HGC: HyperGraph based Clustering scheme for power aware wireless sensor networks

Jocelyn Edinio Zacko Gbadouissa[a,c], Ado Adamou Abba Ari[a,b,*], Chafiq Titouna[d], Abdelhak Mourad Gueroui[b], Ousmane Thiare[e]

[a] LaRI Lab, University of Maroua, P.O. Box 814 Maroua, Cameroon
[b] LI-PaRAD Lab, Université Paris Saclay, University of Versailles Saint-Quentin-en-Yvelines, 45 Avenue États-Unis 78035 Versailles cedex, France
[c] African Institute for Mathematical Sciences (AIMS-Cameroon), P.O. Box 608, Limbé, Cameroon
[d] LIPADE Lab, University Paris Descartes, 45 Rue des Saints Pères, 75006 Paris, France
[e] Department of Computer Science, Gaston Berger University of Saint-Louis, P.O. Box 234 Saint-Louis, Senegal

Abstract

Due to the energy constraints of sensors owing to the limitation of their built-in batteries, the lifespan of Wireless Sensor Networks (WSNs) are significantly affected. These particular ad-hoc networks have a huge number of applications including surveillance and target tracking. Unfortunately, since sensor nodes are limited in terms of power resources, efficient utilization of these resources is an important goal to design power-aware WSNs. This led researchers to propose numerous methods, such as clustered WSNs, in order to effectively manage the power resources. In this work, we proposed a heuristic clustering based on the hypergraph theory, and called HyperGraph Clustering (HGC) that aims at optimizing the energy of sensor nodes. Theoretical evaluation highlighted that this clustering protocol consumed less energy during the cluster formation phase and the selection of the cluster head. In addition, we evaluated the performance of the proposed HGC and the results showed the effectiveness of our scheme to those we compared in terms of the number of nodes alive, residual energy and the total consumption of the network.

Keywords: Clustering, Wireless Sensor Networks, Hypergraph modelling, Simulation.

1. Introduction

1.1. Background

With the recent advances of Micro Electro Mechanical Systems, especially in semiconductors' technologies that led to the design of smart sensors, networking, and material science technologies have enabled a significant shift in Wireless Sensor Network (WSN) [1, 2]. The new wave of research in WSNs has been attracting more attention, especially in networking techniques suitable for highly dynamic ad hoc environments. Therefore resource-constrained sensor nodes have been the focus [3].

Furthermore, a WSN is a paradigm of Internet of Things (IoT) that comprises tens to thousands of battery power-limited devices called wireless sensors, spatially dispersed over an environment and which are interconnected through a wireless medium. Indeed, a sensor node is an electronic device composed of a sensing unit for detecting events or changes in its environment, a processing unit for processing the collected information, a communication unit for transmitting information and a power supply unit [4]. These features are also driving the ubiquitous deployment of large-scale WSNs [5].

Moreover, due to their convenient deployment and self-organization features, WSNs have been designed and implemented for several applications such as environmental or habitat tracking, home automation, healthcare, medical systems, traffic management, and smart battlefields [6]. Furthermore, despite its large applications, WSN faces several problems including security, fault tolerance, data redundancy, data transmission, and energy efficiency [7]. The redundancy is generally due to the random, and dense deployment of sensor nodes, which may generate redundant data while fault-tolerant and data transmission mechanisms aim at improving the lifetime of WSN [8]. In addition, the need to design a suitable security mechanism for WSNs aims at addressing the major attacks on WSNs such as hijacking, eavesdropping, rushing and disruption [4, 9]. While fault-tolerant mechanisms are required in a case where due to harsh conditions, some nodes can discontinue operation given to the malfunctioning and physical damage that lead to the loss of important data.

However, according to the generally harsh environment of the deployment of some WSNs' applications, it is not possible to change a sensor battery. Indeed, in a given WSN, the gathered information on sensors is processed and transmitted to one or more Base Stations (BS) and eventually remote locations for analysis purpose. Thus, sensors consume energy while processing and transmitting the gathered data to another sensor or to the BS; and also when receiving data from a neighbor sensor. In other words, in WSNs, communication and network management protocols consume more energy [10]. Therefore, energy utilization of the limited power supply of sensors, which depends on the way traffic flows from sensors to BS as well as the choice of hopping between sensors, is the most important factor in the design and deployment of WSNs. Thus, avoiding sensor power wastage while gathering data from a WSN becomes the key challenge responsible for enhancing the network lifetime.

*Corresponding author
Email address: adoadamou.abbaari@gmail.com (Ado Adamou Abba Ari)

As one of the most important energy control mechanism is WSN, clustering has significantly improved the performances of sensor networks in numerous applications that enable data dissemination and aggregation. According to the successful topology design and control technique as well as data dissemination and aggregation allowed by clustering, a number of researches adopted this technique in order to significantly improve performances in numerous applications that require gathering data from WSNs. Concretely, in a clustered WSN, sensors are separated into a number of groups called clusters [11, 12] in which data gathering and transferring occurs. Each cluster has a number of cluster members, i.e., sensors and a cluster leader called Cluster Head (CH). It is responsible for collecting gathered data from its members in an intra-cluster communication way and can cooperate with other CHs to report data to a BS by adopting an inter-cluster communication. However, since the CHs have in the most cases the same features with the other cluster members, selecting CHs appropriately is needed to conserve the energy and therefore prolong the network lifetime. Indeed, unbalanced energy consumption on CHs may cause a quick death of nodes and thus cut down the lifetime of the network [13]. This work aims at addressing the life time maximization problem by tackling the network traffic and nodes' energy consumption issues.

1.2. Author's contribution

The purpose of this paper is to address the problem of the WSN lifetime by adopting the clustering. We formulated the k-clustering as a k-partitioning graph theory problem. We proposed a clustering scheme called HGC based on hypergraph concepts. We evaluated the performance and compared it with the Particle Swarm Optimization Clustering (PSO-C) [14], Energy-Balanced Unequal Clustering (EBUC) [15], and the particle swarm optimization based unequal and fault-tolerant clustering protocol (PSO-UFC) [16]. In brief, our contributions can be summarized as follows:

- Hypergraph theory formulation of the k-clustering problem.
- Proposition of a heuristic to partition nodes into clusters.
- CH selection that guarantees the balance of the loads and energy.
- Definition of a multi-hop routing tree that minimized the packet loss ratio.
- Simulation and comparison of the proposed scheme.

1.3. Organization of the paper

The rest of the paper is organized as follows: Section 2 surveys the schemes proposed in the literature to improve the WSN lifetime; we describe the key concepts necessary to the definition of our scheme in Section 3. Section 4 presents the proposed clustering scheme, then we evaluate its performance in Section 5. Finally, we conclude the work in Section 6.

2. Related work

Several routing protocols have been proposed and can be classified in terms of location, mobility, Quality of Service (QoS), hierarchy, etc. In location-based protocols, the location of nodes in the WSN is used for computing distance between nodes for data transmission. Moreover, there are single-path routing protocols that use only one path for data transmission, while multipath schemes define several paths [17]. Otherwise, Hierarchical routing protocols gather nodes into groups called clusters and select a CH as a leader in each group. CHs send data received from nodes to the BS [11].

