Increase of the Lifetime of Wireless Sensor Network using Clustering Algorithm and Optimal Path Selection Method

Wadii JLASSI [1], Rim HADDAD [2], Ridha BOUALLEGUE [3], Raed SHUBAIR [4]

[1] Innov'Com Laboratory, National Engineering School of Tunis, University Tunis El Manar, Tunis, Tunisia
[2] Laval University, Quebec, Canada
[3] Innov'Com Laboratory, High School of Communication of Tunis University of Carthage, Tunis, Tunisia
[4] New York University Abu Dhabi, UAE

wadii.jlassi@gmail.com, rim.haddad@dti.ulaval.ca, ridha.bouallegue@gmail.com, raed.shubair@nyu.edu

Submitted January 2, 2022 / Accepted May 19, 2022 / Online first July 12, 2022

Abstract. *By the recent improvement of the internet of things (IoT), the need to implement wireless networks is increasing. It is a challenge to balance between battery lifetime of the different sensors and network lifetime. Many studies proved the importance of using clustering and Mobile Data Collectors (MDCs) to extend the operating time of sensor nodes. A mobile data collector is used to gather the data recorded by the nodes over a short transmission range. The proposed approach aims to decrease the energy consumption of each sensor node by using the Genetic Algorithm (GA) and mobile data collector. So, we suggest a clustering algorithm to find suitable Cluster Heads and form clusters. Then, we employ the genetic algorithm to construct an optimal data gathering path for MDC. Computer simulation proves that the proposed approach outperforms existing ones.*

Keywords

Wireless sensor network, genetic algorithm, hierarchical agglomerative clustering algorithm, clustering, mobile data collector

1. Introduction

The development of Wireless Sensor Networks (WSNs) plays several roles in numerous aspects of human life, such as environmental conservation, domestic applications, military surveillance, air traffic control, medical care and so on [1], [2]. The WSN architecture consists of a large number of sensor nodes. These are randomly deployed in an environment with limited power, processing and computing resources [2], [3].

Sensor nodes are dispersed in a hostile environment with limited power, processing and compute resources [3]. Sensor nodes aim to capture, process and forward information from the source nodes to their requested destination, also known as sink nodes [4]. WSN parameters such as acquisition data, number of nodes, the lifetime of sensors, energy consumption, geographic location of the sensor, environment and context are critical in WSN design [5].

Clustering [6] is an efficient method to overcome unbalancing energy consumption between sensor nodes and to extend the network lifetime. In clustering, the network is partitioned into numerous clusters. Each cluster has one node, which is called cluster head (CH). This node receives packets from other cluster members, then sends these sensor data to the base station (BS) directly or multi-stage through other cluster heads. The direct transmission results in high energy dissipation, which causes the premature death of sensor nodes. The death of a node can cause a coverage hole that leads to lose influential data from the area. In multi-hop transmission, data are transmitted by one sensor node to the base station through one or more intermediate sensor nodes to reach the base station. Since the sensing environment is wide and the base station is far from the sensor area, this re-transmission process leads to extreme node depletion. According to the existing research, there are two approaches for the selection of CHs in WSN. In the self-organization technique [7], sensor nodes make their own decision and set up their status to become a normal node or a CH node, according to the weight, the residual energy and the position of sensors in the network. In this case, sensor nodes have to communicate with each other to avert the election of more than one CH in each cluster and minimize the number of packets to transmit. In the second approach, Base Station performs the selection of CHs and cluster formation. Every round, each node in the network transmits its residual energy and its location information to BS. This last evaluated the chances of each sensor node and selects the best one to be CH. Then it informs nodes about their status and the cluster [8], [9].

Several researchers have demonstrated the efficiency of MDC for data gathering and energy conservation in WSNs [9]. Starting from the BS and visiting the area of interest, MDC collects data from polling points and forwards it to the BS. This mission has benefits such as minimizing data collection time, conserving energy and extending network lifetime.

DOI: 10.13164/re.2022.0301

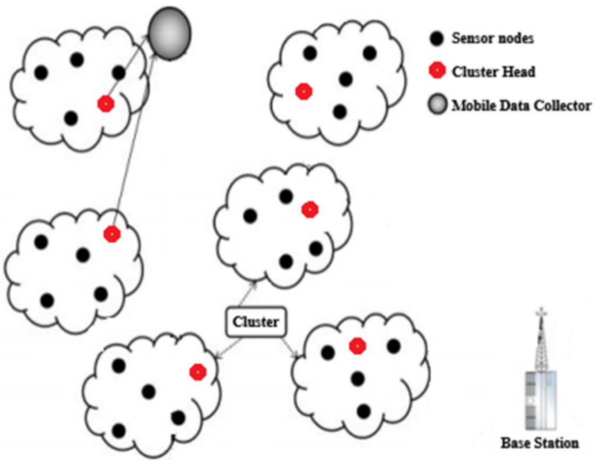

Fig. 1. The architecture of WSN with MDC.

The MDC is a mobile device that contains a rechargeable battery, GPS devices, long-range transceivers and large memory unit that traverses the network to gather data [9]. Cluster heads save the received information and transmit it to the MDC when this device draws near its area. Thus, the energy consumption of each node or cluster head effectively decreases. Periodically MDC returns to BS to upload the gathered data and prepare for the upcoming travel [10], [11].

Our contribution aims to propose a clustering algorithm and optimal data collection path of MDC to conserve energy and overcome the imbalance in energy consumption of the network. It could be brief as follows:

- We propose a clustering algorithm to find the suitable CHs and form clusters, based on the residual energy and the distance between sensor nodes.
- We employ the genetic algorithm to construct an optimal data collection path for MDC.

The paper is organized as follows. In Sec. 2, we discuss related works. In Sec. 3, energy model and problem formulation are presented. The proposed approach is presented in Sec. 4. Simulation results and comparison to other works are given in Sec. 5. Finally, Section 6 concludes the work and presents further works.

2. Related Works

Various data gathering approaches are introduced for energy optimization to improve the operation time of the entire network and decrease power consumption using MDC [12], [13]. The mobility aims to reduce the energy dissipation over the data transmission of sensor nodes that helps to enhance the network lifetime [14], [15]. We distinguish two categories of mobility: random and controlled [16].

In some works, the controlled mobility of MDC is calculated [17, 18, 19, 20]. In [18], the authors estimate the MDC's path in WSN based on the priority ordered dependent nonparametric trees. The dominant node is a node that can be selected as a CH and is computed based on the packet collision index.

In [17], the authors present a matching game-based data collection algorithm with MDCs, which chooses specific CHs to visit based on the distances. The network is partitioned into numerous clusters for multiple MDCs. The trajectory is estimated using mathematical equations based on the position of CHs. The advantage of this approach is reducing the data gathering delay. However, the number of MDCs is high. Every MDC visits a CH to collect data.

