International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 13, No.1 (Jan-2023)

http://dx.doi.org/10.12785/ijcds/130101

Data Aggregation Mechanisms in Wireless Sensor Networks of IoT: A Survey

Suha Abdulhussein Abdulzahra[1] **and Ali Kadhum M. Al-Qurabat**[2]

[1]*Department of Dentistry, Al-Mustaqbal University College, Babylon, Iraq*
[2]*Department of Computer Science, College of Science for Women, University of Babylon, Babylon, Iraq*

Received 19 Apr. 2021, Revised 12 Dec. 2022, Accepted 12 Jan. 2023, Published 31 Jan. 2023

Abstract: In numerous Internet of Things contexts, there is an increasing interest to use wireless sensor technologies. One of the most difficult problems is gathering and analyzing commodity data, given the enormous rise of smart objects and their applications. Sensor nodes are battery-powered, and energy-efficient operations are important. To that end, before transmitting the final data to the central station, remove redundancy from the collected data by neighbouring nodes is beneficial for sensors. Data aggregation is one of the main strategies for reducing data redundancy and improving energy efficiency; it also extends the lifetime of wireless sensor networks. Moreover, network traffic can be minimized by an efficient data aggregation protocol. It may be sensed by more than one sensor when a particular target takes place in a particular area. This article provides an overview of different data aggregation methods and protocols, taking into account the key problems and facets of data aggregation in wireless sensor networks. The structures of data aggregation are grouped into four key classes, namely cluster-based, tree-based, chain-based and grid-based. The thorough comparison of the important approaches of each class often gives a suggestion for more research.

Keywords: Data Aggregation, Energy Consumption, IoT, WSN

1. Introduction and Overview

A vast number of interconnected nodes, from massive network servers and supercomputer clusters to tiny sensor nodes and handheld computers, would be part of the future of the Internet. The Internet of Things (IoT) is the name given to this kind of global network [1], [2]. Wireless sensor networks (WSNs) are amongst the utmost essential contributors to the IoT and play a significant role in people's lives due to their extensive usage in a variety of applications including military, manufacturing, environment, transportation, healthcare, and agriculture [3], [4], [5]. This form of the network has evolved as a result of rapid advancements in wireless networks, electronics, and embedded computing technologies [6].

WSN is made up of a vast number of low-cost, low-energy sensors that are spatially distributed to detect physical or environmental events in a specific region, such as motion, strain, vibration, noise, humidity, temperature, etc. [5]. With minimal battery capacity, storage, computing, and bandwidth capacities, the sensor node can detect, process, store and communicate [6], [7]. These nodes gather information from the tracked area, manipulate it locally, and send it to one or more aggregation points known as a base station or sink (i.e., a point that serves as an interface between users and sensor nodes) for further analysis [8].

The sensor node's battery capacity is the most valuable resource because it impacts the WSN's lifetime. The key function of the power unit (battery) is to provide the sensor node with the necessary energy to accomplish its mission. As the battery life of the sensor node is small, it's difficult, if not impossible, to substitute (or replenish), particularly in a hostile or remote location [9]. The sensor nodes are then installed at high density to take advantage of the spatial similarity between the sensor devices to increase the WSN's network lifetime. In addition, it is necessary to reduce the overall communication cost because it is the most consuming element inside the sensor node. Hence, the maximizing of battery life is one of WSN's greatest problems [10].

Inside the sensor node, the radio system is the primary provenance of energy utilization. Therefore, due to the sensor node battery's small power, the redundancy in sensor information must be deleted to send a little data volume to the sink for the sake of energy conservation and extend the sensor node lifetime [11]. The data collected by the

E-mail address: Suha.Abd@mustaqbal-college.edu.iq, alik.m.alqurabat@uobabylon.edu.iq

node of the sink could be resembling since many sensors are monitoring the same region. Furthermore, dense WSN produces a huge quantity of sensed data that contributes to numerous undesirable problems such as energy consumption, network congestion, data quality deterioration and high load processing on the sink node [3].

Therefore, in order to operate for an extended interval of time, the energy efficiency of the sensor network needs to be increased. This can be accomplished by capturing and processing sensed sensor node data in an energy-efficient manner [12]. Several energy-saving techniques are used in the literature to reduce connectivity costs (in terms of energy) for sensor networks, such as data radio optimization, battery repletion, clustering, routing, scheduling, predictive monitoring, aggregation, and compression [13]. The techniques mentioned above, though, assumed that the sensing and processing of data require very little energy which is negligible compared to the energy spent on the communication of data [11].

Some of the proposed works recommend adopting the reduction of data techniques [14] for lessening data amount collected and transferred while ignoring data precision. They mitigate the volume of data transmitted via every sensor node by utilizing their computational resources to carry out simple computations internally and transmit only required and partially processed data. Energy efficiency and data reduction necessitate adaptive sampling for periodic data gathering and energy-efficient aggregation of data [15]. As the sensor node adjusts its proportion of sampling depending on the adaptive alteration of the tracked phenomenon, energy consumption may be decreased. The predominant aim of data aggregation algorithms, which are effectively implemented to WSN, is to capture and aggregate sensor readings by eliminating duplication from sensed associated data generated by adjacent neighboring sensors, thereby conserving energy and expanding the network's lifetime [1], [12]. This paper's contribution can be summed up as follows:

- Going to review and analyze existing data aggregation methods in WSN of IoT in a systematic manner.
- Going to divide the methods of data aggregation into two types, Hierarchical and Flat Networks.

The rest of this paper is structured as follows: A general outline of data aggregation is provided in Section 2. Section 3 looked at some previous research on data aggregation in WSNs. The numerous data aggregating protocols are defined in Section 4 based on the network architectures involved in the aggregation process. The comparison of data aggregation methods was outlined in Section 5. The data aggregation metrics were discussed in Section 6, and the concluding remarks were presented in Section 7.

2. An Overview of Data Aggregation

Data gathering is the method of taking sensed measurements systematically from a large number of sensor nodes and sending them to a sink for further analysis [16]. Data transmission is a costly process for sensors, in contrast, in-network computations are less expensive in terms of energy consumption and are sometimes considered trivial and negligible [17]. However, the computation needs much less power than the energy used to transmit the data as illustrated in Figure 1. For example, the energy spent to transfer a packet of data of 1 kb and for a distance of 100 meters is roughly equivalent to the implementation of approximately 3 million instructions using a typical microprocessor. Thus, any additional processing that will reduce the size of the data even though one data bit would be useful in terms of energy efficiency. This radical variation amongst transmission and computation indicates the significance of local data processing in decreasing energy consumption in the network of sensors [16].

Neighbouring sensors often generate highly correlated and redundant data [15] the reason for this is that the sensors that are deployed in nearby areas are usually sensing the same phenomena, which would generate many duplicate data. The bandwidth and energy consumed as a result of this duplication in data will be too large [18]. Also, the volume of data produced in large networks of sensors will be huge to be handled by the station [15].

