A Heuristic Crossover Enhanced Evolutionary Algorithm for Clustering Wireless Sensor Network

Muyiwa Olakanmi Oladimeji, Mikdam Turkey, and Sandra Dudley (MIEEE)

School of Engineering, London South Bank University
103 Borough Road, London SE1 0AA
`{oladimm2,turkeym,dudleyms}@lsbu.ac.uk`

Abstract. In this paper, a Heuristic-Crossover Enhanced Evolutionary Algorithm for Cluster Head Selection is proposed. The algorithm uses a novel heuristic crossover operator to combine two different solutions in order to achieve a high quality solution that distributes the energy load evenly among the sensor nodes and enhances the distribution of cluster head nodes in a network. Additionally, we propose the Stochastic Selection of Inactive Nodes, a mechanism inspired by the Boltzmann Selection process in genetic algorithms. This mechanism stochastically considers coverage effect in the selection of nodes that are required to go into sleep mode in order to conserve energy of sensor nodes. The proposed selection of inactive node mechanisms and cluster head selections protocol are performed sequentially at every round and are part of the main algorithm proposed, namely the Heuristic Algorithm for Clustering Hierarchy (`HACH`). The main goal of `HACH` is to extend network lifetime of wireless sensor networks by reducing and balancing the energy consumption among sensor nodes during communication processes. Our protocol shows improved performance compared with state-of-the-art protocols like LEACH, TCAC and SEECH in terms of improved network lifetime for wireless sensor networks deployments.

Keywords: Evolutionary Algorithm, Heuristic Crossover, Wireless Sensors Networks, Clustering, Network Lifetime.

1   Introduction

Recent advances in Micro-Electro-Mechanical Systems (MEMS) have led to the development of autonomous and self-configurable wireless sensor nodes that are densely deployed throughout a spatial region [1]. These sensor nodes have the capability to sense any event or abnormal environmental conditions such as motion, moisture, heat, smoke, pressure etc. in form of data. A large number of these sensors can be networked and deployed in remote and inaccessible areas, hence providing a wireless sensor network (WSN) connectivity. WSNs are usually composed of small, inexpensive and battery-powered wireless sensor nodes. In fact, WSNs have contributed tremendously to a number of military and civil applications such as target field imaging, event detection, weather monitoring, security

and tactical surveillance. However, sensor nodes are constrained by energy supply and bandwidth. Such constraints combined with a typical deployment of large number of sensor nodes pose many challenges to the design and management of WSNs and necessitate energy-awareness at all layers of the networking protocol stack [2]. Thus, *innovative techniques that eliminate energy inefficiencies that would shorten the lifetime of the network are required.* In this work, we present a solution that balances the consuming energy among sensor nodes in order to prolong the network lifetime. The energy balance is done using two mechanisms; firstly, a genetic algorithm (GA) for cluster heads (CHs) selection was used to ensure that well distributed nodes with higher energy will be selected as CHs. Secondly, a mechanism inspired by Boltzmann selection was utilized to select nodes to send into sleep mode without causing an adverse effect on the coverage.

Most existing routing protocols designed to tackle the above challenges are broadly classified into two classes, namely flat and hierarchical. Flat protocols include the old-fashion Direct Transmission (DT) and Minimum Transmission Energy (MTE), which cannot promise a balanced energy distribution among sensors in a WSN. The drawback of the MTE is that a remote sensor uses a relay sensor for data transmission to the sink and this makes the relay sensor the first to die. In the DT, sensors communicate directly with the sink and this causes the remote sensor to die first. Therefore designing energy-efficient clustering protocols becomes a major factor for sensor lifetime extension. Generally, clustering protocols can outperform flat protocols in balancing energy consumption and network lifetime prolongation by adopting data aggregation mechanisms [3, 4]. Theoretically, there are three types of nodes: the cluster-head (CH), member node (MN) and sink node (SN). The member node is responsible for sensing the raw data and employs Time Domain Multiple Access (TDMA) scheduling to send the raw data to the cluster head. The main role of the CH is to aggregate data received from MN and thereby forwards the aggregated data to the SN through single-hop or multi-hop. CHs selection can either be done by the sensors themselves, by the SN or pre-determined by the wireless network designer. In this paper, CH selection is done by the SN because the SN has enough energy and can perform complex calculations. The CH selection problem can be viewed as an optimisation problem where the methods have used GA to solve. From that end, we define an objective function to access the individual solution and propose a novel heuristic crossover that is enhanced by the knowledge of our problem.

