfer until just before Round = 800. This is when the steady node death for DEEC starts to arise, and proposed algorithm has extended stability period beyond this round. The data rate is similar until this point due to the upper hand of energy calculation and probability shortlisting which is initially crucial during CH selection. Using weighted scalarization function helps in more optimal selection and extending the stability period. There is hence a notable boost in total data transmission.

Figure 5 gives the network lifetime against data at BS. This helps to simultaneously understand how extension of lifetime directly affects and increases the data throughput as well after these have been individually compared in terms of protocol functioning previously. LEACH can be seen to have steadiest node death decline curve with the least data received at base station, and the proposed algorithm is shown to have largest stability period as well as net data transmitted at base station.

Figure 6 gives data at base station against system energy. It can be observed from the plot that for each algorithm, the overall initial network energy value is equal as per the defined node parameters. Normal nodes out of a total of 100 nodes have 0.25 J energy and the advanced nodes have 0.5 J energy. As this network energy decreases to reach zero joule, there is a large variation in total data received at the base station for each protocol.

Numerically, the value for proposed algorithm is up to three times when compared to the heterogeneous clustering algorithm. This plot points out it is important to understand that while all algorithms start with the same network energy, the net

data received at base station in case of preexisting algorithms lags that of the proposed algorithm.

This clarifies energy dissipation pattern for the protocols wherein starting from a common value of initial network energy, as the protocol reaches exhaustion, and net energy of zero joule, the individual data received at base station gives the performance measure. This shows the impact of the choice in protocol which directly affects efficiency of a network with any given initial energy.

4.2 Case-2: multi-level heterogeneity

The multi-level heterogeneity simulation has been carried out by setting up varying energy levels within nodes of the network for all the protocol implementations. On comparing network lifetimes as obtained in Fig. 7, the proposed algorithm in the multilevel heterogeneity environment does not perform as well as it was for a two-level heterogeneity environment.

This can be attributed to the fact that proposed algorithm considers multiple parameters in CH selection; however, a high variability in energy alone as a parameter, skews the weighted selection process toward a higher dependency on initial energy values. The data transmitted to base station shows slightly better performance than the two-level heterogeneity environment as shown in Fig. 8.

This is because due to higher variability in energy levels, some remaining nodes with high energy survive toward the end which can transfer more data as compared to two-level heterogeneity.

In Fig. 9, comparison is performed of the data received at base station with all simulations beginning with the same system energy. This figure is an extension of Fig. 6, wherein the multi-level proposed algorithm has been added to the plot in order to draw a comparison between the versions of heterogeneity.