In [20], the authors propose a learning automata approach. In this approach, the network is partitioned into four logical partitions. It conducts the MDC towards each logical partition in a constant period. The trajectory is computed based on MDC's learning automata. On the other hand, random mobility presents several benefits such as simplicity and facility of implementation [21]. However, data collecting latency extended and the MDC's capacity is depressed.

In [22], the authors propose an optimal data gathering method for MDC. They employ the K-means algorithm to form clusters and find optimal CHs. Euler cycle and Hamilton cycle are proposed to discover the MDC's data collection path [24]. However, this approach does not reduce the network lifetime efficiently and maximizes the data gathering trajectory. The MDC will travel a long path to gather data from an important number of elected cluster heads [26].

In [23], the authors present an Energy Aware Path Construction algorithm (EAPC). This approach estimates the performance of each node using two parameters: the distance and the neighbor's weight. It contains the following stages. Firstly, a minimum spanning tree is found and rooted at the BS. Secondly, the EAPC algorithm selects the data collection points (CPs) and constructs the optimal path. Finally, the MDC travels the path and gathers the data from high burdened CPs. However, the CPs exhibit no uniform energy depletion, which minimizes the network lifetime.

In [25], the authors present a residual Energy Aware Mobile Data Collection Scheme in Wireless Sensor Networks (REAMDC). This approach aims to combine multihop sensor nodes and mobile data collector. Firstly, spectral clustering is used [16] then, the suitable node is selected to be the center of each cluster in the network according to two factors: the number and the residual energy of its neighbors [25]. Secondly, the authors select a greedy strategy to discover which cluster should be subdivided to divide the large cluster into tiny clusters. Finally, a data relay tree is constructed based on the residual energy to balance the energy consumption. In every round, the trajectory of MDC is changed. Therefore, the network lifetime can be enhanced. However, the clustering algorithm cannot be employed in large-scale networks. Also, one MDC is not sufficient to prolong the network lifetime efficiently.

In [27], the authors adopt an efficient routing protocol for data transmission in WSNs based on Fog Computing.

This approach aims to forward data from normal sensor nodes to the data center through fog nodes. Also, it proposes an Ant Colony algorithm [26] to find the optimal path to transmit data from sensor nodes. However, in the case of the high number of nodes, the approach was not practical due to the significant computational complexity. Also, the sensor nodes display no uniform energy depletion, which influences the lifetime of the WSN.

The above algorithms confirm the amelioration of energy imbalance and extend the network lifetime. However, most of them do not contemplate the trajectory cost between data gathering points or utilize one MDC to collect data from large-scale networks. Also, these algorithms do not reduce the data gathering time efficiently.

In this study, we propose an energy-efficient data gathering approach for Mobile Data collectors in a wireless sensor network. This approach aims to manage network energy efficiency and constructs the optimal travel path for Mobile Data Collector. Also, we focus on selecting the suitable CHs considering several parameters, which are the residual energy and Euclidean distance of sensor nodes, to ensure the data gathering from the network.

3. Energy Model and Problem Formulation

3.1 Energy Model

The quantity of energy in sensor nodes is almost entirely consumed while transmitting, receiving packets and processing data. The energy model used is the same one used in [28]. As defined in this model, there are two possibilities: the free and the multi-path (MP) space. The energy consumption for transmitting and receiving k-bit packets is given respectively by (1), (2). The distance d represents the distance between the sending and the receiving node:

$$E_{\mathrm{Tx}}(k,d)=\begin{cases}E_{\mathrm{elec}}\cdot k+\mathcal{E}_{\mathrm{fs}}\cdot k\cdot d^{2}, & d<d_{0},\\ E_{\mathrm{elec}}\cdot k+\mathcal{E}_{\mathrm{amp}}\cdot k\cdot d^{4}, & d\geq d_{0},\end{cases} \tag{1}$$

$$E_{\mathrm{RX}}(k)=E_{\mathrm{elec}}\cdot k \tag{2}$$

where E_{elec} is the energy necessary for sending 1-bit and E_{RX} presents the energy usage to receive data. $\varepsilon_{\mathrm{fs}}$ and $\varepsilon_{\mathrm{amp}}$ are the amplified power in the radio model. d_0 is the threshold obtained using (3):

$$d_{0}=\sqrt{\frac{\mathcal{E}_{\mathrm{fs}}}{\mathcal{E}_{\mathrm{amp}}}}\,. \tag{3}$$

In one round, the CH consumes more energy than normal sensor nodes due to receiving data of member nodes and transmitting it to MDC. The CH's energy E_{CH} can be estimated as (4):

$$E_{\mathrm{CH}}=E_{\mathrm{NctoCH}}+E_{\mathrm{CHtoMDC}}, \tag{4}$$

$$E_{\mathrm{NctoCH}}=n\cdot E_{\mathrm{Rx}}(k)\,, \tag{5}$$

$$E_{\mathrm{CHtoMDC}}=m\cdot\left(E_{\mathrm{Rx}}(k)+\mathcal{E}_{\mathrm{amp}}\cdot k\cdot d_{\mathrm{CHtoMDC}}^{2}\right) \tag{6}$$

where the number in one cluster is n. E_{NctoCH} is the energy used by a CH to receive data of member nodes in the cluster and E_{CHtoMDC} is the energy used by the CH to send data of member nodes to MDC.

3.2 Problem Formulation

In this study, the WSN includes n nodes in the $L\times L$ area. We suggest the HAC algorithm to elect the suitable CHs and form clusters. Then, we try to construct an optimal path rooted at BS to transmit the aggregated data via MDC.

After the election of CHs, the BS applies the GA on all the selected CHs to find the optimal MDC travel path. When the MDC finishes collecting data from the CHs, these CHs return to sleep mode to avoid power consumption.

The MDC starts the tour from BS to appropriate CHs to gather information in a single hop at sufficient sojourn time. Periodically the MDC returns to BS and uploads collected data [11]. It moves along a predefined trajectory with a constant speed v. The MDC could have different planning trajectories: the static [29] and the dynamic path [6]. We adopt the static path. During each round, the HAC algorithm is employed for choosing suitable CHs. The genetic algorithm is utilized to obtain the optimal path for each MDC.

4. The Proposed Algorithm

In this section, a Clustering Algorithm with Mobile Data Collector is detailed. In each round, the principal phases contribute as follows:

- Cluster Head election and cluster forming phase,
- Dividing network environment into equal areas,
- Optimal path construction phase.

The first objective is to manage energy efficiency and optimize the network lifetime. To reach this goal, this paper uses clustering and mobile data collectors. We try to construct an optimal path rooted at BS to transmit the aggregated data via MDC. In each round, we find a new subset of CHs and a new traveling MDC's path is established.

