Energy-Efficient Path Construction for Data Gathering Using Mobile Data Collectors in Wireless Sensor Networks

Wadii JLASSI [1], *Rim HADDAD* [2], *Ridha BOUALLEGUE* [3]

[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

wadii.jlassi@gmail.com, rim.haddad@fsg.ulaval.ca, ridha.bouallegue@gmail.com

Submitted September 30, 2022 / Accepted September 4, 2023 / Online first October 17, 2023

Abstract. *Energy is seen as a significant factor in wireless sensor networks (WSNs). It is a challenge to balance between battery lifetime of the different sensors and network lifetime. The main contribution of the proposed approach is to decrease the energy consumption of each sensor node, overcome unbalanced energy usage among sensor nodes, reduce the data gathering time and enhance the network lifetime. To achieve these goals, we combine the Hierarchical Agglomerative algorithm and an optimal path selection method. First, the suitable cluster heads (CHs) are elected based on the Euclidean distance and the residual energy of each sensor node. Then, the base station is situated at the center of the field, which will be partitioned into equal subareas, one for every mobile data collector (MDC). Second, the Kruskal algorithm is used to create an optimal data gathering path from each subset of elected cluster heads. Finally, each mobile data collector travels the optimal path to collect the data from the set of cluster heads of each subarea and returns periodically to the base station to upload gathered data. Computer simulation proves that the proposed approach outperforms existing ones in terms of data gathering time, residual energy and network lifetime.*

Keywords

Wireless Sensor Network (WSN), optimal path, clustering, Mobile Data Collector (MDC), Cluster Head (CH)

1. Introduction

The development of WSNs plays several roles in numerous aspects of human life, such as military surveillance, underwater monitoring, air traffic control, volcanic eruption, medical care, and so on [1], [2].

They have termed ad hoc networks [2]. Sensor nodes are distributed in a specific 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]. In most cases, WSNs consist of static nodes that forward the captured information towards one base station (BS) through direct or more than one hop after each event occurrence [5].

Clustering [6] is an effective method to overcome the imbalance in power consumption in WSN. According to the existing research, there are two approaches for CHs selection 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. In the second approach, the base station performs the election of the suitable CHs and the formation of the clusters. Every round, each node transmits its residual energy and location information to BS. Then, the BS evaluates the probabilities of every sensor node to be CH and announces to nodes concerning their new status (CH or cluster member).

Recently, several researchers have demonstrated the efficiency of MDCs for data gathering and energy conservation in WSNs [8], [9]. Beginning from the BS and visiting the area of interest, MDC collects data from polling points (PPs) and forwards it to the BS. This mission has benefits such as minimizing data collection time, conserving energy and extending network lifetime.

An MDC is a mobile device that contains a rechargeable battery, GPS devices, long range transceivers, and large memory units that traverses the network to gather data [9]. Cluster heads save the received information and transmit them 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].

The main objective is that the previous approaches aim to minimize the energy consumption and optimize the network lifetime. However, most of them do not contemplate the trajectory cost between data gathering points or utilize

DOI: 10.13164/re.2023.0502

one MDC to collect data from large-scale networks. Also, these approaches do not reduce energy consumption and data gathering time efficiently.

We propose a WSN that includes n nodes in the area. Every node has location information obtained during the network deployment time. The appropriate cluster heads (CHs) send their aggregated data when MDC visits them.

The MDC starts the tour from BS to appropriate CHs to gather information in a single hop at sufficient sojourn time. 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]. In the static approach, the path is established before the data collection phase from the elected cluster heads. However, in the second case, the MDC might modify its path throughout its travel. It needs high computational complexity. In our proposed approach, the MDC travels a predefined static path which is constructed by the Kruskal algorithm.

During each round, the Hierarchical Agglomerative Clustering algorithm (HAC) is employed for choosing suitable CHs. It allows to decrease intra-cluster communications since the number of information exchanged to obtain the optimal clustering is reduced and gain energy efficiency for sensor nodes. The Kruskal algorithm is utilized to obtain the optimal path for each MDC to conserve energy, overcome the imbalance in energy consumption of the network and minimize the data gathering time.

