DATA AGGREGATION IN WIRELESS SENSOR NETWORKS: EMERGING RESEARCH AREAS

Ajobiewe, Damilola Nnamaka

Department of Computer Science,
Federal College of Education (Special) Oyo, Nigeria.
ajobiewe.dn@fcesoyo.edu.ng
https://orcid.org/0000-0001-6734-3858

Abstract

Wireless Sensor Networks (WSNs) generate an immense measure of explicit data usage. Such data, which is an exorbitant issue, should be prepared and then transmitted to the base station. Efficient data handling and energy monitoring are primary challenges. since WSN hubs are asset-compelled. The point of wireless sensor networks is not confined to data collection in the present climate. Yet the retrieval of beneficial data still depends on it. The term used for the retrieval of beneficial data is data aggregation. In social affairs, data aggregation helps to collect data in energy-efficient ways to elongate the network's duration. The majority of the detected data by the sensors were seen to be excessive. On the off chance that data repetition can be reduced, it will prompt the organization's extended lifetime and decreased inertness at that point. This paper explains a wide methodological analysis of literature on data aggregation in WSNs. Review of various strategies to reduce data repetition, and specifically through aggregation, as well as scientific categorization of data aggregation, challenges, and broken-down aggregation methods proposed over the last ten (10) years, are discussed.

Keywords: *Data Aggregation, Wireless Sensor Network, Security, Network Lifespan, Energy Efficiency.*

Introduction

Wireless sensor networks (WSNs) consist of a wide-ranging amount of sensor nodes for particular applications distributed in the area of concern, which is typically tiny nodes with networking, connectivity, and sensing capacities. These sensor nodes are identified by (Pourpeighambar et al., 2011; Randhawa & Jain, 2017; Tan & Körpeoğlu, 2003); as well as detailing how they interact through short-range radio signals and work to complete shared tasks. WSNs are critical in a variety of network areas, including environmental control, military applications, health-care applications, industrial process management, home intelligence, safety as well as surveillance, and so on. Therefore, by finding several routes between source and destination, a protocol for routing that remains effective per resources, data aggregation, and energy consumption must be defined. (Tan & Körpeo'lu, 2003). Security concerns, energy usage, latency, data confidentiality, honesty, in addition, once the sensor node is deployed in a hostile environment, data aggregation is crucial (Massad et al., 2008). A study of various data aggregation approaches, the impact of data aggregation within WSN, data aggregation methods, data aggregation security issues in WSN data aggregation are discussed. The article also includes a methodical literature review to assess and identify research problems in the area of data aggregation within WSNs centered on current studies. According to the article, energy savings, as well as latency, are the two most significant variables that influence the efficiency of data aggregation techniques for wireless sensor networks. The latency is associated with aggregating data from local sources and delay is the process of keeping data back across intermediate nodes to merge that to data by distant sources. When evaluating the factors, the two root positioning models are used (energy savings and delay). The two models are defined by the event radius (ER) model with each random source model (Rajagopalan & Varshney, 2006). Important energy benefits are feasible with data aggregation where the origins are grouped around each other or placed arbitrarily, according to the modeling. The benefits are greatest because there are a lot of outlets and they're both close but further away from the base station (Al-Humidi & Chowdhary, 2019; Gatani et al., 2006; Renjith & Baburaj, 2016).

The Literature Review

Chatterjea and Havinga (2003b) suggested Clustered Diffusion with Dynamic Data Aggregation (CLUDDA), using diffusion through a clustering-based data-centric strategy that utilizes in-arrange preparation to maximize data. The grouping strategy is combined with coordinated distribution in this technique. Another curiosity message configuration is defined, in which the inquiry comprehension data is sent along with the query. This instrument is used so that a node may decode the new condition's fluctuating structure query. Intrigue proliferation and data engendering are the two steps of the measurement. When one method fails, it employs an effort to repair the machine in order to restore it. If the field of necessary data sources shifts, so does the data aggregation focus. This technique eliminates unnecessary handling and increases idleness while still requiring a large amount of memory to store intrigue changes and inquiry responses (Chatterjea & Havinga, 2003a). For the most part, this approach is secondary to the fundamental naming system. To reduce data fragmentation induced by spatial similarities between clusters, a range of researchers have proposed using clustering procedure for Decentralized Optimal Compression (DOC) and perhaps a time to reduce coding technique. To lower the cost of intra-cluster contact, the suggested algorithm decides the best proportion allocation inside each cluster. Although an intra-cluster coding procedure was proposed for performing Slepian–Wolf technique within a particular cluster, Wang, Li, et al. (2007) claimed that the technique, while improving communication cost and compression ratio, failed to account for energy consumption.

Zhou et al. (2008) in this study, the authors centered on data aggregation problems of energy-constrained sensor networks. The authors performed research on data aggregation algorithms for wireless sensor networks. Using multiple algorithms, try comparing different output metrics such as lifespan, data precision, and latency. Finally, possible research avenues were discussed.

