Centralized Clustering Evolutionary Algorithms for Wireless Sensor Networks

Kamal S. Hamza
Department of Information Technology
Cairo University
Giza, Egypt
khamza@grad.fci-cu.edu.eg

Fathy A. Amer
Department of Information Technology
Cairo University
Giza, Egypt
f.aamer@fci-cu.edu.eg

ABSTRACT
Hierarchical routing is based on dividing the overall network structure into a set of smaller regions, each is managed via the so-called cluster head (CH). The cluster head is responsible for both inter- and intra-networking across all the sensors in the network. The clustering process greatly impacts the overall performance of the network (e.g., lifetime and energy consumption). As the size of the network increases, so does the complexity of the clustering process. Thus, evolutionary algorithms are typically used to cope with increasing complexity of the process. This paper surveys recent centralized techniques for clustering in both heterogeneous and homogenous WSNs.

CCS Concepts
• Networks → Network design and planning algorithms;

Keywords
WSN; Clustering Algorithms; Evolutionary Algorithm; Centralized Clustering

1. INTRODUCTION
Wireless Sensor Networks (WSN) have been proposed for more than two decades and have been deployed in various real-life applications. The emergence of the Internet of Things (IoT) has renewed a noticeable interest in WSN and other technologies that are expected to be deployed at large-scales in almost all aspects of our daily life [1] [2] [3] [4] [5].

Energy consumption has been always a concern in the design and deployment of such networks. The typical tradeoff between data communication and energy consumption for a challenging design problem in WSN. To this end, hierarchical routing in WSN is increasingly becoming the preferred solution for data networking in WSN due to its efficiency. In hierarchical routing the network is seen as a set of regions [6]; each will be managed through the so-called Cluster Head (CH). CH may handle both inter- and intra-cluster communication across the sensors in the network via single and multi-hop data transmission. It has been shown in the literature that such clustering can yield a better overall network performance in terms of energy consumption.

As a result, several research efforts have focused on developing and examining efficient clustering for sensor networks. The typical problem is defined by assuming a network with N sensor nodes. The nodes have a particular energy model that governs its transmission, reception, and computation of data. The overall objective function to optimize is typically related to the life-time (energy consumption) of the overall network.

Clustering formation can be generally classified into centralized and distributed techniques. In centralized clustering, the control message for cluster computation is received from a central base station (BS) based on information collected from all the sensors in the network. In the distributed approach; however, the decisions related to the formation of the clusters are made by the sensor nodes without the use of a central entity (or BS). In this paper, we focus only on centralized clustering techniques as they are practical for several practical scenarios in WSN.

As the number of sensor nodes grows; so does the complexity and cost of the clustering formation operations. Accordingly, more efficient techniques need to be devised to ensure better network performance and overall good clustering quality. Evolutionary algorithms are known to be more effective for solving large computational problems. Accordingly, there has been several studies in the literature that focused on performing clustering using various evolutionary techniques (e.g., [7], [8], [9]). The fundamental concept of performing clustering using evolutionary algorithms is to randomly generate a set of candidate clusters (solutions), and then compute the best solution based on some objective function. The selected solutions are then used as a seed to a next iteration of selection. Typically, the process continues until a computational limit is reached.

Surveys on clustering algorithms for WSN have been reported in the literature (See for example, [12], [13], [14], [15]). This paper; however, focuses on the centralized clustering algorithms using evolutionary algorithms.

The rest of the paper is organized as follows. Section 2 reviews existing centralized evolutionary clustering algorithms. Classification and comparison of centralized clustering algorithms is given in Section 3. Conclusions are given in Section 4.
2. CENTRALIZED EVOLUTIONARY CLUSTERING ALGORITHMS

Evolutionary algorithms typically share the same underlying concept of processing a population of potential solutions for a particular problem by applying the principle of survival of the fittest to produce good approximations to a solution. Several evolutionary techniques have proposed including the well-known Particle Swarm Optimization (PSO) [16], Genetic Algorithm (GA) [18], Harmony Search Algorithm (HSA) [7], and Bat algorithm (BA) [10]. Several other evolutionary algorithms have been also proposed; however, their use were reported in distributed clustering algorithms. In the following, we review the basic concepts of four main evolutionary algorithms used for centralized clustering in WSN. Figure 1 shows the four classes of the algorithms covered in this paper.