Because sensor nodes have limited energy, it is wasteful for every one of the nodes to send their data directly to the sink, in addition, the reduction of energy consumption in every sensor node has become necessary in order to improve the lifetime of WSN [18]. Consequently, we want strategies for removing redundancy and consolidating data in order to produce information with high quality at the sensor nodes which will reduce the number of packets that will be sent to the sink and thus lead to conserving both energy and bandwidth. In order to achieve this, aggregation of data may be utilized [15]. Data aggregation is described as *"an approach of collecting and accumulating sensor readings by eliminating redundancy from sensed correlated data provided by closely located neighboring sensors in order to reduce energy consumption and prolong the network lifetime"* [1], [2].

Data aggregation is amongst the important energy-saving analysis processes. In the WSN, aggregation is an effective method for conserving finite resources. In this manner, the use of the data aggregation technique will decrease both: the transmissions number and the usage of bandwidth in the network, also, removing the needless consumption of energy in the two processes of receptions and transmissions [19]. Figure 2 shows the work of data aggregation algorithms in general [20].

3. Related Works

This subsection looked at several current related work on data aggregation in WSNs. Many suggested data manage-

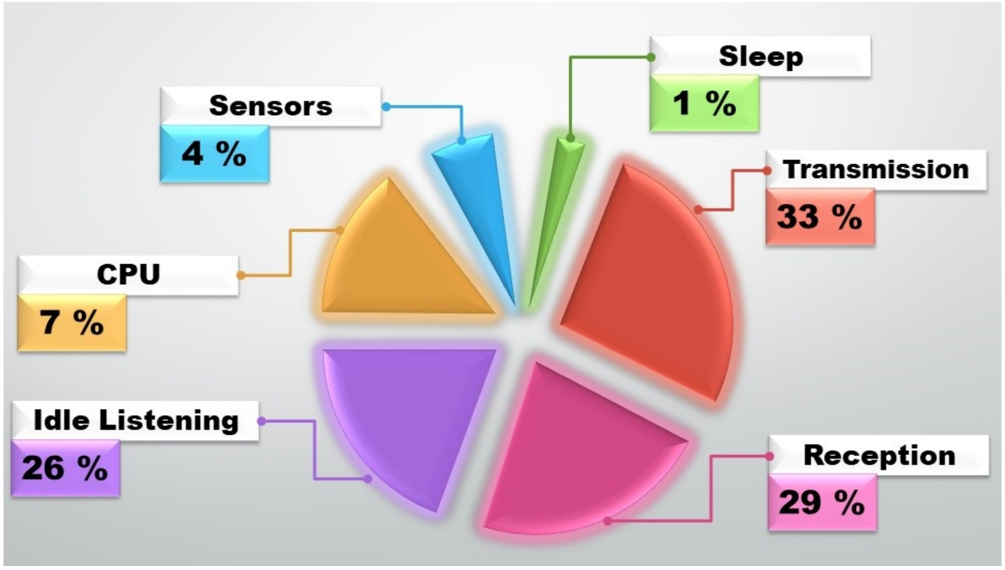

Figure 1. Power consumption of node to receive or transmit messages [17].

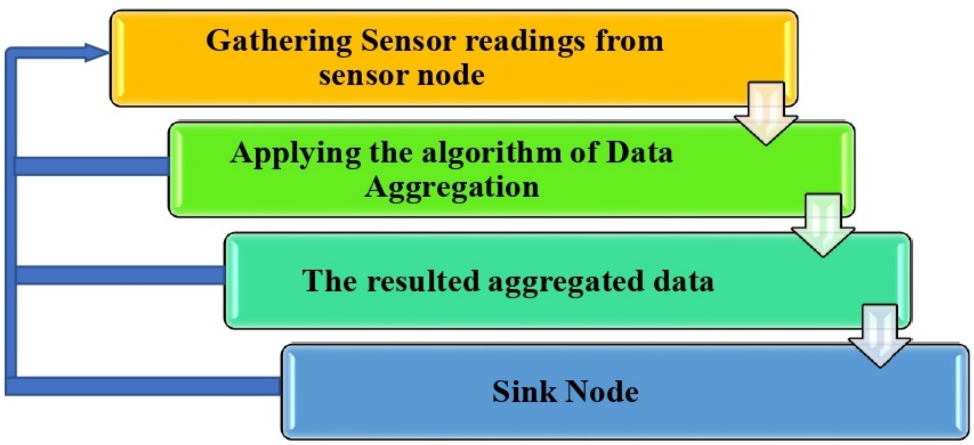

Figure 2. General structure of the data aggregation algorithm [20].

ment projects in WSNs have centered on aggregating data in recent years [1], [2], [3]. The predominant aim of data aggregation algorithms is eliminating duplication from sensed associated data generated by adjacent neighboring sensors, thereby conserving energy and expanding the network's lifetime [10]. However, the works in [21], [22], [23], [24] use basic aggregation methods (such as AVG, MIN, MAX, SUM) for aggregation in order to minimize the transition of data. The association between sensed information is not taken into account in these techniques. While they have a high output of aggregation, the accuracy of the data retrieved is badly poor. Thus, for those systems that need high data precision, these approaches are unacceptable. The LEACH [25] protocol, for instance, splits the network into many clusters. During the setup process, the cluster heads are chosen while the information is aggregated in the steady phase using the AVG approach at each cluster head in order to minimize the traffic of network data.

The work in [26], [27], [28] divides the nodes of the sensor into distinct clusters. One or more nodes for every cluster are chosen as a designated set of nodes for gathering and sending data, the other nodes in the cluster, on the other hand, are turned off. With these approaches, the node's resources can be saved substantially, but due to a huge amount of deactivated nodes, this may result in substantial data loss. The authors of [29] proposed a clustering based on round method for resolving the transmitting of redundant network information in order to extend the network's lifespan. In four-step rounds, the proposed method works: initialization, selection of cluster heads, clustering, and aggregation of results. The proposed clustering scheme decreases energy consumption by resolving much of the redundant data, thereby increasing network throughput.

In [30], [31], the authors use a technique called Prefix-Frequency Filtering (PFF) [30] to explore the issue of filtering collection. In this approach, the aggregation is carried out at two tiers: the tier of the sensor utilizing the processing of data locally and the tier of the aggregator utilizing the Jaccard-similarity PFF method to aggregate related data from the near-adjacent nodes into a single record. Several changes to the PFF system for avoiding duplication from the transmitted data to the sink and reducing data latency [31].

In [1], [2], [3], [4], [5], [6] the authors proposed data aggregation techniques work at two levels of the networks. They used different methods like SAX, Adaptive piecewise Constant Approximation, Hashing table. These approaches are designed to eliminate redundant data at sensor nodes and CH in order to extend the WSN's lifespan.

Table I provides some related works and explains the main features of each research briefly.

