# A performance overview of contemporary hierarchical clustering algorithms in wireless sensor networks

G.V. Shaamili Varsa* and
D. Sridharan

Department of Electronics and Communication Engineering,
CEG Campus,
Anna University,
Chennai-600025, India
Email: shamiliece@gmail.com
Email: sridhar@annauniv.edu
*Corresponding author

**Abstract:** Wireless sensor networks (WSN) is a significant area in various applications and thereby becoming a research hotspot. WSN consists of a substantial number of sensor nodes powered by batteries. Energy efficiency is the most significant thing in WSN because of the non-replaceable batteries in sensor nodes. So energy upkeep is an important task in WSN. Because of the redundant deployment, WSN protocols have to scale to a large number of nodes. One of the prevalent keys to accomplish energy management and frequency reuse is clustering method. Clustering has proven to be an energy efficient method of data transmission and has many aids comprising of scalability and data aggregation. A frequency channel used in intra-cluster communication can be reused in multiple clusters. In this paper, a comparative study of the state of the art clustering algorithms in wireless sensor networks is presented. Initially, the state of the art surveys in the domain of clustering is discussed, which is then followed by the brief explanation of clustering concept, characteristics, design challenges and merits. This work provides the review of topical algorithms in the area of clustering and the classification of the same under various categories including mobility, energy efficiency, optimisation algorithms and fuzzy-based.

**Keywords:** wireless sensor networks; WSN; network lifetime; cluster head selection; survey; clustering; energy efficiency.

**Reference** to this paper should be made as follows: Shaamili Varsa, G.V. and Sridharan, D. (2021) ‘A performance overview of contemporary hierarchical clustering algorithms in wireless sensor networks’, *Int. J. Communication Networks and Distributed Systems*, Vol. 27, No. 1, pp.1–50.

**Biographical notes:** G.V. Shaamili Varsa received her Bachelor’s degree in the stream of electronics and communication engineering in 2011 from Anna University, Coimbatore and ME degree in the stream of communication systems in 2013 from Anna University, Chennai. She is currently a research scholar in the Department of Electronics and Communication Engineering, Anna University, Chennai, Tamil Nadu, India from 2014. Her research interests include wireless sensor networks and energy efficiency.

D. Sridharan obtained his BTech and ME in Electronics Engineering from Madras Institute of Technology, Anna University, Chennai, Tamil Nadu in 1991 and 1993 respectively. He received his PhD in the Faculty of Information and Communication Engineering, Anna University, Chennai, Tamil Nadu in 2005. He is currently working as a Professor in the Department of Electronics and Communication Engineering, CEG Campus, Anna University, Chennai, India. He was awarded the Young Scientist Research Fellowship by SERC of Department of Science and Technology, Government of India. His present research interests include internet technology, network security, distributed computing and wireless sensor networks.

## 1 Introduction

Wireless sensor networks (WSN) is an emerging/developing area in the research which plays an imperative role in many application areas. Basically, WSN comprises of sensors deployed in the physical scenario to sense the required attribute and communicate in a wireless fashion with the goal of transferring the sensed data to the base station (BS). When the deployed sensor senses/detects an event (e.g., change in temperature), it will be reported to the BS for further processing. A number of WSN applications include monitoring applications which include Environmental monitoring for detecting forest fires (Al-Habashneh et al., 2011), smoke detection in buildings, agriculture monitoring for automatic irrigation, predicting the attack of plant diseases, monitoring soil moisture (Shouyi et al., 2013), supervision of highway traffic in Metropolitan areas, Health monitoring, tracking applications including enemy intrusion in military tracking, animal tracking applications for determining the count and monitoring their entry in the residential area, Vehicle tracking and various other applications (Borges et al., 2014). The wide variety of health monitoring applications in WSN leads to the emerging of a new field, wireless body area networks (WBAN) (Movassaghi et al., 2014). Among various design challenges of WSN such as energy efficiency, scalability, lifetime, quality of service, fault tolerance, etc. (Gungor and Hancke, 2009), energy efficiency plays a crucial role in the WSN since power providing battery can neither be replaced nor be recharged in most applications. Though rechargeable batteries came into existence, in applications where humans are hard to reach, using battery in an efficient manner is mandatory. The lifetime of the network can be prolonged by the utilisation of the battery as effective as possible.

Among the several methods of achieving energy efficiency, clustering is one of the proven mechanisms. In clustering, sensor nodes are divided and grouped into various clusters based on some algorithm. Each cluster has a cluster head (CH) acting as a leader node, which performs not only sensing but also collects data from all of its cluster members and forwards to the BS (Mehrani et al., 2010). The CH performs data aggregation on collected data from its members and transmits the aggregated data to the BS (Younis et al., 2006). The data aggregation method reduces the energy wastage in the transmission of a large amount of data by transmitting the aggregated data alone. The primary goal of data aggregation is to accumulate and aggregate data in an energy efficient manner so as to enhance the lifetime of the sensor network (Rajagopalan and Varshney, 2006). The earliest cluster formation protocol is the low energy adaptive clustering hierarchy (LEACH) protocol where CHs are formed based on the randomised selection (Heinzelman et al., 2000). Some clustering mechanisms address the formation of isolated nodes during cluster formation. In large networks, if each sensor starts to communicate and engage in data transmission in the network, an excessive network congestion, and collisions of data will be experienced, which results in draining of a large amount of energy in the network. Node clustering is necessary in such cases. Congestion control algorithms in WSN have been surveyed by Ghaffari (2015).

In this paper, various issues and the state of the art energy efficient mechanisms for prolonging the network lifetime in WSN are discussed. It also discussed some of the recent clustering techniques which provide various methods of cluster formation for utilising energy in an efficient manner. The algorithms are classified according to various criteria and the comparison of the same is provided. The rest of the paper is structured as given below. Section 2 deals with related works which include some state of the art of surveys in the area of clustering in WSN. Section 3 provides brief introduction to clustering method which includes their objectives, design challenges, and clustering parameters. A review of various recent clustering protocols is given in Section 4. A comparison of the surveyed concepts is tabulated in Section 5. Section 6 concludes the work and Section 7 discusses the future directions.

## 2 Related work

In this section, some of the survey papers related to clustering in WSN are discussed. Younis et al. (2006) worked on node clustering in WSN and discussed its developments and deployment challenges. In this work, clustering techniques were classified based on CH election and nature of the clustering algorithm (probabilistic/iterative).

Various distributed clustering algorithms were compared, which are more suitable for large-scale networks. Several challenges in the clustering process were also discussed.

Abbasi and Younis (2007) produced a survey on clustering algorithms for WSN in which algorithms were grouped into variable convergence time and constant convergence time algorithms. Kumarawadu et al. (2008) presented a survey on algorithms for node clustering in WSN where the classification was done based on four categories which includes identity-based clustering, neighbourhood information-based clustering, probabilistic-based and biologically inspired clustering algorithms. Authors have discussed several algorithms under each category and also addressed some issues.

Jiang et al. (2009) performed a short survey of clustering algorithms, in which they have listed certain attributes as classification criteria among CH properties. Based on the classification criteria, few clustering algorithms including hierarchical control clustering algorithm (HCA), LEACH, algorithm for cluster establishment (ACE), energy efficient clustering scheme (EECS), energy efficient unequal clustering (EEUC) and power efficient and adaptive clustering hierarchy (PEACH) were discussed. Finally, a comparison was given based on the classification criteria.

Boyinbode et al. (2010) presented a survey on some clustering algorithms in WSN and compared those algorithms on the basis of CH selection and Cluster formation for non-CH nodes. Lotf et al. (2010) presented a short survey on some of the hierarchical routing protocols in WSN. It included the advantages and disadvantages of surveyed protocols and comparison were provided for these protocols with respect to energy efficiency and network lifetime.

Maimour et al. (2010) provided a survey on cluster-based routing protocols for Energy efficiency in WSN. Here, protocols are divided under two groups namely, pre-established cluster based and on-demand cluster-based routing protocols. Ramesh and Somasundaram (2012) has done a survey on various CH selection algorithms in WSN. Algorithms were classified under various categories such as Deterministic schemes, BS Assisted Adaptive schemes, fixed parameter probabilistic schemes, resource-adaptive probabilistic schemes and combined metric schemes.

Singh and Sharma (2015) presented a survey of various cluster-based routing protocols in WSN. In this work, clustering protocols were classified into block cluster-based, grid cluster-based and chain cluster-based protocols.

Comparison has made between popular clustering protocols with respect to various parameters including energy efficiency, cluster stability, complexity of the algorithm, etc.

Mishra et al. (2014) presented a similar survey on some clustering algorithms in WSN where some of the clustering problems were discussed with the classification of single-level and hierarchical algorithms. Later, LEACH, hybrid energy efficient distributed (HEED) clustering and Hausdorff clustering algorithms were discussed. This paper does not include the summarisation or the differentiation of the various clustering algorithms. Savitha and Patel (2014) presented a short survey on some of the clustering techniques in WSN. Some algorithms under probabilistic and non-probabilistic algorithms were discussed and comparison between these methods was done with the based on various clustering parameters discussed.

Afsar and Tayarani (2014) presented a detailed survey of clustering in sensor networks. In this work, clustering approaches were first divided into equal and unequal clustering approaches, which are further subdivided into probabilistic, deterministic and preset methods. The probabilistic algorithms are classified into random and hybrid methodologies, the deterministic into weight-based, fuzzy-based, heuristic-based and compound algorithms. A detailed explanation of various algorithms was given under each category. Also, comparison was given for first attempts clustering approaches and LEACH--based clustering protocols.

Anastasi et al. (2009) performed a survey on some energy conservation mechanisms in WSN. Xu and Gao (2011) performed a survey on some hierarchical routing protocols which are clustering protocols. The surveyed hierarchical protocols include LEACH, power efficient data gathering in sensor information systems (PEGASIS), threshold sensitive energy efficient sensor network (TEEN) protocol, adaptive TEEN (APTEEN) and two tier data dissemination (TTDD) protocol. Finally, a comparison between the above-listed protocols was given.

Liu (2012) presented a review on routing protocols related to clustering in WSN. A comparison of previous surveys was given and clustering protocols are categorised into cluster construction-based clustering routing protocols and data transmission-based clustering routing protocols. Comparison of these protocols has been done based on clustering characteristics, CH characteristics, clustering process, entire proceeding of the algorithm, Energy efficiency, cluster stability, scalability, load balancing and algorithm complexity. Jindal and Gupta (2013) reported a short survey on energy efficient routing protocols of WSN. Authors have presented a review on advancements in the cluster hierarchy protocol, LEACH and location-based protocol, geographic adaptive fidelity (GAF) protocol.

**Table 1** Comparison of various surveys on clustering in WSN

| *Title* | *Authors and year* | *Classification* | *Review* |
|---|---|---|---|
| Node clustering in wireless sensor networks: recent developments and deployment challenges | Younis et al. (2006) | Classified based on CH election, probabilistic and Iterative clustering techniques | Open issues |
| A survey on clustering algorithms for wireless sensor networks | Abbasi and Younis (2007) | Variable convergence and constant convergence time algorithms | Clustering techniques, objectives, taxonomy of clustering attributes and comparison |
| Algorithms for node clustering in wireless sensor networks: a survey | Kumarawadu et al. (2008) | Identity-based, neighbourhood information based, probabilistic based and biologically inspired clustering algorithms | Design challenges and future directions |
| Towards clustering algorithms in wireless sensor networks – a survey | Jiang et al. (2009) | Classification criteria based on CH properties | Basic concepts and advantages of clustering, comparison of protocols |
| A survey on clustering algorithms for wireless sensor networks | Boyinbode et al. (2010) | Categorised based on cluster head (CH) election and cluster formation for non-CH nodes | Challenges for clustering algorithms, clustering process and comparison |
| Hierarchical routing in wireless sensor networks: a survey | Lotf et al. (2010) | Short survey of hierarchical protocols | Comparison, merits and demerits |
| Cluster based routing protocols for energy efficiency in wireless sensor networks | Maimour et al. (2010) | Pre-established cluster based, on-demand cluster-based | Discussed about clustering and routing in WSNs |
| A comparative study of cluster head selection algorithms in wireless sensor networks | Ramesh and Somasundaram (2012) | Deterministic schemes, base station assisted adaptive schemes, fixed parameter probabilistic schemes, resource adaptive probabilistic schemes and combined metric schemes | Discussed on cluster formation, WSN topologies, clustering strategies and comparison of algorithms |
| Comparison study to hierarchical routing protocols in wireless sensor networks | Xu and Gao (2011) | Discussed LEACH, PEGASIS, TEEN,APTEEN and TTDD algorithms | Comparison |
| Cluster based routing protocols in wireless sensor networks: a survey | Wei et al. (2011) | Based on CH rotation, optimal cluster size, optimum mode of communication | Overview of clustering, challenges for clustering algorithms |
| Survey of extended LEACH based clustering routing protocols for wireless sensor networks | Aslam et al. (2012) | Extended LEACH protocols | Discussed about hierarchical routing protocols and its analytical comparison in terms of energy efficiency |
| A survey on clustering techniques for wireless sensor networks | Mitra and Nandy (2012) | Heuristic algorithms, weighted schemes, Hierarchical schemes and grid based schemes | Clustering algorithms for heterogeneous wireless sensor networks and comparison |

**Table 1** Comparison of various surveys on clustering in WSN (continued)

| *Title* | *Authors and year* | *Classification* | *Review* |
|---|---|---|---|
| A survey on clustering routing protocols in wireless sensor networks | Liu (2012) | Cluster construction based and data transmission based clustering routing protocols | Advantages and objectives of clustering, taxonomy of clustering schemes and comparison |
| Survey on recent clustering algorithms in wireless sensor networks | Mishra et al. (2013) | Single-level clustering, hierarchical clustering, LEACH, HEED and Hausdorff clustering | Clustering problems |
| Study of energy efficient routing protocols of wireless sensor networks and their future researches: a survey | Jindal and Gupta (2013) | Advancements in LEACH and GAF (geographical adaptive fidelity) protocols | Design goals of wireless sensor network |
| Clustering in sensor networks: a literature survey | Afsar and Tayarani (2014) | Equal sized and unequal sized algorithms | Clustering objectives, clustering characteristics, comparison |
| A survey on clustering techniques in wireless sensor networks | Savitha and Patel (2014) | Probabilistic and non-probabilistic algorithms | Discussed about clustering and routing in wireless sensor networks |
| A survey on cluster based routing protocols | Singh and Sharma (2015) | Block cluster, grid cluster and chain cluster based protocols | Design challenges and classification of routing protocols, merits and demerits |
| A survey on centralised and distributed clustering routing algorithms for WSNs | Zanjireh and Larijani (2015) | Distributed and centralised clustering algorithms | Features of WSN and comparison of protocols |
| Energy efficient hierarchical clustering approaches in wireless sensor networks: a survey | Jan et al. (2017) | Cluster-based and grid based hierarchical approaches | Basics of clustering, classification and open issues |
| Review of hierarchical routing protocols for wireless sensor networks | Haque et al. (2018) | General comparison | Challenges and comparison of various hierarchical routing protocols |
| Cluster based routing protocols in wireless sensor networks: a survey based on methodology | Fanian and Rafsanjani (2019) | Classical, fuzzy-based, meta-heuristic and hybrid meta-heuristic and fuzzy-based approaches | Clustering based parameters and methodology-based parameters |
| A survey on HEED based energy efficient clustering algorithms for WSN | Ullah (2020) | HEED based clustering algorithms | Network model and routing issues are discussed and comparison |
| Clustering objectives in WSN: a survey and research direction analysis | Shahraki et al. (2020) | Various clustering objectives | Taxonomy of clustering, statistical analysis of clustering algorithms |

Wei et al. (2011) presented a brief survey on cluster-based routing protocols in WSN in which protocols are sorted into various groups such as rotating the role of CHs, optimal cluster size and optimum mode of communication between SNs and CHs. Aslam (2012) presented a survey of extended-LEACH-based clustering protocols for WSN where the various versions of extended LEACH protocols including multi-hop LEACH, sLEACH and M-LEACH were discussed. Analytical comparison for energy efficiency of these protocols was given.

Zanjireh and Larijani (2015) presented a survey on centralised and distributed clustering-based routing algorithms for WSN. Here, authors have provided the detailed survey on various distributed and centralised clustering algorithms. A comparison was presented between those protocols and advantages and disadvantages were highlighted. Mitra and Nandy (2012) presented a survey on various clustering schemes in WSN. Clustering algorithms are grouped into algorithms, weighted schemes, hierarchical schemes and grid-based schemes. Some of the hierarchical clustering algorithms have been discussed and comparison has been done for few algorithms. A survey on clustering algorithms for cognitive radio networks has also been given by Yau et al. (2014).

Jan et al. (2017) presented a survey of various hierarchical routing protocols where clustering and grid-based algorithms are discussed. Haque et al. (2018) provided a survey of hierarchical routing protocols where the classification of various hierarchical protocols is discussed.

Fanian and Rafsanjani (2019) presented a survey of cluster-based routing protocols in WSN which provides criteria and classification factors, clustering-based parameters and methodology-based parameters. They classify the cluster-based algorithms into classical approaches, fuzzy-based, meta-heuristic-based and hybrid meta-heuristic and fuzzy-based approaches. Comparison is done based on application, cluster size, intra cluster and inter-cluster, method, mobility, node types and CH rotation.

