Sādhanā (2022) 47:166
https://doi.org/10.1007/s12046-022-01939-7

Double attribute based node deployment in wireless sensor networks using novel weight based clustering approach

J SAMUEL MANOHARAN

Department of ECE, Sir Isaac Newton College of Engineering & Technology, Nagapattinam, Tamil Nadu 611102, India
e-mail: drjsm1530@ieee.org

MS received 16 April 2022; revised 20 June 2022; accepted 28 June 2022

Abstract. In recent years, WSNs have become one of the fastest emerging networks. It enables a larger variety of applications in the real-time as well as automation industries. WSN applications are made up of a large count of sensor nodes that are distributed as per the application's requirements. Sensor nodes, depending on its manufacturing rationale, monitor, sense, receive, record, and transfer any type of data. Sensors are inexpensive, tiny, and have limited energy efficiency. Inefficient methods of utilizing this scarce battery power results in the death of nodes which consequently affects the lifetime of the entire network. The failure of nodes because of inadequate routing strategies reduces the network's lifespan and overall quality. Numerous previous research methodologies were applied to improve network lifespan and node connection together with communication dependability. Most of the solutions failed to deliver ideal performance in terms of improving overall QoS, which is a collective characteristic. In this research, a novel WBC approach for data gathering, node clustering, and load balancing in WSN is proposed. The functioning of the proposed model relies on the effective assignment of nodes to the communication task based on their weighted function computed based on the performance characteristic. Load balancing as well as data aggregation, are the two attributes effectively considered in this research work. The performance of the suggested WBC is compared to traditional benchmark techniques using the NS2 program. Multiple measures have been calculated and studied, and in every case, the suggested WBC outperforms.

Keywords. Node deployment; wireless sensor networks; clustering; data gathering; energy efficiency; novel weight based clustering approach.

1. Introduction

WSNs have attracted the interest of academics nowadays due to their widespread application [1]. WSNs are useful in a variety of fields, including control engineering and environmental monitoring. Sensor nodes in WSNs are composed of a variety of elements, including sensors, transceivers, microcontrollers, and memory. Because sensor nodes are powered by batteries, power consumption represents an important component in determining the sensor node's lifespan [2]. Every sensor node accomplishes tasks such as environmental sensing, data gathering, processing, and transmission to various sensor nodes or the BS [3]. As a result, the sensor node uses considerable energy to complete these activities. The energy consumption of sensor nodes will increase when multiple packets are delivered, and vice versa [4]. As one sensor node sends data to some other node in the network, the load on another sensor node grows, causing the severely laden sensor node to die sooner. To solve this issue, the network's total load must be balanced [5].

In WSNs, load balancing solutions include packet routing along an optimum path, clustering of sensor nodes, and so on [6]. We employed cluster- oriented WSNs, where sensor nodes are joined together to create a cluster, with CH serving as the cluster's leader [7]. CHs gather and analyze data from a cluster's individual member sensor nodes. Following that, the processed data is sent to a BS or another CH via an indirect or direct link. All the data is gathered in this manner at the BS or static data sink. For generalized alerts or updates from certain monitoring region, BS is linked to the web [8]. The following are the advantages of cluster oriented WSNs: (1) Energy usage is considerably decreased since only one representation (i.e., CH) per cluster is needed for data aggregation as well as routing [9]. (2) Communication bandwidth may be saved greatly since sensor nodes only interact with their CHs and avoid sending duplicate messages to one another. (3) Cluster administration is simple because they can localize route setting for

Published online: 16 August 2022

sensor nodes using compact routing tables. This, in return, aids network scalability [10].

The surplus energy is allocated to gateways and relay nodes. Gateways have the same capabilities as CHs and are accountable for the same tasks [11]. Batteries are also used to power the gates. As a result, every gateway must save energy in order to extend the network's lifetime [12]. As a result, cluster formation must be precise and correct to reduce CH overload. Quite an overload may enhance communication delay, wasting the CH's high energy and degrading the WSNs' actual quality [13]. As a result, for clustering sensor nodes, load balancing of the CHs is critical. If the data load exceeds the gateway node's handling capabilities, it may expire prematurely owing to increased energy consumption [14]. As a result, they share data with neighboring nodes to balance the overall load and extend the node lifetime [15]. As a result, effective load balancing methods should be used to address this difficulty.

2. Motivation and research contribution

Wireless sensor networks are indispensable in their applications in today's scenario. However, their remote deployment nature followed by limited battery power has opened a large scope of research in finding intelligent methods to utilize the available battery power without network collapse or failure. Nodes are centric to any routing methodology and careful analysis of their energy levels, the proximity of nodes to one another, incoming traffic, etc. are critical parameters in choosing the ideal nodes for communicating from source to destination. These many parameters are however challenging in the simulation level as well as in real-time implementation which has attracted quite a lot of research interest in recent times. The overall objective is to maintain optimal quality of service for a given task.

Keeping in mind the above-stated objective, the research work is systematically categorized as.

- To propose a novel WBC approach for data gathering, node clustering, and load balancing in WSN.
- To perform the clustering, CH selection, and the distance computation as two steps.
- To concentrate on shortest path computation, which decreases energy consumption, packet loss, latency, and load balance by identifying the shortest path for transmitting the data.

The paper is organised as follows. Section 1 is the introduction of node deployment in WSNs. Section 2 deals with motivation and research contribution. Section 3 give literature survey to the related work. Section 4 is related to the proposed work.clustering and data gathering in WSNs. Section 5 deals with results. The conclusion is provided in section 6.

