An Efficient Mobile Data Gathering Method with Tree Clustering Algorithm in Wireless Sensor Networks Balanced and Unbalanced Topologies

Meriem MEDDAH (✉ meddah_meriem@hotmail.fr)

Ecole Nationale d'Ingenieurs de Tunis <https://orcid.org/0000-0002-2260-6025>

Rim HADDAD

high school of communication of tunis

Tahar EZZEDDINE

National Engineering School of Tunis: Ecole Nationale d'Ingenieurs de Tunis

Manuscript

Keywords: Wireless sensor Networks, Minimum Spanning Tree, Residual Energy, Mobile Data Collector, Shortest Path Tour

Posted Date: February 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-193017/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Version of Record: A version of this preprint was published at Wireless Personal Communications on November 1st, 2021. See the published version at <https://doi.org/10.1007/s11277-021-09309-y>.

AN EFFICIENT MOBILE DATA GATHERING METHOD WITH TREE CLUSTERING ALGORITHM IN WIRELESS SENSOR NETWORKS BALANCED AND UNBALANCED TOPOLOGIES

Meriem Meddah¹ Rim Haddad² Tahar Ezzedine³

¹Innov'Com Laboratory, National Engineering School of Tunis,
University Tunis El Manar meddah_meriem@hotmail.fr

²Innov'Com Laboratory, High School of Communication of Tunis,
University of Carthage rim.haddad@supcom.tn

³Sys'Com Laboratory, National Engineering School of Tunis,
University Tunis El Manar tahar.ezzedine@enit.rnu.tn

ABSTRACT

Mobile Data Collector device (MDC) is adopted to reduce the energy consumption in Wireless Sensor Networks. This device travels the network in order to gather the collected data from sensor nodes. This paper presents a new Tree Clustering algorithm with Mobile Data Collector in Wireless Sensor Networks, which establishes the shortest travelling path passing through a subset of Cluster Heads (CH). To select CHs, we adopt a competitive scheme, and the best sensor nodes are elected according to the number of packets forwarded between sensor nodes, the number of hops to the tree's root, the residual energy, and the distance between the node and the closest CH. In simulation results, we adopt the balanced and unbalanced topologies and prove the efficiency of our proposed algorithm considering the network lifetime, the fairness index and the energy consumption in comparison with the existing mobile data collection algorithms.

KEYWORDS

Wireless sensor Networks, Minimum Spanning Tree, Residual Energy, Mobile Data Collector, Shortest Path Tour

1. INTRODUCTION

In the recent years, several studies have considered Wireless Sensor Networks (WSN) as growing technologies. A large number of low-cost and low-power sensor nodes are densely deployed in monitoring area. These nodes provide flexibility and self-organization, in addition to a wide variety of services, to communicate and send data to the Base Station (BS). Previous researches have demonstrated the importance of clustering in WSNs. The way that sensor nodes forward data to the BS differs from an algorithm to another. Generally, sensor nodes are regrouped into clusters, in each one a Cluster Head (CH) receives packets from its cluster's member, then forwards these data to the BS in multi-hop or single-hop manner [1] [2]. However, in large fields, sending data from sensor nodes to the BS becomes difficult. For this reason, using a Mobile Data Collector (MDC) [3] to aggregate data presents several advantages such as saving time, balancing the energy consumption of sensor nodes and conserving network energy.

A MDC is equipped with a rechargeable battery and a powerful transceiver. It achieves energy and time efficient collection starting from the BS and traveling over the network. In the

literature, The MDC planning trajectory could be static or dynamic. In the first case, the path is established before the data gathering phase. Contrary to the dynamic approach in which the MDC could change its trajectory during its fly.

In this paper a new Tree Clustering Algorithm with Mobile Data Collector is proposed (TCAMDC). It is an energy efficient and distributed data gathering algorithm itemized with the following phases: the Initialization phase, and the CH election and the path construction phase. We assume that the BS is the first CH. The MST is divided into sub-trees while selecting the CHs and defining the path traversed by the MDC. The election of the best nodes as CH is a weight-based approach, considering different parameters which are: the residual energy, the number of hops to the root, the number of packets to forward in each round, and the distance to the closest CH. The main goal of our proposed algorithm is to minimize the consumed energy of sensor nodes, and to elect an optimal number of CHs, according to the MDC's capacity.

