**Control and Cybernetics**

vol. **49** (2020) No. 4

# Analysis of data aggregation methods and related issues in Wireless Sensor Networks*

by

**S. Ninisha Nels and J. Amar Pratap Singh**

Department of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Kumaracoil, Thuckalay, Tamil Nadu, India
sninishanels01@gmail.com

**Abstract:** Data aggregation is the process aimed at reducing the transmission count of packets being transmitted in the framework of in-network data processing. It is the data transmission model that takes the information transmitted from different nodes and generates a single data packet after finding and eliminating the redundant packets. Accordingly, this process decreases the transmission count and makes it possible to consume less energy. The major issues in data aggregation mechanism are related to reduction of latency and to energy balancing. Moreover, it is very complex to resolve the issue of packet loss, which is the failure of one or more transmitted packets to arrive at their destination due to the bad and/or congested channel conditions. The present survey involves a collection of 50 research papers dealing with the data aggregation models in wireless sensor networks (WSN). Various data aggregation methods, like the cluster-based approach, structure-free method, tree-based approach, in-network methods, and energy based aggregation model are considered in this survey, regarding the application and the energy usage involved. On the basis of the survey, the issues and drawbacks faced by the respective methodologies are highlighted. In addition, the paper presents simple statistics of the studies considered with respect to the performance measures, simulation tools, publication year, and classification of methods. The future dimensions of the respective research are supposed to be based on the challenges identified in this survey.

**Keywords:** data aggregation, wireless sensor network (WSN), clustering, cluster head (CH), cluster-based data aggregation.

## 1. Introduction

WSN is a collection of sensor nodes such that each node has a radio that is used to perform the process of data transmission within a limited range. The sensor nodes have a battery, which has a limited power. The energy of the battery

*Submitted: February 2020; Accepted: February 2021

gets reduced at every data transmission and reception. Hence, the lifetime of nodes depends on the number of data transmission events that the nodes are involved in (see Kumar and Bharathi, 2020). Data aggregation is the approach used for wireless routing in the sensor networks, intended to integrate the data that come from various sources. The process of data aggregation helps to reduce the number of packet transmissions, removes redundancy, and enables saving of energy. The key role of data aggregation is to minimize the number of transmission events at different levels, so as to minimize the energy consumption. Data aggregation can be performed through signal processing, and can be specifically referred to, in particular, as data fusion (Dhand and Tyagi, 2016). Data fusion combines certain signals and eliminates the signal noise using some specific methods and, in the end, the accurate signal is generated (Dhand and Tyagi, 2016).

Due to the high node density, the same data may get sensed by various nodes, with resulting data redundancy. The redundant data do not offer any valuable information, but rather cause a waste of network resources. With the data aggregation process, the redundancy of data gets reduced in the transmission of the data packets from the source to the base station. Yet, due to the imperfect nature of WSN, sensor nodes are required to be densely positioned in order to increase the quality of data (Sankaralingam, Nagarajan and Narmadha, 2020). There must exist a mechanism of consolidating the data using an energy-efficient way. The adequately designed and functioning aggregation process significantly increases the performance of WSN with respect to network throughput and energy efficiency. Data aggregation is also used in some monitoring applications, like agricultural (Conti et al., 2018) or industrial applications. In general, spatial and temporal correlation of the data incurs a high degree of redundancy (Waykar, 2010). With respect to the spatial interrelations, the nodes that are spatially close to each other tend to contain much more similar data in comparison with the distant nodes. Hence, clustering of them based on the similarity and data density is important for achieving effective data aggregation. Various data aggregation techniques, for instance those using deep learning approaches (see, e.g., Arul et al., 2019) are adopted to minimize the volume of data captured from the nodes before forwarding these data to the sink node (Ullah and Youn, 2019).

This survey is focused on the review of different data aggregation methods. On the basis of the methods introduced, data aggregation techniques are categorized into the cluster-based, tree-based, structure-based, in-network ones, and the energy based ones. Also, the merits and demerits associated with the particular data aggregation methods are considered in this survey. In addition, statistical analysis is carried out based on the tools used, publication year, and performance metrics. In the future, a new approach will be proposed by using an advanced optimization algorithm, meant to overcome the challenges associated with the existing data aggregation techniques.

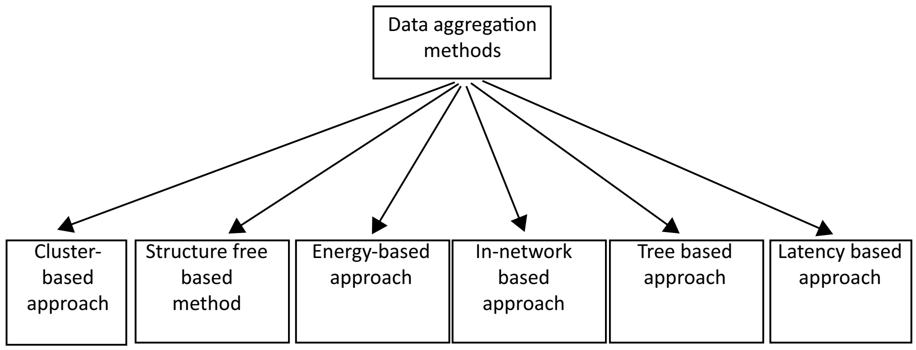


Figure 1: Categorization of data aggregation methods

The present paper is organized as follows: Section 2 formulates the motivation of the survey. Section 3 contains the proper review of data aggregation methods in WSN. Section 4 elaborates on the research gaps and issues, Section 5 provides basic statistics from the survey and finally, the conclusions are provided in Section 6.

## 2. Motivation

In WSN, data aggregation is used to gather the sensed data in order to preserve the energy, as the network functioning is realized using possibly limited energy resources. Data aggregation should provide valid data and remove the redundant data packets. The issues identified for the currently used and developed data aggregation methods motivate the researchers to design the new, energy-efficient and information-wise effective data aggregation methods.

## 3. Related works

This section describes the work on the data aggregation methods and the classification of aggregation techniques in WSN. Figure 1 shows the here utilized categorization of data aggregation methods.

### 3.1. Cluster-based data aggregation approaches

The general scheme of the cluster-based data aggregation approach is shown in Fig. 2.

In terms of the cluster-based data aggregation methods, Krishnan and Kumar (2016) introduced an effective clustering approach with multiple mobile sinks in the heterogeneous WSN to perform data aggregation. This method

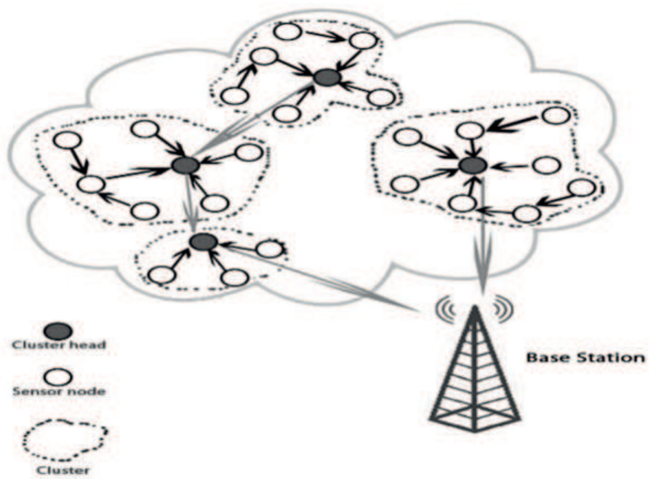


Figure 2: Model of cluster-based data aggregation approach

includes different stages for data gathering, namely tracing of the path to locate the sink, cluster formation, and the data aggregation stage. When a single node is used as a mobile sink in the data gathering process, then retracting the same path would be delayed, this leading to an increase in the latency of data transfer. The sensor region was covered by various trace paths in such a way that the trace paths crossed the region. In this approach, the nodes with high energy were used as cluster heads (CHs), these nodes receiving the data from all other nodes in the cluster. The messages were buffered and were transmitted to the sink node once the given node came into contact. This algorithm increased the network lifetime and reduced energy utilization in an effective manner. The data traffic was reduced.

