Computational Intelligence based Clustering Algorithms for Wireless Sensor Networks: Trends and Possible Solutions

Nitin Mittal

1Dept. of ECE, Chandigarh University, Mohali-140413, Punjab, India
mittal.nitin84@gmail.com

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: A wireless sensor network (WSN) is a state-of-the-art technology for radio communication. A WSN includes several sensors that are arbitrarily distributed in a particular region to detect and track physical characteristics that are hard for humans to observe, like temperature, humidity, and pressure. Because of the nature of WSNs, many issues may arise, including information routing, power consumption, clustering, and cluster head (CH) selection. Although there are still some difficulties in the WSN, owing to its versatility and robustness, it has gained considerable attention among scientists and technologists despite the shortcomings. Various protocols were designed to solve these problems. Low energy adaptive clustering hierarchy (LEACH) is one of the significant hierarchical protocols used to reduce energy consumption in WSNs. This article provides an extensive analysis of LEACH-variant clustering protocols for WSNs. Recent research on Machine Learning, Computational Intelligence, and WSNs has highlighted the optimized WSN clustering algorithms. However, the selection of a suitable paradigm for a clustering solution continues an issue owing to the diversity of WSN applications. In this paper, a comprehensive review of suggested optimized clustering alternatives has been conducted and a comparison of these optimized clustering methods has been suggested based on various performance parameters. The centralized clustering approaches based on the Swarm Intelligence paradigm are observed to be more suitable for the applications in which low energy is required, high information delivery rate, or elevated scalability than algorithms that are based on the other paradigms described.

Keywords: LEACH protocol; cluster formation; WSN; energy consumption; meta-heuristic; computational intelligence; machine learning.

1. Introduction

Real-time implementation has received prominent attention among technocrats and scientists with the latest innovations in the field of sensors. To eliminate the sensor problems, technologists and scientists discovered a solution by deploying Wireless Sensor Network (WSN)[1] real-time applications. The real-time sensors will instantly detect, record and send feedback to the end-user for further processing of the data acquired. WSNs enabling a network-limited delay assurance, vital for packet delivery from end-to-end, are referred to as real-time WSN[2]. WSN is of utmost significance for the search region and data assortment in computer networking. WSN discovers its implementation in several fields, including data storage and tracking. The range of WSN applications has grown enormously due to fast urbanization. WSN has countless sensors that are transmitted in a broad region through packets by transferring information from one sensor to another sensor. WSN is widely used in different areas like military surveillance, healthcare, and other industries[3]. Two types of nodes are there: SNs and gateway node shown in Fig. 1[4]. The data is sensed by SN and the collected information is transmitted by the RF channel to the gateway node. The gateway node is an interface between WSN and the wired world from where sensed data is analyzed using the software. This gateway node is generally called the base station (BS) or sink since it becomes the receiving point of the data gathered from the sensing devices. Any node can work as the sink node in the network but it needs to have the special characteristic of having a higher capacity to receive bulk data. This feature further helps to concentrate the traffic on BS. BS functions as an internal gateway by interconnecting any network to the whole sensor network via the internet. The user node generally has all the information by connecting to the sink node. It should be remembered that a given target area is overall monitored by nodes and BS is in charge of receiving all the collected information[5].

![Fig. 1. WSN System Architecture.](image-url)
Hardware structural design of a Wireless Sensor

Wireless sensor node is a compact device that can measure various physical quantities like pressure, temperature, light. It communicates with collection center where the data is collected, it communicates directly or by the help of other sensor nodes [6]. It became possible to make the microsensor of a few cubic millimeters in volume by doing advancement in microelectronics, in transmission technologies, and in the software. And these microsensors can be operated in the network at a very reasonable price.

**Acquisition unit:** It consists of sensor and analog-to-digital converter (ADC). Function of sensor is gathering digital measurements of the parameters of the environment and then convert it into analog signals. And then these analog signals are converted into a digital signal with the help of ADC.

**Processing unit:** This comprised of two subunits: First is a data storage unit and second is the processor, the function of the processor is to process the data and to control the procedures that allow the sensors to collaborate with the other sensors to carry out the task of acquisition.

**Communication unit:** The function of this unit is to transmit and receive the data. It consists of pair of transmitter and receiver. In this unit, communication is done within the network, but in this case, communication is done by radio frequency (radio waves). And other transmissions are also possible like optical and infrared.

**Power unit:** Its function is to supply energy to all the units, is somewhat similar to battery, it feed sensor, but the resources are limited which makes it a big problem in this kind of network. But recently it has been realized that power from the solar panel can give a solution for enhancing the life of the sensor [5].