3. Related work

In 2017, Edla *et al* [16] have modified the SFLA by altering the frog's population generation as well as offspring generation stages, as well as adding a transfer phase. A new fitness function was also being developed to assess the efficiency of the upgraded SFLA's solutions. With respect to different performance characteristics, we ran comprehensive simulations of the suggested load balancing method. The tested findings were positive, demonstrating the suggested algorithm's efficacy.

In 2021, Hawbani *et al* [17] have suggested a novel technique that incorporates two major components. To begin, every node creates a CZ with a four-cornered regular geometric form. The node's packets will be forwarded through any path inside the CZ. Nodes inside the CZ were explicitly authorized to be chosen as candidates. The network density determines the size of CZ. Next, candidates inside the CZ were selected using the OR metric, which was described as the multiplication of four different distributions: direction, perpendicular distance, transmission distance, and residual energy distribution. The protocol outperformed traditional alternatives with respect to energy consumption, network lifetime, sender waiting time, routing efficiency, and duplicate packets, according to the findings of an exhaustive performance assessment study and modeling of large-scale situations.

In 2017, Chatterjee *et al* [18] have suggested a revolutionary graded node growth approach that produced minimal traffic, just enough for coverage in static WSNs to capture data streams. A distributed, almost load-balanced data collection technique was created on the basis of this node distribution to send packets to the sink node through minimum-hop pathways, which supports to decreasing network traffic. To provide a hypothetical lower constraint on the count of nodes to be implemented, an average case probabilistic analysis is performed based on the perfect matching of random bipartite networks. The suggested methodology, according to analysis as well as simulation tests, leads to a large increase in network lifespan, which greatly outweighs the cost of over-deployment. As a result, this approach provides an energy-efficient and cost-effective and solution for node placement and routing in large WSNs that will run for a long period.

In 2018, Liu and Zhang [19] have suggested a unique load balancing technique for WSN data transmission called super links-oriented data drainage, which takes control of super nodes' more powerful hardware and better communication bandwidth to achieve data traffic redistribution. This was a proactive and early-intervention method, as opposed to traditional passive late-remedy procedures. The main concept was to transmit data from places comparatively far from the sink having a surge in data traffic to those near the sink with low data traffic, and an iterative approach was built to choose acceptable start points and termination points of super connections. The efficacy and benefits of the novel strategy were validated through simulation studies.

In 2020, Adil *et al* [20] have proposed communication architecture on the basis of EGNs to establish a consistent load balancing method for resource-limited networks. The residual energy of the participating nodes, i.e., C I Network, is measured by EGN. Throughout addition, EGN nodes broadcast hop selection information in the network, which regular nodes utilize to modify their routing tables. Ordinary nodes, on the other hand, utilize this information to uni-cast their acquired data to the destination. EGN nodes use a built-in setup to classify their neighbors into categories like normal, powerful, and crucial energy. EGN measures the nearby node's residual energy (E_r) using the SPR and RTT values and those nodes with the highest E_r values were marketed as trustworthy communication pathways. Moreover, EGN sends an RREQ to the network and returns an RREP from each node in its near neighborhood, which was utilized to calculate the E_r energy levels of the nodes in its vicinity. If an adjacent node's E_r value was less than the set categorization threshold value, the node was announced as unavailable for transmitting communication. The suggested method outperforms previous methods with respect to throughput, individual node lifespan, PLR, communication costs, latency, and computing costs, among other metrics. Furthermore, the developed strategy protects WSNs and individual nodes in the operational environment from existing strategies.

In 2020, Zhang *et al* [21] to overcome the problem of energy efficiency, they devised a CH selection method having a MS and proposed relay selection methods using UM-MAB. We suggest a theory called as a VH for MS to gather data with respect to energy efficiency based on the description of residual energy and node density. Furthermore, using two-hop transmission in collaborative power line communication, which works with the long-distance transmission, we inevitably turn the relay selection issue into a permutation issue. In terms of the relay selection issue, we suggest using the MU-MAB machine learning method to handle it using a reward, linked with an increase in energy consumption. Moreover, for the assignment of the final one-to-one optimum pairings to ensure energy efficiency, we use the stable matching theory oriented on marginal utility. The limitation was used to illustrate the performance of MU-MAB utilizing the UCB index for assessing it. Finally, simulation outcomes demonstrate the feasibility and efficiency of the suggested strategies for energy conservation and balance.

In 2019, Li *et al* [22] for WSNs, the EBAR Algorithm has been suggested. To balance the energy consumption of the sensor nodes, EBAR uses a pseudo-random route-finding method and an enhanced pheromone trail updating mechanism. It optimized route formation via an effective heuristic updating technique based on a greedy projected energy cost metric. Furthermore, EBAR employs an energy-oriented opportunistic broadcast technique to decrease the energy consumption generated by control overhead. We analyze EBAR in terms of performance parameters like energy efficiency, energy consumption, and anticipated network lifetime by simulating WSNs in various application situations. In comparison to the traditional methodologies SensorAnt, EEABR, and IACO, the findings of this detailed research reveal that EBAR delivers a substantial enhancement.

In 2019, the GWO technique [23] was used to provide energy-efficient routing and clustering. In addition, two new fitness functions for routing and clustering issues are proposed. The fitness function for routing was designed to reduce the total distance traversed and the count of hops. The clustering fitness algorithm spreads the total load based on the distance between gateways and the BS. When compared to current algorithms such as GA, PSO, and multi-objective fuzzy clustering, the suggested GWO-oriented technique yielded higher values of both routing fitness functions and clustering.