4. Data Aggregation Techniques

One of the things that have an effective role in the implementation of various data aggregation methods is sensor network architecture [15]. Data aggregation a fundamental model for WSNs which aims to accumulate the data incoming from various sources. Among other things that data aggregation methods can achieve are the removal of redundancies, reducing the number of transmissions, and hence, leading to reduced energy exhaustion and provision. The aim of data aggregation methods is to decrease the required communication between different levels in order to decrease total energy expenditure. When the energy consumed for the aggregation is less than the energy used to transporting original data to a higher level, the data aggregation achieved its goal of saving energy [6]. Figure 3 represents the various data aggregation techniques available in WSNs. In the literature, there are different data aggregation algorithms that have been proposed some of them are structured and the others are unstructured which aim to aggregate data in an efficient way without increasing the size of the message.

A. Flat Networks

The role played by every sensor node in flat networks (see Figure 4) is equal and has approximately the same amount of power in the battery [15]. The process of sensing in this network is carried out through the cooperative work of the sensor nodes [19]. In flat networks, data aggregation depends on data-driven routing and a query-based approach where the sink typically transfers a message of the query to the nodes, for instance, through the use of floods and a response messages will transmit to the sink from the nodes whose data matching the query [15].

In addition, these networks use a multihop path, via this path the data is distributed that aiming to aggregate them.

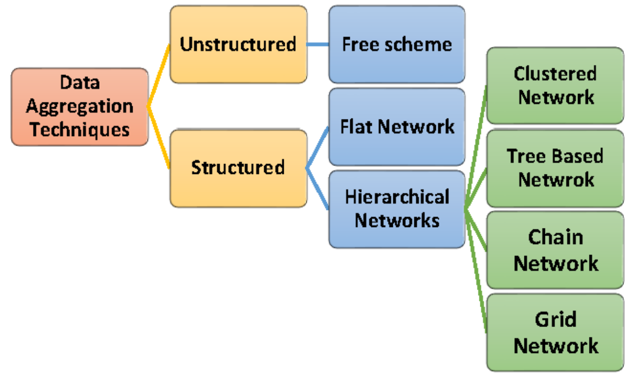

Figure 3. Data aggregation techniques.

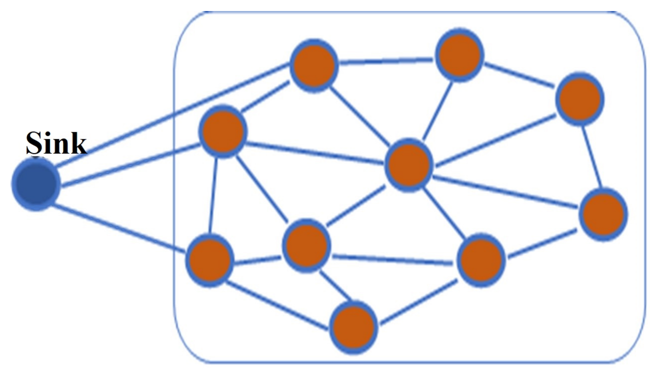

Figure 4. Flat based network.

This topology will increase highly the level of latency in the network. Some examples of routing protocols that carry out the aggregation of data depending upon flat are "flooding", "Sensor Protocols for Information via Negotiation (SPIN)" [32], "Directed Diffusion (DD)" [33], "Rumor Routing (RR)" [34], etc.

In this type of network, no complex switches or routers are utilized, so we do not need much effort to maintain and manage the network. This is a positive advantage for flat-based networks. Also, the quality of the routes between the source and the sink is good in these protocols. Even worse, since flooding is an expensive process, it disregards the conditions imposed by sensor networks. This type of network has disadvantages, including the process of route scheduling, data is aggregated in a single region (sink node) to which the data was transmitted for the sole purpose of aggregating data based solely on competition. As a result, the computation overhead will be greatly increased in the sink, causing to depletion of the energy faster. One of the things that happen as a result of a failure sink is the collapse of the whole network which leads to increased overhead [19].

TABLE I. The Summarization of Literature.

Reference	Authors	Topology	Approach	Objective
[1]	Idress and Al-Qurabat (2020)	Cluster	Fog computation focused on EDaTAP (Energy-efficient Data Transmission and Aggregation Protocol)	To remove redundant related sets of data in sensor devices and fog gateways, use Dynamic Time Warping (DTW) with basic encoding.
[2]	Al-Qurabat and Idress (2020)	Cluster	For optimizing lifetime in PSNs of IoT applications, data gathering and aggregation with selective transmission (DGAST) technique was used.	DGAST functions in rounds. Every round requires two periods, Data collection, data aggregation, selective distribution, and adaptive sampling are the four stages in each period.
[3]	Al-Qurabat and Idress (2019)	Cluster	A protocol of Two-Stage Data Aggregation (TLDA) to extend the life of periodic sensor networks.	There are two levels of data aggregation. The first step includes data collection, the sliding window, and data aggregation using the Adaptive Piecewise Constant Approximation technique. The SAX quantization method based on the chaining hash table is used to group the obtained data sets in the second step.
[5]	Idress et al. (2019)	Cluster	An Improved K-means along with Divide and Conquer approach for energy-efficient data aggregation in WSNs (IDiCoEK).	The measurements are aggregated by IDiCoEK into two groups: node and cluster head levels. At the sensor node, an algorithm for divide and conquer is implemented. The cluster head applies an improved K-means.
[35]	Sarangi and Bhattacharya (2019)	Cluster	Dynamic approach using rendezvous point (RP).	Cluster creation based on the Neural network and optimization based on Ant Colony.
[36]	Yadav and Yadav (2019)	Flat	A method for removing duplication during data aggregation.	To define and eradicate data duplication, the aggregator employs a linear classifier based on SVM.
[37]	Al-Qurabat and Idress (2018)	Cluster	Energy-efficient method of adaptive distributed data processing (EADiDaC).	Delete the redundant data gathered and adjust the sampling rate in compliance with the environmental factors monitored.
[38]	Harb et al. (2017)	Cluster	A powerful data transfer protocol that consists of two data aggregation steps.	Effective two-phase data aggregation, in the first phase, is the similarity between the measurements taken by the sensor used. It uses a distance-based function in the second stage to identify correlations between sets of data obtained.
[39]	Atoui et al. (2016)	Flat	Local aggregation and Prefix filtering.	Two stages: first stage filtering data based on the fitting function; if the data's norm value falls below a certain amount, nothing is sent to the aggregator to be aggregated.
[40]	Karim and Al-kahtani (2016)	Cluster, Tree	Multilayer big data aggregation framework.	A dynamic data aggregation method based on priority proposed for aggregating big data at three levels.
[41]	Makhoul et al. (2015)	Cluster	ANOVA model-based adaptive data collection method.	Combine the ANOVA model with the residual energy for permitting to adjust the proportion of sampling per sensor node in response to varying environmental conditions.

