International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Energy Efficient Data Aggregation Scheme using Improved LEACH Algorithm for IoT Networks

Guguloth Ravi[1], **M. Swamy Das**[2], **Karthik Karmakonda**[3]

Submitted: 24/10/2022 **Revised:** 22/12/2022 **Accepted**: 23/01/2023

Abstract: The Internet of Things makes it possible to have connected buildings, businesses, and intelligent homes by merging embedded technology, wireless sensor networks, control and automation technologies, and wearable gadgets. It is critical to regularly monitor the energy usage of the Internet of Things network since sensor nodes have limited power. In wireless sensor networks, the most significant obstacle is the exhaustion of available energy, and extending the network's lifetime can be accomplished by lowering the amount of energy that is spent. Energy aware routing protocol is highly important in IoT-based networks, although routing protocol that simply considers energy parameter has not performed successfully in managing excessive energy consumption. Energy aware routing protocol is very important in IoT-based network. The emergence of congestion in network nodes results in an increase in the amount of energy consumed and the loss of packets. Routing algorithms should strive for energy efficiency and load balancing across diverse nodes in order to lengthen the lifetime of a network. This will allow for more nodes to participate in the network. Clustering is one of the optimal techniques for efficient data aggregation among the sensor nodes. In a clustered setup, the Internet of Things (IoT) network is partitioned into a predetermined number of smaller networks. One of the most common and widely used clustering methods is LEACH, which stands for low-energy adaptive clustering hierarchy. It is unfortunate that it has some restrictions. In this study, we suggest the use of CAW-LEACH (CONGESTION AWARE - LEACH) as a means of enhancing energy efficiency, CH stability, and the capacity to aggregate data without experiencing congestion. The enhanced protocol that has been proposed takes into account both the depletion energy ratio (DE) and the expected remaining energy (PRE) of the nodes while selecting CH and generating random numbers. Its purpose is to ensure that the CH node that was just recently elected will not be given a second chance in this round. This technique establishes a correlation between the threshold that is utilised in conventional LEACH and each node's energy consumption ratio. The proposed congestion aware data aggregation scheme aggregates the data through traffic free paths by estimating the congestion indicator (CIN) & link error rate (LER), Residual Energy (RE) of the all-available routing paths. By comparing the suggested method to other energy-efficient data aggregation schemes, according to the findings of the experiments, the proposed technique increases the network's durability.

Keywords: *WSN, energy consumption, data aggregation, LEACH, link error rate, congestion indicator, clustering, CAW-LEACH.*

1. Introduction

Many IoT applications also provide access to highly developed software and communication services, in addition to Internet connectivity. The Internet of Things (IoT) is a network of devices, appliances, and other items that were not previously connected to one another. There are many examples of where IoT has been put to use, including in projects that aim to boost computer and network efficiency [1]. Self-configuring wireless networks connect everything that can be connected in a wireless network. There is something in this network that serves as a communication link. In their present form, these open lines provide a wide range of possible methods of contact. Radio-frequency identification chips are the backbone of the Internet of Things (RFID). A Wi-Fi layer, which can be found at the internet's very core, is being used to build out the worldwide infrastructure for RFID tags. The network allows for the exchange of information between the various devices and computers that are linked to one another. These components are a part of a larger, more intricate system. Data is gathered from a variety of sensors to aid in taking temperature readings and other measurements in the immediate vicinity of the sensor. Data is transmitted to nearby sensors so it can be analysed and interpreted in accordance with the requirements of the currently active applications [2].

Many different routing techniques are implemented in order to cut down on the amount of data transfer and, as a result, the required amount of power [3]. When using a WSN, data are transmitted from the sensor node to the

---

[1] *g.raviraja@gmail.com, Research Scholar, Department of Computer Science and Engineering, University College of Engineering (UCE), Osmania University (OU) ,Asst.professor in MRCET,Hyderabad,Telangana,India.*

[2] *msdas_cse@cbit.ac.in, Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad,Telangana,India.*

[3] *Karthikk_cse@cbit.ac.in, Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad,Telangana,India.*

base station, which is often situated in the geographic centre of the network. If the sensor node is placed a significant distance from the base station, then the data packet will need to travel a long distance, which will result in a higher amount of power consumption. In addition, the sensor that is placed near the base station, which is referred to as the bottle neck zone, is subject to significant traffic flow, which quickly depletes the energy stored in the node's battery. Therefore, the nodes in the bottleneck zone die off more quickly, which has an impact on the longevity of the network as well as its connectivity [4].

Routing protocols that simply take energy into consideration as one of their parameters are inefficient. Utilizing a wide variety of parameters not only helps reduce energy consumption but also makes routing protocol more effective [5]. When dealing with various applications, one must take into account a variety of aspects. Management of congestion is considered to be one of the most critical aspects. The network uses more energy when congestion sets in, which is a negative energy-cycle outcome [6]. Congestion can arise in networks for a variety of different reasons. A lack of available storage capacity in the nodes that make up the relay network is one of the primary causes. Congestion happens if a node receives more packets than it can handle, and when this happens, many of the received packets are lost. In wireless sensor networks, congestion developed for almost the same reasons as in wired sensor networks. Congestion can develop, for instance, when multiple nodes all at once make the decision to send packets via a common media [7].

The practise of clustering is one of the promising and effective strategies that can be used to improve energy efficiency [8]. The construction of clusters and the use of a variety of communication channels for the purpose of data transmission have been the approaches that have received the most attention. Figure 1 presents a schematic representation of the conventional clustered network topology.