The MDC's capacity limits the total length of the path. In the path construction phase, a genetic algorithm (GA) [16] is applied to calculate the shortest path passing through the CHs. And this can ensure to find a solution as faster as possible. During the CHs election phase, the algorithm calculates the shortest distance that passes through the elected CHs, then allows electing more and more of CHs if the obtained path is shorter than the MDC's capacity. The shortest path is calculated in each time a new CH is elected. These steps are repeated until the MDC's capacity is attained. And finally the MDC starts the data gathering phase beginning from the BS, and traveling through the elected CHs.

The rest of the paper is organized as follows: section 2 reviews related works. In Section 3 the network environment is presented and the problem formulation is introduced. TCAMDC is detailed in Section 4 and simulation results are discussed in Section 5. Finally, we conclude our work in section 6 giving an overview of our future work.

2. RELATED WORKS

In the last few years, there has been a grown interest in the clustering and the mobile data gathering techniques. Consequently, a several studies in the mobile data gathering and clustering algorithms have been proposed. To extend the network lifetime, the most two important challenges are dividing the network into clusters, and constructing the optimal path for the MDC.

Different approaches have been proposed to group sensor nodes into clusters. In [4] [6] and [7] authors chooses to create the clusters based on nodes density impact factor. Each sensor node will receives the impact factors from other sensor neighbors, then compares the accumulated impact factors and decides either to be elected or not as CH. The node with the highest accumulated impact factor will be selected as CH. The impact factors are calculated according to the distances between sensor nodes and the maximum transmission range of each sensor nodes.

A study [8] proposes a weighted rendez-vous planning (WRP) algorithm where each sensor calculates its weight to be elected as CH. The weight is calculated as follows:

$$W_i = NFD(i) \times H(i, M) \quad (1)$$

The WRP algorithm calculates the weight of a sensor node multiplying the number of packets to be forwarded $NFD(i)$ by its number of hops to the closest CH $H(i, M)$. A node with a highest weight has more chances to be elected as CH. However, the defined travelling path must not exceed the maximum tour length of the MDC which can lead to a shortest network lifetime.

Energy Aware Path Construction algorithm (EAPC) [11], uses the distance between the nodes, its neighbor's weight and the closest CH to evaluate the benefit of each sensor nodes. In this approach, authors use the following equation (2) to evaluate the node's benefit. $W(i)$ represent the weight of sensor node i , B is the subset of sensors which are passed by the new path that

contains the node i but did not belong the tree rooted by the node i and $\text{dist}(I, \text{closestCH})$ denote the distance separating the node I with the closest CH. The node with greatest benefit will be elected as CH. However, in both [11] and [19] sensor nodes with less energy could be elected as CH and this might reduce the lifetime of the WSN.

$$bi = w(i) + \sum_{j \in B} w(j) / \text{dist}(i, \text{closestCH}) \quad (2)$$

To reduce the energy consumption during the data forwarding of the sensor nodes, several studies propose the use a Mobile Data Collector. In [4], the MDC visits all sensor nodes deployed in the network, this conserves efficiently the energy of sensor nodes, however the long constructed path could not be supported by the MDC. A solution was proposed in previous works aiming to adopt multiple MDCs [12][13].

Previous researches propose to construct a static path for the MDC [4][8][11], where the path is computed at the BS, the MDC travels among sensor nodes based on the pre-computed trajectory constructed by the BS. This technique is more suitable for monitoring applications in which data are gathered and communicated back to the sink. In [4] a Hybrid Scheme of client/server and Mobile agent for data aggregation is proposed. Authors propose a weight based CH election scheme that selects the suitable sensor nodes as CHs. Then a Minimum Spanning Tree (MST) is constructed connecting the elected CHs to construct the shortest path for the MDC.