In 2021, Mohajerani *et al* [24] have worked with anycast routing and sink placement to help multi-sink WSNs last longer. To simultaneously handle the issues of multi-sink placement, clustering, and load-balanced anycast routing, two methods were developed: “MPAR” and “EMPAR”. Both EMPAR and MPAR use a two-tiered design, with sensors grouped at the lowest level. A load-balanced data aggregation routing tree connects every sensor to the associated CH. Both approaches employ a modified PSO technique to identify the appropriate position of sinks at the upper level. The ACO approach was used to create a high-level anycast routing tree for every sink. To convey aggregated data from CHs to sinks, every anycast tree employs the hybrid CS approach. The suggested algorithms' effectiveness with respect to energy consumption variation, energy consumption, and network longevity is demonstrated through simulation studies. Because of its CH selection technique, EMPAR outperforms MPAR in terms of performance. To pick the ideal CH for every cluster, EMPAR evaluates both available energy and distance requirements, as well as a rest factor. When contrasted to the MPAR algorithm as well as the energy-aware CS-oriented data aggregation method, EMPAR decreased energy usage by 5.98% and 12.20%, correspondingly, for an average count of clusters. In comparison to the available methods, it enhances network longevity by 12.26% and 30.38%, correspondingly.

In 2021, Yao *et al* [25] in WSNs, SenCar has been used in combination having AODV. This dual function strategy involves computing the node's remaining energy and charging it if it falls below a certain threshold. Finally, the suggested technique was thoroughly validated utilizing numerical results obtained employing the NS2 simulator. In comparison to previous Recharging approaches, these numerical results were given in a graphical style, demonstrating the efficiency of the suggested method.

Furthermore, the suggested model was used for both non-sensor and sensor nodes to find the best and most effective kind.

4. Proposed work

4.1 *Novel weight-based clustering*

Throughout every domain, clustering is used. It is used for data in data mining as well as retrieval methods [26], as well as for WSN nodes. In most cases, fit nodes in the early stages elect to attain Cluster Heads and begin broadcasting to their k-hop neighbors. The level 1 CHs will next attempt to attain level 2 CHs having a probability larger than the network's remaining nodes. This will proceed till the entire nodes have been assigned to one of the clusters. Hierarchical Clustering uses this approach as its foundation [27]. The LEACH was improved in this method CBERP by expanding the count of enhanced nodes that are qualified for CH selection. The life duration of enhanced nodes can be extended in this manner.

Hierarchical clustering begins by forming a tree out of nodes. Clusters are thought to be present in this tree. There are two forms of hierarchical clustering. They are both agglomerative and polarising. Clustering is performed from the bottom up. Clustering is performed top-down in the second kind. The bottom-up strategy is essentially a merging process, where individual clusters join forces with remaining neighbouring clusters to produce a single large cluster that covers entire nodes in the WSN. The next is the polar opposite of the bottom-up method. The whole network is originally treated as a single cluster in this method. Next, depending on specific criteria, it divides the large cluster into multiple smaller clusters. The technique used here for initial clustering is top-down hierarchical clustering.

The Max-Min d-cluster Algorithm [28] creates "D-hop" clusters having a maximum run duration of O(d) rounds. The nodes that exist at a distance of d or less from the CH are grouped together beneath the identical cluster in this strategy. It occurs because of two rounds. During the two rounds, flooding is performed. After each cycle, cluster IDs or counts are recorded. These nodes communicate their IDs to the entire nodes within their reach. This is referred to as flooding. The nodes that receive larger counts of the similar node ID must then pick their CH according on a set of conditions. If any node obtains its self ID after flooding in the second round for the smallest node IDs, it announces itself as the CH. Alternatively, as in the previous round, the CH is determined by the node ID received by the majority of the remaining nodes. Each node must communicate its CH to choose the gateway node. If a node receives distinct CH IDs from the entire of its neighbours, it acts as a gateway node, sending its ID to complete of the CHs via the remaining nodes.

Even though this technique produces clusters from n nodes without keeping any individual nodes left, it cannot be termed an energy-efficient clustering. It was also done with the presumption that each node in the network is aware of the whole network topology. This presumption is hard to consider for, when there is a high count of nodes. The nodes randomly deployed in a position are to be grouped to decrease extra network traffic. Unlike in various domains, complicated clustering algorithms should not be applied here.

The Euclidean distance is used to generate the clusters. Within a short distance, the sensor nodes are gathered and grouped into clusters. Equation (1) can be used to compute the Euclidean distance.

$$di = \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2} \tag{1}$$

Compute the distance by having additional sensor nodes based on the initial sensor node, and in the next group, those nodes that meet the distance criterion group into a cluster. This is performed iteratively, beginning with a random node and ending when all of the nodes in the network are clustered. We decrease the lengths depending on the cluster to the range of [0,1] utilizing min-max normalisation to achieve more appropriate outcomes in the selection of the CH owing to the broad range of estimated distance values.

$$o_{di} = \frac{di_{node} - min_{di}}{max_{di} - min_{di}} \tag{2}$$

Here, di_{node} defines the distance between the node and the cluster centre, min_{di} defines the node's lowest distance from the cluster centre, and max_{di} defines the node's maximum distance from the cluster centre. The following is the amount of energy used by the CM to transfer data:

$$F_{TY}(l, e) = F_{ele} * l + F_{am} * l * e^2 \tag{3}$$

Here, F_{ele} shows the receiver or transmitter's per-bit energy consumption, l shows the package's data size, F_{am} shows the amplifier parameter, and e shows the distance [28].