Besides, the mobility-based protocols address the problem of higher energy consumption by nodes close to the sink, by assuming a time to time mobility (in change of position) of the sink [7]. Wang et al. [18] proposed a clustering scheme based on Particle Swarm Optimization (PSO) for WSN with a mobile sink node. The scheme uses PSO to divide the network into regions, then selects a CH from each region based on the position and residual energy of the node. In the same order of idea, Saranya et al. [19] proposed a clustering algorithm that aims at extending the lifetime of the network and increasing the packet delivered to the mobile sink in the network. Their CH selection is based on criteria such as residual energy of the node, distance, and the data overhead.

Futhermore, hierarchical routing protocols for WSN attracted much attention in recent decades. Low Energy Adaptive Clustering Hierarchy (LEACH) is the first routing protocol for clustering sensor nodes [20]. It selects the CH periodically and circularly to fairly distribute the CH task to each node [20, 21, 22]. Moreover, the CH selection is based on a threshold parameter $T(n)$ defined below in Eq. 1. However, the random selection of the CH leads to the network energy imbalance [11, 18].

$$T(n) = \begin{cases} \dfrac{p}{1 - p \cdot \left(r \mod \frac{1}{p}\right)} & \text{if } n \in G \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Where p is the percentage of CH, r the current round, G the set of nodes which have not become CH in the last $\dfrac{1}{p}$ rounds.

To overcome the network energy imbalance problem, Zang et al. [20] as well as Ari et al. [11], created first a set of sensor node candidates to initial cluster centers based on a threshold parameter. Then, initial cluster centers are selected in such a way that they are furthest away from each other. Also, Ray and Debashis [23] proposed a method called the midpoint algorithm for initial cluster centers selection. For every point, this algorithm finds the distance from the origin, then partitions the points (sorted based on distances) into k groups. The k initial centers are the middle points of each group.

This initial CHs selection affects the results of clustering algorithms like k-means. Thus, in order to improve the results of k-means, some researchers used the seeding technique. Particularly in the k-means++ [24], where the randomized seeding technique is used as an initialization method for the standard k-means. To select the k initial CHs, a center is first chosen in a random uniform way; then, the $k-1$ remaining CHs are selected with the probability $\frac{d(x)^2}{\sum_x d(x)^2}$; where $d(x)$ is the shortest distance from a data point to the closest center already chosen. In this light, Gbadoubissa et al. [25] proposed an initialization scheme based on the geometry of a circle for k-means. The authors consider all the points to be lying in a wide circle, then they select the center of the circle as the first CH. The remaining CHs are picked such that $c_j = (\delta \cos\theta_j,\ \delta \sin\theta_j)$, with $\delta < \frac{r}{2},\ \theta = \frac{2\pi}{k-1},\ (j-1)\cdot\theta \le \theta_j < j\cdot\theta,\quad j = 1,\cdots,k-1.$

Moreover, clustering algorithms based on theoretical mathematics such as graph theory, color, probability, game, and wheel theory for WSNs are widely proposed. In graph theory, the occurence of failures, communication on modalities, and network interference are under control in the cluster network. Consider a unit disk graph (UDG); it is a graph formed from a collection of equal-radius circles, in which two circles are connected by an edge if one circle contains the centre of one other circle. Most researchers use UDG as a network topology; however the use of UDG is avoided when in practice nodes adjust their transmission ranges according to real-time applications. To avoid this issue, Saravanan et al. [26] proposed a distributed semigraph contiguous prevalent set (SCPS) algorithm for the formation of backbone nodes.

Yan et al. [27] proposed two scheduling schemes based on hypergraph theory, named dedicated scheduling and shared scheduling, to maximize the reliability and minimize the transmission delay of packets (divergent objective functions) in a wireless sensing and control network. The dedicated scheduling protocol sets the time slots for each hypernode along the path to the destination, while the shared scheduling allows a hypernode to transmit a packet using the remaining scheduled time slots. Furthermore, hypergraph theory is also used to tackle the interference management issue in device to device (D2D) communications. In [28], Zhang et al. proposed a hypergraph based resource allocation protocol that considers cellular user equipment and D2D pairs as vertices, the interference relation as a hyperedge, and channels as the clusters in the hypergraph. For efficient interference coordination, the hypergraph is partitioned into clusters corresponding to channels.

Chen et al. [29] proposed an interference-free clustering protocol for large-scale WSNs. The sensor nodes use the Time Division Multiple Access (TDMA) protocol to reduce the interference. Indeed, TDMA enables sensor nodes to share the same frequency channel by dividing the signal into different time slots [30]. However, Managing the time slots is a tough task and it is tougher in case of heterogeneous time slots [29, 30]. The protocol proposed by Chen et al. [29] partitions the network into clusters such that the communication within a cluster does not interfere with the communications of others. In [30], Raza et al. tried to tackle the interference issue in industrial WSNs. They designed a scheduling algorithm that takes into account varying timeslots for different nodes of a cluster. The model transforms the time slot to form a deterministic schedule.

The works done by authors in[31, 32, 33] focused on analysis of packet delivery performance in WSN. Their investigations may help in the design of effective multi-objective clustering schemes. While some schemes consider only special attributes of sensor nodes such as remaining energy, distance to BS, etc., to select CHs, authors in [34] proposed a two layer routing scheme which considers cluster load, remaining energy, distance to BS and the degree of neighboring sensor nodes as criteria to select CHs. They proposed a method based entropy to predict the remaining energy of each node at the end of the next round. In the second layer, they proposed a new algorithm to form the routing tree as backbone of the network. In addition, they used a hybrid compressive sensing to compress the aggregated data.

3. Preliminaries

3.1. Background on hypergraph

Definition 1. *Let us denote by* $V = \{v_1, v_2, \cdots, v_n\}$ *the set of vertices, and let* $H = \{h_1, h_2, \cdots, h_m\}$ *be a family of subsets of* V *like defined in Eq. 2.*

$$\bigcup_{i=1}^{m} h_i = V, \ \ and \ \ h_i \neq \emptyset, \ \forall i \in [1, m] \cap \mathbb{N} \tag{2}$$

Thus, the tuple $G = (V, H)$ *is called hypergraph with vertex set* V *and hyperedge* H.

Definition 2. *Let* $G = (V, H)$ *be a hypergraph. The weight matrix of* G *is a matrix* $W(G)$ *with rows and columns representing the vertices of* G *(see Eq. 3).*

$$W(v_i, v_j) = w(e_{ij}) \tag{3}$$

Where e_{ij} *denotes the edge between nodes* v_i *and* v_j. *We can also write it as* $w(e_{ij}) = w_{ij}$.

Proposition 1. *Let* V *be a vertex set, and* H *a hyperedge set. We denote by* $H(v),\ v \in V$ *the set of all hyperedges which contain* v. $|H(v)|$ *is called the degree of the vertex* v.

Proposition 2. *A hypergraph* $G = (V, H)$ *is said k-regular, if each vertex* $v \in V$ *has the same degree* $k > 0$*; but it is said r-uniform when each hyperedge* $h \in H$ *has the same cardinality* $r > 0$.

