OPTIMIZING RELIABILITY AND EFFICIENCY IN DATA AGGREGATION FOR LOW-POWER WSNS

Vijaysinh Uttamrao Bansude
Research Scholar
Dept Of ECE
SJJTU Rajasthan

Dr. Rahulkumar Budania
Supervisor
Dept Of ECE
SJJTU Rajasthan

Dr. S. T. Shirkande
Co-Supervisor
Dept Of ECE
SJJTU Rajasthan

ABSTRACT

In Wireless Sensor Networks (WSNs), data redundancy is a challenging issue that not only introduces network congestion but also consumes a considerable amount of sensor node resources. Data redundancy occurs due to the spatial and temporal correlation among the data gathered by the neighbouring nodes. Data aggregation is a prominent technique that performs in-network filtering of the redundant data and accelerates the knowledge extraction by eliminating the correlated data. However, most of the data aggregation techniques have lower accuracy as they do not cater for erroneous data from faulty nodes and pose an open research challenge. To address this challenge, we have proposed a novel, lightweight, and energy-efficient function-based data aggregation approach for a cluster-based hierarchical WSN. Our proposed approach works at two levels, i.e., at the node level and at the cluster head level. At the node level, the data aggregation is performed using Exponential Moving Average (EMA) and a threshold-based mechanism is adopted to detect any outliers for improving the accuracy of aggregated data. At the cluster head level, we have employed a modified version of Euclidean distance function to provide highly-refined aggregated data to the base station.

Keywords: *Wireless Sensor Networks (WSNs), Data redundancy, Exponential Moving Average (EMA), Data aggregation, cluster head level.*

INTRODUCTION

The Internet of Things carefully connects people/share data and gets them involved doming the increasing digital ecosystem. At the same time, wireless sensor networks (WSNs) have become a separate field of study. They do this by using sensor nodes that work together to collect and analyse a wide range of data from their surroundings. In more interconnected world, the effective handling of data becomes of utmost significance. A key element of both IoT and WSNs is data aggregation. Coordinating the gathering, processing, and transfer of data requires careful planning. This organisation plays a variety of roles with goals that extend beyond basic data handling. Data aggregation's two main goals are to lessen the heavy load that data transmission causes and to effectively save the finite energy resource in networks with limited resources.

The successful functioning of IoT applications and WSNs relies on the strategic and efficient execution of data aggregation. It acts as the cornerstone for creating these networks, guaranteeing that the information gathered is not only precise and trustworthy but also handled with care and caution. Furthermore, data aggregation is essential to improving these networks' effectiveness and enabling seamless, effective operation. Data aggregation is becoming more and more important as IoT applications become more common and diverse, and as WSN becomes more significant in domains like industrial automation, healthcare, and environmental monitoring.

LITERATURE REVIEW

JEYA RANI D (2024) Many applications need (WSNs), which maximise energy consumption and prolong network lifetime via efficient node sensing and data aggregation procedures. Node sensing is the process of gathering information from sensors within each node about various environmental factors, such as the location, behaviour, and history of the node. We presented the ENSA method for node sensing and aggregation in this work. In order to reduce data traffic, save energy, data aggregation gathers and compiles information from several nodes before sending it, which helps with in-network processing. In order to develop a novel technique known as Energy-Efficient Node Sensing and Aggregation (ENSA), this study examines two primary approaches: compressed aggregation with correlation (CACC) and hybrid energy-efficient distributed clustering (HEED). To establish long-spanning networks and balanced energy consumption, HEED is a clustering-based method that elects cluster heads by using node proximity and residual energy. CACC reduces transmission cost and ensures data integrity by compressing aggregated data via data correlation between neighboring nodes.