In the second approach, a dynamic path construction strategy is adopted [5][6][7]. The dynamic approach allows the MDC to decide the next destination during its travelling phase. In [5] authors propose a dynamic itinerary planning approach for Mobile agent. This technique employs the advantages of both Mobile agent and Mobile Server to prolong the network lifetime. The Mobile agent (or MDC) migrates randomly between nodes to gather the sensed data, after being dispatched from the Mobile Server and then return to it. This approach is not suitable in a large network because it is difficult for the MDC to visits all the deployed sensor nodes if the number of sensors increases. Due to node failure, authors in [6] and [7] propose to construct a dynamic path for the MDC, based on fault tolerant. This technique calculates alternative itineraries for the MDC in case of losing the dispatched Mobile Agent, if sensor nodes fail. The clustering method used in [6] and [7] consists on a weight based approach that selects the best sensor nodes as CHs according to the node's impact factor.

In this paper we propose a Tree Clustering Algorithm with Mobile Data Collector in Wireless Sensor Networks aiming to extend the network lifetime and to minimize the tour length of the Mobile Data Collector using the Genetic Algorithm. We focus on the problem of selecting the suitable CHs considering the residual energy of sensor nodes to guarantee the local data aggregation.

3. NETWORK ENVIRONMENT AND PROBLEM FORMULATION

3.1 Network environment

We consider a WSN that comprises a set of n static sensor nodes $N = \{n_1, n_2, n_3, \dots, n_n\}$ distributed over a given region, where the BS is situated in the border of the area. A mobile sensor node called a MDC travels a subset of chosen sensor nodes (CHs) beginning from the BS, to collect the gathered data. A MDC is a mobile node equipped with a rechargeable battery and has a sufficient storage capacity to buffer all sense data from the field. The data collection process is established in each round, considering that each node knows the coordinate of all other sensor nodes. Our objective is to select the best sensor nodes to be elected as CHs, and to determine the shortest path tour that will be traversed by the MDC, visiting each CH to collect sensed data. We assume that the sojourn time of the MDC at each CH is sufficient to gather all stored data. Sub-trees are formed during a network division, considering each tree as cluster rooted by one CH. Each sensor node transmits its packets along the edges of the tree rooted by one CH.

3.1. Problem formulation

Many recent researches have demonstrated the importance to introduce a MDC in WSNs. It extends the network lifetime by efficiently decreasing the energy consumption of sensor nodes. MDC is traveling between predefined positions to aggregate the data from a set of CHs by single-hop communication. The proposed algorithm chooses a set of CHs and then calculates the shortest path that MDC will travels. To attain our goal, and especially reducing the energy consumption we evaluate the best equation as to calculate the node's weight.

Considering a subset of sensor nodes N , the proposed algorithm constructs a MST, then partition the network into sub-trees, in which each CH gathers the data from its cluster's member. The energy dissipated by a sensor node i is calculated in equation 3. We assume that the BS is situated in the field's border. Let us consider that in iteration i , the algorithm chooses a subset of M CHs: $CH_1, CH_2, CH_3, \dots, CH_M$, each one considered as the root of M groups of clusters organized as sub-trees $T_1, T_2, T_3, \dots, T_M$, in iteration $i+1$, the proposed algorithm decides to elect one more CH or not, according to the obtained distance L (Equation 5) and the travelling path that the MDC could perform L_{max} . By this way, our proposed algorithm can guarantee an optimal number of CHs, and the data gathering process.

In each round, sensor nodes upload their packets to their corresponding CHs, along the routing tree. We assume that the number of sensor nodes in a cluster i is $ND(i)$. The number of packets received by CH_i is equal to $ND(i)-1$. CH_i has the job to upload $ND(i)$ packets to the MDC. In the data gathering phase, the MDC will passes through theses CHs subsequently, beginning from the BS.

The MDC moves with a constant speed v and has the task to visit all the CHs before returning to the BS. Considering a maximum allowed packet delay D , we define the maximum traveling path length that MDC could perform in Equation 4. The objective is to elect an optimal number M of CHs and to apply the GA in order to calculate the shortest path L passing through them. This approach guarantees that the MDC will support the total path length and data delivery.

$$E_i = ND(i) \times E_{Tx} + (ND(i) - 1) \times E_{Rx} \quad (3)$$

$$L_{max} = D \times v \quad (4)$$

$$L = \sum_{i=1}^M d_{i, (i+1) \bmod M} \leq L_{max} \quad (5)$$

4. THE PROPOSED TCMDC ALGORITHM

The main objective of our proposed algorithm is to minimize the quantity of energy consumed by sensor nodes. Selecting the CHs, and finding the shortest path for the MDC present two major steps that should be efficiently organized in order to extend the network lifetime. Ideally, to achieve these goals, in TCMDC algorithm we propose to solve these two problems jointly. We attempt to construct a Minimum Spanning Tree rooted at the BS, then to divide iteratively the MST into sub-trees and to construct the MDC's path with GA simultaneously. These steps are established in each round, which means that we obtain a new subset of CHs in each round and a new traveling path is established.