3.2. Operation on hypergraph: weak deletion

Let $G = (V, H)$ be a hypergraph. A *weak deletion* of $v \in V$ from G is to remove v from V and from each hyperedge in $H(v)$. The weak deletion of a vertex involves the creation of a new hypergraph $G_1 = (V_1, H_1)$, with $V_1 = V \setminus \{v\}$, and H_1 the new hyperedge set.

$$W(G) = \begin{array}{cc} & \begin{array}{ccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 \end{array} \\ \begin{array}{c} v_1\\ v_2\\ v_3\\ v_4\\ v_5\\ v_6\\ v_7\\ v_8\\ v_9 \end{array} & \begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} & w_{15} & w_{16} & w_{17} & w_{18} & w_{19} \\ w_{21} & w_{22} & w_{23} & w_{24} & w_{25} & w_{26} & w_{27} & w_{28} & w_{29} \\ w_{31} & w_{32} & w_{33} & w_{34} & w_{35} & w_{36} & w_{37} & w_{38} & w_{39} \\ w_{41} & w_{42} & w_{43} & w_{44} & w_{45} & w_{46} & w_{47} & w_{48} & w_{49} \\ w_{51} & w_{52} & w_{53} & w_{54} & w_{55} & w_{56} & w_{57} & w_{58} & w_{59} \\ w_{61} & w_{62} & w_{63} & w_{64} & w_{65} & w_{66} & w_{67} & w_{68} & w_{69} \\ w_{71} & w_{72} & w_{73} & w_{74} & w_{75} & w_{76} & w_{77} & w_{78} & w_{79} \\ w_{81} & w_{82} & w_{83} & w_{84} & w_{85} & w_{86} & w_{87} & w_{88} & w_{89} \\ w_{91} & w_{92} & w_{93} & w_{94} & w_{95} & w_{96} & w_{97} & w_{98} & w_{99} \end{pmatrix} \end{array} \Rightarrow U(G) = \begin{array}{cc} & \begin{array}{ccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 \end{array} \\ \begin{array}{c} v_1\\ v_2\\ v_3\\ v_4\\ v_5\\ v_6\\ v_7\\ v_8\\ v_9 \end{array} & \begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} & w_{15} & w_{16} & w_{17} & w_{18} & w_{19} \\ 0 & w_{22} & w_{23} & w_{24} & w_{25} & w_{26} & w_{27} & w_{28} & w_{29} \\ 0 & 0 & w_{33} & w_{34} & w_{35} & w_{36} & w_{37} & w_{38} & w_{39} \\ 0 & 0 & 0 & w_{44} & w_{45} & w_{46} & w_{47} & w_{48} & w_{49} \\ 0 & 0 & 0 & 0 & w_{55} & w_{56} & w_{57} & w_{58} & w_{59} \\ 0 & 0 & 0 & 0 & 0 & w_{66} & w_{67} & w_{68} & w_{69} \\ 0 & 0 & 0 & 0 & 0 & 0 & w_{77} & w_{78} & w_{79} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{88} & w_{89} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{99} \end{pmatrix} \end{array}$$

$$\Rightarrow U_s(G) = \begin{array}{cc} & \begin{array}{ccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 \end{array} \\ \begin{array}{c} v_1\\ v_2\\ v_3\\ v_4\\ v_5\\ v_6\\ v_7\\ v_8\\ v_9 \end{array} & \begin{pmatrix} 0 & w_{12} & w_{13} & w_{14} & w_{15} & w_{16} & w_{17} & w_{18} & w_{19} \\ 0 & 0 & w_{23} & w_{24} & w_{25} & w_{26} & w_{27} & w_{28} & w_{29} \\ 0 & 0 & 0 & w_{34} & w_{35} & w_{36} & w_{37} & w_{38} & w_{39} \\ 0 & 0 & 0 & 0 & w_{45} & w_{46} & w_{47} & w_{48} & w_{49} \\ 0 & 0 & 0 & 0 & 0 & w_{56} & w_{57} & w_{58} & w_{59} \\ 0 & 0 & 0 & 0 & 0 & 0 & w_{67} & w_{68} & w_{69} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{78} & w_{79} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{89} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Figure 1: Initialisation phase

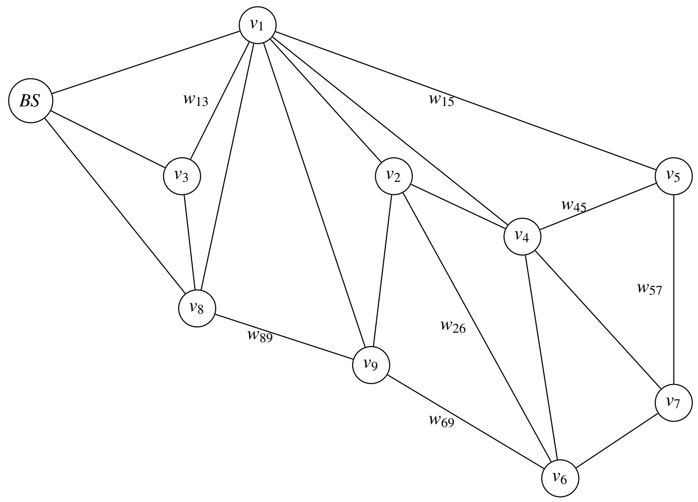


Figure 2: Graph representation of the WSN

3.3. Hypergraph partitioning problem

Given a hypergraph $G = \left(V, H\right)$, the *k−way partitioning* of G is to find a partition $P = \left\{h_1, h_2, \cdots, h_k\right\}$ which divides V into k subset such that:

- $\bigcup\limits_{i=1}^{k} h_i = V$, with $h_i \neq \emptyset$, $i = 1 \cdots k$
- $\bigcap\limits_{i=1}^{k} h_i = \emptyset$
- The cost function $f : H \to \mathbb{R}$ is minimized

4. Proposed protocol

In this section, we propose a clustering algorithm for WSN based on hypergraph theory concepts.

4.1. Network model and assumptions

Our wireless sensor network model consists of a BS and n sensor nodes, randomly deployed in a geographical area. Sensor nodes are static after deployment and the geographical coordinates of sensors are known. Moreover, the distances between sensor nodes and the BS can be estimated via the received signal strength or geographical coordinates of nodes. Each member node periodically performs sensing operations and transmits data to its own CH.

4.2. Design goals

Considering our sensor nodes as vertices and the WSN as a weighted hypergraph G, our first goal is to solve the *k-way partitioning* problem stated in Section 3.3, with the objective function defined in Eq. 5. Then, we design a scheme that selects CHs in such a way that guarantees the balance of loads and energy. The final goal is to define a multi-hop routing tree such that the packet loss ratio is minimized and the network lifetime is improved.

4.3. Overview of the proposed scheme

Let us assume that the number of clusters k is given, and $k < n$. Then, we denote by M and N the sets of marked vertices and unmarked vertices respectively, such that $V = M \cup N$. At the initial step $V = N$, $M = \emptyset$. Thus to generate a hyperedge, we randomly select a vertex (sensor node) v_i. Then, we add every vertex which is such that its inter-node distance with the first vertex is lower than the RSSI of v_i. The size of each hyperedge is equal or lower than $\lfloor \frac{n}{k} \rfloor$. Moreover, after the death of a sensor node, our algorithm performs automatically a weak deletion of this node. In other words, we apply the cluster formation algorithm on the remaining nodes. The pseudocode of this proposed scheme is given in Algorithm 2, and its flow chart is shown in Figure 4a.