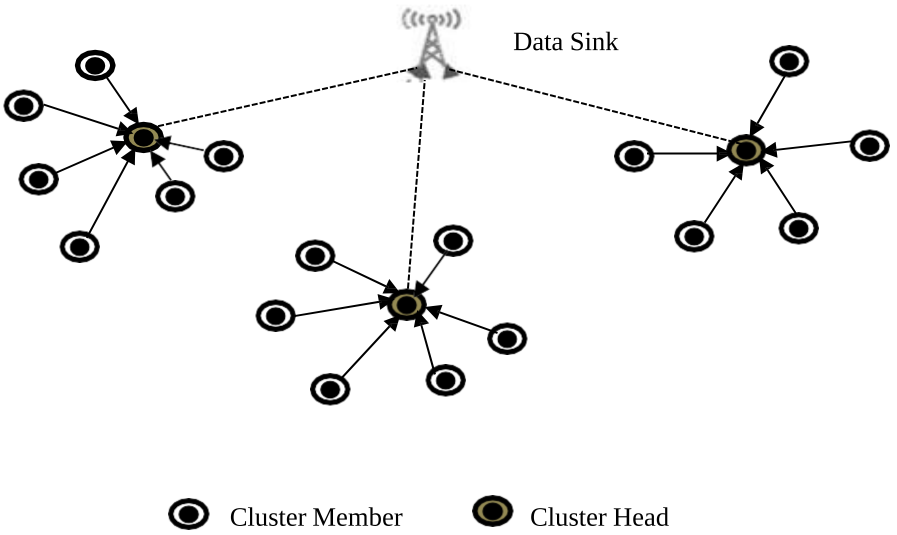

Fig 1: Clustered WSN network topology

Cluster-based routing techniques, in comparison to non-clustering routing protocols, are able to make more effective use of the sensor nodes that are present in the network. One of the responsibilities of a cluster leader, also known as the cluster head (CH), is the elimination of correlated data, which can result in a reduction in the total volume of data [9]. After that, the CH will send the BS the accumulated data that it has just finished processing. Cluster-based routing protocols separate sensor nodes into a large number of clusters with the goal of reducing the amount of energy needed for long-distance communication. Due to the considerable difference in energy depletion between CHs and other nodes, clustering can help reduce the overall amount of energy that is consumed and maintain a workload that is equitable across all of the nodes. Therefore, clustering is a solution that is good for the environment and energy efficiency, as it extends the lifespan of networks and improves energy efficiency. In addition, the vast majority of clustering algorithms make use of optimal CH selection in order to forestall the untimely demise of the sensor nodes and further extend the network's lifetime [10].

Low Energy Adaptive Clustering Hierarchy Protocol

According to [11], the LEACH protocol was the first clustering-based routing method to provide scalability and extend the life of a network. LEACH allows for a notable drop in worldwide energy consumption by continuously dispersing network load to all nodes in varied regions. In a typical setup, clusters are used to organise sensor nodes hierarchically, with one node serving as the CH for its cluster. As the group's central hub (CH), your task is to gather information from the other nodes in your cluster, organise it into reports, and send them off to the sink node. The LEACH algorithm selects a node as the CH if its probability, given by a random number between 0 and 1, is less than a certain threshold, denoted by the symbol TH(n) (Eq 1). In order to join a certain cluster, the remaining nodes pick the CH that requires the least amount of communication energy to reach. By doing so, they are able to join the rest of the cluster. CH's job include switching between all of the sensors to spread out the battery drain.

$$TH(n) = \begin{cases} \frac{P}{1-P\left[\left(r \bmod \frac{1}{P}\right)\right]}, & if\ n \in G \\ 0, & otherwise \end{cases} \qquad \text{Eq (1)}$$

In the equation that was just shown, $TH(n)$stands for the threshold value of n nodes, and P is the abbreviation for the probability value. Every node in the network selects a random integer between 0 and 1, starting with 0.

LEACH operates in iterative rounds, with the first phase of each round being the setup phase, which is in charge of clustering the nodes together [12]. After this, a steady-state phase occurs, during which data is sent to the sink node. In order to cut down on unnecessary overhead, the steady-state phase is significantly longer than the setup phase. Once a node has reached the conclusion that it will function as a CH during the setup phase, it will proceed to transmit an advertisement message. After receiving this message, every node that is not already part of a cluster will make the choice to become part of one of the CHs. Each participating node must participate in data transmission during the steady-state phase. The CH compiles data from all of the cluster nodes and transmits it to the final destination sink. Because of its decentralised nature, LEACH doesn't need any kind of control data from the sink. Furthermore, LEACH can function without any nodes being pre-aware of the global network.

The LEACH protocol makes it possible to keep the network running for longer. On the other hand, there are significant flaws in the new LEACH-based protocols that need to be fixed before they can be considered truly effective. The following are a few of these issues, as well as the answers to them.

- LEACH makes the assumption that all nodes are capable of communicating with the sink, which limits the scalability of the protocol. Incorporating multi-level clusters and providing support for multi-hop routing are two potential solutions to this problem.
- The overhead that is involved, which is caused by fluctuations in CH, results in an inefficient use of energy. One solution to this issue is to cut down on the number of rounds that must be completed during the cluster rebuilding phase.
- The likelihood of picking CHs does not take into account the energy that is still present in the nodes. As a result, nodes that have a low amount of energy left over may be picked as CHs, which results in a rapid loss of energy for these nodes and ultimately leads to the disconnection of the entire cluster.
- The total number of CHs is subject to significant swings. This factor results in uneven cluster partitions, which in turn leads to an increase in the overall amount of energy that is lost over the entire network.

In this paper, we present a unique improved congestion aware LEACH (CAW-LEACH), with the goal of overcoming the shortcomings of traditional approaches and further extending the lifetime of IoT based networks. Each node's depleted energy ratio is taken into account in conjunction with the threshold used in conventional LEACH. This ensures that the node that was previously elected CH will not have another chance to win in the current round. This paper proposes a hierarchical data aggregation scheme as a congestion aware data aggregation scheme to improve energy efficiency and prolong the lifespan of the network. By estimating the congestion indicator (CIN), link error rate (LER), and residual energy (RE) of all available routing paths, the proposed congestion aware data aggregation scheme is able to aggregate the data via traffic-free paths.

Paper contributions

- The threshold and random number generation are related to the depleted energy ratio (DE) of the sensor nodes in order to maintain an appropriate level of energy consumption and to lengthen the lifetime of the network.
- The CHs who were chosen in the round before will not be chosen in the round after that since their DE is significantly higher in comparison to the nodes that are not CHs. It eliminates the possibility that the CHs node that was previously elected will get a second opportunity in following rounds.
- The proposed energy efficient and congestion aware data aggregation determines the routing path for data aggregation with the lowest cost subjected to end to-end delay and congestion.

2. Literature Survey

When designing WSNs, energy conservation is a critical factor to consider. The life of the battery is completely dependent on the longevity of the network. When trying to extend the lifetime of a WSN, one of the tactics that can be implemented is called clustering.

A fresh approach to the problem of constructing for HWSN, an energy-efficient clustering technique has been developed proposed by Santhosh V. Purkar and colleagues [13]. Several parameters, such as residual energy, starting energy, and hot count, as well as the CH selection, are taken into account during the clustering process. Energy efficiency, lifetime, stability, and throughput of the HWSN protocol are improved when using the proposed method as compared to using SEP, DEEC, or LEACH.

Hybrid unequal energy-efficient clustering and layered multi-hop routing was developed by Seyed Mostafa Bozorgi et al. [14] for use in WSNs. Based on the information about the neighbouring nodes in the clustering, a protocol method is defined, and layering is performed. Power consumption and network congestion are both reduced when unused control messages are purged. Based on simulation results, the newly developed HEEC technique is superior to the FHRP, LEACH-ERE, EADUC-II, and HUCL for achieving higher levels of stability.

A method for an energy-efficient clustering approach that utilises an optimised version of the LEACH protocol for data collection and transmission was provided by Salil Bharany et al., [15]. The approach that has been presented for data transmission uses less energy than other methods already in use. When compared to the currently utilised LEACH protocols, the proposed solution yields superior performance outcomes in terms of the ratio of packets successfully delivered and the amount of energy saved.