In this paper, a Heuristic Algorithm for Clustering Hierarchy (HACH) protocol has been developed and it has two mechanisms namely the Heuristic-Crossover Enhanced Evolutionary Algorithm for Cluster Head Selection (HEECHS) and Stochastic Selection of Inactive Nodes (SSIN) mechanisms. HEECHS exploits our knowledge of the problem to develop an effective heuristic crossover that combines genetic material in a unique way to produce improved CHs configuration. This process bears some similarity to a form of optimisation algorithm called Memetic Algorithms (MAs). MAs are a class of stochastic global search heuris-

tics in which Evolutionary Algorithms-based approaches are combined with local search techniques to improve the quality of the solutions created by evolutions [5]. The (SSIN), a mechanism inspired by Boltzmann selection process in GA was used to reduces the number of active nodes in each round by sending some nodes to sleep or inactive mode to conserve energy and extend the network lifetime without adversely affecting coverage. These two mechanisms work collaboratively to maximize network lifetime by balancing energy consumption amongst sensor nodes during the communication process. The balance in energy consumption is achieved by selecting spatially distributed nodes with higher energy as CHs and also sending some nodes into sleep mode without causing an adverse effect on coverage. Protocol presented is a more energy efficient protocol compared with other protocols that employ GA in the sense that it integrates knowledge of the problem into the GA crossover operator.

The remainder of this study is organised as follows. Section 2 discusses related work on the application of genetic algorithms in energy-efficient wireless sensor networks. Section 3 describes the network and radio model underlying the protocol presented. In Section 4 we describe our proposed algorithm under three major stages namely the proposed selection mechanism, clustering algorithm and the energy consumption calculation. Section 5 presents our experimental settings, performance measures, result and discussion. Finally, a conclusion is provided in Section 6.

2   Related Work

Recently, research interests in the area of energy-efficient wireless sensor network has tends towards energy efficient clustering protocols [6–9]. The Low-Energy Adaptive Clustering Hierarchy (LEACH) assumes that the energy of each sensor node is the same at the time of CH selection and non-CH selection. The selection process is carried out probabilistically. The role of the CH is to aggregate data received from its members within its cluster and pass the aggregated data directly to the sink. Problems with this protocol are that the location of the selected CH might be far away from the sink, so it consumes more energy when transmitting to the sink and this can lead to CH nodes to dying faster than other nodes [4]. A two-level LEACH (TL-LEACH) proposed in [10], adds another level to the cluster whereas LEACH that has only one level. This extra level reduces energy consumption most especially for CH far away from the sink. The hybrid energy efficient distributive (HEED) protocol [11] selects CHs using residual energy and the least amount of energy dissipated for communication between the CHs and non-CHs. Data are sent to the sink using a multi-hop communication approach.

In Topology-Controlled Adaptive Clustering (TCAC) protocol [12], a large number of nodes consider themselves candidates for CH nodes and inform other nodes. Each of the candidate CH nodes checks whether the other candidate CH nodes have a higher residual energy level or not. If there are none with higher residual energy, the highest announces itself as the CH. Non-CH nodes select
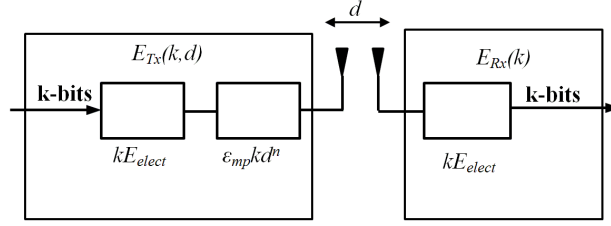
Fig. 1. Radio Energy Dissipation Model

the CH that has the minimum cost distance between itself and the CH to the sink. The TCAC balances the size of the cluster and send the data directly to the sink from the CH. In the scalable energy efficient clustering hierarchy (SEECH) protocol [13], the network nodes are divided into three layers namely the member nodes, CH nodes and relays. Clusters are formed based on centrality of the CH node with minimum intra-cluster energy distribution. A closer node to the sink in a cluster is often selected as the relay node. The relay node helps the CH node to transmit aggregated data to the sink through hop or multi-hop communication.

In [14] a genetic algorithm based energy efficient cluster (GABEEC) protocol that uses static clustering with dynamic CH selection was described. At the end of each round, an associate member node becomes a CH and this decision is based on the residual energy of the current CHs and the average energy of all members in the cluster. A Genetic algorithm approach was used to minimize the communication distance and maximize the lifetime of the network. In [15], a centralized energy-aware cluster-based protocol was discussed to extend the sensor network lifetime by using Optimization (PSO) algorithms. A new cost function that simultaneously takes into account the maximum distance between the non-CH node and its CH, and the remaining energy of CH candidates in the CH selection algorithm was defined.

3   Network and Radio Model

In the proposed HACH protocol, the following assumptions of the network model are assumed:

– The data sink is a stationary and resource-rich device that is placed far away from the sensing field.
– All sensor nodes are homogeneous in terms of energy and are stationary after deployment.
– All sensors have GPS or other location determination devices
– Nodes located close to each other have correlated data.
– Nodes are capable of acting in inactive mode or a low power sleeping mode.