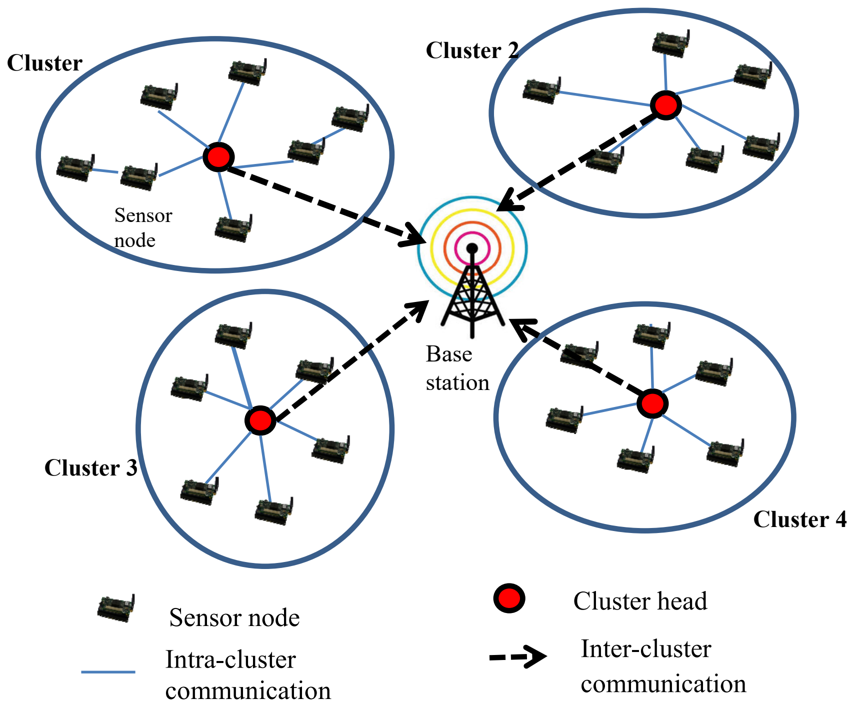
Ullah (2020) provided HEED-based energy efficient clustering protocols in WSN. The classical LEACH and HEED algorithms are discussed in detail, followed by network model. In this work, various HEED-based clustering algorithms are discussed and compared based on various factors such as cluster count, CH role and mobility, objective, communication, network & routing type, deployment, load balancing and communication cost.

Shahraki et al. (2020) presented a survey of clustering objectives in WSN where various objectives related to clustering are discussed in detail. It is provided in the form of research direction analysis. The clustering and its taxonomy are given in detail, the objectives for clustering in WSN are discussed and the statistical analysis is provided for existing algorithms.

## 3 Clustering

The methodology of clustering is particularly useful for large-scale networks, which can reduce communication overhead. Naturally, the grouping of nodes into various clusters has been widely adopted by the research community to satisfy the scalability needs, achieve high energy efficiency and prolong network lifetime in large-scale WSN environments. Many data gathering and routing protocols imply this clustering hierarchy due to the reason that data aggregation and data fusion are possible, which leads to significant energy conservation by transmitting the limited amount of data. Generally, Cluster formation process is a three-level hierarchy where BS forms the highest level, CH nodes form the middle level and cluster members form the lowest level. CH nodes act as gateways between sensor nodes and the BS and transmit to the BS either directly or through multi-hop. The general clustering architecture is shown in Figure 1.

**Figure 1** Clustering mechanism in WSN (see online version for colours)




## *3.1 Characteristics of clustering*

The following are the attributes and characteristics to be considered for designing any clustering algorithm.

- optimal number of clusters to be formed
- intra cluster and inter cluster communication (whether single-hop/multi-hop)
- the mobility of the CH nodes
- the roles of sensor nodes in the network
- cluster formation methodology which considers various parameters including scalability, energy, distance, load balancing, etc.
- CH selection which may be either prior to clustering or selected after clustering, usually with the high energy factor
- clusters overlapping while clustering process.

## *3.2 Design parameters*

The following parameters determine the design of a clustering algorithm in WSN.

- *Number of clusters:* It achieves coverage and load balancing, thereby increasing network lifetime. This parameter depends on the network size and many algorithms predetermine the size of the cluster.
- *Node deployment:* Sensor nodes can either be deployed in a predefined location or in random location. This depends on the application.
- *Energy consumption:* Because in many applications, the main power source, battery cannot be replaced frequently, energy conservation requirements have to be meet. The algorithm should be designed to work for longer time by reducing energy consumption.
- *Fault tolerance:* As applications using WSN are getting higher, it enters into wide area involving military and critical applications. In such cases, failure of sensor nodes may occur and it is difficult to replace. Self-configuration is obligatory in those circumstances.
- *Data aggregation:* Transmission of sensed data from nodes to the BS involves a large amount of energy splurging, overhead and redundant data. Aggregation of data is essential to degrade these parameters. Most sensors have same data results in the transmission of identical data which can be reduced by the aggregation process performed by the CH.
- *Scalability:* The requirement of a number of nodes in a WSN may vary according to applications and coverage area. The hierarchical clustering methodology should prove to be scalable.
- *Communication:* The inter-cluster and intra-cluster communication should be taken into account while designing in a way that it consumes lesser energy. Most inter cluster communication is multi-hop-based which reduces the load on the same node near BS.
- *CH rotation:* The role of CH should be rotated for every round or the energy should be checked to avoid earlier death of CHs.

## *3.3 Objectives*

- network scalability
- data aggregation which reduces redundant data transmission
- reduced energy consumption and delay
- collision avoidance
- load balancing by assigning different roles to CHs and sensor nodes
- maximising network lifetime
- quality of service.

## 4 Classification of clustering algorithms

Clustering algorithms in WSN are emerging day by day with the aim of reducing available problems. This section provides classification of various clustering algorithms and are discussed in brief. Formerly, the classical clustering approaches are discussed.

### *4.1 Classical clustering approaches*

1 *LEACH:* The earliest effort in clustering the WSN nodes is the distributed protocol, LEACH (Heinzelman et al., 2000). In this work, random rotation of high energy CHs is utilised for load balancing. Data fusion is also applied to reduce unwanted data transmission, thereby improving network lifetime and reducing energy dissipation. It runs in two phases, setup phase and steady state (SS) phase. The former involves the cluster formation while the latter involves data transmission. During the development of clusters, each node resolves whether to be or not to be a CH which is based on selecting a random number between 0 and 1. The selected number is compared with the threshold value and is calculated as

$$T(n) = \begin{cases} \dfrac{P}{1 - P * \left( r \bmod \dfrac{1}{P} \right)} & \text{if } n \in G \\ 0 & \text{otherwise} \end{cases}$$

where $P$ is the desired percentage of CHs, $G$ is the set of nodes that have not been CH for the last $1/P$ rounds and r is the present round. If the random number is less than $T(n)$, the node will become CH for the current round. The CH broadcasts the CH-advertisement message using CSMA-MAC protocol. On receiving this message, all non-CH nodes decide to join which CH based on the signal strength of the received message. Each CH creates a TDMA schedule for member nodes, according to which data transmission occurs. A multihop approach of LEACH (M-LEACH) (Mhatre and Rosenberg, 2004) and centralised LEACH protocol (LEACH-C) (Heinzelman et al., 2002) are also emerged with some improvements over LEACH protocol.

2 *PEGASIS:* Lindsey and Raghavendra (2002) proposed a chain-based protocol named power efficient gathering in sensor information systems (PEGASIS). It is an enhancement over LEACH protocol. The main concept is that each node makes to and from transmissions only with the close neighbour thereby saves energy. All nodes are assumed to have the global knowledge of nodes in the network and Greedy algorithm is utilised for the formation of chain. The chain is constructed from the farthest node towards the closest node to the BS. During data gathering, each node transmits to the closest neighbour which fuses data with its own and transmits to the next neighbour in the chain. On the death of a node, the chain will be reconstructed. Here, all nodes will become a leader atleast once.

3 *TEEN:* Another clustering protocol which emerged after LEACH is threshold sensitive energy efficient sensor network protocol (TEEN) (Manjeshwar and Agrawal, 2001). This protocol is reactive where the data will be transmitted to the BS only when an event occurs. It involves two threshold levels specifically, soft and

hard which are used to make decision on when the data should be transmitted to the BS. On sensing the hard threshold, the node should turn on its transmitter and transmit the data to the CH. For soft threshold, the node just turns on its transmitter but it will not send any data to the CH because there is only a little change or no change. In this hierarchical protocol, data are gathered by first level CHs which in turn transmit to the second level CHs. Finally, data will be transmitted to the BS. The main advantage of this data centric protocol is that it reduces the number of transmissions.

4 *APTEEN:* Manjeshwar and Agrawal (2002) introduced an improved version of TEEN known as APTEEN (adaptive threshold sensitive energy efficient sensor network protocol). It is a hybrid protocol which unites characteristics of both reactive and proactive protocols. After the election of CHs, they broadcast four parameters viz. count time, thresholds, schedule and attributed. Those nodes which reach the hard threshold, transmit these factors to the CH. Other nodes will sense the field and send the data to the CH. It is a centralised protocol where the BS performs the CH selection. Some advantages in APTEEN are data aggregation and use of threshold values for earlier transmission of data from hard threshold nodes which helps in the emergency situation. However, the complexity lies in the implementation of threshold values.

5 *HEED:* Another familiar protocol in the area of clustering is HEED clustering protocol which is an iterative-based clustering scheme (Younis and Fahmy, 2004). CHs are elected in three phases namely, initialisation, processing and finalisation. Initially, all nodes determine the probability of being CH as

$$CH_{prob} = C_{prob} * \frac{E_{residual}}{E_{\max}}$$

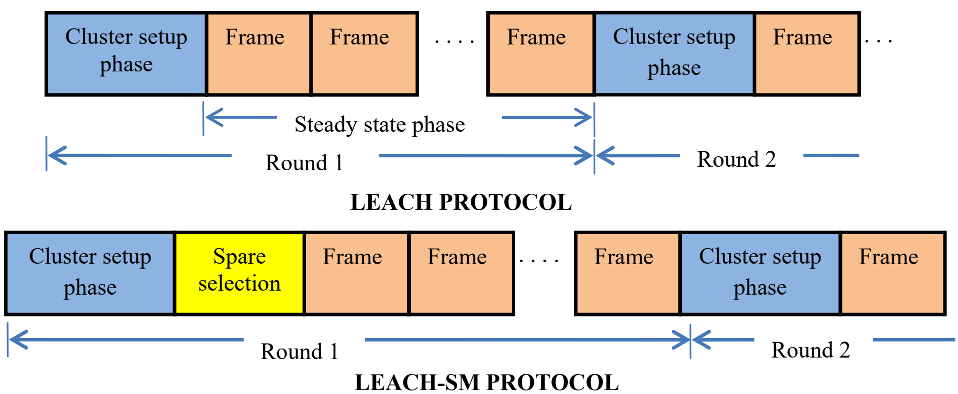
where $C_{prob}$ is the initial percentage of CHs (5%). Depending on this probability, each node selects itself as either final CH (if its value is 1) or tentative CH (if its value is less than 1). The tentative CHs can become common sensor nodes if it is not selected to be CH. Once the final CHs are elected, it broadcasts the status with communication cost. Nodes within the range receive the message and select the lowest cost CH. Nodes that do not receive any advertisement message may become final or tentative CH depends on the probability value. Upon entering the next iteration, each node doubles the value of $CH_{prob}$. The sensed data are transmitted to the BS from CHs in a multi-hop manner. It is a distributed protocol which achieves longer lifetime.

6 *PEACH:* PEACH is introduced by Yi et al. (2007). This protocol supports adaptive multi-level clustering. PEACH develops clusters by overhearing information from neighbour nodes. It backs both location-aware and location-unaware WSNs. The location aware PEACH is a dynamic chain-based clustering protocol which operates in a hierarchical manner. In the location aware case, with the knowledge of location information, global information is transmitted using which the transmission schedule of each node is determined. The merit of this protocol is that it has no overhead compared to other existing approaches.

## 4.2 Leach-based algorithms

1 *LEACH-SM:* Bakr and Lilien (2014) proposed an extended LEACH protocol called LEACH-SM which improves the lifetime of the network by the addition of spare nodes. Spare nodes will be utilised when any of the sensor nodes drain out of energy. Extension of lifetime was achieved by the realisation of *optimal spare selection*, *energy saving management of spare nodes* and *network lifetime estimation*. In the spare selection, two intervals are available namely, sensing range neighbour (SR-neighbour) discovery interval and decentralised energy-efficient spare selection technique (DESST) interval. DESST is a portion of spare management that allows each sensor node to act as either primary node or spare, depends on which they enter into active/passive mode respectively. This protocol provides improvement over the LEACH protocol by increasing the network lifetime achieved by the management of spare nodes and reduces redundant data transmission. It includes little more overhead. The difference in the phases of LEACH and LEACH-SM protocol is shown in Figure 2.

**Figure 2** LEACH and LEACH-SM protocol (see online version for colours)



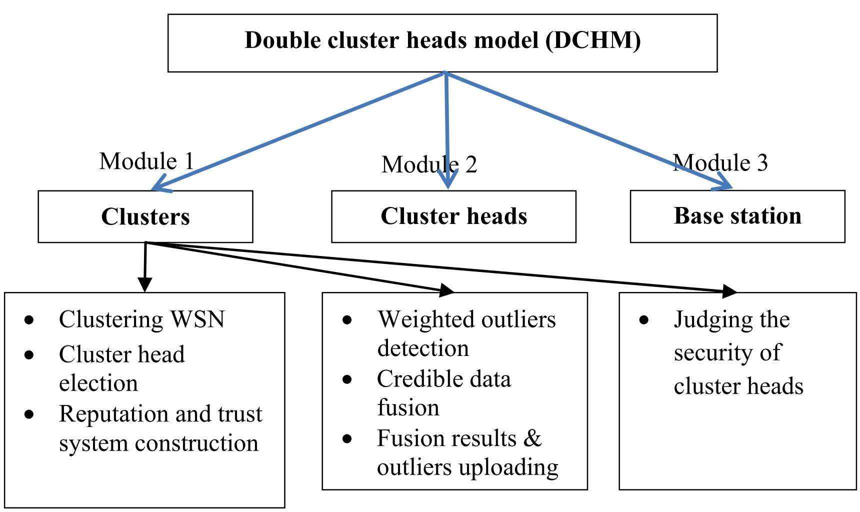

2 *O-LEACH:* Khediri et al. (2014) proposed an optimisation low energy adaptive clustering hierarchy (O-LEACH) to limit energy consumption where clusters are formed based on the residual energy dynamically. The routing process is initiated by the BS. CHs in each round are elected in such a way that the energy of the node must not be ten percent greater than the residual energy of all nodes. CHs make use of the fusion function (aggregation) with the aim of reducing the correlated data. Transmission is carried out after the broadcasting of TDMA table by the CH to its members. Longer stability is achieved compared to LEACH and LEACH-C. The benefit of this method is that it offers optimisation over the LEACH by selecting clusters according to residual energy and stability. The protocol is verified only in static but not in dynamic networks. Moreover, detailed description of the clustering and CH selection process is not provided. The position of the BS plays an important role which once misplaced, will not achieve the performance.

3 *ASLPR:* Shokouhifar and Jalali (2015) proposed a new application specific low power routing (ASLPR) protocol for selecting optimal CHs based on distance from the BS, residual energy and distance between CHs. A hybrid algorithm is proposed

which is based on genetic algorithm (GA) and simulated annealing (SA) to extend the network lifetime. An adaptive threshold value is determined which helps in deciding the CH for the current round. It consists of two phases. During Setup phase, a node picks a random number between 0 and 1. If it is less than the adaptive threshold value, the node becomes CH for the current round. CHs, once selected, advertise its selection and nodes will join the CH based on the minimum distance. In the SS phase, aggregated data transmission occurs according to the TDMA schedule. The optimisation of the controllable parameters can be done by the hybrid GA-SA algorithm. The advantage lies in the fact that it can adaptively tune its parameters to generate maximum lifetime. It also achieves high gain. The computational complexity is high for selecting the optimal CHs. It applies only to static network and multi hop routing technology is not adopted which is necessary in the case of large-scale networks.

4 *DCHM:* Fu and Liu (2015) proposed a novel cluster-based data fusion model called double cluster heads model (DCHM) which is produced by the combination of clustering techniques, reputation and trust systems and data fusion algorithms in order to generate the secure and accurate data fusion method in WSN. The DCHM model is shown in Figure 3. Two CHs are selected by the reputation and trust system. Data fusion will be done independently and results of which are forwarded to the BS. The BS module is responsible for computing the dissimilarity coefficient. This helps in preventing the selection of compromised nodes as CHs. This integration method provides improvement in terms of security and accuracy. The convergence time of the reputation and trust system is shorter. However, attacks such as jamming and information falsifying are not considered.

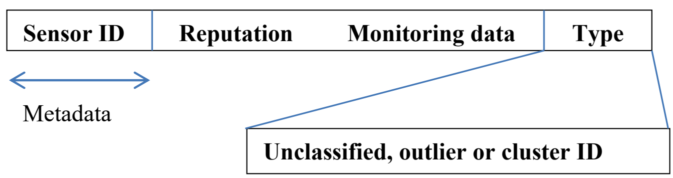
**Figure 3** Double cluster heads model




5 *RDIF:* Ma et al. (2015) proposed a reputation driven information fusion (RDIF) algorithm, where a reputation system is established for each cluster and the CH is designed to perform the information fusion with the reputation values. It is an in-network information fusion algorithm. With the consideration of values collected by

sensor nodes in addition to reputation values, outliers are eliminated before fusion. The two major steps include selecting the credible data from the dataset and fusing the credible data. LEACH-C algorithm is employed for cluster formation and Bayesian information fusion algorithm for fusing the results. This algorithm improves the reliability and accuracy of information fusion. Security is also improved to some extent due to the presence of reputation system. This method does not deal with high dimensional problems. Further, energy conservation or lifetime is not considered which is the major requirement in WSN. Figure 4 depicts the data structure of the CH information.

