

Optimization of data aggregation process in HWSN using adaptive clustering protocols

Zuhair N.mahmood ^{1*}Salah A.Aliesawi ²

^{1,2}College of Computer Science and Information Technology, University of Anbar, Anbar, Iraq
zuh21c1015@uoanbar.edu.iq , salah_eng1996@uoanbar.edu.iq

Cite this paper: Zuhair N.mahmood, Salah A.Aliesawi (2024) Optimization of data aggregation process in HWSN using adaptive clustering protocols. *Frontiers in Health Informatics*, 13 (3), 9559-9576

ABSTRACT

The Wireless sensor networks (WSNs) are intended for use in applications where energy is constrained and therefore, the longevity of the network is an essential aspect to the design. From WSN protocols and data aggregation techniques which reduce the numbers of redundant data packets, save energy and make the network lifespan longer, have led to the increased importance of heterogeneous WSNs; Consequently, there is the need for efficient data gathering protocols that minimize energy consumption, congest time and enhance network lifetime of the HWSNs, This study focuses particularly on clustering algorithms for data collection in HWSNs. Clustering makes it easy to organize the sensor nodes and the operation of the main cluster involves collecting information from the nodes and delivering it to the sink node. Furthermore, this review goes further to explore other types of clustering methods such as LEACH, TEEN and SEP to determine their ability to improve data aggregate performance These reveal the special characteristics of HWSNs and the challenges faced; the network impact is shown and factors that affect the aggregate efficiency such as data collection, clustering and head selection, and routing are identified Advanced techniques like hierarchical clustering, Machine learning and Fuzzy logic based are also discussed as possible solutions Finally the review presents static issues such as security, reliability, and power and overviews major conclusions and future research agenda. The research contributes to the development of more effective ideas to further improve the network based on the assessment of WSN performance and the longevity of the wireless sensor network The research emerges as a handy tool for researchers and professionals in the field, most especially in data aggregation for HWSNs.

Keywords: A new low-power-sensitive threshold-based energy-efficient sensor network protocol, stable.Election protocol, low energy adaptive clustering hierarchy.

1. Introduction

Technological advancements have integrated numerous sensors into the Internet of Things (IOT). Considering this, in 2021, IOT is going to be composed of just over 20 billion connected devices that can generate immense volumes at an unprecedented rate of more than 2.5 quintillion bytes of data per day. This will be driven by massive adoptions of artificial intelligence, whereby people can really leverage and gain value from the data. As a result of this, the sensor devices are finding their deployments in various applications for collecting and analyzing data [1]. A WSN is made up of several tiny sensor nodes operating on batteries. These nodes monitor a wide range of environments and physical factors from different locations, which include motion, temperature, pressure, vibration, noise, and contamination. The network's node collects data from unreachable locations and transmits it to the central location, also called the sinking station. WSNs deploy their mobile or permanent nodes in the required area in a suitable or random manner. Some of the attributes of sensor nodes include but are not limited to size, durability, power level, and energy consumption. The widespread application of WSNs ranges

from traffic control, industrial operations, military monitoring, etc., to medical diagnoses, industrial processes, military surveillance, and traffic management [2].

Data will be gathered, processed, and transmitted to the base stations routinely by the sensor. The number of sensors used in the process or frequency of reporting data depends upon the particular application of interest. Data collection, in general, represents a methodical procedure for collecting sensed data from multiple sensors and sending it to the base station for processing. Because the sensor node has energy limitations, considering the highly correlated data generated by neighboring sensors, it would be inefficient to perform raw data transmission directly from all sensors to the base station. In addition, base stations cannot process the vast amount of data generated by a more extensive sensor network. Therefore, certain networks need to be employed at the intermediate nodes or sensors in order to collect data and provide meaningful information. This can save energy and bandwidth potentially by reducing the number of packets sent to the base station [3].

In fact, the data packets from the surviving sensor nodes ought to be processed alone and combined as well as summarized by the key primitive called DA before they are forwarded to the upper nodes. Or, every single unnecessary or redundant bit will be hooked to the big sensor networks by being disconnected from its corresponding sensor node. First of all, data packets from surviving sensor nodes must be processed, merged, and summarized using the DA called essential primitive before being sent to the upper nodes. Thus, the collected data will be separated from its respective sensor node irrespective of the duplicate data and will get connected to the broad sensor networks. Sensor nodes communicate in a network carrying values with aggregators and other selected nodes to pass on the values. Each aggregator then sums the data before sending it, as in Figure 1 [4].

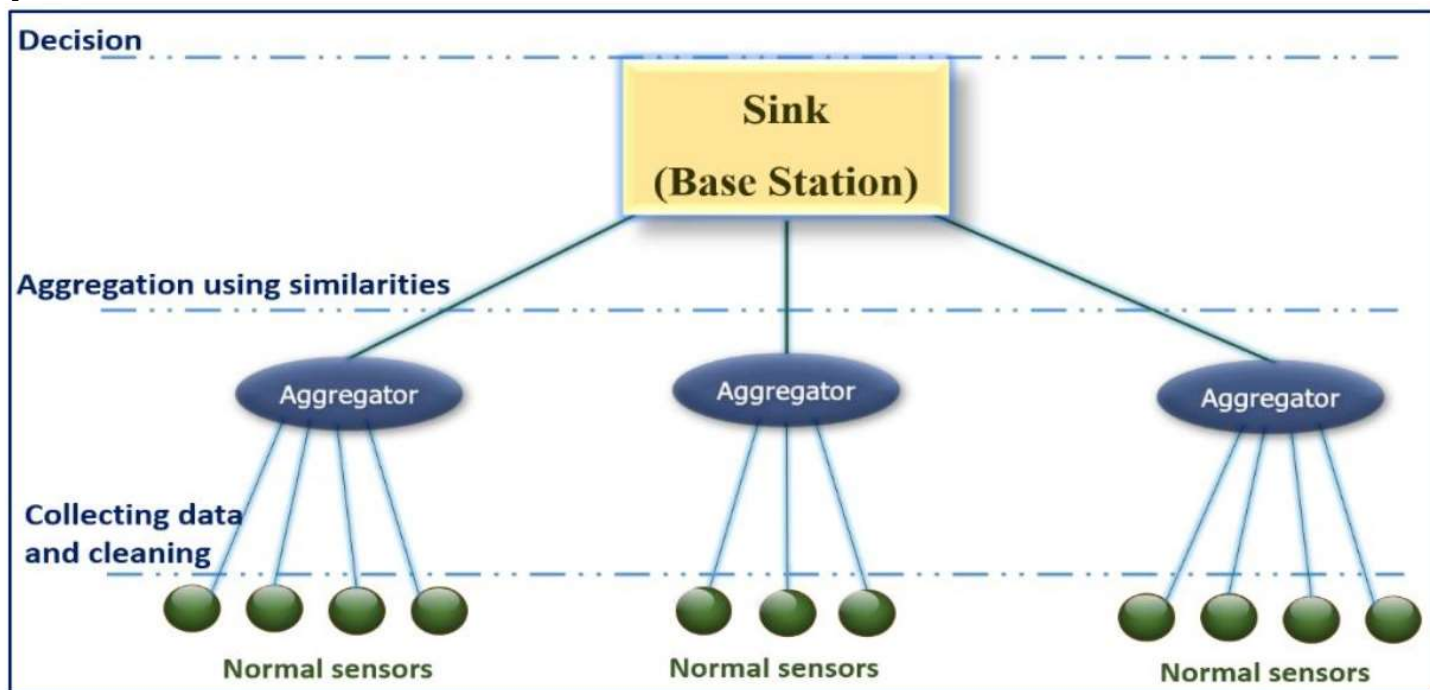


Figure 1 Data Aggregation Scheme [4].

