Routing Optimization Using an Enhanced Protocol for Wireless Sensor Networks with Chain Structure

Ilies Ali Lekhal and Samira Chouraqui
Department of Computer Science, Faculty of Mathematics and Computer Science, University of Sciences and Technology of Oran – Mohamed Boudiaf, Oran, Algeria
Email: lekhal.ilies@gmail.com; s_chouraqui@yahoo.fr

Abstract —A wireless sensor network (WSN) consists of spatially distributed micro-sensors capable of harvesting and transmitting environmental data in an autonomous manner for monitoring physical or environmental conditions. The data captured by the nodes are routed via multi-hop routing to a node called a collection point. For many sensor network applications, the most important features are the network's lifetime, scalability, and equilibrium. Clustering techniques are an effective solution for achieving these objectives. As an alternative technique, chains can be built instead of clusters. In this article, we present a new routing protocol for wireless sensor networks, called "enhanced protocol based on chains and optimized clustering" (EPCOC). The new protocol combines the advantages of the K-means algorithm with MapReduce and a number of WSN technologies such as LEACH, PEGASIS, LEACH-C, and K-LEACH. This approach organizes the network nodes into an optimized cluster of chains by applying both the clustering approach and the chains approach. An analysis of EPCOC's performance shows that it extended the network lifetime 40% to 50% over that of LEACH, 30% to 35% over that of LEACH-C, and 15% to 22% over that of PEGASIS. Furthermore, EPCOC achieved marked improvements over those protocols in terms of energy efficiency and latency.

Index Terms—Wireless sensor network, network lifetime, energy efficiency, K-means, MapReduce, LEACH, PEGASIS, LEACH-C, K-LEACH

I. Introduction

Wireless sensor networks (WSNs) today are used widely in a number of fields and include military, medical, environmental, home monitoring, and industrial applications [1], [2]. A wireless sensor network is a set of communicating nodes that collect information from the environment via sensors, process the information, locally make decisions, and wirelessly communicate with other nodes in the network [3]. Each node consists of four components: transceiver, processor, sensor, and energy unit [4].

Energy is the determining factor for the performance of a wireless sensor network. However, the nodes in WSNs are powered by small batteries with limited energy, which imposes a major constraint that reduces the lifetime and effectiveness of the WSN [5].

Manuscript received May 25, 2018; revised November 25, 2018.
Corresponding author email: lekhal.ilies@gmail.com.
doi:10.12720/jcm.13.12.691-701

Clustering of sensor nodes has been the subject of many studies by researchers seeking to propose solutions for scalability, energy, and lifetime issues of sensor networks. Clustering algorithms aim to limit the communication occurring within a local domain [6]. Clusters are groups of nodes, and local interactions between cluster members are controlled through a cluster head (CH).

To conserve energy, the nodes connect with the cluster head, and their collected data are collected and consolidated by the cluster head. The cluster heads can in turn create another layer of clusters among themselves before attaining the sink. Fig. 1 shows a typical architecture and communication simulation of a WSN [7].

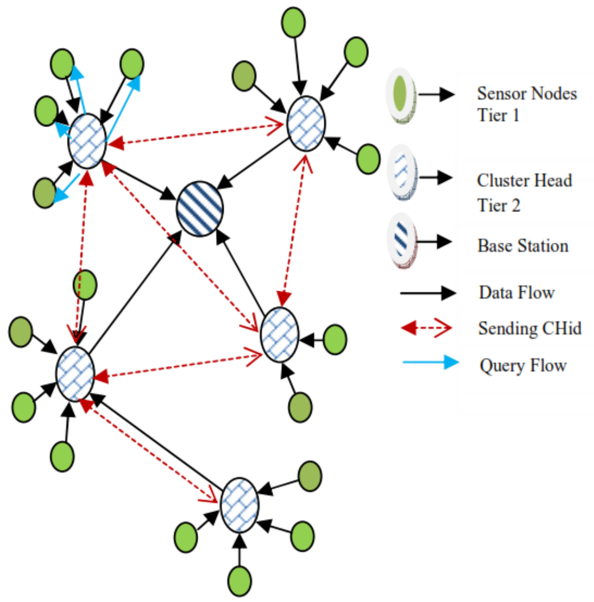

Fig. 1. Flow of data in a WSN with clustering and aggregation [7].

Routing in WSNs is complex because of the characteristics that distinguish WSNs from other wireless networks such as wireless ad hoc networks or cellular networks.

Many new solutions have been proposed, taking into consideration the architectural and application characteristics of WSNs. Routing protocols for WSNs can be categorized according to network structure into the following types: flat network routing, hierarchical network routing, and location-based network routing [8]. Many clustering routing protocols have been suggested.

Low-energy adaptive clustering hierarchy (LEACH) [9], presented by Heinzelman *et al*., is one of the first cluster-based routing protocols; it has a hierarchical structure and uses a self-organized cluster-based approach. The network nodes are divided into clusters. With the clustering technique, the nodes organize themselves into a hierarchical structure. A cluster head (CH) is elected in each cluster. The CHs collect data from the associated member nodes in their clusters based on time-division multiple-access (TDMA) scheduling; then, the aggregated data are sent to the base station (BS) as a data packet. After a predetermined time period, CHs are selected via a message from the BS.

The main purpose of the routing protocol proposed in this article is to prolong the life of a WSN. The idea of the new protocol is to combine the advantages of the K-means algorithm with MapReduce and existing WSN protocols. Based on clustering and chains approaches, this new approach aims to arrange the network nodes in optimized chained clusters.

The remainder of this paper is organized as follows. The discussion of related work is in Section II. In Section III, we introduce our proposed protocol. The simulation results are presented in Section IV, and we finally conclude with Section V.

II. Related Work

A. Routing in WSNs

The main goal of wireless sensor networks is to deliver better data communication with a higher lifetime by preserving the energy of the sensor nodes. Indeed, for obtaining effective communication, there are certain important issues that should be considered before designing any protocol. Among these issues are node placement, the trade-off between energy consumption and precision, the data reporting method, node/link heterogeneity, fault tolerance, scalability, transmission media, connectivity, network dynamics, data collection, and quality of service. These key elements must be examined when designing any wireless sensor networks or protocols for WSNs.

Various protocols have been designed for WSN communication. These protocols are classified into two types: those based on network structure and those based on protocol operation [10]-[12]. These types are described below.

Network-Structure-Based Routing Protocols: These protocols may be further classified into three types: flat network routing, hierarchical network routing, and location-based routing.