Chen et al. (2008); (Zheng et al., 2010) cluster-based data aggregation was investigated, and a circulated data aggregation mechanism was proposed. Their main aim was to resolve the Clustered Slepian–Wolf Coding (CSWC) difficulty and optimize compression gain by using the most disjoint cluster possible to cover the network. Jung et al. (2011) suggested a hybrid data aggregation approach known as mixed clustering-based data aggregation to help with complicated aggregation by combining multiple clustering methods at once. The evaluation method for an appropriate clustering strategy is dependent on the network's status. During the initialization process of this proposed strategy, between the sink node as well as the other network nodes, tree topology is formed. In the next step, cluster head dynamic selection algorithms were being used. In a way, depending on the network status, this hybrid approach provides both dynamic and static clustering. Maraiya et al. (2011a) proposed an Efficient Cluster-head Selection Scheme for Data Aggregation (ECHSSDA), employs a paradigm of cluster-head selection and without depending upon latency as a metric, cluster formation will increase network lifespan and energy efficiency. This technique reduces clustering overhead by choosing a cluster head and a subordinate cluster-head. When a cluster head is overburdened with tasks like sending, receiving, and computations, it may lose control. As a result, anytime a cluster-head dies or fails, a new election is held to choose a new cluster-head, and re-clustering is done to eliminate hot spots. However, the notion of an affiliate cluster-head, who might take responsibility for cluster-head if its energy level fell beyond a certain threshold, might be used to reduce overhead. The cluster setup process, wherein clusters are created, and cluster steady phase is the two phases of the proposed algorithm. The cluster-head is activated in the second step to gather incoming data packets, compile results and send a message to its base station. This technique enhances energy performance then streamlines the cluster choice procedure. A Two-Tier Cluster-Based Data Aggregation (TTCDA) architecture was suggested, which uses temporal and spatial similarity to apply discretely as well as separable aggregation for data packets given by each node (Mantri et al., 2012). The TTCDA is in control of cluster creation, inter-cluster and intra-cluster aggregation are two types of aggregation. Tier 2 then combines the aggregated into one packet depending on the software parameters, utilizing division or additive functionality. However, taking into account node versatility and heterogeneity will strengthen this strategy much further. In the second step, the sensor nodes use additive and divisible aggregation functions in an absolute routine configuration to send transmissions to the cluster head over short distances. The cluster head is assigned to a group based on the nearest distance to both the sink as well as the total packet count. Data packets are collected differently at periodic intervals from a transmitted broadcast message in intra-cluster aggregation. Each cluster head acts as an individual node in the third step. Reduced packet count in data aggregation is the end product at the sink.

Yuea et al. (2012) suggested an Energy-Efficient then Balanced Cluster-Based Data Aggregation Algorithm (EEBCDA) to see if there is an ideal balanced load distributor approach across the whole network where one-hop contact is used to transfer data from cluster-head to base station, eliminating energy dissipation by decreased intra-cluster communication. Any swim-lane remains separated into a set of quadrilateral sections referred to as grids, and each swim-lane is divided into one-of-a-kind sizes of rectangular areas referred to as swim lanes. The network division step is a part of the network structuring process. Each grid elects a cluster-head node with the most electricity. Grids that are further apart from the base station become longer and contain more nodes. The amount of energy used is balanced in this process, since a cluster-head that absorbs more power, more sensor nodes will participate in cluster-head voting. According to Al-Karaki et al. (2009), this approach increases network lifespan and has a stable energy consumption, but in remote grids, latency is higher.

Grouping nodes and clusters for green knowledge aggregation (GCEDA) were introduced by Mantri et al. (2013), in which nodes are clustered mainly based on accessible facts. Despite the higher latency, transmitting costs were lowered. To pass aggregated data to a distant sink, the cluster head employs divisible and additive knowledge aggregation characteristics. Each cluster-head calculates the connection between nodes that are one hop apart. The suggested method divides the whole routing protocol into three stages. Clusters are randomly formed throughout the first section, cluster forming. Based on Euclidean distance and maximal electricity, the cluster-head is chosen. Nodes with similar statistics are discovered and grouped in the intra-cluster section, and then a few aggregation functions are used to depend on the data. Both cluster heads serve as supply nodes in the inter-cluster segment, sending the aggregated records to the base station. Cluster head grouping reduces resource consumption and increases community reliability without sacrificing node heterogeneity or data quality (Mantri et al., 2013).

Sinha and Lobiyal (2013) proposed an energy-efficient data aggregation approach focused on divergence, wherein sensors having sensed the very same thing were first grouped according to certain distinct clusters. The least divergent clusters would be eventually merged since these residual un-clustered sensors approximate their separation throughout comparison to nearby clusters. The route from root to sink is determined by the node's highest utilizable residual capacity. It was discovered that it had a longer total node life.