The principal stages of our work are detailed in Fig. 2.

4.1 Decision of Optimal Visiting Points and Clusters Forming Phase

The most critical challenge in WSN is to balance the consumed energy by each sensor node in the network. The proposed scheme uses a HAC algorithm (Algorithm 1) which divides the network into clusters based on Euclidean distances between nodes and residual energy and aims to find the optimal CHs for each cluster. The proposed scheme is re-executed at each packet delivering round to maintain energy efficiency which means we obtain a new subset of CHs and a new traveling path is established.

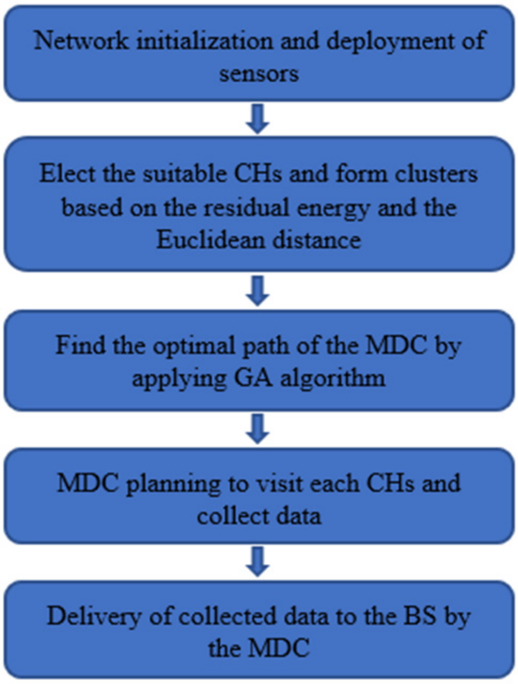

Fig. 2. Flowchart of the proposed approach.

```
Inputs: S= {S1, S2, …, Sn} – set of nodes
Outputs: C = {C1, C2, …, Ck} – clusters with CHs.
                  m: S→ C – set of nodes for each cluster.
Iteration:
 function HAC {
     set C to initial value (k≤ S) – the center is the
       nearest node to the BS.
       foreach Sn in S {
         m(Sn) = minDistance (Si, Ck) – minimum
         distance between node and cluster center.
       }
       while m has modified {
          for each j in {1...k} {
             recalculate the cluster center
          }
          for each Si {
             m(Sn) = minDistance (Si, Ck)
          }
       }
 return C
}
```

Algorithm1: CHs election and cluster formation.

The algorithm accepts two inputs: $\mathbf{N} = \{n_1, n_2, n_3,..., n_n\}$ static nodes with location information and k clusters with k less than n. The output is $\mathbf{C} = \{C_1, C_2, C_3,..., C_k\}$ set of clusters. It executes to group the sensor nodes into k clusters and estimates the center of each cluster. The distance between sensor nodes is calculated using the Euclidean distance formula.

The proposed algorithm is presented as follows:

Initial Clustering: For the first iteration, the network is divided into a predefined number of clusters with k is out of n nodes. In this step, the nearest node to the BS is selected as CH based on the minimum distance estimated $\min(d_i)$ (Fig. 3). d_i presents the distance between node i and the base station.

In two-dimensional space, the distance d_i is calculated as follows (7):

$$d_i = \sqrt{(x_i - x_{\mathrm{bs}})^2 + (y_i - y_{\mathrm{bs}})^2}. \tag{7}$$

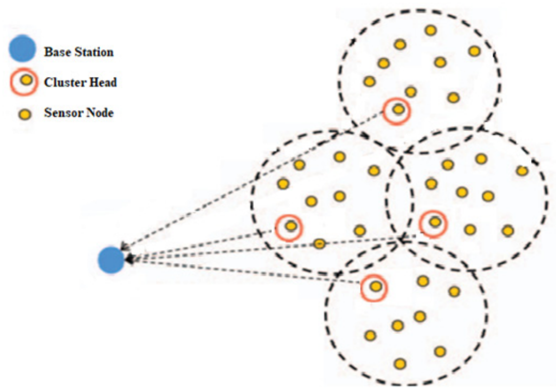

Fig. 3. Initial clustering.

Reclustering: After the association of all sensor nodes to the k clusters, the center point is calculated by the following formula center $(X_{\text{center}}, Y_{\text{center}})$ (8), (9) with:

$$X_{\text{center}} = \frac{1}{\sum_{i=1}^{s} E_i} \cdot \sum_{j=1}^{s} E_j x_j, \tag{8}$$

$$Y_{\text{center}} = \frac{1}{\sum_{i=1}^{s} E_i} \cdot \sum_{j=1}^{s} E_j x_j. \tag{9}$$

The center point of a cluster is a virtual node located at the center of the cluster in the first iteration because all sensor nodes have the same amount of energy initially.

The progress of the algorithm, the center point converges to the node with the highest residual energy. So, the CH is still the nearest node to the center point with the highest residual energy. That prolongs the life of each sensor node and facilitates intra-cluster communication between nodes and CH (Fig. 4).

Choosing the CH: Finally, the nearest node of the center point (x_{center}; y_{center}) is selected as the new CH. For that, we calculate the Euclidean distance between the nodes of each cluster and the novel center point using (10):

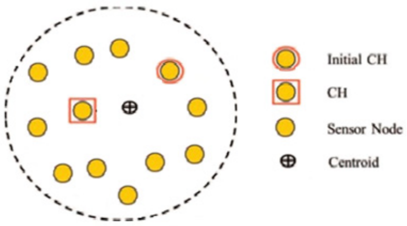

Fig. 4. Reclustering phase.

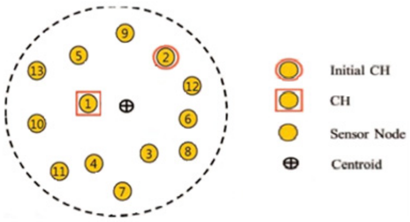

Fig. 5. Ordering of nodes with ID number.

```
Input: sensor nodes
Iteration:
if Residual energy of CH < Energy of threshold
    Cluster Members <- Verify ID ()
    Present CH = ChangeClusterHead ()
    Cluster Members <- InformsMsg ()
Send the aggregate data to the MDC
```

Algorithm 2: Reselection of cluster head.

$$d_i = \sqrt{\left(x_i - x_{\text{center}}\right)^2 + \left(y_i - y_{\text{center}}\right)^2}\,. \tag{10}$$

After the cluster formation, an ID number is assigned to each node of a cluster according to the distance to the center point (Fig. 5). The smaller number is assigned to the closest one. The residual energy of every CH is checked every round to maintain the connectivity of the network. The current CH's power is compared to the preset threshold. If it is smaller than the threshold, the node in the next ID order is selected as a new CH. In this case, the new CH informs other nodes of the updated roles in the cluster (Algorithm 2).