Roy and Chandra (2019) developed an energy efficient scheme for the clustered WSN to improve the energy efficiency of sensor nodes. Once the sensor nodes were initialized, the nodes communicated with each other to form clusters. In this method, the process of clustering was split into two phases, namely the setup and steady phase. The method shows substantial increase in the stability period of the network and its lifetime. The method should be implemented with different clustering techniques for better performance.

Preetha and Sivakumar (2018) developed an energy-efficient sleep scheduling (EE-SS) method to increase the lifetime of the network and to preserve energy. First, the network was divided into clusters and the clusters were merged with respective CHs. CHs were determined based on a linear combination of probability-based selection and local competition. In this approach, the sleep scheduling technique was used for allocating the slots in order to send the data from the source to the base station. The nodes with the higher residual energy were selected as forwarding nodes in such a way that the nodes with the maximal residual energy improved the performance of the network.

Devi, Ravi and Priya (2020) introduced a cluster-based data aggregation mechanism to reduce packet loss. The mechanism consists of two phases, namely construction of aggregation tree and slot scheduling. Each CH uses the compressive aggregation in phase 1 to receive the data from the cluster members. The aggregation tree is constructed by the sink node on the basis of the minimum spanning tree (MST). The latency and the packet loss rate are considered in phase 2 during prioritizing and allocation of time slots to the nodes using aggregated data. The method developed increased the packet delivery ratio and reduced the overhead and latency. It optimized energy consumption and reduced the unnecessary waiting time.

Sasirekha and Swamynathan (2017) developed a cluster chain mobile agent routing (CCMAR) method to increase the performance of the network. CCMAR organized the WSN into clusters and operated in two different phases. The nodes of a cluster were used to form a chain within the cluster to achieve data aggregation in the first phase. A mobile agent was dispatched from the sink node for collecting aggregated data from the chain or CH. Simulations were carried out with respect to network lifetime, transmission delay, and energy consumption. According to the results, the mobile agent offered improved data dissemination.

Ullah and Youn (2020) introduced a data aggregation model based on the extreme learning machine (ELM) and node clustering to minimize erroneous and redundant data. In this method, the instability of the training process was reduced owing to the use of the Mahalanobis distance with a radial basis function. Here, the Kalman filter was used for filtering the data at the sensor node before transmitting the data to CH. The network traffic was reduced and the network's lifetime was increased.

Kamble and Dhope (2016) proposed a Velocity Energy-effective and Link-mindful Cluster-Tree (VELCT) technique, which uses the data collection nodes for collecting data from the CHs inside the clusters, and improves the lifetime of the wireless network by forwarding data in an aggregated format and finding the attacked node and the CH. This method increased the network performance and the accuracy of data transmission.

Shobana, Sabitha and Karthik (2020b) introduced a cluster-based systematic data aggregation method, meant to reduce energy consumption and delay in data transmission. Initially, the network was grouped into clusters and active nodes using sleep state and next the CH was selected on the basis of the ranking method specified for the sensor nodes using two criteria, namely the existing energy level and geographic location.

Zhou et al. (2019) introduced efficient privacy and integrity preserving data aggregation (PIMA) approach for various applications in WSN. In this approach, the homomorphic Paillier encryption was used to protect the data privacy and to verify the integrity of data aggregation. PIMA was used to aggregate the data from hybrid sensors more efficiently and extract the application-specific data from the aggregated cipher texts using the Chinese remainder theorem (CRT). The injected false data were effectively filtered in the aggregator. This model reduced computational cost and communication overhead.

Lakshmi, Velmani and Rani (2019) developed a priority hop based energy-efficient cluster routing (PHECR) algorithm for evaluating the pollution level with air quality data. The lifetime of the network was increased and the resource consumption of the sensor nodes was reduced. This method was used to transfer any category of aggregated data.

Subedi, Lee and Lee (2018) introduced a modified low energy adaptive clustering hierarchy (LEACH) for organizing a large number of sensor nodes and saving the communication resources, in association with the initiation time and the processing time. The efficiency was significantly dependent on the formation of the cluster and the selection of CH.

SriVenkateswaran and Sivakumar (2019) introduced a cluster-based data aggregation model to select the CH for aggregating the data in the sensor networks. To implement the clustering process, the fuzzy c-means and the k-means algorithms were used in this method. The CH was selected from each cluster, and each node transmitted the sensed data to the CH. Here, data aggregation was carried out at the CH. The duplicated data were removed and secure transmission was carried out using elliptic curve cryptography (ECC).

Kadlikoppa, Umarji and Patil (2017) introduced a spin protocol for aggregating the data, generated by sensor nodes. The protocol offered effective data transmission to the base station or sink node. It increased the network's lifetime and saved energy.

Mantri, Prasad and Prasad (2016) introduced a mobility and heterogeneity aware cluster-based data aggregation (MHCDA) to efficiently utilize the bandwidth and to increase the lifespan of the network. This method used the predefined regions for aggregating the packets at CH so as to minimize the communication- and computation-related costs. By using the variable packet generation rate, MHCDA exploited correlation of data packets generated by nodes to reduce energy consumption.

Acharya and Tripathy (2020) developed a reliable fault-tolerant mechanism for aggregating the sensor nodes in the sensor network. This method achieved improved performance in terms of loss probability, detection accuracy, false alarm rate, and energy cost. The gateway nodes were entirely free from unnecessary hassle in storing the cryptographic keys. Moreover, the communication overhead was reduced and the energy was saved.

Otoum, Kantarci and Mouftah (2018) developed an adaptively supervised and clustered model for monitoring the critical infrastructures in the connected sensor clusters. It continuously monitored the behavior of operating characteristics of the receiver and directed the incoming packets to the sensor cluster.

Ren et al. (2017) introduced a cluster-based distributed data aggregation algorithm for reducing the latency of data aggregation in the multi-power and multi-channel WSN. It used less transmission power for transmitting the data packets inside the cluster, whereas more power was used to transmit the data packets between the clusters. This method achieved lower latency and avoided the conflicts using multi-channel WSN.

Abdelhafidh et al. (2018) introduced a bio-inspired clustering and data aggregation algorithm for increasing the lifetime of the network. The k-means algorithm was used to partition the network into a number of clusters, so that the best cluster was selected based on the shortest distance between CH and cluster members. Due to the elimination of redundancy, the energy consumption was reduced.

Zhang et al. (2017) introduced a link based privacy preservation method for aggregating the data. In this method, the data link was formed on the basis of distance and energy consumption. The nodes within the cluster realized the process of data aggregation using CH. Based on the data from the link, the information matrix was generated using CH and performed the homomorphic transformation.

Kamalesh and Kumar (2017) introduced the data aggregation mechanism based on the support vector machine (SVM) to perform loss recovery and failure detection. Based on connectivity, the nodes were partitioned into clusters using the location information. For each cluster, the nodes with maximum connectivity were selected as CH. The aggregator identified the node failure by finding the faulty data with SVM. This method reduced transmission overhead and increased reliability.