Algorithm 1 `HACH` Protocol

    Let *AliveNodes* be the total number of sensor nodes
    Compute the network total coverage.
    while (*AliveNodes* > 0) **do**
        Use algorithm `SSIN` to select inactive nodes. (See Algorithm 2)
        Put selected nodes into sleep mode.
        Apply the proposed `HEECHS` algorithm for CHs configuration. (See Algorithm 3)
        Compute the energy values of E_{CH}, E_{Mem} and E_{Res}. (refer to Section 4.3)
        Find out the number of dead node (node with energy equal or less than 0).
        Update AliveNodes
    end while

– The communication channel is symmetric (i.e., the energy required to transmit a message from sensor node s1 to sensor node s2 is the same as energy required to transmit a message from node s2 to node s1 for a given signal to noise ratio).

For fair comparison with previous protocols [4, 16, 17], we assume the simple model shown in Figure (1) for the radio hardware energy dissipation where the transmitter dissipates energy $E_{Tx}(k, d)$ to run the radio electronics and the power amplifier, and the receiver dissipates energy $E_{Rx}(k)$ to run the radio electronics. For the experiments described here, both the free space (d^2 power loss) and the multipath fading (d^4 power loss) channel models were used, depending on the distance (d) between the transmitter and receiver. The power amplifier is appropriately set in such a way that if the distance is less than a threshold distance, the free space (fs) model is used; otherwise, the multipath (mp) model is used. Thus, to transmit a k-bit message at distance d, the radio expends:

$$E_{Tx}(k, d) = \begin{cases} kE_{elect} + \varepsilon_{mp}kd^4, & \text{if } d > d_0 \\ kE_{elect} + \varepsilon_{fs}kd^2, & \text{if } d < d_0 \end{cases} \tag{1}$$

And to receive k-bit message, the radio expends:

$$E_{Rx}(k) = kEelect \tag{2}$$

where $d_0 = \sqrt{\varepsilon_{fs}/\varepsilon_{mp}}$ denotes the threshold distance and the electronics energy, E_{elect} depends on factors such as the digital coding, modulation, filtering, and spreading of the signal, whereas the amplifier energy, ε_{mp} or ε_{fs} depends on the distance to the receiver and the acceptable bit-error rate.

4   The Proposed Protocol: `HACH`

This protocol performs three sequential operations: sending nodes to sleep, clustering and network operations. At the set-up phase of the algorithm, the sink transmits control packets in order to receive node information in terms of the
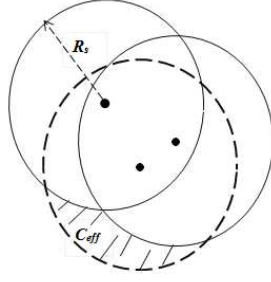
Fig. 2. Illustration of Nodes to Sleep on Coverage Area

nodes ID, location and energy. The proposed `SSIN` protocol helps to dynamically select which nodes to send to sleep by initially generating a candidate list. The candidates list is populated with nodes that have energy smaller than the average energy of all nodes. Using a stochastic process, only a small number of nodes is sent into sleep mode in the candidate list without causing an adverse on the coverage. CH selection using `HEECHS` is then performed on the nodes that are kept active. In this work, the authors formulated clustering as an optimisation problem that is best addressed using GA. The genetic operators used in this approach are the tournament selection, heuristic crossover and mutation operator. The best CH configuration that ensures balanced energy consumption across the networks is selected at every network operation round. At the end of each round, the residual energy of each node is computed. This value is then used to compute the average energy for the next round and this cycle continues until all nodes are dead as shown in Algorithm 1.

4.1   Inactive Node Selection Using `SSIN`

The `SSIN` makes decisions on which nodes to send into inactive mode at the beginning of each network operation round. The candidate list of sleeping nodes is built up by checking the residual energy of nodes that are less than the computed average energy. This inactive node selection mechanism is synonymous to the Boltzmann selection process whereby a method is adopted for controlling the selection pressure [18]. In Boltzmann selection, the temperature parameter is varied in order to control the selection pressure. In this work, the maximum coverage effect, Max_{eff} is used to control the effect of putting nodes to sleep on the WSNs, which is defined as:

$$Max_{eff} = 2 \times \pi \times R_s^2 \tag{3}$$

where R_s is the sensing range of a sensor node (taking the coverage area as a circle with radius R_s), $(pi \times R_s^2)$ signifies the coverage of one node and the value $'2'$ represents coverage of two nodes.

   The effect of putting a node to sleep based on the coverage is defined by the coverage effect C_{eff} as shown in Figure 2. The total coverage effect is calculated

---

Algorithm 2 Proposed `SSIN` protocol

$Acc_{eff} = 0$;
Compute the residual energy, E_{Res} of each node. (refer to section 4.3)
Compute the average energy of all nodes, E_{Avg}.
Generate a candidate list for nodes that satisfies the condition $E_{Res} < E_{Avg}$.
Compute Max_{eff}. (refer to equation 3)
while $(Acc_{eff} < Max_{eff})$ **do**
    Compute probability, P of adding nodes to the sleeping list. (See equation 4)
    if $(rand() < P)$ **then**
        Create list of sleeping node from the candidate list.
        Compute the coverage effect, C_{eff}.
        $Acc_{eff} = Acc_{eff} + C_{eff}$
    end if
end while