Single-hop routing protocol for data transmission: In each round, all nodes of the cluster transmit packets to CH. The proposed algorithm adopts a single-hop routing protocol for the CH to send the gathering data to the MDC at a specific time, which will handle collecting them to the BS.

4.2 Dividing the Network Environment

After selecting the cluster heads and creating the clusters, the field is partitioned into M sub-areas. The sub-areas have equal size. The BS is at the center of the network (Fig. 6). It is aware of the data gathering points elect for each sub-area. We use 2, 4 or 8 MDCs to collect the data from elected CHs. Each MDC is aware of its reserved sub-area. CHs are considered as data gathering points.

```
Initialization: CHs; M MDCs; P_j = {}
The BS splits the network into M equal subareas.
Iteration:
  for i = 1 to CHs {
        for j = 1 to M {
              if  CH_i is located in subarea j
              {  P_j <- {CH_i}
              }
        }
  }
  for j = 1 to M {
  For each subarea, an optimal path P_j is constructed using the Genetic algorithm .
  }
```

Algorithm 3: Partitioning the network environment.

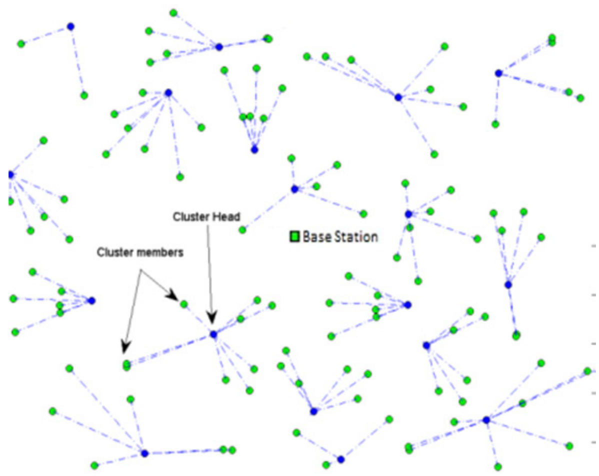

Fig. 6. Wireless sensor network topology.

During the data-gathering tour, each MDC starts the travel tour from the base station. It moves in the region, visiting the subset of CHs belonging to the sub-area. The BS is in the middle of the field. It is aware of the cluster heads elected in each round. The elected CHs in each round are classified into M sets named P. The set P_j comprises the CHs positioned in the jth sub-area (Algorithm 3).

4.3 The Optimal Path Construction Phase for MDC

The genetic algorithm*:* One of the most critical issues is finding the optimal data gathering path for MDC. It takes less data collection time. Several researches [31], [32] prove that GA [19] gives efficient results in terms of distance traveled for small to medium size instances.

The main idea of GA is to generate a novel generation better than the previous. They start from a population of initial potential solutions. The relative performance (fitness) is evaluated. Based on these performances, a new population of potential solutions using simple evolutionary operators: selection, crossover and mutation. A few individuals reproduce, others disappear, and only the best adapted individuals are supposed to survive.

The GA algorithm goes through the repeated application of five main functions, which are presented below:

Initial population: Two methods to generate a chromosome, randomly and heuristically. In this study, the heuristic method is used since the solution is found more quickly.

Fitness function: It should be determined to obtain the most effective solution in GA. The objective of the proposed algorithm is minimizing the length of the trajectory of the MDC. The fitness function that effectively offers the best solution is given in (11). M is the size of a chromosome (the number of elected cluster heads).

$$f_M = \frac{1}{\sum_{j=1}^{M} \operatorname{dist}(j, j+1)}. \tag{11}$$

Selection: The selection allows to statically identify the best individuals of the current population, this operation is based on the performance of the individuals, estimated using the function adaptation.

Crossover: Novel generations are made using the recombination of the solutions chosen in the breeding process.

Mutation: The role of the mutation is to randomly modify, with a certain probability, the value of a gene in a chromosome.

The optimal path selection: After CHs election, we apply the GA algorithm to construct the optimal path of MDC. We take an example of a network composed of 40 nodes; the HAC algorithm makes it possible to elect 5 CHs.

The initial population consists of 5 elected CHs and BS. The individuals are the CHs. The genes are the identifiers of the CHs, with 0 being the identifier of the BS.

The generation of the initial population: Each path is represented by an individual:

Path 1= {0, 16, 17, 28, 20, 39, 0}

Path 2= {0, 28, 20, 16, 17, 39.0}

Path 3= {0, 20, 28, 39, 17, 16.0}

Path 4= {0, 16, 28, 17, 39, 20.0}

Fitness Function: We evaluate these individuals using the fitness function; the objective of this function is to minimize the distance traveled. It calculates the total distance of each road.

Fitness (Path 1) = 1103.9004

Fitness (Path 2) = 996.1384

Fitness (Path 3) = 985.1184

Fitness (Path 4) = 1169.9628

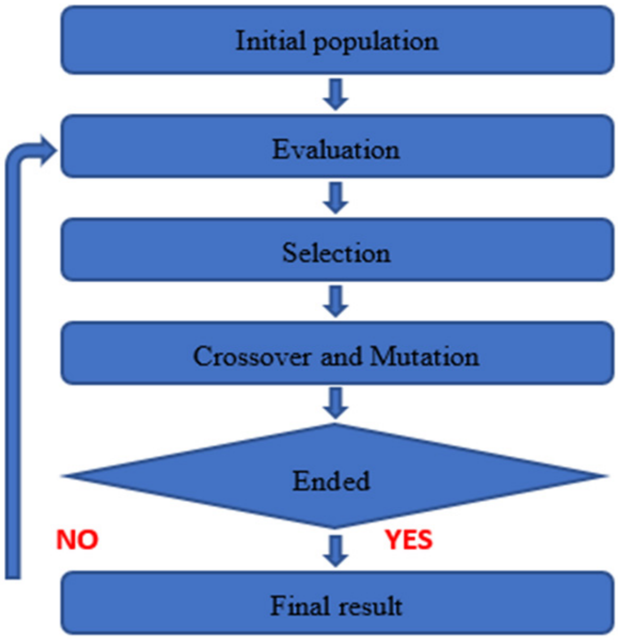

Fig. 7. Flowchart of the genetic algorithm.