Wan et al. (2019) introduced a similarity based data aggregation model using the fuzzy c-means mechanism for clustering and detection of anomalous events. The sensors were organized into clusters by fuzzy c-means. In order to diagnose

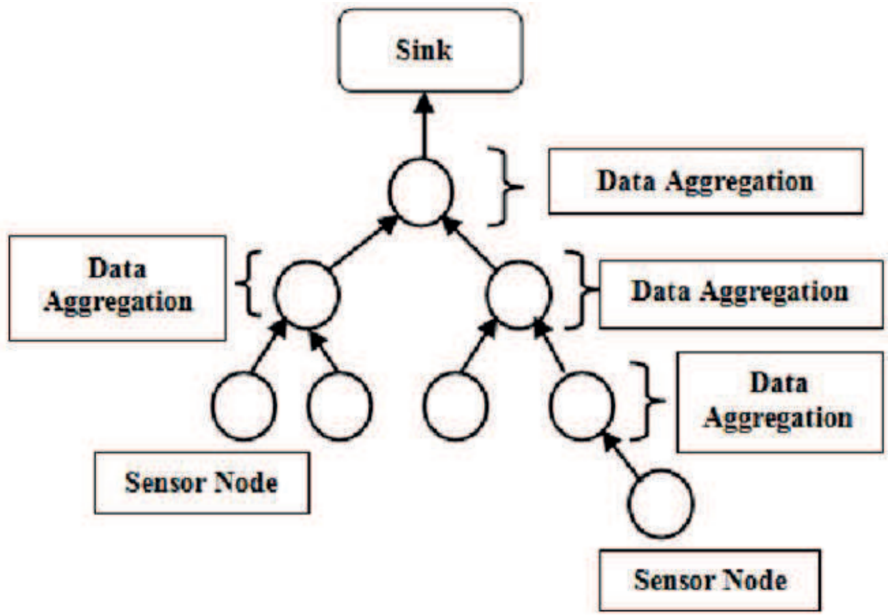


Figure 3: The scheme of the structure free methods

the outliers, support degree function was calculated and the weight of data was obtained using probability distribution characteristics. The performance of this method was evaluated in terms of relative recovery error and outlier detection accuracy.

### 3.2. Structure free methods

Figure 3 presents the general scheme of the structure free methods.

Of the approaches from this category, Mohanty and Kabat (2016) developed an energy-efficient structure free data aggregation and delivery (ESDAD) mechanism for aggregating the redundant information with intermediate nodes. In this mechanism, the waiting time of packets at each neighboring node was computed more accurately, so that the data were aggregated more efficiently in the routing path. The sensed data were transmitted for aggregation to the aggregation point. The mechanism calculates the cost function for the next-hop and structure-free node selection to perfrm the data aggregation process. The buffer at each node is partitioned for maintaining the data flow in order to offer efficient and fair data delivery. The transmission rate was adjusted at the time of congestion. This method results in the reduced energy consumption, reduced miss ratio and end-to-end delay.

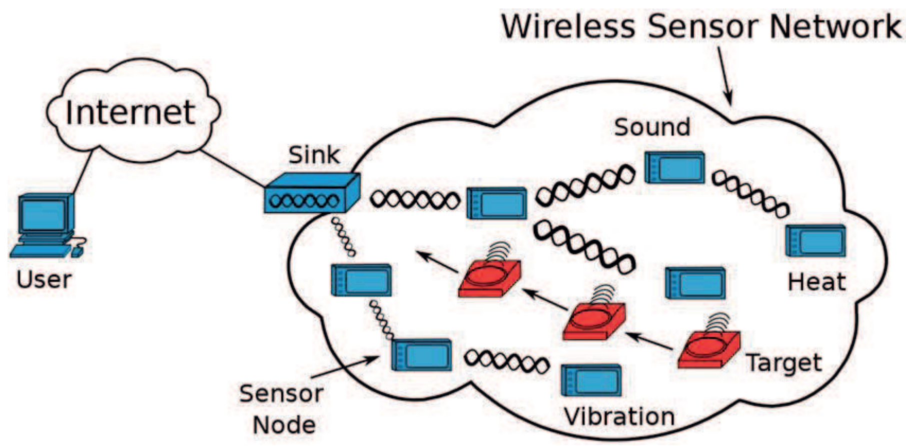


Figure 4: The scheme of the energy based data aggregation methods

Chen et al. (2017) developed a distributed structure free data aggregation approach to generate the conflict free schedule and aggregation tree. In this approach, the utilization ratio of the time slots and aggregation latency were increased.

### 3.3. Energy based data aggregation methods

The general scheme of the energy based data aggregation methods is provided in Fig. 4.

A selection of the energy based data aggregation methods is discussed in this section. Thus, Hua et al. (2018) introduced an energy-efficient Adaptive Slice based secure data aggregation (ASSDA) method to increase the performance of the network under the resource limitations of the nodes. The method includes the construction of an aggregation tree, mixing, slicing, assembling, and finally aggregation. It increased the efficiency of data slicing, minimized the consumption node energy, increased the lifetime of the network, and maintained effective privacy preservation at the same time.

Zhang et al. (2020) introduced an entropy-driven data aggregation method using gradient distribution (EDAGD) to deploy the sensor nodes with the aim of increasing the lifetime of the network. By increasing the lifetime of the network, it reduced the energy hole problem and preserved energy. The data transmission distance was reduced by minimizing the number of hops needed in the transmission path. The entropy-driven model was used to examine the abnormal events using entropy and Choquet integral.

Kumar and Bharathi (2020) introduced an energy-efficient data aggregation and scheduling approach in order to increase the throughput and network lifetime. In this approach, the nodes in the network were scheduled to perform their operations based on such parameters as lifetime, throughput and availability. Based on these factors, the approach estimated lifetime support, throughput support, and data availability support measures, respectively. With these measures, the data aggregation support measure (DAGS) was computed. Based on the estimated measures, the nodes were scheduled for performing the operations. This method achieved prolonged lifetime and improved data aggregation performance.

Wang et al. (2019) introduced a minimum energy encrypted data aggregation (EMEDA) method for reducing the overhead of data encryption. In this method, the data length of ciphertext at each sensor node was reduced using the data length of scheduling quantization of analog/digital (A/D) converter. A heuristic model was used to reduce the overhead caused by the data encryption using the Fibonacci search. Between the sensor node and Edge Computing Node (ECN) the communication overhead is reduced.

Sankaralingam, Nagarajan and Narmadha (2020) introduced an energy-aware decision stump linear programming boosting node classification (EAD-SLPBNC) in WSN. First, the residual energy was computed in order to perform data classification. To increase the margin between training samples of various classes, classification is carried out. The nodes with higher energy and the nodes with lower energy were classified so as to minimize the misclassification error. Upon calculating the distance the lesser energy sensor nodes in WSN transmit the data packets to the neighboring higher energy sensor nodes.

Prabhavathi, Subramanyam and Rao (2016) introduced an energy-aware routing model to increase energy efficiency. Due to the unique routing tactics, this method improves energy consumption and load balancing. Here, the routing agent was used to perform energy oriented optimization and improved energy efficiency was achieved.

Manishankar, Ranjitha and Kumar (2017) introduced an energy-efficient technique for aggregating the data transmitted from the sensor node to the data node. The distance between the data node and the sensor node was computed to re-cluster the nodes based on the remaining energy of nodes. The method effectively reduced energy consumption.

Ashtikar, Javale and Wakchaure (2017) developed an energy-efficient and secure routing approach for detecting the attacks present in the nodes, including aggregator node, CH, and cluster members. The method verified the trustworthiness of data at the CH when the information was received from the cluster member. A hashing algorithm was used to perform integrity checking. This method increased data security, improved memory management and energy utilization.

Jasim et al. (2019) introduced a secure and energy-efficient data aggregation model (SEEDA) using an access control mechanism to solve the energy and security issues. This protocol was intended to improve authentication using a random timestamp and random value based on the secret key. The potentially fake data were verified by the base station after receiving the packets using the secret key. This protocol reduced energy consumption and communication overhead.

Yoon, Kim and Noh (2017) introduced an adaptive data aggregation method for increasing efficiency of energy utilization in the solar-powered WSN. In this method, each node determined the energy budget periodically by considering the residual energy. The method reduced the number of blackout nodes.

Zhou et al. (2019) proposed an energy-efficient and privacy-preserving data aggregation approach to ensure the security of sensor data. In this approach, the network was organized as a tree in such a manner that the leaf nodes of the tree formed various chains. For ensuring privacy, the data sensed by the tail nodes were sliced. This method increased the network's lifetime and reduced energy consumption.