Our contribution could be briefed as follows:

- Propose a clustering algorithm to find the suitable CHs and form clusters, based on the residual energy and the distance between sensor nodes;
- Divide the network area into equal subarea;
- Employ the Kruskal algorithm to construct an optimal data collection path for MDCs.

The paper is organized as follows. Related works are discussed in Sec. 2. An energy model is presented in Sec. 3. The proposed approach is presented in Sec. 4. Computer simulations are described in Sec. 5. In the end, the conclusion and future works are defined in Sec. 6.

2. Related Works

Recently, various data gathering approaches are introduced for energy optimization to improve the operation time of the entire network and decrease power consumption using MDCs. To save maximum energy, one or more mobile data collectors can be used to collect data from CHs within a specific time limit [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]. 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 present a residual Energy Aware Mobile Data Collection Scheme in Wireless Sensor Networks (REAMDC). This approach aims to combine multi-hop 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 [22]. 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 [23], 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.

In [25], 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.

In [27], 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 this study, we propose a clustering approach using MDCs in WSNs. This approach purposes a HAC algorithm that selects the suitable cluster heads considering several parameters like residual energy and the distance between nodes in a single-hop context. The HAC algorithm aims to manage the energy consumption of each sensor node and minimize intra-cluster communications. We gain compared to the intra-cluster communications because the number of information exchanges to obtain the optimal clustering is reduced which explains the gain in energy of the sensors. Also, an optimal path selection method is proposed to construct the shortest data collection path for each MDC to decrease the packet delay time.

3. Power Analysis

The energy of sensor nodes is almost entirely consumed while transmitting, receiving packets, and processing data. The energy model used is the same one used in [28], which defines the following possibilities: free and multipath space. The quantity of energy depleted for forwarding and receiving *k*-bit packets is given respectively by (1), (2):

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

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

where d the distance between a sending node and a receiving node. E_{elec} is the power necessary for sending one bit and E_{RX} is 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):

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

In one round, the elected CH (the center of each cluster) drains a significant amount of energy compared to the other cluster members due to receiving data from it and transmitting it to MDC. CH's energy E_{CH} can be estimated as (4):

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

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

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

where n is the number of cluster member (non-CH sensors) in one cluster. E_{NctoCH} represents the CH's energy to collect transmitted data from cluster members. m is the number of member nodes in one cluster. d_{CHtoMDC} represents the distance between the CH and the MDC. E_{CHtoMDC} represents the CH's energy to forward data from cluster members to MDC.

4. Steps of the Proposed Approach

The principal stages of our work are as follows:

- Elect suitable CHs and form clusters,
- Dividing network environment into equal subareas,
- Construct optimal path.

The first objective is to manage energy efficiency in WSN and reduce the data gathering time. To reach this goal, we suggest the HAC algorithm [31]. Then, we try to construct an optimal path rooted at BS to transmit the aggregated data via MDC.

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 HAC algorithm is proposed (Algorithm 1). It divides the network into clusters based on Euclidean distances between nodes and residual energy. Also, it aims to elect CHs for each cluster. In each round, we acquire a new subset of CHs and a new traveling path.

The proposed algorithm is presented as follows:

Initial Clustering: For the first iteration, the network is partitioned into a predefined number of k clusters. Each cluster contains a several number of sensor nodes. The base station (BS) is aware of the location information of each node in the network. In this step, the HAC algorithm selects the closest node to the BS of each cluster as CH based on the minimum distance estimated min (d_i) (Fig. 1).

The node selected as CH could be modified after the reclustering and choosing the CH phase. After the election of the suitable CHs, the optimal path will be constructed. d_i presents the distance between node i and BS.

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

$$d_{i}=\sqrt{(x_{i}-x_{\mathrm{bs}})^{2}+(y_{i}-y_{\mathrm{bs}})^{2}}\,. \tag{7}$$

Reclustering: After the association of all sensors to the k clusters, the center point (x_{center}, y_{center}) is calculated by the following formula (8), (9) with:

$$x_{\mathrm{center}}=\frac{1}{\sum_{i=1}^{s}E_{i}}\cdot\sum_{j=1}^{s}E_{j}x_{j}, \tag{8}$$

$$y_{\mathrm{center}}=\frac{1}{\sum_{i=1}^{s}E_{i}}\cdot\sum_{j=1}^{s}E_{j}y_{j}. \tag{9}$$

The *center* point is a virtual node located at the center of the cluster in the first iteration.