1.1 . Motivation of the survey

Cluster-based protocols have contributed significantly to many domains in WSN, where most of the sensor energy is utilized efficiently to enhance the lifetime of the network. They make use of different methodologies

and optimization techniques towards effective management of clustering operations. Despite this fact, comprehensive reviews are yet to be presented that can provide more valuable insight into these clustering methodologies and help researchers learn about this topic and consider new research directions. With this objective in mind, the present survey is intended to bring out from different techniques and design characteristics a deep understanding of various clustering protocols that have contributed to network performance enhancement. This survey will provide the researchers with a comprehensive overview of the clustering domain by classifying protocols with regard to methodologies and functionalities and analyzing their strengths and weaknesses.

1.2 Contribution of the survey

This survey outlines clustering protocols' techniques through a critical analysis of relevant features of the available various approaches. The explanation of protocols in the context of performance metrics and methodologies concerning clustering and classification regarding operational principles, network structure, and operational mechanisms is given. Network type and clustering methodologies come up with four categories regarding the clustering protocols surveyed. Major contributions from this survey are:

1. Presenting a new point of view in studying the existing clustering protocols, which are classified by operational principles, network structure, and methodologies.
2. Comparing the performance of different approaches based on feature, performance, and clustering parameters.
3. Analyzing various categories of protocols based on their objectives, key features, advantages, and disadvantages will help the researcher in shaping new research methodologies.
4. Extensive review of clustering parameters, feature methodologies, and performance of various protocols that would lead to a better understanding of clustering models.

1.3. Problem Statement

Heterogeneous Wireless Sensor Networks (HWSNs) are becoming increasingly important for real-world applications. However, traditional data transmission methods in HWSNs lead to:

- High Energy Consumption: Sensor nodes have limited battery life, and frequent data transmissions deplete energy rapidly.
- Network Congestion: Redundant data transmissions overload the network, hindering efficient communication.
- Limited Network Lifespan: The aforementioned issues shorten the overall lifespan of the sensor network.

Therefore, there is a critical need for advancements in data aggregation techniques specifically designed for HWSNs. These advancements should aim to:

1. Reduce data transmission by eliminating redundancy and processing information closer to the source.
2. Optimize energy usage by minimizing data transmissions and utilizing efficient aggregation methods.
3. Extend network lifespan by conserving energy and reducing network congestion.

1.3 Organization of the paper

Overview of Heterogeneous Wireless Sensor Networks in Section 2. Importance of Data Aggregation Techniques in Section 3. Clustering Protocols for Energy Reduction in section 4. and Discussion in Section 5. potential future research direction in section 6 , conclusion In Section 7, the literature on clustering algorithms and these methods for HWSN energy conservation is reviewed.

2. Overview of Heterogeneous Wireless Sensor Networks (HWSNs)

Network types depicted in Figure 2 are heterogeneous and homogeneous wireless sensor networks. Smart and sensor nodes in heterogeneous wireless sensor networks can be distinguished by their processing power and sensing ranges, for example. A high-level energy or communication capability is displayed by the heterogeneous high-level node CH, as seen in Fig 3. Three. Gathering and displaying data from its surroundings is the sensor node's principal job. Advancement nodes may have more memory and more powerful microprocessors and microcontrollers than cluster member nodes.

Longer network lifespans, reduced latency, and improved data transmission reliability are just a few advantages of these HWSNs. Reduced latency, improved data transmission reliability, and extended network life are some benefits of HWSNs[5].

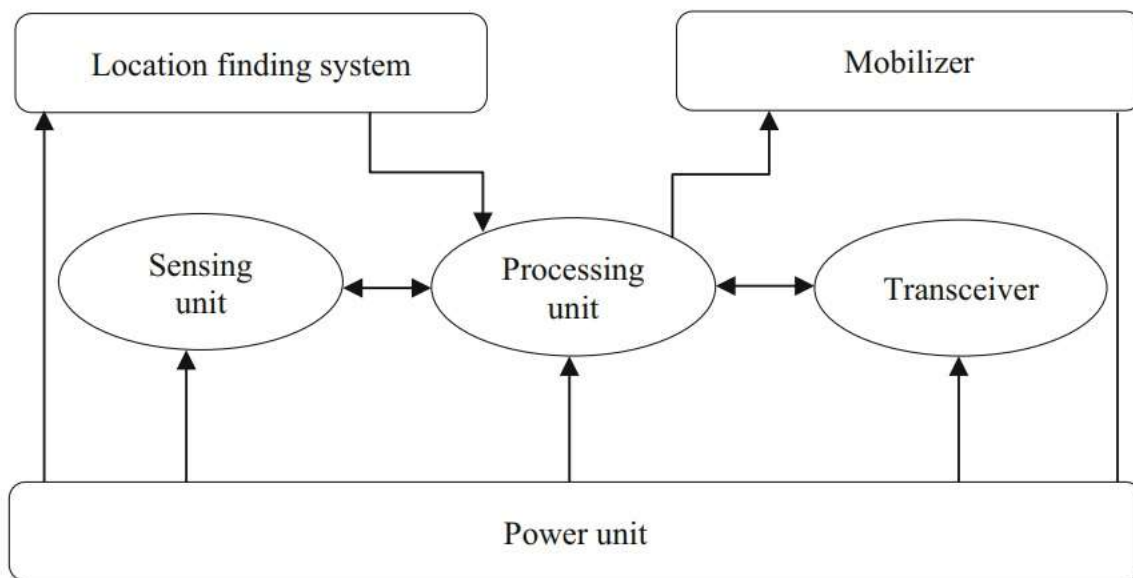


Figure 2 Sensor Node Structure

The sensors positioned in the adjacent region provide duplicate data since these sensors usually identify the same events that must be quantified. The sink node will therefore be mostly in charge of managing the amount of data, bandwidth, and energy used. Data aggregation thus combines information from multiple sensors dispersed over the target environment. The objectives of this technique are to reduce superfluous data transmissions and provide information to the sink node . Figure 4 depicts two wireless sensor network environments: one with data aggregation and the other without[7-9].