The algorithm has two stages: preliminary clustering and final clustering. The nodes that sense the same data are placed in separate clusters in the first step. In the second stage, the residual sensors calculate the deviation from nearest neighbors and join each cluster with the least divergence. The sensed data is mapped in the range [0...1] using a window function. After taking into consideration the latency constraint, this approach is used to increase convergence speeds, aggregation rates, packet size losses, transmission cost, and network lifespan. This approach may be enhanced still further by accounting for node heterogeneity and energy-rich multiple sinks.

To boost energy performance in a cluster-based duty-cycled WSN, Rout and Ghosh (2014) suggested an energy-efficient adaptive knowledge aggregation with group coding (ADANC) (at the bottleneck zone) via hoping away some individual nodes from the clustered head that serve as network decoder nodes while others serve in place of simple spread nodes in the cluster. It's a low-power, cluster-based data aggregation system that divides sensor nodes into two types: simple relay nodes and group coder nodes. Data aggregation is solely the responsibility of the stage of document correlation. Network compression is used where the sum of knowledge similarity issues in the obtained packets is limited. In any case, conventional data aggregation is complete, whereas the data correlation problem is serious. With the obligation interval, energy demand is often minimized at the node level. This method is used to reduce transmission costs and thereby reduce power consumption; however, network latency is not taken into account.

As a result, Banerjee and Bhattacharyya (2014) introduced a fairly balanced distribution-based data aggregation process, wherein fluffy reasoning is often used to choose the cluster head and disperse the same amount of burden through bunches to maximize data transmission. It was planned to develop an energy-efficient routing algorithm that would minimize the overall number of packets transmitted and re-clustering at each round while maintaining network service efficiency. The primary aim of this suggested solution is to resolve the problem of community heads' energy spillage, which arises as a consequence of lopsided burden conveyance. The suggested process is divided into three phases, each of which is divided into three adjusts. The CSMA/CD technique is used to frame bunches during the initial arrangement period. In the standard setup stage, any node transmits its energy status data to the group head, which is also sent to the sink or base station as the case maybe.

Jesus et al. (2014), reported a critical overview of the impact of distributed aggregation algorithms, defining the various forms of aggregation techniques. To reduce the data transmission scale, Xu et al. (2015) proposed an improved data aggregation method focused on signal processing (HDACS), which allows for the complex set of multiple compression levels based on the scale of clusters at various tiers of the data aggregation tree. Rather than setting up the strongest node as the drain, a structure of multi-stage clusters remains built for intermediate data processing. In contrast to other compressive sensing methods, it decreases the data extent in knowledge exchange. The underlying domain is restored using a DCT-based algorithm. The value is often defined by the number of processors as well as the volume of radio energy used. The proposed approach was validated using real-world data and simulated datasets in Sidnetswan's simulation platform.

By focusing on energy consumption, for heterogeneous networks, Mantri et al. (2015) proposed a bandwidth-efficient cluster-based data aggregation (BECDA) method. To provide a solution to the inefficient data collection of the in-network method, an optimized approach of intra- and inter-cluster aggregation at different rates of data generation on randomly distributed nodes was created.

This procedure employs the concept of data link inside the parcel to add aggregation capability to data produced by sensor hubs. Any node may produce erratic data that varies from 0 to 1 by using arbitrary power. Standard hub (20 J), advance hub (30 J), and superhub (40 J) are the three types of nodes, each with a different energy level (40 J). It is created the variable traffic measure. CH is the node in the network of a homogeneous network by far the most energy of both the cluster participants and the most neighbor nodes. This approach decreases data transmission rates, electricity use, and correspondence costs. In either scenario, it results in a decrease in throughput.

Distributed and efficient data aggregation scheduling, proposed by Gao et al. (2019), is a modern approach for scheduling distributed and efficient data aggregation over multi-channel connections (DEDAS-MC). DEDAS-MC lowers latency by transmitting aggregated data to a sink over several channels. DEDAS-MC is a sensor scheduling algorithm that minimizes data aggregation latency and prevents interruptions on a given tree. After that, a distributed algorithm for constructing low-latency data aggregation trees is suggested using the Markov approximation technique. The study centered on the data aggregation latency problem. Two variables determine the latency of data aggregation. Due to the presence of interference, collision-free scheduling is crucial for reducing data aggregation latency. In addition, the tree structure has a major impact on data aggregation latency.

In the area of data aggregation, new research is constantly emerging. However, a systematic literature review is needed to assess and integrate the current studies in this area.

Data Aggregation designed for Wireless Sensor Networks: Security Concerns

In a wireless sensor network, confidentiality and honesty remain the two types of security that are necessary for data aggregation. Data protection, or shielding critical data transfer against passive attacks such as eavesdropping, is the most fundamental security issue. Since the wireless channel is susceptible to eavesdropping via a cryptographic system, data confidentiality is predominantly used in a hostile setting (Maraiya et al., 2011b). Data integrity is the next protection problem. Integrity tends to deter malicious sensor sources including aggregator nodes from compromising the final aggregation value significantly. A sensor node in a sensor network may be easily compromised, changing or discarding messages. There are two strategies for protecting data aggregation: both hop-by-hop encryption nor end-to-end encryption use the same strategy.