2.1 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is a stochastic optimization technique that has been used in various computation and networking problems. In the following, we review two PSO-based algorithms used for centralized clustering in WSNs.

2.1.1 Particle Swarm Optimization based Clustering (PSOBC)

Conventional PSO algorithms are used for continuous search spaces. The Particle Swarm Optimization based Clustering (PSOBC) algorithm [8] extends the conventional PSO by defining a new operator to make it applicable for the discrete search space. Each single solution in PSO will be considered as a Particle. Particles fly through the searching area of solution looking for the global optimum position that produces the best fitness of an objective function. Each particle keeps tracking of personal best position and the global best position of the whole swarm best [17]. The PSOBC is studied under three different scenarios where the base station (BS) location is changed in the network. The location of the BS is shown to impact the overall number of the alive nodes in the network; and hence, its lifetime. Simulation results showed that the PSOBC can greatly improve the network lifetime as compared to the LEACH and the LEACH-C algorithms.

2.1.2 Two-Tier Particle Swarm Optimization for Clustering and Routing Protocol (TPSO-CR)

The Two-Tier Particle Swarm Optimization for Clustering and Routing Protocols (TPSO-CR) [11], as the name suggests, make use of two PSO-based protocols. The first protocol is used for clustering by finding the optimal set of CHs in the network. Whereas, the second algorithm is used for routing by finding the optimal routing tree for inter-cluster communications. The simulation results show that the proposed protocol minimized the average number of non-clustered nodes in addition to improving the data rate in inter-cluster communication, increase network coverage and maintain energy consumption.

2.2 Genetic Centralize Dynamic Clustering (GCDC)

The Genetic Centralize Dynamic Clustering (GCDC) algorithm [9] was proposed to perform centralized clustering in sensor networks. GCDC is based on the well-known Genetic Algorithm (GA). GA is based on a natural selection process that mimics biological evolution. Generally speaking, a GA codifies a typical WSN as a chromosome, where a set of chromosomes is called a generation, and they are kept or dropped based on their fitness functions [18]. The GCDC attempts to find the optimal combinations of CH selection parameters related to the residual energy, distance of inter-clustering, and intra-clustering communication. The dynamic clustering changes the CHs over the time, equalizing the energy consumption across all nodes and, thus, extending the network lifetime.

The reported results for the GCDC shows improvement in the network coverage due to determining the optimal number of CHs and their position. This is a fundamental difference between the GCDC and the well-known LEACH algorithm, which considers only a fixed percentage of CHs without identifying their positions. That can lead to overlapped clusters. Shown simulation results demonstrate improvements in the percentage of the live nodes and the overall residual energy as compared to the conventional LEACH algorithm.

2.3 Harmony Search Algorithm (HSA)

Harmony search HS is an optimization method that is
inspired by some concepts related to the music field [7]. The key inspiring idea comes from the process of searching of musical instruments for the best harmony by polishing the pitches. The generic HS algorithm consists of five main phases: (1) problem formulation, (2) parameter setting, (3) harmony improvisation, (4) memory update, and (5) termination. In problem formulation, the objective function with an appropriate set of constraints is defined and expressed. In the parameter setting phase, similar to typical optimization methods, the values of the key HS algorithm parameters are defined. These key parameters include the following:

- **Harmony Memory Size (HMS)**: Represents the number of harmonies (solution vectors) in the harmony memory HM.

- **Harmony Memory Considering Rate (HMCR)**: In order to use the HM, a variable called harmony memory considering rate is used. The HMCR takes values that represents the probability at which the HS selects one of the harmonies in the HM.