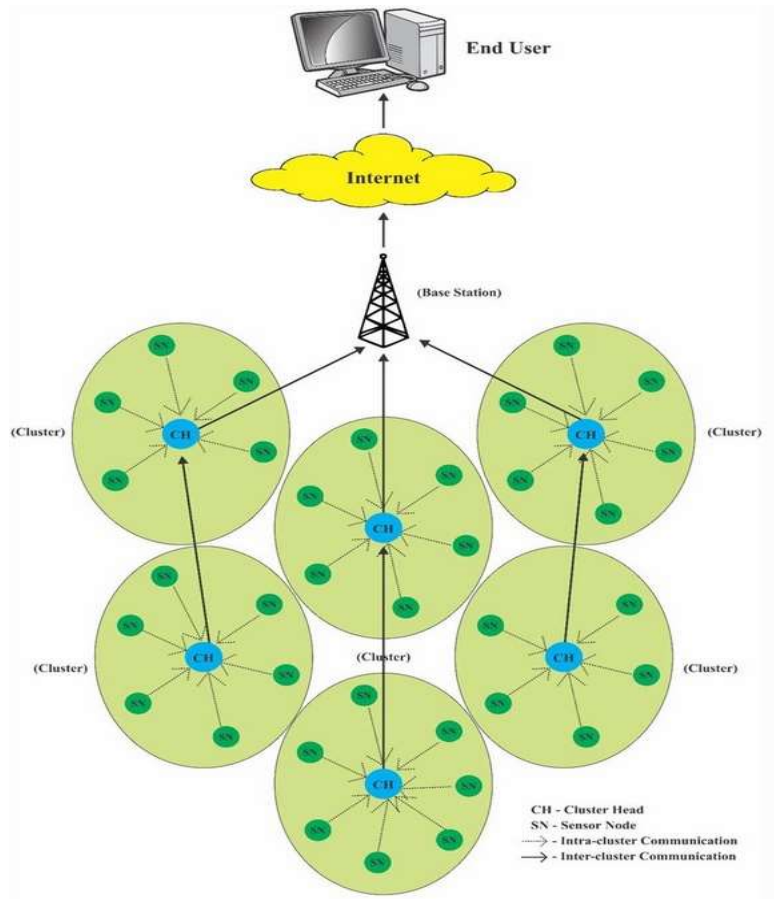
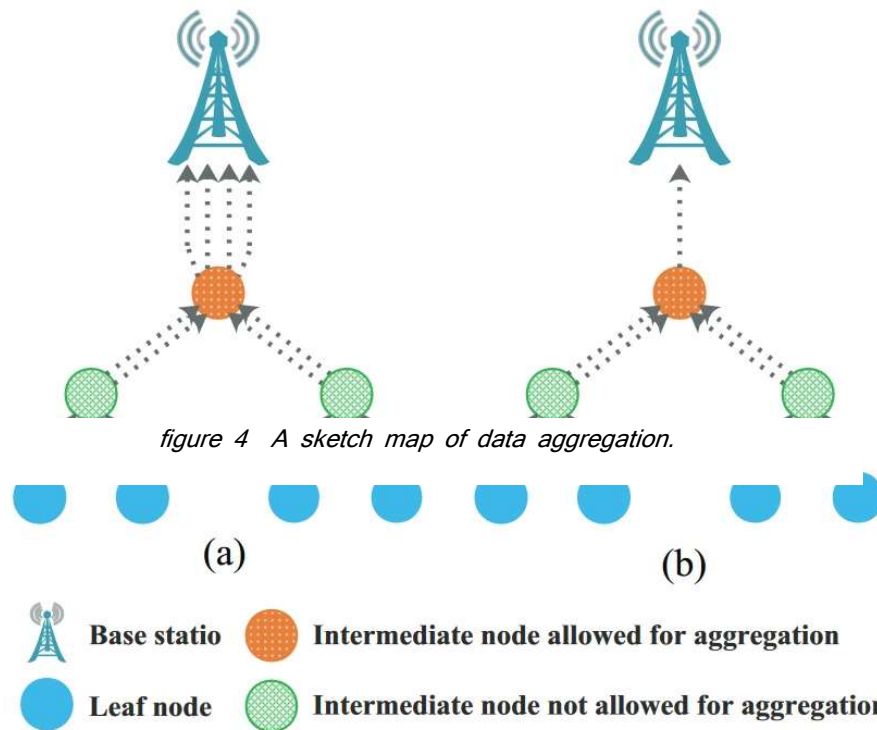


Figure 3 Heterogeneous high-level node CH [6]



The heterogeneous WSN is even more versatile, since it is basically built from sensor nodes possessing different features, including a different kind of communication and sensing range-meaning also different kinds of sensors. As an example, one could build a WSN with nodes made up of diverse types of sensors. Assuming that there are two different kinds of sensor nodes, the higher-end ones would be more expensive and take longer to process, but they would also be much cheaper and have lower processing, communication, and sensing power. Mixing high-end and low-end nodes could result in an execution balance at a cost-effective level[10].

Hence, energy optimization has become a major research factor in HWSN. Secondly, some of the sensors may also forward the data in an HWSN by acting like relays. These, as compared to the other sensors, would die out way sooner. Applications running on HWSNs that require sending pictures or videos use a great amount of energy. As a result, HWSN might split into a number of independent network branches, where the Data transfer becomes challenging.

The optimal approach to reducing data transmission process energy consumption and healing split network branches "must be found to enhance HWSN energy efficiency." The goal of the new relay nodes is to recover connectivity losses in HWSNs. However, as identified by earlier research, the placement of relay nodes is NP-hard[11].

It is challenging to conduct energy efficiency research because packets are sent over unreliable networks. Various approaches have been proposed with respect to energy usage. Nevertheless, clustered routing algorithms are among the most reliable protocols for energy efficiency, load balancing, communication cost, and packet delivery from the sink node to the base station (BS). Prior research has largely focused on energy consumption; however, energy efficiency is an important aspect of WSN, and several energy-efficient algorithms have been created.

However, none of these are really effective. Many routing algorithms have been created to handle different problems, such as energy usage, transmission costs, and load balancing. The review of the literature examines the different cluster-based routing protocol techniques with a focus on hierarchical based routing [12].

Clustering is one of the widely used techniques for energy consumption reduction and enhancement of network lifetime in HWSNs, as can be seen in works such as [10, 13-17][21-18] :

The High-Quality Clustering Algorithm (HQCA) uses fuzzy logic to select optimal Cluster Heads (CHs) based on criteria like residual energy, minimum and maximum energy in each cluster, and minimum and maximum distances between nodes and the base station [13].

In this regard, a Dynamic Uneven Clustering Protocol, DUCP autonomously clusters the Energy Harvesting Sources in energy harvesting wireless sensor networks, to enable CHs energy management and reduce energy consumption[14].

Therefore, the UCT2TSK algorithm, which is based on the theory of Interval Type-2 TSK Fuzzy Logic, takes into consideration relative distance to the base station, residual energy, and node density to optimize CHs and cluster sizes for a balance in energy consumption and an enhancement in energy efficiency[15].

The Clustering Election Game-Based and Two-Level Management Protocol (CEGT) divides the network into layers, selects layer heads based on location and energy, and then performs a two-stage screening and election games to dynamically select CHs, ensuring reasonable energy utilization and efficient load balancing [10].

