Multi-Objective Decision Making Approach based Load Balanced Clustering in WSNs to increase the lifetime of network

Lekhraj¹, Alok Kumar², Anoj Kumar³
Email- lekhraj@mnnit.ac.in¹
alokkumar@mnnit.ac.in²
anojk@mnnit.ac.in³
¹, ², ³MNNIT Allahabad, Prayagraj, India

Abstract
A network of wireless sensors (WSN) is an outstanding technology that can aid in the various applications. Batteries run the sensor nodes those are used in WSN. The battery is impossible to charge or repair, so the most valuable resource for wireless sensor networks is power. Over the years, several strategies have been invented and used to preserve this precious WSN resource. One of the most successful approach for this purpose has turned out to be clustering. The aim of this paper is to suggest an effective technique for choosing cluster heads in WSNs to increase the lifetime of the network. To accomplish this task, Grey Wolf Optimizer (GWO) technique has been used. The general GWO was updated in this paper to meet the particular purpose of cluster head selection in WSNs. In this article, we have considered eleven attributes in the fitness function for the proposed algorithm. The simulation is carried out under different conditions. The results obtained show that the proposed protocol is superior in terms of energy consumption and network lifetime by evaluating the proposed protocol (i.e. CH-GWO protocol) with some well-existing cluster protocols. The suggested protocol forms energy-efficient and scalable clusters.

Keyword: Clustering, LEACH, GWO, MODM, WSN

1. Introduction
Facts have proved that a wireless sensor network (WSN) is one of the most promising reception mechanisms. WSNs gives an opportunity to process and transmit data from the remote environment to the Base Station (BS) [1]. In WSN BS acts as a central node or central data processing node. They found use in various tasks such as forwarding of data, broadcasting of information, and routing of the network to transmit the information from one place to another. Typical WSN includes hundreds of less sized and fixed power battery sensor nodes and all the nodes are independent to other nodes for sensing and performing the operations over data. These sensor nodes are usually deployed in remote areas, so the battery cannot be easily replaced. So it is very difficult to recharge the batteries of the deployed sensor nodes in remote areas. In traditional WSNs all the nodes are directly connected to the BS node. When any nodes sensed information from their surroundings, forward this information to the BS node. This technique losses energy of the nodes very fast. This is where energy-saving work in wireless sensor networks plays an important role. To maintain WSN, it is important to decrease energy consumption in data transmission. To achieve a low power consumption in data transmission, clustering gives an effective and simple enough way to maintain this. Clustering divides nodes into clusters and the representation of data transmission in WSN using clustering is shown in figure 1.
In the past few decades, various hierarchical clustering algorithms for topology control have been proposed in the literature. Specifically, LEACH gets the best attention in all hierarchical clustering approaches for maintaining the best topology for data transmission [4]. LEACH focuses on randomly selecting cluster head nodes in a round-robin manner, and the energy load of the entire network is evenly distributed to each sensor node [5]. It also focuses on the improvement of the overall network life cycle. However, the randomness involved in the selection of cluster heads in LEACH leads to the imbalance of remaining energy between nodes, and also leads to the earlier death of some sensor nodes in the network [6]. In addition, a centralized LEACH (LEACH-C) is proposed as an extended version of LEACH, in which the selection of the best cluster head node is achieved by executing a simulated annealing algorithm [7].

In essence, the cluster is to group multiple nodes of the WSN and inveterate transmission. Communication is carried out between the group and the BS node. In each group, there is one leader who communicates to the BS called the cluster head (CH) [2, 3]. Clusters selection and the corresponding CH itself is an arduous and difficult task. Over the years, many approaches have been applied to the best choice of CHs selection. In past times, various approaches [8] have been used to choose optimal combination CH groups over the randomly deployed network. Several optimization approaches give a better analysis in the selection of optimal CHs. These approaches include types of harmony search algorithms [13], types of particle swarm optimization [12], fuzzy logic [9], types of genetic algorithms [11], and types of ant colony optimization [10], and so on.

This paper uses Gray Wolf Optimization (GWO) approach [14] to give the solution of the set of optimal cluster head selection problems. The solution of the cluster head selection problem is calculated by simulating the gray wolf itself. Alpha is the best solution, the rest is the subsequent level of the solution.

The rest of this article is organized as follows: Summarization of the related work done in the cluster field so far is discussed in section 2. Section 3 contains a detailed explanation of the GWO approach. Section 4 contains information related to the radio model and assumptions about the network are described. Attributes used in the proposed fitness function detail discussion is in section 5. The proposed approach analysis is discussed in section 6. Section 7 introduces the simulation results. Section 8 contains the conclusion and future scope.

2. Literature Survey
To provide a powerful businesslike approach towards a WSN is a difficult task in the current environment. So there is a need for such an approach that gives a long duration to the WSN network in the data collection
or communication. Because in traditional network data communication is done by the direct, which is not efficient direct communication limits the lifetime of the WSN because it consumes more energy to transmit the data to the Base Station (BS). To overcome this limitation clustering approaches give a better understanding. Various clustering approaches are suggested by many researchers [15, 16, and 17]. Several work has been done to aggregate data on a wireless sensor network, which reduces power consumption. Clustering in WSN is an efficient procedure for reducing the power consumption of sensor nodes. In cluster routing algorithms for wireless networks, LEACH is known for its simplicity and efficiency. In LEACH, CHs are randomly selected and all non-CHs are generated based on the received signal strength from the CHs. In LEACH, every node can become a CH, there is no CH selection scheme, and all nodes have the same chance of becoming a CH, so LEACH is ineffective. The CHs are chosen at random, and the energy is distributed equally among all the nodes. CHs collect all sensed information from its cluster [15]. LEACH protocol used clustering for exchanging the information from one sensor to another. It makes a cluster by considering the signal strength of the sensors. In LEACH every cluster send the data to BS after performing aggregation over the data. CH is selected periodically among the cluster nodes. It forms a cluster based on individual perception of the nodes and this type of clustering is belongs to distributed category. In this approach, every CH has a direct link to the sink. In LEACH protocol some limitations are that CHs distribution are not uniformly and according to this theory energy consumption of each CHs are equal so the nonuniform distribution of CHs limits the lifetime of the network. To overcome this limitation another protocol low energy adaptive clustering hierarchy- centralized (LEACH-C) is given by another researcher. LEACH-C [15] selects CHs based on the residual energy of the nodes and average energy of the network, suggest better clustering as compare to the LEACH. In LEACH-C a node is only eligible for acts as a CHs if its remaining energy is higher than the average residual energy of the whole network.