The BS is situated in the field's border, and each sub-tree has a CH as root. Figure 1 presents the following steps. As we can see, in Figure 1.a, the MST is constructed. In the first iteration, the first CH is selected and the edge connecting the new selected CH with its parent is automatically removed, as we can see in Figure 1.b. Then the MDC's path's length is calculated, at the first iteration's end, this length is equal to the distance between the BS and the

first elected CH. If this distance is smaller than the predefined maximum length path L_{max} , the operation is repeated and the selection of one more CH is allowed, until we attain the maximum length path L_{max} (Figure 1.c). The problem to find the best nodes depends essentially on the residual energy of sensor nodes. To select CHs, we propose a weight-based tree technique. Each sensor node calculates its weight and competes to be elected as CH. Thus, each CH evaluates its weight according to Equation 8. In this section we present the proposed TCMDC algorithm which is characterized by three principal phases: The Initialization phase, the CHs election and path construction phase, and the data gathering phase.

4.1. Initializing phase

We model the WSN as a graph $G(V,E)$, where V is the set of sensors and $E=\{e_1,e_2,e_3,\dots\}$ the set of edges connecting them. If the distance separating a node i to a node j is lower than the communication range R , the edge e_{ij} will be added to the set of edges in $G(V,E)$. The initialization phase takes as input $G(V,E)$ and applies the Prim's algorithm [18] to output a Minimum Spanning Tree rooted at the Base Station. The Prim's algorithm selects the shortest edges connecting the sensor nodes to build the tree. We construct a MST to ensure that the sensor nodes send data along the shortest distance, as to extend the network lifetime. After constructing the MST, the CHs election and path construction phase begins. As we can see, in Figure 1.a we see that all the sensor nodes are initially rooted by the BS.

Figure 1 Example of CHs election and clusters creation steps

4.1. CHs election and path construction phase

To select the suitable sensor nodes, our TCMDC algorithm propose a weight based tree clustering technique. It compares the node's weights (Equation 6) which are calculated considering four principal parameters: the residual energy of sensor nodes $E(i)$, $ND(i)$ is the number of forwarder packets by sensor node i , $HTR(i,CH_i)$ which represents the number of hops between the node i to its actual root CH_i and finally the $dist_to_closest_CH(i)$ which is the distance separating the node i with the closest CH. All these parameters have an important influence on the choice, on the other hand, WRP and EAPC algorithms only use $ND(i)$ and

HTR(i,Chi) as parameters to calculate the node's weight, but did not consider the residual energy as parameters in the weight's calculation. Therefore, to extend the network lifetime, our TCMDC algorithm allows to nodes with higher amount of residual energy to be elected as CHs. In addition, CHs consume more energy comparing with the other sensor nodes because they forward the sensed data of their cluster's member to the MDC. Sensor nodes with more data packets to forward are more eligible to be elected as CHs than other nodes. Therefore, our proposed algorithm can ensure a uniform energy consumption in the network. The parameters ND(i) and HTR(i,Chi) plays an important role in the node's weight, more the node has to forward packets (ND(i)), or is far away from his actual CH (HTR(i,Chi)), more it has the chance to be elected as Ch.

The algorithm presented below is performed in each round. We denote CH the subset of chosen CHs and S the non-chosen one, namely the cluster's members. In each round, the BS selects the best node depending on the weight, affects it to the subset CH, and calculates with GA a new path. In the first iteration, we assume that the BS plays the first chosen CH's role (line 1), Lt presents the actual path length obtained during simulation.