Data Aggregation Classification

In this paper, data aggregation overview involved lifetime of the network, data accuracy, energy efficiency, latency, and data aggregation rate as suggested by (Dagar & Mahajan, 2013; Li et al., 2011)

(a) Energy Efficiency: each sensor will use the same amount of energy during each data collection round, but sensor nodes use different quantities of energy for data transmission. A data aggregation procedure in WSNs remains energy efficient if it has the greatest versatility when using the least amount of energy. Energy efficiency is described as the ratio of the amount of data efficiently transmitted in a sensor network to the total energy required to transmit that data.

Energy efficiency is calculated using Equation 1.

$$\sum_{i=1}^{n}\left(\frac{Total\ number\ of\ successfull\ data\ transferred\ in\ a\ sensor\ network}{Total\ amount\ of\ energy\ consumed\ to\ transfer\ data}\right) \qquad (1)$$

In a sensor network, *n* represents the number of sensor nodes, and *i* represents the number of iterations.

(b) Network Lifespan represents the amount of data aggregation rounds performed before the first sensor node's energy is drained. In other terms, it is described as the period (number of rounds) before the first sensor node or group of sensor nodes in the network runs out of energy (battery power) or network disconnection due to the failure of one or more sensors, as (Equation 2):

$$NS_n^n = \min_{v \in V} NS_v \qquad (2)$$

Where the network lifetime NS_n^n terminates as soon as the first node fails and NS_v is the lifespan of node v, V is the node-set without the sink node.

(c) Data Accuracy: Depending on the application for which the sensor network is designed,data accuracy is defined in different ways. The close approximation of goal location at the sink, for example, determines data precision in the target localization dilemma. Data consistency is characterized as the proportion of correctly transferred data to total data submitted (Equation 3).

$$Data\ Accuracy = \frac{Sum\ of\ data\ transferred\ successfull}{Total\ sum\ of\ data\ sent} \quad (3)$$

(d) Latency: the duration between the data packets received at the sink and the data packets produced at the source nodes is referred to as latency. To put it another way, latency refers to the period it takes a sensor node to send and receive results. As shown in (4)

$$Latency_i = \sum_{i=1}^{n} (Time\ taken\ to\ receive\ data - Time\ of\ sending\ data) \quad (4)$$

(e) In WSNs, data aggregation rate is the method of gathering and integrating valuable knowledge in a specific area of interest. Data aggregation is characterized in terms of data aggregation rate and can be called a fundamental processing technique to minimize energy usage and conserve limited resources. The data collection frequency is described as the proportion of successfully aggregated data to the total amount of data sensed (Equation 5).

$$Data\ Aggregation\ Rate = \frac{Sum\ of\ the\ successfull\ data\ aggregated}{Total\ sum\ of\ Sensed\ data} \times \frac{100}{1} \quad (5)$$

Present State of Data Aggregation Strategies in WSNs

For aggregating valuable data in WSNs, the Data Aggregation Technique (DAT) is important (Krishnamachari et al., 2002). Data is stored at intermediate nodes in this system to conserve resources and reduce processing time. Since it seeks to minimize energy usage at any node, this method, therefore, extends the network lifespan. Lossy aggregation including packet size reduction as well as lossless data aggregation without packet size reduction are the remaining two methods for in-network aggregation. Data is collected from different source nodes and then a category feature such as number (), count (), limit (), and minimum () is added to the gathered data in lossy aggregation (). Since only the measured value of the aggregate feature is introduced into the packet after compression, rather than submitting the whole packet of each node, the size of the packet is minimized in this technique. Consider a forest fire monitoring device, where a simple average or maximum temperature reading is needed. Lossy aggregation is expected in such applications since it reacts to the base station in a timely fashion. Each packet is combined into a single packet without being compressed in lossless aggregation.

Data Aggregation Methodology

Data aggregation methodologies include clustered, tree-based, cluster-based approaches, and in-network aggregation, as seen in Figure 1.