The CH uses the following amount of energy to receive the data:

$$F_{RY}(l, e) = (F_{ele} + F_{DB}) * l \tag{4}$$

Here, EDA denotes data aggregation energy consumption [28]. The CH uses the following amount of energy to transport the acquired data to the BS [28]:

$$F_{TY}(l, e) = (F_{ele} + F_{DB}) * l + F_{am} * l * e^2 \tag{5}$$

The sensor node's leftover energy is as below:

$$F_{residual} = F_0 - F_{total} \tag{6}$$

Here, F_0 shows the node's starting energy and F_{total} shows the node's total energy [28]. As per (7), the CH is chosen as the node having the largest g value:

$$g = (F_{residual} - o\_di) \tag{7}$$

Here, the novel WBC is employed as criterion for the CH selection. On the basis of the method, this weight is employed to calculate energy, the distance between nodes and CH, and the number of times a node has become CH. In every clustering repetition, every node determines its weight. In a WSN, clusters are constructed in such a manner that energy consumption is minimised. The novel WBC describes a non-homogeneous network clustering approach. Its effective clustering algorithm selects the superior CHs, increasing the lifetime as well as throughput of WSN. The CH is chosen using this novel WBC so that the CH always contains the maximum residual energy. Residual energy is the energy that remains in a node after it has completed its processing as well as information activities. It prevents low-energy sensor nodes from being chosen as CHs. It extends the WSN's life. In addition to residual energy, the CH is chosen based on various parameters such as the count of living neighbours and the distance from the BS.

The clustering is applied to the whole network nodes. As a result, after the loop is completed, all the nodes will form a cluster. For each node, a Boolean flag parameter is provided. For entire nodes, the parameter value is "False" by default. The flag of that node is set to "True" after it is allocated to a cluster. The clustering algorithm's pseudo code is as follows:

```
For k = 1
For k = 1:1; o
   dis = ((T(j)·ye) − Yl)^2 + ((T(j)·ye) − Zl)^2;
                    e = sqrt(di);
   If (e < di_max && (T(j)·flag == false))
                    T(j)·D = 1;
                  T(j)·flag = true;
      Plot (T(j)·ye, T(j)·ze, 'n.');
      Node_Id_in_Cluster1=T(j)·id
   end
end
For j = 1:1: o
   If (T(j)·flag == false)
                    Yl = T(j)·ye;
                    Zl = T(j)·ze;
   end
end
end
```

Two loops are utilised here. The cluster iteration is done in the initial loop. The second loop iterates across the nodes in that cluster; $T(j)$ specifies the node ID; D indicates the cluster ID; Yl and Zl identify the initial node's position. The distance from the initial node (a random node chosen as the initial node) to the neighbour node is determined utilising Euclidean distance formula. If the distance between the nodes is less than 25 metres and the neighbour node's flag parameter is "false," that node is joined to the first, establishing a cluster.

The flag value "false" is assigned to the initial node that does not meet the criterion. Yl and Zl have been allocated to it. It serves as the second cluster's initial node. It proceeds until all the nodes in the network's flags have the value "true." This algorithm is used to do numerous iterations.

The node weighting parameter is given a random value when the cluster is formed. The CH is initially chosen for the initial round based on its weight. The nodes are next weighted depending on their load M and residual energy F from the following cycle onwards. The weight α is computed employing the below equation (8). (in which j specifies the node ID and u specifies the time instant),

$$\alpha = LF_{iu} + \frac{1}{M_{iu}} \tag{8}$$

Each cluster has its own schedule, which is constructed in ascending order of. If the cluster's nodes have any data to relay, they utilize the slots as per the timetable. If a node in that cluster has little to communicate, the energy of that node is initially verified. If it possesses positive energy, it is assigned a slot based on its value. It sends its load (packet) to the node with the next higher value in the queue in that slot. The data transferred from the nodes within each cluster is transferred to the CH at the end of each round. The CH describes the node that gets the last slot in the schedule. The CH with the highest value is this one. The data from the node is routed through the cluster according to this timetable. The data from CH is forwarded to a neighbouring node in the neighbour cluster, which then passes it on to its CH. The data eventually reaches the Sink node (or BS) through incremental forwarding. The clustering in WSN is shown in figure 1.

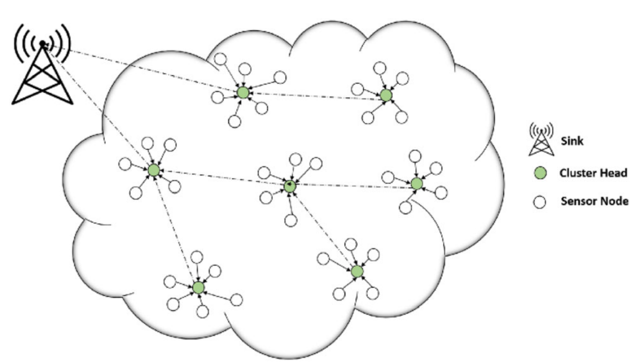

Figure 1. Clustering in WSN.

4.2 *Data gathering*

In WSNs, mobile sinks are widely employed for data collection. This strategy eliminates the energy consumption inefficiencies created by multi-hop transmission, but it may result in a longer delay time. The length of the travelling path in ROI as well as the speed of the mobile sink determines the data aggregation delay time. As a result, the key emphasis of studies in this field is on how to design a shorter travelling path in the ROI for the mobile sink having motion speed control. Multiple data aggregation tasks can be merged if the sensors can be encompassed by a disc with a radius no bigger than the mobile sink's communication range. Because the mobile sink does have a shorter travel route, it may complete the data aggregation operation in less time, reducing the delay time.

4.3 *Load balancing in WSN*

The method's major goal is to effectively cluster sensor networks surrounding a few high-energy gateway nodes. Clustering allows the network to scale to a high count of sensors while also extending the network's life by permitting sensors to save energy by communicating with nearby nodes and spreading the load within the gateway nodes. The cost of communicating with every sensor in the network is assigned by gateways. The cost of connectivity together with the load on the gateways is used to construct clusters.