4.4. Clustering mechanism

4.4.1. Initialization phase

Considering the WSN as a complete graph $G = (V, H)$ (see Figure 2), then

- *Step 1* : Generate the weight matrix $W(G)$ (see Figure 1. $W(G)$ is a square matrix of rank n.
- *Step 2* : Transform $W(G)$ to an upper triangular matrix $U(G)$ by removing every element under the diagonal, that is $w_{ji} = 0$. Since $w_{ij} = w_{ji},\ \forall i, j \in \{1, \cdots, n\},\ i \neq j$.
- *Step 3* : Transform the matrix $U(G)$ to a strictly upper triangular matrix $U_s(G)$ by replacing the diagonal by nil ($w_{ii} = 0,\ i = 1, \cdots, n$).

4.4.2. Cluster generation

The cluster generation is given in Algorithm 1 and the corresponding flowchart is presented in Figure 4b. For illustration purpose, Figure 3 illustrates the functioning of this algorithm on 10 nodes.

4.4.3. Setup phase (or clusters formation)

The cluster setup phase is given in Algorithm 2. The variable used in this algorithm is defined hereinafter.

- $n_i = |h_i|$ denotes the number of vertex within the hyperedge h_i.
- w_{ij} is the weight of the inter-node distance between v_i and v_j.
- s_i is the signal strength indicator of node v_i.
- f_i is the arithmetic mean of inter-node distances (see Eq. 4).

$$f_i = \frac{1}{\binom{n_i}{2}} \sum_{v_j \in h_i,\ v_j \neq v_i} w_{ij} \tag{4}$$

Algorithm 1: Cluster_generation($U_s(G)$, v_r)

input : $U_s(G)$ strictly upper triangular matrix of G, a vertex v_r
output: cluster h_i, and a reduced stricty upper triangular matrix $U_s(G)$

From $U_s(G)$, merge corresponding column and row of v_r; that is, merge the non-nill elements w_{rj} and w_{ir} to a list $Temp_r$.
Select the $\lfloor \frac{n}{k} \rfloor$ shortest elements of $Temp_r$, and add their corresponding vertices to the hyperedge h_i; i.e $h_i = \{v_r, \cdots, v_j, \cdots\},\ j \neq l,\ j \in \{1, \cdots, n\}$.
From $U_s(G)$, delete columns and rows corresponding to elements of h_i
Return h_i and the reduced $U_s(G)$

- $f(H)$ is the objective function defined in Eq. 5.

$$f(H) = \sum_i f_i \tag{5}$$

Algorithm 2: Cluster formation scheme

input : $V = M \cup N$ set of n vertices $v \in \mathbb{R}^d$ and an integer $k \geq 2$
output: clusters $h_1, h_2, \cdots, h_k, \cdots$

Initialization phase
Select the closest vertex v_c to the BS
Call *Cluster_generation($U_s(G)$, v_c)*
repeat
  for $i = 2, \cdots, k$ **do**
    Select randomly a vertex v_i from N
    Call *Cluster_generation($U_s(G)$, v_i)* ;
    Add the remaining vertices to the closest clusters
until *The objective function $f(H)$ is minimized*;

4.5. CH selection process

Let us denote by b_{t_i} the estimated residual battery lifetime of node i, and p_{l_i} the packet loss ratio of node i. Then the coefficient defined in Eq. 6 represents the likelihood of node i to be a CH (ch).

$$ch_{r_i} = \frac{1}{100} \cdot \frac{b_{t_i}}{p_{l_i}} \tag{6}$$

4.5.1. Proposition

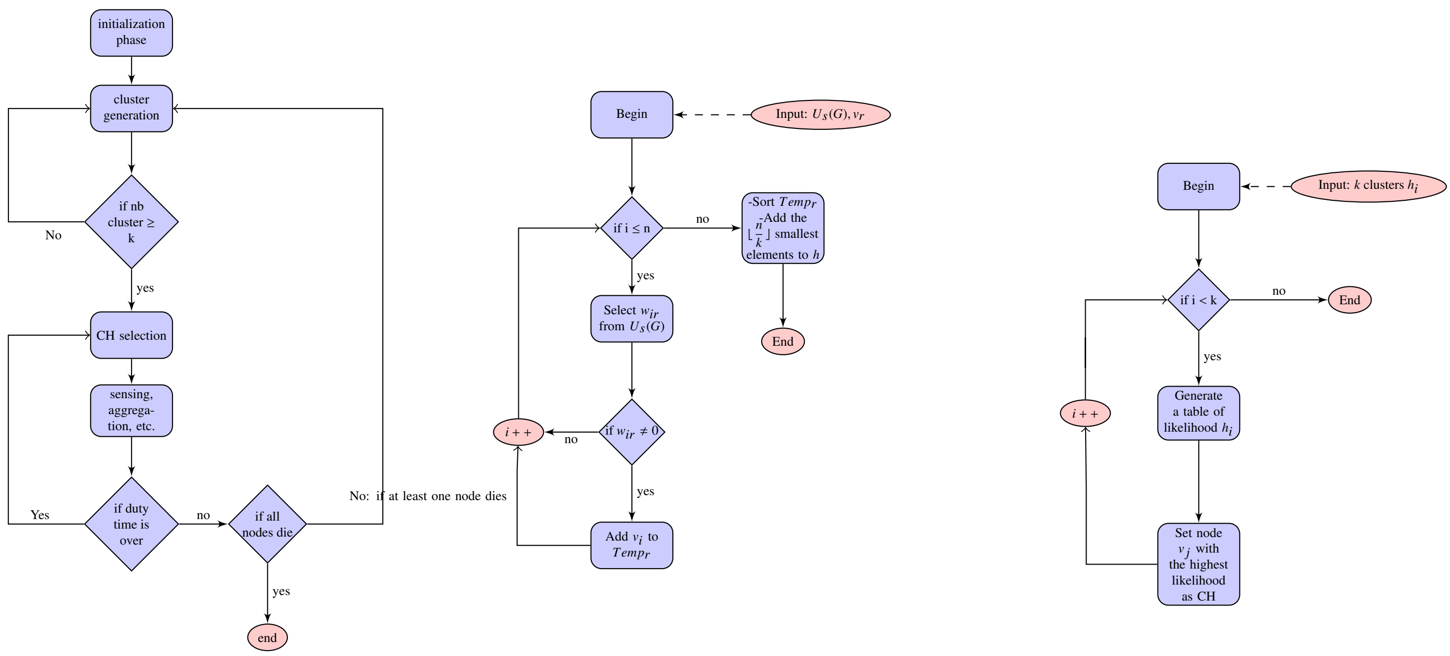
1. Let h_i be a hyperedge. We denote by T_i given in Table 1, the likelihood to be CH of nodes of h_i.