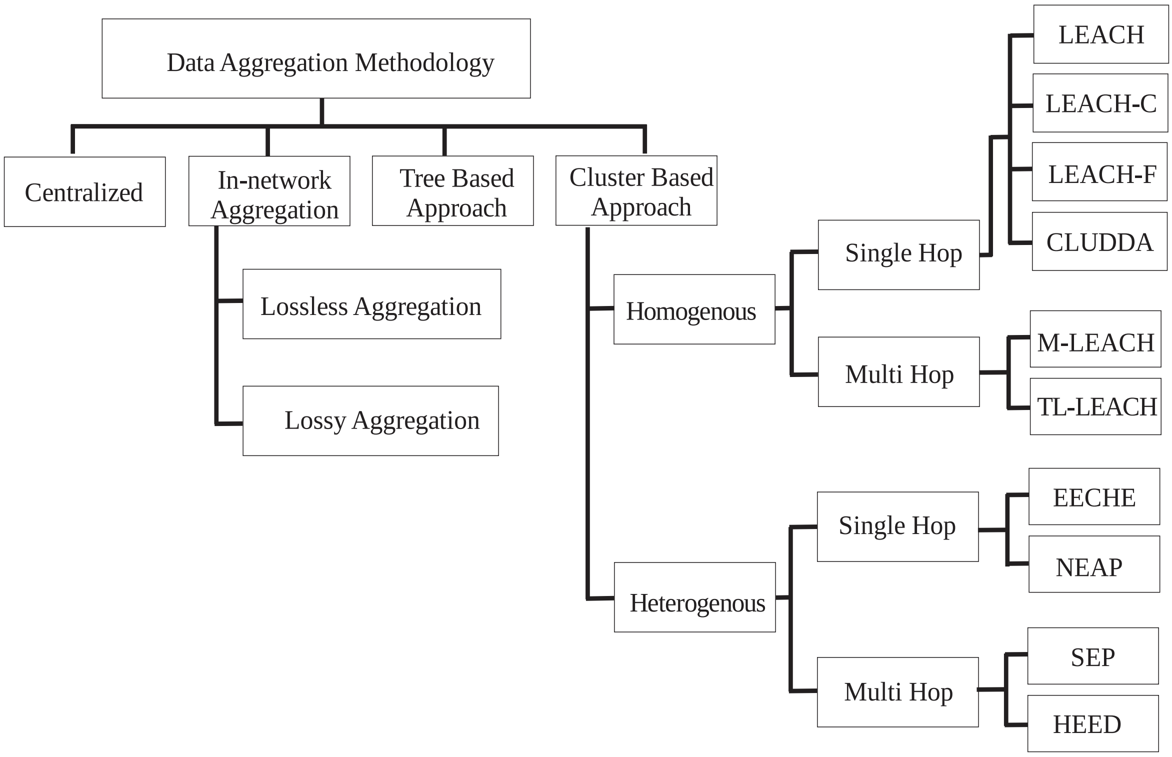

Figure 1: Data Aggregation Methodologies

Source: Author's construct (2020)

The data aggregation process is carried out using a specific routing protocol. The goal is to collect data to reduce energy consumption. As a consequence, sensor nodes will route packets based on data packet content and choose the next hop to make network aggregation easier. Routing protocols are divided by network configuration, which is why they are focused on careful considerations (Krishnamachari et al., 2002).

Tree-Based Approach

The tree-based method requires building an aggregation tree to describe aggregation. The tree is a limited spanning tree in which the sink node acts as the foundation and the root node serves as the leaves. The leaves node sends data to the drain, which is the root node (base station). Wireless sensor networks, as we already know, are vulnerable to failure. If a data packet is lost at any level of the tree, the data is lost not just for that level, but also for all subtrees that are connected to it. This method can be used to create the best aggregation techniques. Madden et al. (2005), created the Tiny Aggregation (TAG) approach, which is a data-centric protocol. The operation of TAG is split into two phases: spread and set. First, an aggregation tree, which is usually a minimum spanning tree, is constructed. The root node serves as the base station, while the leaf nodes serve as the source nodes and the intermediate nodes serve as the parent nodes in this tree. In a route discovered between leaf node and base station, the leaf nodes give their sensed node to their parent node.

Cluster-Based Approach

It is impractical for sensors to relay data directly to the sink in energy-constrained sensor networks of significant scale in such scenarios. Cluster-based approaches are hierarchical. The whole network is split into many clusters in the cluster-based method. A cluster-head is chosen from among cluster representatives for each cluster. Cluster-heads serve as aggregators, combining data obtained from cluster participants on a local level and transmitting the result to the base station (sink). Several cluster-based network organization and data-aggregation protocols for wireless sensor networks have recently been suggested. Figure 1 illustrates the organization of a sensor network focused on clusters. Long-range transmissions or multi-hopping via other cluster heads enable the cluster heads to interact directly with the drain. The maximum lifetime data aggregation (MLDA) algorithm was suggested, which finds data gathering schedules based on sensor node and base-station position, data packet size, and sensor node capacity. For each round, data-gathering schedule determines how data packets are gathered from sensors and sent to the base station. A plan is basically a set of aggregation trees. Dagar and Mahajan (2013), suggested a heuristic-greedy clustering-related MLDA based on the MLDA algorithm in (Dagar & Mahajan, 2013). They split the network into clusters and related to each cluster as a super-sensor in this way. They then calculate the super-sensors' maximum lifetime schedule and use it to create aggregation trees for the sensors. A two-phase clustering (TPC) scheme is presented by (Wang, Li, et al., 2007; Wang, Zheng, et al., 2007). Step I of this scheme generates clusters with both a cluster-head, but each node inside the cluster links directly to the cluster-head. Move 1 necessitates dispersed cluster-head rotating dependent on the residual energy level of sensor nodes, reducing sensor period volatility and conserving energy by eliminating unnecessary cluster-head rotation. In phase 2, each node in the cluster searches for a data relay position, which is a neighbor who is nearby to the cluster-head and creates a data relay link. Sensor nodes in a cluster now transmit data to the cluster-head through a direct connection or a data relay channel, which is a much more energy-efficient scheme. The data relay point aggregates data before sending it to another data relay point or cluster-head. TPC phase II may create an unwanted data relay connection between neighbors in the case of a high network density, causing sensors to close together to feel the same data, causing a loss of energy. Using this form, the field of concern is separated into a set of clusters. Each cluster chooses a cluster head, whom sole task is to compile the data. Instead of sending data to the base station directly, each node detects the required phenomenon and reports it to the cluster's CH. As a result, it saves a significant amount of electricity in a network.