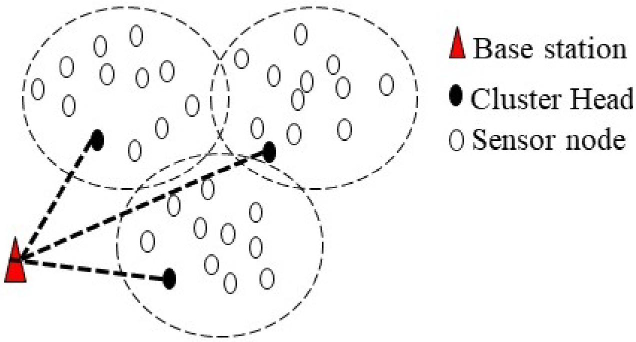

Fig. 1. Initial clustering.

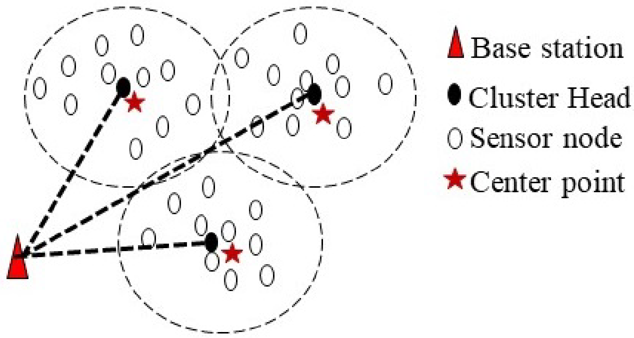

Fig. 2. Reclustering phase.

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

Algorithm 1. CHs election and cluster formation.

As the algorithm progresses, the center point converges to the node with the largest residual energy. The CH is still the nearest node to the center point with the highest residual energy. That prolongs the life of every node and facilitates communication between nodes and CH (Fig. 2).

Choosing the CH: Finally, the node closest to the center point (x_{center}; y_{center}) is selected as the new CH. For that, we estimate the Euclidean distance between the nodes of every cluster and the novel center point based on (10):

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

After the cluster formation, an ID number is attributed to every cluster member according to the distance to the center point (Fig. 3). The smallest number is assigned to the closest one. In every round, the residual energy of each CH

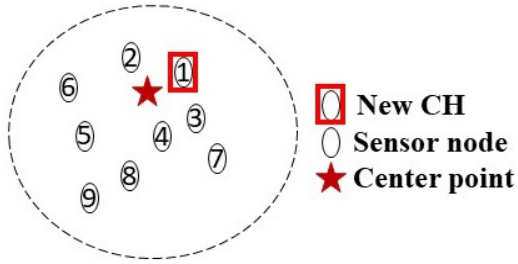

Fig. 3. Sorting of sensor nodes based on the ID number.

```
Input: sensor nodes
Iteration:
if Residual energy of CH < Energy of threshold
{
    Cluster Members ← VerifyID ()
    Present CH = ChangeClusterHead ()
    Cluster Members ←InformsMsg ()
}
Send the aggregate data to the MDC.
```

Algorithm 2. CHs election and cluster formation.

is inspected to maintain network connectivity. The current CH power is compared with the preset threshold value. If it is less than the threshold, the node in the next ID order is nominated as a novel CH. In this case, the new CH reports other nodes of the updated roles in the cluster (Algorithm 2).

The data transmission: The proposed algorithm assumes 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 base station (BS).

4.2 Partitioning the Network Environment and MDCs's Path Construction Phase

Partitioning the network: After selecting the CHs and creating the clusters, the network is partitioned into M subareas. The subareas have equal size. We use 2, 4 or 8 MDCs to collect the data from elected CHs. Each MDC is aware of its reserved subarea. CHs are considered as data gathering points. During the data gathering tour, each MDC begins the travel from the BS and visits the subset of CHs of each subarea. The BS is aware of the CHs elected in each round. The elected CHs in each round are grouped into M sets named P. The set $\mathbf{P}_j$ comprises the CHs situated in the jth subarea (Algorithm 3).

```
Initialization: CHs; M MDCs; Pj = {}
The BS splits the network into M equal subareas.
for i = 1 to CHs {
      { for j = 1 to M
            if   CHi is located in subarea j
            {  Pj ← {CHi}
             }
      }
}
for j = 1 to M {
For each subarea, an optimal path Pj is constructed using the Kruskal algorithm and Hamilton cycle.
}
```

Algorithm 3. Partitioning the network environment.