$$U_s(G) = \begin{array}{c|ccccccccc} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 \\ \hline v_1 & 0 & w_{12} & w_{13} & w_{14} & w_{15} & w_{16} & w_{17} & w_{18} & w_{19} \\ v_2 & 0 & 0 & w_{23} & w_{24} & w_{25} & w_{26} & w_{27} & w_{28} & w_{29} \\ v_3 & 0 & 0 & 0 & w_{34} & w_{35} & w_{36} & w_{37} & w_{38} & w_{39} \\ v_4 & 0 & 0 & 0 & 0 & w_{45} & w_{46} & w_{47} & w_{48} & w_{49} \\ v_5 & 0 & 0 & 0 & 0 & 0 & w_{56} & w_{57} & w_{58} & w_{59} \\ v_6 & 0 & 0 & 0 & 0 & 0 & 0 & w_{67} & w_{68} & w_{69} \\ v_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{78} & w_{79} \\ v_8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{89} \\ v_9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ v_{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \Rightarrow \quad Temp_3 = \boxed{w_{13} \mid w_{23} \mid w_{34} \mid w_{35} \mid w_{36} \mid w_{37} \mid w_{38} \mid w_{39}} \Rightarrow h_1 = \{v_1, v_3, v_8\}$$

$$U_s(G) = \begin{array}{c|cccccc} & v_2 & v_4 & v_5 & v_6 & v_7 & v_9 \\ \hline v_2 & 0 & w_{24} & w_{25} & w_{26} & w_{27} & w_{29} \\ v_4 & 0 & 0 & w_{45} & w_{46} & w_{47} & w_{49} \\ v_5 & 0 & 0 & 0 & w_{56} & w_{57} & w_{59} \\ v_6 & 0 & 0 & 0 & 0 & w_{67} & w_{69} \\ v_7 & 0 & 0 & 0 & 0 & 0 & w_{79} \\ v_9 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \Rightarrow \quad Temp_2 = \boxed{w_{24} \mid w_{25} \mid w_{26} \mid w_{27} \mid w_{29}} \Rightarrow h_2 = \{v_2, v_4, v_9\}$$

$$U_s(G) = \begin{array}{c|ccc} & v_5 & v_6 & v_7 \\ \hline v_5 & 0 & w_{56} & w_{57} \\ v_6 & 0 & 0 & w_{67} \\ v_7 & 0 & 0 & 0 \end{array} \Rightarrow \quad Temp_8 = \boxed{w_{48} \mid w_{68} \mid w_{78}} \Rightarrow h_3 = \{v_8, v_7, v_6\}$$

Figure 3: Illustration of the cluster generation phase with $n = 10$, and $k = 3$



(a) HGC

(b) Cluster generation

(c) CH selection

Figure 4: Flowcharts

Table 1: T_i: Likelihood to be CH

node	i_0	i_1	$\cdots$	$i_{\lfloor \frac{n}{k} \rfloor -1}$
likelihood	$ch_{r_{i_0}}$	$ch_{r_{i_1}}$	$\cdots$	$ch_{r_{i_{\lfloor \frac{n}{k} \rfloor -1}}}$

2. After each duty time τ, the tables of likelihood of hyperedges are updated.

The CH selection algorithm is given in Algorithm 3.

Algorithm 3: CH selection

input : h_i sets of n_i vertices $v \in \mathbb{R}^d$
output: CHs ch_i

foreach *hyperegde* h_i **do**
Generate a table of likelihood T_i
Select a node v_j which has the highest $T_i[j]$ value as CH
Set the duty time (or CH duty time) τ
Restart the process at the end of the duty time τ ;

5. Performance evaluation

5.1. Theoretical analysis of energy consumption

We used the same model as the one adopted in [16]. In this model, the total energy consumed by a node to transmit an s_p-bit packet over a distance d is given by Eq. 7, while the amount of energy consumed to receive an s_p-bit packet is given by Eq. 8.

$$E_{TX}(s_p, d) = \begin{cases} s_p \cdot E_{diss} + s_p \cdot \epsilon_{fs} \cdot d^2, & if\ d < d_0 \\ s_p \cdot E_{diss} + s_p \cdot \epsilon_{mp} \cdot d^4, & if\ d \geq d_0 \end{cases} \tag{7}$$

$$E_{RX}(s_p) = s_p \cdot E_{diss} \tag{8}$$

where:

- E_{diss} = amount of energy dissipated in electronics circuit
- ϵ_{fs} = free space model
- ϵ_{mp} = multipath fading model
- $d_0 = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}}$

Recall the network model, where information such as geographical coordinates, energy level, and packet loss ratio of nodes are known by the BS. The BS performs the cluster generation and the CH selection, then broadcasts a s_p-bit advertisement message to each cluster. The amount of energy consumed is given by Eq. 9.

$$E_{RX}(p) = n \cdot s_p \cdot E_{diss} \tag{9}$$

Let's evaluate the energy consumption of each CH during a round. The choice of *hop* to transmit packet to sink depends on the proximity of each CH to this latter. Let k_i be the number of CH that send their aggregated packets to CH_i, n_i the number of packets sent by CH_i during a round, and $\delta_i j$ the intercluster distance between cluster h_i and h_j. Thus, in the worst-case scenario, the energy consumed by CH to send data to the BS is given by Eq. 10. In the best-case scenario, the energy consumption is presented in Eq. 11.

$$E_{TX_{CH_i}} = n_i \cdot s_p \cdot E_{diss} + \sum_{j \leq k_i} \left(n_j \cdot s_p \cdot \epsilon_{fs} \cdot d_{ij}^2 \right) \tag{10}$$

$$E_{TX_{CH_i}} = n_i \cdot s_p \cdot \left(E_{diss} + \epsilon_{fs} \cdot d_i^2 \right) \tag{11}$$

where, d_i is its distance to the BS.

5.2. Evaluation metrics

We measure the performance of our algorithm based on the following criteria:

- *Network lifetime*: it can be defined in terms of either death of the first node or the death of all the nodes present in the network. In this work, the network lifetime is defined as the number of alive nodes.
- *Residual energy of the network*: it is the sum of remaining energy of alive nodes.
- *Total energy consumption of the network*: it is the amount of energy consumed by the nodes.

5.3. Simulation tool

To evaluate the performance of our algorithm, we use Castalia. It is a simulator for WSN, Body Area Networks and generally networks of low-power embedded devices [35]. It is based on the OMNeT++ platform, which is an object-oriented modular discrete event network simulation framework. It has a generic architecture, so it can be and has been used in various problem domains [36]. Castalia can be used by researchers and developers who want to test their distributed algorithms and/or protocols in realistic wireless channel and radio models, with a realistic node behavior especially relating to access of the radio [35]. We adopted 2000 rounds in the simulation. This value has been properly chosen is a very deterministic way in order to reach the optimal solution is a reasonable time. Indeed, the values of the parameters adopted for the proposed HGC protocol have been carefully adopted according to the previous studies in [14, 15, 16, 37]. The rest of the simulation parameters are given in Table 2.

5.4. Results and discussion

To evaluate the performance of our clustering scheme, we compare it to the PSO-C, PSO-UFC, and EBUC. The PSO-C is a centralized clustering protocol in which the sink node controls the entire CH selection and cluster formation process to enhance the lifetime of the network. The PSO-UFC addresses

Table 2: Castalia simulation parameters

Type	*Parameter*	*Value*
Network	Area	$200 \times 200m^2$
	Number of nodes	200
	Initial energy	0.5 Joule
	Simulation rounds	2000
	Number of CHs	10
Radio model	CC2420	
MAC protocol	TMAC	

imbalanced clustering and the fault tolerance issue in the existing energy-balanced unequal clustering protocol for the long-run operation of the network. The EBUC creates unequal clusters in such a way that the CHs closer to the sink node have smaller cluster sizes in order to preserve their battery for high intercluster relay traffic load. Furthermore, we evaluated the performance in two scenarios. In the first scenario, the sink node is located at the center of the network, and in the second scenario, the sink node is positioned at the bottom right of the area of deployment. The results are presented in the following section.