- Flat network routing. In this category, each node operates similarly to cooperate in performing the sensing task. Because there are numerous nodes, however, this type of protocol is not adapted for allocating a global identifier to each node. This observation has led to the development of data-centric routing in which the base station sends queries to certain areas and waits for the responses to those queries. By contrast, earlier data-centric protocols, such as the sensor protocols for information via negotiation (SPIN) and directed diffusion, operate to save energy using data negotiation and redundant free data.
- Hierarchical network routing. These kinds of networks use the clustering approach. In essence, nodes have one cluster head (CH), which has a higher energy. The CH node will be used for data transmission, while a low-energy node will be used for the sensing process. Usually, hierarchical routing is based on two processes in succession: the process of selecting the cluster head and the transmission/routing of the data [9]. A variety of data collection methods and data aggregation processes are used in hierarchical routing protocols. The various hierarchical routing protocols designed for wireless sensor networks include the low-energy adaptive clustering hierarchy (LEACH) protocol and the power-efficient gathering in sensor information systems (PEGASIS) protocol.
- Location-based routing. There are some applications that continuously track the location of a sensor node. In location-based routing, the sensor nodes are located on the basis of their address, and through the incoming signal strength of the sensors, their neighboring nodes can also be determined. This kind of routing can be easily realized using the Global Positioning System (GPS). When there is no task to be performed in the network, the nodes will be in a sleep state so that energy consumption can be reduced [11]. This type of routing encompasses the geographic adaptive fidelity (GAF) and geographic and energy aware routing (GEAR) protocols.

Protocol-Operation-Based Routing Systems: These are further classified into negotiation-based routing, multipath-based routing, query-based routing, quality of service (QoS)–based routing, and coherent- and non-coherent-data-processing-based routing.

- Negotiation-based routing. This type of routing accomplishes the negotiation of data before transmission. This method eliminates the redundant information identified by multiple sensor nodes in the same network. Such a system will use negotiation mechanisms such as flooding and data collection to ensure that no redundant data will be sent through the network. SPIN is a negotiation-based protocol.
- Multipath-based routing. In this routing system, multiple paths exist between source and destination. Consequently, if any path fails at a given time, the communication can be accomplished through another path [11]. This method ensures the fault tolerance and reliability of the communication network.
- Query-based routing. In this model, the destination node distributes the query for the sensing task. The sensing node with matching results will send the

result back to the destination node. Therefore, all the nodes must have predefined results for the queries. Thus, the communication takes place.

- QoS-based routing. Quality-of-service routing guarantees a balance between the quality of data being sent and the energy consumption needed to send the data. This method of routing takes into consideration several aspects of WSNs such as power consumption, communication bandwidth, and transmission delay. Sequential assignment routing (SAR) is an illustration of QoS-based routing.
- Coherent- and non-coherent-data-processing-based routing. For almost any kind of routing, data processing is the major task to be considered. Many data processing techniques have evolved, which include coherent and non-coherent data processing techniques. When data are processed locally and then forwarded to other nodes for further access, it is called non-coherent data processing; the nodes that are responsible for further processing are called aggregators. Alternatively, when data are sent directly to the aggregators, it is called coherent data processing. On the whole, this method will minimize the computation measure.

B. LEACH

Low-energy adaptive clustering hierarchy (LEACH) was presented by Heinzelman *et al*. [9] as a hierarchical clustering algorithm for sensor networks. LEACH is a cluster-based protocol; it involves the formation of disseminated clusters. In this method, a few sensor nodes are randomly selected as cluster heads. Then, the method distributes the energy load among the sensors in the network. As a result, clusters of sensor nodes are formed based upon their signal strength.

The local cluster heads are regarded as routers to the sink. In LEACH, data from the member nodes are compressed by the cluster heads, which then send an aggregated packet to the BS, thus decreasing of the amount of information transferred to the BS. The Time-Division Multiple-Access / Code-Division Multiple-Access Medium Access Control (TDMA/CDMA MAC) protocol is utilized to lessen inter- and intra-cluster collisions.

Stages: LEACH's operation is divided into two stages: the setup stage and the steady-state stage. The latter takes much more time than the former in order to reduce overhead.

- Setup phase. At this stage, some predetermined fraction of nodes, p, select themselves as cluster heads. A sensor node n selects a random number r between 0 and 1. If the number is lower than a threshold value $T(n)$, that node becomes the cluster head for that round. The threshold value is calculated using the following formula:

$$T(n) = \begin{cases} \dfrac{p}{1 - p\left(r \bmod \dfrac{1}{p}\right)} & \text{if } n \in G \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

  G represents the set of nodes that were not cluster heads in the last $1/p$ rounds. The use of a threshold, $T(n)$, permits the transformation of every node into a cluster head at some point within $1/p$ rounds, yet it is impossible for nodes that have been cluster heads to become a cluster head for a second time within p rounds. After that, any node may (with a probability of $1/p$) become a cluster head in any round.

  After the selection of the cluster heads, an announcement is launched to the rest of the nodes that these are the new cluster heads. Next, the non-clustered nodes choose the cluster that they want to join according to the signal strength of the advertisement. Then, the non-cluster-head nodes notify the cluster heads of their membership.
- Steady-state phase. Based on the number of nodes in the cluster, the cluster head starts to create a TDMA schedule after receiving all the messages from nodes that desire membership in the cluster. Then, it allocates each node a time slot when it can transmit. All the other nodes in the cluster receive the same schedule. In this stable phase, the sensor nodes start sensing and transmitting data to the cluster heads. After receiving all the data, the cluster head node consolidates them to send them to the base station.

After a specific period of time, the network goes back into the setup phase and begins another round of selecting new cluster heads. Each cluster interacts through distinct CDMA codes in order to diminish interference from nodes that are part of other clusters.

Drawbacks of the LEACH Protocol: The various disadvantages of the LEACH protocol are as follows [13]-[15]:

- It relies significantly on the cluster heads rather than the cluster members within the cluster to communicate with the sink. Consequently, it experiences robustness problems such as the failure of cluster heads.
- It generates additional overhead due to the process for the election of the cluster head in each iteration of the communication of information as well as computational overhead, which leads to energy inefficiency for dynamic clustering in the case of large-scale networks.
- There is no inter-cluster communication in the network because CHs communicate directly with the sink. This process requires a high range of transmission power in the network. Because of this, LEACH is not well suited for large-scale networks.
- In LEACH, CHs are not evenly distributed within the cluster, which means that CHs can be positioned at the margin of a cluster.
- In LEACH, CH selection is random and does not take into account the energy consumption of the different nodes within the cluster. This may lead to selecting nodes as CHs in many simultaneous iterations of data processing in the network. It does not work perfectly

with applications that require wide-area coverage with multi-hop inter-cluster communication.

