IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Clustering Using Firefly Algorithm for the Wireless Sensor Network

[1]Dhirendra, [2]Pankaj Kumar Mishra

[1]PG Student, [2]Assistant Professor
[1]Department of Computer Engineering,
[1]GBPUAT, Pantnagar, India

Abstract: Wireless sensor networks have attracted the attention of researchers in recent decades. Due to the rapid change in the technology along with faster-growing usefulness and need in different areas. Thus, the size of the sensors became very small. Sensor's basic operations are sensing, processing, and transferring data to the base station. These operations require energy especially the transmission. Since they have very limited power, that is too difficult to recharge or replace in some Nature-inspired algorithm scenario. So, for uninterrupted services, the utilization of the battery power becomes a more critical concern. In this research paper, for dealing with the problem of better energy utilization a new clustering method has been developed. The nature-inspired Firefly algorithm has been used for the cluster head selection based on nodes centrality, energy, and distance to the base station. Firefly algorithm is proved to be a better technique due to faster convergence. This proposed method has been evaluated against the Low Energy Adaptive Clustering Hierarchy (LEACH) protocol, at various performance metrics. The results show the stability period increased by 23 % and the mortality rate decrease. Also, the overall network life is increased by 18 %. It performs very well results show that it is beneficial to use the Firefly algorithm since it improves the performance of the wireless sensor network.

Index Terms **- Cluster Head Selection, Clustering, Firefly algorithm, Leach, Wireless Sensor Networks.**

1. Introduction

Wireless sensor networks (WSNs) are a widely versatile area of technology. It is continuously being developed and it is very handy to use. It serves various purposes, from monitoring on the battlefield as a surveillance system [1], to medical science as the body monitoring system [2]. It enhances performance and computation by forming the distributed and light-weighted network of sensors. The micro-electro-mechanical system [3], combined with digital electronics ensures the production of light-weighted, tiny, low-cost, low energy, and an unattended sensor. These sensors having the capability of sensing, data processing, and communication. Although they are short-range communication devices, due to their low cost and lightweight, they can be easily deployed. The sensor nodes are deployed where they are needed or near the target, in a large amount. They can be very useful in areas where the condition is not quite good. One of the benefits of using a WSN is that it doesn't have to deploy in an organized or predetermined way. That means protocols using for this kind of network should be self-organizing. There are three main functions of these sensor nodes, first is to monitor their environment, second is computation and processing of the data, and third is to transmit the data to their base station where all the information is collected. Since they are battery operated the use of their battery power wisely is a very critical part, that is the reason the sensors compress and aggregate data before forwarding to the base station because transmission of data requires more energy [4]. And also, replicated and unnecessary data is not transmitted. Only the processed or partially processed data is forwarded that saves energy and cuts off the burden. The WSN having various types like a seismic, thermal, acoustic, and infrared sensor, etc. They have also various applications such as temperature monitoring, humidity, pressure, soil moisture, object detection, noise level, and speed detection. They can be used as a security system in the house or farm, surveillance system on the battlefield, fire detection system in the forest, health monitoring system in the hospitals, communication system at the time of disaster [5].

2. Related work

The process of clustering and the selection of cluster head (CH) through the nature-inspired algorithm is not very new various authors proposed different types of approach for these, out from these works Altakhayneh proposed an algorithm for CH selection by using a Genetic Algorithm (GA) [6]. For comparison, the author uses LEACH [7]. In this algorithm node with the most energy is opted as a CH using a GA. Node energy, CH energy, the distance of nodes from CH, the distance of entire CHs from the sink, and participants in clusters are work as inputs for the Algorithm. Vijayalakshmi proposed an effective way for CH selection by using the multi-objective tabu with the Particle Swarm Optimization (PSO) algorithm [8]. It is a hybrid technique where PSO is combined with tabu search to get the benefit of each other and found a beneficial solution. The local finest position of particles is found using PSO and the information of the global finest position serves as input for the Tabu record. The swapping of the route is done here and from the Tabu list node with the best fitness value and the minimum hop route is selected as CH. This approach improves the routing, besides CH selection, also increases the network lifespan. Ahmad presents a technique for CH selection by using the Artificial Bees Colony (ABC) Optimization technique [9]. In this algorithm, it is assumed that clusters are already formed.