Arash Heidari (2024) A new field of information technology called the (IoT) enables anything to transmit and receive data across a network. Cloud computing, big data, (IWNs), and the Industrial Internet of Things (IIoT) have all made substantial progress in the last many years. A spanning tree or other dependable and efficient data collecting method is necessary while using IIoT. Failure and mobility are ignored by many of the earlier spanning tree techniques in certain cases, data transmission to the base station may be challenging due to damage to the spanning tree. In order to create appropriate trees, this paper suggests a method that combines genetic operators, density correlation degree, and an artificial bee colony to create an ideal spanning tree. Using the mobility probability, residual energy, and hop count distances from the base station, this technique evaluates how well the trees fit the devices. The simulation findings show that the suggested approach outperforms popular techniques such as the Reliable Spanning Tree (RST) construction algorithm and the Hop Count Distance (HCD) based building algorithm in IIoT in terms of gathering trustworthy data.

Beneyaz Ara Begum (2023) Data aggregation refers to the process of combining relevant information from multiple wireless devices and sensors within WSN and IoT networks to interpret the collected data effectively. It also serves as a key strategy for enhancing node energy efficiency, conserving communication bandwidth, and reducing processing time. This article discusses WSN and IoT data aggregation concepts and the components that aggregation techniques use to increase performance. The research evaluates data aggregation methods for WSN and IoT network architecture, interference, fault tolerance, mobility, and security. Due to resource limits, these networks must balance latency, energy efficiency, data quality, freshness, and temporal accuracy. The essay goes into great detail on the requirements and the trade-off strategies that data aggregation methods should use in order to maximize these demands. The study highlights some of the drawbacks of the data aggregation methods currently in use and provides research suggestions to assist bridge these gaps cooperatively.

P. Balamurugan (2019) A collection of tiny, separate devices known as sensor nodes that cooperate to address at least one shared issue is known as a sensor network. Usually, sensor nodes gather, compile, and transmit data to the (BS) in (WSN). One of the biggest problems with WSN is that sensor network nodes don't have a lot of energy. The energy starts to go down while the nodes are collecting data. Because of this, collecting data needs to be dependable, adaptable, and quick. The research tackles energy problems by introducing the Mark Based Data Gathering (MBDG) algorithm, which is an energy-efficient way to gather data that attempts to lower latency and energy utilisation while collecting and sending the fused data to the BS in WSN. A comparison is made between the proposed algorithm and the existing approaches LEACH, PEGASIS, EMLN-DG, and GBE-DG. The result demonstrates that, in comparison to current methods, the suggested approach significantly increases network lifespan while lowering latency and energy usage.

Jacob John (2018) Multiple resource-constrained sensor nodes operating independently for months or even years make up a wireless sensor network. One of the biggest challenges with WSN is energy saving because of its restricted power source. Nodes were static throughout the early phases of WSN development, and multi-hop paths were used for all communication between any two nodes. There is a wealth of literature and research on several static WSN energy conservation techniques. Many connection issues may be resolved by including mobility into WSN. Depending on whether the mobile element is a mobile relay or a mobile sink, we suggest mobility-based energy-saving techniques in this study. We also suggest energy-efficient data collection techniques, such as cross-layer approaches, routing strategies, and clustering schemes in Mobile Wireless Sensor Networks (MWSNs). The different energy-efficient techniques are also compared.

Data Aggregation Meaning

The act of creating a cohesive dataset by combining information from several databases, files, or sources is known as data aggregation. The major objective is to make the dataset easier to comprehend and work with. Aggregated data, which is often summarised, allows for a higher-level perspective and makes managing large datasets easier. In sectors where data analysis and interpretation are necessary, this procedure is essential.

Types of Data Aggregation

Two major categories may be used to classify data aggregation:

Time-Based Aggregation: This technique entails gathering information over a predetermined amount of time. For instance, monthly or annual sales reports can be created by combining daily sales data. This aids in spotting patterns and trends throughout time.

Data that is aggregated according to a geographic or physical location is known as spatial aggregation. For example, regional performance may be examined by combining sales data from several retail locations. Environmental studies, urban planning, and logistics all benefit greatly from this.