Shobana, Sabitha and Karthik (2020a) developed an enhanced soft computing model to achieve secure data aggregation in WSN. First, the network was partitioned into clusters in such a way that the CHs acted as the aggregators and were selected using the fuzzy if-then rule that was applied to make energy consumption possibly small. The method offered data confidentiality, maintaining the aggregation function using fully homomorphic encryption (FHE). In this method, the key generation along with the public key compression was used for condensing the dimensions of the public key. The data integration was performed by using the message authentication code. This method effectively increased the rate of data transmission and reduced energy utilization.

Latha et al. (2019) introduced a trust assisted energy efficient aggregation (TE-EEA) scheme to increase the aggregation precision using limited constraints in aggregation and neighbor reliability. The scheme ensured the selection of trusted neighbors and exploited the tradeoff between security and energy.

### 3.4. The in-network methods

The general scheme of the in-network methods is presented in Fig. 5.

Concerning the in-network type of aggregation methods, Zhang et al. (2018a) introduced a ring based data aggregation scheme to increase energy efficiency, to minimize transmission delay, and also to guarantee transmission reliability. In this scheme, the network was partitioned into rings in such a manner that

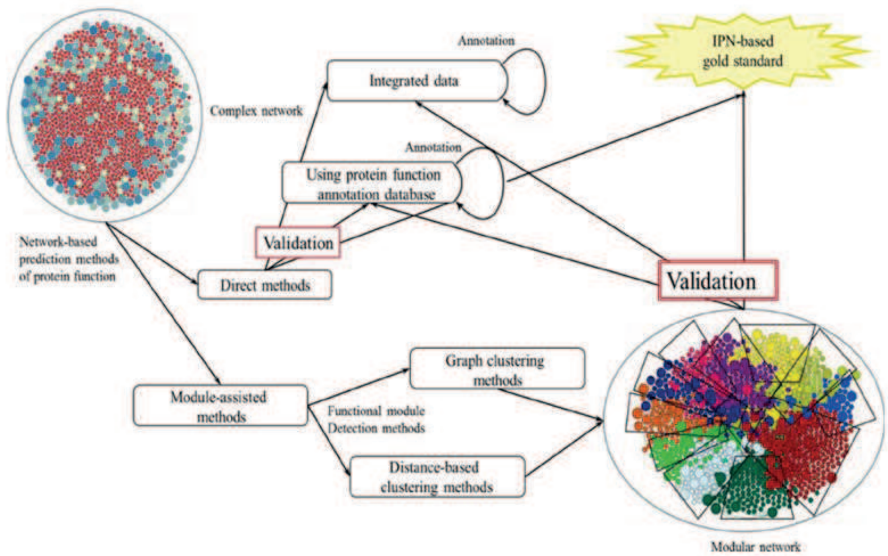


Figure 5: The general scheme of in-network methods

data aggregation was executed from the outside to the inside, ring by ring. With higher residual energy the intermediate node unicasted a number of aggregated packets to the next-hop node. With the higher numbers of unicasted packet copies, the reliability was increased. Yet, sending of more packet copies led to increased energy cost. In this approach, the variable number of packets was inferred by the fuzzy logic system for achieving efficient and effective data aggregation.

### 3.5. Tree-based data aggregation methods

Figure 6 shows the general scheme of the tree-based data aggregation methods.

As regards the tree-based data aggregation methods, Wang, Zhou and Cheng (2019) developed a topology-aware data aggregation (TADA) protocol to improve the efficiency of data aggregation in WSN. The protocol encoded the raw data of nodes using the weight vector and decoded the data at the sink node using the measurement matrix that was designed with the weight vectors. Due to generation of the measurement matrix using the topological structure of WSN, the reconstruction error rate was minimized.

Yestemirova and Saginbekov (2018) proposed a data aggregation method for aggregating the data packets sent from the sensor nodes to the sink nodes in WSN. This method was used to reduce redundant data packets. It effectively solved the data aggregation issues with a limited number of data transmissions.

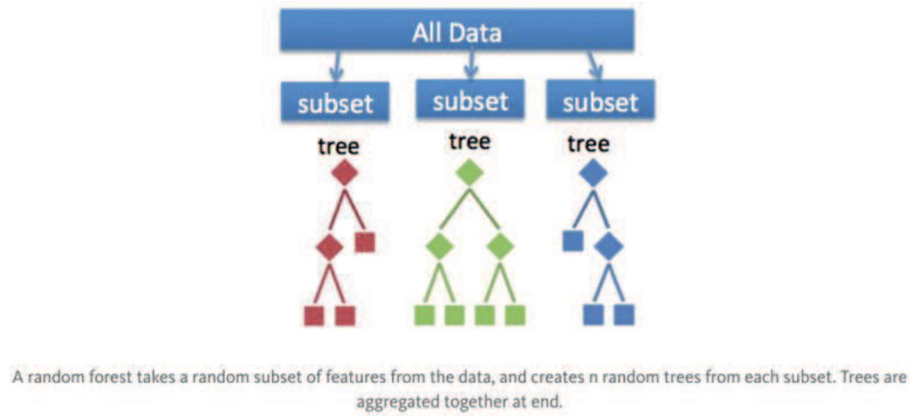


Figure 6: The general scheme of the tree-based data aggregation methods

Atoui et al. (2016) developed a tree-based data aggregation approach using fitting functions to aggregate the data. In this approach, only a part of data was forwarded to the sink node from the aggregator. Due to limited data transmission, energy was saved effectively. The accuracy of performance was measured using the Bayesian belief network.

Jothiprakasam and Muthial (2018) introduced a data aggregation method based on the so called ribbon structure. Data aggregation was divided into two cycles, referred to as the right and left cycles. This method was implemented in the hierarchical topology. It is also used in event detection application. In effect, the network's lifetime was increased.

### 3.6. Latency-based data aggregation methods

Nguyen et al. (2018) introduced an independent set based collision avoidance scheduling (ISBCAS) method for aggregating data into single packets. This method reduced latency and avoided data collision using the relative collision graph. In this approach the network structure was represented by an undirected weighted graph.

### 3.7. Other data aggregation methods

We consider in this section some other methods used for aggregating the data in WNS. Thus, Wang, Zhou and Tong (2019) introduced a Vandermonde matrix-based scalable data aggregation (VSDA) approach to preserve the benefits of a coding based mechanism. In this approach, a new node can join the scaled-up network such that the original weight vectors, activated by the nodes, do not require the regeneration of these vectors (a new node can join the scaled-up network without regeneration). A concise hardware model [VSDA] was designed for quantifying the process of data encoding. The weight vectors exhibited improved scalability, thereby helping to solve the issues caused by network expansion.

Ullah and Youn (2019) introduced a data aggregation method, which used a self-organizing map for reducing the redundant information and for removing the outliers. Cosine similarity was used in the process of clustering of sensor nodes regarding the similarity and density of data. In this method, the interquartile analysis was applied to eliminate the outliers. The method effectively increased the performance of the network and minimized energy consumption.

Jain and Bhola (2018) developed an efficient routing algorithm to increase the lifetime of the network. In this algorithm, the amount of data transmitted to the base station was limited, which reduced the data aggregation load, and thereby the overall performance was enhanced.

Balakrishnan, Vijayalakshmi and Vinayagasundaram (2016) introduced an enhanced iterative filtering method to perform data aggregation in the sensor network. This method is processed by using different techniques like averaging, iterative filtering for secure information agregation. This method is accurate and faster converging.

Nayaka and Biradar (2017) introduced an efficient data aggregation and routing mechanism, featuring enhanced efficiency of packet formatting using a statistical approach. In this mechanism, the smart utility data were used for identifying the network failures and for understanding the patterns that indicate fault, leakage, or theft. The method increased the lifetime of the network and minimized the computational load as well as transmission latency.

Choudhari, Bodhe and Mundada (2017) introduced an iterative filtering algorithm for aggregating the data concurrently from various sources. It was used for rendering the trust assessment, which took the form of weight factors. The method was fast converging and collusion robust. It was tested with the unbiased and biased nodes.