We assume that in the beginning, the initial energy is same for all sensor nodes and is equal to 1j. In the first iteration, the BS calculates the two parameters HTR(i,Chi) and ND(i) then generates the weight for each sensor node (line4-8). Therefore the best node nbest is selected as CH, added in the subset CH and finally removed from S (line10-12). We suppose that in iteration i the BS selected i sensor nodes to play the role of CHs. At this time, we have obtained i clusters each one rooted by one CH (Figure 1.c). The algorithm has attained a distance $L_t < L_{max}$, in this case the selection of a new CH is allowed because L_{max} is not yet attained (Line2). And once the (i+1)th CH is selected, the GA calculates the new shortest path L (Equation 4). Therefore, if the new Length L_t is shorter than L_{max} , the new selected node nbest (CH_{i+1}) is added to the CH subset and removed from S, in the other case, the CH election phase stops and the data gathering phase begins.

When the edge between the new CH nbest and its parent is removed, we obtain a new (i+1)th cluster rooted by the new (i+1)th CH. Sensor nodes belonging to the sub-tree rooted by the last selected CH must update their membership. For this reason, the parameters $dist_to_closest_CH(i)$, HTR(i) and ND(i) must be recalculated in each iteration. The GA calculates the shortest path L_t and finds the best sub sequence of CHs that the MDC will visit. The sub sequence is ordered to minimize the MDC's path length (see Figure 1.d,e and f). The path construction phase is applied simultaneously with the CH election phase presented above. Giving a subset of sensor nodes through which the MDC will pass, the GA finds the best node's order that minimizes the path length.

Assuming that in a specific iteration i, the cluster heads election and the construction path algorithm selects M CHs, the GA will form a routing path namely a chromosome (the sub sequence formed with the M sensor nodes). For example, in Figure 1.e the initial chromosome is {1;3;19;9;5} and in the next step the node 15 is added to the subset (Figure 1.f). The principal phases of a GA are initializing the population, fitness function, selection, cross-over and mutation [16].

$$W(i) = \frac{E(i) \times ND(i) \times HTR(i, CH_i)}{dist_to_closest_CH(i)} \quad (6)$$

$$L_r = dist(BS, CH_1) + \sum_{i=1}^r dist(CH_i, CH_{i+1}) + dist(CH_r, BS) \quad (7)$$

CHs election & path construction Algorithm

Input: L_{max} , $N=\{n_1, n_2, \dots, n_n\}$, $CH=\emptyset$, $S=N$, $L_t=0$;
Output: a set of ordered CHs $CH=\{CH_1, CH_2, \dots, CH_m\}$;

1. $CH=\{BS\}$;
 2. **While** $L_t < L_{max}$ **do**
 3. **for** each $n_i \in S$ **do**
 4. **evaluate** $E(i)$
 5. **evaluate** $HTR(i)$
 6. **evaluate** $ND(i)$
 7. $d(i) = \text{dist_to_closest_CH}(CH, i)$
 8. **evaluate** $W(i)$
 9. **End for**
 10. $n^{best} = \arg(\max(W(i)))$
 11. **call the GA** (reorder the CHs & recalculate the shortest path L_t)
 12. **If** ($L_t < L_{max}$)
 13. $S = S \setminus \{n^{best}\}$
 14. $CH = CH \cup \{n^{best}\}$
 15. remove edge between n^{best} and its parent
 16. **End if**
 16. **End While**
-

We assume that the first chosen chromosome is the subset of the sensor nodes previously ordered in the call of the GA at the previous iteration. The fitness function represents the quality of a chromosome. The main goal of TCMDC algorithm is to minimize the MDC's path length. In this case we define the total distance between the CHs noted L_t as fitness function that the GA will use to define the sequence of travelling path, in order to avoid the delay of data delivery.

Giving a subset of sensor nodes through which the MDC will pass, the GA finds the best node's order that minimizes the path length. Assuming that in a certain iteration i , the cluster heads election and the construction path algorithm selects M CHs, the GA will form a routing path namely a chromosome (the sub sequence formed with the M sensor nodes). For example, in Figure 1.f the first obtained chromosome is $\{1;3;19;9;5;15\}$, and in the next step the node 15 is added to the subset.(Figure 1.f).

The principal phases of a GA are: initializing the population, fitness function, selection, cross-over and mutation. We assume that the first chosen chromosome is the subset of the sensor nodes previously ordered in the call of the GA at the previous iteration.

Initializing the population: In the initialization phase, there is two ways to generate a chromosome, randomly or heuristically. In our TCMDC algorithm we apply the second way because the solution is faster. The first chosen chromosome is the subset of the sensor nodes previously ordered in the call of the GA at the previous iteration.

