EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

Data Aggregation through Hybrid Optimal Probability in Wireless Sensor Networks

S. Balaji[1], S. Jeevanandham[2], Mani Deepak Choudhry[3*], M. Sundarrajan[4], Rajesh Kumar Dhanaraj[5]

[1]Department of CSE, Akshaya College of Engineering and Technology, Coimbatore
[2]Department of IT, Sri Ramakrishna Engineering College, Coimbatore
[3]Department of IT, KGiSL Institute of Technology, Coimbatore
[4]Department of Networking and Communications, SRM Institute of Science and Technology, Kattankalathur, Chennai
[5]Symbiosis Institute of Computer Studies and Research (SICSR), Symbiosis International (Deemed University), Pune

Abstract

INTRODUCTION: In the realm of Wireless Sensor Networks (WSN), effective data dissemination is vital for applications like traffic alerts, necessitating innovative solutions to tackle challenges such as broadcast storms.
OBJECTIVES: This paper proposes a pioneering framework that leverages probabilistic data aggregation to optimize communication efficiency and minimize redundancy.
METHODS: The proposed adaptable system extracts valuable insights from the knowledge base, enabling dynamic route adjustments based on application-specific criteria. Through simulations addressing bandwidth limitations and local broadcast issues, we establish a robust WSN-based traffic information system.
RESULTS: By employing primal-dual decomposition, the proposed approach identifies optimal packet aggregation probabilities and durations, resulting in reduced energy consumption while meeting latency requirements.
CONCLUSION: The efficacy of proposed method is demonstrated across various traffic and topology scenarios, affirming that probabilistic data aggregation effectively mitigates the local broadcast problem, ultimately leading to decreased bandwidth demands.

Keywords: WSN, Data Collection, Energy Efficient, Probabilistic, LEACH, Secure Protocol

Received on 08 November 2023, accepted on 21 January 2024, published on 01 February 2024

doi: 10.4108/eetsis.4996

*Corresponding author. Email: manideepak.c@kgkite.ac.in

1. Introduction

The Internet of Things (IoT) is predicated on the hypothesis that a worldwide network of intercommunicating devices from smart watches and other wearables to sensors and mobile phones could benefit society, stimulate the economy, and raise living standards for those involved. By the beginning of the next decade, it is expected that the Internet of Things will have connected tens of billions of "things" to the Internet, making possible the integration of hitherto unconnected services [1]. The objective is to construct "smart" cities on top of this framework; However, there are several definitions of a "smart" city. Everyone can agree that this technology would improve people's quality of life there. The city's underlying networked infrastructure may greatly improve a wide range of municipal operations and promote elements of social, cultural, and urban growth via the employment of smart devices, computers, sensors, and actuators. While there has been a rise in interest in the concept and implementation of what we now call "smart cities," the academic literature has paid less attention to the issue, initiatives like Chicago's Array of Things (AoT) project, which is made up of a network of modular [2] and interactive sensor boxes that collect real-time data on the city's environment, infrastructure, and activity for the benefit of research and public use [3]. For this vision of the "smart city" to become a reality, sensors and sensor networks are essential because they can gather and transmit data that may contribute to improving public and commercial services [4].

As data packets travel across a network, they may be aggregated at any or all of the nodes along the way. This method reduces the number of data packets sent from sensors to sinks by making use of the interconnections between measurements collected at neighboring nodes [5]. As a result, there is less traffic in the network as a whole and less wait time in the queues of individual nodes. However, data packets are encounters while travelling the network is reduced.

However, most studies still assume that the aggregation is done in real-time, despite the fact that a tradeoff has been identified. We contend that this delay is not inconsequential and must be dealt with in a variety of contexts, especially when significant energy savings are realized, we previously established a general probabilistic paradigm [6] in which each node conducts data aggregation locally and probabilistically. Aggregated packets based on a pre-assigned probability may be sent from each node once a certain amount of time has passed. Numerical assessments in many scenarios have shown the significant performance and operational efficiency gains that can be achieved using this method. Nodes are established at the outset and do not vary while the system operates [7]. In light of these factors, we developed a comprehensive optimization framework to determine the probability at which a node should collect packets and the duration it should wait to reduce power consumption while still achieving stated delay requirements.