**Mobilizer:** It is used to shift node for completing task that is required to be processed.

**Location-Finding System:** Its purpose is to provide required information on the site by help of routing.

Fig. 2. Architecture of a wireless sensor node.

Types of WSNs

Depending on the employment of WSN in various places:

**Terrestrial:** This type of WSN proposed to be placed in terrestrial space. Many sensors are either situated randomly or deployed deterministically in the required place. These WSN is very useful where environmental monitoring or monitoring of natural phenomena is done. But it has limitation in terms of energy management.

**Underground:** These networks contain a particular type of sensor nodes which is recognized for their higher price and for the logistics necessary for their accomplishment and maintenance. The sensor nodes are being installed under the land. Applications of this type of network are in agriculture, in minefield, where soil monitoring is done. A ground node is there whose function is to transmit the data detected by the nodes which are placed underground.
**Underwater:** These networks are placed beneath sea. Cost is higher as compared to the other sensors, wireless communication is acoustic, limited bandwidth, signal loss is frequent, and propagation delays problems are common.

**Multimedia:** This kind of WSN network is specially intended to monitor and trace the data of images, videos, and sound. It has cameras and microphones. In this, good bandwidth and good quality service is required, due to which power consumption is high. Before deployment of these sensors preplanning is done.

**Mobile:** One of the recent types of WSN, in this the network automatically get reorganized by using the mobile nodes. After installation, the nodes disperse to collect the data. Hybrid network are also there that consist combination of mobile sensor and fixed sensors.

2. **DESIgn ISSUES IN WSN**

Energy efficiency poses as one of the prominent challenges in WSN, as SNs operate on non-rechargeable batteries. Energy consumption in WSNs carried out by three processes: sensing, computation and communication. Energy is consumed during communication as transceiver remains active for longer duration. The research in WSN mainly focus on scheming the routing protocols according to a multi-objective optimization problem that focuses on dropping the delay, enhancing accuracy, improving load balancing, support scalability and wide range of densities along with energy efficiency. This means that a routing protocol offers a good performance for WSN data aggregation if it is able to collect the maximum number of data samples with minimum energy consumption and delay[7]. For this reason, the following issues need to be addressed to design a routing protocol and enhance its performance in WSNs[8].

1. **Energy Efficiency:** SN's lifetime is estimated on the basis of available energy and the average consumption of power. Because maximizing the capacity of energy of small-dimensional batteries still need a lot of work to be done, increasing the amount of energy available makes it difficult to achieve the goal of a long lifetime.

2. **Quality-of-Service (QoS):** The two main metrics in QoS are reliability and latency[9]. Usually, to ensure the network's reliability and functionality, the rate of successful data exchange between SNs and BS must be above a certain level. Reliability can be maximized further, but it can be done by increasing the consumption of energy. When it comes to latency, the receipt of sensed data at the sink node is strictly timed for certain types of applications, especially in industrial monitoring applications.

3. **Fault Tolerance:** WSNs need to be fault tolerant (having self-testing, calibrating and recovering abilities) as they are highly prone to failures especially when deployed unattended in harsh environments[10].

4. **Scalability Issue:** WSNs should be scalable as they can be deployed in different sizes of tens, hundreds or thousands. SNs also may be added to the network which is already deployed. There should be no effect on operation of network due to any change in size of the sensor network[11].

5. **Load balancing:** SNs are evenly distributed in order to balance the load occurring on a node. There should not be excess load on a particular node because it leads to quicker death of that node. Therefore, load balancing is as important as other parameters.

6. **Security and Privacy:** Because SNs are sometimes deployed in accessible areas, the threat of technical interference or human intrusion contributes to security privacy issues[12]. Therefore, robust data encryption algorithms, privacy authentication mechanisms and secure data relay routing are needed to protect the entire network from various attacks.

7. **Resource limitation:** Constrained by the limited physical space, the energy supply and memory storage are difficult to improve. Because of the low output transmission, the short communication range and limited coverage are inevitable. Therefore, cost efficiency is a crucial factor to consider as designers are always trying to spend less money on the best possible configurations.

8. **Adaptability:** WSN layout must be highly be adaptive and versatile to be applied in a wide range of application scenarios. Whether the number of nodes is just a handful or from hundreds to thousands, the entire network should be maintained operational. The overall topology of the network changes because of the mobility of SNs and observed events as well as the probability of malfunctioning SNs inside the network. This is why the development of WSNs must be intelligent enough to deal with these dynamic topology scenarios.