---

by invoking a matrix called the Coverage Matrix. The coverage matrix captures the overlapped areas of nodes coverage which allows the identification of nodes that can be sent into sleep mode without affecting coverage as there will be other nodes covering part of the selected node's area. The accumulated Coverage effect Acc_{eff} is defined as the total effect on the coverage as a result of allowing some nodes to sleep. Our algorithm is designed in such a way that the Acc_{eff} value is expected to be less than the Max_{eff} for optimum coverage ($Acc_{eff} < Max_{eff}$). The probability that a node will be added to the sleeping node list can be computed using:

$$P = e^{(-C_{eff}/Max_{eff})/(1-(Acc_{eff}/Max_{eff})))^2} \qquad (4)$$

A randomly generated number is compared with the computed probability, P. A candidate list of inactive node is created if the random number is less than the probability, P. The accumulated frequency Acc_{eff} is computed by adding its current value to coverage effect C_{eff} value. The operations of `SSIN` continues until the Acc_{eff} is greater than the maximum acceptable coverage effect, Max_{eff} as shown in Algorithm 2.

4.2   Cluster Heads Selection using `HEECHS`

Our clustering algorithm involves the task of CH selection and membership association. The proposed `HEECHS` protocol is developed for the CH selection task using a heuristic-based GA. Similar to conventional GA, it runs through a cycle of tasks such as the creation of population strings, evaluation of each string, selection of the best strings and reproduction to create a new population. The significant difference is that it uses a problem-dependent knowledge-based heuristic crossover to find the best CH configuration with the optimum number of well distributed CH nodes. The population, $P(t)$ comprises of individuals with binary representation of 0s or 1s. This represents CH individuals with size p_s and each individual CH has length N_s that is evaluated by computing the fitness value

using Equation 6. Individual CHs with the best fitness value are selected from two randomly selected parent population, $P_x(t)$ and $P_y(t)$ and this continues until the mating pool is filled. Our proposed heuristic crossover is applied to mate the individuals in the pool and subsequently produce a new population of $P(t+1)$. Again, the fitness value of each individual in this new population is computed using Equation 6 and the whole cycle continues until the stopping criterion is achieved. Stopping criterion is achieved when the average fitness of the population does not change further.

4.2.1   Proposed Heuristic Crossover

The core component of the proposed HEECHS is the heuristic crossover and this term simply means a unique form of crossover that uses the knowledge of a problem to combine two candidate solution in order to produce an offspring. A heuristic crossover operator was proposed by Lixin Tang [19] as an operator which utilizes the parents implicit information to produce offspring. In a conventional GA, the selected members of a population are paired, and some two parents undergo crossover to produces pairs of offspring that tends to substitute the parents. Heuristic crossover operator produces only one child from two or more parents, which differs from the canonical approach. Also in the canonical approach, there is no guarantee that an offspring produced would be better than the mating parent [20].

The proposed crossover produces individuals with CH configurations that are well distributed across the sensing field and favors the ones with higher energy. We can see from the Algorithm 3 that the proposed crossover does not allow two CHs to be selected within the same region and gives selection priority to the CH with higher energy. Two individuals are selected from the parent pool and the position of CH genes inside each individual is computed. Each individual has its own array set of CH positions, which is denoted by CH_1 and CH_2. We choose to define the threshold distance between any two neighbouring CH position as $\frac{\sqrt{(x_{max}-x_{min})^2+(y_{max}-y_{min})^2}}{n \times 0.04}$, where the (x_{min}, y_{min}) and (x_{max}, y_{max}) are minimum and maximum xy-coordinates of the sensor fields respectively, $(n \times 0.04)$ represent 4% of the total sensor nodes. The union of CH_1 and CH_2 is represented by $CH_{all} = CH_1 \cup CH_2$. By default, the first CH position $CH_{all}(1)$ in the set CH_{all} is transferred to a newly created set CH_{new}. Each subsequent CHs position in the CH_{all} is compared with the CH_{new} array set in order to ensure decisions based on distance between the CHs and their residual energy (E) as shown in Algorithm 3.

4.2.2   Proposed Objective Functions

In order to solve the CH selection problem, there is a need to develop objective functions because CH selection is dealt with as an optimisation problem. This objective function returns a fitness value which is used to access the quality of a candidate solution. The objective functions is derived by considering the parameters such as the energy of all sensor nodes and the Risk penalty R. The

---

Algorithm 3 Proposed Heuristic Crossover

Select two chromosomes from the parent pool
Compute and store the cluster head position in each chromosomes in sets CH_1 and CH_2.
Compute the threshold distance, T (refer to Section 4.2.1)
Compute the union set CH_{all}. (refer to Section 4.2.1)
Obtain the first cluster position $CH_{all}(1)$ in the set CH_{all}.
Create a new set CH_{new} and transfer the $CH_{all}(1)$ to it.
Compute the distance, D between cluster head positions in the sets CH_1 and CH_2.
while $(D < T)$ **do**
    if (Energy in CH_{all} node is less than CH_{new} node) **then**
        Discard the cluster head node. (i.e. do not add to CH_{new} set)
    end if
    Replace the cluster head node in the CH_{new} set
end while
Add to the cluster head node in the set CH_{all} into the CH_{new} set.