Benefits of Data Aggregation

Improved Decision Making

Organisations are able to make more educated decisions because to the full perspective of information that aggregated data offers. Decision-makers can rapidly comprehend important measurements and trends when they have access to summarised data.

Enhanced Data Quality

To improve the general state of the data, data aggregation often uses data cleaning techniques. Clean, trustworthy data is essential for accurate reporting and analysis.

Increased Efficiency

Time is saved and the possibility of human error is decreased when the data gathering process is automated. Organisations may prioritise data analysis above data collecting due to its effectiveness.

Better Insights

Aggregated data shows patterns and trends that might not be seen in raw data. Both operational enhancements and strategic planning benefit from this greater understanding.

Applications of Data Aggregation

Business Intelligence

One way to use business intelligence is to gather information from different departments. such as marketing, sales, operations, and finance, is data aggregation. Following consolidation, this data is examined to provide insights that enhance output and direct corporate objectives.

Healthcare

In healthcare, data aggregation is bringing together patient data from several places to improve treatment, speed up processes, and help with further research. Trends in illness patterns, treatment effectiveness, and patient outcomes can all be seen in aggregated health data.

Marketing

Data aggregation is a technique used by marketers to gather information from several sources, including website analytics, email marketing, and social media.

Understanding consumer behaviour, gauging the success of campaigns, and tailoring marketing the use of this aggregated data facilitates all endeavors.

Finance

In the financial business, data aggregation is important for managing risk, finding fraud, and following the rules. Banks and other financial organisations collect transaction data to check for anomalous activity, evaluate credit risks, and make sure they are following the law.

Environmental Studies

To track and examine pollution levels, climatic trends, and the use of natural resources, researchers compile data from a variety of environmental sensors. Policy-making and environmental protection depend heavily on this data.

The act of gathering and condensing raw data for analysis is known as data aggregation. Almost every employee participates in data aggregation at some time, even though the phrase is usually connected to technical teams.

Most likely, you have used aggregated data yourself, such as monthly organic visitors, average cost per click, and annual income. Regardless of your function, data aggregation makes data more useful, efficient, and enlightening. Additionally, you probably encounter data aggregation on a daily basis if you don't do it yourself. For instance, the majority of software and apps available today, including business apps, Netflix, an OTT platform, and Spotify, a digital music service.

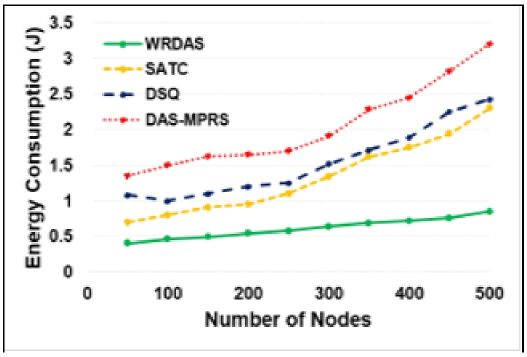
RESEARCH METHODOLOGY

This component constructs an aggregation tree for a given wireless sensor network, shown as an undirected graph where the root is the sink. By taking into account variables like each node's connection strength or remaining energy, it creates routes from every source cluster to the sink.

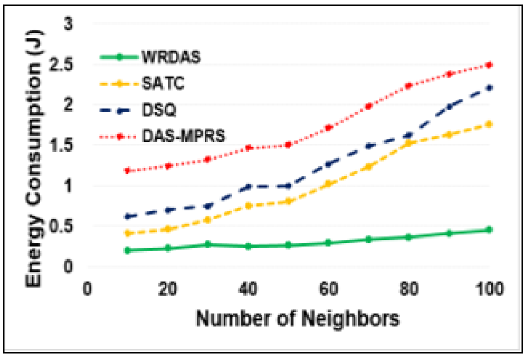
The nodes that are eligible are identified using this neighbour table. Stated differently, nodes that are smaller than the current node's distance from the sink are selected if they have sufficient energy and a strong connection. From the list of eligible nodes, the node with the greatest Q-value is selected as the next hop. The path of the source node is updated with the next hop node after being randomly chosen with a probability value if any of the nodes have the identical Q-values. If there are no appropriate nodes, the next hop is a node with the highest Q-value from the neighbour table. Since this node has no neighbors, all of its neighbors get the message "No-neighbor-to-sink," which prevents data from being sent to this node. For the selected next hop, a reward value is calculated. If the next hop is the sink, then a constant reward is assigned. Otherwise, if the selected next hop has low energy or if there is no next hop, then a negative reward is assigned. If not, a weighted mix of the nodes' leftover energy, distance from the sink, and connection strength is used to determine the reward.