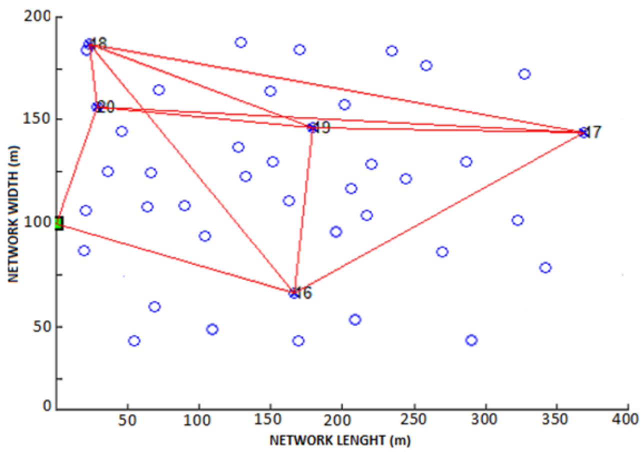

Fig. 8. Path before GA application.

Selection: The binary selection tournament is used to select both parents, where the contestant with better fitness function.

Based on the results obtained by the fitness function, Path 2 and Path 3 are selected. Then we apply the crossover and mutation operators on them.

Crossover: Its role is to combine the individuals selected by the selection operator to produce a new generation.

Two parents:

Path 2= {**0, 28, 20,** 16, 17, 39,0}

Path 3= {**0, 20, 28**, 39, 17, 16,0}

Two children:

S1= {**0, 20, 28**, 16, 17, 39, 0}; fitness(S1) = 1111.2137

S2= {**0, 28, 20,** 39, 17, 16, 0}; fitness(S2) = 1175.5287

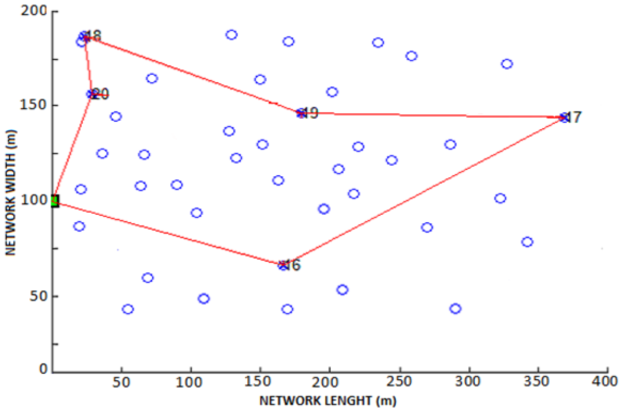

Fig. 9. The path obtained by the GA.

Mutation: We swap the two genes 16 and 39 of individual S1, and of individual S2.

S1= {0, 20, 28, 39, 17, 16, 0}, fitness(S1) = 985.1184

S2= {0, 28, 20, 16, 17, 39, 0}, fitness(S2) = 996.1384

Therefore, the resulting route will be given as {0, 20, 28, 39, 17, 16, 0} whose distance is 985.1184.

Figure 9 shows the shortest path selected to obtain the optimal solution.

5. Simulation and Analysis of Results

In this section, we evaluate and analyze the performance of our proposed approach with a different number of clusters. Then, we compare the proposed algorithm with [22], [23], and [27] algorithms under different scenarios to validate the network performance. For comparison, we use MATLAB R2016a as the simulator. We consider a network of 200 sensor nodes in an area of 200 m × 200 m. Sensor nodes have random uniform distribution, are placed in a fixed point with a unique identifier. We assume that the MDC has a rechargeable battery and a speed equal to 2 m/s as used in [30]. The BS is fixed, situated at the center, and has unlimited energy. The simulations parameters are in Tab. 1.

Parameter	Value
Size of the network	200 m × 200 m
N (Number of deployed nodes)	200
E_0 (initial energy of nodes)	1 J
E_{elec}	50 nJ/bit
E_{cpu}	7 nJ/bit
ε_{fs}	10 pJ/bit/m^2
ε_{amp}	0.0013 pJ/bit/m^4
Position of base station	(100 m, 100 m)
Packet size	4000 bits
velocity of MDC	2 m/s
M	2, 4, 8

Tab. 1. Simulation parameters.

5.1 The Assumptions Considered in the Simulation

The base station is fixed, situated at the center, and has unlimited energy. All sensors in the homogeneous conditions are uniform, and they have equal and limited initial energy.

They have random uniform distribution, are placed in a fixed point with a unique identifier.

The velocity of the Mobile Data Collector is constant. This device is equipped with a rechargeable battery and long-range transceiver.

5.2 Optimal Cluster Numbers

In this subsection, several numbers of simulations are evaluated to verify the performance of our planned approach. Initially, once the number of clusters equals 7% of the total number of nodes, the network is estimable as 582 rounds. Then, the number of cluster increases to 10% and therefore the network lifetime extends to 108 extra rounds. As the number of clusters increases, the number of sensor nodes among the cluster decreases. Additionally, the intra-cluster communication also decreased. In Fig. 10, the network lifetime is raised to 822 rounds once the number of clusters is augmented from 7% to 13%. We tend to induce

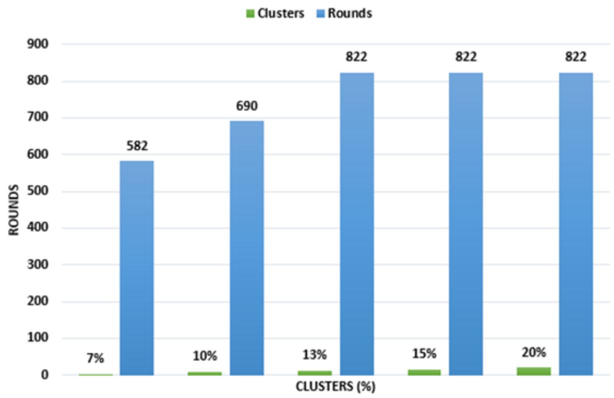

Fig. 10. The percentage of clusters vs. the number of rounds.

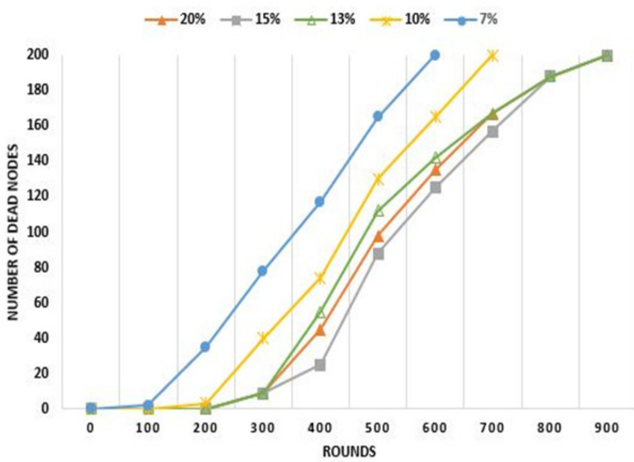

Fig. 11. The number of dead nodes vs. rounds.

identical performance when the number of clusters is given 13%, 15% and 20%. When the number of clusters increased, the distance between CHs and MDC reduces.