TABLE I – *Continued on next page*

TABLE I – Continued from previous page

[42]	Song et al. (2015)	Cluster	Periodicity-and Linear-Based data suppression mechanism (PLB).	Based on the probability of presentation, PLB will choose the most likely linear trend to forecast future results, which increases the success rate of prediction and allows for more predictable data.
[43]	Srbinovski et al. (2015)	Structure-free	Energy-saving algorithm of adaptive data collection.	It collects energy from the sensing area being tracked and changes the sampling proportion based on the residual energy and the observed environment.
[44]	Zhang et al. (2015)	Cluster	Centered on an endocrine regulation technique, an algorithm of adaptive sampling has been developed.	Hormone data is used to monitor whether nodes are in a working or resting state, and to dynamically change gathering frequency.
[29]	Tran and Seung (2014)	Cluster	Clustering scheme and aggregation based on similarity.	The approach improves network lifespan by resolving redundant data transfer in the network.
[30]	Bahi et al. (2014)	Cluster	Adaptive sampling rate based on similarity.	Utilizing prefix frequency filtering (PFF) to remove the duplication of data collected.
[21]	Ren et al. (2013)	Cluster	Data aggregation based on spatial-temporal correlations.	In order to accumulate and relay sensed data to the sink as they reach a threshold, nodes are spatially clustered and temporal correlation is accomplished by the CH.
[45]	Masoum et al. (2013)	Cluster	Adaptive sampling focused on spatio-temporal similarity.	To evaluate the potential sensors responsible for measuring and transmitting, It makes use of spatio-temporal similarity between sensors and the data they detect.
[46]	Yang et al. (2011)	Tree	The order relationship of compound sensor data is used to compress data.	Does an in-network aggregation of data base on the order compression techniques. Based on history, data may be suppressed.
[47]	Alippi et al. (2010)	Cluster	Power aware adaptive sampling method.	The algorithm uses a fast Fourier transform to provide online estimation.
[48]	Law et al. (2009)	Cluster	The Box-Jenkins technique is used in the algorithm of adaptive sampling.	Based on the current sensor readings, forecast potential sensor readings.
[49]	Chatterjea and Havinga (2008)	Cluster	Temporal correlation of sensed data using a sampling algorithm.	The proportion of sampling is determined by the reliability of the controlled environment; when the conditions are unpredictable, the rate increases; otherwise, the rate decreases.
[25]	Heinzelman et al. (2000)	Cluster	Data aggregation using AVG method.	Reduce the network data traffic via aggregating data using average

B. Hierarchical Networks

Some of the negative things that happen in flat networks are the increased communication and the burden of computation in the sink node which in turn leads to the drain of its battery power very quickly. The loss of the sink node to its energy (i.e. death) will result in the breaks down of the network functions completely. Consequently, in the perspective of scalability and the efficiency of energy, many data aggregation methods based on hierarchical topology have been proposed. In hierarchical data aggregation, there is a special node for data combining, which decreases the quantity of transferred messages to the sink. This enhances the efficiency of energy in the network [15]. Hierarchical data aggregation can be additionally partitioned into four types: cluster, chain, grid and tree.

1) Cluster-Based Approach

In the large size WSNs with limited energy, it is wasteful for every one of the nodes to forward directly their data to the sink, in addition, one of the things necessary to improve WSN lifetime is to reduce the consumption of energy per sensor node [18], [19]. Consequently, using efficient ways for reducing the consumption of energy is necessary like clustering.

Clustering is an essential energy-saving mechanism in WSNs. To support the energy-efficient aggregation of WSN data, the WSN can be divided into many clusters. The key component of the cluster is a group of adjacent nodes that collaborate to elect one cluster head. For inter-cluster cooperation and intra-cluster coordination, the cluster head is responsible. The nodes in each cluster collect and aggregate data from the monitored location, which is then sent to the cluster head via multi-hop or one-hop communication. As a consequence, the volume of data sent is decreased dramatically. Furthermore, nodes relay their sensed data within clusters rather than sending it directly to the sink node [18], [19]. Scalability, low complexity, reliability, load balancing, bandwidth performance improvement, energy-saving, and latency reduction, and data replication removal are all advantages of cluster-based approaches.

Figure 5 displays a cluster-based WSN. Some examples for the protocols that designed based on clusters are: “Low Energy Adaptive Clustering Hierarchy (LEACH)” [50], [51], “Hybrid Energy Efficient Distributed Clustering Approach (HEED)” [52], “Clustered AGgregation Technique (CAG)” [53], and “Threshold sensitive Energy Efficient sensor Network protocol (TEEN)” [54].

2) Chain-Based Approach

In cluster-based WSNs, the data is sent from the nodes to the CH, which aggregates it. Hence, if the location of CH is remote from the nodes, they may exhaust for communication more energy. Moreover, enhancements in the efficiency of energy can be acquired if sensor nodes send just to close neighbours [15].

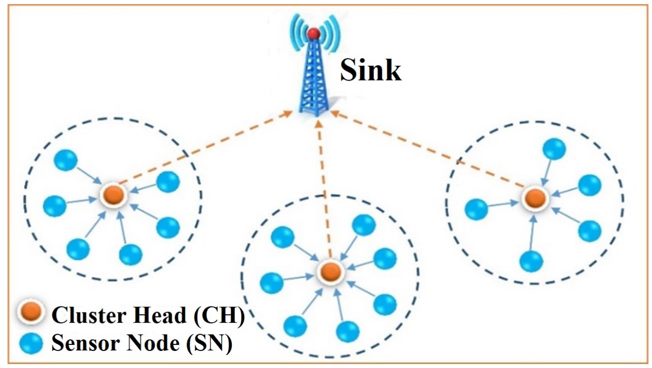

Figure 5. The technique of data aggregation based on clustering.

Chain-based data aggregation represents a kind of hierarchical method that forms the chain structure. The energy in this method is distributed evenly and the communication between every sensor node and its neighbours is possible and every node takes its role as a leader for sending data to the sink. The leader is chosen based on the token passing method. The data will be transferred to the aggregation node after the token is received, and it will finally enter the sink node. The basic idea for the chain-based data aggregation method is that the data is sent between nodes close to each other (i.e., neighbours) only [19].

Some examples for the protocols that designed based on the chain are: “Power-Efficient Gathering in Sensor Information Systems (PEGASIS)” is a famous routing protocol in a chain topology. “Chain Oriented Sensor Network for Efficient Data Collection (COSEN)”, and “Chain-Based Hierarchical Routing Protocol (CHIRON)” are another example of protocols based on chain [19]. Figure 6 displays the chain-based data aggregation technique.

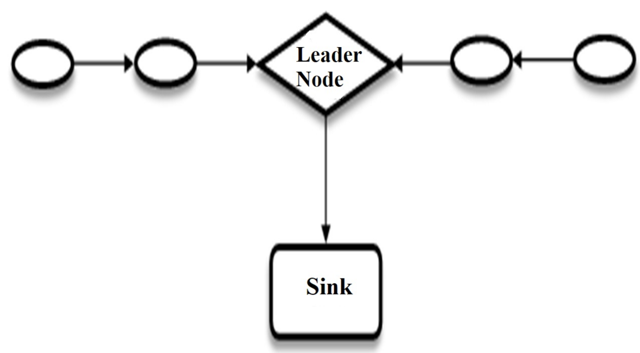

Figure 6. The technique of data aggregation based on the Chain.

