Egyptian Informatics Journal 23 (2022) 483–497

Contents lists available at ScienceDirect

Egyptian Informatics Journal

journal homepage: www.sciencedirect.com

A weighted Markov-clustering routing protocol for optimizing energy use in wireless sensor networks

Leila Abbad [a,*], Azzedine Nacer [b], Houda Abbad [c], Mohammed Taieb Brahim [c], Nadjet Zioui [d]

[a] LIT Laboratory, Electrical Engineering and Industrial Computing Department, ENST High School, Algiers, Algeria
[b] LCDEP Laboratory, Instrumentation Department, University of Science and Technology Houari Boumediene, Algiers, Algeria
[c] EEDIS Laboratory, Computer Science Department, Djillali Liabes University of Sidi Bel Abbes, Sidi Bel Abbes, Algeria
[d] Laboratoire d'Intelligence Artificielle Appliquée (LI2A), Université du Québec à Trois-Rivières, QC, Canada

ARTICLE INFO

Article history:
Received 18 February 2022
Revised 30 April 2022
Accepted 14 May 2022
Available online 27 May 2022

Keywords:
WSN
Markov clustering (MCL)
Sensor abundance
Sensor distribution
Routing protocol
Energy consumption
Network lifetime

ABSTRACT

Interest has increased in wireless sensor networks (WSNs) in fields such as healthcare, industrial control and environmental monitoring. More recently, WSNs have been widely deployed in Internet of Things (IoT) based applications. They are considered as an essential part of IoT networks. The expanded use of WSNs raises daunting technical challenges, not the least of which is sensor battery energy conservation. The design of efficient sensor clustering strategies to reduce energy consumption by data transmission throughout the WSN has become crucial. Several applications of WSNs induce a random deployment of sensors such as the monitoring of conflict zones, the study of natural phenomena in hostile zones or rescue operations. In this context, energy conservation should be ensured according to the non-uniformity of the resulting sensor distribution in different areas of the network. In this study, a novel weighted Markov clustering protocol that considers sensor abundance for cluster-head and queried sensor selections is presented. The new protocol aims to decrease intra-cluster energy consumption by reducing the sending of redundant data in sensor-dense regions. In addition, it attempts to prolong sensor-sparse regions lifetime by limiting the number of queried sensors. This protocol combines a Markov clustering of sensors with a sensor weighting based on residual energy and sensor abundance in the network. The proposed protocol is a significant improvement of an existing unweighted Markov clustering protocol. The unweighted Markov clustering protocol is based on sensor residual energy and sensor location without taking into account the sensor abundance in different areas of the network. Simulations affirm that the new protocol handles more appropriately the non-uniformity of sensor distribution and enhance the durability of wireless sensor networks. Indeed, simulation results show that the proposed clustering protocol outperforms its unweighted ancestor and other well-known clustering protocols in terms of energy conservation and network lifetime. The number of expired sensors and the average dissipated energy are reduced, whereas, the average sensor lifetime is prolonged compared to the unweighted ancestor, or HEED, LEACH, PEGASIS, and TEEN.

1. Introduction

A wireless sensor network (WSN) is an information retrieval system consisting of many sensors distributed over an area of interest [1–10]. The role of each sensor is to collect and transmit data to a base station, which then relays the data to some location for processing. Sensors have a limited energy supply and are often distributed randomly throughout the area covered by the network. This non-uniformity of distribution density produces differences in the amount of data transmitted from each sector. High-density sectors collect highly correlated data whereas low-density sectors have coverage problems due to premature expiry of isolated sensors. Wireless sensor networks have been studied in very different fields such as military [8], environmental monitoring [9], industrial process control [10], domestic monitoring [11–15] and recently the Internet of Things [16–18]. A concern in all WSNs is sensor battery lifetime, which depends on energy consumption by data transmission within the network [19]. The design of low-energy-consumption routing protocols is therefore a major area of R&D. The greatest gains in network longevity so far have been achieved by developing hierarchical (clustering) routing protocols [20–25],

* Corresponding author.
E-mail addresses: leila.abbad@enst.dz, leila.abbad@enst.dz (L. Abbad).

https://doi.org/10.1016/j.eij.2022.05.001
1110-8665/

Nomenclature

Notation

G	Graph
M	Stochastic matrix
D	Diagonal matrix of the matrix M
T_G	Markov matrix
Inf_r	Inflation operator
e	Expansion power (coefficient)
r	Inflation power (coefficient)
E_{elec}	Energy consumed to run transmitter or receiver circuitry
E_{fs}	Amplifier energy needed to transmit signals in a free space model
E_{amp}	Amplifier energy needed to transmit signals in a multipath model
d_0	Threshold distance of the energy model
$E_{T_X}(k,d)$	Energy consumed to transmit a k-bits packet over a distance d
E_{R_X}	Energy consumed to receive a k-bits packet
E_{DA}	Energy consumed for one-bit data aggregation
dis_{min}	Minimal Euclidean distance between sensors in the network
dis_{max}	Maximal Euclidean distance between sensors in the network
dis_{mean}	Average distance of the network
$dis_{threshold}$	Threshold distance of the network
n	Number of sensors in the network
S_i	Sensor number i
dis_{ij}	Distance between sensor S_i and sensor S_j
(x_{s_i}, y_{s_i})	Location coordinates of the sensor S_i
$M_{euclid}(i,j)$	Euclidean distance between sensor S_i and sensor S_j
K	Threshold distance parameter
m	Number of clusters in the network
C_i	Cluster number i
$\|C_i\|$	Number of sensors in the cluster C_i
CH_i	Cluster-head of the cluster C_i
QS_i	Queried sensors of the cluster C_i
N_{s_i}	Number of alive neighbors of a sensor S_i in a specified range
d_{s_i}	Density of a sensor S_i
$d_{threshold}$	Network density threshold
$E_{\exp(i)}$	Minimal residual energy to announce the expiry of sensors in a cluster C_i
E_{s_i}	Residual energy of a sensor S_i
w_{s_i}	Weight of a sensor S_i
$x\%$	Threshold defining tiny clusters
$lp\%$	Percentage of selectable queried sensor candidates in low-density regions
$hp\%$	Percentage of selectable queried sensor candidates in high-density regions

Abbreviations

WSN	Wireless sensor network
MCL	Markov Cluster Algorithm
MCL-BCRP	MCL-Based Clustering Routing Protocol
WMCL-BCRP	Weighted MCL-Based Clustering Routing Protocol
LEACH	Low-energy adaptive clustering hierarchy
LEACH-C	Centralized LEACH
VLEACH	Vice cluster-head LEACH
MH-LEACH	Multi-hop LEACH
TL-LEACH	Two-level LEACH
ESO-LEACH	Enhanced swarm optimization LEACH
μGA-LEACH	micro genetic algorithm LEACH
LEACH-VD	LEACH vector quantization
F-LEACH	Fuzzy LEACH
TEEN	Threshold sensitive energy efficient sensor network
APTEEN	Adaptive threshold-sensitive energy-efficient sensor network
PEGASIS	Power-efficient gathering in sensor information systems
HEED	Hybrid energy-efficient distributed
TDMA	Time division multiple access

which allow sensors to communicate directly with nearby cluster-heads rather than forward all data to a more remote base station.