An optimal path for MDC: One of the most critical issues is finding the optimal data gathering path for MDC. It takes less data collection time. The Kruskal algorithm [19] is adopted to obtain the optimal data gathering path for each subset $\mathbf{P}_j$ (Algorithm 4). It commences with a tree that consists of a single vertex and grows continuously, one edge at a time. It comes to a halt when all the vertices have been attained. After CHs selection and division of the field into equal subareas, the Kruskal algorithm constructs the optimal data gathering path for every subarea. The BS is the first and the arrival point for the MDC tour. The algorithm takes into account all edges of the graph by raising the weight. It is used to discover the minimum weight tour that goes through all sets of cluster heads. The edges of the graph are ordered by raising the weight, and for each of them, it selects it if it does not make a cycle. Once all of the vertices match, an optimal path is found. As far as the path goes across each edge one time without repeating, the Kruskal algorithm path may pass through a vertex more than one time. For example, we suppose 5 vertices (Fig. 4):

BS: the base station is taken as starting point.

B, C, D and E: the Cluster Heads.

The edges are such as BSB = 24.1 and ED = 26.2, etc., are paths connecting two CHs or CH and BS. The Kruskal algorithm is created and the following path is obtained:

BS → B →E →C and D.

However, the algorithm does not give an optimal solution and the BS is not the arrival point for the MDC tour. To solve the problem, a Hamilton cycle is proposed, which passes by every node of the graph once and the BS is the first and the arrival point for the MDC tour as shown in Fig. 5. After adopting the Kruskal algorithm and Hamilton cycle, we obtained the optimal path for mobile data collector as follows:

BS→B→E→D→C→BS.

```
E: = {} set of edges
for each vertex v of G do {
      generate Set(v)
for each edge (u, v) in G sorted by rising weight do {
      if find (u) not equal find (v) then {
            join (u, v) to the set E
            union (u, v)
      }
}
refer E until get a connected graph
}
```

Algorithm 4. The optimal path construction algorithm.

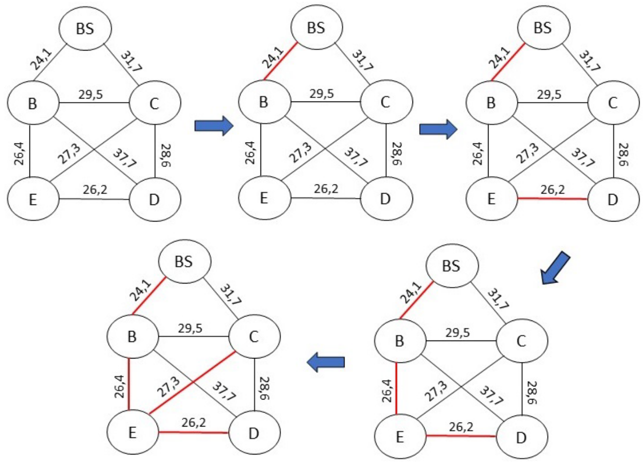

Fig. 4. Steps of the Kruskal algorithm.

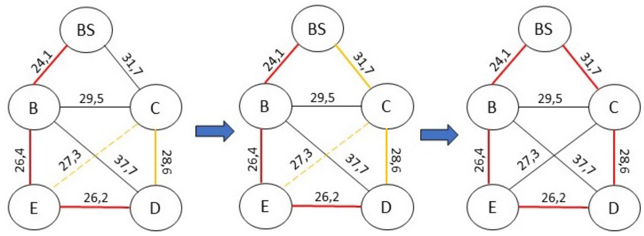

Fig. 5. An optimal path for mobile data collector.