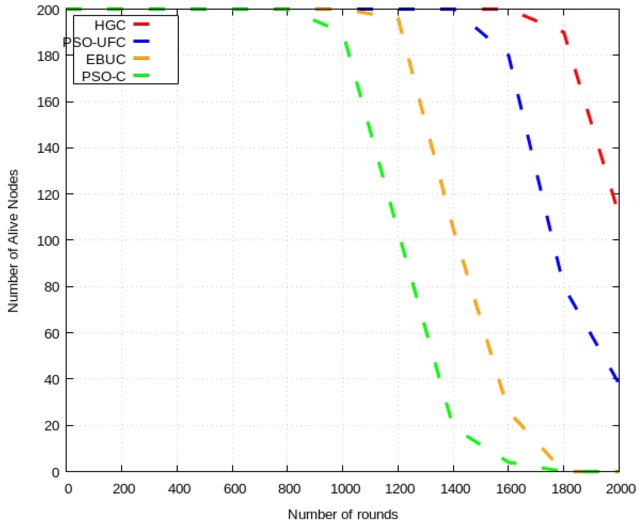
Figure 5a and Figure 5b present the number of alive nodes of the network in scenario WSN#1 and WSN#2 respectively. We can observe that the Hypergraph Clustering algorithm (HGC) has the biggest number of alive nodes. From 200 to 1600 rounds, all the nodes remain alive when HGC is used. For 1800 and 2000 rounds, 68.75% of nodes remain alive, with HGC. This is due to the fact that HGC distributes fairly the role of CH, i.e in each cluster, every node having more energy has the highest probability to become CH.

The residual energy of the WSN#1 and WSN#2 are shown in Figure 6a and Figure 6b. By fairly distributing the CH role, HGC ensures that the nodes save more energy. Moreover, the proposed HGC protocol saves 14.42% of energy in both scenarios compared to the others which save very less. Furthermore, compared to the EBUC and the PSO-C where nodes die gradually, the scheme designed in the clusters formation, especially in the CHs selection process confers on the HGC a high energy efficiency. We compared the protocols based on the total amount of energy consumed by the nodes in networks WSN#1 and WSN#2. The results presented in Figure 7a and Figure 7b show that the HGC protocol consumes less energy than the others, that is 4% of energy consumed with 200 rounds and 85% of energy consumed with 2000 rounds.

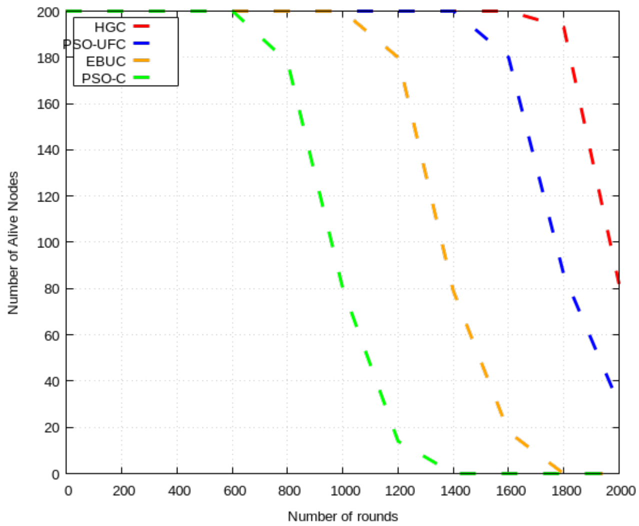
Table 3 represents a comparison in terms of amount of data delivered to the BS per second in the proposed HGC while considering the network WSN#1 and WSN#2. We computed the number of bits which are transmitted to the BS per second, during the rounds. As shown on Table 3, the BS received more packets in the network WSN#1.

6. Conclusion

Energy conservation of WSNs remains one of the main challenges of this kind of network since energy is considered as a



(a) Number of nodes alive: WSN#1

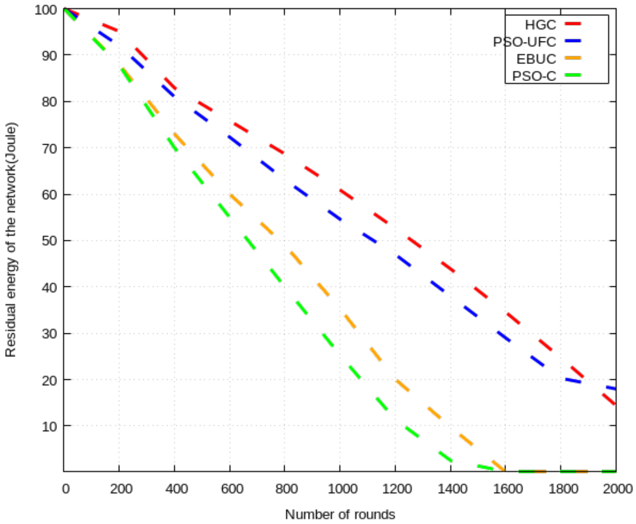


(b) Number of nodes alive: WSN#2

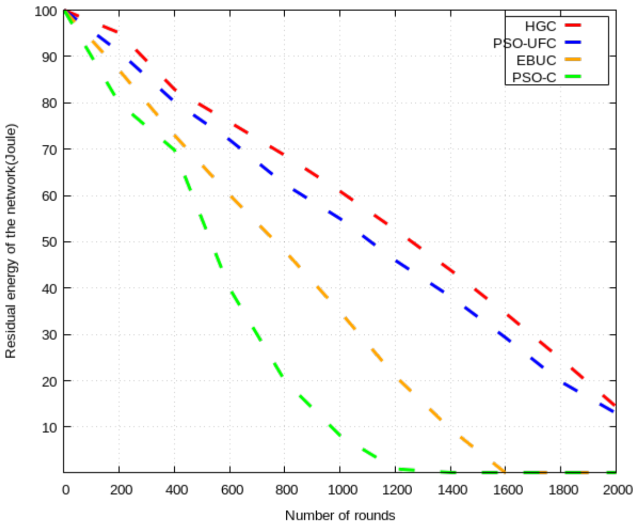
Figure 5: Comparison of number of nodes alive

Table 3: Comparison of throughput of HGC for both network scenarios WSN#1 and WSN#2

	[200; 600[	[600; 1000[	[1000; 1400[	[1400; 1800[	[1800; 2200[
WSN#1 (bit/sec)	2.38×10^4	2.01×10^4	3.4×10^4	3.8×10^4	4.01×10^4
WSN#2 (bit/sec)	1.2×10^4	1.2×10^4	2.4×10^4	2.5×10^4	2.52×10^4

