UAV’s enhanced data collection for heterogeneous wireless sensor networks

Kamel BARKA  
Computer science department  
LaSTIC Laboratory  
Batna 2 University  
Batna, Algeria  
ORCID : 0000-0002-6824-1594

Lyamine GUEZOUILI  
Renewable Energies and New Technologies Department  
LEREESEI Laboratory  
Higher National School of Renewable Energies, Environment & Sustainable Development  
Batna, Algeria  
ORCID : 0000-0002-7406-7633

Assem REZKI  
Computer science department  
LaSTIC Laboratory  
Batna 2 University  
Batna, Algeria  
Email : a.rezki@univ-batna2.dz

Abstract— In this article, we propose a protocol called DataGA-DRF (a protocol for Data collection using a Genetic Algorithm through Dynamic Reference Points) that collects data from Heterogeneous wireless sensor networks. This protocol is based on DGA (Destination selection according to Genetic Algorithm) to control the movement of the UAV (Unmanned aerial vehicle) between dynamic reference points that virtually represent the sensor node deployment. The dynamics of these points ensure an even distribution of energy consumption among the sensors and also improve network performance. To determine the best points, DataGA-DRF uses a classification algorithm such as K-Means.

Keywords—Heterogeneous wireless networks, Unmanned aerial vehicles, Reference point, Collect data, Genetic algorithm.

I. INTRODUCTION

A wireless sensor network is considered heterogeneous if all its sensors have different storage, processing, battery, sensing and communication capabilities.

Heterogeneous sensor networks typically consist of many low-cost sensor nodes for data collection and a smaller number of nodes with more energy for filtering, fusion, and data transport. This type of network design can improve the lifetime and reliability of the system. They are commonly used in real-world applications, as seen in references [1] and [2].

Many energy-efficient protocols for heterogeneous networks have been developed, and most of them rely on the clustering technique, which is effective in terms of scalability and energy conservation for wireless sensor networks.

In [3], the authors propose a protocol that collects data from fixed sensor nodes via mobile UAVs, and this protocol is based on a genetic algorithm, called DGA (Destination selection according to Genetic Algorithm) to control the movement of UAVs between reference points (RPs) where it allows each UAV to determine one point among the reference points as its new destination. These reference points virtually represent the monitored area. The base station determines these points based solely on the communication range value of the sensors and UAVs, without considering the deployment of sensor nodes in the area, this may determine points in a sub-area that does not contain any sensor nodes, resulting in a loss of time when a UAV moves to these points. Related to this is the possibility of locating a single point in a sub-area where sensor nodes are deployed with high density, resulting in a significant loss of sensed data.

Furthermore, during the lifetime of the network, the reference points remain the same, this leads to unequal energy consumption at the sensor nodes, so that sensor nodes that are deployed close to a reference point, consume a lot of energy compared to other sensor nodes, over time the risk of network fragmentation increases.

For this reason, in this article we propose a protocol that collects data from heterogeneous networks according to the following principles:

- The movements of the UAVs between the reference points are oriented via the DGA algorithm
- The reference points are determined based on the deployment of the sensor nodes in the area, the communication range of the sensor node and the UAV.
- Over time, the reference points are redetermined in a smooth and accurate manner (the points are dynamic)

The article is organized as following: The second part presents some related work. In the third part, we present a description of our DataGA-DRF protocol. The simulation results are in part 4. Finally, we present the conclusion of the work with some perspectives.

II. RELATED WORK

A Wireless Sensor Network (WSN) is a collection of small, low-power, autonomous devices called sensor nodes, which communicate with each other wirelessly to cooperatively monitor and collect data from the physical environment. Each sensor node typically consists of a sensing unit, a processing unit, a radio transceiver, and a power source. The nodes in a WSN can be deployed in a heterogeneous environment to measure various parameters, such as temperature, humidity, pressure, vibration, sound, light, and so on. The collected data can be processed and analyzed to provide valuable information for applications such as environmental monitoring, surveillance, healthcare, industrial control, and smart cities. UAVs can enhance data collection in WSNs by providing mobility, flexibility, and extended coverage.