Clustering routing protocols must seek a compromise between extending network lifetime and assuring effective coverage of the area of interest [26]. Many studies have focused on improving network coverage efficiency, scalability, and data routing with minimal energy consumption [2]. The most common clustering routing protocols for wireless sensor networks are LEACH [24] and its descendants TEEN [27], APTEEN [28], PEGASIS [29] and HEED [30]. These routing protocols differ in the means used to form clusters, select cluster-heads, and route sensor data to the base station. Recently, a Markov clustering routing protocol has been proposed (MCL-BCRP) [31], in which the Markov clustering algorithm (MCL) [32] is used to form clusters and select cluster-heads based on sensor location and residual energy. In simulations, this protocol outperforms HEED, LEACH, PEGASIS and TEEN in terms of network lifetime and network coverage.

This manuscript presents an improved version of the Markov clustering protocol. The proposed version is called the Weighted Markov clustering protocol (WMCL-BCRP) where the cluster-head selection is based on weighting sensors. The weighting is applied according to the abundance and residual energy of the sensors. This extended version handles sensor distribution randomness and non-uniformity more suitably in order to improve the resilience of monitoring networks in settings subject to chaotic sensor distributions, for example, battlefields or conflict zones, volcanic activity or other inhospitable areas, rescue operations among others. In all these cases, a drone or a helicopter can launch sensors randomly, and energy consumption can be managed according to the resulting sensor abundance in the different sectors of the network. In addition, the sending of redundant data in sensor-dense regions can be reduced and sensor-sparse regions can be maintained for longer by reducing the network average energy consumption. The number of expired sensors remains manageable, and the network lifetime is prolonged compared to the unweighted protocol (MCL-BCRP), or HEED, LEACH, PEGASIS, and TEEN. The new protocol proposes a restriction on queried sensor assignment to cluster-heads based on distinguishing high sensor density and low sensor density sectors. It also adapts the unweighted Markov protocol (MCL-BCRP) [31], which simulates stochastic flows in graphs by transforming WSNs into stochastic matrices as inputs to provide fast and configurable clustering.

2. Related studies

Hierarchical routing has proven its effectiveness in terms of the scalability and efficiency and especially the optimization of energy consumption by sensor network communications [20–25]. Hierarchical routing is also known as clustering routing since clustering is the main operation underlying the expected increase in energy efficiency. In clustering routing, sensors join clusters according to certain criteria, and some sensors are elected as cluster-heads. Cluster-heads process and transmit sensor readings directly to the base station when it is nearby, otherwise to other cluster-heads, which transmit the data until it reaches the base station [21,23].

Clustering involves two processes: cluster formation and cluster-head selection. The former groups sensors, whereas the latter designates a sensor as group-head according to precisely measurable criteria such as location, density, connectivity, battery energy level, and so on. The head aggregates the data transmitted by sensors in its cluster and transmits the aggregated data to the base station according to a routing scheme also based on measurable criteria. Clustering can be static or dynamic, undergoing updates to balance sensor energy dissipation and thereby increase network longevity. The algorithm used for head selection can be centralized and run at the base station level, distributed at the sensor level, or hybrid. It can also be deterministic, random, or adaptive. Deterministic schemes fix head selection, whereas adaptive schemes select heads on an evolving rational basis [22,23]. Clustering routing protocols designed for WSNs are the fruit of much research on energy efficiency management [33–38]. We summarize below the most studied of these, based on two decades of technical and scientific literature.

Low-energy adaptive clustering hierarchy or LEACH [24] is the most popular clustering protocol devised for WSN applications. It was the first to use hierarchical structuring to organize network sensors into clusters. Head sensors selected randomly are responsible for collecting and aggregating data from the other (nearby) sensors in the cluster. Each head transmits data to the base station and its TDMA schedule to its sensors to organize intra-cluster communication. LEACH has drawbacks, for example, head selection is random, without consideration of sensor residual energy or other adaptive criteria. It is presumed that all heads can reach the base station. When a head becomes non-operational, the network becomes non-functional until a new head is elected. Variants have been proposed to overcome these deficiencies: LEACH-C, a centralized version in which the base station selects a group of cluster-heads from candidates with the highest residual energy using the simulated annealing algorithm [39], VLEACH, which elects heads the same way LEACH does but at the same time designates a vice-head to take over when the principal head power supply is exhausted [40], MH-LEACH and TL-LEACH, which allow multi-hop transmissions between cluster-heads and the base station [41,42]. More recent enhancements include ESO-LEACH, which proposes a *meta*-heuristic particle swarm for cluster formation using the concept of advanced nodes and a more sophisticated set of rules for head selection [43], μGA-LEACH, which uses a micro genetic algorithm to compute an optimal set of cluster-heads [44], LEACH-VD, which explores the shortest paths between all cluster-heads using vector quantization and the DIJKSTRA algorithm [45], and F-LEACH, which introduces a fuzzy-based data aggregation algorithm to prolong network lifetime [46].

PEGASIS provides some improvements over LEACH with the creation of a hierarchical structure. This structure relies on chains instead of clusters [29]. With a nearest neighbor search algorithm, the base station forms chains then notifies each sensor with some information related to its position and direct neighbors in the network. One of the benefits of this protocol is that the chain reconstructs itself when a node vanishes. Collected data are transmitted and aggregated successively in the chain until the leader chain is reached, which transmits the data to the base station. Because there is a single leader chain, PEGASIS is subject to the problem of bottlenecks.