Sadia Firdous and colleagues [16] have presented a clustering-based routing strategy for optimising energy resources while taking load balancing into consideration as part of their research. The rotation of the cluster heads and the calculation of the distance have been handled, together with the energy utilisation at the sensor nodes. In comparison to the LEACH algorithm, the simulation's findings demonstrate that the suggested approach works offers significantly increased performance in terms of both the average amount of energy consumed and the lifetime of the network.

The layered architecture and the method for balancing load between CHs have been presented by Salim El Khediri et al., [17] in order to facilitate the handling of data packets. When it comes to the routing of data packets across sensor nodes, the network is divided into sections of unequal sizes. The new protocol offers significantly enhanced functionality in terms of energy usage, network longevity, and dependability.

In their presentation, Waz et al. [18] covered a wide range of Internet of Things (IoT) security products and approaches. When dealing with many IoT platforms, all of the components are combined into a single stack. This allows the integrity of the components to be maintained. This integration from one stage to another guarantees that there will be no break in the continuity of the security. With the assistance of middleware, complete data from the Internet of Things device may be tracked back user, and the user to the system.

Israr Ahmed et al. [19] described the most significant implications that the Internet of Things has on our everyday lives. It is presumed to be the system that has widespread adoption in both the virtual and the real world. As a result, this article delves into all of the most pressing concerns regarding confidentiality and safety.

According to the results that were obtained [20], this strategy is capable of handling the issues with energy efficiency as well as data delivery-related issues that arise inside heterogeneous systems. The results of the simulation reveal that the proposed method is improved and is superior to the existing ways. These findings are determined by the class's methods. Following the conclusion of each round, the energy consumption models are formed and subsequently decomposed into their respective gates. All in all, this extends the network's lifespan.

Using an enhanced version of the LEACH protocol, Salem and Shudifat [21] looked at the power sensitivity involved in selecting normal nodes as cluster heads when they are closer to the BS.Because of the shortcomings of the LEACH methodology, this was done. The results of the simulations show that the power consumption can be decreased while simultaneously increasing the network lifetime and maximising the number of cycles. In comparison to the LEACH protocol, the proposed method both reduces energy consumption and increases network longevity.

AnshuKumar Dwivedi and colleagues [22] have showed how to get over the bottlenecks that exist in WSNs by using balanced energy dissipation. Fuzzy inference algorithms are used in the clustering process and CH is chosen from those clusters have been studied by the authors. The findings of the experiments demonstrated that the recently proposed method known as EE-LEACH provides superior performance when compared to the older procedures known as DFCR and SCHFTL. The summary of research deficiencies is presented in Table 1.

Table 1: Summary of research gaps

S No	Year	Methodology	Techniques used	Parameter improved	Research gaps
1	2018	Energy efficient Clustering protocol to enhance performance of Heterogeneous wireless sensor network	HWSN, EECP	The node quality index, the length of the stability period, the number of living nodes in each cluster round, and the number of dead nodes in each cluster round	The best CH role is not determined in an appropriate manner. The mobile sink technique was not utilized in the planned
2	2019	Hybrid unequal energy efficient clustering	HEEC	Extended Network Lifespan, Lower Network Overhead, Greater Network Stability, Better Energy Balance	For the purposes of routing, these nodes are not considered energy harvesting nodes. There is no assurance that the network will save energy or have a long lifespan.
3	2021	Energy efficient LEACH	LEACH-EE	Lifespan, amount of energy left over, and packet success ratio	It is necessary to enhance the life cycle of WSNs and reduce the amount of energy required for data transmission.
4	2021	Distance Energy Evaluated approach	DEE	Lifespan of the network, percentage of successfully delivered packets, and number of active node	This work is limited by how long it will last. We have measured how long the node will last because we don't have a lot of resources.
5	2019	Enhanced LEACH	LEACH, CH Selection, E-LEACH	Energy usage at the cluster's head nodes and energy consumption across the network as a whole	This approach does not include any energy-efficient relay selection options. The phase that deals with the transfer of data has a very high energy usage.
6	2021	Fuzzy-LEACH	CH selection, fuzzy based clustering	Enhanced the network's average amount of energy. The mount of throughput during the stable time.	The greater the distance, the higher the energy levels. The data aggregation step has not been modified in accordance with the clusters.

3. System Model

Energy usage model

A sensor node's power consumption is a major factor in establishing the maximum level of performance that can be attained by the system. The network protocols that are put in place are constantly working to lower the amount of energy used by the system in an effort to prolong its lifespan. The energy consumption model is depicted in Figure 2.

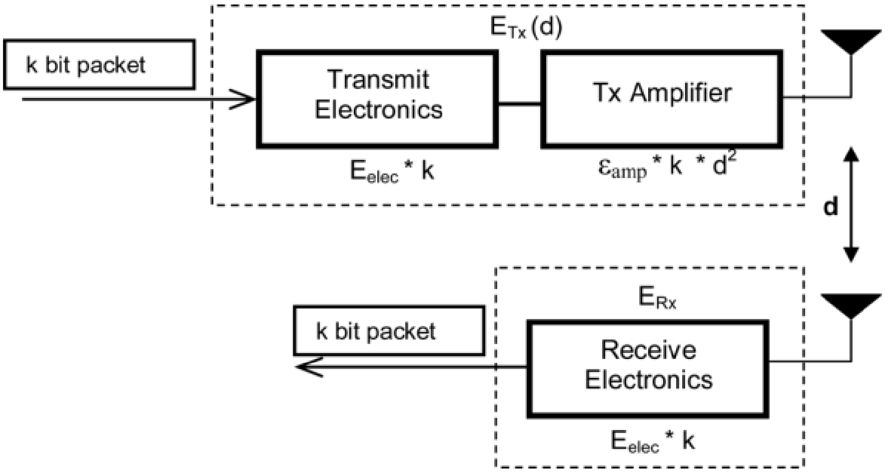

Fig 2: energy usage model in WSN

Radio waves are used for inter-node communication, and an energy model is applied to calculate how much electricity will be expended during the transmission and reception of data. The energy model takes into account not only the free space channel but also the multi-path fading channel. The free space channel is used if the distance (d)between the transmitter and receiver is less than a threshold value, called d_0, and the multi-path fading channel is used otherwise. Simply plug your values into the following equation to find out how much energy the model predicts is needed to send a k-bit message over a d -mile distance.

$$E_T(k,d) = \{ \begin{matrix} kE_{elec} + k \in_{fs} d^2, \ d \leq d_0 \\ kE_{elec} + k \in_{mp} d^4, \ \ d \geq d_0 \end{matrix}$$

where d_0 represents a threshold, E_{elec} represents the amount of energy needed by an electronic circuit, $\in_{fs}$represents the amount of energy needed by free space, and $\in_{mp}$ represents the amount of energy needed by a multi-path channel. The formula for calculating the amount of energy needed to receive k-bits of data is as follows:

$$E_R(k) = kE_{elec}$$

The amount of energy that is used by an electronic circuit, denoted by the symbol (E_{elec}), is determined by a number of different parameters. These factors include the digital coding, modulation, filtering, and spreading of the signal, among others. Both the distance between the transmitter and receiver as well as the bit error rate affect the amount of energy that is used by the amplifier whether operating in free space ($\in_{fs}* \ d^2$) or when operating in multi-path($\in_{mp}* d^4$).