Bootstrapping as well as clustering are the two steps of network setup. Gateways find the nodes inside their transmission range during the bootstrapping process. The commencement of clustering is signaled by gateways sending out a message. All through the clustering procedure, we consider that sensor receivers are open. For avoiding collisions, every gateway begins clustering at a distinct point in time. In response, the sensors send out a message having their maximum transmission power, identifying their position and remaining energy. Every gateway does have a range defined for every node detected during this phase.

Gateways compute the cost of communicating with every node in the range established during the clustering stage. This data is subsequently shared throughout all the gateways. Following the receipt of data from the remaining gateways, every gateway begins clustering nodes depending on communication costs and the network load on the cluster. When the clustering is complete, the entire sensors are given the ID of the cluster to which they correspond. Every sensor contributes to just a single cluster because gateways exchange similar information throughout clustering. The complete traffic is routed via the gateways enabling inter-cluster communication.

Every gateway creates a group of complete nodes with which it may communicate. If a sensor T_k meets the below conditions, it corresponds to the range group *SSet* of gateway H_j:

$$T_k \in SSet_{H_j} \leftrightarrow \left[\left(S_{H_j} > e_{T_k \to H_j}\right) \wedge \left(S_{T_{k,max}} > e_{T_k \to H_j}\right)\right] \quad (9)$$

Here, S_{H_j} shows the gateway H_j's range, $S_{T_{k,max}}$ shows the sensor T_k's maximum range, and $e_{T_k \to H_j}$ shows the distance among both sensor T_k and the gateway H_j. Every node in the range group has a communication cost that the gateway calculates. The computed cost is based on the amount of communication energy used to convey s bits of data more than a $e_{T_k \to H_j}$ distance. The cost D_{k,H_j} linked with sensor T_k, estimated by gateway H_j, utilizing the energy model is:

$$D_{k,H_j} = F_{ty} + F_{ry} = \left(\alpha_1 + \alpha_{amp}\left(e_{T_k \to H_j}\right)^2\right) * s + \alpha_s * s \quad (10)$$

All of the nodes in *SSet* and their related costs are recorded in a per cluster record. All gateways share the record in order to get a global understanding of the network. There are two types of nodes in the system, based on the sensor range: exclusive nodes, which can only interact with single gateway, and multi-gateway nodes, which can connect with several gateways. The count of gateways with which a node may communicate is described as its *reach*. Since exclusive nodes must be handled by a gateway, the initial step toward clustering is to segregate them from the others. To accomplish so, gateways create an exclusive group called *FSet*, which is made up of nodes that meet the below criteria:

$$T_k \in FSet_{H_j} \leftrightarrow \left[\left(T_k \in SSet_{H_j}\right) \wedge \left(\forall l \neq j, T_k \nexists\, SSet_{H_l}\right)\right] \quad (11)$$

Computation load QM_{H_j} and communication load DF_{H_j} owing to sensors in the cluster produce the load on a gateway, which is described as:

$$M_{H_j} = g\left(QM_{H_j}, DF_{H_j}\right) \quad (12)$$

The processing load on a gateway is caused by the energy needed to manage data from the sensors in the cluster. The aggregate of the communication costs of entire sensors in the cluster is used to compute the gateway's communication energy, or DF_{H_j}. That is to say,

$$DF_{H_j} = \sum_{k=0}^{o} D_{k,j} \quad (13)$$

QM_{H_j} of a gateway is linearly proportional to the count of sensors o in the cluster because we presume that the entire sensors are similar and generate information at the similar pace. It indicates that to equalize load in the network, the count of nodes in a cluster and the communication energy needed per gateway must be balanced. We adopt an objective function that minimizes the variance of the cardinality of every cluster in the network to maintain the system near to the average load. That is to say,

$$\sigma^2 = \frac{1}{H}\sum_{j=0}^{H}\left(Y - Y'\right) \tag{14}$$

Here, σ^2 shows the system's load variance, Y shows the cardinality of the gateway H_j, and Y' shows the average cardinality along with the node in question, and H shows the total count of gateways in the network.

4.4 *Shortest path routing in WSN*

By passing through every node, the energy-efficient shortest route method discovers the shortest path to entire nodes. It is primarily made up of two tables: the Distance table and the Sequence table.

Distance table: It is employed to calculate the shortest path within any two nodes.
Sequence table: It is utilized to determine the shortest path within two nodes.

This protocol generates a table of size N (count of Nodes) that it refreshes every second O (N^3). It prevents packet flooding and thus reduces energy use. It operates by selecting a network node as a shortcut repeatedly. The shortest path cost is adjusted if the path via the waypoint is lesser than the existing shortest path. The input to this algorithm is a $E0$-length matrix. If an edge exists among nodes j and k, the length of the edge is stored in the matrix $E0$ at the relevant coordinates.

The diagonal of the matrix is entirely made up of zeros. If no edge connects the edges j and k, the location (j, k) includes positive infinity. To put it another way, the matrix contains the lengths of complete pathways among nodes that do not pass via an intermediary node. The matrix is updated after every iteration of the energy effective procedure, so it now includes the lengths of pathways between entire pairs of nodes utilizing a consistently expanding collection of intermediate nodes. The initial iteration of the algorithm produces the matrix $E1$, which comprises pathways between entire nodes that use precisely one (predefined) intermediate node. $E2$ utilizes two preset intermediate nodes to store lengths, while Eo utilizes o intermediate nodes to store lengths.

$$E[j][k] = Min\{E[j][k], (E[j][l] + E[l][k])\} \tag{15}$$

The variables j, k, and l in this equation represent the row index, column index, as well as sequence count, respectively. Only if the current cost is greater than the l^{th} Sequence index is $E[j][k]$ updated. The sequence count in the sequence tables is changed whenever the cost is modified. The shortest path among two nodes in a network is a series of linked nodes with the least number of edges connecting them. The shortest path routing algorithm is in Algorithm 2.

```
# dis[j,k] shows optimal distances from j^th vertex to k^th vertex
# Begin with entire single edge paths
For j = 1 to o do
For k = 1 to o do
        Dist[j,k] = weight[j,k]
For l = 1 to o do
# l shows the intermediate vertex
For j = 1 to o do
        For k = 1 to o do
          If (dis[j,l] + dis[l,k] < dis[j,k]) then
          # shorter path?
             dis[j,k] = dis[j,l] + dis[l,k]
```

5. Results

5.1 *Experimental setup*

NS2 software is used to test the performance of the suggested WBC approach. With 300 nodes placed in the network, a network model is built. The nodes are distributed at random and periodically. Only three performance metrics were verified in this article to examine network behavior using a WBC approach. The simulation is run five times having a distinct count of nodes each time. Table 1 shows the factors that were considered for the design.