Data Aggregation Progression

The evolution of data aggregation strategies (DATs) in WSNs from 2002 to 2019 is seen. In the year 2002, network lifetime and network density data aggregation approaches based on network lifetime and resources were presented. Clustered diffusion with Dynamic Data Aggregation (CLUDDA) and dynamic data aggregation technique was proposed in 2003 as diffusion and clustering dependent data-centric technique focusing on network lifetime

(Chatterjea & Havinga, 2003). In 2006, a safe pattern-based data aggregation methodology focusing on security and bandwidth was proposed. In 2007, a sparse data aggregation method was proposed, focusing on expense and failure likelihood (Wang et al., 2007).

In the year 2008, a linear distribution-based data aggregation methodology focusing on energy was proposed (Zhou et al., 2008). In addition, parameters-based, energy-oriented distributed and scalable, and dynamic (Zheng et al., 2010) data aggregation strategies were proposed in 2010, with a focus on connectivity expense, network lifetime, energy, aggregation time, latency, and aggregation rate. During this time, researchers sought to optimize the global compression advantage (Zheng et al., 2010).

In 2011, an effective cluster-head selection scheme for data aggregation (ECHSSDA) was introduced, which uses a model of cluster-head selection and cluster creation, estimation, adaptive clustering, and multi-source temporal dependent data aggregation techniques, with a focus on communication redundancy, network lifetime, resources, packet transmission, error rate, and success rate without co-location, with an emphasis on transmission redundancy, network lifetime, energy, packet transmission, error rate, and success rate without (Jung et al., 2011; Maraiya et al., 2011a). In 2012, recoverable hidden two-tier clustering-based mechanisms and tree-based data aggregation methods were discussed, with a focus on connectivity overhead, expense, packet utilization, energy consumption, and energy expense utilizing temporal and spatial correlation (Mantri et al., 2012). Also, in 2012, EEBCDA (energy efficient and balanced cluster-primarily based data aggregation algorithm) was suggested. This approach extends the existence of the network and reduces electricity demand; however, it increases latency in far-flung grids (Yuea et al., 2012). In 2013, data aggregation strategies that centered on network lifespan, energy usage delivery ratio, aggregate ratio, network reliability, and node density were presented (Mantri et al., 2013; Sinha & Lobiyal, 2013). Researchers proposed an energy-efficient adaptive knowledge aggregation utilizing population coding (ADANC), shortest path-dependent, semantic correlation tree (SCT) based adaptive, enhanced distributed, and latency related data aggregation strategies in 2014, network lifetime, propagation overhead, energy consumption, data efficiency, and aggregation latency are all variables to remember. (Banerjee & Bhattacharyya, 2014; Jesus et al., 2014; Rout & Ghosh, 2014). In 2015, Data aggregation strategies focused on bandwidth-efficient cluster-based, delay-aware, confidence management, multi-criterion decision-making, and learning automata were suggested, with a focus on bandwidth, latency, packet distribution ratio, energy usage, network lifespan, aggregation expense, protection, connectivity overhead, and privacy-preserving effectiveness (Mantri et al., 2015; Xu et al., 2015).

In 2016, Atoui et al. (2016), presented a scheme whereby data is filtered using fitting functions at the first level, and if the data's norm value is within a threshold value, it's sent to the aggregator for second-level aggregation. As a result, in Khriji et al. (2018), differential data from sensors reduces the transmission of unnecessary data over successive cycles are used. Aggregators are fixed in this system, while algorithms are dispersed in a cluster network.

Also, in 2019, Kumar and Kim (2019) employ multi-channel TDMA scheduling techniques to decrease collusion and minimize latency, as well as a meta-heuristic approach for energy reduction. The network type is the tree, the aggregator is fixed, and the node type is homogeneous. Unlike Kumar and Kim (2019), Sarangi and Bhattacharya (2019) used cluster generation based on a neural network with ant colony optimization. In this system, aggregators are mobilized using distributed methods, and the deployment model is the cluster.