The following related work focuses on clustering in WSN within the domain of AI-based techniques. In the artificial intelligence (AI) field, there are several approaches that may be employed in clustering, like Particle Swarm Optimization (PSO), Neural Networks (NN), Genetic Algorithms (GA), Ant Colony Optimization (ACO) [4], etc.

Mann et al. [5] proposed a Computational Intelligence (CI) based metaheuristic called BeeSwarm for energy-efficient hierarchical routing. The protocol comprises of three phases:
BeeCluster, BeeSearch, and BeeCarrier. The main advantage of the proposed protocol is that it integrates these three phases for node clustering, data routing and transmission. The performance of the BeeSwarm approach is evaluated against two existing routing protocols, MRP and ERP [6] in terms of several performance parameters. The results indicate that BeeSwarm slightly outperforms the other routing protocols in terms of total number of packets delivered and packet delivery ratio. Additionally, BeeSwarm consumes less energy, thus extending the network lifetime.

The authors in [7] proposed an efficient and environmentally friendly dynamic clustering mechanism that utilizes machine learning to guide the selection of cluster heads based on power demand and information volume. The goal of the dynamic CH scheme is to extend the lifespan of resource-constrained networks. In the proposed scheme, the CHs are selected dynamically to transmit data from the source to the destination devices.

In [8], Sandeep et al. propose a clustering method for wireless sensor networks that utilizes the social behaviors of Rhesus Macaque monkeys, claiming that the technique offers an energy-efficient solution for routing.

In [9], the authors present an enhanced version of Ant Colony AI, which incorporates a new random disturbance factor for use with clusters of N sensor nodes and M cluster heads in WSNs. The positions of the nodes and cluster heads are fixed after distribution, and the clustering is carried out over a number of iterations to reach the optimal solution. The cluster head member nodes are updated using an improved ant colony artificial intelligence technique, which improves the performance of WSN clustering in terms of fitness function convergence and energy consumption.

In [10], an adaptive fuzzy clustering protocol called LEACH-SF was proposed. This protocol uses the fuzzy c-means clustering algorithm to globally cluster all nodes and create a balanced cluster, and then uses the Sugeno fuzzy inference system to identify appropriate CHs. The goal of the LEACH-SF is to prolong the lifetime of the WSN.

The MSO-Tabu algorithm proposed in [11] is based on PSO with Tabu search (TS) as the goal to enhance the WSN CH selection. The suggested approach chooses effective CHs that optimize the route and lengthen the life of the network.

The initial step is to establish the initial locations and energies of the nodes and the base station. The formation of clusters is then determined based on the proximity of the nodes to the base station and the energy level of the nodes. According to the author's results achieved in [11], the suggested MSO-Tabu technique is energy, cluster creation, and lifespan efficient since it reduces average packet loss and end-to-end latency.

In [12], the protocol GAOC (Genetic Algorithm-based Optimized Clustering) is proposed with the aim of optimizing the CH. GAOC increased network lifetime, dead time, network throughput, and remaining energy.

The solution we will propose next is based on one of the artificial intelligence algorithms, namely genetic algorithms.

### III. DATA GA-DRF: PROTOCOL DESCRIPTION

Here we propose a protocol for data collection in a heterogeneous network, called (DataGA-DRF, a Protocol for data collection based on a Genetic Algorithm applied to Dynamic points). DataGA-DRF is an improvement of the DGA protocol. DataGA-DRF uses the DGA algorithm to select a new destination for a drone. Drone destinations are dynamic reference points, this implies that these points change throughout time. In addition, these reference points only represent the sub-areas in which the sensor nodes are deployed and not the entire monitored area. This protocol consists of the following two steps.

In the following, we will use classification to classify the sensor nodes according to their positions, which allows us to define the reference points that represent the area. However, for clustering, we will create clusters to allow the sensor nodes to route data to the UAVs.

#### A. discovery step

The objective of this step is to transport the positions (e.g. coordinates \((x, y, z)\), it is supposed that the sensors are equipped with GPS) of all sensor nodes to the base station.