2. Related Works

An approach called data aggregation has been developed to extend. As a result, many relevant studies have been discovered in the literature, presenting schemes and algorithms for data aggregation with a wide range of aims and purposes. An efficient cluster creation procedure, named LEACH [8]. Data is aggregated and routed to the need for transmission and the amount of energy lost in the process, with data stored locally on each node [9].

To reduce data volume, the suggested architecture [10] elects a leader node to perform the in-network aggregation. However, nodes may also perform partial aggregation. Tiny Aggregating (TAG) [11] is a generic aggregation service for ad hoc networks that uses a similar approach to the one supplied. In this configuration, inquiries from an external user are sent to the sink, and all other network nodes work. Along its path from its source to its destination, data is aggregated at intermediate nodes using a prescribed aggregation function. To account for QoS indicators, in-network processing is performed using certain aggregation approaches [12]. It has been suggested to adapt AIDA (application independent data aggregation).

An additional layer is used between the DLCP and the NETWORK layer [13] – [16] to carry out the aggregation task and to decide whether or not to aggregate data. Results reveal that this technique works well in terms of energy savings and achieving specific QoS parameters; hence, some of the core ideas of this approach are accepted and enhanced inside our framework. Recent academic work has dealt with the aggregation process using optimization algorithms. However, the vast majority of approaches already account for aggregation as data moves across the network and the routing strategy utilized by each approach. In [17], the authors look at the difficulty of extending the time before the first sensor node in a network breaks. They use a technique that simplifies the task into one of the quotas and limited resources to obtain information. They utilize this knowledge to propose an algorithm that solves the problem of how many packets of data need to be sent between two nearby nodes at each round in polynomial time.

High sensing accuracy and low energy consumption are accomplished via the development of a novel method of dynamic energy management that employs an estimate of the target's trajectory based on Radial Basis Function (RBF) [18] – [23]. Sensors' data transit is framed as an optimization problem with maximal information collection at sinks considering. This approach has little networked overhead while giving substantial benefits. The question of the best time to collect data is discussed and answered. In this paper, the authors provide a distributed method for making a conflict-free plan for data collecting utilizing wireless sensor networks based on the concept of maximum independent sets.

We argue that most of these methods ignore an important aspect for delay-sensitive sensor network applications: the time required to perform data aggregation in a tree structure, as seen in **Figure 1**. Few studies, including the ones above, have tried to put a numerical value on the significance of aggregation time. By transforming the collection tree into its binary analogue, we can solve the problem of pinpointing the aggregation sites using a dynamic programming technique in polynomial time. Although not optimum, the authors use a localized approach that is more practical than utilizing worldwide data to determine aggregation delays. Initially, a brute-force approach was proposed to solve this problem; however, its complexity meant that it was unsuitable for usage by typical sensor nodes, which often had limited computing capabilities. Because of this, a heuristic approach was developed in which a single node is selected to aggregate data from, thereby using all available resources, this method ensures that all nodes die at about the same rate.