Each cluster holds a CH, which is selected by ABC optimization. It is also observed in that approach for best results the best position for the base station is the center of the network. Gupta proposed an integrated approach for the clustering along with routing as well, based on Cuckoo Search (CS) and Harmony Search (HS) techniques [10]. All the procedure of clustering is executed on the Base Station (BS). For finding the best positions of CHs improved CS algorithm was applied. In the CS algorithm, the multi-objective function for fitness is run for finding the group of optimal positions for CHs. In the function residue energy of node, its degree, intra-cluster distance, with the CH range considered as the parameters. After CH selection the Cluster formation is done. Then the route discovery process is started by using the upgraded HS algorithm. Each node analyzes hop count, sink distance from the node. Intermediate nodes can be either a cluster or non-CH. Then each node calculates the probability for the next node in its route. Priya proposed an approach for thermal aware CH election and clustering [11]. That involves the Fuzzy and Spider Optimization Algorithm (SOA). The network area is shared into the Fuzzy subsets according to nodes location, the clusters are created by the Fuzzy algorithm using PageRank, and controlling the area of the clusters is depend on the Eigen centrality and cut set method. For the objective function average energy and distance from members to CH, overall energy expenditure, and overall network temperature rise is considered as a parameter. While on the contrary, CHs elected using the SOA, deliberate the temperature and the distance of every node into the cluster. Subramanian presented a hybrid methodology by combining the Grey Wolf Optimization (GWO) with Crow Search Optimization (CSO) for picking CH optimally [12]. The GWO strategy is to change the place of the agents into the search distance, this will lead the premature convergence. To vanquish this limitation, it is merged with the CSO algorithm. GWO was used in the discovery of the potential location for CHs. While CSO helps in locating the optimal CH. The node with the top fitness value and the secondary node with maximum fitness is potential nodes for CHs. This proposed method improved the mean residue energy and decrease the packet latency. Visu proposed a dual CH optimized routing algorithm that uses Krill Herd Optimization [13]. The network area is initially clustered using a k-means centroid algorithm. The primary centroid collects the analyzed data and the secondary centroid provides for each path the detail about Path trust value. The initial centroid node which performs as aggregator and router is only allowed to aggregate data, a node with the utmost residue energy becomes the CH and aggregate data and after aggregation, this data is passed on to the subsequent CH. Using the path trust value, the Krill herd optimization algorithm optimizes the path into the network. This paper deals with the Firefly algorithm [14], for CH selection that minimizes the energy consumption and increases the network's life, as well as good transmission, is achieved.

3. MATERIALS AND METHODS

3.1 Network Model

In the simulation of the proposed work, we deployed 100 sensor nodes randomly in the 100 m X 100 m area. The base station is located in the center (50 m, 50 m) of the area. And the location of sensors can be mapped into x and y-coordinates that are measured into meters. The cluster created in the beginning and will not change throughout the complete operational time. Figure 1 depicted the simulation field.

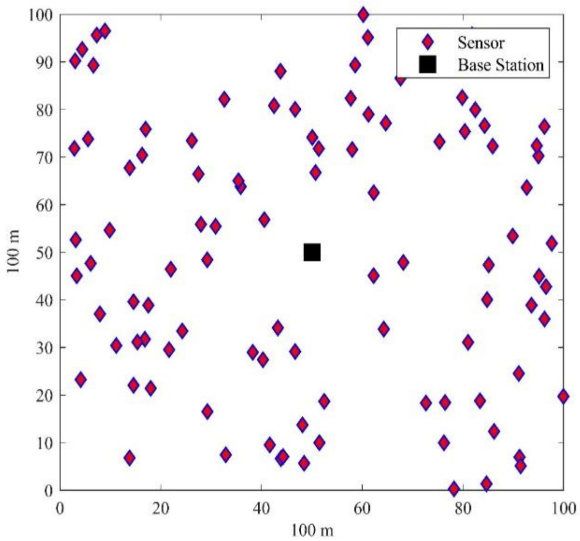

Fig. 1: Wireless Sensor Network Field

3.2 Assumptions

Some assumption is taken into consideration before starting before simulation as feature given below:

- All nodes are static that it will not change their location after the deployment. That means the position of nodes will not change in the entire operation.
- All the nodes are homogeneous and their initial energy is also the same but not limited as they run on battery power.
- Each node has its unique id. So, it can be identified for their role.
- Member nodes of a cluster transmit their data directly to their CH. That means only one hop communication will take place.
- The cluster head directly transmits data to the BS.
- The BS is positioned into the center of the field and has unlimited energy.