Therefore, the network lifetime is increased. However, the data gathering tour for the MDC is prolonged.

Figure 11 shows the number of dead nodes versus rounds with different numbers of selected CHs. We find that the number of CHs elected has a significant impact on the lifetime of the network. We tend to notice the most effective results when the number of CHs is chosen between 13% and 20%. Otherwise, in 7% of the clusters, the primary node dies after 101 rounds. In 10% of the clusters, the first node dies after 189 rounds.

5.3 Performance Measurement in Terms of Maximum Time Spent by MDCs

In this section, we will use a various number of MDCs to guarantee a faster data collection from the elected CHs. For that, we propose to divide the field into sub-areas and to assign one MDC to each area. By this method, we extend the lifetime of the network.

Figure 12 illustrates the time spent by MDCs to collect data from different CHs. We remark that the time spent gathering data for 8 MDCs is less than the time spent for 2 or 4 MDCs. The maximum time spent by MDCs tends to decrease when we increase the number of MDCs.

5.4 Comparison with Other Works

We compare our work to other existing approaches mentioned in [22], [23] and [27]. We simulate 200 randomly deployed sensor nodes at 200 m × 200 m. The number of CHs is sufficient for 13% of the population. This number of CHs is recognized to enable the most energy efficient of the network.

Network lifetime evaluation: Figure 13 shows the comparisons of the proposed approach with the other approaches in terms of dead nodes. The first node depletes energy in rounds 85, 123, 151, and 201, respectively, for [22], [23], [27] and also the proposed algorithm. Each node runs out of energy as the first for each algorithm. Table 2 shows the death of the nodes in 3 cases, the death of the first node (FND), the death of half of the nodes (HNA) and the death of all nodes of the network (LND).

Evaluation of the residual energy: Figure 14 presents the residual energy versus rounds of our proposed approach and approaches mentioned in [22], [23], [27]. We remark the death of all nodes after 650 rounds for the algorithms [22] and [23] while the LND is after 810 rounds for our proposed work. This is due to the use of MDC to collect data sent from an appropriate set of CHs using an optimal path.

Evaluation of number of packets delivered to BS: Figure 15 indicates the number of packets forward to the BS. The curves clearly show the improvement provided by

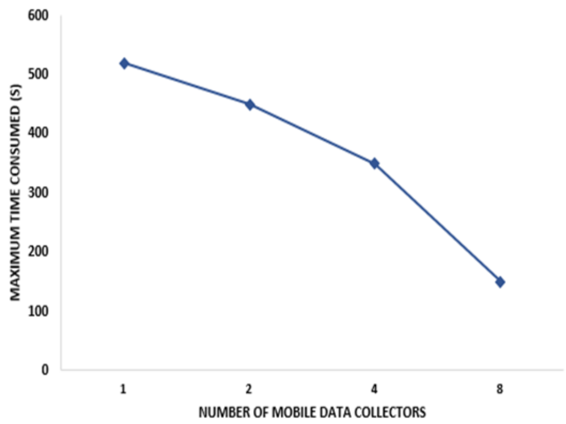

Fig. 12. The number of MDCs vs. time spent for gathering data.

Algorithm	FND	HNA	LND
[22]	85	252	480
[23]	123	354	620
[27]	151	402	750
PROPOSED WORK	201	510	822

Tab. 2. Values of FND, HNA and LND metrics for each algorithm after several rounds.

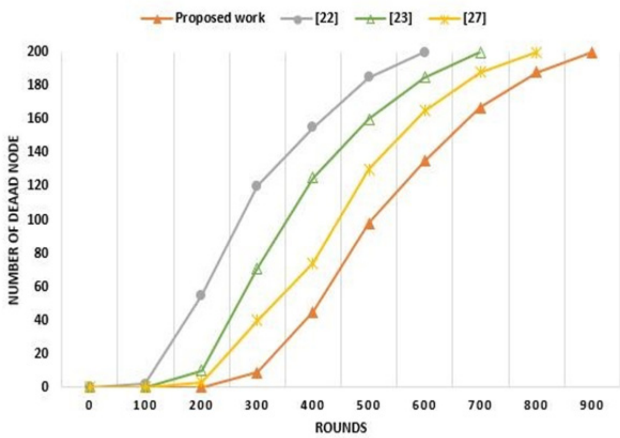

Fig. 13. Distribution of dead nodes vs. rounds.

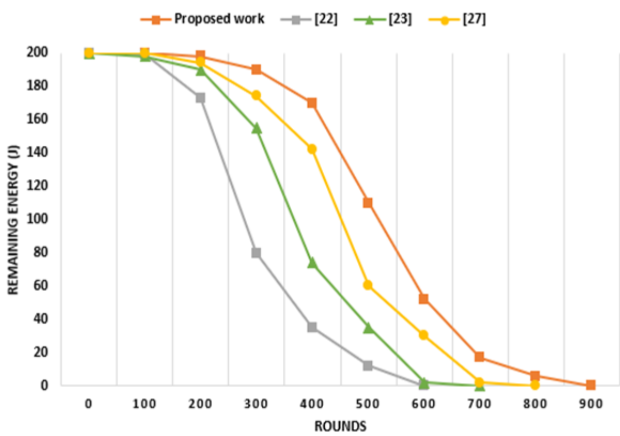

Fig. 14. Variation of residual energy vs. rounds.

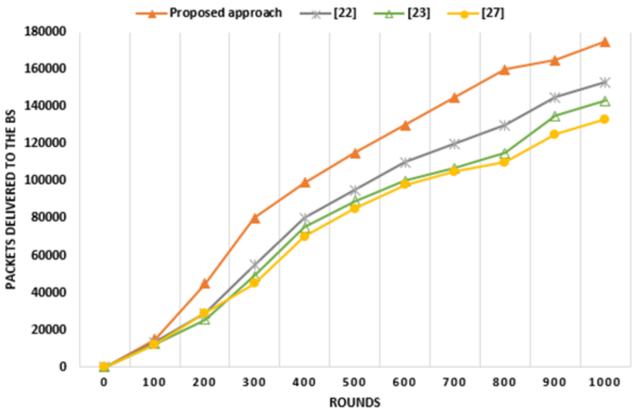

Fig. 15. Number of packets delivered to the BS.