Other energy-efficient clustering and routing protocols have also been developed for HWSNs, namely EDFCM, MCR, EEPKA, LEACH, and SEP to further prolong the lifetime and energy consumption of the networks[10] . Network coding and topology control techniques have also integrated with clustering to further reduce energy consumption in HWSNs[16, 17].

In heterogeneous networks, a variety of topologies are employed, which increases the network's complexity. To put it briefly, a heterogeneous sensor network makes use of two or more distinct types of network nodes, each having a unique set of features and battery life. The real cause of heterogeneous networks is that certain cluster heads need to have more sophisticated hardware and larger batteries in order to reduce the overall hardware cost of the remaining sensor network. However, the inability of the cluster head nodes to correct is being caused by role rotation. the cost of the remaining sensor network. However, role rotation is the reason behind the cluster head nodes incapacity to correct.

Multiple design factors in heterogeneous WSN lead to a successful wireless sensor network. Several of them are displayed as follows:

1. Self-healing: In this aspect, WSNs differ from traditional networks.
2. Memory and Processing Capabilities: It is expected that the sensor nodes' memory and processing capabilities will be limited.
3. Limited Bandwidth Consumption: Processing and sensing often require less energy than communication.
4. Flexibility: The management system must have enough flexibility to quickly adapt to modifications in the topology and dynamics of the network.
5. Scalability: It is expected that a sizable number of sensor nodes will be installed.
6. Production Costs: To reduce the cost of each sensor node, the management system should have the fewest resources possible.
7. Power Consumption: Since most sensor devices are battery-operated, the management program must be able to function on the sensor nodes without consuming an excessive amount of energy[22].

3. Importance of Data Aggregation Techniques

Data collection is a process in which there is regular gathering of information from various sensors and then transmitting it to the base station for analysis. Not all sensors are enabled with the capability to send data to the base station directly or efficiently due to the energy limitation of every sensing node. This is explained due to redundancy caused by the large data overlap in the systems of nearby sensors . In addition, the base station does

not have the capability to handle such a large amount of data generated by a wider diversity of sensors. Such networks can save energy and bandwidth by integrating in-network data and providing meaningful information either at sensors or at intermediary nodes, thus reducing the number of packet transfers to the base station. In-network data aggregation techniques that typically include the process of data integration from multiple sensors[23].

Besides, the energy- and bandwidth-consuming separate data transfer of each node reduces the lifetime of the sensor network. The algorithms for data aggregation have been developed to bypass this problem. Data aggregation aims at reducing network traffic and canceling the unnecessary data transfer with a view to increase energy and longevity while minimizing data overlap [24].

The benefits of merging data include:

- Maintaining energy and reducing traffic load through data aggregation;
- this improves network resilience and reduces redundancy.
- One of the best features of a smart data aggregation method is its ability to reduce node energy consumption while maintaining network lifetime.

Notwithstanding these benefits, data aggregation also results in an increase in packet transmission latency within the network [25].

3.1 Classification of data aggregation protocols and approaches

Data aggregation protocols define the rules of (a) combining sensed data through aggregation functions, (b) managing the transmission of data and control messages, and (c) routing aggregated data toward the base station or internet cloud. The ultimate objectives of these protocols are the minimization of redundant data transmissions, accuracy of data during aggregation, and prolonging the lifetime of WSN/IOT systems. To support these different applications and operation scenarios, we classify data aggregation protocols, with respect to: (a) the structure of WSN/IOT network; (b) the interference and fault-tolerance models employed; (c) the security and privacy concerns; and (d) the mobility. In Figure 5 a classification of some data aggregation protocols according to the described criteria is shown.[26].

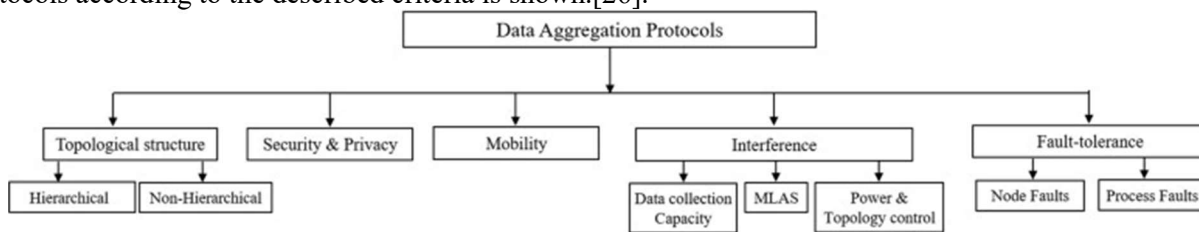


Figure 5: A taxonomy of data aggregation protocols.

4. Clustering Protocols for Energy Reduction

Some of the objectives of clustering include data aggregation, fault tolerance, scalability, and energy efficiency. Clustering improves energy efficiency through the selection of energy-efficient nodes as CHs, which gather, compile, and transmit the data from other nodes. The re-clustering could achieve the goal of fault tolerance if the CH energy goes below the threshold value. Any node that outperforms on many factors can be selected as a CH. When the network forms clusters of equal size, the CHs close to the BS will run out of energy faster. In order to keep load balance and reduce the burden of the nodes closer to the base station, unequal clustering is formed. It is possible to do the clustering in both homogeneous and heterogeneous networks. When different amounts of resources are given to all the nodes in a network, it is called heterogeneous; in a homogeneous

network, every sensor node gets the same amount of resources. Heterogeneity can increase the lifetime of the network and enhance reliability but the cost of implementation is higher as well.

Another important aspect is the optimum percentage of heterogeneous nodes to be added in the network. A heterogeneous node can have more network resources compared to conventional nodes in aspects such as energy, memory, link range, connection capacity, and so on. But there should be fewer such nodes to keep the cost factor in consideration. Energy heterogeneity in heterogeneous networks is addressed using a clustering technique in this work[27].

Clustering offers scalability benefits as the sensor network grows .Moreover, it offers reliability by exploiting the locally adapted solutions. The time to sleep and wake up is scheduled based on clustering techniques that minimize power consumption. In this regard, all sensors are not needed because most of the applications upon WSN use energy at night. As a result, it means that there is a bound on the number of sensor nodes allowed to sleep to save energy. Different academics have proposed many clustering techniques[28].

4.1 Comparison of Clustering Protocols in heterogeneous WSNs

This comparison is focused on the effectiveness of the different clustering protocols in WSNs, with an emphasis on the heterogeneity of sensor nodes in terms of capabilities and characteristics , Table 1 refer to some previews work whishes compared based on (Network Type , Nodes , Aggregation , Algorithm ,Application , Remarks). The different protocols were analyzed using a set of metrics that assessed their performance about how each protocol addressed opportunities and challenges of heterogeneous environments, shown in table 2.