Fitness function: To find the best solution, in GA we must identify a fitness function, which represents the quality of a chromosome. The main goal of TCMDC algorithm is to minimize the MDC's path length. An effective chromosome has the highest fitness value. Therefore, the fitness function that efficiently provides a best solution is defined as follows in Equation 8, where M represents the size of a chromosome (the actual number of elected CHs).

$$f_M = 1 / \sum_{j=1}^M \text{dist}(j, j+1) \quad (8)$$

Selecting and cross over: The GA calculates the fitness value of each chromosome, and then applies the selection phase to pick-up the best one to be reproduced for the next iteration. Thus, this technique improves the average quality of a population by giving the chance to the GA to found a solution as faster as possible. The cross-over mechanism is applied on a selected pair of chromosomes in order to reproduce other chromosomes. For example (in Figure 2), we suppose that we select these two chromosomes: $C1 = \{1; 3; 19; 9; 5; 15\}$ and $C2 = \{1; 19; 3; 9; 15; 5\}$. The crossover generates two other chromosomes $C3$ and $C4$ as demonstrated in Figure 7.

Figure 2 The cross-over phase

Mutation: In the mutation phase, a node is selected randomly from the chromosome. The mutated chromosome represents an alternative route and is generated by combining each partial-chromosome.

5. SIMULATION RESULTS

The performances of our proposed algorithm are evaluated then compared with EAPC, RPCB and WRP which proposes the same approaches as CHEREDC. Matlab R2016a is used as simulator. The numbers of deployed nodes vary between 25 and 300, randomly deployed in an area of 200x200 m². We assume that the MDC has a rechargeable energy harvesting module such as solar power system and a speed equal to 2m/s as used in [9].

5.1. Balanced and Unbalanced deployment

Figure 3 The proposed TCMDC algorithm applied in two experimental scenarios

5.2. The average of the MDC's path obtained with BD and UD scenario

Figure 4 MDC's path length in BD and UD scenario (n=25,50,75)

Figure 4 shows the MDC's path's length obtained in such balanced deployment (BD) and Unbalanced Deployment (UD) (X Topology), using various number of sensor nodes: 25, 50 and 75. We realize that in the three cases ($N=25$, $N=50$ or $N=75$), we obtain a shorter path length with the UD scenario. This can be explained by the fact that in the UD scenario, sensor nodes are positioned in the same area so they are situated close to each other. Therefore, this leads to a shorter distance obtained between the elected CHs which affects the MDC's path length. We decided to choose the interval of round comprised between 200 and 220 because it represents the most stable one in terms of number of still alive sensor nodes as shown in Figure 4.

5.3. Network lifetime vs Number of sensor nodes

In WSNs, the scalability of routing protocols is a critical issue due to the high numbers of sensor nodes and the high node density. In this section, we consider the number of sensor nodes as scalability factor. A good routing protocol has to be scalable and adaptive to the changes in the network topology. In Figure 5, The number of sensor nodes varies between 50 and 200, we compare TCMDC algorithm with other achieved one, regarding to the network lifetime. As we can see, in BD and UD scenarios, we deduce that TCMDC outperforms the other algorithms and extends efficiently the network lifetime. This is because TCMDC algorithm considers the residual energy of sensor nodes and gives more chance to sensor nodes with higher amount of energy to be elected as CH. Therefore, the energy consumption of sensor nodes is reduced.

Figure 5 Network lifetime vs number of sensor nodes

5.4. Fairness index vs number of rounds

In WSNs, we define the fairness index as a metric used to determine whether or not the system resources are fairly shared between sensor nodes. The residual energy is the most important parameter that plays a major role to extend the network lifetime. In Figure 6 the fairness index of the proposed approach TCMDC is compared with the four others (CHEREDC, EAPC, WRP

and CB) in each round, in BD scenario (a) and UD scenario (b). In WRP, the defined travelling path did not consider the distance between a CH to another, this leads to a longer MDC's path length and a data overflow. In EAPC, the residual energy of sensor nodes is not considered in the weight calculation. Therefore the energy is not equitably consumed, so the residual energy of sensor nodes will deplete very quickly. Therefore, TCMDC outperforms the other schemes. We evaluate the fairness index according to Equation 9, where n presents the number of still alive sensor nodes and $E(i)$ the residual energy of sensor node i .