**Figure 4** Data structure of the information stored in CH in RDIF algorithm (see online version for colours)




6 *R-LEACH:* Behera et al. (2019) proposed an efficient CH election based on residual energy suitable for IoT networks. This algorithm has two phases, set up phase and SS phase. In the former phase, the existing LEACH is utilised for the first round of CH election.

At the end of first round data transfer, each node will deficit some amount of energy. Based on this energy spent, the following threshold function is used for the next round of CH selection.

$$T(n) = \begin{cases} \dfrac{P}{1 - P\left(r \bmod \frac{1}{P}\right)} * \dfrac{E_{res}}{E_{init}} k_{opt} & ; \forall n \in G \\ 0;\ otherwise \end{cases}$$

where '*P*' denotes the probability of the node being selected as the CH, 'r' denotes the round, $E_{res}$ and $E_{init}$ represents the residual and initial energies, respectively. $k_{opt}$ denotes the optimal number of clusters. The CH selection is followed by the formation of clusters depending on the signal strength. CHs then transmit the TDMA schedule in which data will be transmitted by nodes and the process repeats. In the SS phase, data transmission occurs and data aggregation will be performed by CHs.

7 *LEACH-SF:* Shokouhifar and Jalali (2017) proposed an algorithm for balanced cluster formation and prolonging the network lifetime. Initially, the WSN is divided into clusters using fuzzy C-means (FCM) clustering algorithm. For CH election, Sugeno Fuzzy model is employed which considers residual energy, distance from the sink and distance from the cluster centroid as the fuzzy inputs. The optimisation of the parameters is done using artificial ant bee colony algorithm. Based on these inputs, impact factor (IF) is determined for every node in the cluster. The node with the highest IF is selected as CH and can be determined as given below

$$IF = \begin{cases} Z(n); if\ E(n) \geq t_e * 1/N \sum_{i=1}^{N} E(i) \\ 0; otherwise \end{cases}$$

where $Z(n)$ is the crisp output determined from the weighted average de-fuzzification method, $E(n)$ is the residual energy of the node, $t_e$ is the threshold parameter which finds whether the node is a candidate CH or not. This algorithm performs better in terms of standard deviation of intra-cluster distance, network lifetime, CPU running time and node deaths compared to the existing LEACH and LEACH-based algorithms.

8 *M-LEACH:* Chaturvedi et al. (2016) introduced an energy efficient CH selection algorithm for WSN. Each node competes for the CH selection in the first phase of the cluster cycle. If any of the nodes wants to be CH, it broadcasts the CH request message to all neighbours with its remaining energy. The BS then compares the node's energy with that of all other nodes. If the requested node's energy is found to be maximum, it replaces the previous CH node. It then broadcasts its ID to BS and all other nodes. If not, the node waits for the new CH announcement. This happens for every cycle. However, this algorithm lags the clustering process of nodes in detail.

### *4.3 Energy efficient clustering algorithms*

1 *REAC-IN:* Leu et al. (2014) proposed a new clustering scheme for WSN called regional energy aware clustering with isolated nodes (REAC-IN). In this technique, CHs are selected based on the residual energy and average regional energy of all sensors within the cluster. Unlike LEACH, in threshold computation, desired percentage of CHs '*p*' is selected based on the residual energy and the average regional energy. The drawback of node isolation is prevented here with the consideration of regional average energy and the node-sink distance in such a way to decide whether the node should transmit directly to the sink or the CH in the previous round. It plays a main role in the prevention of isolated node formation. This protocol provides energy efficiency thereby improving the network lifetime. The major advantage lies in the fact that it solves the isolated node problem and also provides stability.

2 *EDIT:* Thakkar and Kotecha (2014) proposed an energy and delay index for trade-off (EDIT) for optimising energy and delay parameters. The CH and next hop are determined with the consideration of energy and delay requirements of a given application using EDIT. The trade-off between energy and delay is found by considering the two different types of distance between CH and cluster members which are Euclidean distance and hop-count. It involves cluster set up (CS) phase and SS phase with the inclusion of neighbour discovery (ND) Phase in the first round. Algorithm begins with the ND phase originated by the sink. When the energy

of a node reduces below a certain threshold, it declares itself as a Dead node. All nodes broadcast their energy level and compare with others. If its energy is less than any other node, it will become CH. During cluster setup, CHs are selected based on the energy messages received.

$$EDIT = \left(\frac{Neighbours_{total}}{Nodes_{total}}\right)^{\alpha} + \left(\frac{1}{Avg.\,Distance\ from\ Sink}\right)^{\beta}$$

All probable CHs select this 1/EDIT time to announce that it is the CH. The selection process is announced by the joining of member nodes in the range and allotting time slots through TDMA schedule. This EDIT protocol provides energy efficiency and data transmission with bounded delay which is mandatory for time critical applications. End to end delay is lesser for Euclidean distance compared to the Hop count and energy consumption is larger for the same case. Moreover, this is the first protocol providing the trade-off between energy and delay. The drawback is that it provides comparison only between two aspects of distance and does not show the performance comparison with other protocols.

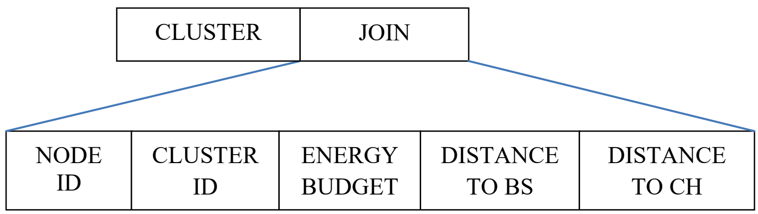
3 *SEECH:* Tarhani et al. (2014) proposed a new distributed algorithm named scalable and energy efficient clustering hierarchy (SEECH) where CHs and relay nodes are selected separately. High degree nodes are assigned the role of CHs whereas low degree nodes are assigned as Relay nodes. To distribute CHs uniformly, a distance-based method is utilised in the process. This protocol includes start phase in the first round, followed by set up and SS phase in all rounds. During CH selection, some nodes are selected randomly using a distributive method as CH candidates. From these candidates, final CHs are selected by executing a simple algorithm. Relay nodes are selected in such a way that these are either closer to the sink or nodes that cannot be CHs. The main benefit is that it balances energy consumption by selecting separate CHs and relay nodes thereby increasing the lifetime of the network. CH distribution is uniform which provides balanced load on CH, balanced energy distribution and less delay.

4 *DHCR:* Sabet and Naji (2015) proposed an algorithm with the objective forming the consumption of energy due to control message transmissions. This protocol involves cluster route-setup and data transmission stages. In the first stage, CHs are specified, and routing tree is constructed followed by formation of clusters. In the second stage, sensed data is transmitted through the routing tree towards the BS. CHs schedule the transmission through TDMA scheduling to avoid collisions. This approach reduces control packets thereby decreasing the transmission cost. The size of all clusters is uniform which achieves energy saving. It performs well in networks with various densities. It avoids the formation of isolated nodes.

5 *ENC:* A distributive energy neutral clustering (ENC) algorithm is proposed by Peng et al. (2015) for forming clusters with the aim to provide continuous network operation. A cluster head group (CHG) is employed to share heavy loads in such a way that a cluster can use multiple CHs. CHG is a group of CH nodes which takes the role of CH in each cluster during the time slot. During its turn, CHG-node plays

the role of CH else turns off its radio. The ENC algorithm includes three phases. It starts with the initialisation phase where certain nodes are randomly selected to be centre nodes (CN) of the initially produced clusters. CN are selected based on a constant which is determined as the ratio of the number of clusters expected to be formed in the network to the total number of sensors in the network. Once these nodes are selected, rest of the nodes turn into the Initial Cluster Members. During the Cluster Formation phase, CN nodes broadcast a 'cluster' message with Cluster-ID. Nodes which receive the message joins the CN with the 'join' message and format is shown in Figure 5.

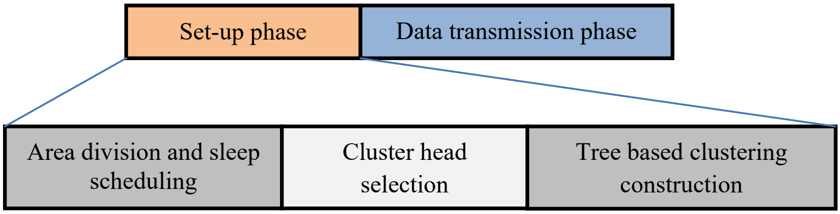
**Figure 5** Cluster-join message format in ENC algorithm (see online version for colours)




During the finalisation phase, the CN estimates the size of the CHG. The neighbour initial cluster member nodes form the CHG and remaining nodes will become final cluster members. CN then transmits the role of every node using the 'schedule' message. An extension of ENC provides the formation of equal sized clusters which in turn allows achieving maximised information gathering. It reduces control message overhead and can achieve maximum information gathering. The major advantage is that it provides perpetual network operation. Use of multiple CHs helps in sharing heavy traffic load. It also prevents cluster failures.

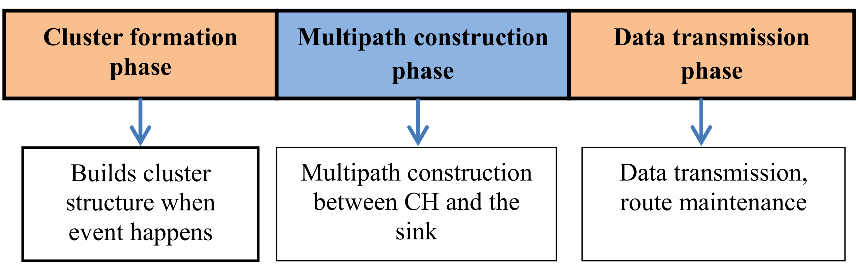
6 *SSTBC:* Tan and Viet (2015) proposed SSTBC algorithm which provides energy preservation by turning off nodes which provide the same information. Various steps in SSTBC protocol is represented in Figure 6. Minimum spanning tree (MST) is used to reduce the energy dissipation. During area division and sleep scheduling step, BS acquires global knowledge of all nodes which are alive, regarding distance and energy. Sensing area is partitioned into grids of size less than the threshold. The node with the maximum residual energy in each grid is set to be in active mode, and others are made to sleep. During CH selection step, all active nodes are partitioned into 5 clusters. All active nodes in each cluster are formed into a MST. The tree is produced using the greedy algorithm and prim algorithm is utilised for achieving the minimum sum of the weight. During data transmission phase, leaf nodes start transmitting to their parent nodes which perform fusion with its own data. On receiving data from all nodes, CH performs data aggregation. This protocol helps in achieving better energy efficiency compared to LEACH and PEGASIS algorithms. It also attains greater network lifetime.

**Figure 6** SSTBC protocol (see online version for colours)




7 *Bee-Sensor-C:* Cai et al. (2015) proposed an energy concerned and scalable multipath routing protocol known as Bee-Sensor-C, which is based on the behaviour of a bee swarm. Using a dynamic clustering scheme, parallel data transmissions are allowed which provides reduced routing overhead and improved scalability. Multipath construction method is used to achieve the balance of energy consumption. Bee-Sensor-C is an event-driven and on-demand routing protocol. Various steps in the protocol are shown in Figure 7.

**Figure 7** Various phases in bee-sensor-C protocol (see online version for colours)

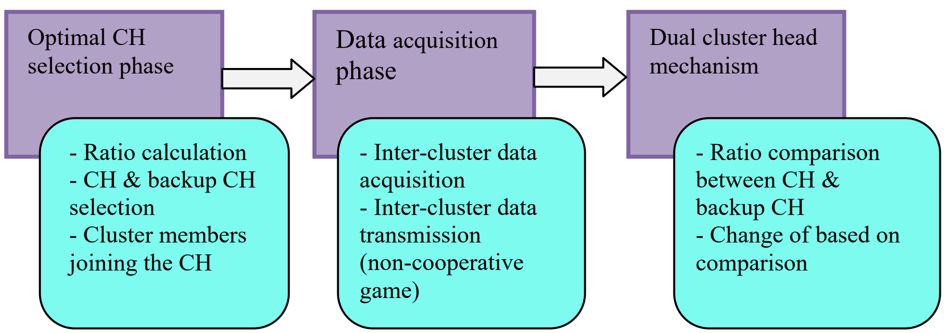



The benefit of this protocol is less overhead and high scalability. It helps in achieving balanced energy consumption. It provides high packet delivery rate and energy efficiency. It is not compatible with dynamic scenarios.

8 *HEBM:* This energy efficient and load balancing protocol involves various phases such as initialisation phase, ND phase, temporary CH election phase, final CH election phase, cluster formation, steady phase and data transmission phase. The ND phase involves the construction of neighbour table with the calculation of node weight. The temporary CH is selected based on the distance between the node and the BS, residual energy, distance between neighbours and node weight. The node with maximum $P_{ch}$, the condition for being selected as CH candidate, is chosen as temporary CH. The selected final CH is responsible for gathering data, creating TDMA schedule and transmitting to neighbour CHs. Clusters are formed based on the selected CHs and then each CH sends the TDMA schedule to their respective members. Apart from that allotted time slot, nodes can enter into sleep mode thereby preserving energy. In the data transmission, nodes transmit data to their CH which in turn transmits to the sink after aggregation (Gherbi et al., 2016).

9 *ECGD:* Lin and Wang (2019) proposed a dual CH mechanism to reduce the energy consumption resulted from the rotation of CHs. The Energy-efficient Clustering algorithm combined game theory and dual CH (ECGD) consists of three phases namely, optimal CH selection, data acquisition and dual CH election phase. In the optimal CH selection phase, the sink first determines the optimal distribution of the CH in each layer where the network is divided into various layers. The ratio of energy to distance metric is introduced which can be determined as the ratio of the residual energy to the distance of the node to the optimal distribution of the CH. Each node in the layer compares its ratio with that of the others and the node with the highest ratio is selected as the CH. The node with the second highest ratio is selected as the Backup CH which is introduced to reduce the energy consumption due to the CH rotation. The second phase involves acquiring data by cluster members within the cluster and inter cluster data transmission. At the end of each phase, the ratio of the CH and backup CH are compared and if found to be lower, then the backup CH takes the role of CH.

**Figure 8** ECGD algorithm constructions (see online version for colours)

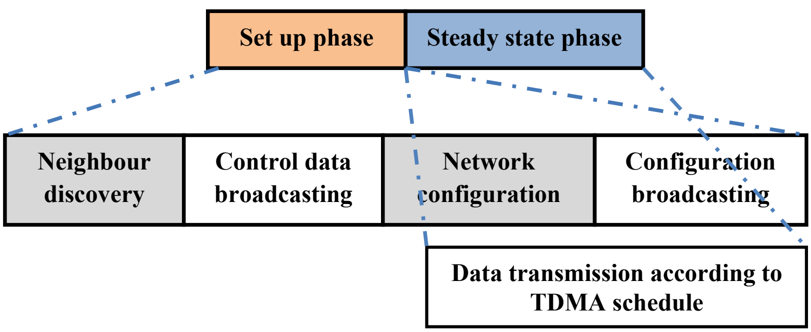



10 *ACP:* Gherbi at al. (2018) proposed a hierarchical approach known as distributed energy efficient adaptive clustering protocol (ACP) for WSN. It involves initialisation phase, phase decision, group formation phase and transmission phase. The first phase involves BS_msg and ND where the neighbours with good quality are discovered. The decision phase involves CH election which is done in two steps, local competition and distance condition. To participate in CH candidate election, each node determines the 'Pch' value, broadcasts it and compares with its neighbours. If the value is found to be maximum, it elects as CH candidate else it joins the node with maximum value. Each temporary CH calculates the distance between the other temporary CH and checks with the threshold value. If it is lesser, it compares the probability value and if found to be greater, it becomes the ordinary node and joins the other CH. In the group formation phase, clusters are formed in a way that neighbour nodes join the CH. After cluster formation, time slots are allocated to nodes for transmission. In addition to energy, other parameters can be included for further improvisation.

### 4.4 Optimisation algorithm-based approaches

1 *ACO-C:* İnkaya et al. (2015) proposed a novel clustering methodology based on ant colony optimisation algorithm. Adjusted compactness and relative separation are the two objective functions used. Two steps involved are neighbourhood construction and data set reduction for extracting the local characteristics of data points and scalability respectively. Figure 9 shows the steps in ACO-C protocol. No prior information is available in this technique and it provides a framework for spatial clustering problem. Clustering evaluation relative to neighbourhood (CERN) and weighted clustering evaluation relative to neighbourhood (WCERN) are two evaluation parameters utilised for perfect clustering. In this ACO-C algorithm, the ant is a search agent which inserts edges between pairs of data points and connected points form a cluster. It uses NC algorithm for the construction of neighbourhood of each data point thereby forming sub-clusters/closures. The major demerit includes that it has longer execution time. It has difficulties with noise detection.