Table 1. Comparison of related techniques

Reference	Focus	Network Type	Nodes	Aggregation	Algorithm	Application	Remarks
[1]	Improve SEP for heterogeneous WSNs	Heterogeneous	Not specified	Not specified	SEP with threshold	Not specified	Focuses on energy efficiency
[2]	Reviews data aggregation techniques	General WSNs	Not specified	Data aggregation methods	Reviews various techniques	Not specified	Not directly comparable to HHSEEC
[3]	Reviews secure data aggregation techniques	General WSNs	Not specified	Secure data aggregation	Reviews various techniques	Not specified	Not directly comparable to HHSEEC
[4]	Discusses WSN challenges and possibilities	General WSNs	Sensor nodes	Not specified	Not specified	Various	Provides context for WSN research

[5]	Surveys WSNs	General WSNs	Sensor nodes	Not specified	Not specified	Various	Provides context for WSN research
[6]	Reviews energy-efficient routing for heterogeneous WSNs	Heterogeneous	Not specified	Not specified	Routing protocols	Not specified	Similar focus on energy efficiency
[7]	Routing method for heterogeneous WSNs	Heterogeneous	Not specified	Not specified	Routing protocol	Not specified	Similar focus on routing in heterogeneous WSNs
[8]	Data aggregation method for WSNs	General WSNs	Not specified	Data aggregation approach	Perceptually important points-based	Not specified	Not directly comparable to HHSEEC
[9]	Data aggregation approach for WSNs	General WSNs	Not specified	Data aggregation approach	Open pits mining-based	Not specified	Not directly comparable to HHSEEC
[10]	Reviews data aggregation algorithms	General WSNs	Not specified	Data aggregation algorithms	Reviews various techniques	Not specified	Not directly comparable to HHSEEC
[11]	Clustering protocol for heterogeneous WSNs	Heterogeneous	Not specified	Not specified	Clustering protocol	Heterogeneous load balancing	Not specified
[12]	Evaluates DEEC-based routing protocol	General WSNs	Not specified	Not specified	DEEC with enhanced threshold	Not specified	Compares to DEEC, a base for HHSEEC
[13]	Compares WSN clustering algorithms	General WSNs	Not specified	Not specified	Clustering algorithms	Comparison of various approaches	Not specified
[14]	Analyzes energy efficiency of clustering protocols	General WSNs	Not specified	Not specified	Clustering protocols	Energy efficiency analysis	Not specified
[15]	Compares DEEC protocol	General WSNs	Not specified	Not specified	DEEC protocol	New version for extended lifetime	Not specified

	versions						
[16]	Reviews LEACH protocol	General WSNs	Not specified	Not specified	LEACH protocol	Review	Not specified
[17]	Clustering algorithm for heterogeneous WSNs with solar energy	Heterogeneous	Solar-powered nodes	Not specified	Clustering algorithm	Solar energy-aware	Not specified
[18]	Clustering scheme for energy-harvesting WSNs	General	Solar-powered nodes	Not specified	Clustering algorithm	Solar energy-aware	Not specified

Energy Efficiency: Energy is one of the critical issues in WSNs, and this heterogeneity becomes graver since nodes in a practical environment may possess different power capacities. This comparison assesses how each protocol maintains energy resources to prolong network operation.

Scalability: Scalability of the clustering protocols is of great essence in accommodating the dynamic nature of WSNs, especially for heterogeneous deployments, where the number or type of nodes might change from one node to the other. Scalable protocols can easily handle network expansion without performance degradation.

Network Lifetime: Network lifetime is considered the most important factor in heterogeneous WSNs, and it has to be extended to enable it to continue collecting and transmitting data for a longer period of time. In this respect, the survivability aspect of a protocol with its effectiveness in the management of energy and network resources is considered.

Security: With the advent of numerous IoT devices, one should ensure data security and privacy. This paper performs a comparison keeping in view the security features each clustering protocol has implemented to protect sensitive information against any kind of potential cyber threat.

Complexity: Clustering protocols' complexities affect the feasibility of realization versus resources utilized. While complex protocols may provide advanced functionalities, large computational resources are taken up, and vice versa-their simpler counterparts focus on efficiency and conservation of resources.

Heterogeneity Adaptability: Heterogeneous WSNs consist of nodes of different capabilities and may be designed for different sensing modalities or communication protocols. This adaptiveness of each protocol against such heterogeneity should be gauged to make sure it can integrate and cooperate seamlessly in a diverse range of environments.

Efficiency of data aggregation: Data aggregation is considered one of the important roles in both avoiding redundant transmissions and saving energy in WSNs. This aspect investigates how each protocol will aggregate and process the sensor data efficiently with minimal latency and overhead.

Table 2. Comparison of Clustering Protocols

Metrics	SEP	TSEP	Z-SEP	DEEC	HNS	PHC	TEEN	EBCS	LEACH
Energy Efficiency	High	High	High	High	Moderate	High	Moderate	High	Moderate
Scalability	High	High	High	High	High	High	High	High	Moderate
Network Lifetime	High	High	High	High	High	High	High	High	Moderate
Security	Moderate	High	High	High	High	High	High	High	Low
Complexity	Moderate	High	High	High	High	High	High	High	Low
Heterogeneity	High	High	High	High	High	High	High	High	Moderate
Data Aggregation	High	High	High	High	High	High	High	High	Moderate

In WSN, various clustering protocols are proposed with the essential objective of improving network efficiency, maximizing node lifetime, and maintaining energy balance across the network. It finds its application in protocols like LEACH (Low-Energy Adaptive Clustering Hierarchy), PEGASIS (Power-Efficient gathering in Sensor Information System), and DEEC (Distributed Energy-Efficient Clustering), which have various methods of selecting cluster heads, aggregating data, and dealing with energy, one of them following :-

3.1. SEP

First to stretch the lifetime of a network comes Stable Election Protocol-SEP. It considers the possibility and its effects of energy heterogeneity; that is, the sensor nodes of the network use different amounts of energy at different instants of time. This may be due to their orientation w.r.t the base station:

- The level of the data gathered in a CH depends on the scale of each cluster.
- Some nodes face malfunction while in transfer due to the location of deployment.
- The network may acquire another set of extra nodes with which to give life anew.
- Renewable energy sources like solar, wind, and other energy sources may be used to power the nodes.

It can be seen that all the foregoing techniques have the potentials to create an energy heterogeneity. The outcome of all this work is the SEP protocol that uses weighted election probabilities of each node for becoming CH based on its remaining energy. The network lifetime of WSNs with the SEP protocol instead of LEACH protocol enhances considerably[28].

4.1.1. TSEP

The Threshold Sensitive Stable Election Protocol, or TSEP, is another hybrid protocol that incorporates the SEP and TEEN protocols. The protocol can be used as an example of a reactive network protocol, which, upon any changes within the data it detects, immediately acts to those changes. The TSEP can be used in heterogeneous networks for the benefit of these. The TSEP method consists of two phases. The first step is the division of people into advance, intermediate, and normal categories. Their initial energies are highest, moderated, and lowest, respectively. $E_{adv} = E_0$ for normal nodes, $E_{adv} = E_0(1+\alpha)$, and $E_{int} = E_0(1+\mu)$ for the intermediate nodes. μ is the extra energy factor $\mu = \alpha/2$ between the intermediate and normal nodes[29].