Energy consumption, delay and throughput are interdependent and tightly coupled parameters. There are number of strategies to achieve high energy efficiency and low latency. One strategy is clustering which provides balanced energy consumption and uses data compression techniques. Another strategy is to use trees whether broadcast or multicast. But it is not much beneficial in WSNs because the data should be delivered to single destination. Moreover, multicast trees provide benefits only in case, if cooperative diversity is used. The third strategy is the sleep mode strategy in which SNs are put into sleep mode to avoid nodes from receiving packets.
that are not intended for them. These types of strategies when combined with MAC protocols provide optimal energy efficiency, but it also incurred delay during transmission.

3. LOW ENERGY ADAPTIVE CLUSTERING HIERARCHY (LEACH)

This subsection gives an overview of the LEACH algorithm. LEACH is a self-organized, adaptive clustering algorithm that utilizes randomization to distribute energy across the network equally[13]. LEACH's operation is split into rounds. Each round comprises of a setup stage and a steady state phase. This choice is separate from the other nodes. The nodes that for the present round have chosen themselves as the CH, broadcast a signal. The CHs define a TDMAschedule for their cluster members (CMs) after the clusters have been formed. The CMs share the gathered information to their CH in the steady state stage. Data aggregation is accomplished by CHs to exclude unnecessary information and reduce the information in single packet and forward information directly to the BS. The next round will begin after a predetermined period. Fig.3 shows the topology of LEACH.

![Fig.3. LEACH protocol topology](image)

Mittal and Singh proposed DRESEP protocol to form the cluster taking the advantageous features of both MTE and LEACH[14]. It utilizes a reactive data-centric approach similar to TEEN. DRESEP protocol brings two enhancements in data transmission phase. First, the protocol uses a reactive approach by employing a threshold-based choice for communication between CMs and CH. Second, the protocol adopts dual-hop transmission strategy between CH and BS instead of direct transmission.

Mittal et al. developed a stability aware approach of DRESEP named SEECP[15]. By implementing this deterministic approach, it is expected to get rid of this issue of uneven distribution of SNs. This protocol certifies preset number of CHs that balances the load effectively in all the nodes. SEECP improves the network performance in terms of energy management and stability period when compared with LEACH and DRESEP.

3. 4. OPTIMIZED HIERARCHICAL BASED ROUTING PROTOCOLS

Traditional hierarchical routing's issue of optimization has led to smart clustering methods. Optimized clustering approaches are new that aim to be smart algorithms that enhance a sensor network's lifetime and make it energy effective. They are based on a latest growth of machine learning (ML)/ Computational Intelligence (CI) as a smart computational methodology to achieve solutions to NP problems[16]. Fuzzy Logic, Genetic Algorithm, Neural Network, meta-heuristic optimization algorithms can be classified as several ML / CI paradigms. The classification of optimized clustering algorithms for WSNs is presented in Fig.4.
4.1. Fuzzy Logic

It is a discipline of mathematics made-up to convey human reasoning. FL enables a measure of uncertainty that characterized by the use of linguistic factors called fuzzy set[17]. As illustrated in Fig. 5, the CH selection parameters like energy remained, node centrality, its distance to the BS and others are input variables for Fuzzy Inferencing System and CH node election is the only output parameter called fitness.

The Fuzzy Logic Cluster-head Election proposed in [18]. It integrates a CH choice with FL approach based on energy, node concentration, and centrality as nodes parameters. In each round, CH is chosen by the BS by identifying the opportunity that each node will become the CH, assisted by the fuzzy descriptors. The FL model
comprises of four steps: input descriptor fuzzification, rule assessment, rule aggregation, and defuzzification. FCH improves the FND better as compared to LEACH and raises the lifetime of the network compared to LEACH using fuzzy logic.

Sert et al. suggested MOFCA protocol using a fuzzy logic strategy to select CHs[19]. It is mainly intended for two main aspects: energy-efficient and lightweight. Based on the node competitive radius, MOFCA chooses the tentative CHs. It utilizes fuzzy logic to calculate the parameters mentioned above. Three fuzzy inputs are used to measure the competitive radius of tentative CHs. MOFCA's primary goal is to solve the hotspot issue resulting from multi-hop communication.

Tamandani et al. improved SEP, called SEPFL based on the fuzzy logic[20]. Three variables are used to improve selection of CH. LEACH-FL considers three attributes of nodes in the selection of the CH[21]. The BS uses if-then rules in fuzzy interface to select nodes. LEACH-FL's remaining operations are the same as LEACH. Moreover, sensor nodes are presumed to be equivalent in the terms of initial energy. Due to its consideration of the three attributes in CH selection, LEACH-FL outperforms LEACH.