**Figure 9** Steps in ACO-C algorithm (see online version for colours)



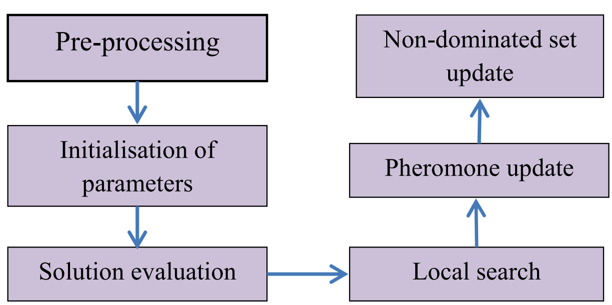

2 *Balanced PSOGSA:* Huynh et al. (2015) proposed a balanced PSOGSA algorithm for improving the lifetime of the WSN. This algorithm introduces three key characteristics. One is combining the PSO and GSA for reducing the probability of trapping in local optimum. For network longevity, election probabilities are weighted by the initial energy of the node related to other nodes. Finally, while updating the position of any particle, the position of neighbouring particles is also considered to improve the objective function. The steps in this algorithm include network initialisation, classification of advanced and normal nodes, probability computation and CH election, clustering, transmission of data from members to CHs and aggregated data transmission, accelerator updation, velocity and position updation for particle.

This algorithm shortens the convergence time of the algorithm. It reduces energy consumption and improves the lifetime of the network. It also helps in better partitioning of the network. The network delay is not balanced which is obligatory in time critical applications.

3 *SNS:* Al-Azzawi et al. (2015) proposed a single node selection (SNS) for energy efficient data delivery based on discrete cuckoo search optimisation (DCSO) algorithm. In SNS, instead of all sensors transmitting the sensed data, only one sensor which is close to the event is selected to deliver the data packet. DCSO algorithm is utilised to choose the best sensor (B_Sens) to deliver data with minimum energy consumption. In DCSO algorithm, input is the set of all sensors in the cluster and the output is the best route (sensor) selected using the fitness function determined based on distance and energy. It helps in choosing the optimal route for data delivery. By selecting the single node, it preserves overall energy thereby increasing the lifetime of the network. Increased throughput is also a significant advantage. The algorithm is little complex.

4 *TPSO-CR:* Elhabyan and Yagoub (2015) proposed a TPSO-CR protocol for maximising the energy efficiency, cluster quality and network coverage. The clustering algorithm for finding the optimal set of CHs and the routing algorithm for finding the optimal routing tree are combined to provide a two tier protocol which is based on particle swarm optimisation (PSO) algorithm. The routing algorithm is constructed with a novel particle encoding scheme and fitness function. The weight sum approach (WSA) is utilised for determining the fitness function. In the clustering algorithm, the BS computes the average energy after collecting information of all nodes. Nodes with energy above the average are eligible to be CHs. It is followed by initialisation of particles and particles evaluation according to the fitness function. The second tier is responsible for construction of routing tree constructed using PSO algorithm. Various steps in this algorithm are shown in Figure 10. This protocol improves the PDR and network coverage with acceptable energy consumption. Only single-hop hierarchical cluster is allowed and energy efficiency can further be enhanced.

**Figure 10** Steps in differet phases of TPSO-CR protocol (see online version for colours)




5 *Enhanced OEERP:* Rejina Parvin and Vasanthanayaki (2015) proposed enhanced-OEERP protocol to prevent the formation of residual nodes. Residual nodes are individual nodes that do not form a part of any cluster during clustering. In this protocol, PSO algorithm is utilised for clustering and gravitational search algorithm (GSA) is employed for routing. During the formation of cluster, a node to assist the CH called cluster assistant (CA) node is selected to reduce the overhead of CH. Initially, number of clusters in a network is predicted, followed by clustering, where BS collects information from all nodes such as position, velocity and energy. The fitness value is calculated for each particle in every iteration and global best and

local best particles are selected. The node with the maximum fitness value in the iteration containing global best value is taken as reference and cluster is formed. CA nodes have higher fitness value next to CH. The GSA algorithm is used to find the best next hop. The route is constructed by calculating the force of attraction determined based on masses of particles and distance between them. Thus, the BS provides the optimal route which helps in the prevention of residual node formation. It provides increased throughput, high packet delivery ratio and improved energy efficiency. This, in turn, enhances the lifetime of the network which is the major requirement in WSN.

6 *ROL:* Hammoudeh and Newman (2015) proposed a new cluster-based route optimisation and load balancing protocol (ROL) to afford a solution for prolonging the network lifetime, improving network robustness and providing timely delivery of messages which is mandatory in case of time sensitive applications. A nutrient-flow-based distributed clustering (NDC) algorithm is developed for load balancing. Various merits provided by this algorithm involve enhanced lifetime, energy efficiency, less delay and excessive packet delivery ratio. It does not provide the optimal path for the movement of mobile sink.

7 *FGF:* Dattatraya and Rao (2019) introduced hybrid-based CH selection for maximising lifetime. In this work, a fitness-based CH selection algorithm is proposed which is a hybrid optimisation algorithm. The fitness-based glowworm swarm with fruitfly algorithm (FGF) is the hybridisation of the glowworm swarm optimisation (GSO) algorithm and fruitfly optimisation algorithm (FFOA). The CH selection is based on the parameters, energy, distance and packet delay. The fitness is evaluated and sorted, among which the best five is chosen. If the index is less than 5, FFOA is updated else GSO is updated. The node with minimum fitness is selected as the CH. The multiple objectives such as energy maximisation and delay minimisation are achieved by reducing negative searchability. The CH performs direct communication with the BS. It achieves efficient network performance by considering QoS parameters.

8 *DTBLO:* Masdari and Barshandeh (2020) proposed a discrete TLBO algorithm (DTLBO) for designing the clustering scheme in WSN. The conventional TLBO algorithm is modified for discrete problems such as clustering. The discretisation is done with the help of genetic operators such as crossover and mutation. This algorithm produces many nested cluster layers. The fitness function is determined with energy required for communication within the clusters and among the clusters. The initialisation phase takes place once for getting location information of nodes. It involves teacher and learner phase for updating the solution. During the setup phase, BS determines the CHs based on the energy costs whereas in SS phase, transmission occurs in sub-phases according to TDMA schedule. Each CH aggregates and forwards to their respective parent CH until it reaches the BS. According to local search algorithm, the CHs are rotated among all cluster nodes depending on the energy and distance and the number of cluster members is compared with the threshold value. The algorithm is better in terms of network lifetime, data delivery and energy usage but being centralised algorithm, the fault tolerance will be lesser.

9 *Thermal aware clustering algorithm*: Priya and Venkatanarayanan (2020) proposed a clustering algorithm for WSN which is a thermal aware solution including temperature factor for CH selection. It is a hybrid of fuzzy and spider optimisation algorithm (SOA) where the fuzzy logic is utilised for clustering process and SOA for the efficient CH selection. The CH is selected by considering thermal heat value and the distance between the nodes. The fuzzy logic approach with α-cut set is employed for clustering and it uses Eigen centrality-based clustering. The objective function is designed using factors such as average distance from member nodes to CH, distance from CH to BS, energy consumption ($E_{Total}$) and temperature rise of the network (T)and is given by,

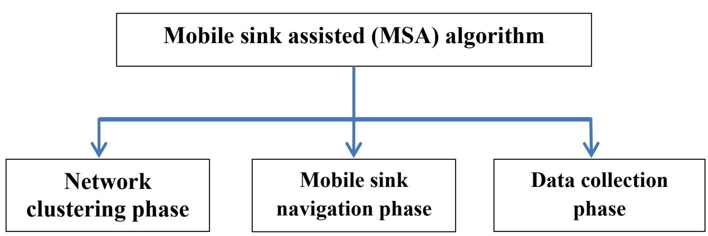
$$F = \alpha dist_1 + \beta dist_2 + \gamma E_{Total} + \sigma T_{Total}$$

where $\alpha$, $\beta$, $\gamma$, $\sigma$ are inertia weights. With the help of vibration and intensity of vibration, the solutions are updated and best is selected. The temperature of the node depends only on the total number of packets but environmental factors can be considered.

## *4.5 Mobility-based clustering algorithms*

1 *MSA:* Jose and Sadashivappa (2015) proposed a mobile sink assisted (MSA) algorithm to enhance the lifetime of the network. The clustering phase performs clustering using LEACH protocol. In the Mobile sink navigation phase, new groups of CHs are formed and a mobile sink is allotted to each group. In the final phase, after deciding the path of the mobile sink, they move traverse the path and collect data from CHs and transmit the aggregated data to the static sink.

**Figure 11** MSA algorithm (see online version for colours)




The static sink, in turn, forwards the data to the required application. In this work, energy hole problem is mitigated and the network lifetime and stability period is expanded. It offers low packet drop ratio and packet delay. Figure 11 provides different phases in MSA algorithm.

2 *MSIEEP:* Abo-Zahhad et al. (2015) proposed a MSIEEP protocol for lightening the energy holes and utilised the adaptive immune algorithm (AIA) for finding the sojourn locations. MSIEEP protocol involves the *prepare phase* for initialising the network by the sink, *setup phase* for network setup by finding the sojourn locations of the sink and optimal CHs and the *SS phase* for transmitting the data from sensors to the CH or sink according to the TDMA schedule. In AIA algorithm, the mobile

sink moves to the centre location and finds nodes that have more energy than average. It forms a population pool of antibodies with them and Roulette wheel selection algorithm is used for antibodies reproduction. Antibodies with greater fitness value are considered as parent antibodies for next generation. The replication is applied for the selection of better CHs by sorting them and the clonal proliferation is used for the reproduction of offspring, which is followed by mutation. Then new pool of antibodies is selected for next generation based on the objective function. In SS phase, the sink moves to the region in its path, makes all nodes in the region to wake up and nodes start sensing the data. The merit of this algorithm is that it achieves scalability and involves low data collection time. It provides energy conservation of 55% over the method using static sink. The drawback of this work lies in determining the polling points of each cluster and scheduling the MIMO data uploading from multiple clusters.

3 *LBC-DDU:* Zhao et al. (2014) proposed a three layer framework which employs distributed load balanced clustering and dual data uploading (LBC-DDU) to achieve scalability, better lifetime and low latency. The framework includes sensor layer, CH layer and mobile collector layer. LBC algorithm is applied in the sensor layer to perform clustering with multiple CHs in each cluster. The concept of multiple CHs is to balance the load and to assist dual data uploading. The data from CH is forwarded to SenCar (mobile collector) through inter-cluster transmissions. LBC algorithm consists of initialisation, Status claim, cluster forming and CH synchronisation. In the Sencar layer, the Sencar stops at every polling points in every cluster to collect data from CHG. This protocol helps in the average distribution of energy. Once formed, there is no change in clusters which helps in energy conservation. It prolongs the stability period which is mandatory for specific applications. However, it does not deal with security.

4 *Energy efficient algorithm with mobile sink:* Wang et al. (2019) proposed an energy efficient scheme for WSN which combines the clustering technique and the sink mobility. The CH selection is based on the residual energy of the node and the distance between the node and the sink. With the help of these two parameters, the weight factor is determined for each node and the one with the highest weight can be selected as the CH. The CH reselection in every round is avoided in this algorithm to reduce the unwanted energy consumption. The reselection occurs only when the weight falls below the threshold. For inter cluster communication between CHs, a chain is constructed using the greedy algorithm and the closest CH to the sink is selected as the leader node. The sink moves in a circular trajectory with constant angular velocity and all nodes are informed about the initial position and the angular velocity of the sink to determine the current position of the sink. However, the distribution of CHs is uneven which will greatly influence the network lifetime. The sink mobility technique should be carefully designed to mitigate some problems.

5 *Data collection method using mobile nodes (MNs):* Pang et al. (2020) introduced a data collection method in WSN using multiple MNs. However, the path of the MN should be properly determined for which the authors have proposed a path equalisation algorithm based on ratio (PEABR). The binary K-means clustering algorithm is utilised for clustering the nodes in the network. The network is

considered to be semi-circle or fan-shaped and is divided into sectors according to the number of MNs. The PEABR algorithm uses the maximum deviation rate to balance the path traversed by MN. The ant colony algorithm is used to determine the order to access the CHs in each fan shaped area. The deviation value of the shortest path of each MN and its length can be determined using this ant colony algorithm. The deviation value can be computed as,

$$\partial_i = \frac{L_{i+1}}{L_i}; i = 1, 2, \ldots.., k$$

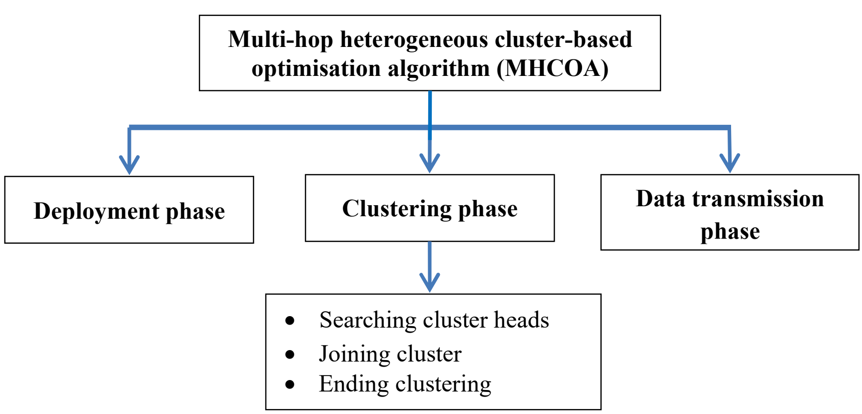
where '$L_i$' denotes the path length and '$k$' denotes the number of MN. In this algorithm, the CH nodes are arranged manually at virtual centres, which are high energy nodes. This might not be possible in all applications. The performance improvement is not shown in comparison with other algorithms.

### *4.6 Algorithms for heterogeneous networks*

1 *SEEC:* Farouk et al. (2014) proposed a stable and energy efficient clustering (SEEC) protocol for heterogeneous networks in which each cluster consists of randomly deployed advanced node (AN) and some NN. NN are allocated the sensing operation and either super nodes (SN)/ANs perform the operation of aggregation on all data from its cluster and transmit to the BS. The multi-level extension of SEEC called M-SEEC is proposed where SN are introduced to cover distant sensing areas. Both SNs and ANs drain out energy slower than the NNs. This results in the balanced energy consumption. NNs cannot act as CHs. Clustering is done in the first round and remains unchanged and thus saves energy.

2 *MHCOA:* Hu et al. (2015) proposed a multi-hop heterogeneous cluster-based optimisation algorithm (MHCOA) to reduce the communication load of heterogeneous WSNs and to preserve the energy of the whole sensor network. The algorithm is showed in Figure 12.

**Figure 12** Phases in MHCOA (see online version for colours)



It uses high energy sensor nodes as CHs which can balance energy consumption. An assumption is made such that higher energy nodes (CHs) are deployed artificially while lower energy nodes undergo random deployment. This algorithm calculates the number of CHs and cluster radius. This algorithm achieves energy saving which in turn extends the lifetime. Another merit is less end-to-end delay. It involves three phases. Also, communication overhead is analysed which includes analysis for deployment phase and clustering phase. The clustering phase includes the communication overhead of CHs statement phase, cluster members respond phase and for CHs' acknowledgement signal to cluster members.

3 *Effective clustering protocol based on network division:* Abidi and Ezzedine (2019) introduced the effective method of clustering for heterogeneous networks. Nodes are deployed in two regions based on their energy, that is, NN in the region surrounding the BS and ANs farther from them. This ensures the maximum working of the network. The CH selection is based on the distance to the sink, residual energy and the neighbours count. The threshold differs from that of LEACH in a way that a cost function is incorporated into that of LEACH and is given by,

$$T(n) = \begin{Bmatrix} \dfrac{P * cost(i)}{1 - P * \left( r \bmod \left( 1/P \right) \right)}; if \ \mathrm{i} \in \mathrm{G} \\ 0; otherwise \end{Bmatrix}$$

and the cost function is determined as,

$$Cost(i) = \alpha * \left( \frac{E_{rem}(i)}{E_{init}} \right) + \beta * \left( \frac{N_{nb}(i)}{N_{alive}} \right) + \gamma * \left( \frac{D_{BS}(i) - D_{BS\min}}{D_{BS\max} - D_{BS\min}} \right)$$

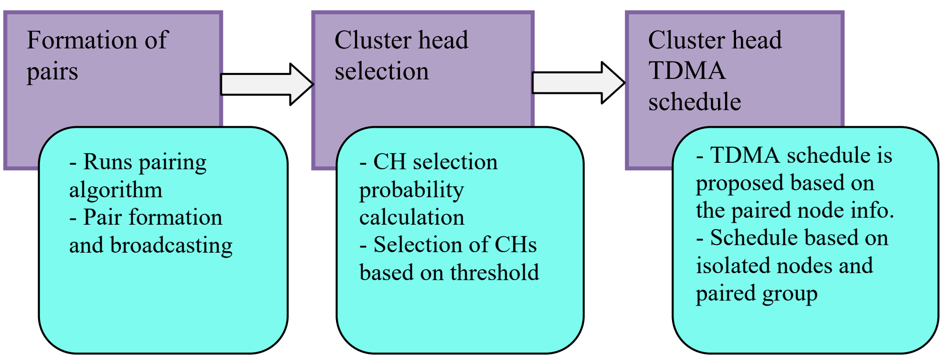
where '$r$' is the current round, '$P$' is the probability of the node to become CH, $E_{init}$ and $E_{rem}$ are the initial and remaining energies respectively, $N_{nb}$ and $N_{alive}$ are the number of neighbours and alive nodes, $D_{BS}(i)$, $D_{BSmin}$ and $D_{BSmax}$ are the distance from the node to the BS, distance from the closest node to BS and BS, distance from the farthest node to the BS. CHs are selected based on this value and CH in the current round cannot become CH for the next $(r+1)/P$ rounds which ensures the distribution of CH role among all nodes. Clusters are formed with the selected CHs and data transmission will be done.