All existing approaches require either knowledge of network density or uniform dispersion of nodes in the field. Eunice and Fahmy. [18] offer hybrid energy-efficient distributed clustering (HEED). HEED makes no assumptions about the network, such as density and size. Each node operates individually in HEED protocol. At the end of the process, each node becomes either a cluster head or a child with respect to the cluster head. The residual energy of the node is the first parameter when choosing the cluster header, and the proximity to its neighbors or the degree of the node is the second. HEED generates a single-level hierarchical clustering structure for intra-cluster communication. HEED is also not giving the guarantee that the selected number of elected CHs would be optimum. In [19, 20], a hybrid approach to aggregation based on clusters is presented, which adaptively selects the appropriate data aggregation function. This document shows the improvement in power consumption versus target speed. Dynamic clustering shows the best performance at high target speeds. Energy-Efficient Hierarchical Clustering Protocol (DWEHC) [21] based on distributed weights is an improvement of HEED. Even if there are some similarities, such as the density and size independent of WSN, the remaining energy is used as the main parameter, but it will The size of the generated CH balance does not overlap. HEED also has some disadvantages. However, HEED provides a good CH distribution, but because the tentative CH is not selected as the CH in the end, there are some undiscovered nodes. According to the implementation of HEED, these nodes are forced to become CH, so it may not be any associated node or maybe in the area of other CHs, resulting in higher CH generation and consuming unbalanced energy. Since a large number of control messages are generated in the iterative process, both HEED and DWEHC will generate a lot of overhead. In [22] another approach is suggested, gives the approximately equal size and non-overlapped clusters are called Fast LOCal Clustering (FLOC). This condition is met: a node with a distance greater than m hops from the CH cannot be connected to the CH, and a node with a one-hop distance from the CH is always connected to the CH.

In heuristic clustering approaches another algorithm is added by the one researcher, which oversees the unbalanced energy consumption in WSN is called energy-efficient clustering algorithms (EECS) [23]. To determine CH, three functions and a weighted cost factor are required. Another modified variant of EECS is energy efficiency unequal clustering (EEUC) [24], which takes into account certain restrictions of EECS, for example, for computing the CH cost each node should be entertain. Since the CH needs to transmit control packets through high transmission power, and the node needs to maintain information about the
CH, the CH and the node have a reason for wasting energy. In article [25] another heuristic clustering approach is discussed is called the energy-efficient heterogeneous clustering scheme of the wireless sensor network (EEHC). In this approach weighted potential probability can be decided on the basis of the residual energy of each node. The Super Heterogeneous Protocol for Balanced Energy Saving Network Integration (BEENISH) [26] gives clustering by considering the remaining power level of specific sensors. In the BEENISH protocol, four power levels are considered to provide efficient clustering in WSNs. In article [27] power-efficient and adaptive clustering hierarchy (PEACH) is introduced by researchers, which supports location-aware as well as location unaware WSNs and also gives the facility to do multi-level clustering. Using some of the attributes in Table 1, some heuristic algorithms have been compared with the proposed algorithm.

In recent studies, optimization approaches give good performance in clustering [28, 29]. Particle Swarm Optimization (PSO) is a way to increase the lifetime of the network. The algorithm proposes to construct paths and distribute routing data through clusters near base stations or the heads of gateway nodes to maximize the network time axis.

Genetic Algorithm (GA) belongs to optimization [30], which can be used for clustering to provide an efficient solution in CHs selection. The algorithm can be changed by selecting many other attributes in the function [31], as selecting a single of the two signal ranges from normal sensors, the state of the sensor node, and selecting the appropriate channel. GA [32] can be select best topology management to obtain less power consumption and achieve balance.

A GWO-based protocol is proposed, which is called Improved Gray Wolf Optimization Based on Fitness Value (FIGWO) [33]. CHs are selected on the basis of the fitness function value. In the FIGWO approach, CHs are selected based on the fitness function value of each node, nodes have the highest energy and nearest to the BS get a chance to become a CHs. In addition, the distance to sending data is recalculated every time when a new CH is selected. However, it suffers from load balancing between CHs.

In this article, the protocol is proposed to provide efficient clustering in the WSNs using GWO. There are certain reasons for selecting the GWO instead of another heuristic as well as metaheuristic approaches. Primary objective: GWO converges faster; Secondary objective is: Fewer decision variables and it avoids local optima.