4.2. Genetic Algorithm

For smart search and optimization, it is an adaptive heuristic method based on biological genetic evolution. GA models the natural evolution by conducting fitness trials to select the finest population on new structures. With GA methods, a population consists of a cluster of chromosomes and these are an answer to a specific issue, and fitness demonstrates the quality of a chromosome depending on specific needs.

Fig.6 presents an example of the GA model. Below are some clustering methods based on GA.

A genetic clustering algorithm (GCA) is proposed for increasing network lifetime with minimum consumption of energy. Fitness function consists of minimizing the number of CHs and minimum network distance. Jin et al. proposed GA to minimize the distance of transmission through clustering in WSNs. GA is used to reduce the entire minimum communication distance in the creation of a number of predefined clusters[22].

GA defines the network's optimum clusters. In GA, chromosomes represent the solution of the problem. Thus, for each chromosome, fitness function is evaluated. The fitness is evaluated by considering energy dissipated and delay occurred in clustering of nodes as factors for CH selection. GA has effectively reduced consumption of energy and improves network lifespan.

Hussain et al. proposed a GA-based clustering method called HCR protocol [23]. To evaluate fitness, parameters to be considered are: the residual energy of each SN, the distance between SN and BS, the distance between SN and CHs, the uniform spatial distribution of SN by calculating cluster distance, and the number of
transmissions at each stage. By combining all of the above-mentioned parameters, the fitness function is assessed. For HWSNs, Attea and Khalil suggested an evolutionary routing protocol (ERP)[24]. For enhanced cluster quality, the parameters to be considered are cohesion and separation error. ERP concentrated on minimizing the intra-cluster separation between SNs and their related CH; and maximizing the minimum CH-CH distance. Khalil and Attea proposed an energy-aware ERP (EAERP) for clustered HWSNs aimed at minimizing overall energy consumption through clustering[25]. Compared to LEACH[26], EAERP is more energy-efficient. Khalil and Attea suggested a stable aware ERP (SAERP) clustering algorithm based on GA for HWSNs. SAERP's goal is to increase the stability period instead of the lifetime of the network. It utilizes evolutionary modeling for CH choice where CHs with well-maintained energy consumption are chosen in a more effective manner. For the initialization of individual solutions, it utilizes energy-aware heuristics and energy-aware mutation to keep higher stable areas[27]. In terms of network stability span, SAERP outperforms LEACH and SEP.

4.3. Neural Network

It is a mathematical model which is inspired by neuronal biological networks. Each neuron is linked to many neurons. It is a network of neurons arranged in input-output layers where NN learns various routes and determines their interrelationships[28]. A straightforward NN model is shown in Fig.7 below. The NNs are used to solve sensor fusion, data mining, and clustering issues.

![Fig.7. A simple Neural Network model.](image)

Elby et al. proposed a new dynamic clustering algorithm based on the Artificial Neural Networks (ANNs) to reduce battery consumption[29]. The concept is to train the network in a reliable way to generate data fusion by diminishing unwanted information flooding. In this protocol, the SNs themselves build a group in such a manner that neighboring sensors form a cluster instead of sending information from each node to other in the next layer. One node nominated as CH within the cluster. SNs send data to the CH on a regular basis. CH performs the data aggregation process. The information can be sent directly to the BS or via other nodes in the cluster in the intermediate layers.

Shafiabadi et al. suggested a protocol for clustering based on the neural network of the self-organizing map (SOM)[30]. The routing-based clustering issue in WSNs was concentrated to diminish energy consumption and maintain network coverage. This protocol, called energy clustering SOM (ECSOM), performs clustering based on each node's energy level and spatial coordinates.

4.4. Reinforcement Learning

It is an ML sub-domain that teaches an agent what to do and how to allocate circumstances to specific behavior [31]. The agent attempts many activities and then has to choose from its experience to increase the
efficiency of the network. In some clustering algorithms, RL-based protocols are used to find optimal paths and to extend the network lifetime.

4.5. Meta-heuristic Optimization

Salehpour et al. presented to use ACO for designing an effective routing algorithm for cluster-based WSNs[32]. CMs transmit data directly to their CH in intra-cluster communication. CHs use ACO to discover a path to BS in inter-cluster communication. Since only CHs are involved in inter-cluster communication, this system more efficiently offers a smooth operation. An energy-sensitive clustering protocol called ACO for Clustering (ACO-C) minimizes and evenly distributes the costs of long-distance communications and data collection.

Wang et al. proposed an uneven clustering algorithm using ACO's dynamic adaptability to get the optimal route from CH to BS[33]. Clusters that are closer to BS have smaller sizes than those that are distant to BS. The closer CHs preserve extra energy for communication between clusters. Compared to LEACH, the algorithm significantly improves average energy consumption and extends the lifespan of the network.