3.3 Radio Energy Model

The first-order radio energy model is used for energy consumption when communication takes place between the sensor nodes. For transmitting *k* bits to the distance between two nodes, energy consumed by the transmitter is according to the equation

$$E_{TX}(k,d) = \begin{cases} kE_{elec} + k\varepsilon_{fs}d^2, d < d_0 \\ kE_{elec} + k\varepsilon_{mp}d^4, d \geq d_0 \end{cases} \quad (1)$$

And the energy consumption is by the receiver is done according to the equation

$$E_{RX}(k) = kE_{elec} \quad (2)$$

where:

E_{TX} = transmitter energy dissipation
E_{RX} = receiver energy dissipation
E_{elec} = energy consumption per bit
ε_{fs} = free space path loss model
ε_{mp} = multipath fading model
d_0 = threshold value

E_{TX} is the transmitter and E_{RX} is the receiver, energy dissipation. E_{elec} are energy consumption per bit to run transmitter or receiver circuit, and free space (ε_{fs}) path loss model and multipath (ε_{mp}) fading model is the energy coefficient for different propagation channels for energy consumption during transmitting and receiving the 1-bit data and d_0 is the threshold value denoted as:

$$d_0 = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}} \quad (3)$$

to differentiate the free space path model from a multipath fading model.

3.4 Firefly Algorithm

It is developed by Dr. Xin-She Yang at Cambridge University in 2007. It is a metaheuristic approach that mimics the conduct of the fireflies flashing. That flashing serves as a communication medium for fireflies. As we are familiar with the light intensity from source to any given distance follows the inverse square rule. That means the lesser the distance brighter the firefly appears and the Attractiveness of the firefly depends on its brightness, fireflies with superior brightness tend to attract other fireflies irrespective of their sex. The purpose of this methodology is to find out the optimum global and local solutions in the search field that results in the best solution for a given fitness function. The three basic rules are considered to illustrate the idea of this algorithm:

- All the fireflies are unisex, whatever their gender, they attract others.
- The attractiveness of Fireflies directly proportionates to brightness, the lesser bright will attract towards the brighter. Attractiveness along with brightness increases or decreases according to the distance among them. And if any of them not dazzling more than a specific firefly then it will travel arbitrarily.
- The brightness is determined and influenced by the considered environment.

In simple cases, the attractiveness (β) of a firefly is determined by its brightness which is acquired from the objective function (*F(x)*). That means the brightness (*I*) of a firefly at the location *x* can be determined by $I(x) \propto F(x)$, although β is relative as we knew light intensity decreases with a larger distance, and light is also absorbed by the medium so the attractiveness varies. We can calculate the light intensity based on distance (*r*) for a given medium where the light absorption coefficient is as

$$I = I_0 e^{-\gamma r} \quad (4)$$

Where:

I_0 = original light intensity

By approximation using Gaussian law we have

$$I(r) = I_0 e^{-\gamma r^2} \quad (5)$$

Now attractiveness (β) of firefly can be calculated as

$$\beta(r) = \beta_0 e^{-\gamma r^2} \quad (6)$$

Where β_0 is attractiveness at distance *r* = 0

3.5 Proposed Methodology

A substantial number of sensors are deployed in the area of study. And randomly generate coordinates are allocated to every sensor for their location in the network and also an equal amount of energy allotted to all sensors. After the initialization process network area is partitioned into the desired number of clusters using the K-means clustering algorithm. For the partition, the K-means algorithm used squared Euclidean distance here, each cluster has a different number of nodes. Here, in this work we want each cluster to have one CH, and 10% of all sensor nodes will become CH in every round.