3) Grid-Based Approach

In data aggregation based on grid, the area to be observed is divided into a grid. A collection of sensor nodes is employed as an aggregator for data in places previously defined in sensor networks. In this network there is no communication among the sensor nodes with each other, and every sensor in a specific grid submits data to the aggregator that belongs to that grid [55]. In this network, the role of aggregator could be played by any sensor node belongs to that grid on the basis of rounds until the last sensor dies. This type of networks is appropriate for military surveillance and weather forecasting. Some examples for the protocols that designed based on chain are: GROUP and ATCBG [56]. Figure 7 displays the technique of aggregation of data based on the grid.

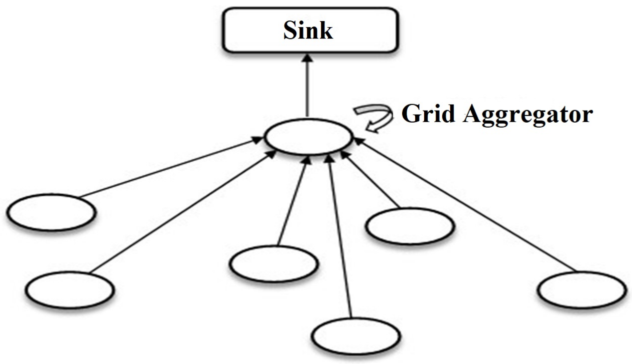

Figure 7. The technique of data aggregation based on the grid.

4) Tree-Based Approach

A tree with the sink as its root shapes the sensor nodes in the tree-based network. The data aggregation procedure is carried out in all mediate nodes around the tree, and a compact type of data is sent to the sink (root) [56]. This kind of networks can be used in applications that include data aggregation in-network which aim to consumption the minimal energy of the network. As an example of applications using this type of network is monitoring the level of radiation in the plant of nuclear in which the greatest importance provides the most useful facts for plant safeness. One of the key advantages of tree-based networks is the design of an energy-efficient tree for aggregation data [15]. Some examples for the protocols that designed based on tree are: "Tree on DAG (ToD)", "Energy aware SPANning tree (ESPAN)", and "Energy-Aware Data Aggregation Tree (EADAT)".

One of the uncomfortable things about this configuration is the high cost and time required for constructing the tree, besides, the dies of one parent node will lead to the die of the entire subtree belonging to it. Sensor nodes in the leaf level have the ability for energy saving more than nodes closest to the root (sink) which consumes more energy in order to passing the received packets from its different children in the sub-tree. Figure 8 displays the tree-based data aggregation technique. However, the load of data packets in some nodes in the tree of aggregation is unfair to the nodes of transmission, this can cause overcrowding in certain nodes, resulting in network failure [55], [56].

C. Structure-less Approach

In data aggregation based on structure-less, there is no need to maintain any structure. This type of networks is very

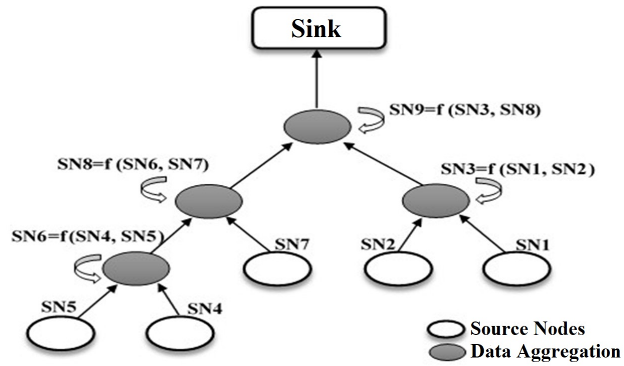

Figure 8. Tree-based data aggregation technique.

appropriate in applications dependent on events because the area of event is usually changing much of the time. The failure of any node will not effect on the network and the reconstruct of the structure is not necessary. However, in this architecture the decisions of routing require to be made on the fly for doing data aggregation. Also, sensor nodes disregard their upper nodes and forwarding their data without waiting for the reception of data from another node. Data-aware anycast (DAA) is considered as a first structure less-based data aggregation method has been proposed, which is a MAC layer protocol for spatial convergence [18], [19]. Other examples on structure-less-based data aggregation method are: "RDAG" and "Structure-Free and Energy-Balance data aggregation protocol (SFEB)" [55].

5. Comparing The Techniques For Data Aggregation

Based on certain typical characteristics for solving aggregation issues, data aggregation strategies in WSN can be compared. Any of the common and important features listed in Table II that should be investigated for every data aggregation method are the methods of data aggregation, the protocols of routing, subtype, objective function, the parameters of optimizing, advantages, disadvantages/open issues and environment.

Owing to the different methods of data aggregation techniques and the lack of criteria, comparative study of techniques of aggregating data is challenging.

Therefore, to determine the successful data aggregation technique, it's crucial to compare data aggregation strategies. The comparison of data aggregation strategies based on hierarchical topology is seen in Table III.

6. Data Aggregation Metrics

The process of designing effective algorithms for data aggregation is not an easy but challenging task. To test the efficacy of algorithms for data aggregation, there are several important performance factors that can be used to calculate algorithm efficiency [57] as follows:

A. Network Lifetime

One of the most important metrics is the lifetime of network which is defined as the number of data aggregation

TABLE II. DATA AGGREGATION MECHANISMS CHARACTERISTICS.

Characteristic	Description
Sub type	Techniques for aggregating data are furthermore categorized into different sub-types.
Routing protocol	In various Data Aggregation Methods, various kinds of routing protocols are utilized, according to the literature. From current literature, routing protocols defined are distributed, cluster-based, tree-based, chain-based and hierarchical.
Objective function	Each data aggregation strategy has an objective feature explicitly tailored for a particular purpose of the mechanism. Reducing latency and energy usage, for example. In fact, the mechanism's key objective feature is to effectively accumulate data by maximizing output parameters.
Optimizing parameter	Any technique of data aggregation has its output parameters specially tailored for the precise purpose of the comparison of outcomes. For example, time, cost, energy, the lifespan of the network, precision of data, etc.
Advantages	This section explains the benefits of data aggregation strategies.
Operational environment	The methodology of data aggregation can be adopted and executed in an operating environment. Different types of methods used in current measurement data aggregation strategies are discussed.
Disadvantages /open issues	This part looks at the drawbacks or other unanswered problems of the technique of data aggregation.

cycles up to the death of $\alpha\%$ of the sensor nodes, where α is determined by the designer of system. For example, if the application impose that all the sensor nodes must work together at the same time which is a vital thing, in this case the definition of lifetime it will be the number of cycles till the first sensor node deplete its energy. The basic idea is that there should be a uniform consumption of energy for the implementation of data aggregation in the network [15]. Network lifetime is expressed by Equation 1

$$NL_n^n = \min_{v \in V}(NL_v) \tag{1}$$

Here, NL_n^n is the lifetime of the network, NL_v is the v node's lifespan, and the node-set without the node of sink is called V.