Threshold-sensitive energy-efficient network (TEEN) protocol is suitable for critical real-time reactive monitoring applications where sensors detect events continuously but transmit no data for long periods [27]. Hierarchical structures are formed where in the sensors in closest proximity are grouped into clusters. Clustering progresses to other levels until it reaches the base station. Cluster-heads do not send a TDMA scheduler to their members but instead a control message containing an attribute, hard and soft thresholds. The attribute represents the type of data being monitored (e.g., temperature), the hard threshold is the critical value beyond which readings must be reported to the cluster-head, and the soft threshold is the smallest change that must be reported. Once a value exceeding the hard threshold is reported, no additional reading is reported unless the difference exceeds the soft threshold. This reduces redundancies and transmissions.

The hybrid energy-efficient distributed or HEED protocol [30] is another extension of LEACH, based on two new metrics for cluster-head selection: sensor residual energy and the cost (energy) of inter-cluster communications. During the initialization phase, the probability of selecting a sensor as a head is computed from the residual energy and the optimal number of clusters. Simulations show that this protocol increases network longevity compared to LEACH because cluster-heads are not selected randomly.

Markov-clustering-algorithm-based routing (MCL-BCRP) [31] is a new protocol based on two novel features, the first comprising segmenting the network into clusters to ensure adequate coverage [32] and the second selecting cluster-heads based on sensor residual energy and location relative to the base station and the cluster centroids. This protocol offers significant improvement in network longevity and coverage compared to LEACH, PEGASIS, TEEN and HEED.

In the present article, we present an improvement of Markov-clustering-based routing, called weighted Markov-clustering-based routing (WMCL-BCRP), based on selection of cluster-heads and sensors in a way that extends the lifetime of networks characterized by non-uniform distribution of sensors. Its adaptive cluster-head selection is based mainly on two parameters: sensor residual energy and distribution density. This weighting at each round makes all sensors in the network potential candidates for selection as cluster-heads. Based on abundance, the lowest-weighted sensors are selected as queried sensors. By distinguishing low and high abundance (density) regions within the network, sparse regions are kept alive for as long as possible and intra-cluster energy consumption is reduced by bypassing redundant data transmissions in dense regions.

Table 1 presents a conceptual comparison between the studied protocols according to main features of hierarchical routing protocols for wireless sensor networks:

3. Energy consumption model

The basic energy consumption model applied to data transmissions in wireless sensor networks has been defined elsewhere [24]. The same model is used in the present study. The sensor radio consumes E_{elec} to run its transmitter or receiver circuitry. The energy required to amplify transmitted signals in the free space model and in the multipath model is denoted respectively E_{fs} and E_{amp}. Thus, to transmit a k-bit message over a distance d, the communication unit expends:

$$E_{T_X}(k,d) = \begin{cases} E_{T_X-elec}(k) + E_{T_X-fs}(k,d), & d < d_0 \\ E_{T_X-elec}(k) + E_{T_X-amp}(k,d), & d \geq d_0 \end{cases} \quad (1)$$

$$= \begin{cases} E_{elec} * k + E_{fs} * k * d^2, & d < d_0 \\ E_{elec} * k + E_{amp} * k * d^4, & d \geq d_0 \end{cases}$$

where $d_0 = \sqrt{E_{fs}/E_{amp}}$ is a threshold distance.

To receive a k-bit message, the communication unit expends:

$$E_{R_X}(k) = E_{R_X - elec}(k)$$

$$= E_{elec} * k = E_{elec} * k \tag{2}$$

In this model, the energy spent on data processing is negligible compared to the expenditure on data transmission. However, data aggregation by each cluster-head is an energy-consuming operation, and for a k-bit message, it is computed as follows:

$$E_{DA}(k) = E_{DA} * k \tag{3}$$

where E_{DA} is the energy consumed per bit for data aggregation in each cluster-head.

4. The proposed approach

4.1. The Markov clustering algorithm

The main concepts used in the Markov graph-based clustering algorithm (MCL) are described below. Graph approaches to clustering involve arranging the nodes of a graph into clusters by drawing links between them. Developed by Stijn Van Dongen [32], Markov clustering (MCL) has been applied in several fields, especially bioinformatics, where it is used to shed light on the biological function of genes and estimate the significance of gene expression. This algorithm does not require that the number of clusters be known in advance with comparison to other graph-based clustering algorithms and unlike k-means clustering [47]. The main idea of Markov clustering is that if a walk starts randomly from one node to another, there is a greater chance of walking within the same cluster than moving to another cluster, since clusters are, by definition, separated by spaces of node sparseness. Clusters therefore can be discovered by exploring changes in the node density of the flow space. Random walks are determined by transforming the flow through the graph to a Markov chain. From a transition matrix as input to the algorithm, stochastic matrices are generated by alternating two complementary mathematical operators, namely expansion and inflation, until a convergence matrix is obtained [32].

The expansion operator expands the flow by computing the powers of the stochastic matrices and enumerating random walks of length greater than 1, which are more frequent within than between clusters. In other words, the probabilities assigned to pairs of vertices belonging to a cluster are relatively large because there are many ways to move among these vertices.

The role of the inflation operator is to increase the walk probabilities within clusters and to decrease them between clusters. In other words, inflation increases flow within clusters and decreases it between them. The inflation operator also allows control of the granularity of the clusters generated by the Markov algorithm [32].

Given a stochastic matrix $M \in R^{k \times l}$ of dimension $k \times l$, and a positive real number e, the application of the expansion operator Exp of coefficient (or power) e to the matrix M is a stochastic matrix $Exp_e(M) \in R^{k \times l}$ defined formally as follows:

$$Exp_e(M) = M^e \tag{4}$$

Applying the inflation operator Inf with coefficient r to the stochastic matrix M yields a stochastic matrix $Inf_r(M) \in R^{k \times l}$ by rescaling each column of M by the coefficient (or power) r. The matrix operator Inf_r is thus defined formally as follows:

$$Inf_r(M)_{pq} = \frac{(M)_{pq}^r}{\sum_{i=1}^{k}\left((M)_{iq}^r\right)} \tag{5}$$