Yadav and Yadav (2019) The aggregator nodes, which are also homogeneous model types, utilize a linear classifier svm - based to identify and eliminate redundant input. With a distributed algorithm, the aggregators are fixed.

The topic of data aggregation latency was a focus for researchers in 2019. Two factors, according to studies, affect data aggregation latency. For instance, reducing data aggregation latency due to disruption needs efficient collision-free scheduling. Second, data aggregation latency is highly influenced by the tree structure (Gao et al., 2019).

Conclusion

This paper focuses on data aggregation, data aggregation advancements, different methods of data aggregation strategies, and a review of data aggregation techniques with and without wireless sensor networks. Rajagopalan and Varshney, as well as Jesus et al., have previously identified data aggregation research issues. This paper utilizes a systemic research method to survey the most recent field study in WSN data aggregation up to 2019. Data aggregation mechanisms and subtypes are thoroughly examined. The data aggregation strategies are compared based on important aspects of data aggregation, as well as the researcher's intent, commitment to science, and various data aggregation algorithms like adaptive, cluster, hidden, resources, latency, network lifetime, network density, nature-inspired optimized, QoS, scheduling, tree, predictor, structure-free real-time, evolutionary game, and a. The majority of the study focused on energy-based data aggregation, with data aggregation with prediction, structure-free real-time, evolutionary game, and hybrid data aggregation techniques still in the early stages of development. Cluster-based data aggregation is the most researched area of WSN after energy-based data aggregation. Data aggregation study was also found to be more widespread in 2011 than in 2002 and 2005, according to the literature. Moreover, during this period, research on data aggregation in WSN concentrated on energy as a QoS parameter, ignoring the fact that reliability and congestion control as QoS parameters still need a considerable amount of work.

References

1. Al-Humidi, N., & Chowdhary, V. (2019). Lightweight Data Transmission Scheme Based onData Aggregation Technique in Wireless Sensor Networks. Proceedings of International Conference on Communication and Information Processing (ICCIP),

2. Al-Karaki, J. N., Ul-Mustafa, R., & Kamal, A. E. (2009). Data aggregation and routing in wireless sensor networks: Optimal and heuristic algorithms. *Computer Networks*, *53*(7), 945-960.

3. Atoui, I., Ahmad, A., Medlej, M., Makhoul, A., Tawbe, S., & Hijazi, A. (2016). Tree-based data aggregation approach in wireless sensor network using fitting functions. 2016

Sixth international conference on digital information processing and communications (ICDIPC),

4. Banerjee, R., & Bhattacharyya, C. K. (2014). Cluster based routing algorithm with evenly load distribution for large scale networks. 2014 International Conference on Computer Communication and Informatics,

5. Chatterjea, S., & Havinga, P. (2003a). CLUDDA-Clustered diffusion with dynamic data aggregation. Ajaccio, Corsica, France: Cabernet Radicals Workshop,

6. Chatterjea, S., & Havinga, P. (2003b). A dynamic data aggregation scheme for wireless sensor networks. *Proc. Program for Research on Integrated Systems and Circuits*.

7. Chen, H., Mineno, H., & Mizuno, T. (2008). Adaptive data aggregation scheme in clustered wireless sensor networks. *Computer Communications*, *31*(15), 3579-3585.

8. Dagar, M., & Mahajan, S. (2013). Data aggregation in wireless sensor network: a survey. *International Journal of Information and Computation Technology*, *3*(3), 167-174.

9. Gao, Y., Li, X., Li, J., & Gao, Y. (2019). Distributed and efficient minimum-latency data aggregation scheduling for multichannel wireless sensor networks. *IEEE Internet of Things Journal*, *6*(5), 8482-8495.

10. Gatani, L., Re, G. L., & Ortolani, M. (2006). Robust and efficient data gathering for wireless sensor networks. Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06),

11. Jesus, P., Baquero, C., & Almeida, P. S. (2014). A survey of distributed data aggregation algorithms. *IEEE Communications Surveys & Tutorials*, *17*(1), 381-404.

12. Jung, W.-S., Lim, K.-W., Ko, Y.-B., & Park, S.-J. (2011). Efficient clustering-based data aggregation techniques for wireless sensor networks. *Wireless Networks*, *17*(5), 1387-1400.

13. Khriji, S., Raventos, G. V., Kammoun, I., & Kanoun, O. (2018). Redundancy elimination for data aggregation in wireless sensor networks. 2018 15th International Multi-Conference on Systems, Signals & Devices (SSD),

14. Krishnamachari, L., Estrin, D., & Wicker, S. (2002). The impact of data aggregation in wireless sensor networks. Proceedings 22nd international conference on distributed computing systems workshops,

15. Kumar, S., & Kim, H. (2019). Energy efficient scheduling in wireless sensor networks for periodic data gathering. *IEEE access*, *7*, 11410-11426.