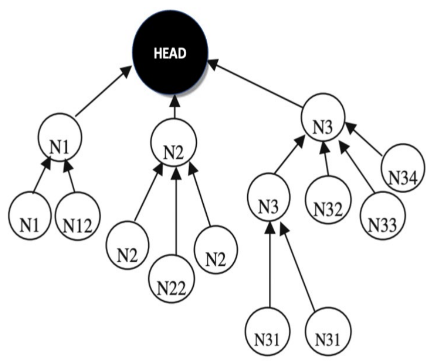

Figure 1. Tree Constructed Aggregation

3. Proposed Methodology

While our work is similar in spirit to that described in related works, it differs significantly from the previous literature in that it can determine ideal values for the aggregation duration and the accompanying aggregation probability to minimize energy consumption. All relevant works mentioned above aggregate at a single node or aggregate for a constant latency. This study provides a novel optimization approach for determining which nodes along a message's routing path should execute aggregation and how much time should be spent at each node performing aggregation. Not only the optimization problem is tackled decentralized, but the may-dual decomposition method is also used. With the end objective of network-wide energy minimization using just local knowledge, each node periodically. As a result of the distributed and dynamic calculation of the aforementioned ideal values at different time intervals, probabilistic data aggregation is assured to be adaptable to network traffic circumstances. We can see the Algorithm for understanding the flow in detail about the proposed model and system flow in **Figure 2**.

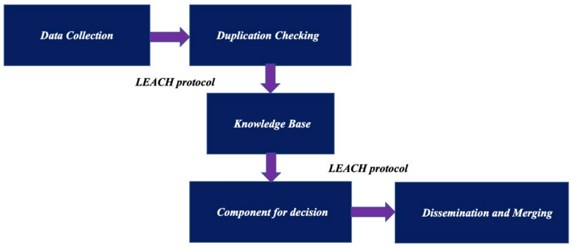

Figure 2. Proposed Flow of Research

<u>*Probabilistic Data Aggregation Algorithm:*</u>
Input: Tree structured Data from the nodes
Set: Dataset Testing area
Aggregation to the given data
If aggregation is OK then
Analysis starts for making the nodes for better communication
else
Test again for better routing values
elif Exact values = settled Value
end if
Performing aggregation for all Data inside the communication
Output: Data Aggregation with good Accuracy Expected

3.1. LEACH protocol for communication

It is a hierarchical protocol where most nodes send their data to the cluster heads, which then compress it and send it to the base station. At the beginning of each round, nodes utilize a stochastic approach to determine if they will take on the role of cluster head for the current round. The idea is that each node has a radio strong enough to direct contact with the home base or the closest cluster head, but doing so at power all the time would be inefficient. The leader of a cluster establishes a timetable for data transmission among the nodes in the cluster.

3.2. Data Collection

Every moving vehicle constantly sends out frames of data about itself and the other cars close by, including speed, X position, Y coordinate, and acceleration. Upon receiving a frame broadcast by another vehicle, each vehicle parses that frame for information and adds it to its own internal database.

3.3. Duplication Checking

As soon as data is entered, take the car as an example, knowledge of the cars immediately surrounds it. Each vehicle then compares its data to others of its kind to see if they are from the same general area. The knowledge base is updated with the new information, and the relevant data is passed to the decision module.

3.4. Knowledge Base

To aggregate data, an aggregation system must determine whether or not numerous data points are sufficiently comparable. This is an easy call for sensor networks; if the data items' sources are in the same cluster or sub-tree of an aggregation tree, they may be combined. Information systems that aggregate data in a segment-oriented fashion determine whether or not two pieces of data share a common spatial or temporal segment. Using hierarchical information aggregation systems, VANETs can make decisions based on the specifics of the data at hand, with fusible data being marked and stored separately.

3.5. Component for Decision

The judgment made on the fusion of data is recorded in the knowledge base, as is the relationship between the various data elements. Making this call aids in determining whether or not to proceed with data item aggregation. Graphs and KD trees are two examples of data structures that could be used to organize a knowledge base.

3.6. Component for Decision

Data from several cars is aggregated, encoded, or compressed to fit into a single frame using the syntactic aggregation technique. Costs are reduced compared to transmitting individual messages or frames. To begin the process of semantic aggregation, data from individual automobiles is first summarized. For instance, merely the fact that there are only five vehicles is transmitted rather than their particular location. In exchange for some precision, this will result in a shorter message. This study will focus on the meaning behind the aggregation because once it has been agreed to combine two data sets, a standard procedure must be put in place to do so.

