

Performance Evaluation of Hybrid Data Aggregation in Wireless Sensor Networks

Nisha Devi

Research Scholar, Maharaja Agrasen University, Baddi, Solan, H.P.

Aparna N. Mahajan

Associate Professor, Maharaja Agrasen University, Baddi, Solan , H.P.

Abstract: In wireless sensor networks (WSN), energy consumption can be significantly reduced by the use of data aggregation (elimination of redundant data). Energy-efficient routing is often employed to determine the optimal path from the source to the destination, thereby avoiding redundant data and conserving energy for relaying the transmitted data. In current research, Energy-efficient routing will be carried out using optimization techniques and the randomization principle. Traditional methods, such as length-based and time-based data aggregation, are simple but result in redundant data, delayed transmissions, and significant energy consumption. The Cluster Head (CH)-based Data Aggregation technique proposed in this research divides the network into clusters, selects energy-efficient CHs, and carries out intelligent in-network data aggregation operations. This approach outperforms time and length-based methods in terms of reducing redundant transmissions and extending network life. From the result analysis, cluster-based aggregation is more efficient at handling data throughput, ensuring that data is transferred at higher rates across the network. The lower delay times with Cluster-based aggregation mean faster communication within the network.

Keywords: Data Aggregation, Redundant data, Energy-efficient data aggregation

1. INTRODUCTION

The advancement of smart sensors has been made easier by the increased global awareness of wireless sensor networks (WSNs) in recent years. Dispersed sensor nodes are capable of gathering data and returning it to the sink and end users. A multi-hop infrastructure-less architecture uses the sink to transport data back to the end user. The task manager node and the sink may communicate via satellite or the internet. There is a protocol stack that serves all purposes for both the sensor nodes and the sink. This stack of protocols manages power and routing awareness, combines information with networking protocols, successfully transfers power across wireless media, and enhances collaborative efforts among sensor nodes. Applications, transport layers, networks, data links, physical layers, mobility management, and task management are part of this protocol stack [1].

A sensor node's primary responsibility in a multi-hop network is to detect and transmit data to the destination, for which a suitable routing path is necessary. Routing protocols are needed in order to compute the appropriate routing path from the source node to the base station. In WSN, a variety of routing protocols are employed, and these are location-based, multipath-based, heterogeneity-based, hierarchical, data-centric, and QoS-based protocols. The routing protocols are designed to minimize the power consumption and resource limitations of the network devices, enhance the

wireless channel's time-varying quality, and lower the probability of packet loss and delay. Sensor nodes basically collect the data from the surroundings and process it locally when they are placed over a region. They can also sense physical surrounding parameters, including humidity, movement, temperature, pressure, and other conditions. After localized processing, the information is forwarded to the sinks. Each node is powered by a battery that has a limited capacity and is very difficult to replace or recharge when the nodes are placed in a certain area because of environmental conditions. For example, a sensor node deployed underwater, referred to as Underwater-WSN, can be supplied with a power supply, as there is no way to do that. Hence, battery prevention becomes a vital issue in WSN to enhance the network lifetime [2].

There are a number of reasons why a node uses more energy than others. These include idle listening, where nodes wake up and listen for incoming messages even when no data is being transmitted; data collisions, where two or more nearby stations want to transmit data and collision consumes more energy; overhearing, where many nodes in the sender's nearby may overhear the transmission of packets even though they are not the intended recipients of these transmissions; and control packet overhead; since control packets use a lot of energy during sending, receiving, and listening, it is advised that fewer control packets be used for data

transmission to minimize the energy consumption. When two nodes are within radio range of one another, they may detect redundant data and send it to the sink node. Redundant data means useless data or any duplicate or unnecessary information that uses resources like energy, bandwidth, and storage but does not contribute to the overall data gathering process. The sink node then finds it difficult to handle such a large amount of data, which causes energy loss. There are many ways to handle the redundant data, such as data aggregation, data compression, Hierarchical routing, redundancy elimination algorithms, and adaptive sampling. The data aggregation algorithm used in the current study aims to eliminate redundant data, reduce energy usage, and enhance the network lifetime. In this approach, data from several sensors is combined by intermediary nodes or cluster heads, who also remove duplicates and summarize the data before sending it to the base station

1.1 Data Aggregation in WSN:

Data collection is a process of obtaining information from the network's source or intermediate nodes. Because the collected data includes duplicated information, bandwidth is wasted if it is sent to the sink node in its current state. Additionally, it causes more energy/battery consumption and lowers the lifetime of the network [3]. Therefore, to increase the lifetime of the network, energy consumption by nodes must be decreased by limiting the number of packets they transmit. So, a data aggregation approach is used for this purpose. Data aggregation removes the unnecessary information from the collected data and forwards it to the sink node. Data aggregation algorithms like MAX, MIN, MEAN, and MEDIAN help to reduce the transmission of redundant data [4]. Data aggregation helps in energy conservation, network robustness enhancement, and redundancy reduction by reducing traffic load [5]. A sensor node or device may not have a large enough buffer to hold the incoming data due to low storage capacity, which causes packet drops once the buffer is filled. In addition to causing data loss, it places an extra load on network traffic since nodes must retransmit the dropped packets. In addition to causing unfavorable outcomes like higher latency, this causes nodes to lose energy more quickly. Data can be handled more efficiently by aggregating it and then sending it to the BS (base station) [6]. Gathering information, aggregating it, and sending the BS useful data are the main goals. The data delivery model and aggregation strategy have an impact on aggregation timing, which is essential for accurately representing data sensed by network nodes [7]. There are numerous ways to classify data aggregation, including network structure, network flow, service quality, and these are:

1.1.1 Data Aggregation Approaches:

1.1.2 1. Structured based Aggregation:

II. Flat Based Aggregation Approach: A network in which each sensor nodes share the have the same battery and functionalities. In these networks, data aggregation is carried out using data-centric routing, in which the sensors receive a query message from the sink. The specific application at issue determines which communication protocol is best [8].