---

reason for considering the energy parameter of sensor nodes is to ensure that nodes with greater energy are given higher priority in CHs selection.

The Risk penalty, R for the CH selection is defined as:

$$R = \begin{cases} Lower - L, & \text{if } L < Lower \\ L - Upper, & \text{if } L > Upper \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where *Lower* and *Upper* are calculated as the 4% and 6% of total number all sensor nodes in the field (n) respectively. R imposes restrictions on the number of CHs (L).

Therefore, the objective function is computed as:

$$F(X) = w_1 * \frac{AvgENCH}{AvgECH} + w_2 * R \quad (6)$$

Where w_1 and w_2 are the weighting factors; the average energy of non-CHs and CHs is given by $AvgENCH = \frac{\sum_{i\varepsilon NCH} E_i}{n - L}$ and $AvgECH = \frac{\sum_{i\varepsilon CH} E_i}{L}$ respectively. The ratio $\frac{AvgENCH}{AvgECH}$ is assigned a higher weighting factor (w_1=0.9) than the Risk penalty, R (w_2=0.1) because much emphasis is placed upon it. (Note: CH and NCH represent the set of all CHs and non-CHs respectively).

4.2.3 Other Operators

Crossover and mutation provide exploration, compared with the exploitation provided by selection. The effectiveness of GA depends on the trade-off between exploitation and exploration [21, 22].

The *tournament selection operator* is used to randomly choose a group of individuals from the current population, and the one with the best fitness is

selected. The tournament size determines the extent of selection pressure. In this work, a tournament size of two is used in order to reduce the selection pressure. This process is repeated as often as desired (usually until the mating pool is filled).

The *mutation operator* alters one individual (parent), to produce a single new individual (child) with a probability of mutation (p_m).

The initial population pool (parents) and offspring individuals that was generated in the previous step are sorted in ascending order based on the objective function values. Then the first CH chromosomes with minimum objective function values are selected to form the population pool for the next generation.

The *stopping criterion* is met when the objective function value of the population does not change further.

4.3   Network Operations and Energy Consumption Computation

Energy consumption at each round can be explained with regards to the two phases in the algorithm such as the set-up phase and the steady state phase.

Set-up Phase The set-up phase involves the transmission and reception of control packets k_{CP} from the sink to all nodes to initialise inter- and intra-communication. In the set-up phase, the sink sends a control packet which contains a short message to wake up and request IDs, positions and energy levels of all sensor nodes in the sensor field. As in Equation 2, the energy $E_{Rx}(k_{CP})$ is dissipated when control packets are received from the sink. All sensors report their IDs, positions and energy levels back to the sink. Also, as in Equation 1, energy $E_{Tx}(k_{CP}, d)$ is consumed when transmitting control packets k_{CP} to the sink. The control packets received from all sensor nodes is processed by the sink in order to make the following vital decision; which nodes to keep active, CH selection, and the associated CH membership. Energy $E_{Rx}(k_{CP})$ is also dissipated for receiving the membership status information from the sink. The energy required for all CHs to transmit the TDMA schedules to their members can be computed using:

$$E_{Tx(ch_i)}(k_{CP}, d_{i-toMem}) = \sum_{i=1} ch_i * \begin{cases} k_{CP}E_{elect} + \varepsilon_{mp}k_{CP}d_{i-toMem}^4, & \text{if } d < d_0 \\ k_{CP}E_{elect} + \varepsilon_{fs}k_{CP}d_{i-toMem}^2, & \text{if } d > d_0 \end{cases}$$
$$(7)$$

Also the energy spent by each member node to receive the TDMA schedule from the CH can also be computed using Equation 2.

Steady Phase During the steady phase, the active sensor nodes start sending data packets k. Each node sends the sensed data to its CH according to the TDMA schedule received. The CH nodes receiver must always be ready to receive packets from its node within its cluster. Data aggregation is performed on all received data at the CH and data are converted into a single data stream. This aggregated data is transmitted from the CH to the sink. During the process the sensor node transceiver consumes energy calculated by Equation 9. The total amount of energy spent by all member nodes to transmit to their respective CHs

is computed using:

$$E_{Rx(m_i)}(k) = \sum_{i=1} m_i k E_{elec} \tag{8}$$

Where m_i denotes the member nodes, which ranges from $i = 1, 2, 3, ..., n - L$. Also, the energy dissipated by the CH for aggregating data received from all its members and itself can be calculated using:

$$E_{DA(m_i+1)}(k) = k E_{DA} * (\sum_{i=1} m_i + 1) \tag{9}$$

Finally, the amount of energy spent by the CH node to transmit to the sink is computed using:

$$E_{Tx(ch_i)}(k_{CP}, d_{i-toSink}) = \sum_{i=1} ch_i * \begin{cases} k_{CP} E_{elect} + \varepsilon_{mp} k_{CP} d_{i-toSink}^4, & \text{if } d > d_0 \\ k_{CP} E_{elect} + \varepsilon_{fs} k_{CP} d_{i-toSink}^2, & \text{if } d < d_0 \end{cases} \tag{10}$$