The matrix M associated with a graph G is defined as $M_{ij} = 1$ if there is a link relating i and j, otherwise as $M_{ij} = 0$. The Markov matrix T_G for graph G is therefore deduced from its associated matrix M and is defined as follows:

$$T_G = MD^{-1} \tag{6}$$

Such as D is a diagonal matrix of the matrix M. Formally, $D_{ii} = \sum_{k=1}^{n}(M_{ki})$ and $D_{ij} = 0$, with $i \neq j$. The basic pseudo-code of the Markov clustering process is presented in the following [31]:

Algorithm 1: MCL(T_G, $e_{(i)}$, $r_{(i)}$)

```
1: T_1 = T_G
2: for k in 1...∞ do
3:     T_2k = Exp_{e_k}(T_{2k-1})
4:     T_{2k+1} = Inf_{r_k}(T_2k)
5:     if T_{2k+1} is doubly idempotent then
6:         break
7:     end if
8: end for
9: Interpret T_{2k+1} as a clustering of the graph G
```

As mentioned above, the MCL algorithm alternates two matrix operators, namely inflation and expansion, until a steady state is reached whereby a stochastic doubly idempotent matrix is obtained with respect to these two complementary operators. The MCL algorithm has not been proven to converge in theory. However, it has been shown that if convergence were to occur, this would be quadratic in a neighborhood of doubly idempotent matrices. In practice, the algorithm usually converges within the first 20 iterations. To extract clusters from the convergence matrix, the set of vertices is divided into attractor and simple vertices. Attractors are defined as vertices th-at attract simple vertices to form clusters. They have

Table 1
Conceptual comparison between LEACH, PEGASIS, HEED, TEEN, MCL-BCRP and WMCL-BCRP protocols.

Protocol	CH selection					Cluster formation	
	Algorithm D: Distributed, C: Centralized	Scenario R: Random, A: Adaptive	E: CH Election, S: CH Selection	Overhead	R: Based on Residual Energy, L: Based on Location, N: Based on Neighbor D: Based on Density	Algorithm D: Distributed, C: Centralized	Frequency ER: Each Round, ED: Each Attractor death
LEACH	D	R	E	Low	–	D	ER
TEEN	D	A	E	High	–	D	ER
PEGASIS	D	A	S	–	–	D	ER
HEED	D	A	E	High	R & N	D	ER
MCL-BCRP	D	A	S	Low	R & L	C	ED
WMCL-BCRP	D	A	S	Low	R & D	C	ED

positive flux values in their corresponding rows within the doubly idempotent matrix, and each simple attracted vertex has a positive value in its corresponding attractor row [32].

4.2. Weighted Markov-clustering routing working process

The aim of the weighted Markov clustering protocol (WMCL-BCRP)is mainly to improve energy management in wireless sensor networks and thereby extend their lifetime, which represents a significant enhancement of the unweighted protocol (MCL-BCRP) [31]. A hybrid strategy is proposed for the cluster formation and cluster-head selection processes. Centralized cluster formation by the MCL algorithm comprises gathering neighboring sensors into clusters based on a threshold distance. Cluster-head selection is an adaptive and a distributed process based on two criteria: sensor residual energy and abundance. These two parameters define a new notion, namely the sensor weighting factor. At each round, candidates with the highest weights are selected as cluster-heads. This selection requires little overhead and thus reduces energy consumption within clusters. Selection of cluster-heads in dense regions (high-density candidates) saves sensor residual energy since the distances between sensors and their cluster-heads are generally short. Unlike most protocols, the weighted Markov clustering protocol allows only a fraction of the sensor members of each cluster, called queried sensors, to send data to the head, as illustrated in Fig. 1. The queried sensors are chosen based on their weights and sensor abundance. This restriction on data transmission to cluster-heads reduces intra-cluster energy consumption at each round. Furthermore, sensors in dense regions would probably send redundant data to their cluster-heads, due to their proximity. Cluster-heads are thus relieved from unnecessary aggregations on strongly correlated data. On the other hand, sensors in low-density regions must send readings frequently to provide adequate coverage since they may detect events that are not sensed in dense regions. The weighted algorithm is thus intended to keep sensors in sparse regions alive for as long as possible by selecting just a fraction of them for transmitting data to cluster-heads.

As shown in Fig. 2, the weighted protocol is designed to function in two main phases:

- A centralized configuration phase run by the base station and triggered each time an attractor dies, requiring reconfiguration of the network.
- A distributed communication phase representing all data transmissions in the network, from queried sensors to their cluster-heads and from cluster-heads to the base station.

4.2.1. Configuration phase

Configuration is the preliminary phase that is run by the base station to organize the hierarchy within the wireless sensor network. This must be completed before communications between sensors and the base station can begin. Configuration consists of three sub-phases: network matrix representation, Markov clustering, and configuration broadcast.

4.2.1.1. Network matrix representation sub-phase. The network matrix is the main input to the MCL algorithm. Using GPS modules, sensors transmit their locations to the base station, and identifiers are assigned to them. The GPS module is disabled once the sensor location has been transmitted. The Euclidean distances between all network sensors are then computed. Mean distance is then calculated as follows:

$$dis_{mean} = (dis_{min} + dis_{max})/2 \quad (7)$$

such that dis_{min} and dis_{max} represent respectively the minimal and the maximal Euclidean distances between the sensor nodes. The threshold distance is a function of sensor dispersion over the sensing- area, and K is a positive integer than can be chosen empirically, such that.

$$dis_{threshold} = dis_{mean}/K \quad (8)$$

Finally, matrix M associated tothe wireless sensor network constituted of n sensors is square, of dimension $n \times n$, defined as follows:

$$M_{ij} = \begin{cases} 1, dis_{ij} \le dis_{threshold} \\ 0, dis_{ij} > dis_{threshold} \end{cases} \quad (9)$$

where $i,j = 1\ldots n$.