III. Cluster Based Aggregation Approach: This technique makes use of reservation-based scheduling, node heterogeneity, and clustering. When using cluster-based aggregation, the network designates a cluster head to handle data aggregation. This method's primary goal is to aggregate data in huge networks in an energy-efficient manner. Instead of sending the data straight to the sink node in this instance, the sensor nodes send the data to the local cluster head, which then sends the aggregated data to the sink node. In large networks, this method dramatically lowers the energy consumption and inefficiency of the energy-constraint sensor nodes [9].

IV. Chain Based Networks: One of the hierarchical techniques of aggregation that makes up chain architecture is chain-based data aggregation. This method distributes energy uniformly, enables neighbour communication between sensor nodes, and rotates who gets to be the leader while sending data to the base station. The method of token passing is employed to select the leader. After receiving the token, that node transmits the information to the aggregator node, which then forwards it to the sink station [10].

V. Grid Based Approach: A group of sensors is designated as data aggregators in specific areas of the sensor network in grid-based data aggregation. Data is directly transmitted from the sensors inside a given grid to the data aggregator within that grid. Because of this, there is no communication between the sensors in a grid. [11].

VI. Tree based aggregation approach: Sensor nodes are arranged into a tree in a tree-based network, with data aggregation occurring at intermediate nodes along the tree and a condensed version of the data being relayed to the root node. Applications that require in-network data aggregation can benefit from tree-based data aggregation. Building an energy-efficient data aggregation tree is one of the main facets of tree-based networks [12]

2. Hierarchical based networks: The sink node in flat networks is overcome by hierarchical networks through data fusion at intermediate or special nodes. By lowering the quantity of messages sent to the sink, it enhances the network's energy efficiency. There are two more

categories for the Hierarchical approach: cluster architectures and chain architectures. There is no particular architecture used in the structure-less data aggregation process. Here, each node on the network can communicate with any other node [13].

3. **Length Based data aggregation:** Sensor nodes buffer and aggregate data according to the size or length of the collected data packets in Wireless Sensor Networks (WSNs) using a technique called length-based data aggregation. The quantity of data accumulated in a buffer is the basis for length-based aggregate, as opposed to time-based aggregation, which depends on a time trigger. The data is transferred and the aggregation is carried out when the buffer reaches a predetermined size threshold. This technique is intended to minimize network transmissions and maximize energy efficiency [11][10].
4. **Time based aggregation technique:** In Wireless Sensor Networks (WSNs), time-based data aggregation is a basic technique where sensor nodes send their aggregated data at predetermined intervals, independent of the quantity of data packets gathered. This technique works particularly well for networks with constant data creation rates when it comes to minimizing transmission redundancy and keeping a regular data reporting schedule [12].
5. **Structure less Aggregation:** No structure is maintained in structure-less data aggregation. In event-based systems, where the event region is subject to rapid changes, it is highly beneficial. Rebuilding the structure is not necessary in the event that any node fails. Making the routing choice to carry out data aggregation is the main disadvantage of structure less data aggregation [14].

In order to give a thorough examination of energy-efficient data aggregation methods in Wireless Sensor Networks (WSNs), this paper is structured into a number of logical and related sections. Every part has a distinct function in constructing the study's overall narrative and research findings.

Following this introductory section, a thorough literature review is presented in **Section 2**, which highlights the contributions made by earlier research on data aggregation in WSNs. Prominent methods such as Length-based Data Aggregation (LDA), Time-based Data Aggregation (TDA), and Cluster-based Data Aggregation (CDA) are covered in the review. It talks about the advantages and disadvantages of each technique and looks at how it has been used in previous studies. Particular focus is placed on the gaps in comparative analysis that this study seeks to fill.

In Section 3, the research's methodology is described. This section describes the buffer-based aggregation decision in LDA, the time-triggered aggregation process in TDA, and the logic of cluster creation and rotation in CDA. The energy model that is used to calculate energy consumption is also introduced, and the energy consumption, throughput, packet delivery ratio (PDR), end-to-end delay, and network lifetime are defined as the primary performance measures.

Section 4 discusses the simulation setup. The simulations used in this study are conducted with MATLAB, a sophisticated numerical computing environment that provides accuracy and versatility in the modeling of Wireless Sensor Networks (WSNs) in which 200 sensor nodes are set up. The packet size, communication range, and beginning energy of every node are all the same. The sink node is placed in the middle. The average throughput, total transmission delay, and number of effectively delivered packets are all monitored by the simulation for each of the three methods. The simulation results and comparative analysis, showing throughput, PDR, and delay graphs and tables for CDA, TDA, and LDA. In general, CDA exhibits faster throughput, PDR, better energy efficiency and a longer network lifetime. While LDA has low latency but may experience packet drops in situations with sparse traffic, TDA has moderate latency and performance.

Section 5 concludes the study, emphasizing the main conclusions and affirming that CDA is more suitable for energy-critical applications

2. LITERATURE REVIEW:

In wireless sensor networks (WSNs), batteries are the primary life indication. The process of transmitting data consumes up the most of energy in wireless sensor networks. Therefore, routing protocols for energy efficiency are required. This is considered by numerous authors to design energy-efficient WSNs. Various type of study are provided in an effort to reduce energy consumption.

The wide range of uses for Wireless Sensor Networks (WSNs) in healthcare, industrial control, military surveillance, and environmental monitoring has attracted a lot of research interest. Since sensor nodes have limited battery life, energy efficiency is still a major design challenge for WSNs. Among other methods, data aggregation algorithms including length-, time-, and cluster-based aggregation have demonstrated potential for increasing network lifetime and reducing energy use. An extensive review of previous studies on energy-efficient data aggregation techniques in WSNs is provided in this part, with

a focus on cluster-based techniques and a comparison with alternative approaches.