5.2 *Energy analysis*

After every round of operation, the energy efficiency is initially tested. The leftover energy, also known as residual energy, is then computed and the results are presented. The energy value is found to be decreasing after every cycle of operation, according to the results. By deducting the total of

Table 1. Simulation Parameters.

Target area	1000×1000 m^2
Node deployment	Random
Bandwidth	19 kbps
Node count	300
BS location	500×500 m^2
Antenna direction	Omni directional
Packet size	3000 bits
Initial energy	1J

Table 2. Energy Analysis.

Methods/Energy (mJ)	Number of rounds having different node count				
	10	20	30	40	50
Cognitive SW-WSN [29]	96	97.1	98.2	98	95.9
PEAR [30]	96.6	97.4	98.9	97.9	96.4
MULE [31]	98.2	98.1	98.9	97	96.7
DCDG-ARW [32]	96	97.2	97.4	95.4	97.4
WBC	96.1	96.9	97.8	96.6	97.2

Table 3. PDR Analysis.

Methods	Number of rounds having different node count				
	10	20	30	40	50
Cognitive SW-WSN [29]	24	29	38	54	61
PEAR [30]	21	27	43	54	60
MULE [31]	22	28	38	47	59
DCDG-ARW [32]	27	30	40	44	60
WBC	28	34	46	55	69

entire expended energy, every node's remaining energy is computed. The nodes' remaining energy is lowered from 99.8% to 96.4% of their original value. An initial energy of 100mJ is given to every node. The data transmission procedure is subjected from the source to the destination based on the WBC model. Table 2 provides the observed metrics of the energy.

5.3 *PDR analysis*

The simulation procedure is repeated numerous times, using various source as well as destination nodes in the network. To analyze network performance, the PDR is computed and validated every time. The count of data delivered to the message collected by the network's destination node is referred to as the PDR. The PDR estimated at every round is shown here. The quantity of PDR increases as the count of nodes in the process increases. The value of PDR is likewise directly related to the count of nodes in the network, according to the results shown in table 3.

5.4 *Throughput analysis*

The network's throughput is estimated and presented in the graph shown here in the same way as the energy together with the PDR are. The entire amount of data transmitted in the network is referred to as throughput. In table, the acquired throughput numbers have been combined. When the number of nodes in the network grows, so does the throughput. The percentage of throughput is directly proportional to the count of nodes engaged in the network function, as shown by the preceding statistics in table 4.

5.5 *Accuracy analysis*

The accuracy analysis for the provided WSN CH selection model is presented here. At distinct node counts, the innovative WBC outperforms the other examined current techniques. At the 50^{th} round, the new WBC outperforms the cognitive SW-WSN, PEAR, MULE, and DCDG-ARW techniques by 0.10%, 0.13%, 0.08%, and 0.18%, respectively. As a result, the accuracy analysis of the innovative WBC is superior to various approaches for the investigated WSN CH selection method. As a result of improved accuracy, weight-based clusters and cluster heads are observed in figure 2 where the bold circles denote the cluster heads of the corresponding node group.

5.6 *Precision analysis*

The accuracy of the newly built WSN CH selection model is examined in this section. Table and the figure show the results of the analysis. The new WBC outperforms various existing approaches at varied node counts. The novel WBC outperforms the cognitive SW-WSN, PEAR, MULE, and DCDG-ARW techniques by 0.07%, 0.11%, 0.07%, and 0.09%, correspondingly, in the 30^{th} round, as shown here. Therefore, for the analyzed WSN CH selection model, the precision analysis of the new WBC is clearly better to the remaining techniques as evident from table 5.

Table 4. Throughput Analysis.

Methods/ Throughput (Kbps)	Number of rounds having different node count				
	10	20	30	40	50
Cognitive SW-WSN [29]	2010	3260	3540	4840	6100
PEAR [30]	1910	3340	3710	5260	6660
MULE [31]	2000	3100	3690	5040	6450
DCDG-ARW [32]	1800	3080	4010	4850	6540
WBC	2140	3600	4500	5940	6090