16. Li, H., Lin, K., & Li, K. (2011). Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks. *Computer Communications*, *34*(4), 591-597.

17. Madden, S. R., Franklin, M. J., Hellerstein, J. M., & Hong, W. (2005). TinyDB: An acquisitional query processing system for sensor networks. *ACM Transactions on database systems (TODS)*, *30*(1), 122-173.

18. Mantri, D., Prasad, N. R., & Prasad, R. (2013). Grouping of clusters for efficient data aggregation (GCEDA) in wireless sensor network. 2013 3rd IEEE International Advance Computing Conference (IACC),

19. Mantri, D., Prasad, N. R., Prasad, R., & Ohmori, S. (2012). Two tier cluster based data aggregation (TTCDA) in wireless sensor network. 2012 IEEE International Conference on Advanced Networks and Telecommunciations Systems (ANTS),

20. Mantri, D. S., Prasad, N. R., & Prasad, R. (2015). Bandwidth efficient cluster-based data aggregation for Wireless Sensor Network. *Computers & Electrical Engineering*, *41*, 256-264.

21. Maraiya, K., Kant, K., & Gupta, N. (2011a). Efficient cluster head selection scheme for data aggregation in wireless sensor network. *International Journal of Computer Applications*, *23*(9), 10-18.

22. Maraiya, K., Kant, K., & Gupta, N. (2011b). Wireless sensor network: a review on data aggregation. *International Journal of Scientific & Engineering Research*, *2*(4), 1-6.

23. Massad, Y., Goyeneche, M., Astrain, J., & Villadangos, J. (2008). Data aggregation in wireless sensor networks. 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications,

24. Pourpeighambar, S. B., Aminian, M., & Sabaei, M. (2011). Energy efficient data aggregation of moving object in wireless sensor networks. 2011 Australasian Telecommunication Networks and Applications Conference (ATNAC),

25. Rajagopalan, R., & Varshney, P. K. (2006). Data aggregation techniques in sensor networks: A survey.

26. Randhawa, S., & Jain, S. (2017). Data aggregation in wireless sensor networks: Previous research, current status and future directions. *Wireless Personal Communications*, *97*(3), 3355-3425.

27. Renjith, P., & Baburaj, E. (2016). EASDAG–Energy Aware Secure Data Aggregation in Wireless Sensor Networks. *Asian Journal of Research in Social Sciences and Humanities*, *6*(11), 1273-1286.

28. Rout, R. R., & Ghosh, S. K. (2014). Adaptive data aggregation and energy efficiency using network coding in a clustered wireless sensor network: An analytical approach. *Computer Communications*, *40*, 65-75.

29. Sarangi, K., & Bhattacharya, I. (2019). A study on data aggregation techniques in wireless sensor network in static and dynamic scenarios. *Innovations in systems and software engineering*, *15*(1), 3-16.

30. Sinha, A., & Lobiyal, D. K. (2013). Performance evaluation of data aggregation for cluster-based wireless sensor network. *Human-Centric Computing and Information Sciences*, *3*(1), 1-17.

31. Tan, H. Ö., & Körpeoğlu, I. (2003). Power efficient data gathering and aggregation in wireless sensor networks. *ACM Sigmod Record*, *32*(4), 66-71.

32. Wang, P., Li, C., & Zheng, J. (2007). Distributed data aggregation using clustered slepian-wolf coding in wireless sensor networks. 2007 IEEE International Conference on Communications,

33. Wang, P., Zheng, J., & Li, C. (2007). Data aggregation using distributed lossy source coding in wireless sensor networks. IEEE GLOBECOM 2007-IEEE Global Telecommunications Conference,

34. Xu, X., Ansari, R., Khokhar, A., & Vasilakos, A. V. (2015). Hierarchical data aggregation using compressive sensing (HDACS) in WSNs. *ACM Transactions on Sensor Networks (TOSN)*, *11*(3), 1-25.

35. Yadav, S., & Yadav, R. S. (2019). Redundancy elimination during data aggregation in wireless sensor networks for IoT systems. In *Recent trends in communication, computing, and electronics* (pp. 195-205). Springer.

36. Yuea, J., Zhang, W., Xiao, W., Tang, D., & Tang, J. (2012). Energy efficient and balanced cluster-based data aggregation algorithm for wireless sensor networks. *Procedia Engineering*, *29*, 2009-2015.

37. Zheng, J., Wang, P., & Li, C. (2010). Distributed data aggregation using Slepian–Wolf coding in cluster-based wireless sensor networks. *IEEE Transactions on Vehicular Technology*, *59*(5), 2564-2574.

38. Zhou, Y., Fang, Y., & Zhang, Y. (2008). Securing wireless sensor networks: a survey. *IEEE Communications Surveys & Tutorials*, *10*(3), 6-28.