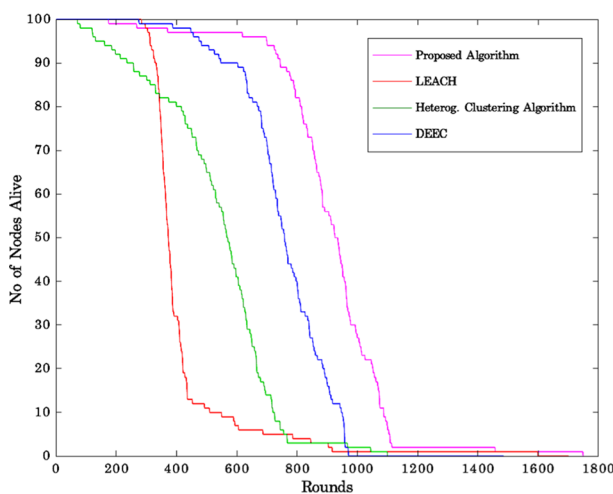


Fig. 7 Comparison of network lifetime

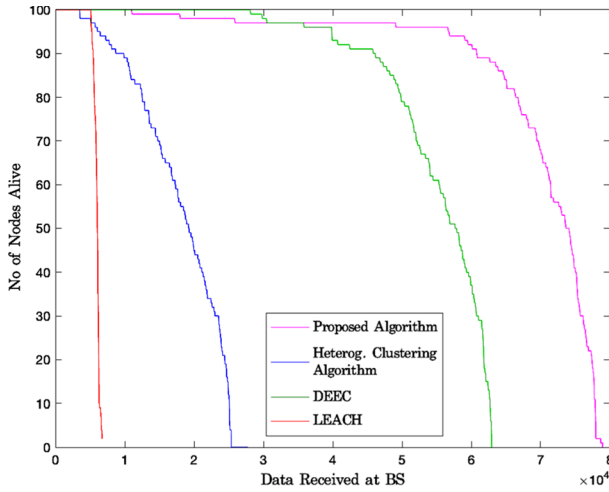


Fig. 8 Comparison of lifetime and data

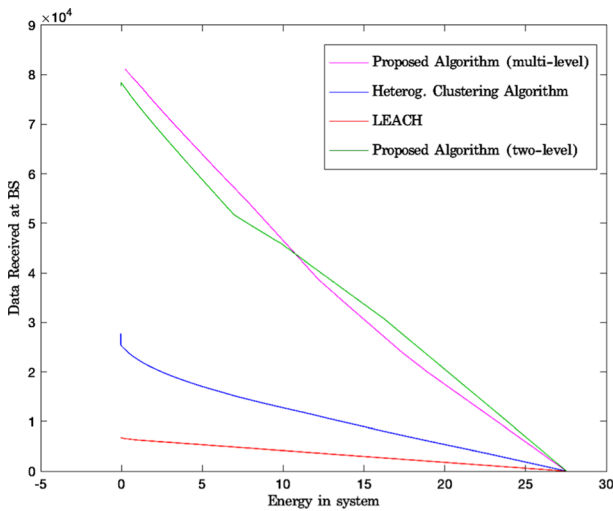


Fig. 9 Comparison of data and system energy

In this case as well, the initial energy of the network for all protocols in comparison is equal, and as the energy dissipates to zero joule for the network, each protocol can transfer some net data value at the base station. The net data for multi-level heterogeneity simulation first lags when compared to two-level heterogeneity as nodes decline at a steeper rate for the former. After a certain point, at a lower energy level after less than half of system energy remains, multi-level heterogeneity surpasses the throughput of the two-level heterogeneity simulation. Thus, it surpasses the improvement in thrice the net data in case of heterogeneous clustering

algorithm which was seen in case of two-level heterogeneity. This is justified by multiple energy levels of which certain high energy nodes survive to transfer greater data toward the end. Overall, both the variations in the simulation provide a more efficient energy dissipation, increase in network lifetime as well as throughput when compared to preexisting protocols with which comparisons have been drawn.

5 Conclusion

The presented paper proposes an algorithm for optimization of the clustering process in weighted scalarization-based clustering algorithms in wireless sensor networks, focusing on crucial WSN challenges- prolonging network lifetime and enhancing throughput.

A model for cluster head selection process is proposed that computes distributed and unique node probability for each node within the network individually, used for shortlisting eligible cluster head nodes. This is the first level of selection that is applied on the node set. This is followed by computation of a weighted scalarization function that quantifiably measures and allocates a value to each eligible node. This function is multi-parameter-based wherein the parameters used are distance to base station, expected lifetime of CH, average of lifetimes of cluster member nodes and maximum power consumed by a member node. An optimal number of cluster heads with least value of the function are chosen as cluster heads.

Within this process, a novel weight selection procedure to be utilized by the weighted scalarization function is also introduced based on analytical hierarchy process and calculates quite accurately the optimal weight values for node parameters. This is one of the main reasons for the enhanced network lifetime and throughput results for the proposed algorithm as it tackles a pervasive problem of prioritized manual weight selection in weighted scalarization function through this process. The results obtained in simulation show greater throughput at base station, with an increase of close to 30% and an increase of 20% in the network lifetime, as measured by last node death. This is in comparison with the considered base protocol of DEEC and other protocols based on similar implementation. However, due to incorporation of node characteristic heterogeneity as well as scalarization weight management, the efficiency of the proposed algorithm in terms of energy as well as data is observed.

The presented work shows the ability of multi-parameter optimization techniques using weight selection processes to have significant impact on enhancement of wireless sensor network lifetime and energy utilization. However, in this study only four parameters have been considered for node heterogeneity. It is expected that an even more thorough analysis for heterogeneity can be done by considering a wider set of node heterogeneity parameters. A future scope for this study includes determination of a universal function that can be used for calculating fitness of nodes by considering multiple node heterogeneity parameters. AHP uses certain network performance indicators for determining parameter preference and weight assignment. Thus, it has a dependency on the accuracy of measurement of these performance metrics as well as the choice of the metric itself.

As a future scope, prediction models for calculating optimal weights might prove to be an advancement to the proposed work. Therefore, it is suspected that there is a scope to use newer techniques and study their application on weighted scalarization function-based clustering algorithms. Further, parameters based on varying node characteristics such as memory and power may be involved in the multi-parametric optimization and may be extended in this manner in future work.