Variants of the LEACH Protocol: The LEACH protocol has many descendant protocol variants that outperform LEACH in terms of power consumption or sensor node lifetime. These categories of approaches are presented and explained in [12]-[16] and are summarized below.

- LEACH-F. In LEACH-F, after the clusters are created, each one is arranged and fixed; there is no setup at the start of each round. Essentially, the same central cluster formation algorithm is used for determining the clusters (as also in LEACH-C, described below). In LEACH-F, it is not possible to add any new nodes to the system or to adjust their behavior based on dying nodes. Additionally, node mobility does not hold in LEACH-F; only the rotation of cluster head position among the nodes inside the cluster can be performed. In fact, LEACH-F may or may not lead to energy savings. In LEACH-F, both stable cluster and rotating cluster concepts are used, and any cluster formed is preserved throughout the network's lifetime to avoid the re-clustering process [17].
- LEACH-C. LEACH-C's fundamental feature is a centralized clustering algorithm. Commonly, the stable state will remain the same because the setup stage of LEACH-C consists of sending the information of each node's current location and energy level to the base station. By operating on the total information of the network, the base station may produce better clusters that require less energy for any data transmission [18]. GPS or another location tracking technique is required. The base station must ensure that only nodes with sufficient energy are permitted to contribute in the selection of the cluster head. Therefore, the base station broadcasts the information to all nodes within the network.

  LEACH-C has a deterministic threshold algorithm that takes into consideration the amount of energy in the node and whether the node was recently a cluster head. The number of cluster head nodes and their placement is not determined. The central control algorithm can be applied in order to form a better distribution of cluster head nodes throughout the network.
- LEACH-B. The LEACH algorithm uses decentralized cluster formation algorithms. In the LEACH-B variant, the cluster node recognizes only its position and that of the destination node. The stages of LEACH-B are, first, the cluster head selection and then the cluster formation and data transmission with multiple accesses. A cluster head that is between the destination and the source will be selected. Consequently, it delivers better energy efficiency [19].
- Energy-LEACH (E-LEACH). This method assists the cluster head selection by making the residual energy of the cluster head nodes a vital element, determining whether these cluster nodes will turn into the cluster head in the subsequent round. Moreover, E-LEACH enhances the cluster head node election process [19]. Consequently, it is characterized by producing a network with a longer lifetime and greater energy savings than the LEACH protocol.
- MS-LEACH. This category combines multi-hop LEACH and single-hop LEACH. Owing to the identification of the critical value for cluster area size, any issue related to multi-hop and single-hop clustering will be resolved. Simulation results for MS-LEACH reveal a 200% improvement in network lifetime [19].
- MH-LEACH. This variant demonstrates a new communication approach from a single hop to multiple hops between cluster head and base station. In this method, the CH will communicate directly with the sink node regardless of the distance measure. In fact, a greater distance is significant as the power consumed will be greater. Therefore, MH-LEACH will adopt the optimal path between CH and base station through multi-hop communication [19].
- Multi-hop LEACH. This approach illustrates the fact that if the network diameter increases, the distance between cluster head and base station will significantly increase. In fact, such a phenomenon is not suitable for any effective communication. Thus, multi-hop communication will be performed in order to reduce the energy consumption of the network [20].
- K-LEACH. The K-LEACH protocol [21] is based on the K-medoids clustering algorithm for obtaining extremely uniform clustering of nodes. It results in a very good choice of cluster heads. It is well recognized that the energy retention of a WSN depends largely on the grouping or clustering of transmission and reception nodes in the setup stage for the first round of communication. K-LEACH considers the minimum distance from the center of the cluster as a primary criterion for node selection as a CH; for the CH selection procedure from the second round onward, K-LEACH is divided into different steps, and each step includes a cluster formation phase and a steady-state phase.

C. PEGASIS Protocol

In the PEGASIS protocol, when a node receives data it transmits them to its closest neighbors and takes turns as a leader for transmission of data to the BS [22]. This approach emphasizes the even distribution of the energy load among the sensor nodes. In addition, the nodes are randomly placed in the field and organize themselves in the form of sequences using a greedy algorithm.

Alternatively, the BS computes this sequence and distributes it to all the nodes. Fig. 2 shows node 0 and node 1 linking to node 3; node 3 is connected to node 1, and node 1 is connected to node 2. When a node is disconnected, the chain is constructed by the same method to bypass the affected node.

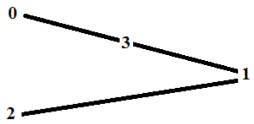

Fig. 2. Chain formation [22].

In gathering data, each node first receives data from a neighbor, combines them with its own data, and then transmits all of them to the next node in the series. In a given round, a simple token-passing approach is initiated by the leader to start the data transmission from the ends of the chain. During this process, the cost is much lower because the quantity of data is very small.

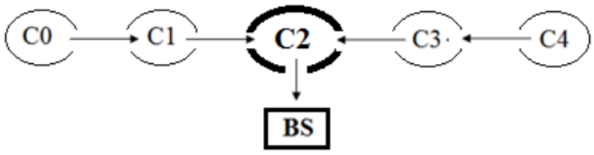

Fig. 3. Token passing [22].

In Fig. 3, node C2 is the leader. In the first step, the data token is transmitted to C0. C0 sends its data to C1. Next, C1 merges its own data with C0's data and sends them all to leader C2. Then, C2 transmits the data token to C4. C4 sends its data to C3. C3 merges its data with C4's data and then transmits them all to leader C2. C2 waits until the data are received from both neighbors, and then it merges its data with its neighbors' data. Finally, the leader transmits a single message to the BS.

In PEGASIS, each node receives and transmits a single packet in each round and becomes the leader at least once in *n* rounds, where *n* is the number of nodes (Fig. 4).

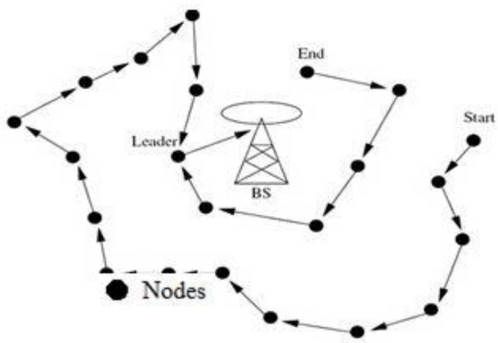

Fig. 4. Illustration of PEGASIS protocol [22].