Proposed system

Proposed CAW-LEACH method

Here we discuss the proposed approach for CH selection and how it stabilises random number generation are broken down and discussed in detail. In the method that has been proposed, the generation of random numbers is multiplied by the depleted energy (DE) and expected remaining energy (PRE) parameters of the network. This is done so that the algorithm is dependent on the energy that is contained within the nodes.

The first step in the CH selection process used by the traditional LEACH algorithm is for the nodes to generate random numbers. After that, the resulting random number is compared to a predetermined cutoff value. In the event that the random value is lower than the threshold, the node will be elevated to CH for the duration of the round. If the random number is less than the cutoff, then this holds true. While there are advantages to using the LEACH method, it cannot guarantee the CH's current residual energy. Standard LEACH-based protocols choose the threshold $TH(n)$ without regard to the nodes' energy levels, instead considering only whether or not the nodes will be selected as the CH. For this reason, CHs are selected at random, and if a node with lower energy is assigned as CH, it will soon die.

The suggested method makes the selection of CH reliant on the depleted energy (DE) and expected remaining energy (PRE) factors, which ultimately results in an increase in the system's performance as well as the network's longevity. The production of the random number is altered in the way that has been proposed to make the selection of CH more energy efficient. The random number is then multiplied by the sensor nodes' depletion energy (DE) and expected remaining energy (PRE) values. As a consequence of the product of these components being multiplied together, the method for producing random numbers is now reliant upon the energy of the nodes. The following is an explanation of the parameters for the suggested depleted energy (DE) and expected remaining energy (PRE):

Depleted energy (DE) ratio: The parameter depleted energy ratio can be defined as the ratio of energy depleted in the sensor rounds in the previous rounds. This parameter is very important to evaluate to prevent the same node is elected as CHs in the subsequent rounds. The parameter DE can be mathematically expressed as follows:

$$DE = \frac{E^T_{init} - E^T_{res}}{(r-1)_T}$$

Where E^T_{init} denotes the starting energy, E^T_{res} denotes the residual energy of each node, r denotes the current round, and $r - 1$is the round that came before this one at time T. The performance of DE in the previous round will serve as the criterion for choosing CH in the next round. In the subsequent round, the role of CH will go to the node that finished the previous round with the lowest DE. Because a CH from the previous round has an extremely high DE in compared to other nodes that are not CHs, it will not be chosen as a CH in the following round.

Predicted remaining energy (PRE): The difference between the initial energy, which is denoted by E_{init}, and the entire energy, which is denoted by E_{ec} , is what constitutes the remaining energy, which is denoted by E_{pre}. Because the prediction of energy avoids participating in CH candidate networks with weaker nodes, E pre is evaluated in this situation rather than residual energy (RE).

The predicted remaining energy of node "n" is given by

$$E^T_{pre}(n) = E^T_{init}(n) - E^T_{ec}(n)$$

Where, E_{ec} of node n can be calculated as follows,

$$E_{ec}(n) = E_{i,n} + E_{n2BS} + E_{elec} + E_{DA}$$

Here, $E_{i,j}$&$E_{node2BS}$ are the transmitting energy for "l"bits from node "i" to "n" and "n"to base station or sink, respectively. Additionally, E_{DA} is the aggregating energy of a datum, and E_{elec} is the energy spent by the reception circuit per bit.

Rand is the symbol used to represent the typical random number in LEACH's classic $rand\ (n)$. In the method that has been suggested, the enhanced procedure for selecting a random number can be described as follows, according to the Equation:

$$rand(n)' = rand(n) * (DE_n + PRE_n)$$

As the enhanced random number's value has been calculated, it can be checked against the sensor nodes' thresholds. When choosing a CH, the threshold function should be given careful thought. The threshold function evaluates the node probabilities and uses that information to select the CH. By judiciously factoring in node energy to the threshold function, network performance can be fine-tuned. In order to keep the network's power consumption stable, each node performs some CH operations. Based on the likelihood of CH selection, there is a one in one possibility of any one of the potential nodes in the network being chosen as the CH. A node's ability to function as a CH in the network is dependent on how much energy it has. The suggested method uses the energy of the nodes from the very first node in the network all the way to the very last node to perform the threshold function.

At this stage, the value of the manipulated random number is evaluated in relation to the threshold function.

$$rand(n)' \leq TH(n)$$

Where $TH(n)$can be represented as:

$$TH(n) = \begin{cases} \frac{P_n}{1 - P_n\ [(r\ mod\ \frac{1}{P_n})]} * (DE + PRE) + \frac{r}{P_n}, & if\ n\ \in G \\ 0, & otherwise \end{cases}$$

In the equation that was just presented, P_n stands for the chance that n will become CH, and r stands for the round number.

Node is promoted to CH status if the random number is less than the threshold; otherwise, the process moves on to the next node in the list.

After the CH selection process is over, the CHs will spread the news to the other nodes in the network that they have been chosen to serve as the CH for this iteration. In order to successfully finish this operation, Each CH node will notify every other node in the network via a broadcast message disguised as an advertising. The strength of the signal of the message that is sent from the CH nodes is used by each member node to determine whether or not it will take part in the process of cluster building.

Proposed Congestion aware intra-cluster routing

In this research, we propose a congestion aware data aggregation technique that is based on congestion indicator (CIN), link error rate (LER), and residual energy (RE) characteristics. Our goal is to achieve more reliable data transmission using this scheme. In order to design the many feasible routing paths from source nodes to CHs, the congestion indicator and the link error rate of all of the available links are calculated for each cycle of communication.

*Congestion indicator (CIN):*Congestion happens when the number of packets arriving at a given node is greater than the number of packets that it can forward. This leads to an increase in the rate at which packets are dropped and a delay in their delivery. The time that elapses between two consecutive packet arrival times at a given node is referred to as the packet arrival time T_{PA} . The packet arrival rate R_{PA} decreases with increasing delay. Inversely proportional is the ratio of packet service time ST_P to packet service rate SR_P. The service time of a packet is the amount of time it takes

delay caused by waiting to forward packets in the queue or for a packet to be retransmitted.

Following are the numbers for the packet arrival rate and the packet service rate:

$$R_{PA} = {}^{1}/_{T_{PA}}$$

The congestion indicator (also known as CIN) is helpful in locating instances of congestion. The ratio of the rate at which packets arrive to the rate at which they are serviced is used to measure it.

$$CIN = \frac{R_{PA}}{({}^{1}/_{S_{TP}})}$$

If $R_{PA} > SR_P$, then the CIN value is greater, which indicates that the node has been congested. The higher the CIN value, the greater the amount of congestion that will result at the particular node. The rate of the packet service that was utilised in order to identify the congestion.