4 *ETASA*: A hybrid energy and traffic-aware sleep-awake (ETASA) is proposed by Shagari et al. (2020) for improving energy efficiency and load balancing in heterogeneous WSN. This algorithm employs the concept of pairing to reduce the redundant data transmissions in case of closely located nodes from sending same sensed data. Here, nodes within 10 m distance are formed as pairs and only one of them can awake and sense/transmit at a time. The CH selection is performed similar to TEAR but with the modified probability where the CH is selected with higher energy, low traffic and higher number of pairs. There are paired groups inside a cluster to which the paired nodes belong to.

**Energy and traffic aware sleep-awake mechanism:**
1. *First round:* Paired nodes transmit data in a round robin fashion.
2. *All other rounds:*
   - Each node computes its average traffic rate of previously sent messages and compares with its pair
   - The current awake node checks whether its energy is below the threshold value and if it is above threshold with lesser traffic rate, it continues to sense and transmit.
   - If the traffic rate is above any of its pair, it goes to sleep after broadcasting the status & the other node enters into wakeup state.
   - In case of a tie in traffic rate, the node with the highest residual energy is made to awake.

The conventional TDMA schedule is modified in such a way that a single slot is allotted to each paired group and all isolated nodes within a cluster instead of all nodes inside the paired group. This is due to the fact that one node in the paired group is allowed to transmit at once. By assigning one time slot to a paired group, the energy during contention and idle listening in traditional TDMA is saved to an extent. Figure 13 provides various steps.

**Figure 13** Steps in the ETASA algorithm (see online version for colours)




5 *MEACBM:* Toor and Jain (2019) introduced mobile energy aware cluster-based multi-hop (MEACBM) protocol for heterogeneous WSN using mobile data collector. As heterogeneous protocol, the algorithm uses three types of nodes such as normal, advanced and intermediate. The network is partitioned into sectors and a MN is placed inside each sector. This protocol reduces the problem of energy holes in the network by introducing MNs. The controlled mobility with adaptive sojourn location is implemented. In the start-up phase, the CH selection is based on the threshold value which is given by,

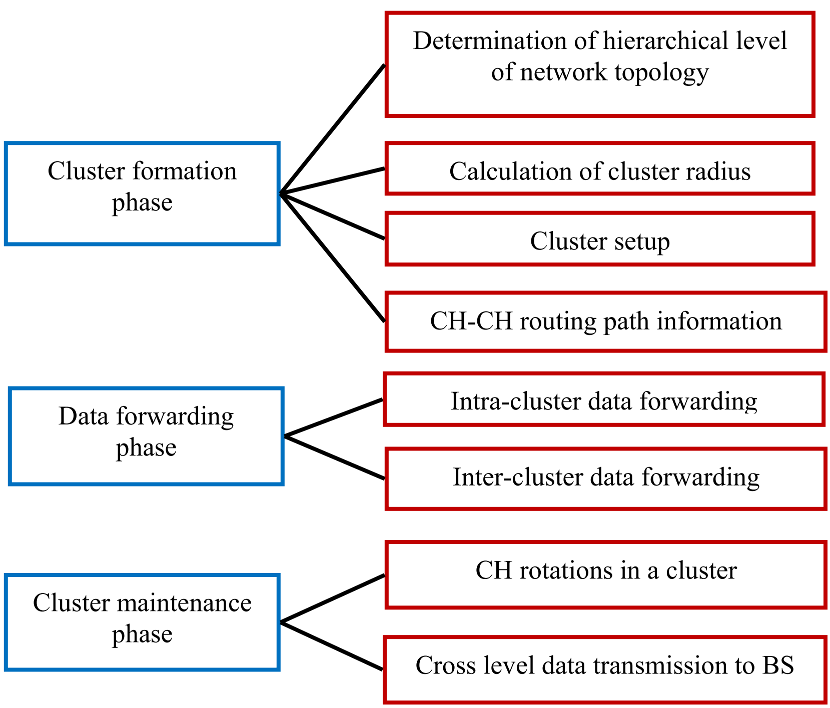
$$T(n_z) = \begin{cases} \dfrac{P_z}{1 - P_z\left(r * \mathrm{mod}\left(1/P_z\right)\right)} * [S(i) \cdot E]; & \textit{if } n \in G^z \\ 0; & \textit{otherwise} \end{cases}$$

where $P_z$ is the probability of nodes being CH and '$z$' may be any among normal, advanced and intermediate nodes. It differs from other protocols in such a way that there is an additional factor, $S(i).E$ which denotes the existing energy of the node. This factor helps in selecting the node with higher energy to be CH. In case of any isolated node, sub clusters are formed with sub-CH. The mobile data collector is used to collect data from CHs and one MN is allocated to each sector. However, the number of cluster member nodes inside each cluster is not considered which provides no load balancing.

### 4.7 *Location-based clustering algorithms*

1 *ACT:* Lai et al. (2012) proposed an arranging cluster sizes and transmission (ACT) ranges protocol with the aim of reducing the cluster size near the BS. Figure 14 illustrates various phases and steps involved in ACT protocol. In this method, all CHs are allowed to consume approximately the same amount of energy to avoid the earlier death of CHs nearer to the sink. ACT protocol consists of cluster formation phase, data forwarding phase and cluster maintenance phase. By reducing the size of clusters near the BS, it reduces rapid energy drain, thereby increasing the network lifetime. The coverage problem is an issue in this work.

**Figure 14** Steps in ACT protocol (see online version for colours)




2 *Hybrid CS:* Xie and Jia (2014) proposed a transmission efficient clustering method which utilises hybrid compressive sensing (CS). In this hybrid method, sensor nodes do not use CS for transmitting data to CH. CHs employ CS for inter cluster

transmission, i.e., data transmission to the BS. Initially, sensor nodes are deployed uniformly using Poisson Point process. After distribution of nodes, clustering takes place and CHs transmit data using CS. Here, instead of transmitting the original data to the sink, CHs compute projections of all nodes and transmit. For intracluster transmission, sensor nodes transmit data through the shortest path to the CH. For intercluster transmission, a minimal cost backbone tree is constructed which connects all CHs to the sink and transmission of projections is made through the tree path. In this work, determination of optimal cluster size is necessary to reduce the number of intracluster transmissions. A mobile agent-based compressive data gathering algorithm (MA-Greedy algorithm) is also proposed to adapt the dynamic change of network topology (Lv et al., 2016). The centralised clustering algorithm is the minimum transmission clustering algorithm which includes selecting the CHs, dividing nodes into clusters and forming the backbone tree. Nodes that are nearer to the centre points of each cluster are selected as CHs. The MST method is used for the backbone tree construction. It follows distributed implementation where the sink calculates the central points of cluster areas and elects CH. Once CHs are elected, they broadcast through advertisement messages. Nodes join the nearby CH. In the backbone tree, each CH has upstream CHs called parent CHs. Each CH transmits projections to its parent CHs till it reaches the sink. This method of using CS reduces the number of data transmissions thereby providing efficiency. It is also fault tolerant and can reorganise the cluster structure during power failures. It achieves high compression ratio. However, this applies only to uniform distribution of nodes.

3 *H-DHAC:* Zhu et al. (2015) proposed a hybrid distributed hierarchical agglomerative clustering (H-DHAC) protocol which is applicable for scenarios where location data may or not be available. It contains both the location data (quantitative) and binary connectivity data (qualitative) for clustering in WSN. The first case is preferred for accurate location scenario, and the latter is preferred if location data is not available. The clustering includes three main steps which are obtaining input data set, calculating resemblance coefficients and executing HAC algorithm. Unweighted pair-group method with arithmetic averages (UPGMA) is utilised here. The transmission power control (TPC) scheme used here is adaptive transmission power control (ATPC). The scheduling and data gathering stage involves two phases which are Intra-cluster transmission phase and CH to BS transmission phase. Finally, during cluster maintenance phase, CH rotation has to be done in order to balance the energy consumption. This method mainly works towards the enhancement of lifetime of the network and increasing the amount of transmitted data. This method works well both in the presence and absence of GPS. It is not applicable for multi-hop communications.

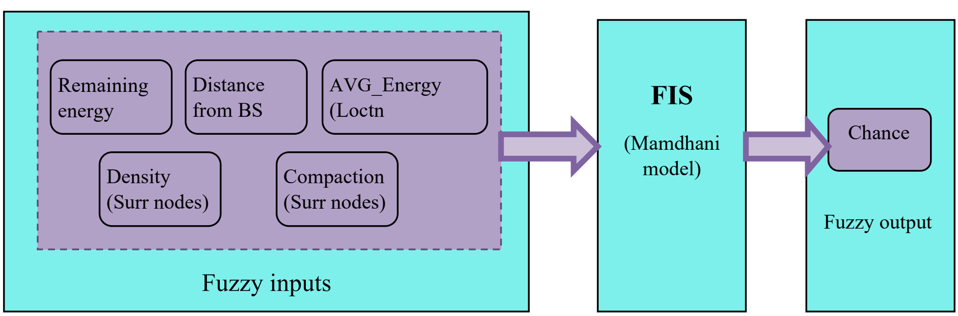
## *4.8 Fuzzy-logic-based clustering approaches*

1 *FMCR-CT:* Mazinani et al. (2019) proposed a fuzzy multi cluster-based routing with a constant threshold (FMCR-CT) for improvisation of the network lifetime. In order to avoid the number of control messages while transmitting CH election, the clustering is not done for every round. The CH selection is performed using different algorithms in each round. The first clustering is done using the fuzzy where the residual energy and number of neighbours are given as fuzzy inputs. The chance for

being CH is determined for each node and the one with the highest output within its neighbouring radius is selected as the CH. The second clustering is performed only when there is necessity since there will not be a drastic change in the energy of the CHs. Then CHs will re-elected in the third clustering based on the residual energy and the distance to the CH and is carried out in rounds 3, 5, 7, and so on. After the selection of CHs, each node tries to send sensory data to the CH, which aggregates data and forwards to the BS via the CH leader node. The CH leader node is determined based on the suitability criterion and is calculated as the ratio of the distance to the BS to that of the residual energy of the node. In the competition radius of the CH, if there is no CH with better suitability criterion, single hop transmission is adopted else multi-hop transmission.

2 *FL-EEC/D:* An energy-efficient fuzzy logic-based clustering technique named fuzzy logic-based energy efficient clustering based on minimum separation distance enforcement (FL-EEC/D) for WSN is proposed by Hamzah et al. (2019). This technique adopts many energy-affecting factors for clustering and utilises fuzzy inference system for achieving better parameter integration results. Five linguistic variables are introduced as fuzzy inputs as shown in the figure which influences the network lifetime by energy consumed by CHs, total energy consumed by non-CH or distribution of energy consumption loads. The output variable ‘chance’ decides the CH in such a way that the node with the highest chance value acts as the CH. The CHs are determined under the condition that there should be minimum separation distance ‘d’ between the newly selected and any of the previously determined CHs.

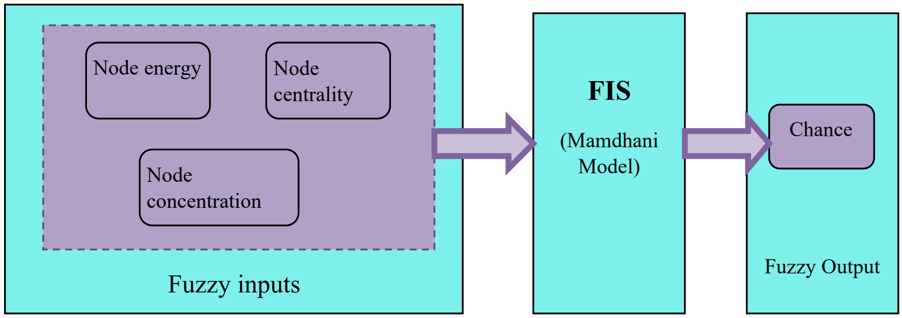
**Figure 15** Fuzzy model of FL-EEC/D algorithm (see online version for colours)




3 *LEACH-FC:* Lata et al. (2020) introduced the LEACH-FC algorithm for improving the network lifetime of WSN. It is a LEACH-based protocol where the fuzzy logic is used for clustering and CH election. Figure 16 shows the fuzzy implementation.

Initially, the CHs are selected with fuzzy inputs as node energy, node concentration and node centrality (NC). The output chance value determines the CH where the node with high energy, high concentration and close centrality has the highest chance of being CH. For the formation of clusters, the fuzzy inputs taken into account are node energy, distance to the CH and the distance to the BS. Mamdani technique is employed for CH, vice CH and clustering whereas the centroid defuzzification is used to obtain crisp value.

**Figure 16** Fuzzy model of LEACH-FC algorithm (see online version for colours)

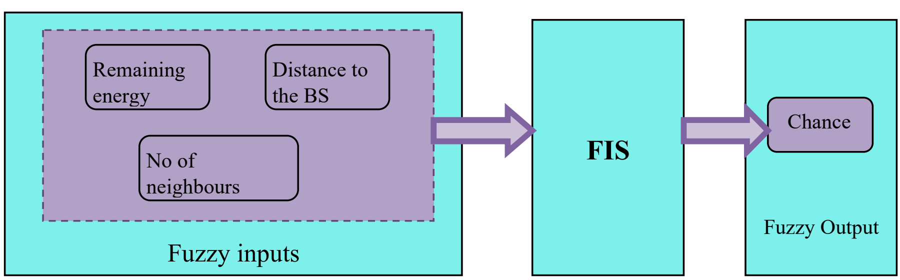



4 *MACHFL-FT:* Mirzaie and Mazinani (2019) proposed a fuzzy logic-based energy-efficient protocol for heterogeneous WSN named MACHFL-FT. Three different algorithms are used to perform clustering in different rounds for selecting better CHs and steps of which are given below.

First algorithm:

- the numbers of neighbours are determined for each node
- the fuzzy output chance is determined with fuzzy inputs such as remaining energy, distance to the BS and the number of neighbours
- the node with the highest chance inside the neighbouring radius is selected as the CH.

**Figure 17(a)** First algorithm of MACHFL-FT (see online version for colours)



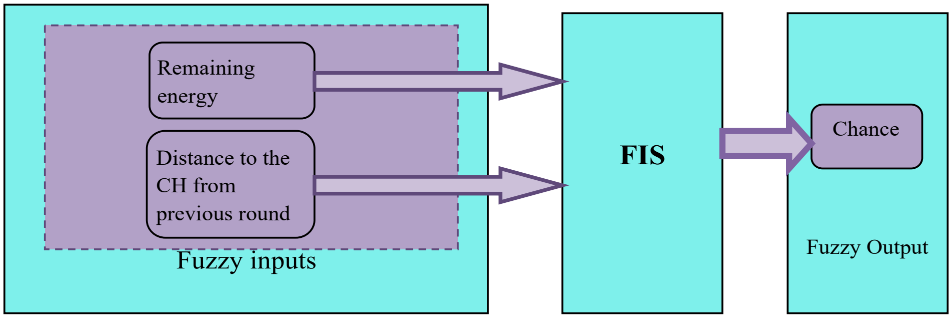

*Second algorithm:*

- the CHs from the previous round are reselected as CHs.

*Third algorithm:*

- the numbers of neighbours are determined
- the fuzzy inputs given to the FIS are remaining energy and the distance to the CH from the previous round
- the node with the highest fuzzy output chance is selected as the CH within the neighbouring radius.

**Figure 17(b)** Third algorithm of MACHFL-FT (see online version for colours)




5 *Multi-criterion FL-based approach:* Ranga et al. (2016) introduced a multi criterion fuzzy logic inter-cluster and intra-cluster-based stable CH election approach for WSN. The most stable nodes using multi-criterion is selected as CHs using fuzzy logic technique. In this approach, the stable nodes which pass the criterion at two levels are only selected as CHs. At level-1, the expected node stability index and number of neighbours are given as fuzzy inputs and stable nodes resulting from fuzzy output are selected as local eligible nodes and are taken to level-2 selection. The distance between CHs, distance of CH to sink through multi-hop and centrality are the parameters taken for level-2 selection and given as fuzzy inputs. Finally, the node which has to be selected as CH should have high stability and low energy consumption for data transmission. However, the unexpected failure of CHs is not addressed which may lead to loss of data.

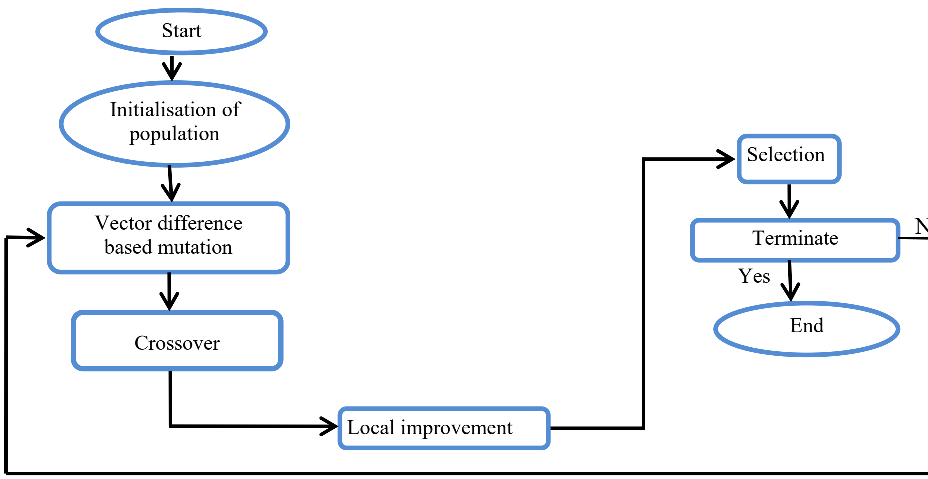
6 *Hybrid LEACH and FCM:* Bouyer et al. (2015) introduced the combination of LEACH and FCM algorithm for optimal selection of CHs. This is proposed to overcome the demerits of the LEACH protocol by combining the LEACH and FCM algorithms. In contrary to LEACH, the node with lesser energy and long distance to BS cannot be selected as CH in this algorithm. The clustering is done using FCM. The node chooses either to be normal or CH based on the fuzzy probability. Each node participates in the CH election using parameters energy, distribution and the centre localisation and compares with neighbours. This hybrid protocol reduces energy consumption due to the combination of LEACH and FCM. However, the hybridisation of both LEACH and FCM in this algorithm should have been detailed clearer for better understanding. Few other parameters might have been taken into account for performance comparison.