B. Energy Efficiency

In each data collection round, each sensor can consume the same amount of energy in an ideal situation, but in actual circumstances, for data processing, sensor nodes use varying amounts of resources. In WSNs, a data aggregation approach is cost-effective if it provides optimal functionality with reduced energy consumption. Energy efficiency is expressed by Equation 2

$$Energy\ Efficiency_i = \sum_{i=1}^{n}\left(\frac{Sent_{Data}}{Total_{Energy}}\right) \tag{2}$$

Where, n is the number of sensors nodes in a sensor network, $Sent_{Data}$ is the amount of data successfully sent in a sensor network and $Total_{Energy}$ is the total energy consumed to transfer those data.

C. Data Accuracy

The data accuracy definition is varying from application to another and relies upon the design of sensor network for a particular application. For example, in an application that tackle the problem of target localization, the data accuracy is determined through estimating the location of target at the sink node. As seen in Equation 3, data accuracy is described as the proportion of successfully transmitted data to total data sent [18, 19].

$$Data\ Accuracy = \frac{Amount\ of\ data\ sent\ successfully}{Total\ amount\ of\ data\ sent} \tag{3}$$

D. Latency

The definition of latency is the delay engaged in the transmission of data, routing and the aggregation of data. Latency may be considered as the delay in time between the process of generating data packet at the sensor nodes and the reception of data packets by the sink [67]. Latency is calculated using is Equation 4 as follows:

$$Latency_i = \sum_{i=1}^{n}(Time_{RX} - Time_{TX}) \tag{4}$$

Where, in a sensor network, n indicates the number of sensor nodes, $Time_{RX}$ is the time of receiving data and $Time_{TX}$ is the time of sending data.

TABLE III. SUMMARIZATION OF PROTOCOLS OF DATA AGGREGATION BASED HIERARCHICAL TOPOLOGY.

Protocol	**Architecture**					**Data delivery model**			**Computation**		**Scalability**			**Application type**							**Overhead**			**Disadvantage**
														Monitoring										
	Flat	**Cluster**	**Chain**	**Grid**	**Tree**	**Event**	**Periodic**	**Query**	**Decentralized**	**Centralized**	**Limited**	**Average**	**Good**	**Habitat**	**Environment**	**Health**	**Home/Office**	**Disaster**	**Battle field**	**Civil**	**Low**	**Average**	**High**	
SPIN	×					×	×		×		×			×							×			All nodes within transmission range of that node must pay a price for that transmission.
DD	×							×	×		×				×						×			Latency is slightly higher
RR	×							×	×				×	×	×						×			
LEACH		×				×				×			×			×							×	Cluster formation is costly.
HEED		×				×				×		×			×									System have limitations imposed due to using some parameters in cluster choosing.
CAG		×				×				×		×		×									×	preservation data centric
TEEN		×				×				×			×				×						×	Does not appropriate for PSN.
PEGASIS			×			×				×		×						×			×			Delay of communication is high because signal chain is long.
COSEN			×			×				×									×					Lot of redundant transmission paths

TABLE III – *Continued on next page*

TABLE III – Continued from previous page

Protocol	Architecture					Data delivery model			Computation		Scalability			Application type: Monitoring							Overhead			Disadvantage
	Flat	Cluster	Chain	Grid	Tree	Event	Periodic	Query	Decentralized	Centralized	Limited	Average	Good	Habitat	Environment	Health	Home/Office	Disaster	Battle field	Civil	Low	Average	High	
CHIRON			×			×				×		×							×	×		×		Multiple short chains
GROUP				×				×		×		×			×							×		Aggregation tree is frequently rebuilt. CH selects based on distance of grid only.
ATCBG				×			×			×	×				×						×			CH selects based on distance only. Tree building based on distance and energy
ToD					×	×			×		×				×							×		Big packets sent to the root due to reduction of aggregation ratio. Dropping rate is high.
ESPAN					×		×		×				×			×					×			
EADAT					×		×			×		×					×					×		The power threshold for broadcasting is not clearly explained how chosen. This work is not compared with existing aggregation algorithms
DAA	Structure less					×				×			×							×			×	Delay of transmission is not taken into account; it must be discussed.

E. Data Aggregation Rate

In WSNs, the aggregation of data is a method in which valuable information is gathered and combined in a single area of interest. To minify energy usage and conserve available resources, data aggregation could be thought of as a basic method of processing and is defined in terms of the rate at which data is aggregated. The rate of data aggregation is defined as the percentage of the amount of data successfully aggregated $Data_{Aggregated}$ to the total amount of data sensed $Total_{Sensed}$ as in Equation 5.

$$Data\ Aggregation\ Rate = \left(\frac{Data_{Aggregated}}{Total_{Sensed}}\right) \times 100 \quad (5)$$

The comparison among the varying classification of data aggregation techniques for the structure-based topology which is taken into account the above metrics is demonstrated in Table IV.

7. Conclusions and Future Work

Wireless sensor networks have a limited amount of resources. Since the key fuel that keeps them running is battery power, it's crucial to conserve energy in these networks. Because most of the energy used for data transmission and reception is significant, the data aggregation process is important and optimization is required. Not only do effective data aggregations offer energy conservation, but they also eliminate redundancy information and thus provide only usable data. When the source node data is sent to sink in a multihop manner by neighbors' nodes by minimizing transmitting and receiving power, the energy usage is low compared to sending data directly to sink, which is aggregation, reducing the transmission of data than without aggregation. This survey illustrates that these approaches not only minimize resource usage but also increase the lifespan of a network.

Acknowledgment

The researchers would like to express their gratitude to the University of Babylon in Iraq for their support.

References

[1] A. K. Idrees and A. K. M. Al-Qurabat, "Energy-efficient data transmission and aggregation protocol in periodic sensor networks based fog computing," *Journal of Network and Systems Management*, vol. 29, no. 1, pp. 1–24, 2021.

[2] A. K. M. Al-Qurabat and A. Kadhum Idrees, "Data gathering and aggregation with selective transmission technique to optimize the lifetime of internet of things networks," *International Journal of Communication Systems*, vol. 33, no. 11, p. e4408, 2020.

[3] A. K. M. Al-Qurabat and A. K. Idrees, "Two level data aggregation protocol for prolonging lifetime of periodic sensor networks," *Wireless Networks*, vol. 25, no. 6, pp. 3623–3641, 2019.

[4] A. K. M. Al-Qurabat, C. Abou Jaoude, and A. K. Idrees, "Two tier data reduction technique for reducing data transmission in iot sensors," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*. IEEE, 2019, pp. 168–173.