If the CIN number is lower, then a node will experience a lower level of congestion. The following formula is used to compute the cost of the route based on the traffic index heading towards CH:

$$CIN_{route} = \sum_{i=1}^{n} CIN_n$$

Link Error rate (LER):The error rate (ER), which has a negative impact on the functioning of the sensor network, is one of the issues that have traditionally been linked with wireless channels. The linkages that have a high mistake rate will result in an increased cost to route, and will therefore be avoided. It is sufficient, then, to take into account the link error rate in order to achieve an improvement in the energy efficiency of the data aggregation process. The following is a description of how the parameter LER between nodes $i\ and\ j$ can be understood:

$$LER = \frac{dist(i,j)}{Buffer_{size}(j)}$$

Data transmission phase

The CH considers both the congestion indicator and the link error rate characteristics while making its assessment of all of the available links. Following an analysis of the optimal paths, the cluster head generates a routing table for every node in the cluster. When determining the most efficient path, the routing table takes into account a unique record. The following fields can be found in the routing table:

Id	Energ	CIN	LER

Cluster head sends routing tables to each node after constructing the routing table. All nodes have routing tables at the end of the process. The primary objective of this phase is to identify the subsequent hop for each packet that arrives at each node. Once a data packet is received, the receiving node act as follows: If the next hop node is CH, the data should be transmitted directly to CH. Else, check for the available neighbours. If neighbours found, then compare the CIN of each neighbour and select the node with lowest CIIN & LER value as forwarder node and forward the data to the selected node. If all the neighbour nodes having similar CIN value, then select the node with LER and transmit the data.

Energy analysis

Let's say that the entire number of sensor nodes that are present in the whole sensor region is denoted by N. Because there are k total sensors in the cluster that was generated, the total number of clusters in the separated region is equal to N/k. The M sensor nodes each send out the same number of bits of data. According to the Radio Energy Dissipation model of the sensor node, the amount of energy that is needed to transmit and receive 'l' bits of data over a distance of D_n is denoted by the symbols $E_{trans}\,(l,d)$ and $E_{recv}\,(l,d)$ accordingly, such that

$$E_{trans}\,(l,d) = \begin{cases} lE_{ec} + l\varepsilon_{fs}d^2, & d \leq d_0 \\ lE_{ec} + l\varepsilon_{mp}\,d^4, & d \geq d_0 \end{cases}$$

$$E_{recv}(l,d) = lE_{ec}$$

The amount of energy that must be used during transmission is proportional to the path loss exponent n, where n equals 2 for open space and 4 for multipath interference. E_{ec} is the amount of energy that is used up by the electronics of the transmitter and receiver during a single bit.'d' denotes the physical separation between the transmitter and the receiver. The amount of energy that is used by the transmitter amplifier may be calculated using the route loss exponent and the formula ε_{fs}, ε_{mp}. Due to the fact that the sensor nodes do aggregation, the'm' bits of data are aggregated into a smaller number of the 'h' bits. The value of E_{CD} represents the amount of energy that is used by the model that utilises centralised data aggregation (also known as Cluster Head). The value given by E_{SD} represents the amount of energy that is used by the suggested approach for performing Data Aggregation in the sensor node. The energy gain, denoted by E_{gain} , for a cluster that is a result of the suggested technique is as follows:

$$E_{SD} = m(E_{ec} + \varepsilon_{fs} d^2) + mE_{ec}$$

$$E_{CD} = h(E_{ec} + \varepsilon_{fs} d^2) + hE_{ec}$$

$$E_{gain} = E_{SD} - E_{CD}$$

$$E_{gain} = (2m - h)E_{ec} + (m - h)\varepsilon_{fs} d^2$$

Because there are N/k clusters in the area surrounding the sensor. The total amount of energy that can be gained via the proposed method is denoted by b.

$$Total_{gain} = (\frac{N}{K}) E_{gain}$$

Algorithm

```
For all the nodes '$n$' where $n \in N$
Divide the nodes as 'k' clusters
End for
CH selection
For all node '$n$' where $n \in k$
Calculate $CER_n$
        Calculate $TH(n)$
            Estimate $rand(n)$
            $rand(n_{new}) = rand(n) * CER_n$
                If $rand\ (n_{new}) \le TH(n)$
                $CH -> n$;
                End if
End for
Data transmission
For all the available paths
        Calculate $CIN, LER$
                If $(CIN_n \,\&\, LER_n = low)$
                Select $n$ as forwarder node
                Else
                $n = n + 1$
                End if
End for
```

4. Result and Discussion

Setup for simulation

The newly proposed CAW-LEACH mechanism is simulated, and its performance is evaluated with the help of NS2 before being compared to EN-LEACH and EE-LEACH. The sensor nodes in the network field are placed there in a haphazard manner. The dimensions of the network area are 1000 metres by 500 metres. All of the sensor nodes have been configured to have an initial energy level of 100 j. The number of nodes in the network can range anywhere from 50 to 200. During data transmission, the CBR traffic agent is utilised to generate consistent traffic. CBR is an abbreviation for "constant bit rate." UDP is the protocol that is used to carry out the data communication. The results of the experiment are presented in table 2, which can be found down below.

Table 2: Simulation parameters

Simulation Parameter	**Value**
Area of network	1000 m x 500 m
Number of sensors	50 to 200
Size of clusters	4
Initial energy of sensor node	100j
Size of packet	1024 bytes
Routing protocol	AODV

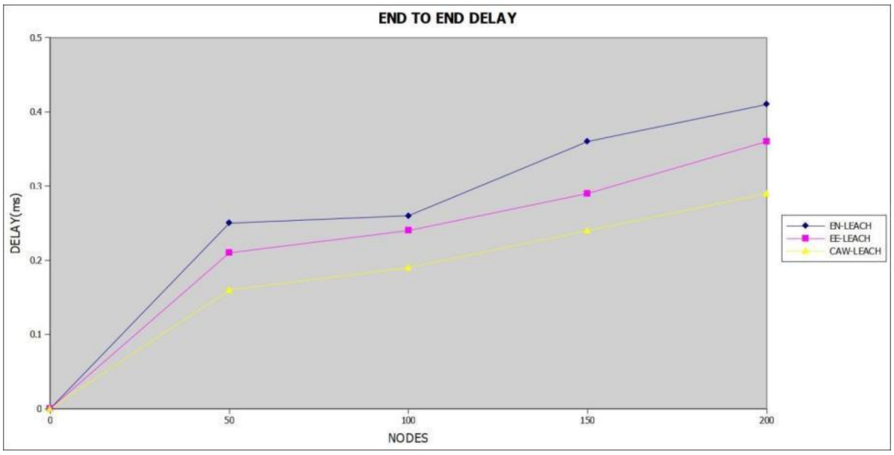

Fig 3: End to End Delay

Table 3: Comparison analysis of CAW-LEACH and existing methods as EN-LEACH and EE-LEACH for Delay

NODE	PROPOSED	EE-LEACH	EN-LEACH
50	0.16	0.21	0.25
100	0.19	0.24	0.26
150	0.24	0.29	0.36
200	0.29	0.36	0.41

The end-to-end delay is the overall time a packet takes to travel through a network. The evaluation of the proposed method's end-to-end delay time is presented in the values that have been mentioned above in table 3. The reduced amount of CH rotation is due to the stable CH selection, the implementation of DE, and the random number. The selection of low overhead relays during the relay selection process, which is conducted utilising quality and energy are connected in the QoS reputation paradigm, is another factor that contributes to the proposed method's ability to reduce end-to-end delay. When compared to the suggested technique, the previous methods encountered significantly higher delays of up to 0.48 milliseconds, while the proposed method only suffered a minimum average delay of 0.16 milliseconds. The graphical representation of the end-to-end delay can be found in Figure 3.