References

1. Yick J, Mukherjee B, Ghosal D (2008) Wireless sensor network survey. *Comput Netw* 52(12):2292–2330
2. Stojcev M, Stamenkovic Z, Dimitrijevic B (2019) Energy conservation and harvesting in wireless sensor networks. *Wireless Commun Mobile Comput* 2019:1–2
3. Basagni S, Naderi MY, Petrioli C, Spenza D (2013) Wireless sensor networks with energy harvesting. *Mobile Ad Hoc Netw* 1:701–736
4. George P (2015) Challenges in the design and implementation of wireless sensor networks: A holistic approach-development and planning tools, middleware, power efficiency, interoperability. In: 2015 4th Mediterranean Conference on Embedded Computing (MECO), IEEE. pp 1–3
5. Jha V, Prakash N, Mohapatra AK (2019) Energy efficient model for recovery from multiple nodes failure in wireless sensor networks. *Wireless Pers Commun* 108(3):1459–1479
6. Ullah I, Youn HY (2020) Efficient data aggregation with node clustering and extreme learning machine for WSN. *J Supercomput* 76(12):10009–10035
7. Priyadarshi R, Gupta B, Anurag A (2020) Deployment techniques in wireless sensor networks: a survey, classification, challenges, and future research issues. *J Supercomput* 76(9):7333–7373
8. Amezzane I, Fakhri Y, El Aroussi M, Bakhouya M (2018) GPU and data stream learning approaches for online smartphone-based human activity recognition. In: Proceedings of the 3rd International Conference on Smart City Applications, pp. 1–7
9. Lee CC (2020) Security and privacy in wireless sensor networks: advances and challenges. *Sensors* 20(3):744
10. Jha V, Verma S, Prakash N, Gupta G (2018) Corona based optimal node deployment distribution in wireless sensor networks. *Wireless Pers Commun* 102(1):325–354
11. Mehmood A, Mauri JL, Noman M, Song H (2015) Improvement of the wireless sensor network lifetime using LEACH with vice-cluster head. *Ad Hoc Sens Wirel Networks* 28(1–2):1–17
12. Jan B, Farman H, Javed H, Montrucchio B, Khan M, Ali S (2017) Energy efficient hierarchical clustering approaches in wireless sensor networks: a survey. *Wirel Commun Mob Comput* 2017:1–14
13. Ramesh K, Somasundaram DK (2011) A comparative study of cluster head selection algorithms in wireless sensor networks. *Int J Comput Sci Eng Surv* 2(4):153–164
14. Wu C-H, Chung Y-C (2007) Heterogeneous wireless sensor network deployment and topology control based on irregular sensor model. *Adv Grid Pervasive Comput Lect Notes Comput Sci* 4459:78–88
15. Sharma D, Ojha A, Bhondekar AP (2018) Heterogeneity consideration in wireless sensor networks routing algorithms: a review. *J Supercomput* 75(5):2341–2394
16. Fei Z, Li B, Yang S, Xing C, Chen H, Hanzo L (2017) A survey of multi-objective optimization in wireless sensor networks: metrics, algorithms, and open problems. *IEEE Commun Surv Tutor* 19(1):550–586
17. Yang X-S (2014) Multi-Objective Optimization. *Nature-Inspired Optimization Algorithms*, 197–211
18. Heinzelman W, Chandrakasan A, Balakrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans Wireless Commun* 1(4):660–670
19. Tong M, Tang M (2010) LEACH-B: an improved LEACH protocol for wireless sensor network. In: 2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), IEEE. pp 1–4