Total Energy Dissipated

In our algorithm, the total energy dissipated by all the CHs can be computed using:

$$E_{CHs} = 2 * E_{Rx}(k_{CP}) + E_{Tx}(k_{CP}, d_{i-toSink}) + E_{Tx}(k_{CP}, d_{i-toMem}) \\ + E_{Rx(m_1)}(k) + E_{DA(m_i+1)}(k) \tag{11}$$

And the energy dissipated by the member nodes is computed as:

$$E_{Mem} = E_{Tx}(k_{CP}, d_{i-toSink}) + E_{Tx}(k_{CP}, d_{i-toCH}) + 3 * E_{Rx}(k_{CP}) \tag{12}$$

Therefore, the overall energy dissipated by all nodes is represented by $E_{TOTAL} = E_{CHs} + E_{Mem}$. (Note: Residual energy E_{Res} of individual node at each round can be computed by subtracting the energy consumption from the current residual energy.

5   Experimental Results

In this work, we evaluate protocols from an energy efficiency perspective by examining the number of nodes alive versus rounds. With the aid of MATLAB tools, a simulation model was developed to test the proposed algorithm and graph results are generated to evaluate the lifetime of the sensor nodes. The proposed technique is scalable and it may lead to energy efficiency improvements across various network topologies. To assess this claim the performance of HACH is compared to three other protocols LEACH, TCAC and SEECH across three scenarios with different network sizes; small, medium and large. TABLE I describes the parameters of proposed scenes in details.

Table 1. Parameter values for each experiment

Experiment	Parameter Settings			
	Dimension	Sink Coordinates	Number of Sensors	Initial Energy
Experiment I	100×100	(50,175)	100	0.5J
Experiment II	100×100	(50,200)	400	0.5J
Experiment III	200×200	(50,350)	1000	1.0J

5.1 Experimental Settings

The common communication parameters used for all three experiments presented in Table 1 below are the electronic energy, E_{elect} =50 nJ/bit, free space loss ε_{fs}=10 pJ/bit/m^2, multipath loss, ε_{mp}=0.0013 pJ/bit/m^4, threshold distance, d_0=87m, data aggregation energy E_{DA}= 5 nJ/ bit/signal, packet size k=4000, and control packet size, k_{CP}=50. The GA parameters are set as population size, p_s=100 and mutation rate, p_m= 0.05.

5.2 Performance Measures

There are many metrics used to evaluate the performance of clustering protocols [19]. These measures again are used in this paper to evaluate the performance of `HACH` protocol:

1. **First Dead Node (FDN)**: This is the number of rounds after which the first node dies. It can also refer to the operational lifetime or stability period of the network.
2. **Last Dead Node (LDN)**: This is the number of rounds from the start of network operation until the last node dies.
3. **Instability Period Length (IPL)**: The round difference between the round at which the last node dies and the first node dies. (i.e. $IPL = LND - FND$).
4. **Average Energy at first node dies (AEFND)**: The average energy of all sensor nodes when the first node dies.

Clearly, the longer the stability period and the shorter the instability period are, the more reliable the clustering process of the wireless sensor network.

5.3 Results and Discussion

The values presented in this section are the averaged results obtained from 100 simulation runs. In each simulation run, a new set of sensor nodes was distributed to a network area. Experimental results shown in Figure 3 depicts the number of alive nodes after each round. For a summarized comparison FND, LND, and IPL measures belonging to the graphs are presented in Table 2. Experiment I (100 nodes) shows that the `HACH` protocol maintains the network operational lifetime

Table 2. Comparison of LEACH, TCAC, SEECH and `HACH` for FND,LND and IPL

Experiment	Protocol	Performance Measure (Round)		
		FND	**LND**	**IPL**
Experiment I (100 Nodes)	LEACH	726	1209	483
	TCAC	933	1006	73
	SEECH	1028	1099	71
	HACH	1064	1167	103
Experiment II (400 Nodes)	LEACH	685	1274	589
	TCAC	948	1071	123
	SEECH	1016	1140	124
	HACH	1235	1307	72
Experiment III (1000 Nodes)	LEACH	672	2014	1342
	TCAC	725	1664	939
	SEECH	1587	2202	615
	HACH	1789	2010	221

Table 3. Average Energy at FND for our proposed `HACH` protocol

	Experiments		
	Experiment I	**Experiment II**	**Experiment III**
AEFND	0.0232	0.0164	0.0650

of 36, 131 and 338 more than the SEECH, TCAC and LEACH respectively. The energy of the sensor nodes are balanced until the FND times and this is observed on the graph by a sharp decrease in the number of alive nodes for the TCAC, SEECH and `HACH`.