An example is a process used in hierarchical aggregation systems to combine two aggregates. The need for complex logic in data fusion prevents any simple approaches from being universal. Average, sum and minimum are all instances of straightforward data aggregation operations.

Once fusing able data is identified and the corresponding data components are aggregated, new data becomes available, which is currently only known to the vehicle that did the aggregation. The vehicle must therefore propagate the newly formed data throughout the network. Flooding, periodic beaconing, geo-broadcasting, and incrementally working one's way up a tree are all viable options for dissemination. The data processed by the fusion module is also screened to determine what should be shared with the rest of the world. We can also see the accuracy of routing in Table 1, 2, 3 and in **Figure 3**. Further **Figure 3** depicts the various comparative analysis of proposed systems 3(a) Routing Accuracy, 3(b) Network Lifetime and 3(c) Processing Time.

Table 1. Routing Accuracy

Model	Results	
	Existing	Proposed
Accuracy	94%	98%
Estimated Speed	6700obs/sec	8500obs/sec
Instruction Time	7.2s	2.9s

Table 2. Network Lifetime

Algorithms	Results	
	Rounds	Time ($* 10^4 s$)
ECRP	3	4.2
HMBCR		4.8
Proposed		6.9
ECRP	7	1.5
HMBCR		2.1
Proposed		1.9
ECRP	11	1.2
HMBCR		1.4
Proposed		1.8