### *4.9 Other clustering approaches*

1 *DECA:* Kuila and Jana (2014) proposed a differential evolution (DE)-based clustering algorithm which prevents the faster death of the highly loaded CHs thereby extending the network lifetime of the WSN. A local improvement scheme is incorporated into the DE algorithm for better performance and faster convergence of the algorithm. Bootstrapping and Clustering are two phases used to setup the network. In bootstrapping process, IDs are assigned to all sensor nodes and

gateways. Gateways collect the ID information from sensor nodes within its range and transfer to the BS. In the clustering phase, the DE-based clustering algorithm is utilised to perform clustering. The DE-based clustering algorithm is depicted in Figure 18. The fitness function is determined with the consideration of parameters including standard deviation of the lifetime of CHs and average cluster distance. Mutation and Crossover are performed, followed by the local improvement phase which is necessary for improving the network lifetime. This can be achieved by reassigning a member sensor node from gateway with the lowest lifetime to the one having maximum. Finally, selection process is carried out which selects the best vector with the maximum fitness value. This algorithm improves efficiency and prevents the earlier death of nodes. The convergence time is faster and it involves lower energy consumption which leads to less number of dead nodes. The demerit is that it does not adopt multi-hop communication.

**Figure 18** DE-based clustering algorithm (see online version for colours)




2 *PDC:* A prediction-based data aware clustering protocol (PDC) is proposed by Ashouri et al. (2015) for reducing transmissions throughout the network thereby minimising energy consumption. It makes use of spatial and temporal correlation to form highly stable clusters. The method utilises a TAG-like tree topology. CHs do not use direct communication, instead they utilise local prediction models for forecasting readings in the network. CHs compute the prediction model whereas members and the sink maintain a copy of it. Two queues for maintaining sensed values (SQ) and predicted values (PQ) are used. The first queue is used only by CHs and second can be used by CH, cluster members and sink node to predict the values. The prediction model is sent to the destination by the CH and both the destination and the CH employ the model for predicting measurements and validating the precisions respectively. CH constantly compares the sensed values with the predicted ones. If the value of the error exceeds the threshold, it updates the model and informs to all nodes. The clustering approach is known as CH centric clustering. Temporal

correlation of data sequences is used to build a prediction model with low computation. It utilises prediction with data aware clustering and formed long lived clusters. This PDC method achieves high accuracy and has low communication and computation cost. It conserves energy to a great extent because of no regular transmission of data.

3 *SCCH:* Izadi et al. (2015) proposed a self-configurable clustering mechanism (SCCH) for earlier detection of disordered CHs, which may fail and cannot undergo transmissions. In this work, Type 2 fuzzy logic system (T2FLS) for acquiring the appropriate CHs in each cluster. The inputs of the fuzzy system include energy, NC and local distance (LD). Based on the levels of these input variables (low, medium and high), the output of the system can be classified into five levels including very low (VLow), low, medium, high and very high (VHigh). The output of the FLS for each sensor node will then be transmitted to the neighbours in the form of beacon message which contains Packet-Type, Node-ID and FLS-output. After comparing the value with its own, the node with the maximum FLS-output value will become the CH. Each selected CHs have backup CHs (BCHs) that can be selected from the lowest prioritised nodes. The CH joining message involves the packet-type, CH-ID and BCH-ID. The member nodes transmit data according to the TDMA schedule. This action leads to depletion of energy. Every time, CHs update their energy level and compares itself with the pre-set energy. If it degrades below the level, CHs inform their nodes and BCH. Lastly, BCH takes the role of CH during its power failure thereby providing consistent operation. It helps in achieving longer lifetime and involves less overhead. The data loss rate is less and also has low traffic overhead. It cannot meet QoS requirements.

4 *A distributed Clustering approach:* Pal et al. (2015) proposed a distributed clustering approach to have balanced clusters. Initially, CHs are selected based on the probabilistic approach and is followed by cluster formation. The cluster formation, includes initial cluster formation phase and rescue phase. The initial cluster formation phase introduces two parameters such as $Th_{cluster}$ and $Th_{distance}$. $Th_{cluster}$ chooses the amount of cluster members for every cluster and can be determined by the ratio of number of nodes in the network to the number of CHs. If member nodes of the CH is less than the $Th_{cluster}$ value, nodes join the cluster else wait for the rescue phase. Condition for Initial Cluster formation Phase: No of nodes in a cluster ≤ $Th_{cluster}$. Nodes which do not join any cluster will enter into rescue phase and join the nearest CH. These nodes will join based on the condition that the CH is positioned at a distance lower than the $Th_{distance}$ and the condition for phase-1. $Th_{distance}$ is computed based on the trade-off between cluster size and total cluster distance. Condition for rescue phase is distance of the node from the CH ≤ $Th_{distance}$. The cluster setup is pursued by the data transmission according to TDMA schedule. All cluster member nodes remain in sleep state except in the time slot allotted to them. Thus, it saves energy. The re-clustering happens after the current round. This scheme achieves better cluster quality and can form balanced clusters. The death rate of nodes is low and have extended lifetime.

5 *EELBCA:* Kuila and Jana (2015) introduced an energy efficient and load balancing clustering algorithm for WSN named EELBCA. This algorithm provides the load-balance of the CHs in terms of the number of cluster members. Another algorithm, parameter-based clustering (PBCA) is proposed by authors which balances the load of the CHs in terms of energy consumption. The EELBCA network setup consists of bootstrapping and clustering. The bootstrapping involves assigning IDs to all sensor nodes and gateways. Sensor nodes are divided into restricted and open nodes where the former one can communicate with only one gateway whereas the latter can communicate with more than one gateway. The min-heap is constructed with all gateways after assigning gateways to restricted nodes. It is built in such a way that the root node is the minimum loaded gateway. In PBCA algorithm, the gateway load and energy consumption of the nodes are considered. The node having the least chance of getting gateway is assigned first which is done by sorting of the possible gateways in non-decreasing order. This provides better load balancing for static networks but cannot be applicable for dynamic networks.

6 *ETPSO-CR:* Merabtine et al. (2019) proposed balanced clustering approach for WSN using three techniques. This approach can be applied to any centralised clustering protocol. The BS runs the dynamic round time calculation at the beginning of each round for calculating its duration with inputs such as ratio of energy, round and the vector of residual energies of CHs. With the help of a threshold value, the optimal round duration is selected which is the minimum among all calculated durations. The final round duration is selected in a way that the each CH's residual energy is lesser than the respective threshold. Apart from periodic traffic alone, all other traffic can be considered such as event-driven and query-driven.

7 *BPA-CRP:* A balanced power aware clustering and routing protocol (BPA-CRP) is introduced by Darabkh et al. (2019) for proper utilisation of limited WSN resources. The network is divided into equal layers and nodes within layers are grouped into clusters. Each cluster has CH and there are forwarders in every layer which collect and forward data from each CH in its layer. The number of forwarders for each layer is selected depends on the crossover distance and layer diagonal. The reference points are determined and nodes close to them are selected as forwarders. The forwarders need to be reselected only if the energy falls below the threshold value. The CHs are selected in a round-robin fashion in order to avoid control overhead in such way the energy is saved. The lifetime of the network is further extended in a way by making the nodes which drain maximum energy to work under 'only normal' mode. Nodes operating under this mode cannot participate in forwarder/CH selection.

## 5 Comparison

This section involves the comparison of the above given clustering algorithms and their objectives and merits are highlighted as shown in Table 2. Table 3 gives the comparison of these clustering algorithms regarding parameters and performance improvement of each surveyed protocols.

**Table 2** Comparison of clustering algorithms

| *Sl. no.* | *Protocol/method* | *Node type* | *Clustering process* | *Mobility* | *Highlights* |
|---|---|---|---|---|---|
| 1 | LEACH-SM | Homogeneous | Distributed | Stationary | Includes spare selection and management of spare nodes to LEACH protocol |
| 2 | O-LEACH | Homogeneous | Centralised | Stationary | CHs execute fusion function to reduce correlated data within clusters |
| 3 | ASLPR | Homogeneous | Distributed | Stationary | Adaptively tune its parameters to prolong network lifetime |
| 4 | DCHM | Homogeneous | Centralised | Stationary | Integration of clustering algorithm, reputation and trust systems and data fusion |
| 5 | RDIF | Homogeneous | Centralised | Stationary | Information fusion algorithm for extraction of credible data |
| 6 | R-LEACH | Homogeneous | Distributed | Stationary | Similar to LEACH but considers residual energy for selection of CHs from second round |
| 7 | LEACH-SF | Homogeneous | Centralised | Stationary | Uses fuzzy C-means algorithm for clustering and Sugeno fuzzy model for CH election |
| 8 | M-LEACH | Homogeneous | Centralised | Stationary | The node's remaining energy is considered for the CH selection |
| 9 | REAC-IN | Homogeneous | Distributed | Stationary | Prevents the formation of isolated nodes |
| 10 | EDIT | Homogeneous | Distributed | Stationary | Analyses energy-delay trade-off |
| 11 | SEECH | Homogeneous | Distributed | Stationary | Selects cluster heads and relay nodes separately |
| 12 | DHCR | Can be heterogeneous | Distributed | Stationary | Constructs minimum spanning tree with BS at the root and CHs at edges |
| 13 | ENC | Heterogeneous | Distributed | Stationary | Provides network-wide energy neutral operation and CHG mechanism reduces cluster re-formation |
| 14 | SSTBC | Homogeneous | Centralised | Stationary | Achieves energy efficiency by turning off radio of unnecessary nodes |
| 15 | Bee-Sensor-C | Homogeneous | Distributed | Stationary | Achieves balanced energy consumption and improves scalability using dynamic clustering scheme |
| 16 | HEBM | Homogeneous | Distributed | Stationary | Uses Adaptive clustering and achieves load balancing |
| 17 | ECGD | Homogeneous | Hybrid | Stationary | Uses the concept of dual CH mechanism for reducing energy consumption |
| 18 | ACP | Homogeneous | Distributed | Stationary | Stable nodes are selected as CHs because initially temporary CHs are selected and then refined to final CHs |
| 19 | ACO-C | - | Centralised | Stationary | Extracts local characteristics of data points and achieves scalability |
| 20 | Balanced PSOGSA | Heterogeneous | Distributed | Stationary | Reduces the probability of trapping in local optimum by combining PSO ad GSA algorithms |

**Table 2** Comparison of clustering algorithms (continued)

| *SI. no.* | *Protocol/method* | *Node type* | *Clustering process* | *Mobility* | *Highlights* |
|---|---|---|---|---|---|
| 21 | SNS | Homogeneous | Distributed | Stationary | Provides efficient data delivery procedure for clustering algorithms using single node selection |
| 22 | TPSO-CR | Homogeneous and heterogeneous | Centralised | Stationary | Employs PSO-based clustering and routing algorithm for finding the optimal set of CHs and routing tree |
| 23 | E-OEERP | Homogeneous | Centralised | Stationary | Eliminates the formation of residual nodes by employing PSO for clustering and GSA for routing |
| 24 | ROL/NDC | Homogeneous | Distributed | Stationary | Load balancing at both network level and cluster level |
| 25 | FGF | Homogeneous | Distributed | Stationary | Optimal CH selection using hybrid algorithm |
| 26 | DTLBO | Homogeneous | Centralised | Stationary | Using discrete version of the table learning optimisation algorithm for hierarchical clustering |
| 27 | Thermal aware clustering | Homogeneous | N/A | Stationary | Uses hybrid fuzzy logic and spider optimisation algorithm for clustering and CH selection |
| 28 | MSA | Homogeneous | Distributed | Mobile sink | Employs both static and mobile sinks for achieving energy efficiency |
| 29 | MSIEEP | Homogeneous | Centralised | Mobile sink | Elimination of energy hole problem |
| 30 | LBC-DDU | Homogeneous | Distributed | Mobile collector | Provides distributed load balanced clustering and dual data uploading for mobile data collection |
| 31 | Energy efficient alg. with mobile sink | Homogeneous | Distributed chain formation | Stationary nodes with mobile sink | The clustering scheme is combined with the sink mobility and a chain is used for intercluster transmission |
| 32 | Data collection using MN, PEABR | Heterogeneous | N/A | Both static and mobile nodes | Uses mobile nodes for collecting data from the CH nodes after clustering the network using binary K-means clustering |
| 33 | SEEC and M-SEEC | Heterogeneous | Distributed | Stationary | Uses advanced nodes (AN), super nodes (SN) and normal nodes (NN) |
| 34 | MHCOA | Heterogeneous | Distributed | Stationary | Introduces multi hop cluster based optimisation algorithm network to save network average energy and to improve the lifetime |
| 35 | Effective clustering protocol | Heterogeneous | Distributed | Stationary | Performs clustering for advanced nodes and normal nodes undergo direct transmission to the BS |
| 36 | ETASA | Heterogeneous | Centralised | Stationary | Nodes are formed as pairs to reduce redundant transmission. TDMA schedule is modified for improving the sleep awake scheme |
| 37 | MEACBM | Heterogeneous | N/A | Static and mobile | It employs remaining energy factor into the threshold value for CH selection which helps in improving the network lifetime |

**Table 2** Comparison of clustering algorithms (continued)

| *SI. no.* | *Protocol/method* | *Node type* | *Clustering process* | *Mobility* | *Highlights* |
|---|---|---|---|---|---|
| 38 | Hybrid CS | Homogeneous | Partially centralised and distributed | Stationary | Uses hybrid compressive sensing (CS) for reducing number of transmissions |
| 39 | ACT protocol | Homogeneous | Centralised | Stationary | Incorporates variable cluster sizes far and near the BS based on relaying the load |
| 40 | H-DHAC | Homogeneous | Distributed | Stationary | Hybrid protocol which utilises both quantitative location data and qualitative connectivity data for clustering |
| 41 | FMCR-CT | Homogeneous | Distributed | Stationary | Clustering is performed in alternate rounds to reduce the energy consumption in CH election |
| 42 | FL-EEC/D | Homogeneous | Distributed | Stationary | Uses fuzzy logic for clustering, focuses on the distribution of CHs and uses Gini index to measure the distribution |
| 43 | LEACH-FC | Both homogeneous and heterogeneous | Centralised | Stationary | Fuzzy logic is used for the CH selection and cluster formation algorithm |
| 44 | MACHFL-FT | Heterogeneous | Centralised | Stationary | Three types of clustering is used for every round |
| 45 | Multi-criterion FL based approach | Homogeneous | Distributed | Stationary | Stable nodes are selected as CHs at two levels with different parameters |
| 46 | Hybrid LEACH&FCM | Homogeneous | Centralised | Stationary | It provides the advantages of LEACH and FCM algorithms |
| 47 | DECA | Heterogeneous | Centralised | Stationary | Prevents the faster death of highly loaded CHs using novel Differential evolution based clustering |
| 48 | PDC | Homogeneous | Distributed | Stationary | Forms highly stable clusters and reduces transmissions |
| 49 | SCCH | Homogeneous | Distributed | Stationary | Detects and replaces disordered CHs using a self-configurable clustering algorithm |
| 50 | A distributed clustering approach | Homogeneous | Distributed | Stationary | Forms balanced clusters by the consideration of thresholds |
| 51 | EELBCA | Homogeneous | Centralised | Stationary | Two algorithms for load balancing is proposed where one with the consideration of number of allotted nodes for CH and other with the energy consumption |
| 52 | ETPSO-CR | Both homogeneous and heterogeneous | Centralised | Stationary | CH rotation is based on the residual energy. It introduces some factor into objective function for clustering |
| 53 | BPA-CRP | Homogeneous | Distributed | Stationary | Batch based clustering algorithm where the CHs are rotated for every round. It uses ‘only-normal’ operation nodes which are draining out of energy |