In order to increase network lifetime and improve data aggregation performance, a number of researchers have proposed many clustering algorithms. Mohamed et al. 2017 presented used two routing methods based on energy efficiency and connectivity aware routing to solve the hole problem in WSNs. From this method, the cost of topology reformation can be reduced and the lifespan of the network increased [15]. Thangaramya et al. 2019 developed an energy-efficient clustering approach based on neuro fuzzy. To employ the resource-efficient clusters and reduce data loss in neural fuzzy, they developed a membership function that took node energy and communication distance into account. [16].

Yazici et al. 2019 suggested a method to reduce the quantity of data that needs to be sent across the wireless multimedia sensor network by using intra-node processing. A new, power-efficient cluster-based routing technique for sensor networks was introduced. The primary techniques for controlling WSN topology, effective data aggregation, and routing clustering are used which uses the less energy [17]. Guo et al. 2019 suggested a reinforcement learning-based routing protocol. To maximize the network's energy efficiency and lifespan, the nodes in network were strengthened to choose the best routing path by using a reward and penalty policy [18]. Khan et al.2020 discussed a method of dynamic routing to address the problem of sensor node mobility in wireless sensor networks. Because of the range of human activities, the locations of sensor nodes on the human body are altered every second. Therefore, when nodes utilise the static routing technique, packet and energy losses happen during transmission. In order to address the issue, the authors employed information on node residual energy, hop count to sink distance, and throughput when nodes select the subsequent hop node to transmit data [19].

Daanoune et al. 2020 stated that Energy consumption in WSN is influenced by a number of parameters, including the transmission protocols, packet data delivery and battery capacity. Due to which the network have a short lifespan. Author suggested a new, enhanced LEACH routing protocol. This protocol uses a root CH with the highest energy capacity and the nearest location to the sink node to collect all data before sending it to the sink [20].

Further refining the clustering approach Umbreen et al. 2020 discussed that many clustering protocols are intended to extend the network lifespan. Certain parameters that significantly affect the energy consumption of the sensor determine which CH should be used. Each node's capacity is determined by taking into account its mobility level, remaining energy, sinking distance, and neighbour density. The inter-cluster communication protocol is multi-

hop/single-hop [21]. Haseeb et al.2020 resolved the security problems that arose when a wide-area Internet of things used the conventional routing method. They presented light-weight structure-based data aggregation routing, a secure protocol that makes use of the conventional routing protocols' in-route data aggregation for data routing [22].

Gandhi et al. 2020 suggested a fruit fly optimization technique which is used to dynamically relocate the mobile sink inside a region of a grid-based clustered network. The simulations showed that the suggested data aggregation method is much better in terms of energy usage and network lifetime [23]. Ullah et al. 2020 suggested that the Mahalanobis distance-based radial basis function is applied to the ELM's projection stage in order to reduce the training process' instability. Prior to being delivered to the cluster head, the data from each sensor node is filtered using a Kalman filter [24]. Yun et al. 2021 proposed the Energy efficient DA scheme for optimal communication between the nodes. The authors developed the Q-learning, reinforcement learning, ML technique for communication. The energy consumption using the proposed technique is 3.878J using the Grid topology [25]. Bhushan et al. 2021 discussed that to aggregate different types of data packets and to increase the energy efficiency, FAJIT has priority to resolving the parent node selection issue. Based on the candidate nodes with the least amount of dynamic neighbours, parent nodes are chosen. When there are an equal number of dynamic neighbours, fuzzy logic is used [26].

Jain et al. 2022 proposed two-vector normalized quintile regression (NQR) based data prediction model, this method is used to effectively achieve energy savings in synchronous data collection cycles. The proposed NQR methodology is demonstrated to have better energy efficiency, greater prediction accuracy, more positive predictions with good data quality, and longer network lifetime when compared to state-of-the-art methodologies [27].

William et al. 2022 stated that in order to achieve scalability and QoS optimization with the least amount of money spent (SCADA-ML), a new routing protocol utilizing machine learning was designed. The optimal CH for a specific cluster is determined by a given sensor node parameter, such as residual energy, distance from the base station, and allotted bandwidth, which is used to guide the ANN architecture. Through efficient data aggregation on each cluster's CH nodes (ICA), machine learning is employed to reduce cluster energy consumption [28].

Raj et al. 2023 represented a special unsupervised neural network design known as Partly-Informed Sparse Auto encoder (PISAE), which tries to reconstruct all sensor values from selected prime numbers. Here CORP (cross-layer based opportunistic Routing protocol) approach is used to determine the most effective course of action Idrees, A. K et

al. 2019 discussed the better data aggregation based on two level data aggregation scheme. At the sensor node, the divide and conquer method is used to eliminate the redundant data from the measured data before sending it to the cluster head. The cluster head uses an improved K-means technique to aggregate the sensor node data sets it has received into clusters of closely related data sets before sending the best representative set from each group to the base station [30].

Babu et al. 2020 discussed machine learning while using fuzzy logic for formation of CH. The Cluster Head (CH) is also elected while the Integration of Distributed Autonomous Fashion with Fuzzy If-then Rules (IDAFFIT) algorithm is proposed for clustering. ASLPP-RR, an adaptive source location privacy preservation technique, is proposed for routing in this context. Additionally, end-to-end secrecy and integrity are maintained while using the Secure Data Aggregation based on Principle Component Analysis (SDA-PCA) algorithm [31]. Nguyen, N. T et. al. 2016 stated the Maximum Lifetime Data Aggregation Tree Scheduling (MLDATS) problem. in which a fixed number of data are allowed to be aggregated into one packet, the scheduling of virtual data aggregation trees to maximize network lifespan [32]. Pham, V. T. et. al. 2022 discussed the problem of using minimum latency scheduling to aggregate and report data to the sink without data collision in multiple-data-type WSNs having unidirectional links, The Relative-Collision-Graph-Based Scheduling Algorithm (RCGBSA) is used in the attributes of data flows among sensors that are found and integrated into the relative collision graph [33].