The choice of the threshold distance is crucial to guarantee low energy dissipation during intra-cluster communications. Indeed, this distance determines whether two sensors are in the same

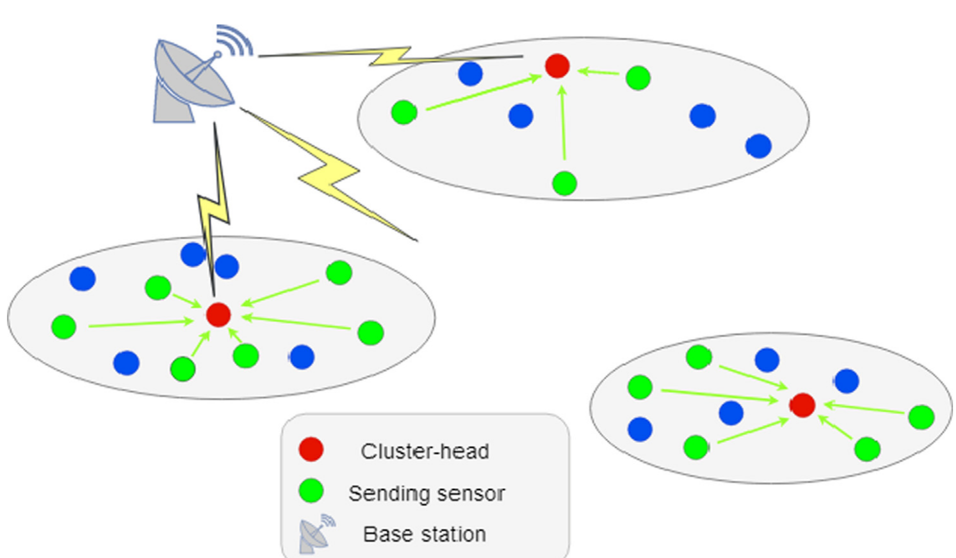

Fig. 1. Weighted markov clustering protocol architecture.

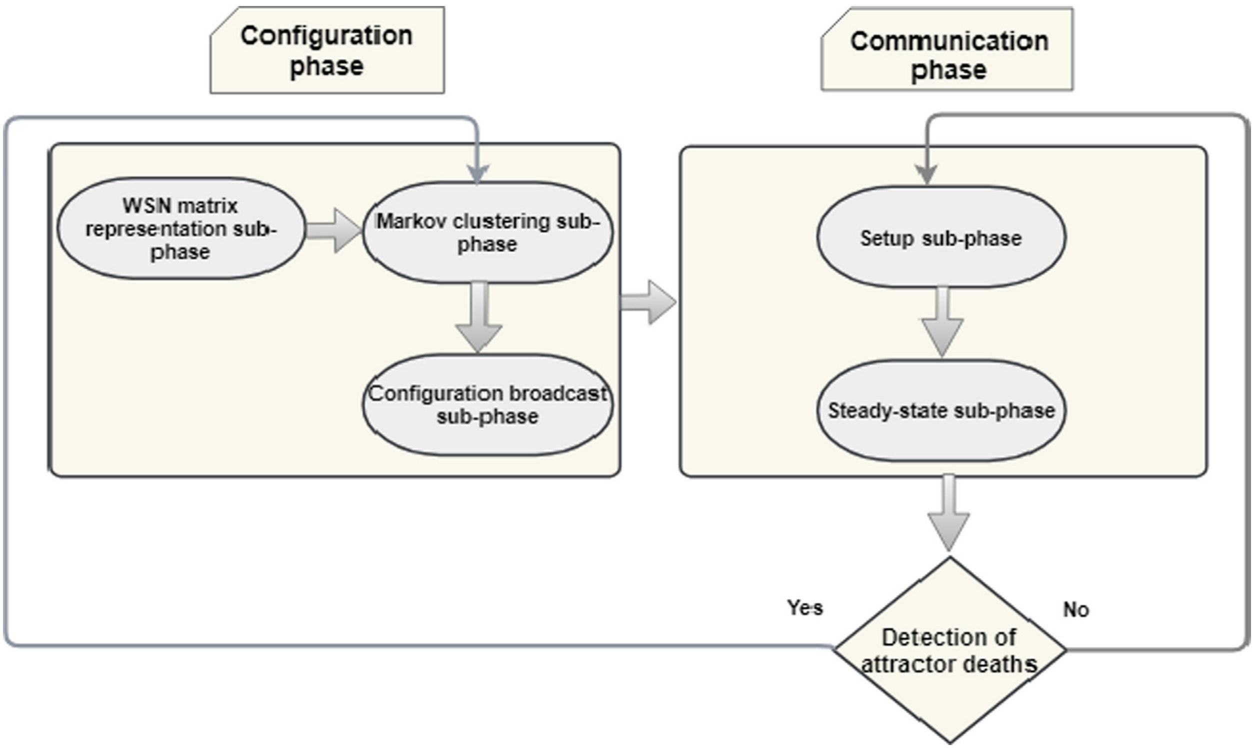

Fig. 2. Flowchart of the weighted Markov clustering functional protocol.

neighborhood or not and can communicate at low energy expenditure. The underlying algorithm of the network matrix representation on a wireless sensor network *WSN* is the following:

Algorithm 2: Network Matrix Representation(*WSN*)

for i in $1\ldots n$ do
  Receive the location coordinates $(\boldsymbol{x}_{s_i}, \boldsymbol{y}_{s_i})$ of the sensor S_i
end for
for i in $1\ldots n$ do
  for j in $1\ldots n$ do
    $M_{\text{euclid}}(\mathrm{i},\mathrm{j}) = \sqrt{\left(\mathrm{x}_{\mathrm{s_i}} - \mathrm{x}_{\mathrm{s_j}}\right)^2 + \left(\mathrm{y}_{\mathrm{s_i}} - \mathrm{y}_{\mathrm{s_j}}\right)^2}$
  end for
end for
$dis_{min} = min(M_{euclid}(i,j))_{1\le i,j\le n}$
$dis_{max} = max(M_{euclid}(i,j))_{1\le i,j\le n}$
$dis_{mean} = \frac{(dis_{min}+dis_{max})}{2}$
$dis_{threshold} = \frac{dis_{mean}}{k}$
for i in $1\ldots n$ do
  for j in $1\ldots n$ do
    $M(i,j) = \begin{cases} 1, & \text{if } dis_{ij} \le dis_{threshold} \\ 0, & \text{if } dis_{ij} > dis_{threshold} \end{cases}$
  end for
end for

4.2.1.2. *Markov clustering sub-phase.* The Markov clustering sub-phase consists of applying the MCL algorithm to structure the network as clusters, by computing the Markov matrix induced from the matrix *M*. The MCL algorithm produces different clusters depending on sensor distribution over the sensing area. Simulations show that this algorithm rarely gives overlapping clusters. Such clusters occur only in symmetrical networks, which rarely exist in practice. In any case, the base station can overcome this problem by attributing overlapping sensors to clusters with the nearest attractor. Once the clusters are formed, the base station checks whether they are of a reasonable size relative to the network. Since simulations have shown that very small clusters tend to die prematurely, the base station rejects all clusters smaller than a specified percentage of network size and assigns their sensors to neighboring clusters with the nearest attractor. The pseudo-code of the Markov clustering sub-phase is presented in the following:

Algorithm 3: Markov Clustering(*M*)

for i in $1\ldots n$ do
  for j in $1\ldots n$ do
    $D(i,j) = \begin{cases} \sum_{k=1}^{n}(M_{ki}), & if \quad i \ne j \\ 0, & if \quad i = j \end{cases}$
  end for
end for
$T_G = MD^{-1}$
$MCL(T_G, e_{(i)}, r_{(i)})$
Extract the set $(C_i)_{1\le i\le m}$ of clusters from the convergence matrix
Resolve overlappings in $(C_i)_{1\le i\le m}$
Reject clusters in $(C_i)_{1\le i\le m}$ smaller than $x\%$ of the network size
Assign sensors of rejected clusters to neighboring clusters with the nearest attractor

4.2.1.3. *Configuration broadcast sub-phase.* For each cluster, the base station computes a minimal energy below which a sensor must announce its expiry to rest of the cluster (equation (1)). In the configuration broadcast sub-phase, the cluster-head and queried sensor designations are based on sensor abundance. For this purpose, the base station calculates a threshold value that distinguishes sensor-sparse and sensor-dense regions. It then attributes to each sensor an initial density as follows:

$$d_{s_i} = 1 + (N_{s_i}/n) \tag{10}$$

where N_{s_i} is the number of alive neighbors of sensor S_i within a specified range, and $i = 1 \ldots n$. It is presumed that all network sensors are initially live. This definition implies that the density associated with each sensor decreases as its neighbors die off, down to 1, when no live sensors remain within the specified range. The network density threshold is deduced as follows:

$$d_{threshold} = \frac{\sum_{i=1}^{n} d_{s_i}}{n} \tag{11}$$

This threshold is updated each time the network is reconfigured. Finally, the base station broadcasts to all sensors a configuration message containing the parameters previously determined such as the sensor identifiers and locations, the set of clusters as well as their estimated threshold energies for expiry announcements, and the network density threshold. The following algorithm summarizes the successive steps of the configuration broadcast sub-phase:

Algorithm 4: Configuration broadcast($(\mathbf{C_i})_{1 \le \boldsymbol{i} \le \boldsymbol{m}}$)

```
for i in 1...m do
    compute E_exp(i) the minimal energy
  to announce the expiry of sensors in C_i
end for
for i in 1...n do
d_si = 1 + (N_si / n)
end for
d_threshold = (sum_{i=1}^{n} d_si) / n
configuration_message = [d_threshold]
for i in 1...n do
  configuration_message = configuration_message +
  [(S_i,(x_si, y_si))]
  end for
  for i in 1... m do
  configuration_message = configuration_message +
  [(C_i,E_exp(i))]
  end for
  Broadcast configuration_message to all sensors of the
  network
```

4.2.2. Communication phase

In hierarchical structure WSNs, two modes of communication can be distinguished: inter-cluster and intra-cluster. In intra-cluster mode, communication between the different sensors and their cluster-heads is enabled by the TDMA access method. Each sensor is assigned a time slot to transmit readings to the cluster-head. In intra-cluster mode, the CDMA access method is used instead to allow different cluster-heads to transmit aggregated data to the base station. The communication phase is carried out as repetitive cycles of fixed duration called rounds. Each round comprises a setup sub-phase and a steady-state sub-phase.

4.2.2.1. Setup sub-phase. During setup, five main tasks are accomplished before proceeding to communication:

- Residual energy broadcast: Each sensor announces its residual energy by broadcasting it according to the previous TDMA scheduling or announces its expiry when its residual energy reaches the threshold calculated during configuration. Sensors in each cluster thus update the residual energies and hence their cluster table by deleting members declared expired.

- Computing weights: Each sensor computes its current weight to submit its candidacy as a cluster-head or a queried sensor for the current round. Its current density is computed using equation (11), and the weight then calculated from the density and residual energy as follows:

$$w_{s_i} = E_{s_i} * d_{s_i} \tag{12}$$

where E_{s_i} is the residual energy of the sensor S_i, and $i = 1 \ldots n$. The highest weights correspond to the sensors with the highest residual energies in high-density regions, and the lowest weights to the sensors with the lowest residual energies in the low-density regions.

- Cluster-head selection: The highest weighted sensor in a cluster is selected as the head. This simple selection process does not require any overhead.
- Queried sensors selection: The main purpose of queried sensor selection is to reduce power consumption inside clusters. It also reduces sending of strongly correlated data to cluster-heads in high-density regions and thereby extends the coverage of low-density regions. This selection is based on sensor regional density. In low-density regions ($d_{s_i} \le d_{threshold}$) only sensors among *lp%* of the weakest in the cluster can send data, while only *hp%* of the weakest in high-density regions ($d_{s_i} > d_{threshold}$) can send data to cluster-heads. The low-density and high-density queried candidate percentages *lp%* and *hp%* are crucial for the coverage and the energy management of the network. It can be observed easily that a high *hp%* reduces the average sensor lifetime in high-density regions, while a small *lp%* limits the coverage of low-density regions.
- TDMA scheduling: Each time a cluster table is updated after sensor expirations, the cluster-head broadcasts a new TDMA schedule to its members before starting the next sub-phase. Each member then communicates with its cluster-head during its allocated time slot.