Where DataGA-DRF uses the mechanisms used in [DGA] as follows:

1. Initially, the base station creates the set \((Z)\) of reference points based only on the communication range of the sensor and the drone, such as:

\[
Z = \left\{ \left( P_{ij} \left( \frac{\sqrt{2(r^2 - r^2)}}{2}, \frac{j\sqrt{2(r^2 - r^2)}}{2}, 2 \right) \right), \right. \quad \left. \begin{array}{l}
i, j \in Z \\
r \leq z \leq r \end{array} \right\}
\]

With \(r\) and \(R\) is the communication range of the sensor and the drone respectively.

2. Each drone selects a point in the \(Z\)-set as a new destination by applying the DGA algorithm, and then moves towards it.

3. When a drone arrives at its destination the first time, it creates a cluster with the sensors and then collects the data of each member sensor in this cluster. The creation of the cluster is performed as in [3].

Finally, this step ends when the base station receives the coordinates of all sensor nodes deployed in the area, which means that the drones have passed over all reference points (including the base station).

#### B. Improvement step

The objective of this step is to optimize the set of reference points \((Z)\) to represent only the sub-area where the drones are deployed, and to update the \((x, y)\) coordinates of these points over time. Where DataGA-DRF uses the following mechanisms:

1. After receiving the coordinates of the sensors, the base station classifies these sensors into a group of classes \((C)\) by applying a classification algorithm.

\[
C = \{c_i; i \in \mathbb{N} \text{ and } c_i \text{ is a class}\}
\]

2. The base station creates the set \(Z^2\) composed of reference points \(P_i (x_i, y_i, z_i)\), where the coordinate values \(x_i\) and \(y_i\) are equal to the coordinate values of a single sensor node belonging to the class \(c_i \in C\), and the coordinate value \(z_i\) is smaller than the value of the sensor's
communication range \((r)\). Thus, each class of sensor node is represented by a reference point in \(\mathbb{Z}_2\) (i.e., \(|\mathbb{Z}_2| = |C|\)).

\[
\left\{ \begin{array}{l}
Z_2 = \{ P_i(x_i, y_i, z_i), 1 \leq i \leq |C| \} \\
x_i and y_i \in c_i \\
r / 2 \leq z \leq r 
\end{array} \right.
\]

(3)

3. Each drone selects a point in the \(Z\)-set as a new destination by applying the DGA algorithm, and then moves towards it.

4. When a drone arrives at its destination, it creates a cluster with the sensors and then collects the data of each member sensor in this cluster. The creation of the cluster is done as in [3].

5. Over time, the base station updates the \(Z_2\) set by choosing new \(x_i\) and \(y_i\) values corresponding to the coordinate values of another sensor of the same class. In this case we can take the passage of the UAVs twice over the same reference point as an update criterion.

The steps of this protocol are explained in the two organigrams in figures 1 and 2.

IV. SIMULATION RESULTS

This section presents simulation results to evaluate the performance of DataGA-DRF using the NS2 simulator.

The HWSN simulation includes 100 stationary sensors randomly placed in a 300x300 meter event space. A central base station is also located in the area, along with a set of UAVs. The range of communication for the sensors, base station, and UAVs is 30, 70, and 70 meters respectively. The sensor nodes collect data and send it to UAV nodes, which create clusters through hop-by-hop communication. The UAVs move at a constant velocity of 10m/s (The effect of speed on performance has already been discussed in [3], 10m/s is the average) using the DGA mobility model and pause for 2 minutes. The key parameters of the network are summarized in Table 1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topology</td>
<td>random (300x300)</td>
</tr>
<tr>
<td>Base station (BS)</td>
<td>One in center area</td>
</tr>
<tr>
<td>Static Sensor Nodes</td>
<td>144</td>
</tr>
<tr>
<td>UAV Nodes</td>
<td>1 and more</td>
</tr>
<tr>
<td>classification algorithm</td>
<td>K-Means</td>
</tr>
<tr>
<td>Transmission range of the sensor</td>
<td>30m</td>
</tr>
<tr>
<td>Transmission range of the UAV</td>
<td>70m</td>
</tr>
<tr>
<td>Transmission range of the BS</td>
<td>70m</td>
</tr>
<tr>
<td>Velocity of the UAV</td>
<td>10m/s</td>
</tr>
<tr>
<td>Pause time ((T_p))</td>
<td>2 min</td>
</tr>
<tr>
<td>Switchover time at the BS ((T))</td>
<td>6 min</td>
</tr>
<tr>
<td>Crossing Probability (P_c) for DGA</td>
<td>1</td>
</tr>
<tr>
<td>Mutation Probability (P_m) for DGA</td>
<td>0.05</td>
</tr>
</tbody>
</table>