4.3 Execution of Data Collection

After CHs selection and optimal path construction, the MDC saves the location information of the data gathering points on its memory. It starts from the BS and visits all the new data gathering points via the predefined trajectory. In each round, the MDC travels in a predefined path to gather data from the CHs. When it arrived in the range of each CH, it brings to a halt and transmits a message. This message includes an ID number and password. When the CHs receive this message with the corresponding ID number, they send an accept message to MDC. After the reception of the confirmation message, a communication channel is obtained between the CHs and the MDC and each CH sends the aggregate data. After each tour, the MDC arrives at BS and uploads aggregate data. After that, MDC prepares the network to travel for the next round.

5. Simulation and Analysis of Results

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, 27, 17, 25] algorithms under different scenarios to validate the network performance. For comparison, we use the metrics used in [28] and 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.

5.1 Optimal Cluster Numbers

In this subsection, several scenarios are evaluated to check the performance of our planned approach. The network is partitioned into a predefined number of clusters. k is chosen depending on the total number of sensors predefined before simulations.

Figure 6 shows the percentage of clusters versus the number of rounds. Initially, when the number of clusters is

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.

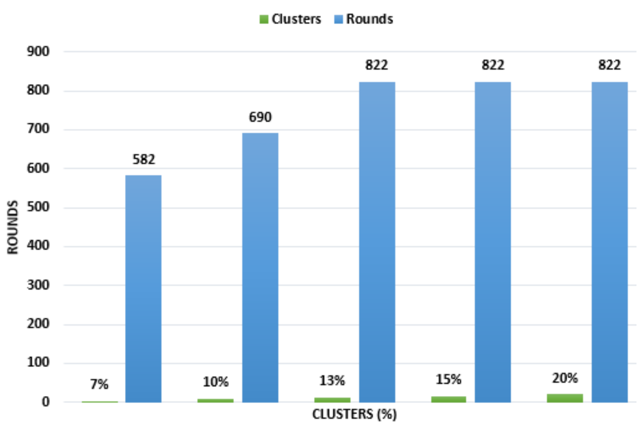

Fig. 6. The percentage of clusters vs. the number of rounds.

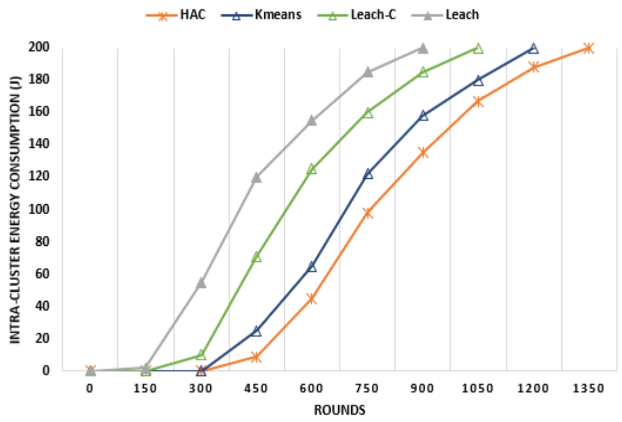

Fig. 7. The intra-cluster energy consumption vs. rounds.

equal to 7%, the network is estimable as 582 rounds. Then, the number of clusters increases to 10% and therefore the network lifetime increases to 108 extra rounds. Additionally, the intra-cluster communication also decreased. As the cluster number increases, the size of the cluster decreases.

The network lifetime is raised to 822 rounds once the number of clusters is augmented from 7% to 13%. We tend to induce identical performance when the number of clusters is given 13%, 15%, and 20%.

When the number of clusters will be increased, the distance and the communications between sensor members and the CH of each cluster will be decreased and therefore the network lifetime increases. The number of rounds increases up to a point where it remains constant. This point is the lifetime of the WSN for each number of k clusters.

5.2 Intra-Cluster Energy Consumption

Figure 7 presents the intra-cluster energy consumption versus rounds of our proposed approach, k-means [23], Leach-C [2], and Leach [3]. We can see that the HAC algorithm consumes less intra-cluster energy compared to k-means, Leach-C, and Leach. This is due to the fact that the number of information exchanged to obtain the optimal clustering is reduced. Therefore, intra-cluster communications are decreased.

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 faster data collection from the elected CHs. For that, we partition the field into subareas and attribute one MDC to every area.