PEGASIS expends less energy than LEACH in the following ways. Firstly, in the local gathering process, the distance transmitted by most of the nodes is much less than that by CH in LEACH. Secondly, the leader normally receives only two messages from its neighbors, in contrast to LEACH (in which, for example, if there are 10 nodes for each cluster and 50 nodes in the network, it will receive 10 messages). Finally, in each round of communication, one node sends the message to the BS.

The PEGASIS protocol has its major applications in characterizing and monitoring environments [23]. In such applications, the sensing nodes are influenced by environmental factors such as temperature, humidity, and pressure. Each node combines its sensed data with those of the adjacent node. Finally, when the CH has all the sensed data, they will be sent to the BS. Although PEGASIS has advantages in these environments, the protocol has a major disadvantage in the high latency that is introduced by the long chain of nodes.

D. MapReduce Framework

The MapReduce framework was originally developed by Google, but with its wide adoption, it has today become a standard tool for data analysis at the large scale of companies [24]. The MapReduce programming model was defined by Dean and Ghemawat [25]. It consists of two functions, Map and Reduce. Both of these functions are defined with a data structure of (key, value) pairs.

According to the format of the (key1, value1) pairs, the Map function is applied to each item in the input dataset; each call produces a list (key2, value2). All the pairs that have the same key in the output lists are gathered with the Reduce function, which generates either a one-value list (value3) or an empty list [11]. This Map-and-Reduce process is demonstrated in Fig. 5 using the following operations [24], [25]:

- Map(key1, value1) → List(key2, value2)
- Reduce(key2, value2) → List(value3)

As an illustration, consider counting the number of occurrences of each word in a large dataset. First, the Map function generates each word along with its number of occurrences in the dataset. Then, the Reduce function sums all the occurrence counts generated by the Map function for each word.

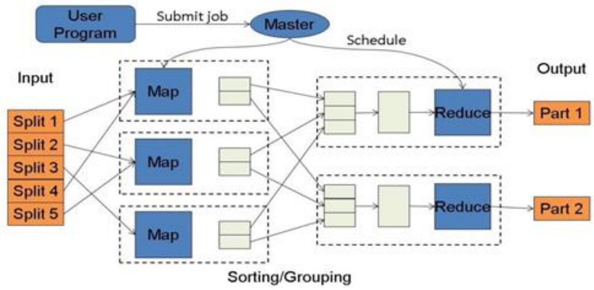

Fig. 5. MapReduce processing [24].

E. K-Means Algorithm

K-means is a partitioning-based clustering technique that attempts to determine a specified number of clusters (*k*). They are represented by their centroids through minimization of the square error function specially developed for low-dimensional data. In the case of high dimensional data, however, such an algorithm does not work well, and the result may very often be inaccurate because of outliers.

Two approaches are possible for the cluster center initialization: randomly selecting the initial values or choosing the first *k* samples of the data points. As an alternative, different sets of initial values are proposed (from the data points), and the set that is closest to optimal is chosen [7].

III. Our Proposed System

A. Overview

The main motivation for proposing our enhanced protocol based on chains and optimized clustering (EPCOC) is to combine the advantages of the K-means algorithm with the MapReduce framework and WSN routing.

EPCOC is primarily based on the advantages of the clustering techniques of LEACH and those of chain construction within the clusters of PEGASIS. Here, all nodes within a cluster communicate only with their closest neighbors and not directly with the CH, which reduces the number of nodes communicating with the CH, thereby saving energy and prolonging the CH's life (Fig. 6).

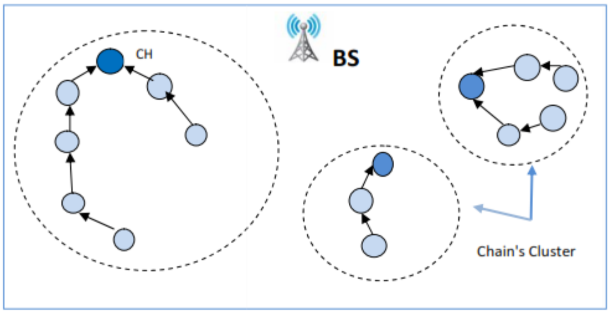

Fig. 6. Organization of chains and clusters in EPCOC.

The principal idea for implementing this architecture in EPCOC is the use of a centralized K-means algorithm for the first clustering operation, which ensures uniform clustering, followed by a parallelized algorithm for the construction of the chains within the clusters. This entire process is performed using the MapReduce framework to reduce the setup time and to improve the clustering. Fig. 7 illustrates the principle of this method.

The motivation for choosing the K-means algorithm is the use of the Euclidian distance in the selection of the cluster head and the assurance of obtaining the best clustering that always provides the most energy efficient solution.

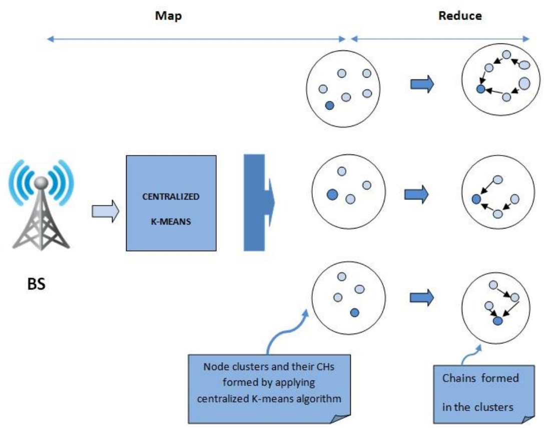

Fig. 7. EPCOC clustering process.

B. Description of Proposed Protocol

EPCOC has two essential phases: The setup phase and the steady-state phase, as shown in Fig. 8.

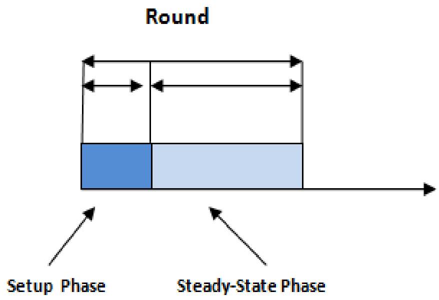

Fig. 8. EPCOC round.