Metan and Murthy (2018) introduced a secure data aggregation method to identify the external and internal attacks in WSN. The public key was generated in this method using the elliptic curve cryptosystem and this resulted in a recursive process that led to an increase in the message size. The method effectively balanced communication performance and security demands. Energy consumption, delay and computational burden were all decreased owing to this method, which also increased the performance of the process.

Zhong et al. (2018b) introduced a secure data aggregation approach based on the homomorphic encryption and signature technique. In this case, energy consumption for unnecessary transmission was reduced by using data filtering. The method ensured the authorized aggregation for identifying the origin and for validating the received messages. The base station recovered the original data and performed aggregation operations. The results showed that the method reduced communication overhead, energy consumption, delay, and computational overhead.

Table 1 shows the objectives, which were encountered in the papers surveyed, related to the assessment of the introduced methods and approaches.

## 4. Research gaps and issues

We discuss in this section the research gaps and issues faced by the data aggregation methods surveyed.

The cluster-based data aggregation methods face the following issues and challenges: The security and the fault tolerance are the major problems for the mobile agent techniques. In particular, the security issues at both the agent and node levels were not investigated by Sasirekha and Swamynathan (2017). Data aggregation at the CH can be efficient only when similar data are grouped and processed together, which is not done in the Mahalanobis distance-based radical basis function (MDRBF) referred to as MELM. This method has still to be tested with complex datasets and operation condition, see Ullah and Youn (2020). Data aggregation for WSN in MDRBF failed to provide security in the system for information authentication during data transmission (Kamble and Dhope, 2016). The OSM-EFHE method fails to provide privacy preserving data security in data management meant to improve the real time implementation (Shobana, Sabitha and Karthik, 2020a). The aggregation algorithm in Zhou et al. (2019) failed to cover the wide range of aggregate queries in terms of consideration of the integrity preserving data aggregation scheme in the context of a strong adversarial model. The MHCDA method in Mantri, Prasad and Prasad (2016) fails to consider the parameters like mobility and heterogeneity of both the nodes and the sink for improving the connectivity with IoT. The deterministic WSN cluster deployment model fails to evaluate the reliability of different random cluster deployment techniques for different sizes of WSNs. In order to make this method more robust and application-oriented, the cluster deployment model have to tested in the context of real WSN application (Acharya and Tripathy, 2020). The Adaptively Supervised and Clustered Hybrid IDS (ASCH-IDS) failed to account for the impact of cluster sizes on the performance of data aggregation (Otoum, Kantarci and Mouftah (2018). The link-based privacy-preserving data aggregation scheme (LPDA) failed to provide data integrity and did not form the data link based on residual energy (Zhang et al., 2017). The data aggregation method, proposed in Wan et al. (2019) failed

Table 1: Comparison of literature references regarding objectives of the study

| *No.* | *Reference* | *Objective(s)* |
|---|---|---|
| 1 | Roy & Chandra (2019) | To improve energy efficiency of the sensor nodes |
| 2 | Chen et al. (2017) | To generate the conflict free schedule and aggregation tree |
| 3 | Kumar & Bharathi (2020) | To increase the throughput and network lifetime |
| 4 | Zhang et al. (2018) | To increase energy efficiency<br>To minimize the transmission delay<br>To guarantee transmission reliability |
| 5 | Yestemirova & Saginbekov (2018) | To reduce the redundant data packets |
| 6 | Nguyen et al. (2018) | To reduce latency and avoid data collision |
| 7 | Balakrishnan, Vijayalakshmi & Vinayagasundaram (2016) | To achieve data aggregation in the sensor network |
| 8 | Abdelhafidh et al. (2018) | To minimize energy consumption and reduce the amount of transmitted data |
| 9 | Acharya & Tripathy (2020) | To secure better performance and longer network lifetime |
| 10 | Arul et al. (2019) | To improve the SE performance of CI users |
| 11 | Ashtikar, Javale & Wakchaure (2017) | To improve memory management, energy utilization and security of data, together with attack detection |
| 12 | Atoui et al. (2016) | To save energy with less of data transmission and improve accuracy |
| 13 | Choudhari, Bodhe & Mundada (2017) | To check the trustworthiness of data and the reputation of the sensor nodes |
| 14 | Conti et al. (2018) | To mitigate the conflict between the public bodies in management of wildlife and the farmers |
| 15 | Devi, Ravi & Priya (2020) | To enhance network performance |

Table 1, continued (part 2)

| | | |
|---|---|---|
| 16 | Dhand & Tyagi (2016) | To study the importance of data gathering and compare various hierarchical clustering methods |
| 17 | Hua et al. (2018) | To improve the efficiency of data slicing and reduce energy consumption of nodes |
| 18 | Jain & Bhola (2018) | To classify the data aggregation design goals |
| 19 | Preetha & Sivakumar (2018) | To minimize energy dissipation by the redundant data |
| 20 | Sasirekha & Swamyanathan (2017) | To check LEACH, PEGASUS and other similar routing algorithms, and the energy efficient cluster chain based protocol |
| 21 | Ullah & Youn (2020) | To reduce the redundant and erroneous data efficiently |
| 22 | Kamble & Dhope (2016) | To identify the cluster head and the attacked node |
| 23 | Shobana, Sabitha & Karthik (2020b) | To minimize energy consumption and transmission delay, and increase network lifespan |
| 24 | Zhou et al. (2019a) | To protect data privacy and to check the aggregate data integrity |
| 25 | Lakshmi, Velmani & Rani (2019) | To evaluate the pollution level with air quality data from the network |
| 26 | Subedi, Lee & Lee (2018) | To increase the network lifetime |
| 27 | SriVenkateswaran & Sivakumar (2019) | To remove the duplicate data and make transmission secure |
| 28 | Kadlikoppa, Umarji & Patil (2017) | To increase network lifetime and improve data gathering |
| 29 | Mantri, Prasad & Prasad (2016) | To exploit correlation of data packets with a variable packet generated by nodes |
| 30 | Otoum, Kantarci & Mouftah (2018) | To detect malicious behavior in the sensor network |
| 31 | Ren et al. (2017) | To ensure lowest average latency |

Table 1, continued (part 3)

| | | |
|---|---|---|
| 32 | Zhang et al. (2017) | To protect data privacy and to reduce energy consumption and computational overhead |
| 33 | Kamalesh & Kumar (2017) | To minimize transmission overhead and increase reliability |
| 34 | Wan et al. (2019) | To improve performance in terms of data outlier detection and to minimize relative recovery error |
| 35 | Mohanty & Kabat (2016) | To improve energy efficiency, on-time delivery ratio and reliability |
| 36 | Zhang et al. (2020) | To develop sensors network aiming at maximizing the network lifetime |
| 37 | Wang et al. (2019) | To reduce the overhead from data encryption and to minimize the cipher text data length |
| 38 | Sankaralingam, Nagarajan & Narmadha (2020) | To reduce energy consumption and aggregation time and increase the accuracy |
| 39 | Prabhavathi, Subramanyam & Rao (2016) | To obtain improved energy efficiency the within minimal data aggregation duration |
| 40 | Manishankar, Ranjitha & Kumar (2017) | To perform energy efficient data aggregation |
| 41 | Yoon, Kim & Noh (2017) | To reduce the number of nodes forced to black out |
| 42 | Jasim et al. (2019) | To enhance authentication by generating a random timestamp and a random value with a secret key |
| 43 | Shobana, Sabitha & Karthik (2020a) | To ensure secure data aggregation using the energy efficient data processing in large-scale networks |
| 44 | Latha et al. (2019) | To improve the overall aggregation precision with limited constraints |

Table 1, continued (part 4)

| | | |
|---|---|---|
| 45 | Wang, Zhou & Cheng (2019) | To verify the routing protocols |
| 46 | Jothiprakasam & Muthial (2018) | To reduce energy consumption and to enhance the energy saving modes |
| 47 | Wang, Zhou & Tong (2019) | To validate the performance in terms of the number of transmissions, energy consumption, and storage space |
| 48 | Ullah & Youn (2019) | To enhance network performance |
| 49 | Nayaka & Biradar (2017) | To increase the lifetime of integrated smart utility grid and to improve public utility usage efficiency |
| 50 | Metan and Murthy (2018) | To develop a novel key management system |

to consider the scheduling of links efficiently without any empty packets and the arrangement of sensor nodes to increase the charging efficiency, and data collection volume is not discussed.