In each cluster, the Firefly algorithm is applied to get the optimum location for the CH. For applying the firefly algorithm on a particular cluster, fireflies' location is initialized by the location of nodes of that cluster, the number of fireflies will be the same as the number of nodes in that cluster, and to get the light intensity for these fireflies, a cost function is used. For any Firefly "*i*" cost function is given below, where "*i*" is the id no. of a particular Firefly:

$$Cost_i = E_i * W_1 + D_i * W_2 + C_i * W_3 \quad (7)$$

Where:

E_i = remaining energy of node *i*
D_i = distance from the base station
C_i = centrality of a node *i*
W_1, W_2, W_3 = weights

The Sum of all the weights must be 1. These are given as input, and the output from the cost function becomes the intensity values of these fireflies. After that, the firefly algorithm runs, until the termination condition is reached. It is ensured in each iteration that the movement of fireflies remains within the boundary of that cluster. After the termination of the firefly algorithm, results are evaluated and the best firefly mapped into the cluster, if the node is not selected in the previous round and has energy then it will be selected as CH otherwise second-best node is selected after applying these conditions. If a node is selected as a CH then it will not be selected until all the nodes become CH ones.

3.6 Evaluation Parameters

For evaluating the performance of the proposed work, some performance metrics are used. From the study of the previous works, it is observed that these metrics that are mentioned below, are a base for determining the efficiency and it works as resilient for showing the findings of work. The opted parameters are defined as below:

3.6.1 Stability period

The stability period can be described as "The period between the starting of operation and the death of the first sensor node of the network." This is also known as the stable region. It is a crucial parameter. This period should be as long as possible for a reliable network. A long stability period shows how well-balanced energy has been utilized.

3.6.2 Half node dead period

The Half Node Dead (HND) period can be stated as "The time between the starting of operation, to the time when half number of total nodes dies." It is also an important parameter to keep track of the energy consumption of nodes, and for how much time, it can provide reliable information.

3.6.3 Network lifetime

It is an important parameter for the performance evaluation of any clustering protocol as the core objective of clustering is the growth of the lifetime of the network. Network lifetime can be described as "The overall time of operation from start to end until the energy of all the sensor nodes of the network has exhausted completely."

3.6.4 Number of nodes alive per round

It is used to determine the alive nodes in every round or the other words those nodes that have not yet expended their complete energy. The number of Alive nodes indicates that they can be used for the further completion of their task.

3.6.5 Number of nodes dead per round

This parameter defines the number of dead nodes in every round that cannot be used further to accomplish any task, as they drained out their energy completely.

3.6.6 Throughput

This parameter represents the number of packets transmitted to the Base station that received successfully. Every time a packet is transmitted, it consumes energy. So, the lifetime of the network is related to throughput. Long network life means high throughput and vice versa is also true.

4. RESULTS AND DISCUSSION

The performance after applying the firefly algorithm against the LEACH algorithm was evaluated. The modified algorithm used for the proposed work is implemented on MATLAB. The parameters considered [15], are shown in Table 1.

Table 1: Parameter Values

Symbol	Parameter	Quantity
n	Total number of nodes	100
E_0	The initial energy of each node	0.5 J
E_{elec}	Transmitter and Receiver energy	50 nJ/bit
ε_{fs}	Amplification coefficient of free space	10 pJ/bit
ε_{mp}	Amplification coefficient of multipath	0.0013 pJ/bit
E_{DA}	Data aggregation energy	5 nJ/bit
k	Data packet size	4000 bits

4.1 Comparison of the stability period

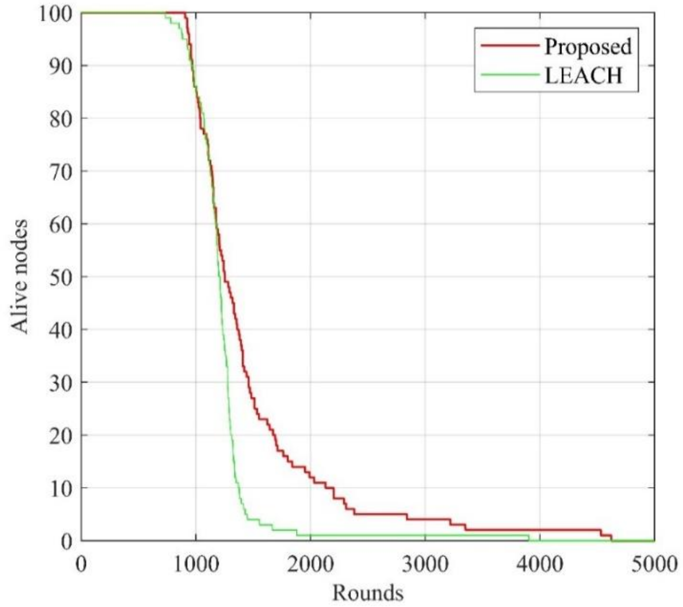

Fig. 2: Comparison of alive nodes (Number of rounds vs Alive nodes)