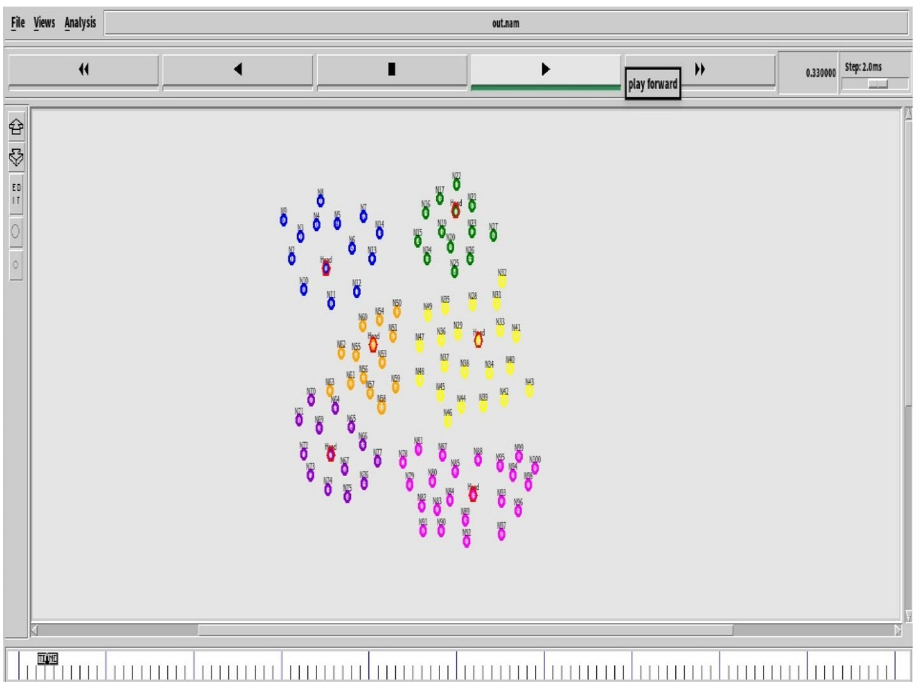

Figure 2. Clustering and clusterhead formation using proposed WBC.

Table 5. Precision Analysis.

Methods	Number of rounds having different node count				
	10	20	30	40	50
Cognitive SW-WSN [29]	0.88	0.85	0.89	0.91	0.91
PEAR [30]	0.91	0.87	0.93	0.88	0.94
MULE [31]	0.89	0.89	0.94	0.92	0.90
DCDG-ARW [32]	0.90	0.88	0.91	0.90	0.93
WBC	0.92	0.90	0.94	0.93	0.95

6. Conclusion

A unique WBC technique for data gathering, node clustering, and load balancing in WSN was suggested in this study. The WBC method was broken down into three steps or modules. The first stage focused on clustering and selecting a CH. The emphasis of the second stage was distance calculations. The third stage focused on shortest path calculation, which reduced energy usage, packet loss, latency, and load balancing by determining the shortest path for data transmission. Using the NS2 tool, the performance of the recommended WBC was compared to standard benchmark approaches. Multiple metrics have been generated and analyzed, and the proposed WBC consistently surpassed the competition.

Abbreviations

WSN	Wireless Sensor Network
QoS	Quality of Service
WBC	Weight Based Clustering
CH	Cluster Head
SFLA	Shuffled Frog Leaping Algorithm
CZ	Candidates Zone
EGNs	Energy Gauge Nodes
RTT	Round Trip Time
SPR	Strength of Packet Reply
RREQ	Route REQuest
MS	Mobile Sink
UCB	Upper Confidence Bound
RREP	Route REPly
UM-MAB	Multi-User Multi-Armed Bandit
PLR	Packet Loss Ratio
VH	Virtual Head
EBAR	Energy-Efficient Load Balancing Ant-based Routing
GWO	Grey Wolf Optimization
MPAR	Multi-sink Placement and Anycast Routing
GA	Genetic Algorithm
EMPAR	Extended Multi-sink Placement and Anycast Routing
PSO	Particle Swarm Optimization
AODV	Ad hoc On-Demand Distance Vector
ACO	Ant Colony Optimization
CBERP	Cluster Based Energy Efficient Protocol
BS	Base Station
CS	Compressive Sensing
ROI	Region of Interest
PDR	Packet Delivery Ratio
PEAR	Predictive Energy-Aware Routing
SW-WSN	Small-World Wireless Sensor Network
DCDG-ARW	Dynamic Compressive Data Gathering using Angle-based Random Walk
MULE	Mobile Ubiquitous Local Area Network Extensions

Author's contribution: The author is solely responsible for the experimental works conducted in this paper, drafting of the paper and presentation of all the sections.

Funding: The author did not receive support from any organization for the submitted work.

Availability of Data & Material The author hereby declare that no specific data sets are utilized in the proposed work.

Code availability: Since, future works are based on the custom codes developed in this work, the code may not be available from the author.

Declarations

Conflicts of interest: The author has no relevant financial or non-financial interests to disclose.

References

[1] Kamal A and Hamid Md Abdul 2017 Supervisory routing control for dynamic load balancing in low data rate wireless sensor networks. *Wireless Networks.* 23: 1085–1099

[2] Low C P, Fang C, Ng J M and Ang Y H 2008 Efficient load balanced clustering algorithms for wireless sensor networks. *Computer Communications.* 31(4): 750–759

[3] Kuila P, Gupta S K and Jana P K 2013 A novel evolutionary approach for load balanced clustering problem for wireless sensor networks. *Swarm and Evolutionary Computation.* 12: 48–56

[4] So J and Byun H 2017 Load-Balanced Opportunistic Routing for Duty-Cycled Wireless Sensor Networks. *IEEE Transactions on Mobile Computing.* 16(7): 1940–1955

[5] Kacimi R, Dhaou R and Beylot A L 2013 Load balancing techniques for lifetime maximizing in wireless sensor networks. *Ad Hoc Networks Journal.* 11(8): 2172–2186

[6] Selvakumar K and Pattabirani G 2019 A clustered fuzzy and dynamically well organized load balancing algorithm (CFDLB) for network life time enhancement in wireless sensor networks. *International Journal of Innovative Technology and Exploring Engineering.* 8(4): 472–479

[7] Palani U, Amuthavalli G and Alamelumangai V 2020 Secure and load-balanced routing protocol in wireless sensor network or disaster management. *IET Information Security.* 14(5): 513–520

[8] Wang Tianshu, Yang Xichen, Kongfa Hu and Zhang Gongxuan 2021 A Distributed Load Balancing Clustering Algorithm for Wireless Sensor Networks. *Wireless Personal Communications.* 120(4): 3343–3367