Kuila et al. also presented a clustering algorithm to significantly extend network lifespan by balancing CHs’ lifetime. For CH selection, a clustering algorithm based on DE is used using an efficient vector encoding scheme[34]. The outcomes illustrate that improved clustering algorithm has better convergence compared to traditional clustering algorithms based on DE and GA.

4.6. Hybrid Algorithms

Shokouhifar and Jalali presented ASLRP protocol to extend the network’s operating life[35]. To achieve better performance, the tuning of controllable parameters is carried out using the GA-SA hybrid approach. The proposed protocol effectively balances node energy consumption, improves operating lifespan and receives more data packets at BS.

A novel energy efficient protocol named improved fuzzy unequal clustering routing (IFUC) is proposed with an aim to increase in energy efficiency and achieve load balancing[36]. It is used to determine one node’s chance to become CH and estimate the CH radius using ACO. The proposed clustering scheme shows better results than LEACH and EEUC.

Tomar et al. suggested a FIS based clustering protocol. The goal is to provide the probabilistic CH selection with an alternative solution. Fuzzy-logic and ACO are used in this strategy to select CH and determine an ideal route between node and BS[37].

Obaidy et al. proposed a smart hybrid algorithm based on GA-PSO to increase the lifespan of mobile WSNs[38]. PSO used to manage distance and making WSN self-organized. Final results showed that it is an effective method for minimizing distance.

4. Inferences

The above literature reveals the fact that CH election is probabilistic in most LEACH-like clustering protocols and thus a low-energy SN has possibility to become CH. A deterministic way for CH selection needs to be considered, such as DEC, which uses node residual energy as a parameter to elect CH. Instead of using only the concept of residual energy, other factors are also considered for balancing load between SNs. Furthermore, some CHs may also be located near each other. This shows that CHs in the network are not well distributed. Therefore, the basic need is to design an energy-efficient cluster-based routing algorithm.

Many clustering algorithms use proactive approaches for periodic transmissions. A number of WSN applications, however, require reactive approaches, e.g. military applications in which nodes only transfer their statistics when an intruder is detected. It is therefore of prime consideration to design clustering algorithms for event-driven networks (such as TEEN, APTEEN etc.).

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Ref.</th>
<th>Algo. used</th>
<th>Optimization Criteria</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSO-Clustering</td>
<td>Guru et al. (2005)</td>
<td>PSO</td>
<td>Min. intra-cluster distance</td>
<td>extended network lifetime</td>
</tr>
</tbody>
</table>
CHs are assumed to have a long communication range in most protocols that allows them to forward information directly to the BS. This statement is not always realistic since BS is often directly inaccessible to all nodes due to signal propagation problems. The most suitable solution to this problem is data transmission using multi-hop communication, which leads to hot spot problems in the network. Therefore, consideration of an alternative solution to handle this problem is required.

Scalability is a significant feature used to save network energy. The protocol that is used to save energy for small networks with few nodes may also be applicable for large networks, i.e. it should be adaptive to adapt to the larger network domain. There is another parameter named stability that cannot be ignored while designing an energy efficient routing protocol. Most protocols in the literature enhance the network's operating life by ignoring its effect on the robustness or stability period. There is a trade-off between the network's operating life and the network's stability period, which must be handled wisely.

Various parameters involved in communication protocols of WSNs are data transmission, routing type, network type, load balancing, energy efficiency, latency, stability, scalability, and communication cost.

5. Conclusion

Routing protocol design for WSNs is a challenging task that needs to take energy efficiency, robustness, and scalability into account. Proper formation of clusters and choice of CH nodes is very essential in WSNs. The major research in this direction focuses on how to minimize energy consumption during the extraction process of essential data in resource constraint systems like WSNs. The literature indicates that in many issues related to CH selection in WSNs, many evolutionary algorithms have outperformed deterministic methods. Proper choice of evolutionary algorithms together with suitable fitness function can effectively balance node energy consumption and maximize the lifetime of the network.
References


13. W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-Efficient Communication Protocol for Wireless Microsensor Networks P & SeQVUIRxG 5SC7STQVUWRXFe G 5S79 9 \$ \$ STQaXcbd G5S7A9 G5S7bhSeQVUIRx 7A9 \$ \$ B \$ pi Qbhr 6i Q i R & G @ 7 e S QaXI F G @ 7A9 \$ \$ \$ B \$ sSeQaX G @ 7 STQaX F 6,” 2000.