Figure 6. The fairness Index vs rounds

$$f_{index} = \frac{(\sum_{i=1}^n E(i))^2}{\sum_{i=1}^n (E(i)^2)} \quad (9)$$

5.5. Energy consumption

In Figure 6, we remark that in both BD and UD scenario the energy consumption per round increases as the number of sensor nodes increases. Our proposed algorithm presents better results than other algorithms for two principal reasons. The first one is because in the selection of CHs, the residual energy of sensor nodes is considered, contrary to EAPC, WRP and CB algorithms which chooses CH taking into consideration only the number of packets to be forwarded and the number of hops to the root. In addition, the MDC's path length plays an important role in the algorithm which gives more opportunity to elect more CHs. Equation 10 presents the energy consumption per round calculation (nbr is the total number of rounds).

$$\xi = \sum_{r=1}^{nbr} \left(\sum_{i=1}^N E(i) \right) / nbr \quad (10)$$

Figure 7 Energy consumption per round vs number of sensor nodes in BD scenario

5.6. Network Lifetime

In this section, we compare our proposed TCMDC with CHEREDC, EAPC, WRP and CB algorithms in terms of network lifetime. We use 260 sensor nodes deployed in a balanced topology. Figure 8 presents the lifetime extension of the proposed approaches versus rounds, and notices the performance of our TCMDC algorithm. This is due to the use of the MDC for the data gathering process which decreases efficiently the amount of packets forwarded by CHs, so the consumed energy is limited.

Figure 8 Network lifetime vs rounds

CONCLUSION

In this paper, a new Tree Clustering Mobile Data Collector algorithm is proposed, we construct a Minimum spanning and divide the population into sub-clusters. In this algorithm, CHs are elected, and the MDC's path is established iteratively. The proposed algorithm chooses nodes with highest weight as CHs and applies the GA to calculate the MDC's path. The way that the MDC visits the elected CH balances the energy consumption of sensor nodes. According to simulation results, our TCMDC algorithm gives better results than other achieved approaches in terms of network lifetime. The energy consumed by sensor nodes is considerably minimized, so the network lifetime is efficiently extended.

As future work, we plan to apply multiple MDCs to gather the sensed data in case of a large WSN and to divide the network into equal areas and assign a MDC in each one.

DECLARATIONS

FUNDING

This work was funded by Innov'Com Laboratory (Innovation of Communication and Cooperative Mobiles Laboratory), Created in 2011 under Code LR-11 / TIC-03, its former name is 6 'such research unit created in 2005 under Code UR / 11-05 / 05). The objectives of this Lab revolve around its contribution to the innovation of intelligent (cognitive), communicating, cooperative and mobile systems.

CONFLICTS OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AVAILABILITY OF DATA AND MATERIAL

The data supporting the conclusions of this paper will be made available by the authors.

CODE AVAILABILITY

Not Applicable

AUTHORS CONTRIBUTION

M-M and R-H developed the methodological proposal and data analysis, described the conclusions and references, realized the literature review and wrote the part of the theoretical frame. T-E read and approved the final manuscript.