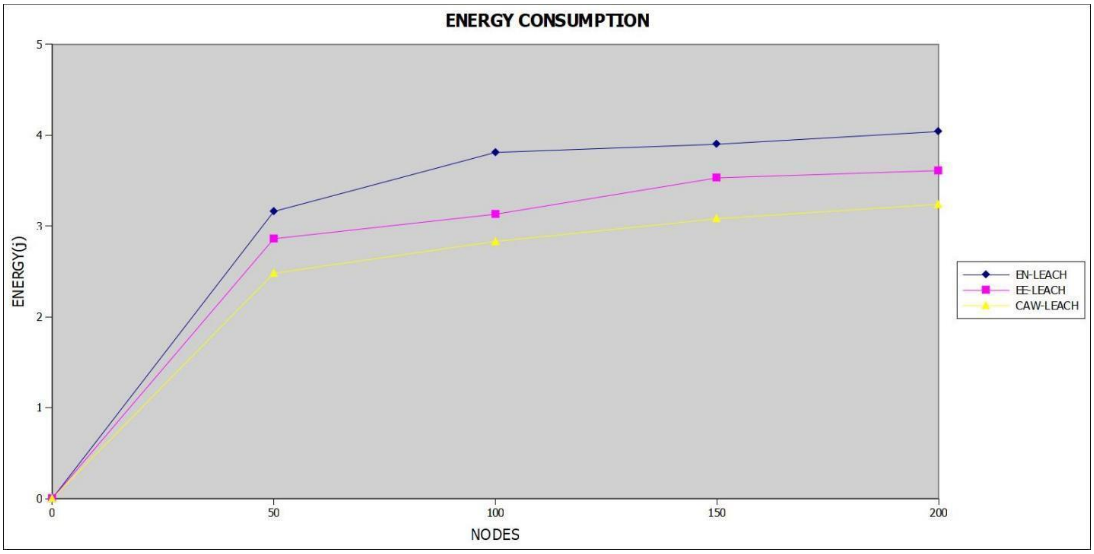

Fig 4: Energy consumption

Table 4: Comparing proposed and existing energy consumption strategies

NODE	PROPOSED	EE-LEACH	EN-LEACH
50	2.48	2.86	3.16
100	2.83	3.13	3.81
150	3.08	3.53	3.90
200	3.24	3.61	4.04

In order to participate in network operations, each sensor node is provided with 100 j of initial energy to use. Every time there is activity on the network, the energy level drops. In every network, the energy should be optimised for maximum efficiency in order to sustain network activity. The elimination of the need for retransmission and other activities that use up a lot of energy in the network is made possible thanks to the CH selection using DE and the selection of relay nodes with less congestion indicator values. This leads to a low amount of energy consumption in the proposed network. The method that was suggested had a documented rate of energy consumption that was an average of 2.9j. The graphic representation of the amount of energy used is shown in Figure 4. The findings of the experiment are presented in table 4, which can be found up above.

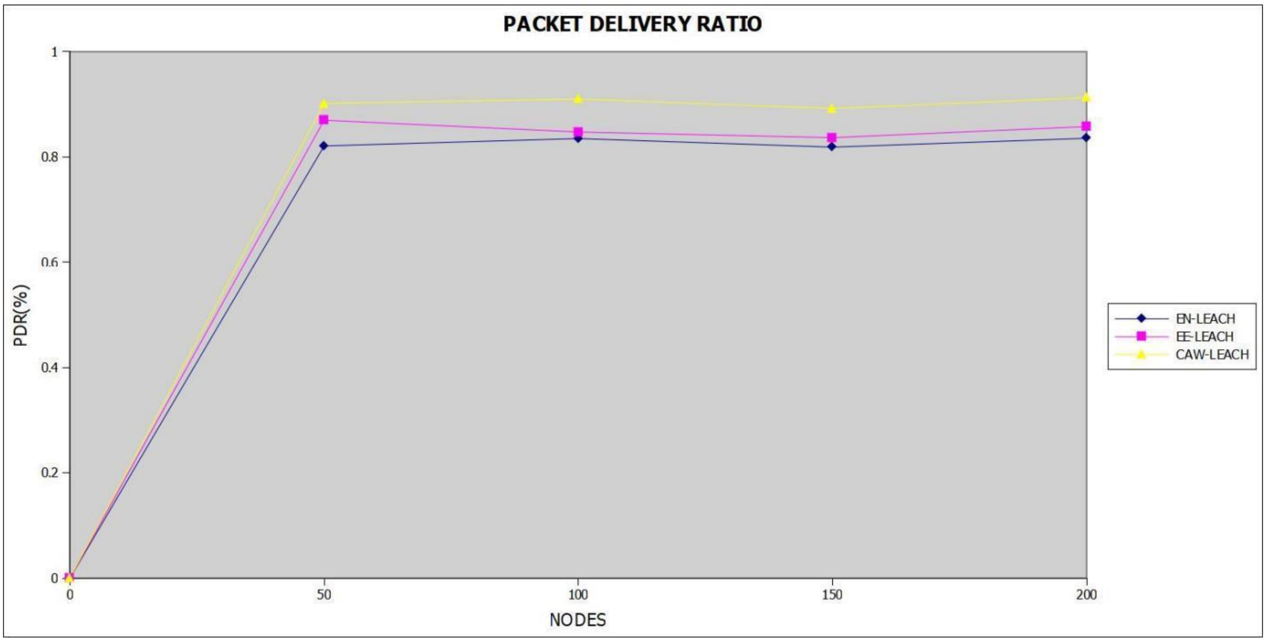

Fig 5: Packet delivery ratio

Table 5: Analysis of proposed and current PDR approaches in comparison

NODE	PROPOSED	EE-LEACH	EN-LEACH
50	0.9015	0.8696	0.8208
100	0.9099	0.8472	0.8348
150	0.8918	0.8365	0.8185
200	0.9125	0.8575	0.8353

The packet delivery rate (PDR) is defined as the fraction of data packets that reach their destination out of the total number of packets delivered. This ratio can be thought of as the proportion of data packets that are successfully received. The delivery rate was significantly enhanced because to the aggregation of data and efficient relay selection. The sensor nodes are able to convey the data with a high level of success thanks to the careful choice of channels that are reliable and well-balanced, in addition to the most suitable relays based on quality of service reputation. The prior approaches maintained an average PDR rate of 0.82, which is a rather low PDR rate. In contrast, the proposed method reached a maximum PDR of 0.91%, whereas the existing methods only achieved a PDR rate of 0.82 on average. The graphical representation of the packet delivery ratio can be seen in Figure 5. The outcomes of the experiment are presented in table 5, which can be found up above.