Concerning the structure free methods let us mention that the one from Mohanty and Kabat (2016) was not tested and evaluated in the real time dynamic environment.

The energy based approaches can be said to face the following issues. First, the privacy preservation and the information of the edge computing nodes were not considered in the work of Wang et al. (2019) on Energy Minimum Encrypted Data Aggregation (EMEDA). The study of Sankaralingam, Nagarajan and Narmandha (2020) fails to reduce the redundancy during the data aggregation process. The data were not offloaded to the centralized storage systems in Manishankar, Ranjitha and Kumar (2017). The data aggregation model from Nayaka and Biradar (2017) failed to minimize message cost and to increase failure tolerance. The aggregation protocol, proposed in Jasim et al. (2019) failed to prevent the attacks and solve the security issues while adding mobile nodes to the network. Due to insufficient energy, some data may be lost by the forwarding nodes in the approach proposed by Yoon, Kim and Noh (2017). The soft computing model as proposed by Shobana, Sabitha and Karthik (2020a) was not validated with the real time implementations and actually failed to increase data security.

The research gaps and issues faced by the tree-based methods are the following ones: the wireless caching model was not incorporated into the TADA approach from Wang, Zhou and Cheng (2019) for increasing security and transmission reliability. Concerning the work of Yestemirova and Saginbekov (2018), the method fails to incorporate a node failure tolerance algorithm and the experiments on real testbeds for high number of packet transmission are missing. The approach of Atoui et al. (2016) failed to minimize the number of transmitted parameters by analyzing the correlation between the variables. The method from Jothiprakasam and Muthial (2018) fails to implement a distributed scheduling algorithm for WSN.

No efficient dimensionality reduction and classification algorithm was employed in Shobana, Sabitha and Karthik (2020b) to increase the performance of data aggregation. The CSDAM approach from Wang, Zhou and Tong (2019) faces a major challenge for implementing the data aggregation mechanism on actual hardware. Zhang et al. (2018) in their proposed approach failed to evaluate the performance using membership function and fuzzy output space. The proposal from Ashtikar, Javale and Wakchaure (2017) failed to recover the compromised node and use it in the network rather than leaving the node. The iterative filtering model from Chen et al. (2017) did not satisfactorily enhance the security of data aggregation.

## 5. Survey statistics

This section provides simple statistics on the present survey of the data aggregation methods for WSN, based on the performance metrics, simulation tool, publication year, and data aggregation methods.

### 5.1. Statistics based on data aggregation methods

Figure 7 shows the distribution of data aggregation techniques in the surveyed sample according to the here adopted categorisation. The diagram shows that the cluster-based data aggregation methods are used in 21 research papers, while the energy-based approaches are used in 13 research papers. The remaining kinds of approaches, that is - the tree-based ones, the structure-free ones, in-network and latency-based ones appear in much smaller numbers of papers, namely, respectively, in 4, 2 and then the last two in just a single paper each.

### 5.2. Statistics based on simulation tool

Figure 8 shows the distribution of the papers here surveyed according to the simulation tools used to analyze the performance of the data aggregation process. As the diagram in this figure states, the network simulator-2 (NS-2) was used in ten research papers, being followed by the MATLAB software library, used in seven research papers. Then, we have Java, used in three research papers, and the network simulator-3 (NS-3), appearing in two papers reviewed. Other tools, that is - Python, Castalia, OriginPro, QuaNet, and TinyOS (TOSSIM) - are used each in only a single research work here considered.

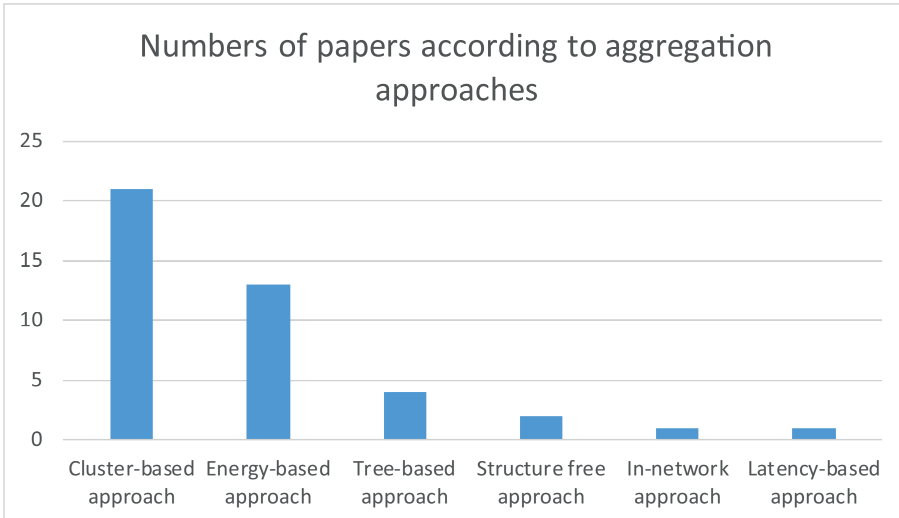


Figure 7: Distribution of papers surveyed according to the data aggregation techniques used in the them

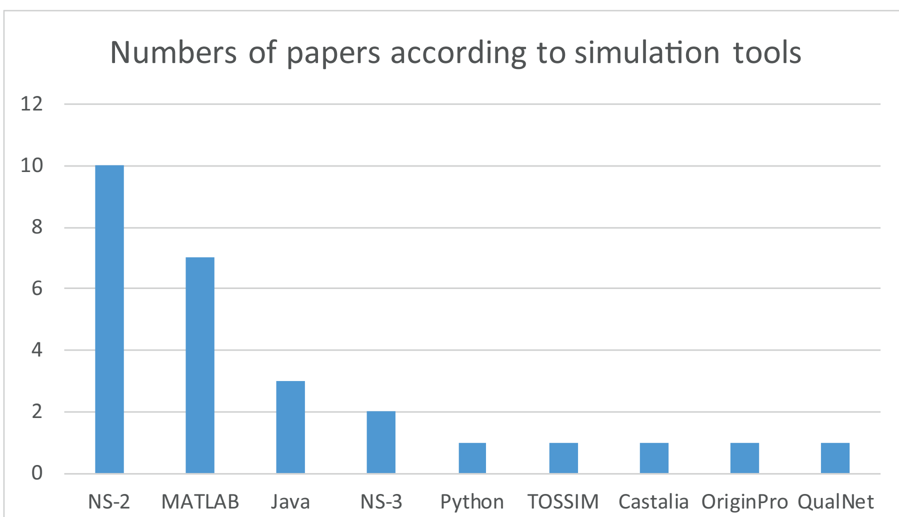


Figure 8: Distribution of the research papers surveyed based on the simulation tools used