Table 1: Comparison of existing CH selection approaches with proposed algorithm

| Protocol name | Type of topology | Network Type | CH selection | Attributes used in fitness function | Clustering objective |
|---------------|-----------------|--------------|--------------|--------------------------------------|----------------------|
| LEACH         | Single hop      | Homogeneous  | Random       | None                                 | Energy consumption   |
| LEACH-C       | Single hop      | Homogeneous  | Random       | None                                 | Energy consumption   |
| HEED          | Single hop      | Homogeneous  | Random with iteration | None                             | Energy consumption   |
| FIGWO         | Single hop      | Homogeneous  | Random with selection | Two                              | Energy consumption   |
| GWO-C         | Single hop      | Homogeneous  | Random with selection | Four                             | Energy consumption   |
| CH-GWO (Proposed) | Single hop | Homogeneous  | Random with selection | Eleven                           | Energy consumption   |

3. Grey Wolf Optimizer

GWO is a meta-heuristic algorithm that can solve many optimization problems. GWO drew inspiration from the leadership that naturally exists in the hunting mechanism of gray wolves. It is the most studied algorithm [14] and a new optimization method compared to other optimization methods. In terms of implementation and use, the algorithm is similar to the genetic algorithm. In this article [14], algorithm mathematical equations have come from the pack hunting mechanisms observed patterns. Then, adapt the equation to the current problem to find an optimal feasible solution.
Candidate solutions can also be categorized based on the hierarchy of gray wolves' social dominance. Therefore, alpha ($\alpha$) makes the best and most optimized solution, where beta ($\beta$) and delta ($\delta$) form the second and third best solutions, accordingly [14]. All other solutions are regarded as ($\Omega$) solutions and are the most unsuitable solutions. The optimization process is based on the hunting process, and it works over the guidance of $\alpha$, $\beta$, $\delta$, and $\Omega$ parameters.

During the hunt, the gray wolf surrounds its prey [14]. This enclosed behavior of wolves can be described mathematically as:

$$\vec{Z} = |\vec{K}| \cdot \vec{Y}_p(t) - \vec{Y}(t) \quad (1)$$

$$\vec{Y}(t + 1) = \vec{Y}_p(t) - \vec{A} \cdot \vec{Z} \quad (2)$$

$$\vec{A} = 2\vec{b} \cdot \vec{r}_1 - \vec{b} \quad (3)$$

$$\vec{K} = 2 \cdot \vec{r}_2 \quad (4)$$

$$\vec{Z}_\alpha = |\vec{K}_1 \cdot \vec{Y}_\alpha - \vec{Y}| \quad (5)$$

$$\vec{Z}_\beta = |\vec{K}_2 \cdot \vec{Y}_\beta - \vec{Y}| \quad (6)$$

$$\vec{Z}_\delta = |\vec{K}_3 \cdot \vec{Y}_\delta - \vec{Y}| \quad (7)$$

$$\vec{Y}_1 = \vec{Y}_\alpha - \vec{A}_1 \cdot (\vec{Z}_\alpha) \quad (8)$$

$$\vec{Y}_2 = \vec{Y}_\beta - \vec{A}_2 \cdot (\vec{Z}_\beta) \quad (9)$$

$$\vec{Y}_3 = \vec{Y}_\delta - \vec{A}_3 \cdot (\vec{Z}_\delta) \quad (10)$$

$$\vec{Y}(t + 1) = \frac{\vec{Y}_1 + \vec{Y}_2 + \vec{Y}_3}{3} \quad (11)$$

Where $\vec{A}$ and $\vec{K}$ are coefficient vectors and calculated by using equation 3 and 4 respectively, current iteration is denoted by $t$, $\vec{Y}_p$ represents prey vector position, $\vec{Y}$ represents grey wolf vector position, $\vec{r}_1$ and $\vec{r}_2$ denotes two random vectors and the value of these two random vectors should be in between 0 and 1, $\vec{b}$ components decreases its value from 2 to 0 when iteration is repeated.

In this process, the location of the prey is presumed to be unknown. The search process is led by the best alpha and beta candidate solutions and the least relevant member (i.e., omegas updates its position based on the information provided by the best search agent) [14] alpha and beta. The mathematical description is modelled in terms of equation 5-11 for this purpose:
4. Energy model and Presuppositions

4.1. System Model

In WSN, the sensor nodes sense the environment near to it and sends data to the corresponding CH, and the CH sends the aggregated information to the receiver after collecting the information. For this kind of transmission, to adjust the power consumption, we need representation. As shown in Figure 2, we have adopted an ordinary model of hardware energy consumption.

![Energy Model Diagram]

Fig 2: Energy Model

We assume that fading is in multipath and free-space for experimental purposes, and depends on the distance from the transmitter to the receiver node. For example, if the transmission distance is smaller than the threshold $d_0$, the power amplifier can decrease the power loss and control the power, only the free space model is used, and if its value is larger than the threshold, then multipath model is used for data transmission. To transmit p bit data at distance $d$, required energy is:

$$E_{TX} = E_{(TX-elec)}(p) + E_{(TX-amp)}(p - d)$$  \hspace{1cm} (12)

Energy consumption to transmit the p bit data at transmitter end:

$$E_{TX} = p * E_{elec} + p * e_{fs} * d^2 \hspace{1cm} \text{if } d \leq d_0$$

$$E_{TX} = p * E_{elec} + p * e_{amp} * d^4 \hspace{1cm} \text{if } d \geq d_0$$  \hspace{1cm} (13)

Energy consumption to receive the p bit data at receiver end:

$$E_{RX} = p * E_{elec}$$  \hspace{1cm} (14)

Optimal number of CHs are calculated for the simulation as [15] by equation 15.

$$K_{optimal} = \sqrt{\frac{n}{2\pi}} * \sqrt{\frac{e_{fs}}{e_{amp}}} * \frac{M}{d_{to \ BS}^2}$$  \hspace{1cm} (15)

4.2. Presuppositions

1. Nodes are distributed randomly in a 2 dimensional space.
2. All nodes become stationary after a deployment.
3. All nodes are homogeneous.
4. Sensor nodes sense the data from its surrounding and communicate the sensed information to the respective CHs.
5. BS has infinite energy for communication.
6. Sensor nodes can change their mode from active to sleep and sleep to active.