Table 3. Processing Lifetime

Algorithms	Results
	Processing Time
ECRP	0.232
HMBCR	0.211
Proposed	0.298

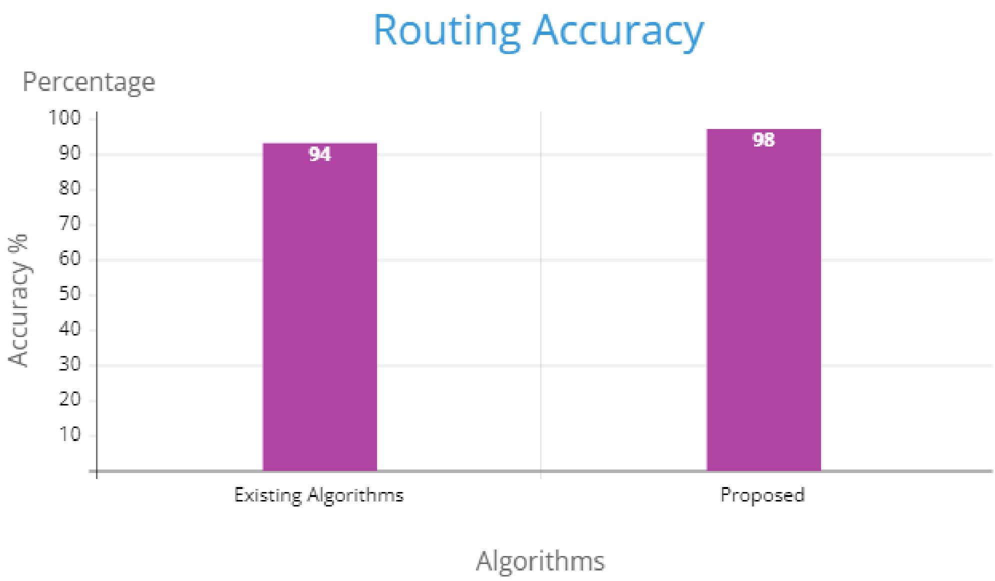

Figure 3 (a). Accuracy %

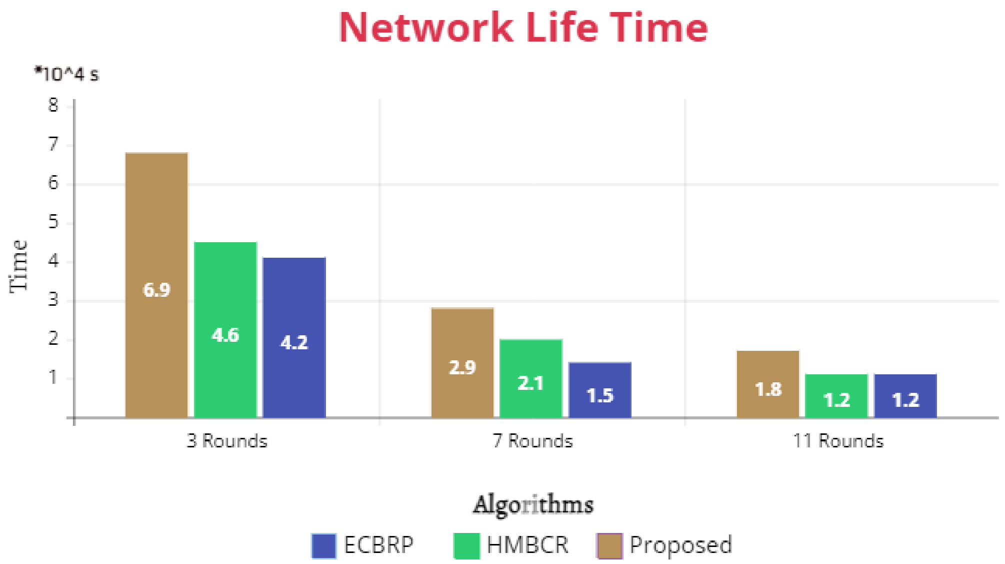

Figure 3 (b). Network Life Time

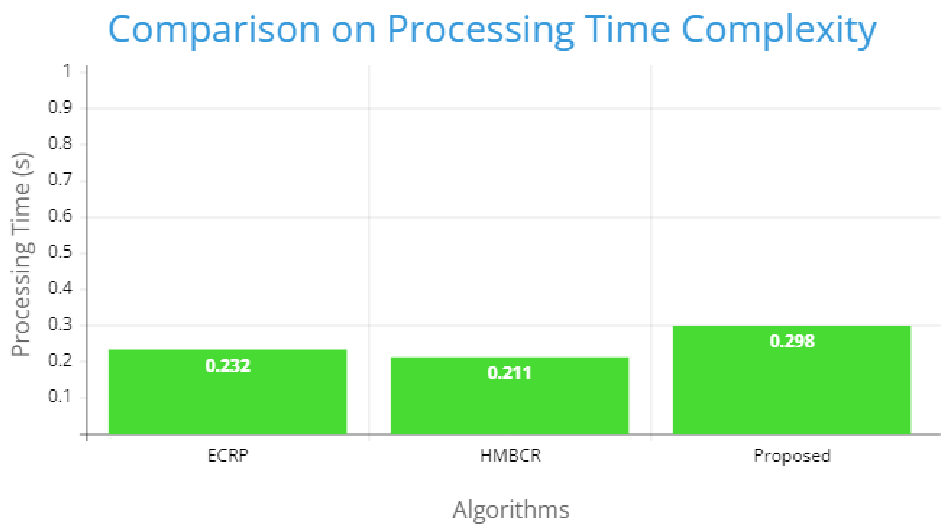

Figure 3 (c). Processing Time Complexity of Algorithms

4. Result Evaluation

Two methods with a similar goal to reduce overall energy consumption. This was done so that our suggested framework could be better situated within the current literature by serving as a benchmark and providing a full comparative review. The Algorithm enforces a policy in which the total delay encountered along a route is proportionally distributed among its nodes. **Figures 4 and 5** depict the second approach based on a straightforward aggregation methodology that takes advantage of data-centric routing. With this method, the allowable aggregation delay is spread out evenly over the entire path of nodes.