**Table 3** Performance comparison of clustering algorithms

| *SI. no.* | *Protocol* | *Simulation tool* | *Objectives* | | | | *Performance* |
|---|---|---|---|---|---|---|---|
| | | | *Energy efficiency* | *Network lifetime* | *Scalability* | *Others* | |
| 1 | LEACH-SM | - | High | High | High | Reduces redundant data transmission | WSN lifetime –23% better than LEACH |
| 2 | O-LEACH | Matlab | High | High | Medium | Longer stability period | More efficient than LEACH and LEACH-C |
| 3 | ASLPR | Matlab | High | High | Medium | High stability, higher computational complexity | Outperforms LEACH, LEACH-DT and LEACH-EP in terms of lifetime ad received data packets |
| 4 | DCHM | NS-3 | Low | Low | Medium | Provides security and accuracy | Improves accuracy of data fusion |
| 5 | RDIF | NS-3 | Low | Low | High | Reliability, security and accuracy of fusion | Better performance in terms of security and accuracy |
| 6 | R-LEACH | Matlab | High | High | Medium | Improved throughput and increase in no of packets to BS | Better performance compared to LEACH, IGHND, GHND and CBDAS |
| 7 | LEACH-SF | - | Medium | High | High | Higher number of data packets received in sink, intracluster distance | Performance improvement over LEACH, LEACH-DT, LEACH-ERE, LEACH-FL, ASPLR protocols |
| 8 | M-LEACH | NS-2 | Low | Medium | Low | Slightly improved throughput | Better performance compared to LEACH protocol. |
| 9 | REAC-IN | Network Simulator-2 (NS-2) | High | High | Medium | Improves stability, higher amount of data received | Better |
| 10 | EDIT | Castalia | High | High (for hop count) | Medium | - | Less delay ad high energy consumption for Euclidean distance compared to hop-count |
| 11 | SEECH | Matlab | High | High | High | Smaller intra cluster energy dissipation, less delay | Lifetime –41% better than LEACH and 10% better than TCAC |
| 12 | DHCR | NS-2 | High | High | Medium | Better distribution of CHs | Lifetime –65% and 69% improvement in random and non-uniform distribution |
| 13 | ENC | Matlab | High | Medium | Medium | Less control message overhead and more total amount of information bits | Maximised network information gathering and better performance compared to LEACH |

**Table 3** Performance comparison of clustering algorithms (continued)

| *SI. no.* | *Protocol* | *Simulation tool* | *Objectives* | | | | *Performance* |
|---|---|---|---|---|---|---|---|
| | | | *Energy efficiency* | *Network lifetime* | *Scalability* | *Others* | |
| 14 | SSTBC | NS-2 | High | High | Medium | - | Lifetime –250% and 33% improvement over LEACH and PEGASIS<br>Energy efficiency –23% better than PEGASIS |
| 15 | Bee-Sensor-C | - | High | High | High | Less control overhead, high packet delivery rate | Control overhead reduces by 40.8%, 53% and 66% compared to Bee-sensor, IEEABR and FF-ant |
| 16 | HEBM | NS-2 | High | Medium | Low | Lesser percent of dead nodes, less latency, higher throughput | Achieves 25%–28% lifetime compared to FEMCHRP and 32%–40% higher compared to DEEAC. Improved transmission delay and more number of rounds |
| 17 | ECGD | NS-2 | Medium | High | Medium | Larger amount of data received by sink | Better performance compared to LEACH-ERE, EIRNG, PEGASIS and TEEN |
| 18 | ACP | NS-2 | Medium | Medium | High | Better CH distribution, number of clusters | Better performance compared to PEGASIS and DEEAC protocols |
| 19 | ACO-C | Matlab | Medium | Medium | High | Accuracy of clustering solution, longer execution time | Better performance compared to k-means, single-linkage and DBSCAN |
| 20 | Balanced PSOGSA | Matlab | High | High | Medium | Reduces the convergence time, high delay | Better compared to PSO and GSA |
| 21 | SNS | Matlab | High | High | Medium | High network throughput and better robustness | Better performance compared to LEACH and E-LEACH |
| 22 | TPSO-CR | Castalia | High | High | Medium | Better network coverage, improves packet delivery rate, high network throughput | Better performance compared to LEACH-C,GA-C and PSO-C in terms of throughput and energy efficiency |
| 23 | Enhanced OEERP | NS-2 | High | High | Medium | Better load balancing, high throughput, higher PDR | Performance improvement compared to existing protocols |
| 24 | ROL | Dingo | High | Medium | Medium | Good load balancing, better end-to-end delay | Improvement in performance over LEACH and reduces end-to-end delay upto 56%, data delivery ratio upto 0.98% over Mires++ |

**Table 3** Performance comparison of clustering algorithms (continued)

| SI. no. | Protocol | Simulation tool | Objectives | | | | Performance |
|---|---|---|---|---|---|---|---|
| | | | *Energy efficiency* | *Network lifetime* | *Scalability* | *Others* | |
| 25 | FGF | Matlab | Medium | Medium | Medium | Better convergence rate | Better over GA, PSO, ABC, GSO, ALO and CS, GAL-LF, FFOA and GOA with respect to alive node analysis |
| 26 | DTLBO | Matlab | High | Medium | Medium | Better local search and data packets at BS | Better performance over exiting algorithms |
| 27 | Thermal aware clustering | Matlab, Zigbee and PIC | Medium | Medium | Medium | Better node mortality, higher number of data messages received | Better performance compared to LEACH and FCPSO |
| 28 | MSA | Matlab | High | High | Medium | High packet delivery ratio, less delay | Performance improvement over LEACH, modified LEACH and artificial bee colony |
| 29 | MSIEEP | Matlab | High | High | Medium | Higher stability period, lower delay. High packet delivery ratio | Lifetime-43.84% over LEACH,30.44% over LEACH-GA, 4.93% 0ver A-LEACH |
| 30 | LBC-DDU | Matlab | High | High | High | Load balancing | Data collection time reduces by 20%, CH energy conservation –60% |
| 31 | Energy efficient alg. with mobile sink | Matlab R2016a | Medium | Medium | Medium | Number of nodes alive, lifetime with different weight calculation, radius are discussed | Better performance compared to CCMAR and ECDRA in terms of lifetime and total energy consumption |
| 32 | Data collection using MN, PEABR | Matlab | N/A | N/A | Medium | Packet loss rate, deviation rate | Shortest path determination for MN |
| 33 | M-SEEC | Matlab | High | High | Medium | High throughput, high stability period, less overhead | Energy consumption –35% of the total energy of the network (smaller than M-DEC –96% and EDDEEC –85%) |
| 34 | MHCOA | NS-2 | High | High | Medium | Less end to end delay | Saves average network energy by 16.7%, lifetime improvement by 38% |
| 35 | Effective clustering protocol | Matlab | High | High | Medium | High stability and high throughput | Better performance compared to LEACH, SEP, DEEC and Z-SEP protocols |
| 36 | ETASA | Matlab | Medium | High | Low | Slightly better throughput | Improvement in performance compared to TEAR and SEED |

**Table 3** Performance comparison of clustering algorithms (continued)

| *Sl. no.* | *Protocol* | *Simulation tool* | *Objectives* | | | | *Performance* |
|---|---|---|---|---|---|---|---|
| | | | *Energy efficiency* | *Network lifetime* | *Scalability* | *Others* | |
| 37 | MEACBM | Matlab | High | High | Medium | Higher number of alive nodes, throughput, number of dead nodes | Better performance compared to EACBM, NDCM and EEHPMDC |
| 38 | Hybrid CS | - | Low | Low | High | Reduction in the number of transmissions | Reduction by 60% in number of transmissions |
| 39 | ACT protocol | NS-2 and Mannasim | High | High | Medium | Balances the energy dissipation in CHs | Longer lifetime compared to LEACH, BCDCP and MR-LEACH |
| 40 | H-DHAC | NS-2 | High | High | Medium | Low cost, higher amount of transmitted data, high reliability | Lifetime –105.84% better than LEACH, 37.02% than LEACH-C and 9.18% from DHAC-CON (until 90% nodes die) |
| 41 | FMCR-CT | N/A | Medium | High | High | Lesser number of dead nodes compared to other algorithms | Better performance compared to MOFCA, EAFCA, EAUCF, FLECH, DECUC, FUCA and DUCF in terms of scalability, network lifetime and energy consumption |
| 42 | FL-EEC/D | .Net | Medium | High | Medium | Higher number of alive nodes, better Gini index | Better performance compared to LEACH, k-means LEACH, FL[1]/D |
| 43 | LEACH-FC | Matlab | Medium | Medium | Low | Number of alive nodes is higher when compared to other algorithms | Better performance compared to LEACH for homogeneous, 2-level and 3-level heterogeneous scenario and multilevel scenario in terms of alive nodes, energy consumption and bits received |
| 44 | MACHFL-FT | N/A | Medium | High | Medium | Lesser number of dead nodes | Better performance compared to IFUC, FEMCHRP, DUCF, and FBUC |
| 45 | Multi-criterion FL based approach | Matlab | High | Medium | High | Better stability period | Outperforms LEACH-E, CHEF and FZLP |
| 46 | Hybrid LEACH and FCM | - | Low | Medium | Low | Lesser number of dead nodes | Better performance compared to LEACH protocol |

**Table 3** Performance comparison of clustering algorithms (continued)

| *SI. no.* | *Protocol* | *Simulation tool* | *Objectives* | | | | *Performance* |
|---|---|---|---|---|---|---|---|
| | | | *Energy efficiency* | *Network lifetime* | *Scalability* | *Others* | |
| 47 | DECA | Matlab | High | High | Low | Faster convergence, less number of death sensor nodes | Better performance compared to DE, GA, LBC and GLBCA and inferior performance to EELBCA |
| 48 | PDC | - | High | Medium | Medium | High accuracy, low computation and communication overhead, efficiency | Performance –15% improvement on average |
| 49 | SCCH | Matlab | High | High | Low | Low data loss rate, less traffic | Outperforms PDD and DBCH-LEACH-C |
| 50 | A distributed clustering approach | Matlab | High | High | Medium | Low node death rate, better cluster quality | 8% improvement over LEACH for 50% node death |
| 51 | EELBCA | Matlab | Medium | Medium | Medium | Better load balancing, packets received and active gateways. | Better performance compared to LBCA and LBC protocols |
| 52 | ETPSO-CR | NS-3, GenSeN tool | High | High | Medium | Better number of data packets received at the BS | Better performance compared to LEACH-C and TPSO-CR in terms of energy consumption, lifetime and data packets received at BS |
| 53 | BPA-CR | Matlab R2014a | High | High | Medium | Better network utilisation | Exceptional performance compared to UCR, COCA and EA-CRP |

## 6 Conclusions

WSN is composed of numerous sensor nodes deployed in an environment to sense the physical parameter The battery charged sensor nodes are difficult to recharge in unmanned areas, once they drain out of energy. This necessitates the utilisation of energy efficiently without wasting. Hierarchical clustering has proved to be one of the energy efficient mechanisms. In clustering, the entire network is divided into various clusters with each cluster having a CH. The non-CH nodes transmit data directly to their respective CH nodes which in turn perform data aggregation over the collected data. The aggregated data will be transmitted by the CHs to the BS. Several mechanisms are available for clustering. Some algorithms perform the CH selection initially followed by the formation of clusters. Some other mechanisms group nodes into clusters based on some criteria and then select CH for each cluster. Although there are various methods of clustering algorithms provided by researchers, there is still research going on in many directions.

In this work, a state-of-the-art review of various clustering algorithms for WSN developed in recent years is provided. Initially, a short introduction is given followed by state-of-the-art surveys of clustering algorithms in WSN. Later, attributes and design challenges of clustering have been presented. The earliest and familiar clustering approaches are discussed. Later, various other algorithms are explained in brief and grouped into LEACH-based, energy-efficient, optimisation, fuzzy logic based and mobility. A performance comparison of those algorithms has also been presented. Most of these algorithms performed well in improving energy efficiency and enhancing the lifetime of the network as these are the major goals in WSN. Recently many algorithms have been emerging using fuzzy-based clustering. This work provides knowledge about the recent techniques and research issues in the field of clustering for researchers in the area of WSN.

## 7 Future directions

Although clustering has been extensively used in various applications and environments, still some areas and problems have not been properly examined. This section provides some future directions in the area of clustering in WSN. One of the research challenges is the effect of mobility in the WSN. Most algorithms consider only static nodes which cannot be applied to MNs. In this literature, many mobility-based algorithms consider only mobility of sink node most of which mitigates the energy hole problem. Only few algorithms are available for dealing with mobile sensor nodes and so further development in this area should be done for handling the topology changes which in turn changes the routing. Because of the development of many applications using mobile sensor nodes, it is challenging for research regarding the formation of clusters with MNs and also maintaining topology changes.

Another section is regarding energy efficiency which is one of the mandatory requirements in this area. Many algorithms are emerging for achieving energy efficiency, but they provide trade off with other parameters. Though some provide a better algorithm, they do not support multi hop communication. Future work can be done to establish a better energy efficient algorithm with the consideration of all the possible parameters. One of the recently emerging techniques is energy harvesting sensor networks which can be extended according to necessary requirements.

Clustering in heterogeneous WSN is another direction where future research can be carried out. It is still a challenge in the localisation and deployment of heterogeneous nodes and also cluster formation in heterogeneous WSN and so future work can be done.

## References

Abbasi, A.A. and Younis, M. (2007) 'A survey on clustering algorithms for wireless sensor networks', *Computer Communications*, Vol. 30, Nos. 14–15, pp.2826–2841.

Abidi, W. and Ezzedine, T. (2019) 'Effective clustering protocol based on network division for heterogeneous wireless sensor networks', *Computing*, Vol. 102, No. 2, pp.413–425.

Abo-Zahhad, M., Ahmed, S.M., Sabor, N. and Sasaki, S. (2015) 'Mobile sink-based adaptive immune energy-efficient clustering protocol for improving the lifetime and stability period of wireless sensor networks', *IEEE Sensors Journal*, Vol. 15, No. 8, pp.4576–4586.

Afsar, M.M. and Tayarani, N.M.H. (2014) 'Clustering in sensor networks: a literature survey', *Journal of Network and Computer Applications*, Vol. 46, pp.198–226.

Al-Azzawi, M.K., Luo, J. and Li, R. (2015) 'Inspired energy efficient data delivery based on redundant data elimination using discrete cuckoo search optimization', *Int. J. Control Autom.*, Vol. 8, No. 2, pp.417–428.

Al-Habashneh, A.A.Y., Ahmed, M.H. and Husain, T. (2011) 'Reliability analysis of wireless sensor networks for forest fire detection', *2011 7th International Wireless Communications and Mobile Computing Conference*, pp.1630–1635, IEEE, July.

Anastasi, G., Conti, M., Di Francesco, M. and Passarella, A. (2009) 'Energy conservation in wireless sensor networks: a survey', *Ad Hoc Networks*, Vol. 7, No. 3, pp.537–568.

Ashouri, M., Yousefi, H., Basiri, J., Hemmatyar, A.M.A. and Movaghar, A. (2015) 'PDC: prediction-based data-aware clustering in wireless sensor networks', *Journal of Parallel and Distributed Computing*, Vol. 81, pp.24–35.

Aslam, M., Javaid, N., Rahim, A., Nazir, U., Bibi, A. and Khan, Z.A. (2012) 'Survey of extended LEACH-based clustering routing protocols for wireless sensor networks', *2012 IEEE 14th International Conference on High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems*, pp.1232–1238, IEEE, June.

Bakr, B.A. and Lilien, L.T. (2014) 'Extending lifetime of wireless sensor networks by management of spare nodes', *Procedia Computer Science*, Vol. 34, pp.493–498.

Behera, T.M., Mohapatra, S.K., Samal, U.C., Khan, M.S., Daneshmand, M. and Gandomi, A.H. (2019) 'Residual energy based cluster-head selection in WSNs for IoT application', *IEEE Internet of Things Journal*, Vol. 6, No. 3, pp.5132–5139.

Borges, L.M., Velez, F.J. and Lebres, A.S. (2014) 'Survey on the characterization and classification of wireless sensor network applications', *IEEE Communications Surveys & Tutorials*, Vol. 16, No. 4, pp.1860–1890.

Bouyer, A., Hatamlou, A. and Masadari, M. (2015) 'A new approach for decreasing energy in wireless sensor networks with hybrid LEACH protocol and fuzzy C-means algorithm', *Int. J. Communication Networks and Distributed Systems*, Vol. 14, No. 4, pp.1–13.

Boyinbode, O., Le, H., Mbogho, A., Takizawa, M. and Poliah, R. (2010) 'A survey on clustering algorithms for wireless sensor networks', *2010 13th International Conference on Network-Based Information Systems*, pp.358–364, IEEE, September.

Cai, X., Duan, Y., He, Y., Yang, J. and Li, C. (2015) 'Bee-sensor-C: an energy-efficient and scalable multipath routing protocol for wireless sensor networks', *International Journal of Distributed Sensor Networks*, Vol. 11, No. 3, pp.1–14.

Chaturvedi, A., Goswami, D.N. and Singh, S. (2016) 'Energy efficient cluster head selection for cross layer design over wireless sensor network', *Int. J. Communication Networks and Distributed Systems*, Vol. 16, No. 4, pp.335–351.

Darabkh, K.A., El-Yabroudi M.Z. and El-Mousa, A.H. (2019) 'BPA-CRP: a balanced power-aware clustering and routing protocol for wireless sensor networks', *Ad Hoc Networks*, Vol. 82, pp.155–171.

Dattatraya, K.N. and Rao, K.R. (2019) 'Hybrid based cluster head selection for maximizing network lifetime and energy efficiency in WSN', *J. King Saud Univ. – Comput. Inf. Sci.*, Vol. 1, No. 1, pp.34–46.

Elhabyan, R.S. and Yagoub, M.C. (2015) 'Two-tier particle swarm optimization protocol for clustering and routing in wireless sensor network', *Journal of Network and Computer Applications*, Vol. 52, pp.116–128.

Fanian, F. and Rafsanjani, M.K. (2019) 'Cluster-based routing protocols in wireless sensor networks: a survey based on methodology', *Journal of Network and Computer Applications*, Vol. 142, pp.111–142.

Farouk, F., Rizk, R. and Zaki, F.W. (2014) 'Multi-level stable and energy-efficient clustering protocol in heterogeneous wireless sensor networks', *IET Wireless Sensor Systems*, Vol. 4, No. 4, pp.159–169.