RESULTS AND DISCUSSIONS

Graph 1 displays the results of this metric, which calculates the overall energy used to schedule the aggregate tree's nodes without collision. Compared to other algorithms, WRDAS uses a very low amount of energy in scenario 1. The energy consumption of SATC, DSQ, and DAS-MPRS is just 53%, 46%, and 34%, respectively, compared to WRDAS. The primary cause is because, in contrast to other algorithms, WRDAS has a very small number of state changes. Furthermore, it schedules the nodes using minimum number of time slots which consequently increases the aggregation ratio. Significant energy savings are achieved as a consequence of the reduction in the quantity of data packets sent. Because SATC, DSQ, and DAS-MPRS require more time slots for scheduling than WRDAS, their energy usage rises approximately linearly as the number of nodes grows. Because DAS-MPRS builds a data aggregation tree with many discontinuous pathways, it uses the most energy of all the techniques.



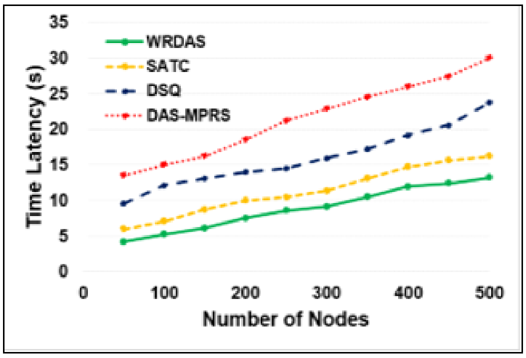
(a). Scenario 1



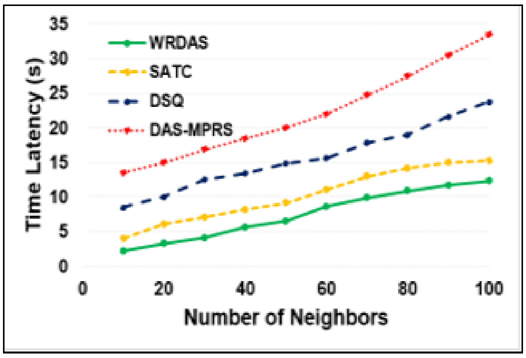
(b). Scenario 2

Graph 1: Energy Consumption

The time needed to integrate data from the leafy nodes to the sink is known as the delay, and it is expressed in terms total the total number of time slots. The delay findings in relation to the number of neighbours and nodes are shown in Graph 2. Graph 2 (a) illustrates that WRDAS consistently performs better than SATC, DSQ, and DAS-MPRS, despite the fact that latency rises with the number of nodes. This superior performance is attributed to WRDAS's consideration of three key parameters: residual energy, hop distance, and number of neighbors. Since SATC gives priority to time latency, its performance is close to WRDAS, whereas the high latency achieved by DAS-MPRS is due to the randomness in aggregator selection. As the number of nodes rises, DAS-MPRS's latency grows significantly. In other words, its temporal delay is four times greater than WRDAS on average. Though DSQ considers hop distance in allocation of time slots, its time latency is slightly higher than SATC.