REFERENCES

- [1] Meddah, M., Haddad, R., & Ezzedine, T. (2017, June). An energy efficient and density control clustering algorithm for wireless sensor network. In *Wireless Communications and Mobile Computing Conference (IWCMC)*, 2017 13th International (pp. 357-364). IEEE.
- [2] Meddah, M., Haddad, R., & Ezzedine, T. (2018, May). Residual Energy and Density Control Aware Cluster Head Election in Wireless Sensor Network. In *2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA)* (pp. 141-146). IEEE.
- [3] El Fissaoui, M., Beni-hssane, A., Ouhmad, S., & El Makkaoui, K. (2020). A Survey on Mobile Agent Itinerary Planning for Information Fusion in Wireless Sensor Networks. *Archives of Computational Methods in Engineering*, 1-12.
- [4] El Fissaoui, M., Beni-Hssane, A., & Saadi, M. (2016, November). Energy aware hybrid scheme of client-server and mobile agent models for data aggregation in wireless sensor networks. In *International Conference on Hybrid Intelligent Systems* (pp. 227-232). Springer, Cham
- [5] Dong, M., Ota, K., Yang, L. T., Chang, S., Zhu, H., & Zhou, Z. (2014). Mobile agent-based energy-aware and user-centric data collection in wireless sensor networks. *Computer networks*, 74, 58-70.
- [6] El Fissaoui, M., Beni-Hssane, A., & Saadi, M. (2019). Energy efficient and fault tolerant distributed algorithm for data aggregation in wireless sensor networks. *Journal of Ambient Intelligence and Humanized Computing*, 10(2), 569-578.
- [7] Fissaoui, M. E., Beni-Hssane, A., & Saadi, M. (2018). Multi-mobile agent itinerary planning-based energy and fault aware data aggregation in wireless sensor networks [J]. *EURASIP Journal on Wireless Communications and Networking*, 2018(1), 92.
- [8] G. Xing, T. Wang, Z. Xie, and W. Jia, "Rendezvous planning in wireless sensor networks with mobile elements," *IEEE Trans. Mobile Comput.*, vol. 7, no. 12, pp. 1430–1443, Dec. 2008.
- [9] Tashtarian, F., Moghaddam, M. H. Y., Sohraby, K., & Effati, S. (2015). On maximizing the lifetime of wireless sensor networks in event-driven applications with mobile sinks. *IEEE Transactions on Vehicular Technology*, 64(7), 3177-3189.
- [10] Salarian, H., Chin, K. W., & Naghdy, F. (2014). An energy-efficient mobile-sink path selection strategy for wireless sensor networks. *IEEE Transactions on vehicular technology*, 63(5), 2407-2419.
- [11] Wen, W., Zhao, S., Shang, C., & Chang, C. Y. (2018). EAPC: Energy-aware path construction for data collection using mobile sink in wireless sensor networks. *IEEE Sensors Journal*, 18(2), 890-901.

- [12] Abdolkarimi, M., Adabi, S., & Sharifi, A. (2018). A new multi-objective distributed fuzzy clustering algorithm for wireless sensor networks with mobile gateways. *AEU-International Journal of Electronics and Communications*, 89, 92-104.
- [13] Zhang, C., & Fei, S. (2020). A Matching Game-Based Data Collection Algorithm with Mobile Collectors. *Sensors*, 20(5), 1398.
- [14] Rao, X., Huang, H., Tang, J., & Zhao, H. (2016). Residual energy aware mobile data gathering in wireless sensor networks. *Telecommunication Systems*, 62(1), 31-41.
- [15] Ghosh, N., Banerjee, I., & Sherratt, R. S. (2017). On-demand fuzzy clustering and ant-colony optimisation based mobile data collection in wireless sensor network. *Wireless Networks*, 1-17.
- [16] Sert, S. A., Bagci, H., & Yazici, A. (2015). MOFCA: Multi-objective fuzzy clustering algorithm for wireless sensor networks. *Applied Soft Computing*, 30, 151-165.
- [17] Mehr, M. (2014). Cluster head election using imperialist competitive algorithm (chei) for wireless sensor networks. *Int J Mobile Netw Commun Telemat*, 4(3), 1-9.
- [18] Zhu, C., Wu, S., Han, G., Shu, L., & Wu, H. (2015). A tree-clusterbased data-gathering algorithm for industrial WSNs with a mobile sink. *IEEE Access*, 3, 381-396.
- [19] Younes, A., Badawi, U. A., Farag, T. H., Alghamdi, F. A., & Salah, A. B. (2018). A Genetic Algorithm to Find the Minimum Cost Paths Tree with Bandwidth Constraint in the Computer Networks. *International Journal of Applied Engineering Research*, 13(10), 7472-7476.

Figures

Figure 1

Example of CHs election and clusters creation steps

Figure 2

The cross-over phase

Figure 3

The proposed TCMDC algorithm applied in two experimental scenarios

Figure 4

MDC's path length in BD and UD scenario ($n=25,50,75$)

(a) BD Scenario

(b) UD Scenario

Figure 5

Network lifetime vs number of sensor nodes

(a) BD Scenario

(b) UD Scenario

Figure 6

The fairness Index vs rounds

Figure 7

Energy consumption per round vs number of sensor nodes in BD scenario

Figure 8

Network lifetime vs rounds