Figure 8 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 path length of every MDC will decrease as the number of MDCs increases and therefore maximum time spent by every MDC tends to diminish when it travels the path to gather data from CHs of each subarea.

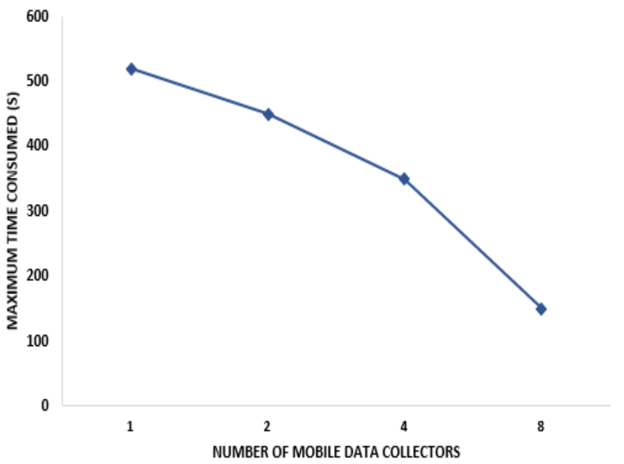

Fig. 8. The number of MDCs vs. time spent for gathering data.

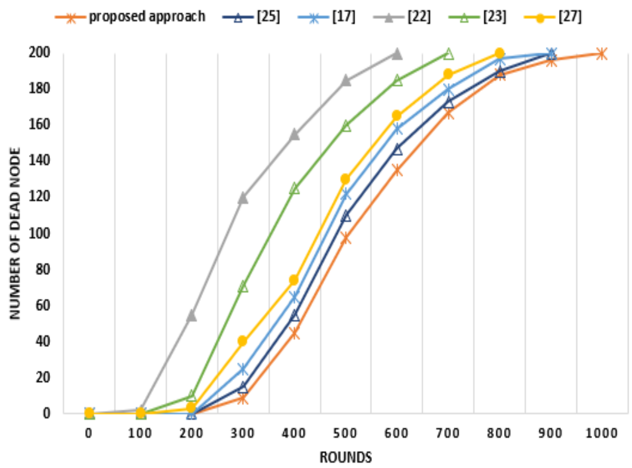

Fig. 9. Distribution of dead nodes vs. rounds.

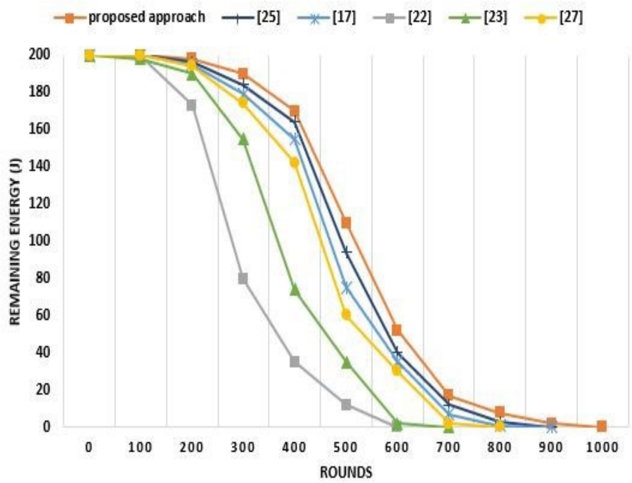

Fig. 10. Variation of residual energy vs. rounds.

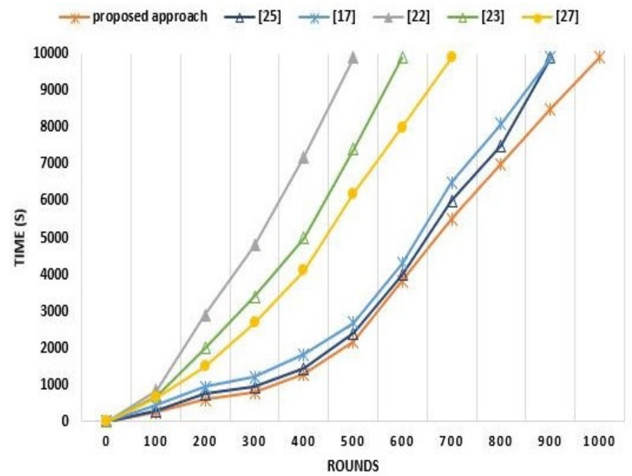

Fig. 11. Variation of residual energy vs. rounds.