The result for the stability period and alive nodes is shown in Figure 2. Where the x-axis is labeled as the number of rounds and the y-axis represents the number of nodes that are alive. The changing slopes of the curve of the alive nodes are illustrated in the figure. shows that the nodes in the proposed method remain alive for a longer period compared with the LEACH method, the curve slopes of the proposed method are smaller which shows that the dying process is relatively slower. The first node dead (FND) is happening in 906 rounds in the proposed method while in the LEACH curve it can be seen that the first node dies early at 734 rounds. The first node indicates the end of the stability period. So, from the figure, it can be seen that the proposed work increases the stability period. This is the result of the consideration of energy, centrality, and distance in the proposed algorithm, which utilizes the energy consumption and prevents deplete their energy too early.

4.2 Comparison of half-dead nodes

The comparison death of half number of total nodes is illustrated in Figure 3. The x-axis denotes rounds while the y-axis denotes the number of dead nodes. From the comparison, it can be noticed that 50% of nodes died in both the algorithm at 1200 rounds in LEACH and 1251 rounds in the proposed work respectively. The slope of the graph shows the quick death of nodes in the LEACH algorithm after a point speed of nodes death increased rapidly while in the proposed work speed for the death of nodes is slightly slow. Also, the time for half amount of node death is increased in the proposed work which is a sign of wise selection of CHs and efficient utilization of nodes energy.

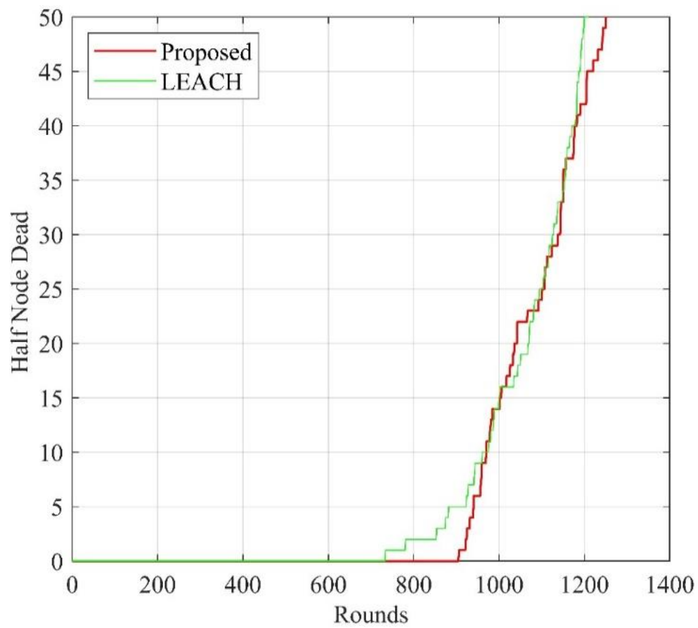

Fig. 3: Comparison of Half node dead (Number of rounds vs Half node dead)

4.3 Comparison of dead nodes and network lifetime

The experiment result for the dead node per round is shown in Figure 4. It can be evaluated that the proposed algorithm outperforms the LEACH algorithm. In the figure, the x-axis represents the number of rounds while the y-axis represents the number of dead nodes. The proposed algorithm can prevent nodes to die for a longer period than LEACH. The rapid growth in the curve of LEACH represents the faster death of nodes as it selects the CHs on the probability basis if an unlucky node is elected for the CH, the energy of that consumed fastly as the responsibility of data gathering from neighbor nodes added. And it will cost additional energy expense and creates a load on it. The last node died (LND) at the 4623 rounds of proposed work but the LEACH last node dies very early at 3904 rounds in comparison. The death of the last node indicates the end of the operation and represents the lifetime of the network. It can be seen that the proposed algorithm increases the network lifetime very efficiently. The changing slope of the curve of the alive nodes is illustrated in the figure. Compared with the LEACH method, the curve slopes of the proposed method are smaller which shows that the dying process is relatively slower. This is the result of the consideration of energy, centrality, and distance in the proposed algorithm, which utilizes the energy consumption to every node efficiently, and ensuring that nodes will not deplete out of their energy too early.