Table 2: Assumption for simulation

| Parameter Description                        | Symbol | Parameter Value |
|----------------------------------------------|--------|-----------------|
| Nodes                                        | N      | 100, 200        |
| Deployment Area                              | M*M    | 100*100, 200*200|
| Sink Position                                | -      | Variable        |
| Data packet size                             | -      | 500 byte        |
| Control packet size                          | -      | 100 bit         |
| Amplifier energy consumption at free space   | $\epsilon_{fs}$ | 10 pj/bit/m$^2$ |
| Amplifier energy consumption at multi space  | $\epsilon_{amp}$ | 0.0013 pj/bit/m$^4$ |
| Sensor nodes initial energy                  | -      | 0.5 j           |
| Energy consumed in the circuit.              | $E_{elec}$ | 50 nj/bit       |
| Number of CHs                                | $K_{opt}$ | Calculated      |

5. Attributes used in fitness function
In WSN, there are several attributes those are useful in data collection. We're looking at 11 attributes. Among these, few in nature are beneficiaries and few are non-beneficiaries. Beneficiary criteria need a higher value for energy conservation, but a lower value for non-beneficiary criteria is good for having the best solution. In Table 3 definition of the attributes is located.

Table 3: Attributes Definition

| S.no. | Attribute Name          | Nature of Attribute | Definition                                                                 |
|-------|-------------------------|---------------------|---------------------------------------------------------------------------|
| 1     | Coverage_of_CHs         | Beneficial          | Greater value is needed, in order for maximum nodes to beat lower distances to the respective CH |
| 2     | CH_BS_Bearing           | Beneficial          | Greater value is needed, in order for maximum CHs to beat lower distances to the BS |
| 3     | Avg_Eresidual           | Beneficial          | Greater value is needed so that chosen CHs energy will be high             |
| 4     | BS_Max_Distance         | Non-Beneficial      | Lower value is needed so that chosen CHs would not have more distance to BS |
| 5     | CHs_Avg_Distance        | Non-Beneficial      | Lower value is needed so that nodes would not have more distance to respective CH |
| 6     | BS_Avg_Distance         | Non-Beneficial      | Lower value is needed so that from CHs to BS average distance will be less |
| 7     | Node_Energy             | Non-Beneficial      | Lower value is needed so that lifetime of the nodes will be higher         |
| 8     | CH_Energy                | Non-Beneficial      | Lower value is needed so that lifetime of the chosen CHs will be higher    |
| 9     | CHs_Avg_lifetime        | Beneficial          | Greater value is needed so that maximum CHs will have higher lifetime      |
| 10    | Avg_Eres_of_Conn_CH     | Beneficial          | Greater value is needed so that connected CHs will have high average energy |
| 11    | Avg_Eres_of_Disconn_CH  | Beneficial          | Greater value is needed so that disconnected CHs will have high average energy |
1) **Coverage_of_CHs**
This indicates the percentage of sensor nodes whose distance from the corresponding CH is less than or equal to d0. Its maximum value confirms that a large number of nodes are nearer to the respective CH. This confirms that the demand of electricity will be decreased. The formula for the estimation is:

\[
Coverage \text{ of } CHs = \frac{\text{Count}_{Nodes}(\text{Min}(Distance_{Node-v-CHs}) \leq d0)}{n} \times 100
\]  

Where \( d_0 = \sqrt{\frac{\epsilon_s}{\epsilon_m}} \) \( Distance_{(Node-v-CHs)} \) shows distance in between nodes and CHs & \( \text{Count}_{Nodes_v} \) count number of nodes_v.

2) **CH_BS_Bearing**
This indicates that how many CHs in total have a distance less than d0 from the BS. A higher value indicates that the distance from the selected CH to the receiver is shorter, which means that the less power is required for data transmission from CH to BS. The formula for the estimation is:

\[
\text{CH-BS-Bearing} = \frac{\text{Count}_{CHs}(\text{Distance}_{(CH-v-BS)} \leq d0)}{\text{Tot}_\text{Optimal}_CHs}
\]

Where \( Distance_{(CH-v-BS)} \) indicates CH ‘v’ to BS distance, \( \text{Count}_{CHs_v} \) indicate count number of CHs, \( \text{Tot}_\text{Optimal}_CHs \) shows optimal number of CHs.

3) **Avg_E_residual**
This indicates the average value of the remaining power of CHs. The larger value of this attribute indicates that the selected CHs have higher remaining energy, which means that more information can be collected and sent to the BS by using these CHs. The calculation formula is:

\[
\text{Avg E residual} = \frac{\sum_{v=1}^{\text{Tot}_\text{Optimal}_CHs} \text{E residual}_{CH_v}}{\text{Tot}_\text{Optimal}_CHs}
\]

Where \( \text{E residual}_{CH_v} \) shows selected CH remaining energy.