When the FND is achieved or during the IPL, lots of nodes begin to die due to the limited residual energy. The IPL values of various protocols are shown in Table 2 with `HACH` having a very low IPL values in Experiment II(400 nodes) and III (1000 nodes) except in Experiment I which has 30 rounds more than TCAC. This shows that our algorithm performs better in a dense network. Using our proposed `HACH` protocol, the AEFND values for Experiments I,II and III are shown in Table 3. It can be observed that the AEFND value for each of the three experiments is approximately zero at the time the first node dies. For example, Figure 4 shows that at FND time of 1064, the AEFND of sensor nodes is 0.0232J: which means we were able to used almost all of our initial energy till the death of first node.

From the above, it can be deduced that the proposed `HACH` decrease in energy consumption and energy balancing can increase the network lifetime. At each round, our proposed `HACH` protocol conserves energy by selectively allowing some nodes to become inactive before network operation. Also, it uses an adaptive heuristic approach to select the best CHs configuration by ensuring that CH are optimally distributed over the sensor field.
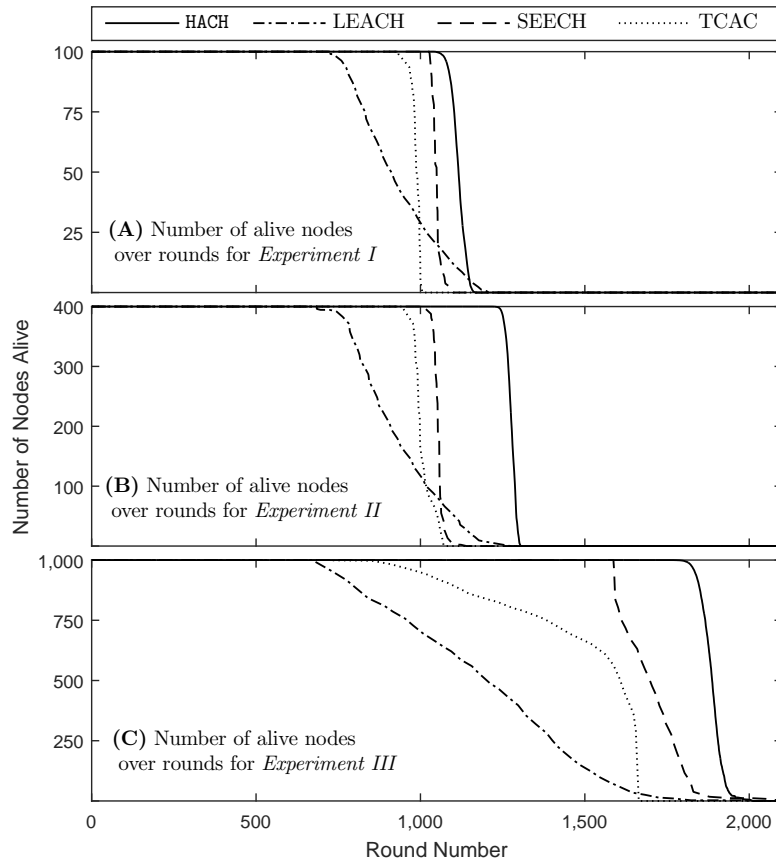
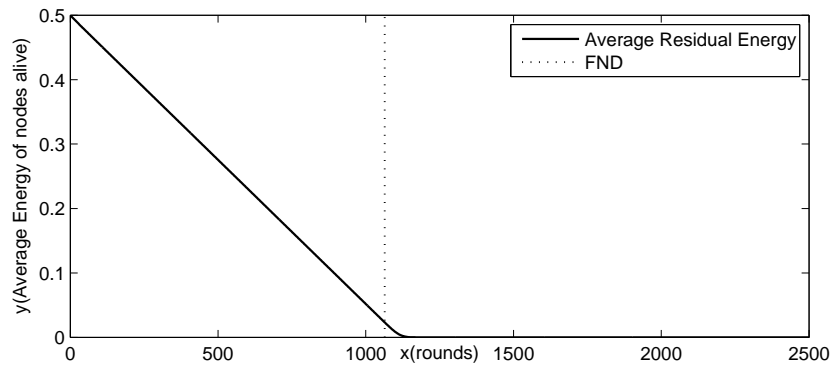**Fig. 3.** Network Lifetime Comparison of HACH with LEACH, SEECH, TCAC

Fig. 4. Average Energy of alive node versus Rounds (refer to Experiment I)

6   Conclusion

In this paper, a Heuristic Algorithm for Clustering Hierarchy (`HACH`) protocol was proposed where the major two operations included sending some nodes into sleep mode and the cluster head selection. We proposed the Heuristic-Crossover Enhanced Evolutionary Algorithm for Cluster Head Selection protocol (`HEECHS`), a mechanism that employs genetic algorithms for a cluster head selection that balance energy consumption and enhances better distribution of cluster heads. To achieve the selection process, an objective function to access the quality of each solutions was designed. A problem-dependent heuristic crossover to produce a better cluster head configuration was implemented. Also, we employed a new mechanism called Stochastic Selection of Inactive Node (`SSIN`) inspired by Boltzmann selection process to stochastically select some nodes to send into sleep without causing an adverse effect on coverage. The two proposed mechanisms work collaboratively to reduce and balance energy consumption by selecting distributed nodes with high energy as cluster heads in order to prolong network lifetime. From the three experiments, simulation results demonstrated that our proposed `HACH` outperforms other protocol in terms of energy consumption, network lifetime and stability period in all the three experiments.