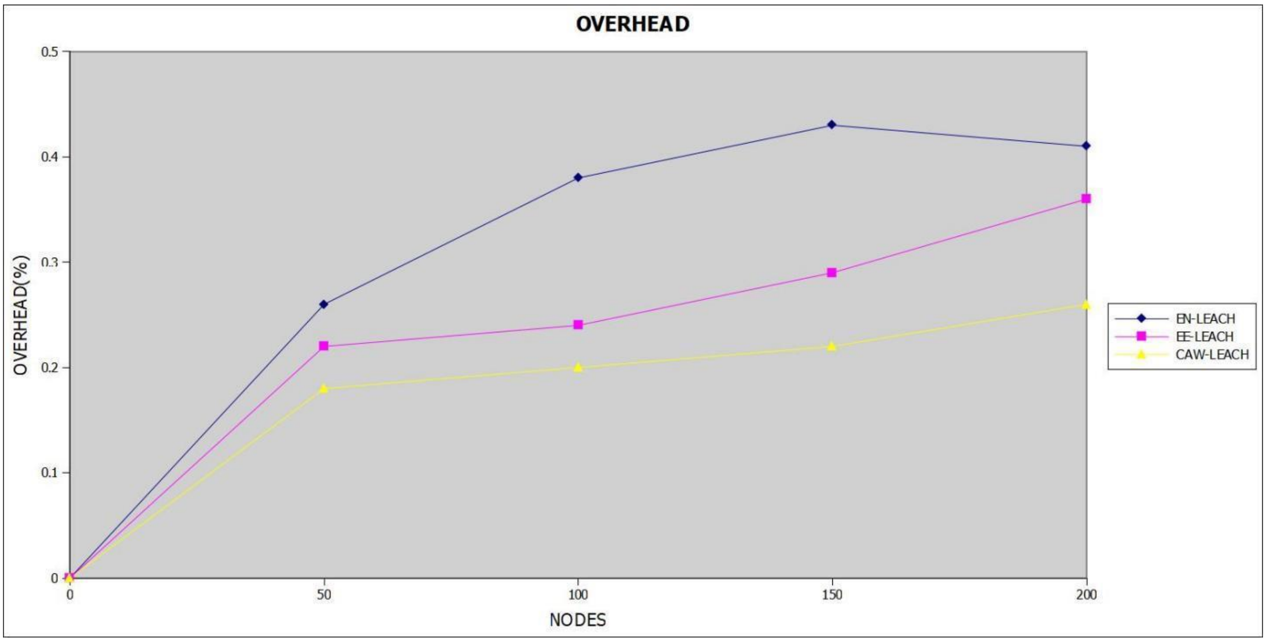

Fig 6: Routing overhead

Table 6: An evaluation of the proposed and existing approaches to Overhead

NODE	PROPOSED	EE-LEACH	EN-LEACH
50	0.18	0.22	0.26
100	0.2	0.24	0.38
150	0.22	0.29	0.43
200	0.26	0.36	0.41

The amount of control packets that are disseminated throughout the network to facilitate data transfer is directly connected to the overhead. The proposed approach had an overhead that ranged from 0.21 to 0.26 on average, whereas the existing methods had overheads that ranged from 0.26 to 0.40 on average. The error-free operation of the specified pathways is ensured by the utilisation of optimal relay selection in conjunction with congestion aware relay selection. Because of this, the proposed technique has a relatively minimal amount of overhead. The graphical representation of the routing overhead is provided in Figure 6. The findings of the experiment are presented in table 6, which can be found up above.

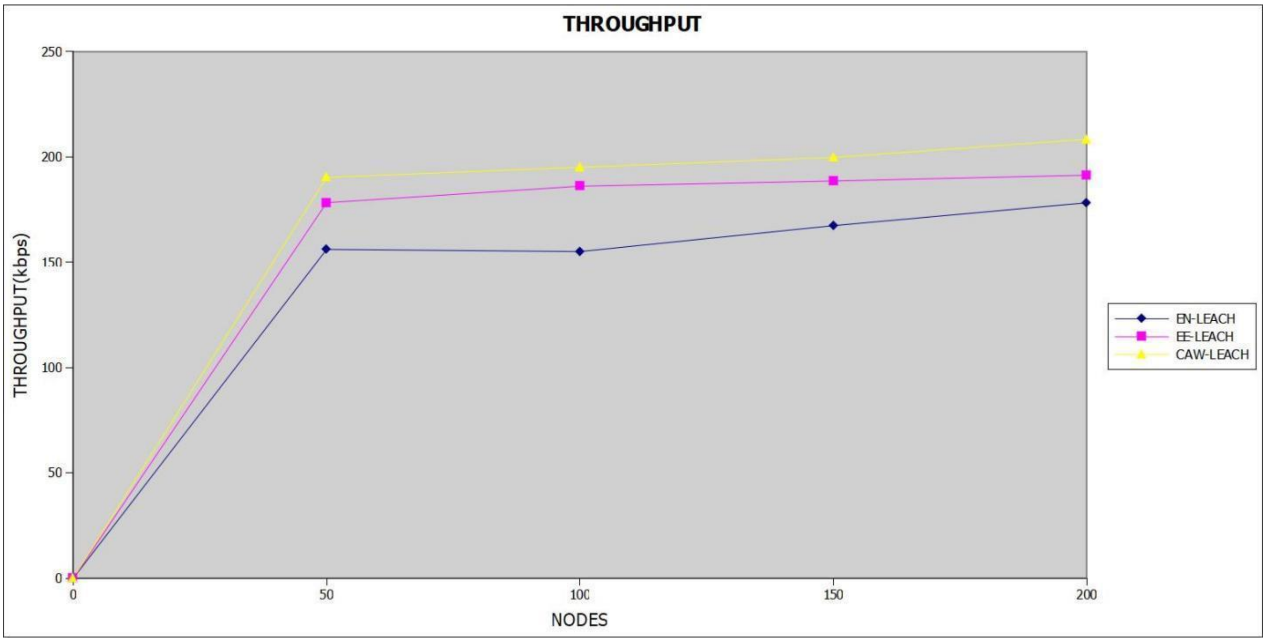

Fig 7: Throughput

Table 7: Throughput Comparison analysis of CAW-LEACH and existing methods as EN-LEACH and EE-LEACH

NODE	PROPOSED	EE-LEACH	EN-LEACH
50	190	178	156
100	195	186	155
150	199	188	167
200	208	191	178

The total number of data units that a node is able to handle in a specific amount of time is referred to as its throughput. The optimal selection of relays and energy-constrained CH selection through the use of DE both contribute to the best data aggregation. Table 7 shows that the proposed method has a higher throughput than existing methods. During the test, the suggested approach kept the average throughput rate as high as 389 kbps, whereas the existing methods kept a throughput rate that was lower than the proposed one. The graphical representation of throughput can be seen in Figure 7.