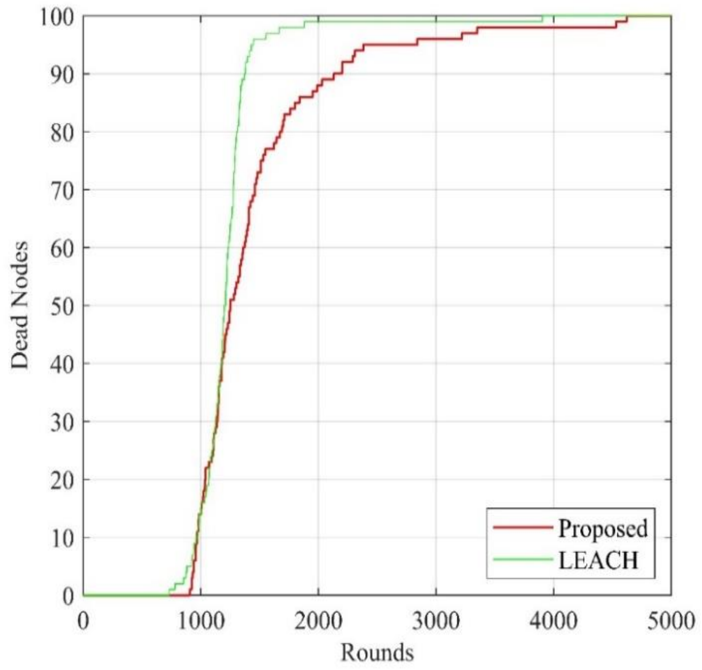

Fig. 4: Comparison of dead nodes (Number of rounds vs Alive nodes)

4.4 Comparison of throughput

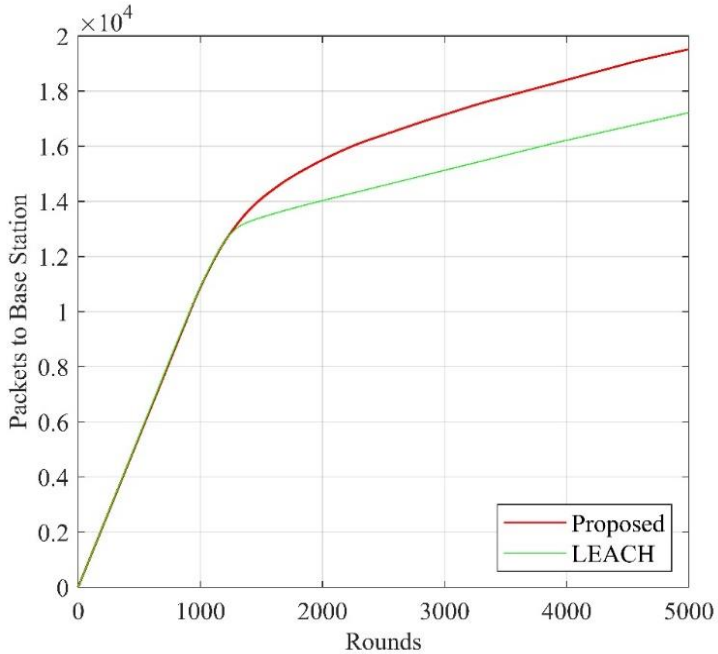

Fig. 5: Comparison of Packets send to Base Station (Number of rounds vs Packets to BS)

The simulation result for the comparison of packets received successfully by the sink is shown in Figure 5. the x-axis in the figure represents the number of rounds and the y-axis represents the number of packets received by BS. The reception of packets by BS is called throughput. The figure shows the comparison of packets that are sent by the proposed work and the LEACH algorithm. It can be noticed that 1.9×10^4 packets were received in the proposed algorithm and 1.7×10^4 packets were received at the BS through the LEACH algorithm. The proposed method transfers more packets to the BS than the LEACH algorithm. The high number of packets indicates a longer life with good communication. For the accuracy and the integrity of information, a high number of data is essential. The high amount of sense data helps in better analysis of the field. And allows the end-user for taking appropriate action according to it. So, from the figure, it can be evaluated that the throughput of the proposed work is increased.

Table 2: Conclusive results

Protocol	FND	HND	LND
LEACH	734	1200	3904
Proposed	906	1251	4621

From Table 2 it can be analyzed that our proposed firefly algorithm method outperforms the traditional LEACH algorithm, whether it is the case of stability period, time of First Node Dead, Half Node Dead time, or whole networks life till Last Node Dead.