4) **BS_Max_Distance**
It demonstrates the maximum distance to the sink from any selected CH. Its minimum value shows that all CHs are needed to transmit information below this distance. This implies that total power consumption is limited. It is determined as:

\[
\text{BS-Max_Distance} = \text{Max}(Distance_{CH_v-BS})
\]

5) **CHs_Avg_Distance**
It demonstrates the mean value of the distances between all sensors and the respective CHs. Its smaller value means that the data packet requires a shorter distance to be transmitted. It also demonstrates the lower dissipation of electricity. It is determined as:

\[
\text{CHs_Avg_Distance} = \frac{\sum_{v=1}^{n} \text{Distance}_{Node-v-NearCH}}{n}
\]

6) **BS_Avg_Distance**
The mean value of all the distances from the chosen CHs to the BS is shown. The lower value means the CHs are closer to the BS. It is determined by its value as:

\[
\text{BS-Avg_Distance} = \frac{\sum_{v=1}^{\text{Tot}_\text{Optimal}_CHs} \text{Distance}_{(CH_v-BS)}}{\text{Tot}_\text{Optimal}_CHs}
\]
7) **Node Energy**
This demonstrates the energy needed by the sensors to transmit the data to the correspondence CHs. The lower value of this factor indicates that the maximum total number of nodes is closer to their correspondence CH. It is determined by its value as:

\[
Node_{Energy} = \sum_{v=1}^{n} E_T(Node_{v} - CH)
\]  

(22)

Where \( E_T(Node_{v} - CH) \) shows energy required to transmit the data from node \( v \) to respective CHs.

8) **CH Energy**
This demonstrates the energy needed by the CHs to transmit the data to the BS. The lower value of this factor indicates that the maximum total number of CHs is closer to the BS. It is determined by its value as:

\[
CH_{Energy} = \sum_{v=1}^{Tot\_Optimal\_CHs} E_T(CH_{v} - BS)
\]  

(23)

Where \( E_T(CH_{v} - BS) \) shows energy required to transmit the data from CH \( v \) to BS.

9) **CHs Avg lifetime**
The mean lifetime of CHs is shown. It illustrates how much time the CHs will continue the process of transmission. Large value means that more data can be processed by the CHs, meaning that CHs have more life. It is determined by its value as:

\[
CH_{s\_Avg\_lifetime} = \sum_{v=1}^{Tot\_Optimal\_CHs} \left( \frac{E_{res_{CH_{v}}}/Avg\_Transmission\_Energy\_required}{Tot\_Optimal\_CHs} \right)
\]  

(24)

Where \( Avg\_Transmission\_Energy\_required \) shows energy required to transmit the data from CH to BS.

10) **Avg_Eres_of_Conn_CH**
Connected CH is defined as the CHs those have a less distance from the BS. It measures the mean residual power of CHs those have a less distance to BS. The larger value of this factor indicates that the distances of the CHs to the BS is less than 'd0' and that the residual energy is higher. It is determined by its value as:

\[
Avg\_Eres\_of\_Conn\_CH = \frac{\sum_{v=1}^{Tot\_Optimal\_CHs} (Distance(CH_{v} - BS) \leq d_0)}{Tot\_Optimal\_CHs}
\]  

(25)

11) **Avg_Eres_of_Disconn_CH**
Disconnected CH is defined as the CHs those have a higher distance from the BS. It measures the mean residual power of CHs those have a higher distance to BS. The greater value of this factor means that the CHs that are far from the BS have higher energy, showing that CHs survival time will be longer. It is determined by its value as:

\[
Avg\_Eres\_of\_Disconn\_CH = \frac{\sum_{v=1}^{Tot\_Optimal\_CHs} (Distance(CH_{v} - BS) > d_0)}{Tot\_Optimal\_CHs}
\]  

(26)

6. **Proposed Work**
In this section, a detailed description of the proposed protocol is given. The process of clustering is divided into two parts: the selection step for CHs is the first segment, and the creation stage for clusters is the second segment.

| Abbreviation | Attributes                  |
|--------------|-----------------------------|
| C1           | Coverage of CHs             |
| C2           | CH BS Bearing               |
| C3           | Avg Eresidual               |
6.1. CH Selection
The selection of CH in the network is done through GWO technology. Usually, CHs are elected according
to the parameters of the sensor network (such as distance from the BS, remaining energy of the sensor nodes
and coverage of the sensor nodes). Choosing CHs through a single parameter will not be profitable for an
effective communication process because the energy usage and lifetime of the entire network are also can
influenced by several other competing factors. Therefore, certain other attributes considerations need to be
addressed and there is also a need for cooperation between them. Eleven parameters as given in Table 4
are considered in this proposed protocol, i.e. CH-GWO, to elect the optimal CHs. All the sensor nodes sends
their exact location and residual energy to the BS after the initialization of the network. BS runs a cluster
protocol based on GWO. The fitness function is just established based on 11 considered attributes as given
[36], Coverage_of_CHs, CH_BS_Bearing, Avg_Eresidual, BS_Max_Distance, CHs_Avg_Distance,
BS_Avg_Distance, Nodes_Energy, CH_Energy, CHs_Avg_lifetime, Avg_Eres_of_Conn_CH,
Avg_Eres_of_Disconn_CH.
The primary aim of the CH-GWO is to select the CHs in the network to extend the life of the network.
Eleven factors are considered to effectively select the CHs. The GWO approach [14] is used to provide the
optimal lifetime and balanced energy consumption by choosing the best set of CHs. The coordination
between these factors verifies that the balance between conflicting attributes helps to create favorable
conditions for optimal energy consumption. Instead of minimizing each fitness function equation (16-26)
separately, minimize the combination of these considered factors by using fitness function equation 27.

\[
\text{Fitness} = k_1 \times c_1 + k_2 \times c_2 + k_3 \times c_3 - k_4 \times c_4 - k_5 \times c_5 - k_6 \times c_6 - k_7 \times c_7 - \\
k_8 \times c_8 + k_9 \times c_9 + k_{10} \times c_{10} + k_{11} \times c_{11} \\
(27)
\]

Where \(k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, k_9, k_{10}, k_{11}\) are constant and \(k_1 + k_2 + k_3 + k_4 + k_5 + k_6 + \\
k_7 + k_8 + k_9 + k_{10} + k_{11} = 1\) and weightage of each attribute for optimal CHs set selection is mentioned
in Table 5.