TABLE 1. SIMULATION PARAMETERS

![Fig. 1. DataGA-DRF steps for UAV](image1)

![Fig. 2. DataGA-DRF steps for Base Station (BS)](image2)
Figure 3 a and b shows the creation of the reference points set process by the base station in a network consisting of 45 sensors and one UAV. We notice that in the discovery step (see Figure 3 a) there are 18 reference points because the BS only relies on the communication ranges of the sensor and the UAV in order to calculate the coordinates of these points. But in the improvement step, the number of points becomes only 8 points (see Figure 3.b) because the BS has classified the sensors in 8 classes by applying the K-Means classification algorithm, which allows to reduce the number of UAV movements between the points.

A. Delivery Ratio Evaluation

The purpose of the Delivery Ratio Evaluation is to examine the impact of increasing the number of UAVs deployed on the delivery ratio, which is the proportion of events that reach the BS out of the total number of events generated within a given period.

Figure 4 illustrates the average delivery ratio as the number of UAVs in the network increases. We observe that the introduction of many UAVs into the environment clearly allows for the optimization of data receipt. Nonetheless, what is noteworthy is that even a small number of UAVs is able to achieve a 100% reception rate.

We notice that even with dynamic reference points, we can have the same delivery results as with classical reference points.

B. Latency evaluation in cluster

Here, we analyze the latency as the number of UAV nodes deployed in the network is increased. Latency is defined as the average amount of time it takes for a message from a source static node to reach a UAV node.

One of the comparison points between DataGA-DRF and DGA is the size of a cluster. Indeed, a cluster with DGA is smaller than a cluster with DataGA-DRF, since with DataGA-DRF the approach is interested in creating clusters in places where sensors are grouped, this explains the large size of a cluster in DataGA-DRF. This explains the increased latency in DataGA-DRF within the cluster.

C. Latency evaluation in network

In this section, we are examining the latency when there are 4 UAV nodes deployed in the network. Why exactly with 4 UAVs? because in Figure 4 we have reached an average delivery ratio equal to 100% from the injection of the 4th UAV. We define the latency as the average time required for a message from a source node (static sensor) to the base station.

We notice in Figure 6 that in both cases, before the 8th minute the latency values are the same in both cases, which means that the DataGA-DRF protocol works like DGA, so DataGA-DRF is in the discovery step. However, after the 8th minute, in the case of DGA, the latency value remains fixed, but in the case of DataGA-DRF, the latency value is minimized, which means that the DataGA-DRF protocol is in the improvement step. Note also that before the 4th minute,
the latency value is zero which means that no message arrives at the base station.

![Graph](image-url)

**Fig. 6.** The latency in network for DataGA-DRF and DGA

V. CONCLUSION

Our study, which utilizes Genetic algorithms and Unmanned Aerial Vehicles (UAVs) for clustering, shows the effectiveness of using Genetic algorithms for collecting data through dynamic reference points. Simulation results indicate that our solution DataGA-DRF is supreme compared to DGA. Substantial improvements in latency have been demonstrated in a series of experiments. Eventually, Such a cross-study can open doors for intelligent data collection in heterogeneous WSNs.

As a perspective, we would like to replace the ranking algorithms with deep learning-based methods to determine the reference points.

REFERENCES