After an extensive review, it came to light that the majority of studies focused on using decision tree, enhanced routing, and clustering techniques to achieve energy-efficient data transmission. However, there is still an absence of dependable node-to-node links and efficient bandwidth use using these techniques. Even though a two-level data aggregation scheme-based on advanced data aggregation method was used, every time when system perform calculations and data aggregation process was carried out, which was very time-consuming and resulted in high computational complexity. Machine learning and data analysis are two modules which are require to make the physical objects and environment autonomous. Data analysis would analyze all the data that is created over time to figure out the previous patterns and be more effective in the future, whereas machine learning would develop strategies to assist learning in various devices of the network to make them automated and self-standing and provide enhanced network lifetime[34][35][36]. In WSN energy efficiency is the primary issue for the effective operation of node because they have limited energy and are hard to recharge once installed. The sensor nodes cannot be continuously recharged due to the environmental conditions, as it is not feasible to manually access the locations where the sensor nodes are installed. For example, a sensor node placed underwater cannot be repeatedly recharged since it is not possible to get there. As a result, battery prevention becomes a crucial consideration for WSN.

Table:1 literature review

Author & Year	Method / Approach	Focus Area	Key Contribution / Outcome
Mohemed et al. (2017)	Energy-efficient & connectivity-aware routing	Hole problem in WSNs	Reduced topology reformation cost, improved network lifespan
Thangaramya et al. (2019)	Neuro-fuzzy clustering	Energy & distance-based membership function	Resource-efficient clusters, reduced data loss
Yazici et al. (2019)	Intra-node processing + power-efficient cluster-based routing	Wireless multimedia sensor networks	Reduced transmission data, energy-efficient clustering
Guo et al. (2019)	Reinforcement learning routing	Reward & penalty policy	Optimal path selection, increased energy efficiency & network lifetime
Khan et al. (2020)	Dynamic routing	Sensor node mobility (human activity)	Reduced packet & energy loss via energy, hop count, and throughput-aware routing
Daanoune et al. (2020)	Enhanced LEACH protocol	Root CH selection (highest energy, nearest sink)	Improved energy efficiency, prolonged lifetime
Umbreen et al. (2020)	Multi-parameter CH selection	Mobility, energy, distance, density	Improved energy efficiency, flexible inter-cluster communication

Gandhi et al. (2020)	Fruit Fly Optimization (mobile sink)	Grid-based clustered WSN	Better energy usage & extended lifetime
Ullah et al. (2020)	Mahalanobis RBF + Kalman filter	Data filtering before aggregation	Reduced instability, improved data quality
Yun et al. (2021)	Q-learning ML-based aggregation	Grid topology	Energy-efficient DA scheme (3.878J), optimized communication
Bhushan et al. (2021)	FAJIT + fuzzy logic	Parent node selection	Energy-efficient aggregation with dynamic neighbour consideration
Jain et al. (2022)	Normalized Quintile Regression (NQR)	Data prediction for aggregation	Better energy efficiency, prediction accuracy, longer lifetime
William et al. (2022)	ML-based SCADA-ML protocol (ANN)	Scalability & QoS	Efficient CH selection, reduced cluster energy
Raj et al. (2023)	PISAE (unsupervised neural net) + CORP	Opportunistic routing	Enhanced energy efficiency using sparse auto-encoder
Idrees et al. (2019)	Two-level aggregation (improved K-means)	Redundant data elimination	Reduced transmission, efficient aggregation
Babu et al. (2020)	Fuzzy logic + IDAFFIT + SDA-PCA	CH selection & secure aggregation	Improved clustering, energy-saving, data security
Nguyen et al. (2016)	MLDATS scheduling	Tree-based aggregation	Maximized network lifespan with scheduling
Pham et al. (2022)	RCGBSA scheduling	Latency & collision reduction	Collision-free, efficient aggregation with multiple data types

Table 1 gives a detailed literature review on WSN clustering and data aggregation which reflects a strong trend of technique progression over the years. The initial research (2016–2019) concentrated on enhancing routing and clustering via energy-conscious protocols (LEACH optimizations, connectivity-based routing, fuzzy/neuro-fuzzy clustering) to eliminate redundancy and extend network lifetime. Between 2019–2020, studies moved towards reinforcement learning, mobility dynamic routing, optimization techniques (fruit fly, Kalman filter), and light-weight secure aggregation that enhanced energy efficiency, security, and reliability. Current research (2021–2023) focuses on machine learning and AI-based solutions (Q-learning, ANN, NQR regression, autoencoders) that optimize prediction accuracy, smart CH selection, scalability, and energy optimization.

Overall, although these approaches manage to enhance energy efficiency, reliability of data, and network lifespan, discrepancies exist in providing reliable node-to-node connections, bandwidth usage, and minimizing computation complexity in real-time aggregation.

3. RESEARCH METHODOLOGY:

WSNs face many challenges like energy efficiency, security, scalability, limited bandwidth, complexity, cost and longevity due to their deployment conditions and limited sources. As different data aggregation schemes were discussed in previous section like cluster based, chain based, grid based, length based, time based etc.

Length-Based Aggregation (LDA)

- Each CH maintains a buffer B_i .
- Packets from member nodes are added to B_i .
- When $B_i \geq \theta_p$ packets are aggregated and sent to the sink.

Time-Based Aggregation (TDA)

- Each CH starts a timer T .
- Incoming packets are stored in a buffer.
- When $t \geq T$, aggregation occurs regardless of buffer size.

Hybrid(Cluster-Based) Aggregation

- Combines both thresholds (θ_p and T).

The **Energy Model** is used for calculating transmission and reception energy:

$$E_{rt}(k, d) = E_{ee} \cdot k + \epsilon_a \cdot k \cdot d^n$$

$$E_{rt}(k) = E_{ee} \cdot k$$

- where: k = packet size (in bits)
- d = distance between nodes
- E_{ee} = electronics energy per bit
- ϵ_a = amplifier energy
- n = path loss

- Aggregation occurs when **either** condition is satisfied.