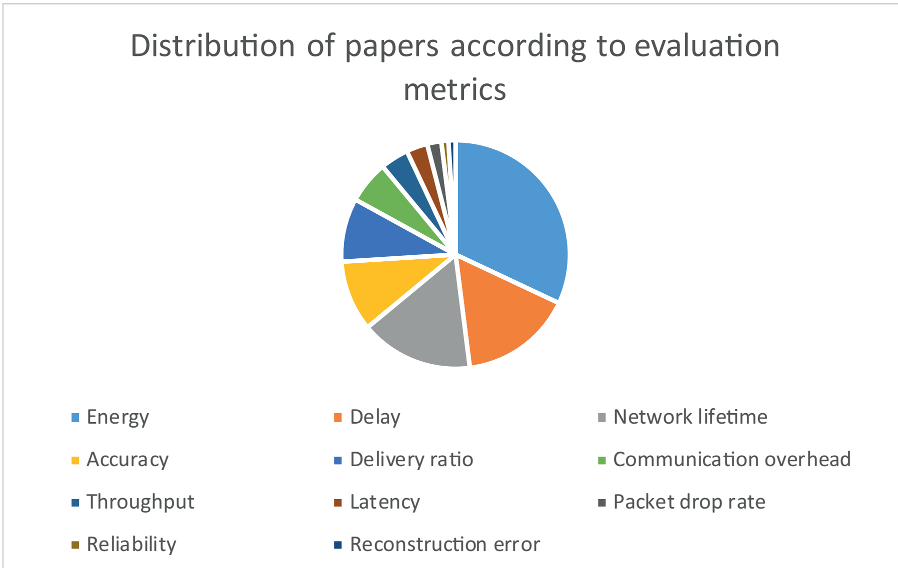


Figure 9: Distribution of papers surveyed based on evaluation metrics

### 5.3. Statistics based on evaluation metrics

Figure 9 presents the distribution of the papers surveyed according to the evaluation metrics used in them to show the performance of the respective data aggregation methods. In an obvious correlation with the methods proposed and developed, commented upon in Section 5.1 (Fig. 7) the biggest number of the research papers used energy metrics for analyzing the performance. The energy metric is followed by delay, network lifetime, accuracy, delivery ratio and communication overhead, each of these having been used by several studies among those reviewed. Then, the remaining metrics, like reconstruction error, latency, packet drop rate, throughput and reliability are used only marginally, some of them in just a single study here reviewed.

### 5.4. Statistics based on publication year

Figure 10 displays the diagram, showing the numbers of papers according to their publication year. It can be easily seen that we deal here, indeed, with quite an up-to-date review. Namely, no paper dating back earlier than the year 2016 was considered, and the vast majority of the surveyed papers come from the years 2017-2019, with quite a proportion of those from the year 2020. Given the timing of this survey, this certainly demonstrates its current and valid character.

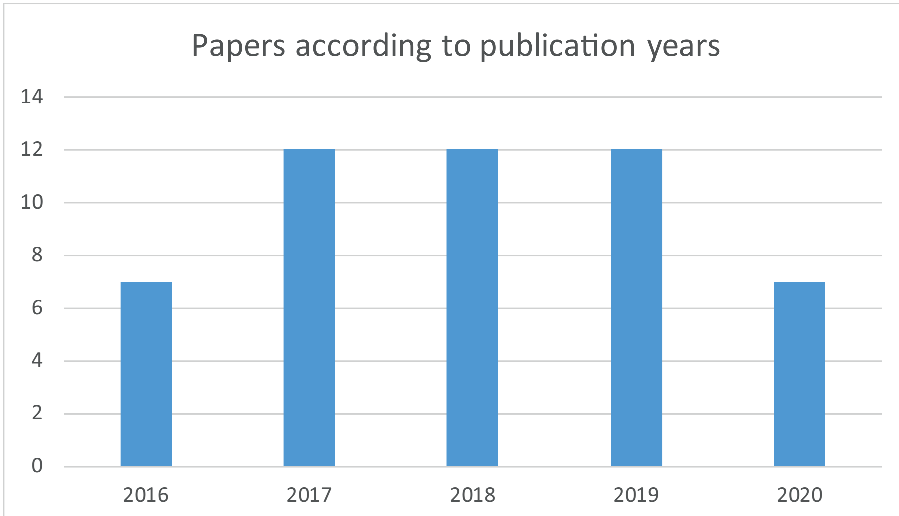


Figure 10: Distribution of papers surveyed based on publication year

## 6. Conclusion

Data aggregation in WSN is the process of gathering data from the sensor nodes to be sent to the sink node. WSNs contain a large number of sensor nodes, which are capable of sensing and forwarding the data on target events. In this paper, a detailed survey is reported, performed on the basis of 50 research papers, among which 21 concern the cluster-based data aggregation methods, 13 concern the energy-based data aggregation methods, four refer to the tree-based aggregation methods, while two research papers are based on the structure free methods. We provide the survey along with the specificaton of the merits and demerits of the techniques, proposed and presented in these research papers. The papers, used in this survey, have been collected from different sources, such as Science Direct, Google scholar, IEEE, and so on. Finally, we also present simple statistics of the papers surveyed, based on the simulation tool used, performance metrics, publication year, and the categorization of data aggregation methods.

## References

Abdelhafidh, M., Fourati, M., Fourati, L.C., Mnaouer, A.B. and Zid, M. (2018) Lifetime maximization for pipeline monitoring based on data aggregation and bio-inspired clustering algorithm. In: *IEEE 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, 666-671.

Acharya, S. and Tripathy, C.R. (2020) A reliable fault-tolerant ANFIS model based data aggregation scheme for Wireless Sensor Networks. *Jour-*

*nal of King Saud University-Computer and Information Sciences*, **32**(6), 741–753.

Arul, V.H., Sivakumar, V.G., Marimuthu, R. and Chakraborty, B. (2019) An Approach for Speech Enhancement Using Deep Convolutional Neural Network. *Multimedia Research (MR)*, **2**(2): 37-44.

Ashtikar, R., Javale, D. and Wakchaure, S. (2017) Energy Efficient & Secured Data Routing Through Aggregation Node in WSN. In: *IEEE International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, 1-6.

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

Balakrishnan, C., Vijayalakshmi, E. and Vinayagasundaram, B. (2016) An enhanced iterative filtering technique for data aggregation in WSN. In: *IEEE International Conference on Information Communication and Embedded Systems (ICICES)*, 1-6.

Chen, Q., Gao, H., Cheng, S., Li, J. and Cai, Z. (2017) Distributed non-structure based data aggregation for duty-cycle wireless sensor networks. In: *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, 1-9.

Choudhari, E., Bodhe, K.D. and Mundada, S.M. (2017) Secure data aggregation in WSN using iterative filtering algorithm. In: *IEEE International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 1-5.

Conti, L., Bartolozzi, S., Racanelli, V., Guerri, F.S. and Iacobelli, S. (2018) Alarm guard systems for the prevention of damage produced by ungulates in a chestnut grove of Middle Italy. *Agronomy Research*, **16**(3).

Devi, V.S., Ravi, T. and Priya, S.B. (2020) Cluster Based Data Aggregation Scheme for Latency and Packet Loss Reduction in WSN. *Computer Communications* 149, 36-43.

Dhand, G. and Tyagi, S.S. (2016) Data aggregation techniques in WSN: Survey. *Procedia Computer Science*, 92, 378-384.

Hua, P., Liu, X., Yu, J., Dang, N. and Zhang, X. (2018) Energy-efficient adaptive slice-based secure data aggregation scheme in WSN. *Procedia Computer Science*, 129, 188-193.

Jain, K. and Bhola, D.A. (2018) Data Aggregation Design Goals for Monitoring Data in Wireless Sensor Networks. *Journal of Network Security Computer Networks* **4**(3): 1-9.

Jasim, A.A., Idris, M.Y.I.B., Azzuhri, S.R.B., Issa, N.R., Noor, N. B.M., Kakarla, J. and Amiri, I.S. (2019) Secure and Energy-Efficient Data Aggregation Method Based on an Access Control Model. *IEEE Access* 7, 164327-164343.

Jothiprakasam, S. and Muthial, C. (2018) A method to enhance lifetime in data aggregation for multi-hop wireless sensor networks. *AEU-International Journal of Electronics and Communications*, 85, 183-191.

Kadlikoppa, P., Umarji, I. and Patil, S. (2017) Data Aggregation & Transfer in Data Centric Network Using Spin Protocol in WSN. *International Journal on Recent and Innovation Trends in Computing and Communication*, **5**(8): 142-149.

Kamalesh, S. and Ganesh Kumar, P. (2017) Data aggregation in wireless sensor network using SVM-based failure detection and loss recovery. *Journal of Experimental & Theoretical Artificial Intelligence*, **29**(1): 133-147.

Kamble, S. and Dhope, T. (2016) Reliable routing data aggregation using efficient clustering in WSN. In: *IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, 246-250.