The setup phase is divided into two steps. In the first step, the base station forms sensor node clusters and assigns the roles to each node. This operation is referred to as the Map protocol. The second step consists of forming the chains within the clusters by applying a chaining algorithm. This phase is referred to as the Reduce protocol.

Furthermore, to reduce the duration of the initialization phase, we have made some modifications for the proper operation of our algorithm: In the original MapReduce framework, the Map step is parallel in nature. In EPCOC, we use a centralized Map algorithm at the BS, but the Reduce step is parallelized to optimize the construction of the final chain's cluster. This parallelization confers a substantial advantage for reducing the clustering time.

In the step for constructing the chains within the clusters, EPCOC adopts the same closest-neighbor principle used in PEGASIS. The construction starts with the node furthest from the BS. (If there is more than one node the same distance from the BS, one of them will be chosen arbitrarily.) This node represents the head of the chain. Then, the node nearest to the head of the chain is selected to be added and to become the new head of the chain. The operation is repeated until all nodes are part of the series, as demonstrated in Fig. 9 and Algorithm 1.

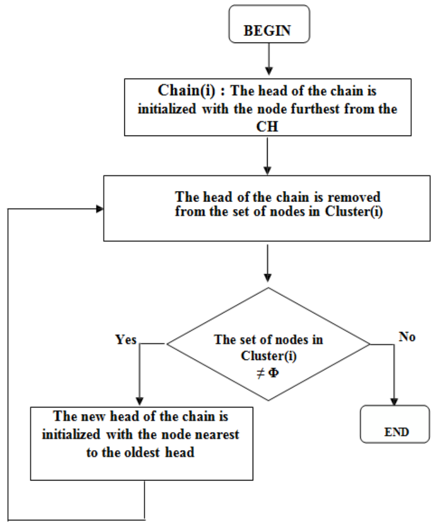

Fig. 9. Flow chart for construction of the chains within the clusters.

Algorithm 1. Chain Algorithm

```
1. procedure Chains_Construction
2. Chain(i): Chain of cluster i
3. Cluster_Node: The set number of the same-cluster nodes
4. Head_Node: The head of the chain
5. Furthest(): Returns the furthest node
6. Nearest(): Returns the nearest node
7. Dist_Node(): Calculates the distance between nodes
8. i = 1        /*ID of the chain of cluster 1*/
9. Remove(Cluster_Node,CH)
10. Chain(i) = { }
11. Head_Node = Furthest(CH) /*Returns the node furthest from the CH*/
12. Remove(Cluster_Node,Head_Node)
13. Repeat
14. Neighbor = Nearest(Head_Node) /*returns the node nearest to the head of the chain*/
15.     if (Dist_Node(Head_Node,Neighbor) < Dist_Node(Head_Node,CH)) then
16.     Head_Node = Neighbor
17.     Add(Chain(i),Head_Node)
18.     Remove(Cluster_Node,Head_Node)
19.     else
20.     Add(Chain(i),CH)
21.     i = i + 1
22.     Goto 10         /*Chain Error: reconstruct the chain*/
23.     End if
24. Until (Cluster_Node = Ø)
```

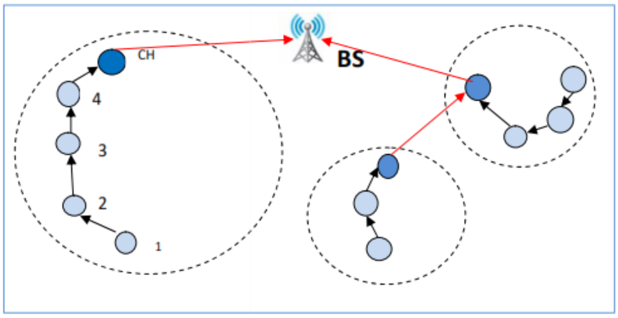

Fig. 10. Data transmission in EPCOC.

Fig. 10 shows the data transmission schema using the new node organization. Node 1 transmits its data to its closest neighbor node 2; node 2 combines the data with its own and sends them all to another neighbor. This process continues until the data reach the CH, which transmits them to the BS.

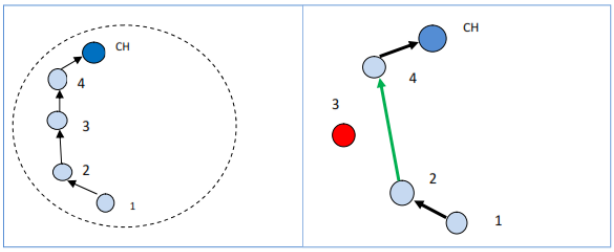

Fig. 11. Updating of a cluster's chain organization.

Fig. 11 depicts a cluster's chain organization before and after updating. All alive nodes have an energy level greater than 0, and a node dies after its energy has been depleted.

Thus, in this new technique, the number of nodes communicating with the CH is considerably reduced, and the communication between cluster nodes and the cluster head will be reduced, thus preserving the energy and ensuring long lifetimes for the CHs.

EPCOC uses a multi-hop technique for transmitting through the CHs' neighbors to limit power consumption. Each CH sends the data collected within the cluster to the BS. To preserve node energy reserves, EPCOC reduces the exchange of data between nodes and their CH by aggregating the data of each node in the cluster's chain. Additionally, EPCOC avoids collision and interference problems when accessing the medium within a cluster by using the TDMA mechanism, which defines the exact time when a node can transmit its collected data. Like LEACH, EPCOC uses the concept of the random rotation of the CH role, which controls energy dissipation and prevents the premature death of nodes chosen as CHs. In most scenarios, intra-cluster communication does not scale well when the number of nodes increases. In our approach, however, the transmission distances are shortened and the number of nodes communicating with the CH is considerably reduced, resulting in greater energy savings and an increase in CH lifetime. The latency introduced by the long chain of nodes in the PEGASIS protocol is significantly reduced because the chains formed in the clusters are small and operate at the same time.

For the first communication cycle, during installation, we use the K-means algorithm along with MapReduce for cluster formation, which ensures uniform clustering. The K-means algorithm is based mainly on the Euclidian distances, and the cluster formation by the K-means algorithm ensures the best clustering and selection of the cluster head. Using the Euclidian distance to the nearer or at the center of a cluster always gives the most energy efficient solution in WSN.

As in the K-LEACH protocol, from the second round onward EPCOC selects the next cluster heads based on the next node closest to the first-round cluster head, and the process continues thus. This technique is efficient and optimizes the selection of the CH with a lower energy cost since the nearest node in the chain is already known.