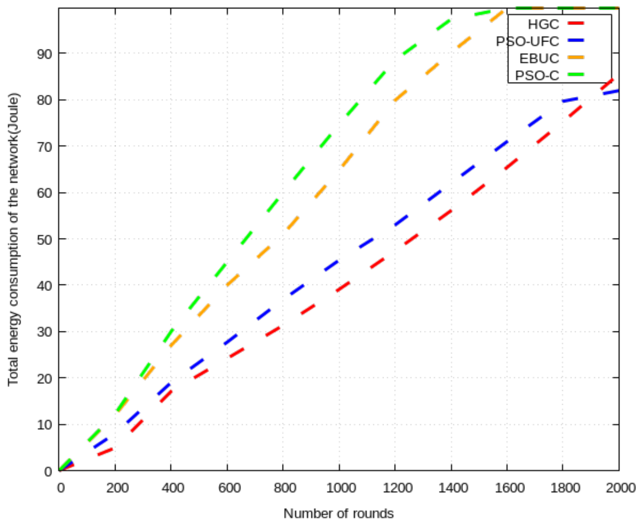


(a) Residual energy of the network: WSN#1

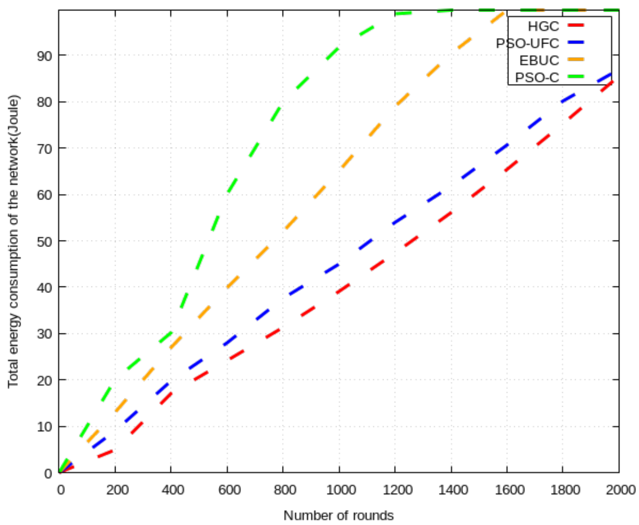


(b) Residual energy of the network: WSN#2

Figure 6: Comparison of residual energy of the network



(a) Energy consumption of the network: WSN#1



(b) Energy consumption of the network: WSN#2

Figure 7: Comparison of energy consumption of the network

scarce source for a sensor node. Thus, optimizing the energy of sensor nodes is vital to extend the network lifetime. In this paper, we proposed a heuristic clustering scheme called HyperGraph Clustering (HGC) in order to minimize the use of energy in WSN. The proposed scheme consisted of three phases -initialization - cluster formation - and CH selection, based on concepts of hypergraph theory to generate clusters. Theoretical evaluation of the performance of HGC showed that it consumes less energy during the cluster formation and CH selection phases. Moreover, we made intensive simulations of the proposal and compared the HGC of some relevant protocols. The results showed that HGC outperforms the compared protocols, in the different scenario setups. As future work, we intend to adapt the HGC to mobile wireless sensors networks.

References

[1] C. Titouna, F. Naït-Abdesselam, A. Khokhar, DODS: A Distributed Outlier Detection Scheme for Wireless Sensor Networks, Computer Networks 161 (2019) 93–101.

[2] O. Senouci, Z. Aliouat, S. Harous, MCA-V2I: A Multi-hop Clustering Approach over Vehicle-to-Internet communication for improving VANETs performances, Future Generation Computer Systems 96 (2019) 309–323.

[3] Q. Wang, I. Balasingham, Wireless sensor networks-an introduction, Wireless Sensor Networks: Application-Centric Design (2010) 1–14.

[4] V. Sharma, A. Pughat, Introduction to energy-efficient wireless sensor networks, in: Energy-Efficient Wireless Sensor Networks, CRC Press, 2017, pp. 1–26.

[5] A. Djedouboum, A. Abba Ari, A. Gueroui, A. Mohamadou, Z. Aliouat, Big data collection in large-scale wireless sensor networks, Sensors 18 (12) (2018) 4474.

[6] R. Hamidouche, Z. Aliouat, A. M. Gueroui, A. A. A. Ari, L. Louail, Classical and bio-inspired mobility in sensor networks for iot applications, Journal of Network and Computer Applications 121 (2018) 70–88.

[7] G. Wang, B. Lee, J. Ahn, G. Cho, A UAV-assisted CH election framework for secure data collection in wireless sensor networks, Future Generation Computer Systems.

[8] S. Rani, S. H. Ahmed, Multi-hop routing in wireless sensor networks: an overview, taxonomy, and research challenges, Springer, 2015.

[9] F. Khedim, N. Labraoui, A. A. A. Ari, A cognitive chronometry strategy associated with a revised cloud model to deal with the dishonest recommendations attacks in wireless sensor networks, Journal of Network and Computer Applications 123 (2018) 42–56.

[10] J. F. Myoupo, B. P. Nana, V. K. Tchendji, Fault-tolerant and energy-efficient routing protocols for a virtual three-dimensional wireless sensor network, Computers & Electrical Engineering 72 (2018) 949–964.

[11] A. A. A. Ari, B. O. Yenke, N. Labraoui, I. Damakoa, A. Gueroui, A power efficient cluster-based routing algorithm for wireless sensor networks: Honeybees swarm intelligence based approach, Journal of Network and Computer Applications 69 (2016) 77–97.

[12] A. A. A. Ari, A. M. Gueroui, C. Titouna, O. Thiare, Z. Aliouat, Resource allocation scheme for 5G C-RAN: a Swarm Intelligence based approach, Computer Networksdoi:10.1016/j.comnet.2019.106957.

[13] J. Wang, Y. Gao, W. Liu, A. K. Sangaiah, H.-J. Kim, An improved routing schema with special clustering using pso algorithm for heterogeneous wireless sensor network, Sensors 19 (3) (2019) 671.

[14] N. A. Latiff, C. C. Tsimenidis, B. S. Sharif, Energy-aware clustering for wireless sensor networks using particle swarm optimization, in: 2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, IEEE, 2007, pp. 1–5.

[15] C.-J. Jiang, W.-R. Shi, X.-l. TANG, et al., Energy-balanced unequal clustering protocol for wireless sensor networks, The Journal of China Universities of Posts and Telecommunications 17 (4) (2010) 94–99.

[16] T. Kaur, D. Kumar, Particle swarm optimization-based unequal and fault tolerant clustering protocol for wireless sensor networks, IEEE Sensors Journal 18 (11) (2018) 4614–4622.

[17] N. Kumar, Y. Singh, Routing protocols in wireless sensor networks, in: Handbook of Research on Advanced Wireless Sensor Network Applications, Protocols, and Architectures, IGI Global, 2017, pp. 86–128.

[18] J. Wang, Y. Cao, B. Li, H.-j. Kim, S. Lee, Particle swarm optimization based clustering algorithm with mobile sink for wsns, Future Generation Computer Systems 76 (2017) 452–457.

[19] V. Saranya, S. Shankar, G. Kanagachidambaresan, Energy efficient clustering scheme (EECS) for wireless sensor network with mobile sink, Wireless Personal Communications 100 (4) (2018) 1553–1567.

[20] Y. Zhang, H. Sun, J. Yu, Clustered routing protocol based on improved k-means algorithm for underwater wireless sensor networks, in: 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), IEEE, 2015, pp. 1304–1309.

[21] M. Stein, D. Lerbs, M. Hassan, M. Schnee, I. Schweizer, K. Weihe, M. Mühlhäuser, Evaluation study for clustering in wireless sensor networks, in: 2016 10th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), IEEE, 2016, pp. 1–6.

[22] G. Y. Park, H. Kim, H. W. Jeong, H. Y. Youn, A novel cluster head selection method based on k-means algorithm for energy efficient wireless sensor network, in: 2013 27th International conference on advanced information networking and applications workshops, IEEE, 2013, pp. 910–915.

[23] A. Ray, D. De, Energy efficient clustering protocol based on K-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network, IET Wireless Sensor Systems 6 (6) (2016) 181–191.