(a). Scenario 1



(b). Scenario 2

Graph 2: Time Latency

In case 2, as seen in Graph 2 (b), WRDAS exhibits noticeably better performance as the number of neighbours increases. This is because the higher node density enhances the concurrency of scheduled nodes, allowing WRDAS to utilize available time slots more efficiently. This resulted in a smaller number of time slots allocated which reduces its time latency. Of all the schemes, DAS-MPRS performs very poor, showing a steep rise when the number of neighbors increases, as the multiple disjoint paths needs more time slots.

CONCLUSIONS

Numerous tiny, widely dispersed, inexpensive sensor nodes make up wireless sensor networks (WSNs). These nodes collect crucial data on occurrences that are associated in space and time. While there are certain benefits to the dense deployment of these nodes, there are also drawbacks, such as data redundancy. Repetitive data transmission and sensing causes bottlenecks in networks and shortens their lifespan. Sensor nodes may be positioned in inaccessible locations and contain limited, irreplaceable batteries in addition to data redundancy. Resource-constrained nodes transmit the raw data to the base station after it has been collected. However, because of the high cost of transmission, sending such raw data uses a lot of energy and might cause delays, packet loss, or network congestion. Consequently, the energy resources of the network are used more quickly than anticipated. Reducing data traffic inside the network is essential for energy conservation. Developing energy-efficient data aggregation techniques that employ lossy aggregation methods to preserve data reliability and integrity while increasing the aggregation ratios and reducing the number of packets by eliminating unneeded information is the main objective of WSN research. Recently, a variety of data aggregation strategies have been put forward. For instance, the authors of introduced a cluster-based aggregation method that uses similar functions, data value weights, and a distance-based metric at the cluster head.

REFERENCES

1. *JEYA RANI D (2024), "enhancing node sensing and aggregation efficiency in wireless sensor networks (ensa) ", Journal of*

Theoretical and Applied Information Technology, ISSN:1817-3195, Vol.102, Issue.19

2. *Arash Heidari (2024), "A reliable method for data aggregation on the industrial internet of things using a hybrid optimization algorithm and density correlation degree", Cluster Computing, ISSN: 1573-7543, Volume 27, pages.7521–7539*
3. *Beneyaz Ara Begum (2023), "ata aggregation protocols for WSN and IoT applications – A comprehensive survey", Journal of King Saud University - Computer and Information Sciences, ISSN: 2213-1248, Volume.35, Issue.2*
4. *P. Balamurugan (2019), "Reliable and Energy Efficient Data Gathering Protocol in Wireless Sensor Networks", International Journal of Recent Technology and Engineering, ISSN: 2277-3878, Volume-7, Issue-5S2*
5. *Jacob John (2018), "Energy Efficiency Enhancement Methods for Mobile Wireless Sensor Networks: A Survey", IOSR Journal of Computer Engineering (IOSR-JCE), ISSN: 2278-0661, Volume. 20, Issue.2, Pages.59-66*
6. *Hemant Sethi (2011), "EIRDA: An Energy Efficient Interest based Reliable Data Aggregation Protocol for Wireless Sensor Networks", International Journal of Computer Applications, ISSN 0975-8887, Volume. 22, Issue.7*
7. *Hevin Rajesh Dhasian (2013), "Survey of data aggregation techniques using soft computing in wireless sensor networks", Information Security", IET Information Security, ISSN: 1751-8717, vol.7, issue. 4,*
8. *Hiren Thakkar (2012), "A Power Efficient Cluster-based Data Aggregation Protocol for WSN (MHML)", International Journal of Engineering and Innovative Technology, ISSN: 2277-3754, Volume.1, Issue. 4,*
9. *Hongxing Li (2014), "Latency-minimizing data aggregation in wireless sensor networks under physical interference model" Ad Hoc Networks, ISSN :1570-8713, Volume.12, Pages.52-68.*
10. *Huseyin Ozgur Tan (2011), "Computing Localized Power-Efficient Data Aggregation Trees for Sensor Networks", IEEE Transactions on Parallel and Distributed Systems, ISSN: 1558-2183, vol.22, issue. (3), pages.489-500. DOI: 10.1109/TPDS.2010.68.*