Table 5: Attribute Weightage in CH selection

| Attribute constant name | Attribute constant weightage |
|------------------------|----------------------------|
| K1                     | 0.12                       |
| K2                     | 0.12                       |
| K3                     | 0.12                       |
| K4                     | 0.056                      |
| K5                     | 0.056                      |
| K6                     | 0.056                      |
| K7                     | 0.056                      |
| K8                     | 0.056                      |
| K9                     | 0.12                       |
| K10                    | 0.12                       |
| K11                    | 0.12                       |
6.2. Cluster Formation

After receiving the notification of the CHs, the member node selects the nearest CH and uses the CSMA MAC protocol to reply the Join packet to the corresponding CH. It also measures the distance between the member node and the respective CH.

The sensors can be scheduled to sense the atmosphere at various samples or time intervals at the sensor node level and sleep as much as possible to save power of the batteries. Although some information may be lost, this is an efficient way of maximizing the lifetime of the sensor nodes [34]. If the distance between the member node and the CH is greater than the distance between it and the BS, the member node can interact directly with the BS at a fixed time slice. Otherwise, to form a cluster, it will join the cluster based on the closest distance (Euclidean distance). Use the distance matrix MD (m x n) to re-cluster nodes according to the distance of the selected cluster head, as equation 28:

\[
MD = \begin{bmatrix}
d_{ch1,y1} & d_{ch2,y2} & \cdots & d_{chn,yn} \\
d_{ch2,y1} & d_{ch2,y2} & \cdots & d_{ch2,yn} \\
\vdots & \vdots & \ddots & \vdots \\
d_{chm,y1} & d_{chm,y2} & \cdots & d_{chn,mn}
\end{bmatrix}
\]

(28)

Where d is a Euclidean distance between a node location and a CH location. If a \((x, y)\) and b \((x, y)\) are two nodes then Euclidean distance between these nodes are calculated by using equation 29:

\[
d_{a,b} = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2}
\]

(29)

Each value \(d_{i,j}\) in MD matrix is a distance value of the \(i^{th}\) cluster head and the \(j^{th}\) node. The column containing the minimum value indicates the cluster number to be joined by the corresponding node. For example, if \(d_{Ch1,y2}\) is the minimum value in the second column, then in this case, node \(y2\) will associated with that cluster whose cluster head is \(Ch1\).

Once a cluster is created, CH assigns a time slot for each member after receiving all CH join emails from all nodes. The responsibility for collecting data from all nodes in the cluster are of respective CHS. The CH sends the frames to the base station after applying data aggregation while collecting data frames from all members. While the cluster member nodes can enter sleep mode when those have not any data to sense, the CH must remain in active condition. It should be noted that in the LEACH protocol, the re-clustering technique is often followed where CHs are chosen using the probabilistic approach instead of a deterministic method. The re-clustering and data transfer process will continue for several rounds until all the nodes not get die.
Working model or framework of proposed clustering protocol is as figure 3:

![Diagram of proposed energy efficient clustering protocol working model]

6.3. Algorithm
GWO technique gives faster convergence and it also uses few variables and avoids to stick in local minima. On the basis of these properties we used this algorithm in wireless sensor network to select the optimal CHs. The pseudo code of the proposed CH-GWO algorithm is described below:

**Input:** Number of alive sensor nodes in each round.
Number of groups: \( N_g \)
Number of CHs in a group: Calculated based on equation 15 from the alive nodes whose remaining energy is higher than the average remaining energy of the all nodes.

**Output:** Optimal CHs set

| Initialize Groups \( g_i \) |
|-----------------------------|
| for \( i = 1: N_g \) |
| a. Calculate Fitness of each node present in this group \( (g_i) \) by using equation 27 |
| b. \( g_{best} = \{ \min \text{ (Fitness (}g_k\text{)), } k \} \text{ - For each node present in this group} |
| End for |
| Gbest = \{ \text{gbest}_i | \text{Fitness (gbest}_i\text{)} = \min \text{ (Fitness (gbest)_i)}, i, 1 < i \leq N_g \} |
| for itr = 1: Maxitr |
| for \( i = 1: N_g \) |
| a. Calculate Fitness of each node present in this group \( (g_i) \) by using equation 27 |
b. Choose the leader node $\alpha$, $\beta$ and $\delta$ based on the best three fitness value in this group

c. Update the position of the best CHs using equations 8, 9 and 10

d. Take an average of the updated CHs position in above step and update the position for best CH

e. Choose the closest sensor node to the position calculated in above step to become a final CH in this group

f. if \( \text{Fitness (Final CH)} < \text{Fitness (gbest)} \)
   
   \[ \text{gbest}_i = \text{Final CH} \]
   
   End if
   
   else
   
   gbest\(_i\) unchanged
   
   End else

g. if \( \text{Fitness (Final CH)} < \text{Fitness (Gbest)} \)
   
   \[ G\text{best} = \text{Final CH} \]
   
   End if
   
   else
   
   G\text{best} unchanged
   
   End else
   
   End for
   
   End for

Stop

7. Simulation and Results

The experimentation of proposed CHGWO based CH selection in WSN is carried out using MATLAB, can provide practicality and effective modeling of WSN. As specified in [15], it uses the same operation model and uses a round operation method. Each round covers two things, primary is clustering and secondary one is data transmission. GWO approach select CHs in the clustering phase from the active sensor nodes, and in the data transmission cluster member transmit its sensed data to respective CHs. Finally, after conducting the aggregation, the CHs transmit that information to the sink node.