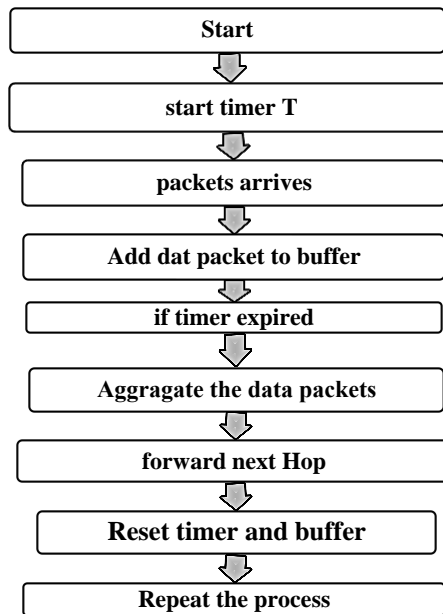


Figure 1: Flowchart Hybrid data based aggregation

Figure 1 shows Cluster-based Aggregation technique which combines time and length based data aggregation model. Every CH carries a buffer and a timer, and either threshold can cause aggregation. This technique saves energy by lowering the transmission frequency. Aggregation takes place at predetermined intervals, and incoming packets are kept in a buffer. After that, the combined packet is sent on to the following hop. Although this method offers reliable aggregation

4. Simulation results and comparative analysis:

MATLAB was used to perform the simulation for a 200-node . Key performance parameters energy consumption, packet delivery ratio (PDR), end-to-end delay, and throughput, were used to assess three aggregation strategies: cluster-based (CH-based), packet length-based (LDA), and time-based (TDA).

Aggregation Algorithms in proposed work:

The proposed work aims to improve the energy efficiency and network lifetime of WSN using CH based data aggregation over the time based and length based data aggregation techniques. length-based and time-based aggregation techniques have many drawbacks like: high energy usage as a result of inefficient or frequent transmissions, data loss and increased delay, inability to adjust to changing environmental conditions, low scalability in networks with a high density or size.

Algorithm 1 Packet Length-Based Aggregation

Require: Packet buffer B_i , Threshold θ_p

- 1: **for** each Cluster Head (CH) **do**
- 2: Initialize $B_i \leftarrow 0$
- 3: **for** each incoming packet p **do**
- 4: $B_i \leftarrow B_i + 1$
- 5: **if** $B_i \geq \theta_p$ **then**
- 6: Aggregate B_i packets
- 7: Forward aggregated packet to next hop
- 8: Reset $B_i \leftarrow 0$

```

9:     end if
10:  end for
11: end for
    
```

The aggregator node receives packets from the sensor nodes once they have sensed environmental data. At the aggregator node, a buffer is used to temporarily hold the incoming packets. The aggregator keeps track of the buffered packets (B_i) overall length. As long as the buffer's total packet length remains below the packet length threshold (θ_p), aggregation will continue. When the threshold is exceeded by adding another packet, the aggregator sends the buffered packets as a single aggregated packet. The packet may also be sent by the aggregator if a pre-established timeout happens before it reaches the threshold in order to avoid excessive delays. Next node or sink receives the aggregated packet. The aggregator repeats the procedure after clearing the buffer.

Algorithm 2 Time-Based Aggregation

Require: Timer T , Current Time t , Buffer B_i : **for** each Cluster Head (CH) **do**

```

2:     Start timer  $T$ 
3:     while true do
4:         if packet  $p$  arrives then
5:              $B_i \leftarrow B_i + 1$ 
6:         end if
7:         if  $B_i \geq \theta_p$  or  $t \geq T$  then
8:             Aggregate packet sin  $B_i$ 
9:             Forward aggregated packet to next hop
10:            Reset  $B_i \leftarrow 0$ , Restart timer  $T$ 
11:        end if
12:    end while
13: end for
    
```

The aggregation process is individually operated by each Cluster Head (CH) node in the network. The CH initiates the aggregation window (waiting period) by setting a timer T . For continuous functioning, the CH keeps an eye out for incoming packets (infinite loop). A packet p is saved in the buffer B_i for aggregation at a later time if it reaches the CH. The CH determines if the aggregate time window has ended when the current time t reaches or surpasses the timer value T .

The CH: - Aggregates all of the packets in B_i if the buffer is not empty.

-Transfers the combined packet to the subsequent sink or node.

- Resets the timer T for the subsequent aggregation cycle and clears buffer B_i .

For continuous packet aggregation, the procedure is repeated indefinitely.

Hybrid(Cluster Based) Aggregation Algorithm (Packet Count + Time-Based):

Algorithm3 Hybrid Aggregation

Require: Packet buffer B_i , Threshold θ_p , Timer T , Current Time t

```

1: for each Cluster Head (CH) do
2: Initialize  $B_i \leftarrow 0$ , Start timer  $T$ 
3:     while true do
4:         if packet  $p$  arrives then
5:              $B_i \leftarrow B_i + 1$ 
6:         end if
7:         if  $B_i \geq \theta_p$  or  $t \geq T$  then
8:             Aggregate packet sin  $B_i$ 
9:             Forward aggregated packet to next hop
10:            Reset  $B_i \leftarrow 0$ , Restart timer  $T$ 
11:        end if
12:    end while
13: end for
    
```

The CH initiates the timer T and sets the packet buffer B_i to zero. The CH can process packets for as long as it wants once continuous operation starts. The number of packets is increased if a packet p arrives and is added to the buffer. Aggregation can be done in either of the two cases : -The number of packets in the buffer reaches or above the packet count threshold, or θ_p and A timeout happens when the timer T expires. All of the packets in buffer B_i are combined after being triggered. The combined packet is sent to the sink node or subsequent hop. The timer is restarted for the subsequent aggregation cycle when buffer B_i has been cleared. The procedure is repeated for ongoing data aggregation.