[5] I. D. I. Saeedi and A. K. M. Al-Qurabat, "An energy-saving data aggregation method for wireless sensor networks based on the extraction of extrema points," in *AIP Conference Proceedings*, vol. 2398, no. 1. AIP Publishing LLC, 2022, p. 050004.

[6] 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. 35, p. 0875, 2022.

[7] A. K. M. Al-Qurabat, A. K. Idrees, and C. Abou Jaoude, "Dictionary-based dpcm method for compressing iot big data," in *2020 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 2020, pp. 1290–1295.

[8] A. K. Idrees and A. K. M. Al-Qurabat, "Distributed adaptive data collection protocol for improving lifetime in periodic sensor networks." *IAENG International Journal of Computer Science*, vol. 44, no. 3, 2017.

[9] A. K. M. Al-Qurabat and A. K. Idrees, "Energy-efficient adaptive distributed data collection method for periodic sensor networks," *International Journal of Internet Technology and Secured Transactions*, vol. 8, no. 3, pp. 297–335, 2018.

[10] A. M. K. Abdulzahra and A. K. M. Al-Qurabat, "A clustering approach based on fuzzy c-means in wireless sensor networks for iot applications," *Karbala International Journal of Modern Science*, vol. 8, no. 4, pp. 579–595, 2022.

[11] S. A. Abdulzahra, A. K. M. Al-Qurabat, and A. K. Idrees, "Data reduction based on compression technique for big data in iot," in *2020 international conference on emerging smart computing and informatics (ESCI)*. IEEE, 2020, pp. 103–108.

[12] A. K. M. Al-Qurabat, "A lightweight huffman-based differential encoding lossless compression technique in iot for smart agriculture," *International Journal of Computing and Digital System*, 2021.

[13] W. B. Nedham and A. K. M. Al-Qurabat, "An improved energy efficient clustering protocol for wireless sensor networks," in *2022 International Conference for Natural and Applied Sciences (ICNAS)*. IEEE, 2022, pp. 23–28.

[14] A. K. M. Al-Qurabat, H. M. Salman, and A. A. R. Finjan, "Important extrema points extraction-based data aggregation approach for elongating the wsn lifetime," *International Journal of Computer Applications in Technology*, vol. 68, no. 4, pp. 357–368, 2022.

[15] A. K. M. Al-Qurabat and S. A. Abdulzahra, "An overview of periodic wireless sensor networks to the internet of things," in *IOP Conference Series: Materials Science and Engineering*, vol. 928, no. 3. IOP Publishing, 2020, p. 032055.

[16] I. F. Akyildiz and M. C. Vuran, *Wireless sensor networks*. John Wiley & Sons, 2010, vol. 4.

[17] W. Dargie and C. Poellabauer, *Fundamentals of wireless sensor networks: theory and practice*. John Wiley & Sons, 2010.

[18] S. Sirsikar and S. Anavatti, "Issues of data aggregation methods

TABLE IV. Taxonomy of Data Aggregation.

Network Topology	Pros	Cons
Flat	• The quality of routes between the source and sink is good. • Not need much effort to maintain and manage the network topology.	• Not respected energy restriction. Latency is very high. • The sink node's failure will bring the whole network down.
Tree	• Consumes less power. • Energy improvement.	• Power consumption is unbalanced between nodes close to the sink and nodes that are in the sub tree. • Data sent from the leaf to the root node takes a long time to arrive. • Building the tree is considered time consuming and expensive. • The maintenance overhead of tree is high.
Cluster	• Energy consumption enhancement. • Scalability optimization. • Improving channel bandwidth utilization due to the efficient data aggregation. • Cluster head failure does not affect the entire network.	• The non-unified clustering leads to the lack of a cluster head on some side of the network. • The sink node's failure will bring the whole network down. • When data transmitted to the sink using multihop will result in high latency. • The connectivity of network may not be ensured. • Also, if they are part of the same cluster, the amount of energy consumption varies considerably from one sensor to the next. As a result, energy distribution is unequal.
Chain	• Energy optimization. • Energy distribution is balancing. • Lengthening network lifetime.	• Management overhead of topology is high. • Too much delay for collection of data.

in wireless sensor network: A survey," *Procedia Computer Science*, vol. 49, pp. 194–201, 2015.

[19] A. Tripathi, S. Gupta, and B. Chourasiya, "Survey on data aggregation techniques for wireless sensor networks," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, no. 7, pp. 7366–7371, 2014.

[20] S. Sirsikar and S. Anavatti, "Issues of data aggregation methods in wireless sensor network: A survey," *Procedia Computer Science*, vol. 49, pp. 194–201, 2015.

[21] F. Ren, J. Zhang, Y. Wu, T. He, C. Chen, and C. Lin, "Attribute-aware data aggregation using potential-based dynamic routing in wireless sensor networks," *IEEE transactions on parallel and distributed systems*, vol. 24, no. 5, pp. 881–892, 2012.

[22] H. Luo, H. Tao, H. Ma, and S. K. Das, "Data fusion with desired reliability in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 3, pp. 501–513, 2010.

[23] X. Xu, X.-Y. Li, P.-J. Wan, and S. Tang, "Efficient scheduling for periodic aggregation queries in multihop sensor networks," *IEEE/ACM Transactions on networking*, vol. 20, no. 3, pp. 690–698, 2011.

[24] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on wireless communications*, vol. 1, no. 4, pp. 660–670, 2002.

[25] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd annual Hawaii international conference on system sciences*. IEEE, 2000, pp. 10–pp.

[26] J. Yuan and H. Chen, "The optimized clustering technique based on spatial-correlation in wireless sensor networks," in *2009 IEEE Youth*

Conference on Information, Computing and Telecommunication. IEEE, 2009, pp. 411–414.

[27] G. Li, S. Peng, C. Wang, J. Niu, and Y. Yuan, "An energy-efficient data collection scheme using denoising autoencoder in wireless sensor networks," *Tsinghua Science and Technology*, vol. 24, no. 1, pp. 86–96, 2018.

[28] T. D. Le, N. D. Pham, and H. Choo, "Towards a distributed clustering scheme based on spatial correlation in wsns," in *2008 International Wireless Communications and Mobile Computing Conference*. IEEE, 2008, pp. 529–534.

[29] K. T.-M. Tran and S.-H. Oh, "Uwsns: A round-based clustering scheme for data redundancy resolve," *International Journal of Distributed Sensor Networks*, vol. 10, no. 4, p. 383912, 2014.

[30] J. M. Bahi, A. Makhoul, and M. Medlej, "A two tiers data aggregation scheme for periodic sensor networks." *Adhoc & Sensor Wireless Networks*, vol. 21, no. 1, 2014.

[31] H. Harb, A. Makhoul, R. Tawil, and A. Jaber, "A suffix-based enhanced technique for data aggregation in periodic sensor networks," in *2014 international wireless communications and mobile computing conference (IWCMC)*. IEEE, 2014, pp. 494–499.

[32] J. Kulik, W. Heinzelman, and H. Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks," *Wireless networks*, vol. 8, no. 2, pp. 169–185, 2002.