In this type of modelling number of rounds is a parameter to compare the performance of the network. The FIRST\(_N\)-DEAD mean that at which round first sensor nodes get die. The HALF\(_N\)-DEAD means that at which round 50 percent of the sensor nodes get die. The network dead mean that 90 percent of the sensors die at which round. Network dead means that WSN eventually runs out of the data collection process when this condition is satisfied. Here, using the three parameters mentioned, the suggested approach is compared with several existing approaches. We also compare the outcomes produced in WSN by various existing CHs selection approaches. Comparing the FIRST-N-DEAD, HALF-N-DEAD, and NET-DEAD values for different algorithms, FIRST-N-DEAD means that the first node dies at what round, while the HALF-N-DEAD and NET-DEAD round values show the round when half of the network nodes die and the round when the whole network dies.

The comparisons were made between five algorithms for scenario numbers 1 to 6. These scenarios are described in Table 6. Table 7 shows the algorithms and corresponding simulation abbreviations, and the generated results are shown using figures 4, 5, 6, 7, 8 and 9.

| Table 6: Considered scenario details |
|--------------------------------------|
| **Scenario** | **Nodes (N)** | **Deployment Area (M* M)** | **BS Location** |
| 1            | 100           | 100*100                     | (50, 50)        |
| 2            | 100           | 100*100                     | (50, 150)       |
| 3            | 100           | 200*200                     | (100, 100)      |
| 4            | 100           | 200*200                     | (100, 300)      |
| 5            | 200           | 200*200                     | (0, 0)          |
| 6            | 200           | 200*200                     | (100, 100)      |
Table 7: Algorithms abbreviation table

| Algorithms                  | Abbreviations |
|-----------------------------|---------------|
| CH-GWO (Proposed)           | W1            |
| GWO-C [35]                  | W2            |
| FIGWO [33]                  | W3            |
| LEACH [15]                  | W4            |
| LEACH-C [15]                | W5            |

Figure 4: Scenario 1 comparison graph

Figure 5: Scenario 2 comparison graph
Figure 6: Scenario 3 comparison graph

Figure 7: Scenario 4 comparison graph

Figure 8: Scenario 5 comparison graph
The results show that with CHGWO, the algorithm has certain advantages when considering 11 attributes. CHGWO always gives the best result for the scenario when we see the results produced by the proposed CHGWO algorithm for scenario (1-6). Current approaches in some cases have decent performance for few comparison parameters, but overall proposed algorithm gives better results for each scenario.

The FIRST-N-DEAD value in one existing algorithm is better than the proposed algorithm in scenario 5, but the values of the other two comparison parameters HALF_N_DEAD and NET_DEAD are worse. In this scenario, the CHs are chosen in a way that, they have the shortest distance from the node but according to the other parameters they are not fine. Our aim is to balance the overall consumption of energy, so we focus in a combined way on all the parameters. For all three FIRST N DEAD, HALF N DEAD and NET DEAD comparison parameters, we need the optimal value. Finally, we can assume that, by using GWO, the proposed algorithm is better in all aspects.

The detailed results review as well as the use of simulation have proven this proposed CHGWO approach gives better results in all aspects. This also decreases the downside of data missing from such specific areas because of the dead nodes. Proposed CHGWO provides balanced energy consumption at all areas, thereby it reduces the probability of dead nodes from one specific area, but other algorithms may encounters with this problem.

8. Conclusion
The GWO is a relatively new technique with a wide variety of possibilities open for its improvement. CHs election and cluster formation approach is discussed in this article. A proper fitness function is built that takes into account important network parameters. For the six typical scenarios of the BS case, the results were compared, and the GWO was found to produce consistently better results compared to the existing GWO-C, FIGWO, LEACH and LEACH-C. Three basic metrics of FIRST_N_DEAD, HALF_N_DEAD, and NET_DEAD were checked for the performance for all six scenarios. In all the cases and in all the metrics, the GWO was able to have better performance. The suggested work for the WSNs with static sensor nodes was idealized, implemented, and evaluated. Furthermore, the work can be applied to networks with mobile sensor nodes, i.e. sensors that can alter their real-time locations.

References
[1]. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E. Wireless sensor networks: a survey. Computer networks. 2002;38(4):393-422.
[2]. Younis O, Krunz M, Ramasubramanian S. Node clustering in wireless sensor networks: recent developments and deployment challenges. IEEE network. 2006;20(3):20-25.
[3]. Abbasi AA, Younis M. A survey on clustering algorithms for wireless sensor networks. Computer communications. 2007;30(14-15):2826-2841.

[4]. Senthil, M., Rajamani, V., & Kanagachid, G. (2014). Energy-efficient cluster head selection for lifetime enhancement of wireless sensor networks. Information Technology Journal, 13(4), 676–682.

[5]. Kaur, H., & Seehra, A. (2014). Performance evaluation of energy efficient clustering protocol for cluster head selection in wireless sensor network. International Journal of Peer to Peer Networks, 5(3), 1–13.

[6]. Noori, M., & Khoshtarash, A. (2013). BSDCH: New chain routing protocol with best selection double cluster head in wireless sensor networks. Wireless Sensor Network, 05(02), 9–13.

[7]. Gupta, V., & Sharma, S. K. (2014). Cluster head selection using modified ACO. Advances in Intelligent Systems and Computing, 1(1), 11–20.

[8]. Kulkarni RV, Forster A, Venayagamoorthy GK. Computational intelligence in wireless sensor networks: A survey. IEEE communications surveys & tutorials. 2011;13(1):68-96.

[9]. Agrawal D, Pandey S. FUCA: fuzzy-based unequal clustering algorithm to prolong the lifetime of wireless sensor networks. International Journal of Communication Systems. 2018;31(2):e3448.

[10]. Ye Z, Mohamadian H. Adaptive clustering based dynamic routing of wireless sensor networks via generalized ant colony optimization. Iieri Procedia. 2014;10:2-10.