The processes of the construction of the chains repeat each setup phase with a new CH elected. By using the K-means algorithm, EPCOC ensures a perfect clustering and a better cluster head selection procedure.

C. Details of the Cluster Setup Phase Processes

In EPCOC, enhanced installation clustering is optimized using the MapReduce algorithm. The key and value types for the proposed method are as follows:

- key1: List of the initial set of k centroids selected
- value1: List of all other nodes along with their location and energy level information
- key2: List of the cluster heads
- value2: List of clusters with their member nodes
- key3: List of the cluster heads
- value3: List of all the clusters with chains

Map (First Clustering Step): In this Map process, the initial set of randomly selected k centroids is input as key1, and the list of all other nodes along with their location and energy level information is input as value1. By using the mapper (key1, value1) protocol, the Map phase produces a list of the cluster heads as key2 and a list of all the clusters with their member nodes as value2. This mapper protocol is given as Algorithm 2.

Reduce (Second Clustering Step), Construction of the Chains in the Clusters: In this Reduce process, the results produced by the mapper protocol are input as key2 to generate the cluster heads, and the list of all the clusters with their node members is input as value2. By using the reducer (key2, value2) protocol, the Reduce phase produces a list of the final clusters with their chains and cluster heads as (key3, value3). This reducer protocol is given as Algorithm 3.

```
Algorithm 2. Map Protocol
    Protocol mapper (key1, value1)
1. K-Means (key1, value1)  // Formation of the clusters with their
cluster heads and their member nodes by the BS
2. Generate output key2, value2
```

```
Algorithm 3. Reduce Protocol
    Protocol reducer (key2, value2)
1. Read (value2);  // Build chains in the clusters
2. Repeat
3. procedure Chains_Construction;
4. Until no change;
5. Produce key3,value3;
```

IV. Simulation and Results

A. Simulation Setup

A performance analysis of our protocol was carried out using the network simulator NS2 [26].

In the simulation, we compared the performance of our proposed EPCOC with the LEACH, LEACH-C, and PEGASIS protocols.

Our experimentation model was established on 100 sensor nodes randomly distributed in an area of 100 m × 100 m, shown in Fig. 12.

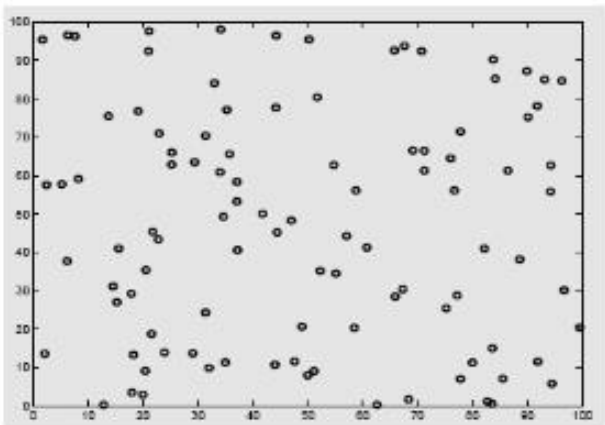

Fig. 12. Experimentation model.

Table I shows the simulation parameters. The BS was located at x = 175, y = 50. The bandwidth of the data channel was set to 1 Mbps, the length of data messages was 500 bytes, and the packet header for each type of packet was 25 bytes. The transmission and reception latency of a data packet was 25 µs. When a node's energy became lower than the threshold, it could no longer send data and was considered to be a dead node.

Before presenting the simulation results, we state the assumptions that were made:

TABLE I: SIMULATION PARAMETERS

Parameter	Value
Surface area of the network	100 m × 100 m
Position of the BS	(175,50)
Number of nodes	100
Number of clusters	5
Initial energy	2 J
Size of data packets	500 bytes

- All nodes have a fixed position throughout the simulation period.
- Both sensor nodes and base station are stationary after the base station is positioned outside the area of the sensor nodes.
- The wireless sensor network contains uniform sensor nodes.
- All sensor nodes have the same initial energy.
- There are no limits on the base station's energy, memory, or computational power.
- The radio channel is symmetric in that the energy consumed in transmitting data from node X to node Y is the same as that for transmitting from node Y to node X.

For the radio model (Fig. 13), we assumed a simple model wherein the transmitter consumes energy to run radio electronics and a power amplifier, and the receiver dissipates the energy of running the radio electronics. For our experiments, we used both the free space (d^2 power loss) and the multipath fading (d^4 power loss) channel models, depending on the distance d between the transmitter and the receiver. A free space (fs) model was used if the distance was less than a threshold d_0; otherwise, the multipath (mp) model was used. ε_{fs} and ε_{mp} are the amplifier energies in the free space and multipath models, respectively. The following equations represent the amount of energy consumed in transmitting a packet with k bits over a distance d:

$$E_{TX}(k,d) = \begin{cases} kE_{elec} + k\,\varepsilon_{fs}\,d^2 & \text{if } d \le d_0 \\ kE_{elec} + k\varepsilon_{mp}d^4 & \text{if } d > d_0 \end{cases} \quad (2)$$

$$d_0 = \sqrt{\frac{E_{fs}}{E_{mp}}} \quad (3)$$

E_{elec} is the electronics energy.

All the nodes of the network began the simulation with an initial energy of 2 J and an unlimited quantity of data to transmit to the base station.

The cluster heads changed and the cluster chains were rebuilt every 20 s.

The electronics energy (E_{elec}) depends on factors such as the digital coding, modulation, filtering, and spreading of the signal, whereas the amplifier energy, $\varepsilon_{fs}d^2$ or $\varepsilon_{mp}d^4$, depends on the distance to the receiver and the acceptable bit-error rate. For the experiments described in this paper, the communication energy parameters were set as follows: E_{elec} = 50 nJ/bit, ε_{fs} = 10 pJ/bit/m^4, ε_{mp} = 0.0013 pJ/bit/m^4, and the energy for data aggregation E_{DA} = 5 nJ/bit/signal. In addition, to receive a k-bit message, the radio expends an energy of

$$E_{RX}(k) = kE_{elec} \tag{4}$$

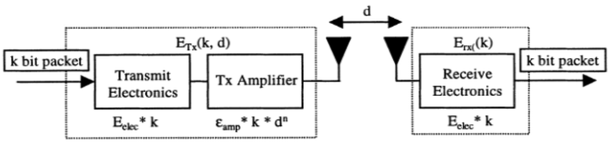

Fig. 13. Radio energy dissipation model [9].