Fu, J.S. and Liu, Y. (2015) 'Double cluster heads model for secure and accurate data fusion in wireless sensor networks', *Sensors*, Vol. 15, No. 1, pp.2021–2040.

Ghaffari, A. (2015) 'Congestion control mechanisms in wireless sensor networks: a survey', *Journal of Network and Computer Applications*, Vol. 52, pp.101–115.

Gherbi, C., Aliouat, Z. and Benmohammed, M. (2016) 'An adaptive clustering approach to dynamic load balancing and energy efficiency in wireless sensor networks', *Energy*, Vol. 114, pp.647–662.

Gherbi, C., Aliouat, Z. and Benmohammed, M. (2018) 'Energy Efficient with time synchronized and service coverage guarantee in wireless sensor networks', *Int. J. Communication Networks and Distributed Systems*, Vol. 21, No. 1, pp.56–79.

Gungor, V.C. and Hancke, G.P. (2009) 'Industrial wireless sensor networks: challenges, design principles, and technical approaches', *IEEE Transactions on Industrial Electronics*, Vol. 56, No. 10, pp.4258–4265.

Hammoudeh, M. and Newman, R. (2015) 'Adaptive routing in wireless sensor networks: QoS optimisation for enhanced application performance', *Information Fusion*, Vol. 22, pp.3–15.

Hamzah, A., Shurman, M., Al-Jarrah, O. and Taqieddin, E. (2019) 'Energy-efficient fuzzy-logic based clustering technique for hierarchical routing protocols in wireless sensor networks', *Sensors*, Vol. 19, No. 3, p.561.

Haque, M., Ahmad, T. and Imran, M. (2018) 'Review of hierarchical routing protocols for wireless sensor networks', *Intelligent Communication and Computational Technologies*, pp.237–246, Springer, Singapore.

Heinzelman, W.B., Chandrakasan, A.P. and Balakrishnan, H. (2002) 'An application-specific protocol architecture for wireless microsensor networks', *IEEE Transactions on wireless Communications*, Vol. 1, No. 4, pp.660–670.

Heinzelman, W.R., Chandrakasan, A. and Balakrishnan, H. (2000) 'Energy-efficient communication protocol for wireless microsensor networks', *Proceedings of the 33rd annual Hawaii International Conference on System Sciences*, pp.10–19, IEEE, January.

Hu, S., Han, J., Wei, X. and Chen, Z. (2015) 'A multi-hop heterogeneous cluster-based optimization algorithm for wireless sensor networks', *Wireless Networks*, Vol. 21, No. 1, pp.57–65.

Huynh, T.T., Dinh-Duc, A.V., Tran, C.H. and Le, T.A. (2015) 'Balance particle swarm optimization and gravitational search algorithm for energy efficient in heterogeneous wireless sensor networks', *The 2015 IEEE RIVF International Conference on Computing & Communication Technologies-Research, Innovation, and Vision for Future (RIVF)*, pp.175–179, IEEE, January.

İnkaya, T., Kayalıgil, S. and Özdemirel, N.E. (2015) 'Ant colony optimization based clustering methodology', *Applied Soft Computing*, Vol. 28, pp.301–311.

Izadi, D., Abawajy, J. and Ghanavati, S. (2015) 'An alternative clustering scheme in WSN', *IEEE Sensors Journal*, Vol. 15, No. 7, pp.4148–4155.

Jan, B., Farman, H., Javed, H., Montrucchio, B., Khan, M. and Ali, S. (2017) 'Energy efficient hierarchical clustering approaches in wireless sensor networks: a survey', *Wireless Communications and Mobile Computing*, Vol. 2017, pp.1–4.

Jiang, C., Yuan, D. and Zhao, Y. (2009) 'Towards clustering algorithms in wireless sensor networks-a survey. In *2009 IEEE Wireless Communications and Networking Conference*, pp.1–6, IEEE, April.

Jindal, P. and Gupta, V. (2013) 'Study of energy efficient routing protocols of wireless sensor networks and their further researches: a survey', *International Journal of Computer Science and Communication Engineering*, Vol. 2, No. 2, pp.2, No. 2, pp.57–62.

Jose, D.V. and Sadashivappa, G. (2015) 'Mobile sink assisted energy efficient routing algorithm for wireless sensor networks', *World of Computer Science & Information Technology Journal*, Vol. 5, No. 2, pp.16–22.

Khediri, S.E., Nasri, N., Wei, A. and Kachouri, A. (2014) 'A new approach for clustering in wireless sensors networks based on LEACH', *Procedia Computer Science*, Vol. 32, pp.1180–1185.

Kuila, P. and Jana, P.K. (2014) 'A novel differential evolution based clustering algorithm for wireless sensor networks', *Applied Soft Computing*, Vol. 25, pp.414–425.

Kuila, P. and Jana, P.K. (2015) 'Heap and parameter-based load balanced clustering algorithms for wireless sensor networks', *Int. J. Communication Networks and Distributed Systems*, Vol. 14, No. 4, pp.413–432.

Kumarawadu, P., Dechene, D.J., Luccini, M. and Sauer, A. (2008) 'Algorithms for node clustering in wireless sensor networks: a survey', *2008 4th International Conference on Information and Automation for Sustainability*, pp.295–300, IEEE, December.

Lai, W.K., Fan, C.S. and Lin, L.Y. (2012) 'Arranging cluster sizes and transmission ranges for wireless sensor networks', *Information Sciences*, Vol. 183, No. 1, pp.117–131.

Lata, S., Mehfuz, S., Urooj, S. and Alrowais, F. (2020) 'Fuzzy clustering algorithm for enhancing reliability and network lifetime of wireless sensor networks', *IEEE Access*, Vol. 8, pp.66013–66024.

Leu, J.S., Chiang, T.H., Yu, M.C. and Su, K.W. (2014) 'Energy efficient clustering scheme for prolonging the lifetime of wireless sensor network with isolated nodes', *IEEE Communications Letters*, Vol. 19, No. 2, pp.259–262.

Lin, D. and Wang, Q. (2019) 'An energy-efficient clustering algorithm combined game-theory and dual cluster-head mechanism for WSN', *IEEE Access*, Vol. 7, pp.49894–49905.

Lindsey, S. and Raghavendra, C.S. (2002) 'PEGASIS: power-efficient gathering in sensor information systems', *Proceedings, IEEE Aerospace Conference*, Vol. 3, pp.1125–1130, IEEE, March.

Liu, X. (2012) 'A survey on clustering routing protocols in wireless sensor networks', *Sensors*, Vol. 12, No. 8, pp.11113–11153.

Lotf, J.J., Hosseinzadeh, M. and Alguliev, R.M. (2010) 'Hierarchical routing in wireless sensor networks: a survey', *2010 2nd International Conference on Computer Engineering and Technology*, Vol. 3, pp.650–654, IEEE, April.

Lv, C., Wang, Q., Yan, W. and Shen, Y. (2016) 'Energy-balanced compressive data gathering in wireless sensor networks', *Journal of Network and Computer Applications*, Vol. 61, pp.102–114.

Ma, T., Liu, Y., Fu, J. and Jing, Y. (2015) 'A reliable information fusion algorithm for reputation based wireless sensor networks', *International Journal of Future Generation Communication and Networking*, Vol. 8, No. 1, pp.281–298.

Maimour, M., Zeghilet, H. and Lepage, F. (2010) 'Cluster-based routing protocols for energy-efficiency in wireless sensor networks', in *Sustainable Wireless Sensor Networks*, pp.167–188, InTech.

Manjeshwar, A. and Agrawal, D.P. (2001) 'TEEN: a routing protocol for enhanced efficiency in wireless sensor networks', *15th IEEE International Symposium on Parallel and Distributed Processing*, Vol. 1, pp.2009–2015, April.

Manjeshwar, A. and Agrawal, D.P. (2002) 'APTEEN: a hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks', *International Symposium on Parallel and Distributed Processing*, IEEE, April.

Masdari, M. and Barshandeh, S. (2020) 'Discrete teaching-learning-based optimization algorithm for clustering in wireless sensor network', *J. of Ambient Intelligence and Humanized Computing*, https://doi.org/10.1007/s12652-020-01902-6.

Mazinani, A., Mazinani, S.M. and Mirzaie, M. (2019) 'FMCR-CT: an energy-efficient fuzzy multi-cluster-based routing with a constant threshold in wireless sensor network', *Alexandria Engineering Journal*, Vol. 58, No. 1, pp.127–141.

Mehrani, M., Shanbehzadeh, J., Sarrafzadeh, A., Mirabedini, S.J. and Manford, C. (2010) 'FEED: fault tolerant, energy efficient, distributed clustering for WSN', *2010 The 12th International Conference on Advanced Communication Technology (ICACT)*, Vol. 1, pp.580–585, IEEE, February.

Merabtine, N., Djenouri, D., Zegour, D., Boumessaidia, B. and Boutahraoui, A. (2019) 'Balanced clustering approach with energy prediction and round-time adaptation in wireless sensor networks', *Int. J. Communication Networks and Distributed Systems*, Vol. 22, No. 3, pp.245–274.

Mhatre, V. and Rosenberg, C. (2004) 'Homogeneous vs. heterogeneous clustered sensor networks: a comparative study', *2004 IEEE International Conference on Communications (IEEE Cat. No. 04CH37577)*, Vol. 6, pp.3646–3651, IEEE, June.

Mirzaie, M. and Mazinani, S.M. (2019) 'MACHFL-FT: a fuzzy logic based energy-efficient protocolto cluster heterogeneous nodes in wireless sensor networks', *Wireless Networks*, Vol. 25, No. 8, pp.4597–4609.

Mishra, N.K., Jain, V. and Sahu, S. (2013) 'Survey on recent clustering algorithms in wireless sensor networks', *International Journal of Scientific and Research Publications*, Vol. 3, No. 4, pp.1–4.

Mitra, R. and Nandy, D. (2012) 'A survey on clustering techniques for wireless sensor network', *International Journal of Research in Computer Science*, Vol. 2, No. 4, pp.51–57.

Movassaghi, S., Abolhasan, M., Lipman, J., Smith, D. and Jamalipour, A. (2014) 'Wireless body area networks: a survey', *IEEE Communications Surveys & Tutorials*, Vol. 16, No. 3, pp.1658–1686.

Pal, V., Singh, G. and Yadav, R.P. (2015) 'Balanced cluster size solution to extend lifetime of wireless sensor networks', *IEEE Internet of Things Journal*, Vol. 2, No. 5, pp.399–401.

Pang, A., Chao, F., Zhou, H. and Zhang, J. (2020) 'The method of data collection based on multiple mobile nodes for wireless sensor network', *IEEE Access*, Vol. 8, pp.14704–14713.

Peng, S., Wang, T. and Low, C.P. (2015) 'Energy neutral clustering for energy harvesting wireless sensors networks', *Ad Hoc Networks*, Vol. 28, pp.1–16.

Priya, R.K. and Venkatanarayanan, S. (2020) 'Implementation of thermal aware wireless sensor network clustering algorithm based on fuzzy and spider optimized cluster head selection', *J. of Ambient Intelligence and Humanized Computing*, pp.1–11, https://doi.org/10.1007/s12652-020-02002-1.

Rajagopalan, R. and Varshney, P.K. (2006) 'Data aggregation techniques in sensor networks: a survey', *IEEE Communications Surveys & Tutorials*, Vol. 8, No. 4, pp.48–63.

Ramesh, K. and Somasundaram, D.K. (2012) 'A comparative study of clusterhead selection algorithms in wireless sensor networks', *Int. J. of Computer Science & Engineering Survey*, Vol. 2, No. 4, pp.153–164.

Ranga, V., Dave, M. and Verma, A.K. (2016) 'A multi-criterion fuzzy logic intra-cluster and inter-cluster based stable cluster-head election approach in large scale WSN', *Int. J. Communication Networks and Distributed Systems*, Vol. 17, No. 4, pp.1–17.

Rejina Parvin, J. and Vasanthanayaki, C. (2015) 'Particle swarm optimization-based clustering by preventing residual nodes in wireless sensor networks', *IEEE Sensors Journal*, Vol. 15, No. 8, pp.4264–4274.

Sabet, M. and Naji, H.R. (2015) 'A decentralized energy efficient hierarchical cluster-based routing algorithm for wireless sensor networks', *AEU-International Journal of Electronics and Communications*, Vol. 69, No. 5, pp.790–799.

Savitha, M. and Patel, S.D. (2014) 'A survey on clustering techniques in wireless sensor networks', *Int. J. of Emerging Tech. and Adv. Eng.*, Vol. 4, No. 1, pp.28–32.

Shagari, N.M., Idna Iris, M.Y., Salleh, R.B., Ahmedy, I., Murtaza, G. and Shehadeh, H.A. (2020) 'Heterogeneous energy and traffic aware sleep-awake cluster-based routing protocol for wireless sensor network', *IEEE Access*, Vol. 8, pp.12232–12252.

Shahraki, A., Taherkordi, A., Haugen, O. and Eliassen, F. (2020) 'Clustering objectives in WSN: a survey and research direction analysis', *Computer Networks*, (In Press), https://doi.org/10.1016/j.comnet.2020.107376.

Shokouhifar, M. and Jalali, A. (2015) 'A new evolutionary based application specific routing protocol for clustered wireless sensor networks', *AEU-International Journal of Electronics and Communications*, Vol. 69, No. 1, pp.432–441.

Shokouhifar, M. and Jalali, A. (2017) 'Optimized sugeno fuzzy clustering algorithm for wireless sensor networks', *Engineering Applications of Artificial Intelligence*, Vol. 60, pp.16–25.

Shouyi, Y., Leibo, L., Renyan, Z., Zhongfu, S. and Shaojun, W. (2013) 'Design of wireless multi-media sensor network for precision agriculture', *China Communications*, Vol. 10, No. 2, pp.71–88.

Singh, S.P. and Sharma, S.C. (2015) 'A survey on cluster based routing protocols in wireless sensor networks', *Procedia Computer Science*, Vol. 45, pp.687–695.

Tan, N.D. and Viet, N.D. (2015) 'SSTBC: Sleep scheduled and tree-based clustering routing protocol for energy-efficient in wireless sensor networks', *The 2015 IEEE RIVF International Conference on Computing & Communication Technologies-Research, Innovation, and Vision for Future (RIVF)*, pp.180–185, IEEE, January.

Tarhani, M., Kavian, Y.S. and Siavoshi, S. (2014) 'SEECH: scalable energy efficient clustering hierarchy protocol in wireless sensor networks', *IEEE Sensors Journal*, Vol. 14, No. 11, pp.3944–3954.

Thakkar, A. and Kotecha, K. (2014) 'Cluster head election for energy and delay constraint applications of wireless sensor network', *IEEE Sensors Journal*, Vol. 14, No. 8, pp.2658–2664.

Toor, A.S. and Jain, A.K. (2019) 'Energy aware cluster based multi-hop energy efficient routing protocol using multiple mobile nodes (MEACBM) in wireless sensor networks', *AEU-International Journal of Electronics and Communications*, Vol. 102, pp.41–53.

Ullah, Z. (2020) 'A survey on hybrid, energy efficient and distributed (HEED) based energy efficient clustering protocols for wireless sensor networks', *Wireless Personal Communications*, Vol. 112, No. 4, pp.41–53.

Wang, J., Gao, Y., liu, W., Sangaiah, A.K. and Kim, H.J. (2019) 'Energy efficient routing algorithm with mobile sink support for wireless sensor networks', *Sensors*, Vol. 19, No. 7, p.1494.

Wei, C., Yang, J., Gao, Y. and Zhang, Z. (2011) 'Cluster-based routing protocols in wireless sensor networks: a survey', *Proceedings of 2011 International Conference on Computer Science and Network Technology*, Vol. 3, pp.1659–1663, IEEE, December.

Xie, R. and Jia, X. (2014) 'Transmission-efficient clustering method for wireless sensor networks using compressive sensing', *IEEE Transactions on Parallel and Distributed Systems*, Vol. 25, No. 3, pp.806–815.

Xu, D. and Gao, J. (2011) 'Comparison study to hierarchical routing protocols in wireless sensor networks', *Procedia Environmental Sciences*, Vol. 10, pp.595–600.

Yau, K.L.A., Ramli, N., Hashim, W. and Mohamad, H. (2014) 'Clustering algorithms for cognitive radio networks: a survey', *Journal of Network and Computer Applications*, Vol. 45, pp.79–95.

Yi, S., Heo, J., Cho, Y. and Hong, J. (2007) 'PEACH: power-efficient and adaptive clustering hierarchy protocol for wireless sensor networks', *Computer Communications*, Vol. 30, Nos. 14-15, pp.2842–2852.

Younis, O. and Fahmy, S. (2004) 'HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks', *IEEE Transactions on Mobile Computing*, Vol. 3, No. 4, pp.366–379.

Younis, O., Krunz, M., and Ramasubramanian, S. (2006) 'Node clustering in wireless sensor networks: recent developments and deployment challenges', *IEEE Network*, Vol. 20, No. 3, pp.20–25.

Zanjireh, M.M. and Larijani, H. (2015) 'A survey on centralised and distributed clustering routing algorithms for WSNs', *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp.1–6, IEEE, May.

Zhao, M., Yang, Y. and Wang, C. (2014) 'Mobile data gathering with load balanced clustering and dual data uploading in wireless sensor networks', *IEEE Transactions on Mobile Computing*, Vol. 14, No. 4, pp.770–785.

Zhu, J., Lung, C.H. and Srivastava, V. (2015) 'A hybrid clustering technique using quantitative and qualitative data for wireless sensor networks', *Ad Hoc Networks*, Vol. 25, Part A pp.38–53.