Krishnan, A.M. and Kumar, P.G. (2016) An effective clustering approach with data aggregation using multiple mobile sinks for heterogeneous WSN. *Wireless Personal Communications*, **90**(2): 423-434.

Kumar, N.K. and Bharathi, A. (2020) Real time energy efficient data aggregation and scheduling scheme for WSN using ATL. *Computer Communications*, 151, 202-207.

Lakshmi, M., Velmani, P. and Rani, P.A.J. (2019) Design and Implementation of Priority Hop Based Energy Efficient Cluster Routing Algorithm for WSN Data Aggregation. *International Journal of Applied Engineering Research*, **14**(11): 2698-2703.

Latha, A., Prasanna, S., Hemalatha, S. and Sivakumar, B. (2019) A harmonized trust assisted energy efficient data aggregation scheme for distributed sensor networks. *Cognitive Systems Research*, 56, 14-22.

Manishankar, S., Ranjitha, P.R. and Kumar, T.M. (2017) Energy efficient data aggregation in sensor network using multiple sink data node. In: *IEEE International Conference on Communication and Signal Processing (ICCSP)*, 0448-0452.

Mantri, D.S., Prasad, N.R. and Prasad, R. (2016) Mobility and heterogeneity aware cluster-based data aggregation for wireless sensor network. *Wireless Personal Communications*, **86**(2): 975-993.

Metan, J. and Murthy, K.N. (2018) FSDA: Framework for Secure Data Aggregation in Wireless Sensor Network for Enhancing Key Management. *International Journal of Electrical & Computer Engineering*, **8**(6).

Mohanty, P. and Kabat, M.R. (2016) Energy efficient structure-free data aggregation and delivery in WSN. *Egyptian Informatics Journal*, 17(3): 273-284.

Nayaka, R.J. and Biradar, R.C. (2017) Data aggregation and routing scheme for smart city public utility services using WSN. In: *IEEE Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 1-8.

Nguyen, N.T., Liu, B.H., Pham, V.T. and Liou, T.Y. (2018) An efficient minimum-latency collision-free scheduling algorithm for data aggregation in wireless sensor networks. *IEEE Systems Journal*, **12**(3): 2214-2225.

Otoum, S., Kantarci, B. and Mouftah, H. (2018) Adaptively supervised and intrusion-aware data aggregation for wireless sensor clusters in critical infrastructures. In: *IEEE International Conference on Communications (ICC)*, 1-6.

Prabhavathi, S., Subramanyam, A. and Rao, A.A. (2016) Energy efficient dynamic reconfiguration of routing agents for WSN data aggregation. In: *Emerging Research in Computing, Information, Communication and Applications*, Springer, New Delhi, 291-301.

Preetha, M. and Sivakumar, K. (2018) An Energy Efficient Sleep Scheduling Protocol for Data Aggregation in WSN. *Taga Journal*, **14**, 1748–0345.

Ren, M., Li, J., Guo, L., Li, X. and Fan, W. (2017) Distributed data aggregation scheduling in multi-channel and multi-power wireless sensor networks. *IEEE Access*, 5, 27887-27896.

Roy, N.R. and Chandra, P. (2019) EEDAC-WSN: Energy Efficient Data Aggregation in Clustered WSN. In: *IEEE International Conference on Automation, Computational and Technology Management (ICACTM)*, 586-592.

Sankaralingam, S.K., Nagarajan, N.S. and Narmadha, A.S. (2020) Energy aware decision stump linear programming boosting node classification based data aggregation in WSN. *Computer 3Communications*, **155**, 130–142.

Sasirekha, S. and Swamynathan, S. (2017) Cluster-chain mobile agent routing algorithm for efficient data aggregation in wireless sensor network. *Journal of Communications and Networks*, **19**(4): 392-401.

Shobana, M., Sabitha, R. and Karthik, S. (2020a) An enhanced soft computing-based formulation for secure data aggregation and efficient data processing in large-scale wireless sensor network. *Soft Computing*, **24**, 1-12.

Shobana, M., Sabitha, R. and Karthik, S. (2020b) Cluster-based systematic data aggregation model (CSDAM) for real-time data processing in large-scale WSN. *Wireless Personal Communications*, 117, 2865–2883

SriVenkateswaran, C. and Sivakumar, D. (2019) Secure cluster-based data aggregation in wireless sensor networks with aid of ECC. *International Journal of Business Information Systems*, 31(2): 153-169.

Subedi, S., Lee, S. and Lee, J. (2018) A New LEACH Algorithm for the Data Aggregation to Improve the Energy Efficiency in WSN. *International Journal of Internet, Broadcasting and Communication*, 10(2): 68-73.

Ullah, I. and Youn, H.Y. (2019) A novel data aggregation scheme based on self-organized map for WSN. *The Journal of Supercomputing*, 75(7): 3975-3996.

ULLAH, I. AND YOUN, H.Y. (2020) Efficient data aggregation with node clustering and extreme learning machine for WSN. *The Journal of Supercomputing*, 76, 10009–10035.

WAN, R., XIONG, N., HU, Q., WANG, H. AND SHANG, J. (2019) Similarity-aware data aggregation using fuzzy c-means approach for wireless sensor networks. *EURASIP Journal on Wireless Communications and Networking*, 1, 59.

WANG, X., ZHOU, Q. AND CHENG, C.T. (2019) A UAV-assisted topology-aware data aggregation protocol in WSN. *Physical Communication*, 34, 48-57.

WANG, X., ZHOU, Q. AND TONG, J. (2019) V-Matrix-Based Scalable Data Aggregation Scheme in WSN. *IEEE Access*, 7, 56081-56094.

WANG, Z., LIU, Z., WANG, F., CHEN, L. AND SHANG, W. (2019) Energy minimum encrypted data aggregation scheme for WSN in smart grid. In: *AIIPCC'19 Proceedings of the International Conference on Artificial Intelligence*, *Information Processing and Cloud Computing.* Article no. 73, 1-6.

WAYKAR, S. (2010) Dynamic routing protocol for ad-hoc network. In: *ICWET 2010 at TECT Mumbai.* Kluwer Academic Publishers.

YESTEMIROVA, G. AND SAGINBEKOV, S. (2018) Efficient data aggregation in wireless sensor networks with multiple sinks. In: *IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA)*, 115-119.

YOON, I., KIM, H. AND NOH, D.K. (2017) Adaptive data aggregation and compression to improve energy utilization in solar-powered wireless sensor networks. *Sensors*, **17**(6): 1226.

ZHANG, K., HUANG, H., WANG, Y. AND WANG, R. (2017) Link-based privacy-preserving data aggregation scheme in wireless sensor networks. In: *International Conference on Industrial IoT Technologies and Applications*, Springer, Cham. 119-129.

ZHANG, J., HU, P., XIE, F., LONG, J. AND HE, A. (2018) An energy efficient and reliable in-network data aggregation scheme for WSN. *IEEE Access*, 6, 71857-71870.

ZHANG, J., LIN, Z., TSAI, P.W. AND XU, L. (2020) Entropy-driven data aggregation method for energy-efficient wireless sensor networks. *Information Fusion*, 56, 103-113.

ZHONG, H., SHAO, L., CUI, J. AND XU, Y. (2018) An efficient and secure recoverable data aggregation scheme for heterogeneous wireless sensor networks. *Journal of Parallel and Distributed Computing*, 111, 1-12.

ZHOU, L., GE, C., HU, S. AND SU, C. (2019a) Energy-Efficient and Privacy-Preserving Data Aggregation Algorithm for Wireless Sensor Networks. *IEEE Internet of Things Journal*, **7** (5), 3948–3957.

Zhou, Q., Qin, X., Liu, G., Cheng, H. and Zhao, H. (2019b) An Efficient Privacy and Integrity Preserving Data Aggregation Scheme for Multiple Applications in Wireless Sensor Networks. In: *IEEE International Conference on Smart Internet of Things (SmartIoT)*, 291-297.