[11]. Bara'a AA, Khalil EA. A new evolutionary based routing protocol for clustered heterogeneous wireless sensor networks. Appl Soft Comput. 2012;12(7):1950-1957.

[12]. Singh B, Lobiyal DK. A novel energy-aware cluster head selection based on particle swarm optimization for wireless sensor networks. Human-Centric Computing and Information Sciences. 2012;2(1):13.

[13]. Agrawal, D. and Pandey, S., 2017, November. FLHSCBC: fuzzy logic and improved harmony search based clustering algorithm for wireless sensor networks to prolong the network lifetime. In International Conference on Ubiquitous Computing and Ambient Intelligence (pp. 570-578). Springer, Cham.

[14]. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Advances in engineering software. 2014;69:46-61.

[15]. Heinzelman WB, Chandrasan AP, Balakrishnan H. An application-specific protocol architecture for wireless microsensor networks. IEEE Transactions on wireless communications. 2002;1(4):660-670.

[16]. Younis, O., & Fahmy, S. (2004). HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Transactions on Mobile Computing, 3(4), 366–379. IEEE.

[17]. Baker, D., Ephremides, A., & Flynn, J. (1984). The design and simulation of a mobile radio network with distributed control. IEEE Journal on Selected Areas in Communications, 2(1), 226–237. IEEE.

[18]. S. Younis, S. Fahmy, “Distributed Clustering in Ad-hoc Sensor Networks: A Hybrid, EnergyEfficient Approach”, in Proceedings of IEEE INFOCOM, March, Hong Kong, China, 2004.

[19]. W. C. Y. Lee, Mobile Cellular Telecommunications. McGraw Hill, 1995.

[20]. Sakya, S. (2020). Design of Hybrid Energy Management System for Wireless Sensor Networks in Remote Areas. Journal of Electrical Engineering and Automation (EEA), 2(01), 13-24.

[21]. Ding, P., Holliday, J. A., & Celik, A. (2005). Distributed energy-efficient hierarchical clustering for wireless sensor networks, distributed computing in sensor systems (pp. 466–467). Berlin: Springer.

[22]. Demirbas, M., Arora, A., & Mittal, V. (2004). FLOC: A fast local clustering service for wireless sensor networks. Workshop on dependability issues in wireless ad hoc networks and sensor networks (DIWANS/DSN 2004) (pp. 1–6).

[23]. Ye, M., Li, C., Chen, G., & Wu, J. (2005). EECS: An energy efficient clustering scheme in wireless sensor networks. In 24th IEEE international performance, computing, and communications conference IPCCC 2005 (pp. 535–540). IEEE.

[24]. Yuan, H.-Y., Yang, S.-Q., & Yi, Y.-Q. (2011). An energy-efficient unequal clustering method for wireless sensor networks. In 2011 international conference on computer and management (CAMAN) (pp. 1–4). IEEE.

[25]. Kumar, D., Aseri, T. C., & Patel, R. B. (2009). EEHC: Energy efficient heterogeneous clustered scheme for wireless sensor networks. Computer Communications, 32(4), 662–667. Elsevier.

[26]. Qureshi, T. N., Javaid, N., Khan, A. H., Iqbal, A., Akhtar, E., & Ishfaq, M. (2013). BEENISH: Balanced energy efficient network integrated super heterogeneous protocol for wireless sensor networks. Procedia Computer Science, 19, 920–925. Elsevier.

[27]. Yi, S., Heo, J., Cho, Y., & Hong, J. (2007). PEACH: Power-efficient and adaptive clustering hierarchy protocol for wireless sensor networks. Computer Communications, 30(14), 2842–2852. Elsevier.

[28]. Camilo, Tiago, et al. "An energy-efficient ant-based routing algorithm for wireless sensor networks." International Workshop on Ant Colony Optimization and Swarm Intelligence. Springer Berlin Heidelberg, 2006.

[29]. Zeng, Bing, and Yan Dong. "An improved harmony search based energy-efficient routing algorithm for wireless sensor networks." Applied Soft Computing 41 (2016): 135-147.
[30]. Jin, S., Zhou, M., & Wu, A. S. (2003). Sensor network optimization using a genetic algorithm. In Proceedings of the 7th world multiconference on systemics, cybernetics and informatics (pp. 109–116).
[31]. Ferentinos, K. P., Tsiligiridis, T. A., & Arvanitis, K. G. (2005). Energy optimization of wireless sensor networks for environmental measurements. In Proceedings of the international conference on computational intelligence for measurement systems and applications (CIMSA)51 (pp. 1031–1051).
[32]. Lee, D., Lee, W., & Kim, J. (2007). Genetic algorithmic topology control for two-tiered wireless sensor networks. Computational Science-ICCS, 2007, 385–392. Springer.
[33]. Zhao X, Zhu H, Aleksic S, Gao Q. Energy-efficient routing protocol for wireless sensor networks based on improved grey wolf optimizer. KSII Transactions on Internet & Information Systems. 2018;12(6):4.
[34]. Sridhar, P., Madni, A. M., & Janshidi, M. M. (2008). Multi-criteria decision making in sensor networks. IEEE Instrumentation and Measurement Magazine, 11(1), 24–29.
[35]. Agrawal, D., Wasim Qureshi, M. H., Pincha, P., Srivastava, P., Agarwal, S., Tiwari, V., & Pandey, S. (2020). GWO-C: Grey wolf optimizer-based clustering scheme for WSNs. International Journal of Communication Systems, 33(8), e4344.
[36]. Rajpoot, P., & Dwivedi, P. (2019). Multiple parameter based energy balanced and optimized clustering for WSN to enhance the Lifetime using MADM Approaches. Wireless Personal Communications, 106(2), 829-877.