- **Pitch Adjusting Rate (PAR)**: After selecting the harmony to be improvised, the pitch adjustment process is performed. The pitch adjusting rate (PAR), gives the probability by which the HS performs changes to the selected harmony.

- **Maximum Improvisation (MI)**: In the HS algorithm, improvisation is performed on a single harmony at a time. The number of improvisations to be made is predefined and stored in the maximum improvisation variable.

- **Fret Width (FW)**: Is a small random arbitrary length only for continuous variable, which was formerly called the bandwidth (BW).

The HS algorithm was intensively used for various routing problems in WSN. In centralized clustering, the main objective of deploying the HS algorithm is to minimize the intra-cluster distance and optimize the overall energy consumption of the network. The selection of the CHs according to their residual energy of nodes, attached in data sent by each node to the BS, will be repeated in each round of data exchange. Typically, the HSA is used to optimize the selection of CHs by finding optimal k CHs. The central BS sends information of CHs and identify which cluster to join to the rest of nodes. After the forming of the clusters, nodes send their data to BS through their CHs.

Based on the results reported in the literature, the optimal cluster distribution of the nodes resulted from the adoption of the HSA showed a remarkable improvement in the average energy consumption, especially in large-scale networks. This is because of the optimal distribution of clusters in the network. Moreover, HSA also avoids the shortcoming of premature convergence of GA method. In addition, the HSA overcomes the inability of the PSO to maintain the desired levels of population diversity and the balance between local and global searches.

### 2.4 Bat algorithm (BA)

The Bat Algorithm (BA) [10] is an emerging algorithm that focuses on minimizing the total communication distance and energy consumption based on the loudness parameter of bats taking. The algorithms takes into consideration that CHs that are close to the BS will typically deplete their energy and die faster than other nodes in the network. This is due to the heavy rely-rate occurring; a typical problem known as the Hot-Spot problem.

To deal with the Hot-Spot problem, network nodes are clustered into clusters with unequal sizes; known as the “unequal clustering method”. The work reported in [10] claims that the conventional PSO and Harmony Search algorithm (HSA) discussed above can be seen as special cases of the proposed Bat Algorithm, which attempts to combine the main advantages of the PSO, GA, and the HSA.

The typical process of the BA can be summarized in the following four steps.

- **Step 1**: Initialize the bat population, the pulse rates, the loudness, and define the pulse frequency.
- **Step 2**: Update the velocities to update the location of the bats, and decide whether detonate the random walk process.
- **Step 3**: Rank the bats according to their fitness value, find the current near best solution found so far, and then update the loudness and the emission rate.
- **Step 4**: Check the termination condition to decide whether go back to step 2 or end the process and output the result.

The study reported in [10] studied the BA under two cases. The first case proposes the BA-WSN algorithm that did not
Table 1: Comparison of Centralized Evolutionary Algorithms for clustering in WSN

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Optimization Approach</th>
<th>Clustering Objectives</th>
<th>Mobility</th>
<th>Optimization Objective</th>
<th>Location Awareness</th>
<th>Nodes Capability</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSOBC [8]</td>
<td>Particle Swarm Optimization (PSO)</td>
<td>Optimize intra cluster and inter cluster communication energy</td>
<td>No</td>
<td>Clustering</td>
<td>Yes</td>
<td>Homogeneous</td>
</tr>
<tr>
<td>GCDC [9]</td>
<td>Genetic Algorithm (GA)</td>
<td>Reduces the energy depletion rate Improve Network coverage</td>
<td>No</td>
<td>Clustering</td>
<td>Yes</td>
<td>Homogeneous</td>
</tr>
<tr>
<td>HSA [7]</td>
<td>Harmony Search Algorithm (HSA)</td>
<td>Minimize the intra-cluster distance Reduce power consumption Extend network life time</td>
<td>No</td>
<td>Clustering</td>
<td>Yes</td>
<td>Homogeneous</td>
</tr>
<tr>
<td>BA-WSN [10]</td>
<td>BAT Algorithm</td>
<td>Better performance (convergence/accuracy)</td>
<td>Yes</td>
<td>Clustering</td>
<td>Yes</td>
<td>Homogeneous</td>
</tr>
</tbody>
</table>