5. Conclusion

This work proposes an enhanced LEACH-based clustering approach to improve IoT based network data aggregation efficiency, conserving energy, and extending the network's lifetime. In the classic LEACH approach, the selection of cluster heads (CHs) is accomplished by the use of a random integer. Because LEACH accords the same priority to nodes with low residual energy as it does to nodes with high residual energy, this can lead to some nodes with low residual energy passing away before nodes with high residual energy. In addition, as IoT based networks are energy limited networks, congestion is the primary problem that needs to be addressed in order to achieve efficient data aggregation. We propose an improved LEACH protocol (CAW-LEACH) that takes into account the congestion indicator and estimated remaining energy of the nodes for CH selection and random number generation. This will help increase the energy efficiency as well as the stability of the CH. Its purpose is to ensure that the CH node that was just recently elected will not be given a second chance in this round. In addition, a congestion-aware data aggregation scheme is presented for the purpose of making data aggregation more effective. The simulation and analysis results show that CAW-LEACH can significantly outperform existing approaches for IoT-based networks in terms of energy efficiency, network lifetime, and data throughput.

References:

[1] Chen, J.I.Z. and Yeh, L.T., 2021. Graphene based Web Framework for Energy Efficient IoT

Applications. *Journal of Information Technology*, *3*(01), pp.18-28.

[2] Mamdiwar, S.D., Shakruwala, Z., Chadha, U., Srinivasan, K. and Chang, C.Y., 2021. Recent advances on IoT-assisted wearable sensor systems for healthcare monitoring. *Biosensors*, *11*(10), p.372.

[3] Maheshwari, P., Sharma, A.K. and Verma, K., 2021. Energy efficient cluster based routing protocol for WSN using butterfly optimization algorithm and ant colony optimization. *Ad Hoc Networks*, *110*, p.102317.

[4] Yuvaraj, N., Karthikeyan, T., Praghash, K. and Reddy, K.H., 2021. Binary flower pollination (BFP) approach to handle the dynamic networking conditions to deliver uninterrupted connectivity. *Wireless Personal Communications*, *121*(4), pp.3383-3402.

[5] Luo, J., Chen, Y., Wu, M. and Yang, Y., 2021. A survey of routing protocols for underwater wireless sensor networks. *IEEE Communications Surveys & Tutorials*, *23*(1), pp.137-160.

[6] Rani, K.S. and Balasubadra, K., 2022. Cross Layer Controlling Algorithm to Overcome Congestion in Underwater Wireless Sensor Network. *Wireless Personal Communications*, *122*(4), pp.3741-3756.

[7] Alfouzan, F.A., 2021. Energy-efficient collision avoidance MAC protocols for underwater sensor networks: Survey and challenges. *Journal of Marine Science and Engineering*, *9*(7), p.741.

[8] Priyadarshini, R.R. and Sivakumar, N., 2021. Cluster head selection based on minimum connected dominating set and bi-partite inspired methodology for energy conservation in WSNs. *Journal of King Saud University-Computer and Information Sciences*, *33*(9), pp.1132-1144.

[9] Jain, K., Mehra, P.S., Dwivedi, A.K. and Agarwal, A., 2022. SCADA: scalable cluster-based data aggregation technique for improving network lifetime of wireless sensor networks. *The Journal of Supercomputing*, pp.1-29.

[10] Roberts, M.K. and Ramasamy, P., 2022. Optimized hybrid routing protocol for energy-aware cluster head selection in wireless sensor networks. Digital Signal Processing, 130, p.103737.

[11] Singh, H. and Singh, D., 2021. Hierarchical clustering and routing protocol to ensure scalability and reliability in large-scale wireless sensor networks. The Journal of Supercomputing, 77(9), pp.10165-10183.

[12] Farahzadi, H.R., Langarizadeh, M., Mirhosseini, M. and FatemiAghda, S.A., 2021. An improved cluster formation process in wireless sensor network to decrease energy consumption. Wireless Networks, 27(2), pp.1077-1087.

[13] C. Gulzar, AmeenaYasmeen 2016. Maximum network lifetime with load balance and connectivity by clustering process for wireless sensor networks: International Journal of Computer Engineering In Research Trends ,3(7),pp.375-383.

[14] Bozorgi, S.M. and Bidgoli, A.M., 2019. HEEC: A hybrid unequal energy efficient clustering for wireless sensor networks. Wireless Networks, 25(8), pp.4751-4772.

[15] A Damayanthi ,Mohammad Riyaz Belgaum, 2022. A Study of Heterogeneity Characteristics over Wireless Sensor Networks . International Journal of Computer Engineering in Research Trends, 9(12), pp.258-262.

[16] Firdous, S., Bibi, N. and Wahid, M., 2021, December. An Energy-Efficient Cluster Based Routing Algorithm for Wireless Sensor Network. In 2021 International Conference on Frontiers of Information Technology (FIT) (pp. 182-187). IEEE.

[17] Khediri, S.E., Nasri, N., Khan, R.U. and Kachouri, A., 2021. An improved energy efficient clustering protocol for increasing the life time of wireless sensor networks. Wireless Personal Communications, 116(1), pp.539-558.

[18] Waz, I.R., Sobh, M.A. and Bahaa-Eldin, A.M., 2017, December. Internet of Things (IoT) security platforms. In 2017 12th International Conference on Computer Engineering and Systems (ICCES) (pp. 500-507). IEEE.

[19] Navdeep Kumar Chopra, Rajesh Kumar Singh., 2019, An energy-aware clustering approach for routing mechanism in WSN using Cuckoo Search. International Journal of Computer Engineering in Research Trends ,6(7),pp. 340-345.

[20] Mahapatra, C., Sheng, Z. and Leung, V.C., 2016, May. Energy-efficient and distributed data-aware clustering protocol for the Internet-of-Things. In 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) (pp. 1-5). IEEE.

[21] Salem, A.O.A. and Shudifat, N., 2019. Enhanced LEACH protocol for increasing a lifetime of WSNs. Personal and Ubiquitous Computing, 23(5), pp.901-907.

[22] Dwivedi, A.K. and Sharma, A.K., 2021. EE-LEACH: Energy Enhancement in LEACH using Fuzzy Logic for Homogeneous WSN. Wireless Personal Communications, 120(4), pp.3035-3055.