4.1.2 Z-SEP

The base of the Z-SEP protocol is zone partitioning. The network contains more than one zone; for example, zone 0, zone 1, and zone 2. The standard node exists in network zone 0 while the advance nodes exist in zones 1 and 2. The nodes that are not participating in the clustering process send data directly to the base station. It uses Zone 1 and 2 to select cluster heads. This protocol enhanced the stability period of the network and reduced the time period of instability introduced by direct transmission, but the power in conventional nodes gets exhausted in a very short span of time, which leaves a large area of coverage unexplored. The power in

conventional nodes dissipates in a very short span of time by making use of a wide area of coverage[30].

4.1.3 DEEC

The DEEC protocol uses a heterogeneous routing technique to determine cluster leaders by dividing the average energy of the network by the residual energy of each sensor node. DEEC's two-level nodes are similar to SEP's in terms of node kind. The primary issue with this strategy is that when advanced nodes' energy levels fall to those of a regular node, they typically experience unfavorable outcomes. Thus, the goal of our research is to improve this tactic. ALBIRI developed an energy-efficient heterogeneous network technique based on the DEEC protocol. The DDEEC protocol solves a fundamental issue with the DEEC methodology, which penalizes sophisticated nodes. Nevertheless, when deciding which option was best, the algorithm ignored the distance between each node and the base station[31].

4.1.3 LEACH

LEACH was the first hierarchical routing system primarily designed for sensor networks. This adaptive clustering technology performs an even distribution of the energy load among the sensor nodes of a network by means of randomized cluster head rotation. This type of stochastic process has in common the characteristic of self-organizing, hence highly flexible. The layout used by the LEACH protocol is one-hop. Regular cluster nodes and a cluster header form every WSN cluster.

The protocol randomly chooses the cluster head, which operates with the sensor nodes when monitoring the balance of power in the network. This approach divides the operation of LEACH into multiround, whereby at every round, there is the Setup and Steady phase; the cluster head, directly linked to the monitoring station, switches off as many nodes as possible to ensure energy conservation.

Following are the major disadvantages of LEACH:

- Inappropriate cluster head selection: As the LEACH algorithm does not consider each node's initial energy, all nodes have equal opportunities to be a cluster head. Choosing a low-energy node will cause it to malfunction quickly, which impairs the performance of the whole network.

It selects cluster heads randomly using the unproven procedure, hence the cluster head distribution in the network is not proper; therefore, resulting in an unbalanced network load. Therefore, it is a serious matter. Also, it ignores the distance between the cluster head and the main station while the process of creation of clusters is concerned, thus threatening the optimization of energy within the network.

Some other roles of the cluster head node will be that it will collect data from different ordinary nodes and, with a single hop, will send them to the monitoring station. Everybody knows that its energy gets drained out sooner compared to the rest of the nodes. The entire cluster will be finished in case the cluster leader dies.

It is incompatible with long running applications; the latency makes it impossible to connect it to a large-scale sensor network. As we mentioned earlier, routine updates to the monitoring station could be sent either in the form of information about an event or a question or both. LEACH technique is suitable for those conditions. Vary in size from cluster to cluster: since some clusters took place by chance, most probably some of them have a high concentration of sensing nodes and others have a low concentration.[32].

4.1.4 PHC

PHC is such a clustered routing technique proposed for data aggregation in heterogeneous WSN to enhance the life and efficiency of such networks.

The PHC technique is based upon the method of forming clusters among sensor nodes; each cluster will have one Cluster Head that performs data aggregation responsibility, along with transmission[33-35].

The CH selection is optimized using techniques such as Genetic Algorithm (GA) and Prairie Dog Optimization (PDO) to balance energy consumption and network lifetime [35, 36].

It follows therefore that PHC is also concerned with mechanisms and protocols for data aggregation and routing capabilities securely, even using Micro-segmentation, Batch Authentication, and Elliptic Curve Cryptography (ECC), so as to contribute to energy efficiency and safety for general data [34, 37, 38].

The PHC technique has been shown to outperform existing routing protocols in terms of network lifetime, energy consumption, throughput, and security [33-38].

4.1.5 ES-NMPBFO

The algorithm is designed to alleviate high computing by:

the complexity of the standard bacterial foraging optimization (BFO) algorithm by introducing periodic BFO. Further, two learning strategies, the global best individual (gbest), the personal best of each candidate solution and personal historical best individual, are used in the chemotaxis operation. This is done in order to enhance the convergence speed and drive the bacteria towards the optimum position. Two elimination-dispersal operations are also proposed to avoid falling into local Optima and improve the diversity of solutions. The proposed algorithm is compared to Compared with five other algorithms on six validity indexes on two clustering cases of data, nine general benchmark datasets, and four credit risk assessment datasets. The experimental results show that the proposed algorithm significantly outperforms the other rival approaches[39].

4.1.6 EBACS

EBACS is an entropy information theory-based clustering routing technique in heterogeneous WSN. EBACS categorizes the network into three layers: one for data transmission, another for cluster formation, and the last layer for cluster head selection. More precisely, EBACS considers a set of variables such as the distance of the node from the gateway device, its residual energy, and node density distribution in the monitoring field to select nodes operating as cluster heads with the help of certain entropy schemes.

The experiment results show that EBACS outperforms LEACH on the grounds of throughput and network lifetime. As shown by Hau, the intelligent agents can facilitate the AODV routing protocol for energy efficiency. Tan et al. suggested a clustering-based sector tree routing protocol for energy efficiency. According to this protocol the MST that links them provides the facility of communication to the nodes in a cluster. GAMAL et al. They proposed the technique for the combination of K-means clustering hierarchy with hybrid particle swarm optimization for creating clusters to enhance the lifetime of WSNs. This method has used K-means clustering hierarchy, where the fuzzy logic theory selects CHNs based on two parameters mentioned above: the level of residual energy and the distance between CHNs and BS. The scientists combined the technique of K-means clustering with the technique of particle swarm optimization in order to create clusters. Khatrinagbaf proposed the best ant colony algorithm for maintenance[40].

4.1.7 TEEN

TEEN It proposes threshold-based communication techniques to avoid useless transmission in energy-limited networks. The different threshold settings of TEEN can be selected by the sink node. It allows communication only between those nodes that can measure the variation in the wireless network atmosphere. When no variation is detected consecutively for numerous rounds, the majority of alive nodes are considered dead by it[41].

It is unique in having two different thresholds for the amount of data that TEEN collects, namely Hard Threshold and Soft Threshold. Once it reaches HT, the cluster node is allowed to send data to the nearby station. ST indicates that sensor node will continue to gather data until it reaches or exceeds this threshold. Hence, additional threshold values were set for this protocol. Additional parameters are also set for the case of heterogeneous sensor network[42].

It proposes TEEN protocol, a clustering technique that aims to point out abrupt changes in sensing attributes such as temperature. While the nodes periodically sense their environment, they consume lower energy than that proactive networks since data transmission is highly controlled and less frequent owing to two thresholds.