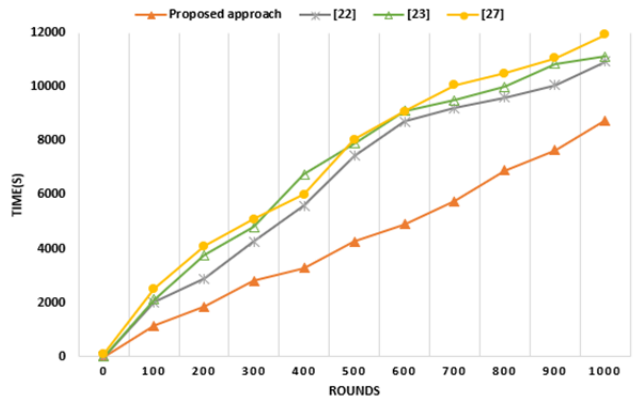

Fig. 16. Variation of residual energy vs. rounds.

the proposed approach compared to [22], [23] and [27] in terms of the number of packets received by the BS. This improvement is explained by the efficient management of the energy reserves in the proposed approach (the nodes longer live, sensors transmit more packets).

The number of packets delivered to BS in the proposed approach is equal to 175000 while the number of packets delivered to BS in [22], [23], and [27] is 153000, 147000 and 135000, respectively.

Time evaluation of the delay in sending packets to the BS: The delay time for sending packets to the BS depends on the distance traveled by the MDC. The delay time is high if the distance is large. Otherwise, it is low if the distance is small.

Figure 16 demonstrates that our proposed approach can significantly reduce the packet delay time compared to [22], [23] and [27].

This is mainly justified by the decrease in the travel length of MDC due to the use of GA to find the optimal path of MDC.

6. Conclusion and Perspectives

A novel approach using MDC is proposed to aggregate data from CHs and construct the optimal path.

The suitable CHs are elected for each cluster. Then, we propose the GA algorithm to select the best path for MDC. The MDC's path is modified iteratively in each round. The aggregation of data from CHs enhances the network lifetime.

To prove the effectiveness of our proposed approach, we compare our work to recent algorithms. We found that our proposed approach exceeds the existing ones relating to network lifetime and the number of packets delivered to BS.

In future contributions, we project to apply the machine learning process and the intelligent artificial IA to select the optimal path. Also, we plan to divide the network and employ more than one MDC for each subarea.

Acknowledgments

We are grateful for the support of the Department of Electrical and Computer Engineering at the New York University of Abu Dhabi (NYU).

References

[1] ROCHA, A., PIRMEZ, L., DELICATO, F., et al. WSNs clustering based on semantic neighbourhood relationships. *Computer Networks Journal*, 2012, vol. 56, no. 5, p. 1627–1645. DOI: 10.1016/j.comnet.2012.01.014

[2] KULAKOWSKI, P., CALLE, E., MARZO, J. L. Performance study of wireless sensor and actuator networks in forest fire scenarios. *International Journal of Communication Systems*, 2013, vol. 26, no. 4, p. 515–529. DOI: 10.1002/dac.2311

[3] YICK, J., MUKHERJEE, B., GHOSAL, D. Wireless sensor network survey. *Computer Networks,* 2008, vol. 52, no. 12, p. 2292– 2330. DOI: 10.1016/j.comnet.2008.04.002

[4] MANJUNATHA, T. N., SUSHMA, M. D., SHIVAKUMAR, K. M. Security concepts and Sybil attack detection in wireless sensor networks. International Journal of Emerging Trends and Technology in Computer Science, 2013, vol. 2, no. 2, p. 383–390. ISSN: 2278-6856

[5] CHENG, S., CHANG, T. Y. An adaptive learning scheme for load balancing with zone partition in multi-sink wireless sensor network. *Expert Systems with Applications*, 2012, vol. 39, no. 10, p. 9427–9434. DOI: 10.1016/j.eswa.2012.02.119

[6] EL FISSAOUI, M., BENI-HSSANE, A., SAADI, M. Energy efficient and fault tolerant distributed algorithm for data aggregation in wireless sensor networks. *Journal of Ambient Intelligence and Humanized Computing,* 2019, vol. 10, no. 2, p. 569–578. DOI: 10.1007/s12652-018-0704-8

[7] ABDOLKARIMI, M., ADABI, S., SHARIFI, A. A new multi-objective distributed fuzzy clustering algorithm for wireless sensor networks with mobile gateways. *AEU-International Journal of Electronics and Communications*, 2018, vol. 89, p. 92–104. DOI: 10.1016/j.aeue.2018.03.020

[8] SHAH, R. C., ROY, S., JAIN, S., et al. Data MULEs: Modeling a three-tier architecture for sparse sensor networks. In *Proceedings of IEEE Workshop on Sensor Network Protocols and Applications (SNPA)*. Anchorage (AK, USA), 2003, p. 30–41. DOI: 10.1109/SNPA.2003.1203354

[9] CHEN, T., CHEN, T., WU, P. On data collection using mobile robot in wireless sensor networks. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 2011, vol. 41, no. 6, p. 1213–1224. DOI: 10.1109/TSMCA.2011.2157132

[10] ZHAO, M., MA, M., YANG, Y. (2008). Mobile data gathering with space-division multiple access in wireless sensor network. In *Proceedings of 27th Conference on Computer Communications (INFOCOM)*. Phoenix (AZ, USA), 2008, p. 1957–1965. DOI: 10.1109/INFOCOM.2008.185

[11] SALARIAN, H., CHIN, K.-W., NAGHDY, F. An energy-efficient mobile-sink path selection strategy for wireless sensor networks. *IEEE Transactions on Vehicular Technology,* 2014, vol. 63, no. 5, p. 2407–2419. DOI: 10.1109/TVT.2013.2291811

[12] DONG, M., OTA, K., YANG, L. T., et al. LSCD: A low-storage clone detection protocol for cyber-physical systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems,* 2016, vol. 35, no. 5, p. 712–723. DOI: 10.1109/TCAD.2016.2539327

[13] DUAN, X., ZHAO, C., HE, S., et al., Distributed algorithms to compute Walrasian equilibrium in mobile crowdsensing. *IEEE Transactions on Industrial Electronics*, 2017, vol. 64, no. 5, p. 4048–4057. DOI: 10.1109/TIE.2016.2645138

[14] GHOSH, N., BANERJEE, I., SHERRATT, R. S. On-demand fuzzy clustering and ant-colony optimisation based mobile data collection in wireless sensor network. *Wireless Networks*, 2019, vol. 25, no. 4, p. 1829–1845. DOI: 10.1007/s11276-017-1635-6

[15] SERT, S. A., BAGCI, H., YAZICI, A. MOFCA: Multi-objective fuzzy clustering algorithm for wireless sensor networks. *Applied Soft Computing,* 2015, vol. 30, p. 151–165. DOI: 10.1016/j.asoc.2014.11.063

[16] SOMASUNDARA, A. A, RAMAMOORTHY, A., SRIVASTAVA, M. B. Mobile element scheduling for efficient data collection in wireless sensor networks with dynamic deadlines. In *25th IEEE International Real-Time Systems Symposium.* Lisbon (Portugal), 2004, p. 296–305. DOI: 10.1109/REAL.2004.31