The suggested method relies on a number of QoS metrics such as throughput, PDR, and delay. The result compares the effectiveness of two current approaches—time-based and length-based data aggregation with the performance of CH based data aggregation.

- a. Throughput: Amount of useful aggregated data transmitted per unit time

$$\text{Throughput} = \text{Total aggregated data}_{\text{received}} / \text{time}$$

- b. PDR: It is the proportion of packages received at the sink to the number of packets sent by nodes

$$\text{PDR} = \frac{\text{packets received at sink}}{\text{packets transferred}}$$

- c. Delay: time (in seconds) required for a packet to travel from its sender to its recipient.

$$\text{Delay} = \frac{\text{time (total packet received)}}{\text{packet transferred}}$$

Analyzing the results for throughput, Packet Delivery Ratio (PDR), and delay across different aggregation methods for nodes ranging from 50 to 200 nodes gives us a clear picture of which method performs best. When it comes to throughput, the Cluster Head (CH) based aggregation method consistently delivers better performance compared to the length-based and time-based aggregation methods

noticeably higher than the length-based aggregation, which averages about 5,098 units, and the time-based aggregation, which comes in at around 4,916 units. This means that CH based aggregation is more efficient at handling data throughput, ensuring that data is transferred at higher rates across the network.

Table 2: Comparison based on throughput for different data aggregation methods

No of nodes	Throughput CH Based	Throughput Length Based	Throughput Time Based Aggregation
50	5247.9340	4884.2156	4741.4385
100	5557.9275	4687.6686	4775.8032
150	5538.1305	5159.1067	5180.4014
200	5579.6102	4790.3623	5047.0876

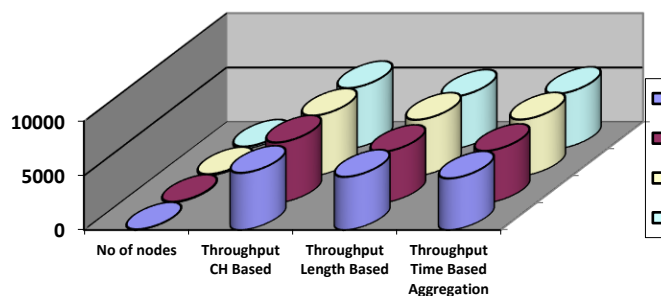


Figure3: Bar Graph for Comparison based on throughput for different data aggregation method

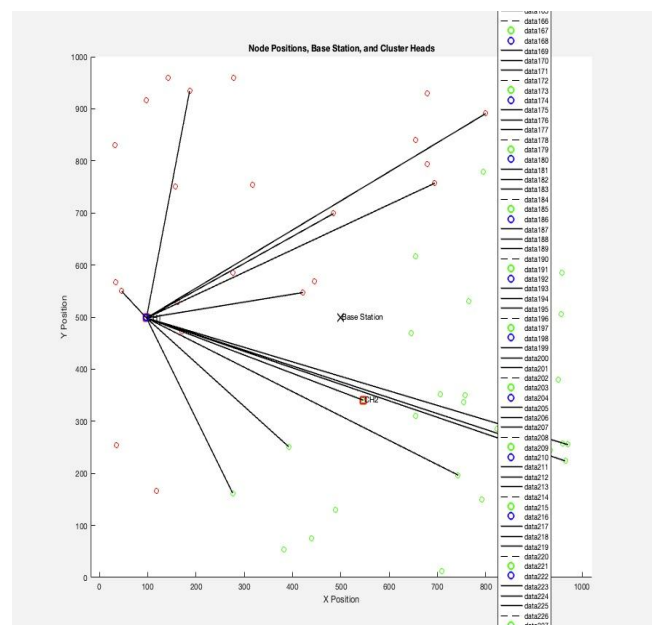


Figure 2: Clustering Implemented scenario

This Figure 2 illustrates the cluster-based data aggregation technique, in which CHs aggregate data from member nodes prior to transmitting it to the BS, thus decreasing transmission cost, increasing network lifespan and energy efficiency.

A comparative analysis is conducted against two studies for each of the three QoS parameters employed in the study in order to support the efficacy of the suggested work. Table 1 shows the throughput analysis for three investigations, including the proposed and current ones. Following this are Tables 2 and 3, which present the parametric values of the delay analysis and PDR, respectively.

Table 2 and figure 3 shows on average, the throughput for CH based aggregation is around 5,631 units. This is

For Packet Delivery Ratio (PDR), in table 3 and figure 4, CH based aggregation again proves to be the most effective. The average PDR for CH based aggregation is roughly 0.477, whereas the length-based aggregation averages about 0.451, and the time-based aggregation is slightly lower at around 0.429. This indicates that CH based aggregation does a better job at ensuring packets successfully reach their destinations, making it more reliable than the other methods.

Table 3: Comparison based on Packet delivery rate for different data aggregation methods

No of nodes	PDR CH Based	PDR Length Based	PDR Time Based Aggregation
50	0.4577	0.4369	0.4307
100	0.4781	0.4594	0.4607

150	0.4638	0.4202	0.4010
200	0.5070	0.4887	0.4233

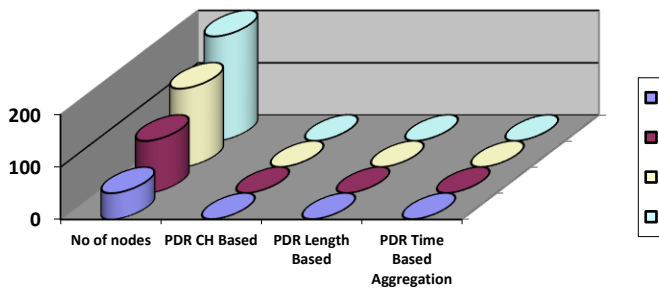


Figure 4: Bar Graph for Comparison based on PDR for different data aggregation methods