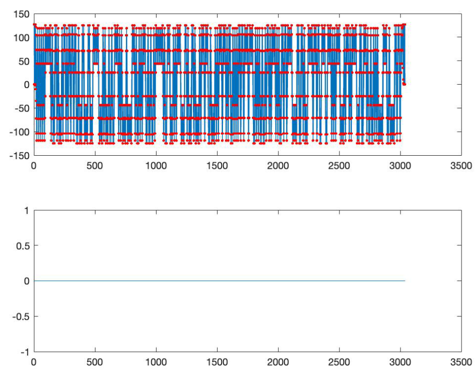

Figure 4. Difference in Data Aggregation

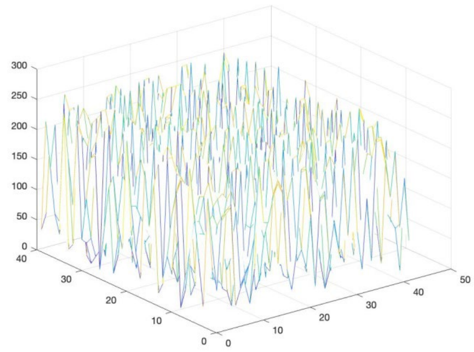

Figure 5. Total Number of Sample Data Aggregation

Specifically, our method demonstrates significantly better performance than the two competing techniques in terms of aggregate gain. Since the Aggregation Gain increases in proportion to the volume of network traffic, the benefits of aggregation have been shown, most notably under heavy traffic loads, when the quantity of messages in the network decreases significantly. **Figure 6** further demonstrates that the recommended Algorithm significantly outperforms the other examined methods under all simulated conditions, and that it does not evenly distribute the permissible aggregation delay over all nodes. As can be seen in **Figure 7**, the suggested optimization framework instead selects and minimizes the total amount of transmissions.

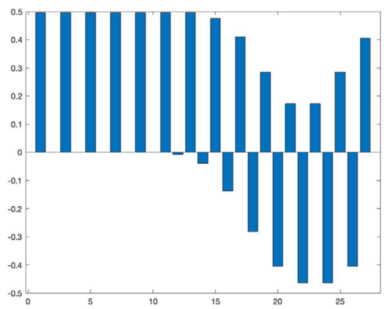

Figure 6. Effective Aggregation and Number of Transmissions

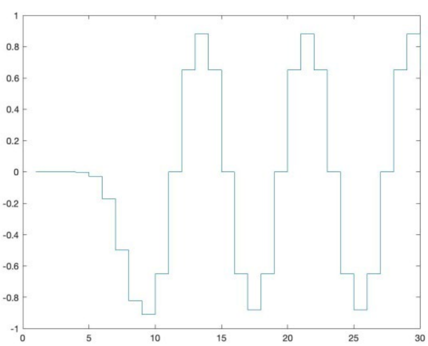

Figure 7. Data Transmissions in Variations

Finally, the suggested framework's adaptability to future traffic volumes and application needs is assessed. This is accomplished by considering the case below. For an air quality application, data are collected via a sensor network of fifty randomly placed environmental sensor nodes in a 70 m and 70 m region, as shown in **figure 8**. With a time-delay limitation of 20 seconds, sensor nodes in this scenario produce packets at a rate of 0.7 per second, including data on the ambient air quality.

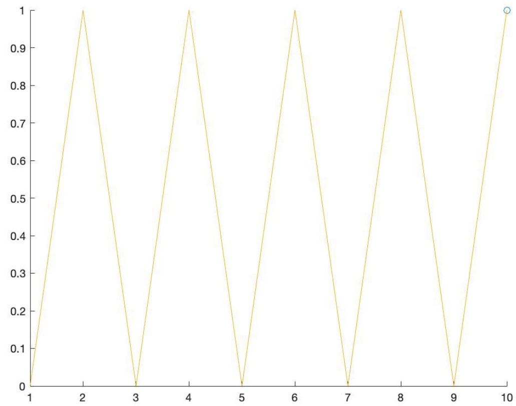

Figure 8. Delay Limitations

We again evaluate our technique compared to several other implementations, both those with a constant aggregation probability and those with a fixed aggregation period. Our method outperforms all others by a wide margin, as shown in **figure 9**, and it can successfully transmit 90% of the created packets even when a stringent delay constraint is enforced, and traffic volumes are high. Since the method uses nearly the same amount of energy for transferring much more packets, it achieves higher energy consumption efficiency than the other alternatives.