The previous distributed actions, which are executed by all sensors of the network during the setup sub-phase, are formulated in Algorithm 5:

Algorithm 5: Setup($(\mathbf{C_i})_{1 \le \boldsymbol{i} \le \boldsymbol{m}}$)

```
for i in 1...m do
   for j in 1...|C_i| do
      if (E_sj <= E_exp(i)) then
         broadcast expiry_message to all sensors in C_i
      else broadcast E_sj to all sensors in C_i
   end if
end for
end for
for i in 1...m do
   for j in 1...|C_i| do
         Delete expired sensors from the local cluster table of
   the sensor S_J
         Update residual energies of alive sensors in the local
   table of the sensor S_J
         end for
   for i in 1...m do
   for j in 1...|C_i| do
      d_sj = 1 + (N_sj / n)
   w_sj = E_sj * d_sj
```

(continued on next page)

a (*continued*)

Algorithm 5: Setup($(\mathbf{C_i})_{1\leq i\leq m}$)

```
end for
end for
for i in 1...m do
   for j in 1...|C_i| do
      if (w_{S_j} = max(w_{S_k})_{1≤k≤|C_i|}) then
   CH_i = S_j
   end if
end for
end for
for i in 1...m do
   QS_i = []
   for j in 1...|C_i| do
   for j in 1...|C_i| do
      if (d_{S_j} ≤ d_threshold) then
         if (S_j belongs to lp% weakest sensors in C_i) then
            QS_i = QS_i + [S_j]
         end if
      else
         if (S_j belongs to hp% weakest sensors in C_i) then
            QS_i = QS_i + [S_j]
      end if
   end if
end for
end for
for i in 1...m do
   CH_i broadcast a TDMA schedule to all sensors
 belonging to C_i
end for
```

4.2.2.2. Steady-state sub-phase. The following considerations are applicable during the steady-state sub-phase:

- Only selected sensors can send their readings to their cluster-head. The transmission is performed during the time slots allocated to them in the previous sub-phase. The remaining sensors are switched to idle status to conserve their battery power.
- The cluster-heads aggregate the selected sensor readings and then transmit them to the base station in CDMA mode. Any sensor expiry detected during the round is also reported.
- The base station processes the data received from all network cluster-heads to make decisions and transmit information to the end users. The base station thus monitors the network, cluster attractor expirations particularly, and updates the threshold energies and network density, computes the new clusters, and then broadcasts the new configuration to the whole network. Algorithm 6 describes the distributed interactions performed by all sensors in the network and the base station:

Algorithm 6: Steady-state($(\mathbf{C_i})_{1\leq i\leq m}$)

```
Sensors:
for i in 1...m do
   for j in 1...|C_i| do
      if (S_j belongs to QS_i) then
         S_j sends data to CH_i in TDMA mode
      else
   if (S_j = CH_i) then
```

a (*continued*)

Algorithm 6: Steady-state($(\mathbf{C_i})_{1\leq i\leq m}$)

```
         S_j transmits aggregated data to the base station in
 CDMA mode
         S_j reports sensor expiry occurred in C_i
      else
         S_j switches to idle status
      end if
      end if
   end for
   end for
Base station:
Receives data from (CH_i)_{1≤i≤m}
if (attractor expiry is detected) then
   Reconfigure the network
end if
```

5. Results and discussion

The performance of the proposed weighted Markov-clustering-based routing protocol (WMCL-BCRP) is compared below to its unweighted ancestor (MCL-BCRP)[31] as well as to the TEEN [27], PEGASIS [29],LEACH [24] and HEED [30] protocols. The authors chose these protocols for simulations because, in addition to the availability of their codes in Matlab programming language, they are the most used as performance benchmarks in the literature. The simulations were carried out on a battery of tests containing 100 random wireless sensor networks. After presenting the simulation parameters and their selected performance comparison criteria, we discuss performance in terms of the round during which the first and last sensor expirations occurred, the number of expired sensors per round, average sensor lifetime, and the average energy dissipation by the network per round.

5.1. Simulation settings

The code of the proposed weighted Markov-clustering-based routing protocol (WMCL-BCRP) and its unweighted ancestor (MCL-BCRP) as well as the simulation scripts to compare the performances with LEACH, PEGASIS, TEEN and HEED protocols were programmed with the Matlab programming language. The simulations do not require any particular specification at the physical and data link levels. The simulations were run with wireless networks of 100 sensors each with a GPS module and the same initial battery energy.

The main limitation of the proposed protocol as well as its unweightedancestor [31] is the use of sensors equipped with GPS modules constituting an additional load for sensors batteries. This constraint is unavoidable in order to configure the network using the MCL algorithm. GPS modules are known to be energy consumers [26], however, this is not a problem for the energy management of sensors, as GPS modules are only used once during the network configuration phase for a very short time and disabled once the sensor location has been transmitted.

The sensors were distributed randomly in a 10,000 m^2 square area, which produced a network with sparsely and densely covered regions. Fig. 3 presents an example of a network with a random distribution of sensors. The base station was placed at the center of the area. The same reading frequency is considered for all sensors. The transmission of readings to the base station is considered continual. Moreover, the sensors in a cluster are supposed to be able to communicate directly with the head. The head can also

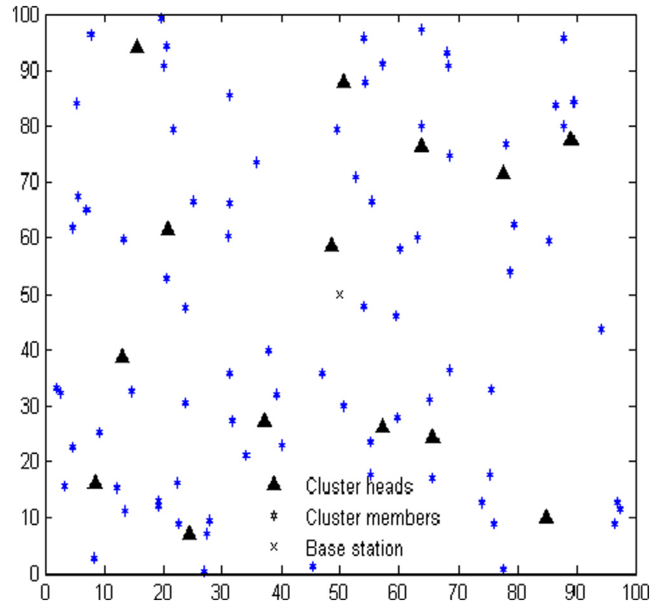

Fig. 3. An example of a 100 sensors network randomly distributed over a 100 × 100 sensing area.