[9] Smys S and Haoxiang Wang 2021 A secure optimization algorihtm for Quality-Of-Service improvement in Hybrid wireless networks. *IRO Journal on Sustainable Wireless Systems.* 3(1): 1–10

[10] Chanak P, Banerjee I and Rahaman H 2015 Load management scheme for energy holes reduction in wireless sensor networks. *Computers and Electrical Engineering.* 48: 343–357

[11] Wang E, Li H and Zhang S 2019 Load balancing based on cache resource allocation in satellite networks. *IEEE Access.* 7: 56864–56879

[12] Zhao M, Yang Y and Wang C 2015 Mobile data gathering with load bal anced clustering and dual data uploading in wireless sensor networks. *IEEE Transactions on Mobile Computing.* 14(4): 770–785

[13] Kuila P and Jana P K 2012 Improved load balanced clustering algorithm for wireless sensor networks. In: *International Conference on Advanced Computing, Networking and Security, ADCONS 2011*, Springer, Berlin, Heidelberg, 2012, pp. 399 – 404, 2012

[14] Edla D R, Kongara M C and Cheruku R 2019 SCE-PSO Based clustering approach for load balancing of gateways in wireless sensor networks. *Wireless Networks.* 25: 1067–1081

[15] Edla D R, Lipare A and Cheruku R 2018 Shuffled complex evolution approach for load balancing of gateways in wireless sensor networks. *Wireless Personal Communications.* 98(4): 3455–3476

[16] Edla D R, Lipare A, Cheruku R and Kuppili V 2017 An Efficient Load Balancing of Gateways using Improved

Shuffled Frog Leaping Algorithm and Novel Fitness Function for WSNs. *IEEE Sensors Journal.* 17(20): 6724–6733
[17] Hawbani A, Wang X, Sharabi Y, Ghannami A, Kuhlani H and Karmoshi S 2021 LORA: Load-Balanced Opportunistic Routing for Asynchronous Duty-Cycled WSN. *IEEE Transactions on Mobile Computing.* 18(7): 1601–1615
[18] Chatterjee P, Ghosh S C and Das N 2017 Load Balanced Coverage with Graded Node Deployment in Wireless Sensor Networks. *IEEE Transactions on Multi-Scale Computing Systems.* 3(2): 100–112
[19] Liu X and Zhang P 2018 Data Drainage: A Novel Load Balancing Strategy for Wireless Sensor Networks. *IEEE Communications Letters.* 22(1): 125–128
[20] Adil M, Khan R, Almaiah M A, Binsawad M, Ali J, Al-Saaidah A and Ta Q T H 2020 An Efficient Load Balancing Scheme of Energy Gauge Nodes to Maximize the Lifespan of Constraint Oriented Networks. *IEEE Access.* 8: 148510–148527
[21] Zhang J, Tang J and Wang F 2020 Cooperative Relay Selection for Load Balancing With Mobility in Hierarchical WSNs: A Multi-Armed Bandit Approach. *IEEE Access.* 8: 18110–18122
[22] Li X, Keegan B, Mtenzi F, Weise T and Tan M 2019 Energy-Efficient Load Balancing Ant Based Routing Algorithm for Wireless Sensor Networks. *IEEE Access.* 7: 113182–113196
[23] Yunjian Tang, Weiren Shi, Jun Yi and Yanxia Wang 2011 Dynamic Load-balancing Algorithm of WSN for Data Gathering Application. *Computer Engineering and Applications.* 47(6): 122–126
[24] Mohajerani A and Gharavian D 2016 An Ant colony optimization based routing algorihtm for extending network lifetime in wireless sensor networks. *Wireless Networks.* 22(8): 2637–2647
[25] Yao Y, Cao Q and Vasilakos A V 2015 EDAL: An energy efficienty, delay-aware, and lifetime-balancing data collection protcol for heterogeneous wireless sensor networks. *IEEE/ACM Transactions on Networking.* 23(3): 810–823
[26] Gowri S, Anandhamala G S and Divya G 2014 Enhancing the Digital Data Retrieval System Using Novel Techniques. *Journal of Theoretical and Applied Information Technology.* 66(2): 481–489
[27] Janani S, Ramaswamy M and Samuel Manoharan J 2018 Clustered HEED scheme for congestion avoidance in cognitive radio sensor network. *Journal of Theoretical and Applied Information Technology.* 96(17): 5674–5684
[28] Sharma R, Vashisht V and Singh U 2019 EEFCM-DE: energy-efficient clus tering based on fuzzy C means and differential evolution algorithm in WSNs. *IET Communications.* 13(8): 996–1007
[29] Pandey O J and Hegde R M 2018 Low-Latency and Energy-Balanced Data Transmission Over Cognitive Small World WSN. *IEEE Transactions on Vehicular Technology.* 67(8): 7719–7733
[30] Jecan Eusebiu Eusebiu, Pop Catalin, Ratiu Ovidiu and Puschita Emanuel 2022 Predictive Energy-Aware Routing Solution for Industrial IoT Evaluated on a WSN Hardware Platform. *Sensors.* 22(6): 2107. https://doi.org/10.3390/s22062107
[31] Pandey O J, Mahajan A and Hegde R M 2018 Joint Localization and Data Gathering Over a Small-World WSN With Optimal Data MULE Allocation. *IEEE Transactions on Vehicular Technology.* 67(7): 6518–6532
[32] Shima Pakdaman Tirani, Avid Avokh and Jamshid Abouei 2022 Dynamic Compressive Data Gathering using Angle-based Random Walk in Hybrid WSNs. *Ad Hoc Networks.* 127 https://doi.org/10.1016/j.adhoc.2021.102770