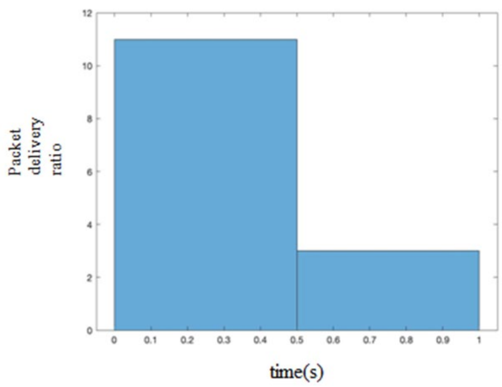

Figure 9. Delay Constraints from 0 to 0.9s

5. Conclusion

Wireless sensor networks have been considered a prospective and realistic solution that may play a vital part in the sensing, gathering, and monitoring of data in many physical settings for many smart city applications, for both public and private usage. Furthermore, data aggregation methods are considered a practical and efficient paradigm for decreasing the volume of traffic in sensor networks, extending their useful life, and alleviating their processing load. This research presents and assesses a framework for optimizing data collection in restricted sensor networks. With the proposed framework, we can calculate the optimal likelihood for a node to aggregate packets and the optimal aggregation length for a node to wait in order to reduce total energy consumption while meeting specific mandated delay restrictions. Extensive simulation results show the effectiveness and adaptability of the proposed framework by showing how it uses primal-dual decomposition to solve the corresponding optimization problem. Based on the reported study, the suggested Algorithm, Aggregation-based Energy Management, was compared to existing approaches that either don't aggregate data or aggregate data with constant probabilities and for specified durations across a variety of traffic loads and application situations. The suggested framework provides a reliable paradigm that minimizes energy consumption, maximizes the likelihood of message delivery, and expands knowledge while incurring just a negligible hit to precision, as shown by the performance evaluation procedure and corresponding simulation results. Last but not least, our framework's ability to rapidly adjust to new situations is a major selling point that makes it well-suited for use in challenging and ever-evolving settings and Programs.

Acknowledgements.

I thank my coauthors Mr. S. Balaji, Mr. S. Jeevanandham, Dr. M. Sundarrajan and Dr. Rajesh Kumar Dhanaraj for their expertise and assistance throughout all aspects of our study and for their help in writing the manuscript.

References

[1] Alkwai LM, Mohammed Aledaily AN, Almansour S, Alotaibi SD, Yadav K, Lingamuthu V. Vampire Attack Mitigation and Network Performance Improvement Using Probabilistic Fuzzy Chain Set with Authentication Routing Protocol and Hybrid Clustering-Based Optimization in Wireless Sensor Network. MPE. 2022; 2022:1-11.

[2] Bhaskarwar RV, Pete DJ. Energy efficient clustering with compressive sensing for underwater wireless sensor networks. PPNA. 2022; 15;2289-2306.

[3] Balamurugan A, Janakiraman S, Priya MD, Malar ACJ. Hybrid Marine predators optimization and improved particle swarm optimization-based Optimization cluster routing in wireless sensor networks (WSNs). CC. 2022; 19(6):219-247.

[4] Yadav R, Sreedevi I, Gupta D. Bio-Inspired Hybrid Optimization Algorithms for Energy Efficient Wireless Sensor Networks: A Comprehensive Review. Electronics. 2022; 11(10):1-22.

[5] Senthil GA, Raaza A, Kumar N. Internet of Things Energy Efficient Cluster-Based Routing Using Hybrid Particle Swarm Optimization for Wireless Sensor Network. WPC. 2022; 122(3):2603-2619.