Table 4: Comparison based on Delay for different data aggregation methods

No of nodes	Delay CH Based	Delay Length Based	Delay Time Based Aggregation
50	54.959	55.950	61.580
100	46.102	56.685	52.361
150	46.848	49.886	52.916
200	44.497	47.038	53.161

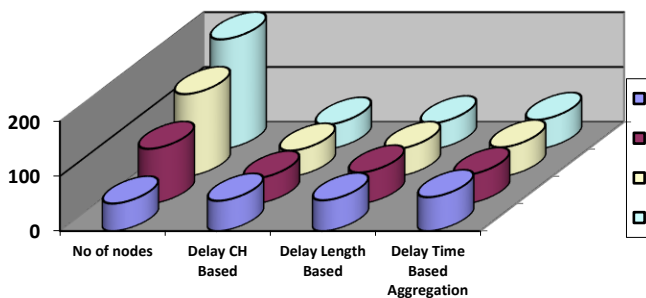


Figure 5: Bar Graph for Comparison based on Delay for different data aggregation methods

In table 4 and figure 5, for Packet Delivery Ratio (PDR), CH based aggregation again proves to be the most effective. The average PDR for CH based aggregation is roughly 0.477, whereas the length-based aggregation averages about 0.451, and the time-based aggregation is slightly lower at around 0.429. This indicates that CH based aggregation does a better job at ensuring packets successfully reach their destinations, making it more reliable than the other methods.

When we look at delay, CH based aggregation shows lower delay times, which highlights its efficiency even further. The

average delay for CH based aggregation is about 48.101 milliseconds. This is significantly lower than the length-based aggregation, which averages around 52.889 milliseconds, and the time-based aggregation, which averages about 54.503 milliseconds. The lower delay times with CH based aggregation mean faster communication within the network.

5. Conclusion and future scope:

Energy conservation is a critical factor due to the limited power resources of sensor nodes. By minimizing the amount of data transmitted and processed, energy consumption is significantly reduced, leading to extended network lifetime and improved efficiency.

The current research suggests a Cluster Head-Based Data Aggregation method that significantly improves the conventional length-based and time-based approaches. The framework allows real-time applications, lowers energy usage, and increases network lifetime by integrating energy-aware CH selection, dynamic rotation, and in-network aggregation. In the future, machine learning may be used at CHs for dynamic thresholding and adaptive data fusion. The effectiveness of the suggested work will be assessed by simulation, and network performance will be assessed through analysis of the simulated outcome.

6. Conflicts of interest declaration

All the authors declare that they have no conflicts of interest in this research.

7. Declaration of generative AI and AI-assisted technologies

Statement: During the preparation of this work the author(s) used MATLAB in order to assist in data analysis, code generation, or report writing. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Statement: The author(s) declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