3. COMPARISON OF CENTRALIZED ALGORITHMS

Despite the fact that most Evolutionary Algorithms reported in the literature for clustering in WSN are based on similar and related concepts; however, the lack of a benchmark makes the quantitative evaluation and comparison of the various algorithms difficult. For the purpose of this paper, we compare the various algorithms based on the following two key properties:

- **Optimization Objective:** The limited resources of sensor nodes naturally make energy consumption; and hence, the network life-time, a prime objective for most optimization formulation in WSN. Various approaches can be used to optimize the energy consumption such as reducing the average energy consumption, increasing alive nodes over time and average fitness value. In addition to energy efficiency, few algorithms have considered additional optimization objectives such as: throughput, delay, coverage, robustness, link quality, data delivery ratio, and overhead.

- **Mobility:** The ability of all or some of the senor nodes in the network to move can be critical to the overall clustering process. In stationary networks, nodes are randomly deployed and cannot move from one location to another. When mobility is assumed, nodes can move from one position to another after deployment in order to find optimal clustering solution at the cost of an increased computation complexity.

- **Location awareness:** The ability of a node to recognize its location can be useful for clustering, especially, if this feature is combined with the mobility capability. In such a case, nodes can compute and change their location to improve the overall clustering solutions. Clearly; however, such enhanced capabilities comes with an increased cost for the overall network as most location awareness techniques assumes the use of high-cost and energy-hungry devices such as the GPS.

- **Node capabilities:** In some WSN, sensor nodes are not assumed to be of equal capability in terms of, for instance, energy, computation capability (data aggregation) or link communication (transmit power (range), and bandwidth. Such networks are said to be heterogenous. Heterogeneity adds new complexity to the overall clustering problem, but with the potential of improving the overall quality of the solution.

Table 1 summarizes the five algorithms for centralized clustering discussed in this paper in terms of the four aforementioned properties. The following main observations can be deduced from the shown table about centralized evolutionary clustering algorithms.

- Most algorithms assume the nodes are location aware. This can simplify and enhance the clustering process; however, this property may not be practical for large-scale real-life deployment of WSNs.

- Most algorithms assumes stationary nodes with fixed locations. This can negatively impact the optimization of the clusters, but it can simplify the optimization process.

- Most algorithms focus on the optimization of the clustering as an objective function without considering the routing optimization. This simplify the optimization problem, but can lead to impractical clustering solutions.

- Most algorithms assume homogenous sensor nodes with same and equal capabilities. This can greatly reduces the overall clustering complexity, but may not reflect...
the reality of emerging WSN deployments where heterogeneous nodes are needed to fulfill applications nodes.

4. CONCLUSIONS

This paper presents a brief survey on the use of Evolutionary algorithms for computing the clustering in Wireless Sensor Networks (WSNs). Four classes of evolutionary algorithms are considered; namely, the Particular Swarm Optimization (PSO), Genetic Algorithm (GA), Harmony Search Algorithm (HSA), and the Bat algorithm (BA). Five algorithms under these four classes are presented and their key operational concepts are explained. The main properties of the five algorithms are summarized and analyzed. It is clear that the use of evolutionary algorithms in centralized clustering is limited and requires further research. In particular, algorithms that deal with mobility and heterogeneity need to be developed to deal with the real-life scenarios of deploying large-scale WSNs. Moreover, while the assumption made by most existing algorithms about the awareness of nodes about their location can be very useful for finding optimal clustering solution; however, this assumption may not be practical for large-scale WSNs due to the high cost and energy consumption incurred by location finding devices.

5. REFERENCES