5.4 Comparison with Other Works

We compare our work to other existing approaches mentioned in [22, 23, 27, 17] and [25]. We simulate 200 randomly deployed sensor nodes at 200 × 200 meters. 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: 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). Figure 9 shows the comparisons of the proposed approach with the other approaches in terms of dead nodes. We remark that the proposed approach has the lowest number of dead nodes from the beginning to all depletion of nodes. The first node depletes energy in rounds 96, 135, 163, 182, 193, and 207, respectively, for [22, 23, 27, 17, 25]. However, the first node depletes from 201 rounds in the proposed approach. After 865 rounds, all nodes of the network die with [22, 23, 27, 17, 25]. However, about 25 nodes are still alive with the proposed approach.

Evaluation of the residual energy: Figure 10 presents the residual energy versus rounds of our proposed approach and approaches mentioned in [22, 23, 27, 17, 25]. We remark the death of all nodes after 600 rounds for the algorithms [22] and [23] and after 900 rounds for the algorithms [27], [17], and [25]. However, some sensor nodes still have residual energy until 1000 rounds with the proposed approach. This is due to the decrease in intra-cluster communications and the use of MDCs to collect data sent from an appropriate set of CHs using an optimal path.

Approaches	FND	HNA	LND
[22]	96	255	554
[23]	135	357	622
[27]	163	405	754
[17]	182	458	848
[25]	193	489	869
PROPOSED APPROACH	207	513	953

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

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 11 demonstrates that our proposed approach can significantly reduce the packet delay time compared to [22, 23, 27, 17, 25].

This is mainly justified by the decrease in the travel length of each MDC due to the use of the optimal selection method to find the shortest path of each MDC.

6. Conclusion and Perspectives

One of the most critical issues in wireless sensor networks is conserving energy to enhance the network lifetime. In this study, we employ a clustering algorithm and an optimal path selection method using mobile data collectors to reduce the data gathering path, decrease power consumption, and extend the network lifetime.

We use the HAC algorithm to make clusters and choose suitable cluster heads (CHs). Then, the BS partitioned the network into equal subareas. We propose the Kruskal algorithm and the Hamilton cycle to select the best path for each MDC of each subarea. The MDC's path is mounted iteratively in each round to gather the data from the set of cluster heads. The aggregation of data from CHs aims to minimize the energy consumption of sensor nodes and enhance network lifetime.

We resolve the overall performance of our proposed technique and evaluate it to recent algorithms. We found that our proposed approach exceeds the existing ones relating to data gathering time, residual energy, and network lifetime.

As future contributions, we project to apply the machine learning process to form clusters and elect the data gathering points according to the residual energy, the number of packets transferred between nodes, and the density. Also, we attempt to evaluate a method that optimizes the number of MDCs and utilize artificial intelligence (AI) to construct a controllable trajectory.

Furthermore, we aim to propose a method for detecting and preventing possible attacks in the approach which we have already designed in the presence of one or multiple MDCs. The proposed method allows to avoid the loss of valuable data, minimize the unnecessary traffic caused by the attacks and ensure the proper functioning of the protocols and therefore preserve energy and maximize the network lifetime.

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 neighborhood 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 to 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] 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

[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. 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] 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

[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] 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

[23] 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

[24] TASHTARIAN, F., YAGHMAEE MOGHADDAM, M. H., 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

[25] 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

[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] WEN, W., ZHAO, S., SHANG, C., et al. 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

[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] 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

[30] JLASSI, W., HADDAD, R., BOUALLEGUE, R., SHUBAIR, R. Increase the lifetime of wireless sensor network using clustering algorithm and optimal path selection method. *Radioengineering,* 2022, vol. 31, no. 3, p. 301–311. DOI: 10.13164/re.2022.0301

[31] 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

About the Authors …

Wadii JLASSI (corresponding author) was born in Tunisia. He received the M.S. degree in Telecommunication from the Faculty of Science of Bizerte in 2016. He is currently pursuing the Ph.D. degree in 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 an 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 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.