[6] Gamal M, Mekky NE, Soliman HH, Hikal NA. Enhancing the Lifetime of Wireless Sensor Networks Using Fuzzy Logic LEACH Technique-Based Particle Swarm Optimization. IEEE Access. 2022; 10:36935-36948.

[7] Sahoo B M, Pandey H.M, Amgoth T. A Genetic Algorithm Inspired Optimized Cluster Head Selection Method in Wireless Sensor Networks. SEC. 2022; 75.

[8] Yadav RK, Mahapatra RP. Hybrid metaheuristic algorithm for optimal cluster head selection in a wireless sensor network. PMC, 2022; 79.

[9] Prakash PS, Kavitha D, Reddy PC. Delay-aware relay node selection for cluster-based wireless sensor networks. Measurement: Sensors. 2022; 24.

[10] Amutha J, Sharma S, Sharma SK. An energy-efficient cluster based hybrid optimization algorithm with static sink and mobile sink node for Wireless Sensor Networks. ESA. 2022; 203.

[11] Vanitha CN, Malathy S, Dhanaraj RK, Nayyar A. Optimized Pollard Route Deviation and Route Selection using Bayesian Machine Learning Techniques in Wireless Sensor Networks. Computer Networks. 2022; 216.

[12] Muthukumar S, Rajesh DH. Optimal Cluster Based Routing Technique for Wireless Sensor Networks using Hybrid Optimization Algorithm for Maximizing Life of Sensors. WPC. 2022; 125:3479-3500.

[13] Khalaf OI, Romero CAT, Hassan S, Iqbal MT. Mitigating hotspot issues in heterogeneous wireless sensor networks. JS. 2022; 2022:1-14.

[14] Hegde K, & Dilli R. Wireless Sensor Networks: Network Life Time Enhancement using an Improved Grey Wolf Optimization Algorithm. ES. 2022; 19:186-197.

[15] Manoharan G, Sumathi A. Efficient routing and performance amelioration using Hybrid Diffusion Clustering Scheme in heterogeneous wireless sensor network. IJCS. 2022; 35(13).

[16] Gautam AK, Yadav R. Energy efficient hybrid routing protocol for Wireless Sensor Networks Using AI Technique. EAI ETSIS. 2022; 9(35):1-9.

[17] Khera S, Turk, N., Kaur, N: HC-WSN - A Hibernated Clustering based framework for improving energy efficiency of wireless sensor networks. MTA. 2022; 82(3):3387-3894.

[18] Joshi P, Kumar S, Raghuvanshi AS. A performance efficient joint clustering and routing approach for heterogeneous wireless sensor networks. ES. 2023; 40(5).

[19] Rajesh L, Mohan HS. Adaptive Group Teaching Based Clustering and Data Aggregation with Routing in Wireless Sensor Network. WPC. 2022; 122(2):1839-1866.

[20] Alom, M. K., Hossan, A., Choudhury, P. K: Improved Zonal Stable Election Protocol (IZ-SEP) for hierarchical clustering in heterogeneous wireless sensor networks. E-Prime-AEEE. 2022; 2.

[21] Prasanth A, Sabeena G, Sowndarya, Pushpalatha, N. An artificial intelligence approach for energy-aware intrusion detection and secure routing in internet of things-enabled wireless sensor networks. CCPE. 2023; 35(23).

[22] Paruthi Ilam Vazhuthi P, Prasanth A, Manikandan SP, Devi Sowndarya KK. A hybrid ANFIS reptile optimization

algorithm for energy-efficient inter- cluster routing in Internet of Things-enabled Wireless Sensor Networks. PPNA. 2023; 16:1049-1068.

[23] Vazhuthi P, Manikandan SP, Prasanth A. An Energy-Efficient Auto Clustering Framework for Enlarging Quality of Service in Internet of Things- enabled Wireless Sensor Networks using Fuzzy Logic System. CCPE. 2022; 34(25).