20. Abdellah E, Benalla SAID, Hssane AB, Hasnaoui ML (2010) Advanced low energy adaptive clustering hierarchy. *Int J Comput Sci Eng* 2(07):2491–2497
21. Beiranvand Z, Patooghy A, Fazeli M (2013) "I-LEACH: an efficient routing algorithm to improve performance & to reduce energy consumption in Wireless Sensor Networks." In: The 5th Conference on Information and Knowledge Technology, IEEE. pp 13–18
22. Smaragdakis G, Matta I, Bestavros A (2004) "SEP: a stable election protocol for clustered heterogeneous wireless sensor networks." In: Second international workshop on sensor and actor network protocols and applications (SANPA 2004), vol 3
23. Faisal S, Javaid N, Javaid A, Khan MA, Bouk SH, Khan ZA (2013) "Z-SEP: zonal-stable election protocol for wireless sensor networks". *J Basic Appl Sci Res*, 3
24. Nurlan Z et al (2021) EZ-SEP: extended Z-SEP routing protocol with hierarchical clustering approach for wireless heterogeneous sensor network. *Sensors* 21(4):1021
25. Qing L et al (2006) Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. *Comput Commun* 29(12):2230–2237
26. Jarchlo EA, Bazlamaçci CF (2014) "Life time sensitive weighted clustering on wireless sensor networks." In *SENSORNETS*, pp 41–51
27. Chatterjee M, Das SK, Turgut D (2002) WCA: a weighted clustering algorithm for mobile ad hoc networks. *Clust Comput* 5:193–204
28. Yang H, Li Z (2015) "A clustering optimization algorithm based on WCA in MANET," 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp 1–5
29. Aissa M, Belghith A, Drira K (2013) New strategies and extensions in weighted clustering algorithms for mobile ad hoc networks. *Procedia Comput Sci* 19:297–304
30. Chlamtac I et al (2003) Mobile ad hoc networking: imperatives and challenges. *Ad Hoc Netw* 1(1):13–64
31. Ouchitachen H, Hair A, Idrissi N (2017) Improved multi-objective weighted clustering algorithm in wireless sensor network. *Egypt Inform J* 18(1):45–54
32. Essa A, Al-Dubai AY, Romdhani I, Esriaftri MA (2017) "A new weight based rotating clustering scheme for WSNS," 2017 International Symposium on Networks, Computers and Communications (ISNCC), pp 1–6
33. Amine D, Nasr-Eddine B, Abdelhamid L (2015) A distributed and safe weighted clustering algorithm for mobile wireless sensor networks. *Procedia Comput Sci* 52:641–646
34. Babu NV, Boregowda B, Puttamadappa C, Davanakatti SS (2012) An optimized weight based clustering algorithm in heterogeneous wireless sensor networks. *Comput Sci Inf Technol* 2:185–195
35. Ding P, Holliday J, Celik A (2005) Distributed energy-efficient hierarchical clustering for wireless sensor networks. In: *International Conference on Distributed Computing in Sensor Systems*, pp 322–339
36. Chand S, Singh S, Kumar B (2014) Heterogeneous HEED protocol for wireless sensor networks. *Wireless Pers Commun* 77(3):2117–2139
37. Zeb A, Islam AKMM, Zareei M, Mamoona IA, Mansoor N, Baharun S, Katayama Y, Komaki S (2016) Clustering analysis in wireless sensor networks: the ambit of performance metrics and schemes taxonomy. *Int J Distrib Sens Netw* 12(7):4979142
38. Prasai K, Ghimire S (2016) Selection of Weighting Factors in Weighted Clustering Algorithm in MANET, 2016 Proceedings of the World Congress on Engineering and Computer Science, 1
39. Gunantara N (2018) A review of multi-objective optimization: methods and its applications. *Cogent Eng* 5(1):1502242
40. Liu Z, Mou X, Liu H-C, Zhang L (2021) Failure mode and effect analysis based on probabilistic linguistic preference relations and gained and lost dominance score method. *IEEE Trans Cybernet*. <https://doi.org/10.1109/TCYB.2021.3105742>
41. Liang X, Wu X, Liao H (2020) A gained and lost dominance score II method for modelling group uncertainty: case study of site selection of electric vehicle charging stations. *J Clean Prod* 262:121239
42. Song B, Kang S (2016) A method of assigning weights using a ranking and non hierarchy comparison. *Adv Decis Sci* 2016:1–9
43. Liu HC, Chen XQ, You JX, Li Z (2021) A new integrated approach for risk evaluation and classification with dynamic expert weights. *IEEE Trans Reliab* 70(1):163–174

44. Yadav AK, Rajpoot P, Kumar P, Dubey K, Singh SH, Verma KR (2019) Multi parameters based heterogeneous clustering algorithm for energy optimization in WSN. In: 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE. pp 587–592
45. Mishra R, Jha V, Tripathi RK, Sharma AK (2020) Corona based node distribution scheme targeting energy balancing in wireless sensor networks for the sensors having limited sensing range. *Wireless Netw* 26(2):879–896

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Vivekanand Jha¹  · Rashika Sharma¹

✉ Vivekanand Jha
vivekanandjha@igdtuw.ac.in

Rashika Sharma
09rs01@gmail.com

¹ Indira Gandhi Delhi Technical University for Women, Kashmere Gate, New Delhi, Delhi 110006, India