[33] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: A scalable and robust communication paradigm for sensor networks," in *Proceedings of the 6th annual international conference on Mobile computing and networking*, 2000, pp. 56–67.

[34] D. Braginsky and D. Estrin, "Rumor routing algorthim for sensor networks," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, 2002, pp. 22–31.

[35] K. Sarangi and I. Bhattacharya, "A study on data aggregation techniques in wireless sensor network in static and dynamic scenarios," *Innovations in systems and software engineering*, vol. 15, no. 1, pp. 3–16, 2019.

[36] S. Yadav and R. S. Yadav, "Redundancy elimination during data aggregation in wireless sensor networks for iot systems," in *Recent trends in communication, computing, and electronics*. Springer, 2019, pp. 195–205.

[37] S. A. Abdulzahra, A. K. M. Al-Qurabat, and A. K. Idrees, "Compression-based data reduction technique for iot sensor networks," *Baghdad Science Journal*, vol. 18, no. 1, pp. 0184–0184, 2021.

[38] H. Harb, A. Makhoul, D. Laiymani, and A. Jaber, "A distance-based data aggregation technique for periodic sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 13, no. 4, pp. 1–40, 2017.

[39] I. Atoui, A. Ahmad, M. Medlej, A. Makhoul, S. Tawbe, and A. Hijazi, "Tree-based data aggregation approach in wireless sensor network using fitting functions," in *2016 Sixth international conference on digital information processing and communications (ICDIPC)*. IEEE, 2016, pp. 146–150.

[40] L. Karim and M. S. Al-kahtani, "Sensor data aggregation in a multi-layer big data framework," in *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 2016, pp. 1–7.

[41] A. Makhoul, H. Harb, and D. Laiymani, "Residual energy-based adaptive data collection approach for periodic sensor networks," *Ad Hoc Networks*, vol. 35, pp. 149–160, 2015.

[42] Y. Song, J. Luo, C. Liu, and W. He, "Periodicity-and-linear-based data suppression mechanism for wsn," in *2015 IEEE Trustcom/BigDataSE/ISPA*, vol. 1. IEEE, 2015, pp. 1267–1271.

[43] B. Srbinovski, M. Magno, B. O'Flynn, V. Pakrashi, and E. Popovici, "Energy aware adaptive sampling algorithm for energy harvesting wireless sensor networks," in *2015 IEEE Sensors Applications Symposium (SAS)*. IEEE, 2015, pp. 1–6.

[44] J. Zhang, L. Ren, Y. Ding, and K. Hao, "Adaptive sampling algorithm with endocrine regulation mechanism for wireless sensor network," in *2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*. IEEE, 2015, pp. 502–507.

[45] A. Masoum, N. Meratnia, and P. J. Havinga, "An energy-efficient adaptive sampling scheme for wireless sensor networks," in *2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*. IEEE, 2013, pp. 231–236.

[46] C. Yang, Z. Yang, K. Ren, and C. Liu, "Transmission reduction based on order compression of compound aggregate data over wireless sensor networks," in *2011 6th International Conference on Pervasive Computing and Applications*. IEEE, 2011, pp. 335–342.

[47] C. Alippi, G. Anastasi, M. Di Francesco, and M. Roveri, "An adaptive sampling algorithm for effective energy management in wireless sensor networks with energy-hungry sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 2, pp. 335–344, 2009.

[48] Y. W. Law, S. Chatterjea, J. Jin, T. Hanselmann, and M. Palaniswami, "Energy-efficient data acquisition by adaptive sampling for wireless sensor networks," in *Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly*, 2009, pp. 1146–1151.

[49] S. Chatterjea and P. Havinga, "An adaptive and autonomous sensor sampling frequency control scheme for energy-efficient data acquisition in wireless sensor networks," in *International Conference on Distributed Computing in Sensor Systems*. Springer, 2008, pp. 60–78.

[50] K. Dasgupta, K. Kalpakis, and P. Namjoshi, "An efficient clustering-based heuristic for data gathering and aggregation in sensor networks," in *2003 IEEE Wireless Communications and Networking, 2003. WCNC 2003.*, vol. 3. IEEE, 2003, pp. 1948–1953.

[51] F. Fanian and M. K. Rafsanjani, "A novel routing efficient algorithm based on clustering in wsns," *Indian Journal of Science and Technology*, vol. 6, no. 12, pp. 5542–45, 2013.

[52] O. Younis and S. Fahmy, "Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on mobile computing*, vol. 3, no. 4, pp. 366–379, 2004.

[53] S. Yoon and C. Shahabi, "The clustered aggregation (cag) technique leveraging spatial and temporal correlations in wireless sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 3, no. 1, pp. 3–es, 2007.

[54] A. Manjeshwar and D. P. Agrawal, "Teen: Arouting protocol for enhanced efficiency in wireless sensor networks." in *ipdps*, vol. 1, no. 2001, 2001, p. 189.

[55] C. Ramar and K. Rubasoundar, "A survey on data aggregation techniques in wireless sensor networks," *International Journal of Mobile Network Design and Innovation*, vol. 6, no. 2, pp. 81–91, 2015.

[56] H. Kumar and P. K. Singh, "Power transmission analysis in wireless sensor networks using data aggregation techniques," *International Journal of Information System Modeling and Design (IJISMD)*, vol. 9, no. 4, pp. 37–53, 2018.

[57] A. K. M. Al-Qurabat, Z. A. Mohammed, and Z. J. Hussein, "Data traffic management based on compression and mdl techniques for smart agriculture in iot," *Wireless Personal Communications*, vol. 120, no. 3, pp. 2227–2258, 2021.

Ali Kadhum M. Al-Qurabat Ali Kadhum M. Al-Qurabat received his Ph.D. in Software (Wireless Sensor Networks) in 2018 from the University of Babylon, Iraq. He received the M.Sc. degree in Information Technology from Universiti Tenaga Nasional (UNITEN), Malaysia, in 2012. He received the B.Sc. degree in computer science from the University of Babylon, Iraq, in 2002. He joined the Department of Computer Science, University of Babylon, Iraq, in 2006, where he is currently an Assistant Professor. His research interests include WSN, Data Aggregation, IoT, Data Mining, Data Compression and E-Procurement. Dr. Al-Qurabat has been a TPC Member of several national and international networking conferences and a Reviewer for several international journals. He is the author of a great deal of research studies published in national and international journals, and at conference proceedings.

Suha Abdulhussein Abdulzahra Suha Abdulhussein Abdulzahra obtained a Bachelor's Degree in Computer Science from the University of Babylon-College of Science-Computer Department in 2006 and then got a Master's Degree in Networks from the University of Babylon-College of Science for Women-Computer Department in 2020. She is currently working towards her Ph.D. in the Network Department, College of Information Technology, University of Babylon, Iraq. She works as an assistant Lecturer at the Ministry of Education since 2007 until now. Her research interests include IoT, Wireless Sensor Networks.