8. REFERENCES:

- [1] M. A. Matin and M. M. Islam, "Overview of wireless sensor network," *Wirel. Sens. Netw.-Technol. Protoc.*, vol. 1, no. 3, 2012
- [2] W. Dargie and C. Poellabauer, *Fundamentals of Wireless Sensor Networks: Theory and Practice*. John Wiley & Sons, 2010
- [3] Ahmed A, Abdullah S, Bukhsh M, Ahmad I, Mushtaq Z. An energy-efficient data aggregation mechanism for IoT secured by blockchain. *IEEE Access*. 2022
- [4] Dagar M, Mahajan S. Data aggregation in wireless sensor network: a survey. *International Journal of Information and Computation Technology*. 2013
- [5] Begum BA, Nandury SV. Data aggregation protocols for WSN and IoT applications—A comprehensive survey. *Journal of King Saud University-Computer and Information Sciences*. 2023, <https://doi.org/10.1016/j.jksuci.2023.01.008>
- [6] Gavel S, Charitha R, Biswas P, Raghuvanshi AS. A data fusion based data aggregation and sensing technique for fault detection in wireless sensor networks. *Computing*. 2021, <https://doi.org/10.1007/s00607-021-01011-y>
- [7] Krishnamachari L, Estrin D, Wicker S. The impact of data aggregation in wireless sensor networks. In *Proceedings 22nd international conference on distributed computing systems workshops 2002 Jul 2* (pp. 575-578). IEEE.
- [8] Pandey V, Kaur A, Chand N. A review on data aggregation techniques in wireless sensor network. *Journal of Electronic and Electrical Engineering*. 2010 Aug;1(2):01-8.
- [9] Maurya MK, Shivhare A, Ali A, Mishra A, Kumar M. Cluster based smart random walk for data aggregation in wireless sensor network. In *2022 IEEE 15th International Symposium on Embedded Multicore/Many-core Systems-On-Chip (MCSoc) 2022 Dec 19* (pp. 98-104). IEEE.
- [10] Du K, Wu J, Zhou D. Chain-based protocols for data broadcasting and gathering in the sensor networks. In *Proceedings International Parallel and Distributed Processing Symposium 2003 Apr 22* (pp. 8-pp). IEEE.
- [11] Wang NC, Chen YL, Huang YF, Chen CM, Lin WC, Lee CY. An energy aware grid-based clustering power efficient data aggregation protocol for wireless sensor networks. *Applied Sciences*. 2022 Sep 30;12(19):9877.
- [12] Wang NC, Lee CY, Chen YL, Chen CM, Chen ZZ. An Energy Efficient Load Balancing Tree-Based Data Aggregation Scheme for Grid-Based Wireless Sensor Networks. *Sensors*. 2022 Nov 29;22(23):9303.]
- [13] Nalayini P, Prakash RA. Hierarchical Data Aggregation with Data Offloading Scheme for Fog Enabled IoT Environment. *Computer Systems Science & Engineering*. 2023 Mar 1;44(3).
- [14] Sasirekha S, Swamynathan S. A comparative study and analysis of data aggregation techniques in WSN. *Indian Journal of Science and Technology*. 2015 Oct;8(26):1-0.
- [15] R. E. Mohamed et al., "Energy-efficient routing protocols for solving energy hole problem in wireless sensor networks," *Comput. Netw.*, vol. 114, p. 51-66, 2017, <https://doi.org/10.1016/j.comnet.2016.12.011>.
- [16] K. Thangaramya et al., "Energy aware cluster and neuro-fuzzy based routing algorithm for wireless sensor networks in IoT," *Comput. Netw.*, vol. 151, pp. 211-223, 2019, <https://doi.org/10.1016/j.comnet.2019.01.024>.
- [17] A. Yazici et al., "A fusion-based framework for wireless multimedia sensor networks in surveillance applications," *IEEE Access*, vol. 7, pp. 88418-88434, 2019, <https://doi.org/10.1109/ACCESS.2019.2926206>.
- [18] W. Guo et al., "Optimizing the lifetime of wireless sensor networks via reinforcement-learning-based routing," *Int. J. Distrib. Sens. Netw.*, vol. 15, no. 2, 2019, <https://doi.org/10.1177/1550147719833541>.
- [19] R. A. Khan et al., "RK-energy efficient routing protocol for wireless body area sensor networks," *Wirel. Pers. Commun.*, vol. 116, p. 1-22, 13, Aug. 2020.
- [20] I. Daanoun et al., "An enhanced energy-efficient routing protocol for wireless sensor network," *Int. J. Electr. Comput. Eng.* (2088-8708), vol. 10, no. 5, [p. 12],
- [21] S. Umbreen et al., "An energy-efficient mobility-based cluster head selection for lifetime enhancement of wireless sensor networks," *IEEE Access*, vol. 8, pp. 207779-207793, 2020, <https://doi.org/10.1109/ACCESS.2020.3038031>.
- [22] K. Haseeb et al., "LSDAR: A light-weight structure based data aggregation routing protocol with secure internet of things integrated next-generation sensor networks," *Sustain. Cities Soc.*, vol. 54, p. 101995, 2020, <https://doi.org/10.1016/j.scs.2019.101995>.
- [23] G. Sanjay Gandhi et al., "Grid clustering and fuzzy reinforcement-learning based energy-efficient data aggregation scheme for distributed WSN," *IET Commun.*, vol. 14, no. 16, pp. 2840-2848, 2020, <https://doi.org/10.1049/iet-com.2019.1005>.
- [24] I. Ullah and H. Y. Youn, "Efficient data aggregation with node clustering and extreme learning machine for WSN," *J. Supercomput.*, vol. 76, no. 12, pp. 10009-10035, 2020, <https://doi.org/10.1007/s11227-020-03236-8>.
- [25] W. K. Yun and S. J. Yoo, "Q-learning-based data-aggregation-aware energy-efficient routing protocol for wireless sensor networks," *IEEE Access*, vol. 9, pp. 10737-10750, 2021, <https://doi.org/10.1109/ACCESS.2021.3051360>.
- [26] S. Bhushan, M. Kumar, P. Kumar, T. Stephan, A. Shankar, & P. Liu, "FAJIT: A fuzzy-based data aggregation technique for energy efficiency in wireless sensor network," *Complex Intell. Syst.*, vol. 7, no. 2, pp. 997-1007, 2021, <https://doi.org/10.1007/s40747-020-00258-w>.
- [27] K. Jain and A. Singh, "A two-vector data-prediction model for energy-efficient data-aggregation in wireless sensor network," *Concurrency Comput. Pract. Experience*, vol. 34, no. 11, p. e6898, 2022, <https://doi.org/10.1002/cpe.6898>.
- [28] P. William et al., "Analysis of data aggregation and clustering protocol in wireless sensor networks using machine learning" in *Evolutionary Computing and Mobile Sustainable Networks: Proceedings of ICECMSN*. Singapore: Springer Singapore, 2022, pp. 925-939, https://doi.org/10.1007/978-981-16-9605-3_65.
- [29] V. P. Pandiya Raj and M. Durairandian, "Energy conservation using PISAE and cross-layer-based opportunistic routing protocol (CORP) for wireless sensor network," *Eng. Sci. Technol. An Int. J.*, vol. 42, p. 101411, 2023, <https://doi.org/10.1016/j.jestch.2023.101411>.
- [30] A. K. Idrees, et al., "Yaseen, W.L." in *15th International wireless communications & mobile computing conference (IWCMC)*, vol. 2019. IEEE, 2019, Jun
- [31] M. V. Babu et al., "An improved IDAF-FIT clustering based ASLPP-RR routing with secure data aggregation in wireless sensor network," *Mob. Netw. Appl.*, vol. 26, no. 3, pp. 1059-1067, 2021, <https://doi.org/10.1007/s11036-020-01664-7>
- [32] N. T. Nguyen et al., "On maximizing the lifetime for data aggregation in wireless sensor networks using virtual data aggregation trees," *Comput. Netw.*, vol. 105, pp. 99-110, 2016, <https://doi.org/10.1016/j.comnet.2016.05.022>.
- [33] V. T. Pham et al., "Minimizing latency for data aggregation in wireless sensor networks: An algorithm approach," *ACM Trans. Sens. Netw.*, vol. 18, no. 3, pp. 1-21, 2022, <https://doi.org/10.1145/3450350>.
- [34] S. Gupta et al., "Optimizing Wireless Sensor Networks through Ant Colony-Based Localized Mesh Topology," *Journal of The Electrochemical Society*, vol. 171, no. 5, p. 057503, 2024
- [35] S. Gupta, "Reliable gamma-interconnection network for data analysis in sensor networks: design and performance evaluation," *ECS Sensors Plus*, vol. 2, no. 3, p. 034801, 2023.
- [36] N. Devi et al., "IoT Optimization for Smart Cities and Mobility in Smart Cities". In *Applications of Optimization and Machine Learning in Image Processing and IoT*. (pp. 54-66). Chapman and Hall/CRC, 2023