Data are transmitted only when required; hence, much less energy can be consumed. APTEEN: Adaptive Threshold sensitive Energy Efficient sensor Network technology: is an extension to TEEN that is also a hybrid protocol approach but adjusts the threshold values within TEEN based on the changing nature with respect to the application and user interest. APTEEN has a query diffusion system supporting three types of queries: historical, on-time, and persistent queries[43].

5. Discussion

This survey paper deals with the use of Google Scholar for searching many keywords related to data aggregation on a heterogeneous network based on different capabilities, processing, and energy in order to detect new papers, conferences, and articles in the same field.

Article It is worth mentioning doesn't contains comments, book surveys ,reports, monitoring editorial, and the non-English languages materials that are written in this field environment. This examines area purposes, take place, preferences, and impediments of data conglomeration instruments on networks. based on that, the open polices and future patterns through the information conglomeration components on networks are explored.

These protocols for clustering have emerged as one of the powerful techniques in the management of energy consumption in WSN, though there are areas for improvement along with exciting possibilities for the future.

It is limited by the focus on explicit metrics . Although network lifetime and/or energy efficiency have significant importance, most protocols aim at this without considering other critical issues such as data accuracy, latency, or security. Furthermore, with most of the protocols, it has been assumed that deployment of nodes would be static. Real applications, however, include mobile nodes, which implicitly require dynamic cluster and CH selections. Moreover, the efficacy of applying some protocols decreases due to large-scale deployment because of increased control overhead and complex cluster management. Finally, heuristic optimization techniques have their own set of challenges that are usually in conflict with each other regarding computational cost. A balance between attaining desired objectives and reducing computational burden presents a dilemma.

6. Future Directions

The future development of WSN clustering protocols lies in multi-objective optimizations. This means protocols should be able to design various factors optimally all at once, such as network lifetime, energy consumption, data accuracy, latency, and security. Besides, the protocols must adapt to dynamic topologies owing to mobile sensor nodes.

This calls for effective cluster reconfiguration mechanism and techniques for CH selection. Scalability and distributing clustering is also an important feature in dealing with deployments on a large scale. This may involve techniques in the formation of distributed clusters and hierarchical clustering structures. Finally, heuristic optimization coupled with machine learning methods might drive much better performance both in cluster formation and CH selection. This will enhance the adaptiveness , thereby rendering WSN clustering protocols more robust.

6. Conclusion

In summary, improving network performance and extending network lifetime requires the use of clustering protocols to reduce energy consumption in heterogeneous wireless sensor networks (HWSNs). To address energy consumption issues in HWSNs, several algorithms and protocols have been proposed.

Longer choice Set head is an essential factor to improve HWSN performance because it has a significant impact on data collection efficiency and network lifetime.

To address these limitations, can proposed a new approach based on the Hierarchical Ultra Energy Clustering Protocol . This approach includes features of established protocols such as SEP, TSEP and DEEC. The techniques integrates super nodes or super ultra-node to improve the network hierarchy and adjust the selection of cluster heads based on energies remaining to improve energy consumption .

The development process involved modifying the existing protocols or hybridization to create novel algorithm to introduce adaptation of cluster head election mechanism. This resulted in a foundational script for this mechanism, which can be further refined and tested in specific simulation environments.

These findings suggest that significant gains in energy efficiency and network stability for HWSNs can be achieved by means of creative algorithm development and optimization techniques. In order to protect networks from security threats, future research should give special attention to how mobile nodes affect cluster head selection mechanisms and how to integrate security measures.

To further improve the performance of heterogeneous wireless sensor networks, clustering algorithms for isomorphic sensor networks with energy replenishment and energy consumption optimization must be developed.

References

- [1] A. S. Abdalkafor and S. A. Aliesawi, "Applying of (SOM, HAC, and RBF) algorithms for data aggregation in wireless sensors networks," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 354-363, 2022.
- [2] S. Kumar, S. K. Verma, and A. Kumar, "Enhanced threshold sensitive stable election protocol for heterogeneous wireless sensor network," *wireless personal communications*, vol. 85, no. 4, pp. 2643-2656, 2015.
- [3] I. D. I. Saeedi and A. K. M. Al-Qurabat, "A systematic review of data aggregation techniques in wireless sensor networks," in *Journal of Physics: Conference Series*, 2021, vol. 1818, no. 1: IOP Publishing, p. 012194.
- [4] M. S. Abood, H. Wang, H. F. Mahdi, M. M. Hamdi, and A. S. Abdullah, "Review on secure data aggregation in Wireless Sensor Networks," in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1076, no. 1: IOP Publishing, p. 012053.
- [5] A. Rathee, R. Singh, and A. Nandini, "Wireless sensor network-challenges and possibilities," *International Journal of Computer Applications*, vol. 140, no. 2, pp. 1-15, 2016.
- [6] R. S. Rathore, S. Sangwan, S. Prakash, K. Adhikari, R. Kharel, and Y. Cao, "Hybrid WGWO: whale grey wolf optimization-based novel energy-efficient clustering for EH-WSNs," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, pp. 1-28, 2020.
- [7] A. S. Abdalkafor and S. A. Aliesawi, "Data aggregation techniques in wireless sensors networks (WSNs): Taxonomy and an accurate literature survey," in *AIP Conference Proceedings*, 2022, vol. 2400, no. 1: AIP Publishing.
- [8] A. S. Abdalkafor and S. A. Aliesawi, "Efficient data aggregation strategy in wireless sensor networks: challenges and significant applications," in *Proceedings of International Conference on Computing and Communication Networks: ICCCN 2021*, 2022: Springer, pp. 131-139.
- [9] A. S. Abdalkafor and S. A. Aliesawi, "The impact of data aggregation strategy on a performance of wireless sensor networks (WSNs)," in *2021 14th International Conference on Developments in eSystems Engineering (DeSE)*, 2021: IEEE, pp. 217-222.
- [10] P. Vadicherla and D. Vadlakonda, "Study on energy efficient routing protocols scheme in heterogeneous wireless sensor networks (network & mobility)," *Materials Today: Proceedings*, vol. 47, pp. 4955-4958, 2021.
- [11] J. Xie, B. Zhang, and C. Zhang, "A novel relay node placement and energy efficient routing method for heterogeneous wireless sensor networks," *IEEE Access*, vol. 8, pp. 202439-202444, 2020.
- [12] K. S. Adu-Manu, N. Adam, C. Tapparello, H. Ayatollahi, and W. Heinzelman, "Energy-harvesting wireless sensor networks (EH-WSNs) A review," *ACM Transactions on Sensor Networks (TOSN)*, vol. 14, no. 2, pp. 1-50, 2018.