[24] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, in: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[25] J. E. Z. Gbadoubissa, A. A. A. Ari, A. M. Gueroui, Efficient k-means based clustering scheme for mobile networks cell sites management, Journal of King Saud University-Computer and Information Sciences.

[26] S. Saravanan, R. Poovazhaki, N. Shanker, Cluster Topology in WSN with SCPS for QoS, Wireless Personal Communications 99 (3) (2018) 1295–1314.

[27] M. Yan, K.-Y. Lam, S. Han, E. Chan, Q. Chen, P. Fan, D. Chen, M. Nixon, Hypergraph-based data link layer scheduling for reliable packet delivery in wireless sensing and control networks with end-to-end delay constraints, Information Sciences 278 (2014) 34–55.

[28] H. Zhang, Y. Ji, L. Song, H. Zhu, Hypergraph based resource allocation for cross-cell device-to-device communications, in: 2016 IEEE International Conference on Communications (ICC), IEEE, 2016, pp. 1–6.

[29] Z. Chen, H. Lin, L. Wang, B. Zhao, Interference-free clustering protocol for large-scale and dense wireless sensor networks., KSII Transactions on Internet & Information Systems 13 (3).

[30] M. Raza, H. Le Minh, N. Aslam, S. Hussain, Deterministic scheduling for energy efficient and reliable communication in heterogeneous sensing environments in industrial wireless sensor networks, EAI Endorsed Transactions on Energy Web 3 (11) (2017) 152764.

[31] C. Cirstea, M. Cernaianu, A. Gontean, Packet loss analysis in wireless sensor networks routing protocols, in: Telecommunications and Signal Processing (TSP), 2012 35th International Conference on, IEEE, 2012, pp. 37–41.

[32] K. Srinivasan, P. Dutta, A. Tavakoli, P. Levis, Understanding the causes of packet delivery success and failure in dense wireless sensor networks, in: Proceedings of the 4th international conference on Embedded networked sensor systems, ACM, 2006, pp. 419–420.

[33] J. Zhao, R. Govindan, Understanding packet delivery performance in dense wireless sensor networks, in: Proceedings of the 1st international conference on Embedded networked sensor systems, ACM, 2003, pp. 1–13.

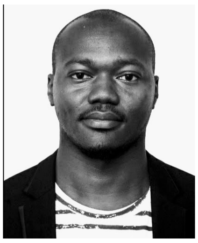
[34] W. Osamy, A. M. Khedr, A. Aziz, A. A. El-Sawy, Cluster-tree routing based entropy scheme for data gathering in wireless sensor networks, IEEE Access 6 (2018) 77372–77387.

[35] A. Boulis, Castalia: A simulator for wireless sensor networks and body area networks, NICTA 41 (2011) 1–120.

[36] O. Andras Varga, Omnet++: User guide Version 5.4.1 (2016) 1–166.

[37] R. Hamidouche, Z. Aliouat, A. A. A. Ari, A. M. Gueroui, An efficient clustering strategy avoiding buffer overflow in iot sensors: a

bio-inspired based approach, IEEE Access (2019) 1–19doi:10.1109/ACCESS.2019.2943546.



Jocelyn Edinio Zacko Gbadouissa is a Teaching Assistant at African Institute for Mathematical Sciences (AIMS-Cameroon). He graduated from the University of Ngaoundere in Cameroon with a Master degree in Computer Engineering in 2016, and from the AIMS-Cameroon with a Master degree in Mathematical Sciences in 2017. Prior to that, J.E.Z. Gbadouissa received the Bachelor of Science degree in Mathematics and Computer Science in 2013 from the University of Ngaoundere, Cameroon. He is actually involved in a co-supervised Ph.D. Student program in Computer Science with the University of Maroua, Cameroon and the University of Versailles Saint-Quentin-en-Yvelines, France. His current research includes Wireless Networks, Mathematical Modeling and Artificial Intelligence.

Ado Adamou Abba Ari is a Research Associate at the LI-PaRAD Lab of the University of Versailles Saint-Quentin-en-Yvelines, France and at the LaRI Lab of the University of Maroua, Cameroon. He received his Ph.D. degree in computer science in 2016 from Université Paris-Saclay in France with the higher honors. He also received the master degree of business administration (MBA) in 2013, the Master of Science (M.Sc.) degree in computer engineering in 2012 and the Bachelor of Science (B.Sc.) degree in mathematics and computer science in 2010 from the University of Ngaoundéré, Cameroon. He serves/served on several journals and conferences program and reviewing committees. Thus, he achieved the outstanding reviewer status with the Elsevier Computer Networks (IF: 3.030). Moreover, his a recognized reviewer of a number of journals including IEEE Access (IF: 4.098), Journal of Network and Computer Applications (IF: 5.273), Computer Communications (IF: 2.766), Remote Sensing (IF: 4.118), Sensors (IF: 3.031), Symmetry, (IF:2.143), Electronics (IF: 1.764), Telecommunication Systems (IF: 1.707), Wireless Personal Communications (IF: 0.929), Sustainable Computing: Informatics and Systems (IF: 1.800), etc. His current research is focused on bio-inspired computing, Wireless Networks, IoT, 5G and the Cloud Radio Access Network.

Chafiq Titouna received his Ph.D. degree in computer science in 2017 from University of Bejaia, Algeria, a Magister degree in computer science in 2012 from the University of M'Sila and the Higher National School of Computer Science, Algeria. He is currently an Associate Professor at the University of Batna 2, Algeria. His current research is focused on Wireless Sensor Networks, VANET, 5G and the Cloud Radio Access Network.

Abdelhak (Mourad) Gueroui is an associate professor in computer engineering at the University of Versailles Saint-Quentin-en-Yvelines, France. He received his M.Sc. degree from Université Paris 6 and his Ph.D. in networking and computer engineering in 2000 from the University of Versailles Saint-Quentin-en-Yvelines, France. He has authored more than 100 publications in international conferences and journals, as well as book chapters, including ACM, IEEE, ELSEVIER and has both chaired and served in numerous program committees in prestigious international conferences. He received several awards, including best papers and serves/served on several journals and conferences executive committees. His research interests are in the area of wireless networks (WATM, WIMAX, LTE, 5G, Cloud Computing, WLAN, MESH, VANET and WSNs), particularly performance evaluation and QoS provisioning.

Ousmane Thiare is a full Professor at the department of computer science of the Gaston Berger University of Saint-Louis, Senegal since 2015. He obtained the Ph.D. in Computer Science at the University of Cergy-Pontoise, France in 2007. He obtained a Diploma of Advanced Studies (DEA) in Computer Science at the Polytechnic School of the François Rabelais University of Tours (Polytech'Tours) and a DEA in Applied Mathematics in 2000 at the Gaston Berger University of Saint-Louis, Senegal. From 2013 to 2018 he served as Director of the Ousmane Seck Computer Center at the Gaston Berger University of Saint-Louis, Senegal. Professor Ousmane Thiare has also been the Director of the African Centre of Excellence for Mathematics, Informatics and Information and Communication Technologies (ACE-MITIC) since 2017. ACE-MITIC is an ambitious World Bank project with a budget of US$8 million. Professor Ousmane Thiare was appointed, in May 2018, Rector of Gaston Berger University of Saint Louis, President of the University's Assembly. He is active in research in his preferred fields, namely: distributed systems, distributed resource management, computing grids, cloud computing, mobile ad hoc networks, wireless sensor networks and the Internet of Things (IoT).