References

1. Soroush Naeimi, Hamidreza Ghafghazi, Chee-Onn Chow, and Hiroshi Ishii. A survey on the taxonomy of cluster-based routing protocols for homogeneous wireless sensor networks. *Sensors*, 12(6):7350–7409, 2012.
2. Ayon Chakraborty, Swarup Kumar Mitra, and Mrinal Kanti Naskar. Energy efficient routing in wireless sensor networks: A genetic approach. *CoRR abs/1105.2090*, 2011.
3. Ameer Ahmed Abbasi and Mohamed Younis. A survey on clustering algorithms for wireless sensor networks. *Computer communications*, 30(14):2826–2841, 2007.
4. Wendi B Heinzelman, Anantha P Chandrakasan, and Hari Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. *Wireless Communications, IEEE Transactions on*, 1(4):660–670, 2002.
5. William E Hart, Natalio Krasnogor, and James E Smith. *Recent advances in memetic algorithms*, volume 166. Springer Science & Business Media, 2005.
6. Sang H Kang and Thinh Nguyen. Distance based thresholds for cluster head selection in wireless sensor networks. *Communications Letters, IEEE*, 16(9):1396–1399, 2012.
7. LiCheng YeMao, ChenGui Fa, and WuJie Hai. An energy efficient clustering scheme in wireless sensor networks. *Ad Hoc & Sensor Wireless Networks (to be published)*.
8. Nikos Dimokas, Dimitrios Katsaros, and Yannis Manolopoulos. Energy-efficient distributed clustering in wireless sensor networks. *Journal of parallel and Distributed Computing*, 70(4):371–383, 2010.
9. Shan Lin, Jingbin Zhang, Gang Zhou, Lin Gu, John A Stankovic, and Tian He. Atpc: adaptive transmission power control for wireless sensor networks. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, pages 223–236, 2006.

10. V Loscri, G Morabito, and S Marano. A two-levels hierarchy for low-energy adaptive clustering hierarchy (tl-leach). In *IEEE Vehicular Technology Conference*, volume 62, page 1809. IEEE; 1999, 2005.
11. Ossama Younis and Sonia Fahmy. Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *Mobile Computing, IEEE Transactions on*, 3(4):366–379, 2004.
12. Dahlila P Dahnil, Yashwant Prasad Singh, and Chin Kuan Ho. Topology-controlled adaptive clustering for uniformity and increased lifetime in wireless sensor networks. *IET Wireless Sensor Systems*, 2(4):318–327, 2012.
13. Mehdi Tarhani, Yousef S Kavian, and Saman Siavoshi. Seech: Scalable energy efficient clustering hierarchy protocol in wireless sensor networks. *Sensors Journal, IEEE*, 14(11):3944–3954, 2014.
14. Selim Bayrakli and Senol Zafer Erdogan. Genetic algorithm based energy efficient clusters (gabeec) in wireless sensor networks. *Procedia Computer Science*, 10:247–254, 2012.
15. NM Latiff, Charalampos C Tsimenidis, and Bayan S Sharif. Energy-aware clustering for wireless sensor networks using particle swarm optimization. In *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, pages 1–5. IEEE, 2007.
16. Jenn-Long Liu, Chinya V Ravishankar, et al. Leach-ga: Genetic algorithm-based energy-efficient adaptive clustering protocol for wireless sensor networks. *International Journal of Machine Learning and Computing*, 1(1):79–85, 2011.
17. K Go Vijayvargiya and V Shrivastava. An amend implementation on leach protocol based on energy hierarchy. *International Journal of Current Engineering and Technology*, 2(4):427–431, 2012.
18. D. Dumitrescu, B. Lazzerini, L.C. Jain, and A. Dumitrescu. *Evolutionary Computation*. International Series on Computational Intelligence. Taylor & Francis, 2000.
19. Tang Lixin. Improved genetic algorithms for tsp. *JOURNAL OF NORTHEASTERN UNIVERSITY (NATURAL SCIENCE)*, page 01, 1999.
20. Basela S Hasan, Mohammad Khamees, Ashraf S Hasan Mahmoud, et al. A heuristic genetic algorithm for the single source shortest path problem. In *Computer Systems and Applications, 2007. AICCSA'07. IEEE/ACS International Conference on*, pages 187–194, 2007.
21. Rajesh Halke and VA Kulkarni. En-leach routing protocol for wireless sensor network. *International Journal of Engineering Research and Applications*, 2(4):2099–2102, 2012.
22. JS Brunda, BS Manjunath, BR Savitha, and P Ullas. Energy aware threshold based efficient clustering (eatec) for wireless sensor networks. *Energy*, 2(4), 2012.