B. Results

An analysis of the simulation results presented in Fig. 14 show that EPCOC extended the network lifetime 40% to 50% over that of LEACH, 30% to 35% over that of LEACH-C, and 15% to 22% over that of PEGASIS. We conclude that EPCOC increases the energy gain and prolongs the network's life over those of LEACH, LEACH-C, and PEGASIS. These results can be explained as a consequence of EPCOC's architecture, which is based on effective clustering and communication on small chains whereby all nodes within a cluster communicate only with their closest neighbors and not directly with the CH. This approach reduces the number of nodes communicating with the CH, thereby saving energy and prolonging the CH's life.

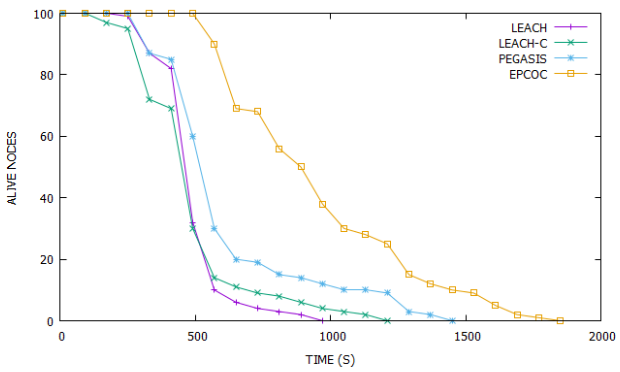

Fig. 14. Number of alive nodes over time.

In the results plotted in Fig. 15, we see the number of data messages received by the base station over the network's lifetime. We deduce that the better result demonstrated by EPCOC can be explained by the ease of communication between the nodes and the BS and the number of alive nodes over the lifetime of the WSN.

In Fig. 16, we show the number of alive nodes plotted against the number of data messages received by the base station. With its data aggregation approach, EPCOC provides better performance than LEACH and LEACH-C because this solution facilitates the transmission of data over the chains, reduces the quantity of data transmitted to the CHs, and distributes the load evenly among the cluster nodes. The CHs work less because each node aggregates the data before transmission.

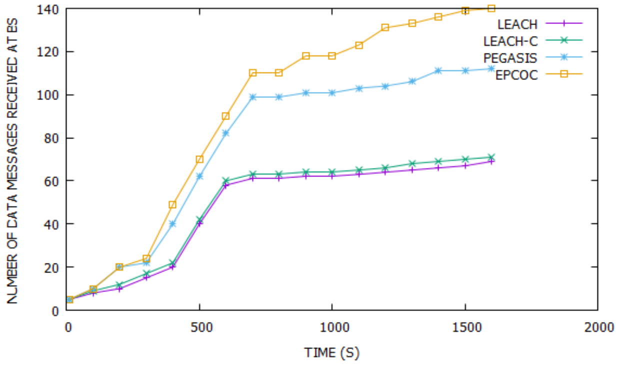

Fig. 15. Number of data messages received by the base station (BS) over time.

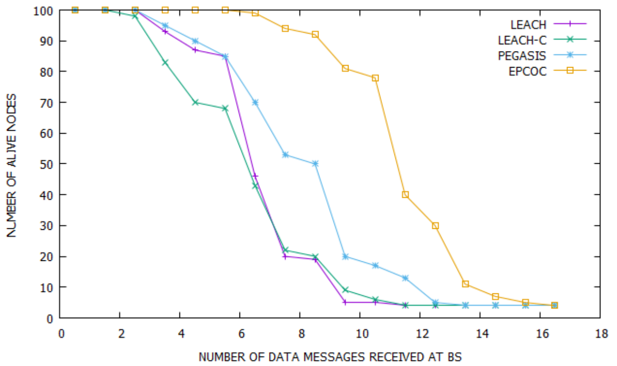

Fig. 16. Number of alive nodes plotted against the number of data messages received by the base station (BS).

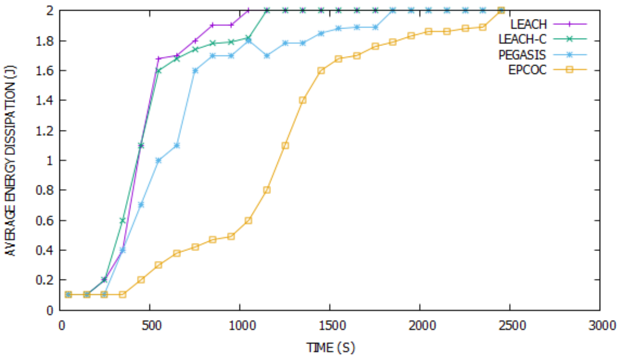

Fig. 17. Average energy dissipation over the network's lifetime.

Fig. 17 shows the average energy dissipation of the tested protocols over the network's lifetime, demonstrating the excellent results achieved by EPCOC. These results can be explained by the fact that EPCOC minimizes the consumption of energy by decreasing the communication distance to the CH and by aggregating data in the chains.

In data communications, the average time spent to send a packet with aggregated data to the BS is known as latency. The results obtained for the latency for each protocol are presented in Fig. 18. The plot shows the excellent results achieved by EPCOC in terms of the reduction in latency over PEGASIS, LEACH-C, and LEACH.

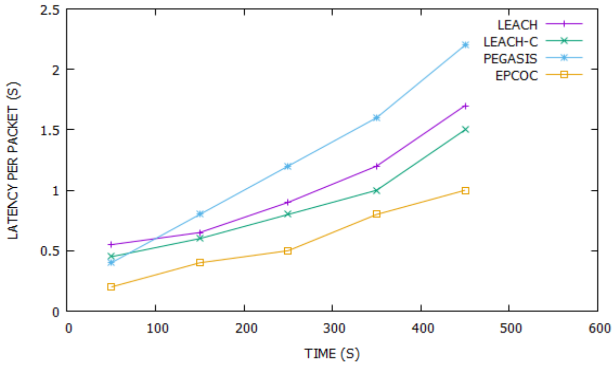

Fig. 18. Average latency per packet.

V. Conclusions

In wireless sensor networks, the main purpose in designing an energy efficient routing protocol is to increase the network lifetime. This paper presents an enhanced WSN protocol based on a hybrid topology using clustering and multiple chains. This proposed technique is influenced by resource constraints, such as energy, time, and computational complexity, as well as networking and architectural factors and network management issues.