5. Conclusions

For the evaluation of this work different parameters are considered on which the proposed method performs better than the LEACH algorithm. The simulation results show the efficiency of the proposed method, it increases the stability period by 23.43% and the number of half-dead nodes by 4.25%, also increases the lifetime of the network by 18.36%. The increment in the stability period indicates that the selection of CHs is being done efficiently compared to the LEACH. The packets received to the base station are also increased. High packet reception means there is less energy consumption and also an increased network lifetime with high throughput.

References

[1] M. Usman, M.A. Jan, X. He, Cryptography-based secure data storage and sharing using HEVC and public clouds, *Inf. Sci.* Vol. 387, pp. 90-102, 2017.

[2] P. Halapeti, S. Patil, Healthcare monitoring system using wireless sensor networks, *Int. J. Adv. Res. Comput. Sci. Technol., Vol.* 2, pp. 443-446, 2014.

[3] L.P. Clare, G. Pottie, J.R. Agre, 1999. Self-Organizing Distributed Sensor Networks. In *SPIE Conference on Unattended Ground Sensor Technologies and Applications* (pp 229–237).

[4] Y. Wang, P. Shi, K. Li, Z. Chen, An energy efficient medium access control protocol for target tracking based on dynamic convey tree collaboration in wireless sensor networks. *Int. J. Commun. Syst.*, Vol. 25, pp.1139-1159, 2012.

[5] F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless Sensor Networks: A Survey, *Elsevier Sci. B. V. Comp. Networks.* Vol. 38, pp. 393–22, 2002.

[6] W.A. Altakhayneh, M. Ismail, M.A. Altahrawi, M.K. AbuFoul, 2019. Cluster Head Selection Using Genetic Algorithm in Wireless Network, In *2019 IEEE 14th Malaysia International Conference on Communication (MICC), Selangor, Malaysia* (pp 13-18).

[7] W.R. Heinzelman, A. Chandrakasan, B. Hari, 2000. Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In *Proceedings of the 33rd Annual Hawaii International Conference on IEEE* (pp 1-10).

[8] K. Vijayalakshmi, P. Anandan, A multi-objective Tabu particle swarm optimization for effective cluster head selection in WSN. *Cluster Comput.*, Vol. 22, pp. 12275-12282, 2019.

[9] T. Ahmad, M. Haque, A.M. Khan, 2018. An Energy-Efficient Cluster Head Selection Using Artificial Bees Colony Optimization for Wireless Sensor Networks. In *Advances in Nature-Inspired Computing and Applications. EAI/Springer Innovations in Communication and Computing. Springer, Cham* (pp 189-203).

[10] Govind Gupta, Sonu Jha, Integrated clustering and routing protocol for wireless sensor networks using Cuckoo and Harmony Search based metaheuristic techniques, *Eng. Appl. Artif. Intell.*, Vol. 68, pp. 101-109, 2018.

[11] R.K. Priya, S. Venkatanarayanan, Implementation of thermal aware wireless sensor network clustering algorithm based on fuzzy and spider optimized cluster head selection. *J. Ambient Intell. Human. Comput.* Vol.11, pp. 1-11, 2020.

[12] P. Subramanian, J.M. Sahayaraj, S. Senthilkumar, A Hybrid Grey Wolf and Crow Search Optimization Algorithm-Based Optimal Cluster Head Selection Scheme for Wireless Sensor Networks, *Wireless Pers. Commun.*, Vol. 113, pp. 905-925, 2020.

[13] P. Visu, T.S. Praba, N. Sivakumar, Bio-inspired dual cluster heads optimization routing algorithm for wireless sensor networks, *J. Ambient Intell. Human. Comput.* (2020).

[14] X.S. Yang, Firefly algorithm, *Nature-Inspired Metaheuristic Algorithms*, 2nd ed. (Luniver Press, UK, 2010). pp. 81-89.

[15] Q. Ni, Q. Pan, H. Du, C. Cao, Y. Zhai, A Novel Cluster Head Selection Algorithm Based on Fuzzy Clustering and Particle Swarm Optimization, *IEEE/ACM Trans. Comput. Biol. Bioinform*. Vol. 14, pp. 76-84, 2017.