[17] ZHANG, C., FEI, S. A matching game-based data collection algorithm with mobile collectors. *Sensors,* 2020, vol. 20, no. 5, p. 1–16. DOI: 10.3390/s20051398

[18] YALÇIN, S., ERDEM, E. A mobile sink path planning for wireless sensor networks based on priority-ordered dependent nonparametric trees. *International Journal of Communication Systems*, 2020, vol. 33, no. 12, p. 1–19. DOI: 10.1002/dac.4449

[19] YOUNES, A., BADAWI, U. A., FARAG, T. H., et al. A genetic algorithm to find the minimum cost paths tree with bandwidth constraint in the computer networks. *International Journal of Applied Engineering Research,* 2018, vol. 13, no. 10, p. 7472–7476. ISSN: 0973-4562

[20] KAMAREI, M., PATOOGHY, A., SHAHSAVARI, Z., et al. Lifetime expansion in WSNs using mobile data collector: A learning automata approach. *Journal of King Saud University - Computer and Information Sciences,* 2018, vol. 32, no. 1, p. 65–72. DOI: 10.1016/j.jksuci.2018.03.006

[21] ALPARSLAN, D. N., SOHRABY, K. Two-dimensional modeling and analysis of generalized random mobility models for wireless ad hoc networks. *IEEE/ACM Transactions on Networking*, 2007, vol. 15, no. 3, p. 616–629. DOI: 10.1109/TNET.2007.893873

[22] WU, Q., SUN, P., BOUKERCHE, A. A novel data collector path optimization method for lifetime prolonging in wireless sensor networks. In *2019 IEEE Global Communications Conference (GLOBECOM)*. Waikoloa (HI, USA), 2019, p. 1–6. DOI: 10.1109/GLOBECOM38437.2019.9013989

[23] WEN, W., ZHAO, S., SHANG, C. EAPC: Energy-aware path construction for data collection using mobile sink in wireless sensor networks. *IEEE Sensors Journal*, 2020, vol. 18, no. 2, p. 890–901. DOI: 10.1109/JSEN.2017.2773119

[24] HA, I., DJURAEV, M., AHN, B. An optimal data gathering method for mobile sinks in WSNs. *Wireless Personal Communication,* 2017, vol. 97, p. 1401–1417. DOI: 10.1007/s11277-017-4579-3

[25] RAO, X., HUANG, H., TANG, J., et al. Residual energy-aware mobile data gathering in wireless sensor networks. *Journal of Telecommunications Systems,* 2016, vol. 62, p. 31–41. DOI: 10.1007/s11235-015-9980-1

[26] SUN, Y., DONG, W., CHEN, Y. An improved routing algorithm based on ant colony optimization in wireless sensor networks. *IEEE Communications Letters,* 2019, vol. 21, no. 6, p. 1317–1320. DOI: 10.1109/LCOMM.2017.2672959

[27] ABIDOYE, A. P., KABASO, B. Energy-efficient hierarchical routing in wireless sensor networks based on fog computing. *EURASIP Journal on Wireless Communications and Networking*, 2021, p. 1–26. DOI: 10.1186/s13638-020-01835-w

[28] JLASSI, W., HADDAD, R., BOUALLEGUE, R., et al. A combination of K-means algorithm and optimal path selection method for lifetime extension in wireless sensor networks. In *International Conference on Advanced Information Networking and Applications*. 2021, p. 416–425. DOI: 10.1007/978-3-030-75078-7_42

[29] EL FISSAOUI, M., BENI-HSSANE, A., SAADI, M. Energy aware hybrid scheme of client-server and mobile agent models for data aggregation in wireless sensor networks. In *Proceedings of the 16th International Conference on Hybrid Intelligent Systems (HIS 2016).* 2016, p. 227–232. DOI: 10.1007/978-3-319-52941-7_23

[30] TASHTARIAN, F., MOGHADDAM, M. H. Y., SOHRABY, K., et al. On maximizing the lifetime of wireless sensor networks in event-driven applications with mobile sinks. *IEEE Transactions on Vehicular Technology,* 2015, vol. 64, no. 7, p. 3177–3189. DOI: 10.1109/TVT.2014.2354338

[31] MAHESHWARI, P., SHARMA, A., VERMA, K. Energy efficient cluster-based routing protocol for WSN using butterfly optimization algorithm and ant colony optimization. *Ad Hoc Networks*, 2021, vol. 110, p. 1–15. DOI: 10.1016/j.adhoc.2020.102317

[32] LAXMA REDDY, D., PUTTAMADAPPA, C., SURESHB, H. N. Merged glowworm swarm with ant colony optimization for energy efficient clustering and routing in wireless sensor network. *Pervasive and Mobile Computing,* 2021, vol. 71, p. 1–18. DOI: 10.1016/j.pmcj.2021.101338

About the Authors …

Wadii JLASSI (corresponding author) was born in Tunisia. He received M.S. degree in Telecommunication from the Faculty of Science of Bizerte in 2016. He is currently pursuing the Ph.D. degree in the National Engineering School of Tunis. His research interests include energy optimization algorithms, IT network security and internet of things.

Rim HADDAD was born in Tunisia. She received the B. Eng. Degrees in Telecommunication from the High School of Telecommunication of Tunis. From 1997 to December 2006, she was Engineer in Telecommunication in Tunisie Telecom. Since January 2007 she was a university assistant in the High School of Computing. She joined the Higher Institute of Communications Iset'Com as an assistant professor in telecommunication in 2010. Her research interests include telecommunication networks, IT network security, VoIP technology, new generation networks, wireless networks, cloud computing, big data and mobile computing.

Ridha BOUALLEGUE was born in Tunisia. He received the M.S degree in Telecommunications in 1990, the Ph.D. degree in Telecommunications in 1994, and the HDR degree in Telecommunications in 2003, all from the National School of Engineering of Tunis (ENIT), Tunisia. Director and founder of the National Engineering School of Sousse in 2005. Director of the School of Technology and Computer Science in 2010. Currently, Prof. Ridha Bouallegue is the director of the Innovation of Communicant and Cooperative Mobiles Laboratory, INNOV'COM Sup'COM, Higher School of Communication. His current research interests include mobile and cooperative communications, access technique, intelligent signal processing, CDMA, MIMO, OFDM and UWB systems.

Raed SHUBAIR Full Professor of Electrical Engineering and Regional Director of IEEE; Founding Chair of IEEE APS Educational Initiatives Program; Founding Member of six IEEE society chapters in UAE; Founder & Chair IEEE at NYU Abu Dhabi. Research interests: wireless; communications; antennas; signal processing; machine learning; terahertz; biomedicine; mobile health. IEEE Distinguished Educator Award nominee; Recipient of 2020 IEEE Award of the Year.