The proposed algorithm is powered by the K-means algorithm along with a MapReduce framework to improve the clustering operation. A simulation showed excellent results achieved by the proposed protocol. EPCOC outperformed LEACH, PEGASIS, and LEACH-C not only in terms of network lifetime but also in total system energy consumption.

The main reason for this result is the more even distribution of energy consumption. The proposed protocol also solves the problem of high latency introduced by the long chain of nodes in the PEGASIS protocol and the poor energy dissipation in the LEACH protocol. However, there are areas where the performance of the proposed protocol can be further improved. Possibilities include the creation of localized chains, whereby all the chains are restricted to precise areas to enhance efficiency in order to avoid interferences, and the designing of a secure topology to provide security to the aggregated data in the WSN.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, no. 4, pp. 393-422, Mar. 2002.

[2] S. Zhang and H. Zhang, "A review of wireless sensor networks and its applications," in *Proc. IEEE International Conference on Automation and Logistics (ICAL)*, 2012, pp. 386-389.

[3] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: A survey," *Ad Hoc Networks*, vol. 7, no. 3, pp. 537-568, May 2009.

[4] J. Chen and F. Yu, "A uniformly distributed adaptive clustering hierarchy routing protocol," in *Proc. IEEE International Conference on Integration Technology*, 2007, pp. 628-632.

[5] O. Younis, M. Krunz, and S. Ramasubramanian, "Node clustering in wireless sensor networks: Recent developments and deployment challenges," *IEEE Network*, vol. 20, no. 3, pp. 20-25, June 2006.

[6] S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *Proc. IEEE INFOCOM 2003*, 2003, pp. 1713-1723.

[7] P. Sasikumar and S. Khara, "K-means clustering in wireless sensor networks," in *Proc. Fourth International Conference on Computational Intelligence and Communication Networks (CICN)*, 2012, pp. 140-144.

[8] Y. M. Lu and V. W. S. Wong, "An energy-efficient multipath routing protocol for wireless sensor networks," *International Journal of Communication Systems*, vol. 20, no. 7, pp. 747-766, July 2007.

[9] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Annual Hawaii International Conference on System Sciences (HICSS '00)*, 2000, pp. 1-10.

[10] Reena, S. Phulre, and R. Pandit, "Wireless sensor network and its routing protocol: A survey paper," *International Journal of Computer Science and Mobile Computing*, vol. 2, no. 6, pp. 60-64, June 2013.

[11] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: A survey," *IEEE Wireless Communications*, vol. 11, no. 6, pp. 6-28, Dec. 2004.

[12] S. K. Singh, M. P. Singh, and D. K. Singh, "A survey of energy-efficient hierarchical cluster-based routing in wireless sensor networks," *International Journal of Advanced Networking and Applications*, vol. 2, no. 2, pp. 570-580, Aug. 2010.

[13] Neetika and S. Kaur, "Review on hierarchical routing in wireless sensor networks," *International Journal of Smart Sensors and Ad Hoc Networks*, vol. 2, nos. 3-4, pp. 85-90, 2012.

[14] P. Kaur, M. Katiyar, "The energy-efficient hierarchical routing protocols for WSN: A review," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 11, pp. 194-199, Nov. 2012.

[15] J. Gnanambigai, N. Rengarajan, and K. Anbukkarasi, "Leach and its descendant protocols: A survey," *International Journal of Communication and Computer Technologies*, vol. 1, no. 3, pp. 15-21, Sep. 2012.

[16] A. Braman and G. R. Umapathi, "A comparative study on advances in LEACH routing protocol for wireless sensor networks: A survey," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, no. 2, Feb. 2014.

[17] F. Zhao, Y. Xu, and R. Li, "Improved LEACH routing communication protocol for a wireless sensor network," *International Journal of Distributed Sensor Networks*, vol. 8, no. 12, Article ID 649609, Dec. 2012.

[18] R. V. Biradar, S. R. Sawant, R. R. Mudholkar, and V. C. Patil, "Multihop routing in self-organizing wireless sensor networks," *IJCSI International Journal of Computer Science Issues*, vol. 8, no. 1, Jan. 2011.
[19] V. Kumar, S. Jain, and S. Tiwari, "Energy efficient clustering algorithms in wireless sensor networks: A survey," *International Journal of Computer Science Issues*, vol. 8, no. 5, pp. 282-291, Sep. 2011.
[20] X. H. Wu and S. Wang, "Performance comparison of LEACH and LEACH-C protocols by NS2," in *Proc. Ninth International Symposium on Distributed Computing and Applications to Business Engineering and Science (DCABES)*, 2010, pp. 254-258.
[21] P. Bakaraniya and S. Mehta, "K-LEACH: An improved LEACH protocol for lifetime improvement in WSN," *International Journal of Engineering Trends and Technology*, vol. 4, no. 5, pp. 1521-1526, May 2013.
[22] H. Rana, S. Vhatkar, and M. Atique, "Comparative study of PEGASIS protocols in wireless sensor network," *IOSR Journal of Computer Engineering*, vol. 16, no. 5, ver. I, pp. 25-30, Sep.-Oct. 2014.
[23] S. Vhatkar and M. Atique, "Design issues, characteristics and challenges in routing protocols for wireless sensor networks," *International Journal of Computer Applications*, vol. 1, no. 2, pp. 42-47, Apr. 2013.
[24] S. Jeon, B. Hong, J. Kwon, Y. Kwak, and S. Song, "Redundant data removal technique for efficient big data search processing," *International Journal of Software Engineering and Its Applications*, vol. 7, no. 4, pp. 427-436, Jul. 2013.
[25] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107-113, Jan. 2008.
[26] M. Greis. Tutorial for the Network Simulator NS2. [Online]. Available: http://www.isi.edu/nsnam/ns/tutorial/

Ilies Ali Lekhal received his Master's degree in systems, networks, and database technologies at the University of Sciences and Technology of Oran – Mohamed Boudiaf, Algeria, in 2012. Currently, he is a Ph.D. student in the Computer Science Department at the same university. His research interests include wireless networks, embedded systems, artificial intelligence, multi-agent systems, and distributed computing.

Samira Chouraqui has been a full professor in the Department of Mathematics and Informatics at the University of Sciences and Technology of Oran – Mohamed Boudiaf (USTO-MB) since January 2017. She received her M.Sc. degree in satellite communications in 2002 at Surrey University in the UK and her Science Doctorate degrees in computer engineering and applied mathematics in 2010 at USTO-MB. Her current research interests focus on the intelligence of dynamic systems such as robots and vehicles.