- [13] A. A. Baradaran and K. Navi, "HQCA-WSN: High-quality clustering algorithm and optimal cluster head selection using fuzzy logic in wireless sensor networks," *Fuzzy Sets and Systems*, vol. 389, pp. 114-144, 2020.
- [14] A. Puviarasu, M. Balaji, R. Thirukkumaran, A. S. Kumar, and M. Premkumar, "Dynamic uneven clustering protocol for efficient energy management in EH-WSNs," *Materials Today: Proceedings*, vol. 57, pp. 2092-2095, 2022.
- [15] Y. Tao, J. Zhang, and L. Yang, "An unequal clustering algorithm for wireless sensor networks based on interval type-2 TSK fuzzy logic theory," *IEEE Access*, vol. 8, pp. 197173-197183, 2020.
- [16] S. Doostali and S. M. Babamir, "An energy efficient cluster head selection approach for performance improvement in network-coding-based wireless sensor networks with multiple sinks," *Computer Communications*, vol. 164, pp. 188-200, 2020.
- [17] T. Luo, J. Xie, B. Zhang, Y. Zhang, C. Li, and J. Zhou, "An improved levy chaotic particle swarm optimization algorithm for energy-efficient cluster routing scheme in industrial wireless sensor networks," *Expert Systems with Applications*, p. 122780, 2023.
- [18] Y. Yang, Y. Wu, H. Yuan, M. Khishe, and M. Mohammadi, "Nodes clustering and multi-hop routing protocol optimization using hybrid chimp optimization and hunger games search algorithms for sustainable energy efficient underwater wireless sensor networks," *Sustainable Computing: Informatics and Systems*, vol. 35, p. 100731, 2022.
- [19] W. Fei, B. Hexiang, L. Deyu, and W. Jianjun, "Energy-efficient clustering algorithm in underwater sensor networks based on fuzzy C means and Moth-flame optimization method," *IEEE Access*, vol. 8, pp. 97474-97484, 2020.
- [20] S. Joshi, T. Anithaashri, R. Rastogi, G. Choudhary, and N. Dragoni, "IEDA-HGEO: improved energy efficient with clustering-based data aggregation and transmission protocol for underwater wireless sensor networks," *Energies*, vol. 16, no. 1, p. 353, 2022.
- [21] L. Cai, R. Huang, Z. Li, L. Luo, Z. Xiong, and Y. Chen, "A clustering election game-based and two-level management protocol for wireless sensor networks," *IEEE Internet of Things Journal*, 2023.
- [22] M. Elhoseny, A. E. Hassanien, M. Elhoseny, and A. E. Hassanien, "Hierarchical and clustering WSN models: their requirements for complex applications," *Dynamic Wireless Sensor Networks: New Directions for Smart Technologies*, pp. 53-71, 2019.
- [23] I. D. I. Saeedi and A. K. M. Al-Qurabat, "Perceptually important points-based data aggregation method for wireless sensor networks," *Baghdad Science Journal*, vol. 19, no. 4, pp. 0875-0875, 2022.
- [24] H. Ramezanifar, M. Ghazvini, and M. Shojaei, "A new data aggregation approach for WSNs based on open pits mining," *Wireless Networks*, vol. 27, no. 1, pp. 41-53, 2021.
- [25] M. Kaur and A. Munjal, "Data aggregation algorithms for wireless sensor network: A review," *Ad hoc networks*, vol. 100, p. 102083, 2020.
- [26] B. A. Begum and S. V. Nandury, "Data aggregation protocols for WSN and IoT applications—A comprehensive survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 2, pp. 651-681, 2023.
- [27] S. Kaur, R. N. Mir, A. Khamparia, P. Rani, D. Gupta, and A. Khanna, "Heterogeneous load balancing clustering protocol for Wireless Sensor Networks," *Cognitive Systems Research*, vol. 70, pp. 10-17, 2021.
- [28] J. Bholra, M. Shabaz, G. Dhiman, S. Vimal, P. Subbulakshmi, and S. K. Soni, "Performance evaluation of multilayer clustering network using distributed energy efficient clustering with enhanced threshold protocol," *Wireless Personal Communications*, vol. 126, no. 3, pp. 2175-2189, 2022.
- [29] A. Almasri and K. A. Darabkh, "A comparative analysis for WSNs clustering algorithms," in *2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC)*, 2020: IEEE, pp. 263-269.
- [30] S. Hassan and M. Ahmad, "Energy heterogeneity analysis of heterogeneous clustering protocols," *Int. Arab J. Inf. Technol.*, vol. 19, no. 1, pp. 45-54, 2022.

- [31] H. H. El-Sayed and Z. M. Hashem, "Comparison of the new version of DEEC protocol to extend WSN lifetime," *EURASIP Journal on Wireless Communications and Networking*, vol. 2023, no. 1, p. 56, 2023.
- [32] N. Qubbaj, A. A. Taleb, and W. Salameh, "Review on LEACH protocol," in *2020 11th International Conference on Information and Communication Systems (ICICS)*, 2020: IEEE, pp. 414-419.
- [33] L.-L. Hung, F.-Y. Leu, K.-L. Tsai, and C.-Y. Ko, "Energy-efficient cooperative routing scheme for heterogeneous wireless sensor networks," *IEEE Access*, vol. 8, pp. 56321-56332, 2020.
- [34] P. Konduru and N. Nethravathi, "SECURE and energy-efficient routing protocol based on micro-segmentation and batch authentication," *Computer Networks*, vol. 248, p. 110293, 2024.
- [35] M. Y. Aalsalem, "An effective hotspot mitigation system for Wireless Sensor Networks using hybridized prairie dog with Genetic Algorithm," *Plos one*, vol. 19, no. 4, p. e0298756, 2024.
- [36] R. Ramteke, S. Singh, and A. Malik, "Optimized routing technique for IoT enabled software-defined heterogeneous WSNs using genetic mutation based PSO," *Computer Standards & Interfaces*, vol. 79, p. 103548, 2022.
- [37] M. S. Khan and M. Daneshmand, "I-SEP: An Improved Routing Protocol for Heterogeneous WSN for IoT based Environmental Monitoring."
- [38] M. Swapna and G. Satyavathy, "Energy-aware optimal clustering and secure routing protocol for heterogeneous wireless sensor network," *International Journal of Computer Networks and Applications*, vol. 9, no. 1, pp. 12-21, 2022.
- [39] C. Guo, H. Tang, and B. Niu, "Evolutionary state-based novel multi-objective periodic bacterial foraging optimization algorithm for data clustering," *Expert Systems*, vol. 39, no. 1, p. e12812, 2022.
- [40] N. D. Tan and H.-N. Hoang, "Energy-Efficient Distributed Cluster-Tree Based Routing Protocol for Applications IoT-Based WSN," in *RICE*, 2022, pp. 213-218.
- [41] M. Bilal, E. U. Munir, and F. K. Alarfaj, "Hybrid clustering and routing algorithm with threshold-based data collection for heterogeneous wireless sensor networks," *Sensors*, vol. 22, no. 15, p. 5471, 2022.
- [42] A. Krasnov, M. Y. Prakhova, and U. Novikova, "Mathematical simulating qualitative parameters of routing and clustering protocols in wireless data gathering networks," in *2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, 2020: IEEE, pp. 1-7.
- [43] Z. u. A. Jaffri *et al.*, "TEZEM: A new energy-efficient routing protocol for next-generation wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 18, no. 6, p. 15501329221107246, 2022.