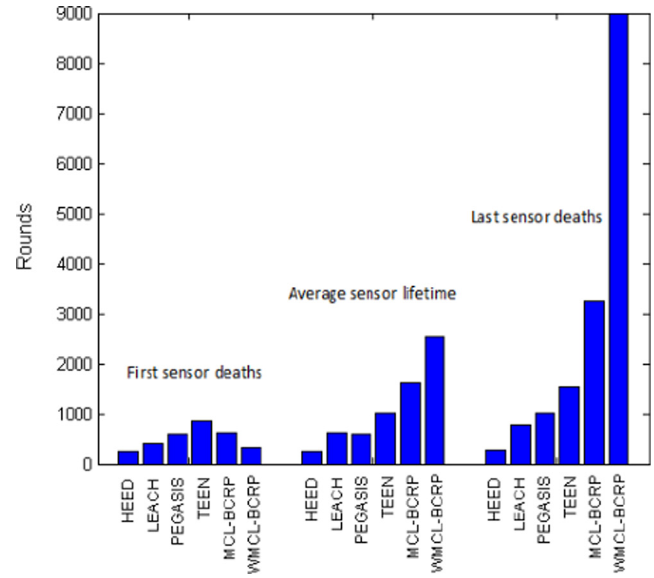

Fig. 4. First sensor expiry, average sensor lifetime and last sensor expiry in WSNs managed by the proposed weighted Markov-clustering-based (MCL-PCRP), unweighted ancestor (MCL-BCRP), TEEN, PEGASIS, LEACH and HEED routing protocols, based on simulation.

Table 2
WSN simulation settings and parameters.

Parameter	Value
Sensing area	100 × 100 m^2
Base station × coordinate	50 m
Base station y coordinate	50 m
Number of sensors	100
Size of data packet	4000 bits
Initial energy E_0	0.5 J
Threshold distance parameter K	4
Expansion power coefficient e	2
Inflation power coefficient r	2
Density range	5 m
Queried candidate percentage in low-density regions ($lp\%$)	25 %
Queried candidate percentage in high-density regions ($hp\%$)	50 %
Threshold defining tiny clusters ($x\%$)	5 %
Electrical energy per bit (E_{elec})	50 nJ/m^2
Data aggregation energy (E_{DA})	5nJ
Amplifier energy consumption (E_{amp})	0.0013 pJ/bit/m^2
Amplifier energy consumption (E_{fs})	10 pJ/bit/m^4

Table 3
Comparison of First sensor expiry, average sensor lifetime and last sensor expiry in WSNs managed by WMCL-PCRP, MCL-BCRP, TEEN, PEGASIS, LEACH and HEED routing protocols based on simulation.

Protocol	First sensor expiry	Average sensor lifetime	Last sensor expiry
WMCL-BCRP	at round 328	2886.4100	at round 8976
MCL-BCRP	at round 628	1622.9400	at round 3238
TEEN	at round 852	1019.7700	at round 1542
PEGASIS	at round 596	601.4100	at round 1014
LEACH	at round 401	612.3100	at round 772
HEED	at round 255	258.0300	at round 265

communicate directly with the base station. The simulation settings are summarized in Table 2.

5.2. First and last sensor expirations and average sensor lifetime

The comparison of the proposed protocol, with its unweighted ancestor and the other protocols, in a random non-uniform sensor distribution setting is shown in Fig. 4 and detailed in Table 3.

The simulations show clearly that the weighted Markov-clustering-based routing (WMCL-BCRP) gives the longest average sensor lifetime and the latest last sensor expiry but not the latest first sensor expiry. From Table 3,it can be observed that the proposed protocolextends the average sensor lifetime by about 156 % and 249 % respectively compared to the unweighted Markov-clustering-based (MCL-BCRP) and TEEN routing protocols and extends this average more than four times compared to PEGA-

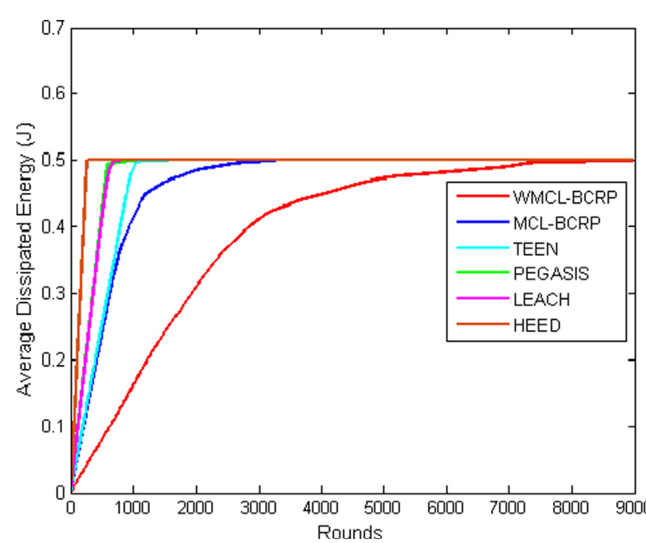

Fig. 5. Average Dissipated Energy with WMCL-BCRP, MCL-BCRP, TEEN, PEGASIS, LEACH and HEED protocols.

Table 4
Average energy (J) dissipation by different WSN clustering routing protocols.

Round	WMCL-BCRP	MCL-BCRP	TEEN	PEGASIS	LEACH	HEED
100	0.015933	0.047647	0.050519	0.083540	0.082045	0.184421
133	0.021347	0.063335	0.067080	0.111109	0.109052	0.249053
150	0.024091	0.071448	0.075674	0.125311	0.123122	0.280181
200	0.031964	0.095274	0.100996	0.167081	0.164225	0.377182
250	0.040025	0.119048	0.126124	0.208851	0.205416	0.478014
265	0.042365	0.126179	0.133730	0.221382	0.217798	0.500000
299	0.047746	0.142380	0.150893	0.249786	0.245728	-
304	0.048584	0.144767	0.153427	0.253963	0.249903	-
499	0.078846	0.237562	0.249738	0.416867	0.407537	-
525	0.082911	0.249890	0.262961	0.438587	0.427052	-
650	0.102647	0.308588	0.325207	0.495504	0.492692	-
772	0.122600	0.355743	0.386620	0.497016	0.500000	-
1014	0.163213	0.416277	0.490103	0.500000	-	-
1542	0.245318	0.469136	0.500000	-	-	-
1574	0.249964	0.470359	-	-	-	-
2000	0.309643	0.485300	-	-	-	-
2500	0.369135	0.493700	-	-	-	-
3000	0.411158	0.498400	-	-	-	-
3238	0.425455	0.500000	-	-	-	-
4000	0.448661	-	-	-	-	-
5000	0.473135	-	-	-	-	-
6000	0.482157	-	-	-	-	-
7000	0.490861	-	-	-	-	-
8000	0.497964	-	-	-	-	-
8500	0.499008	-	-	-	-	-
8976	0.500000	-	-	-	-	-

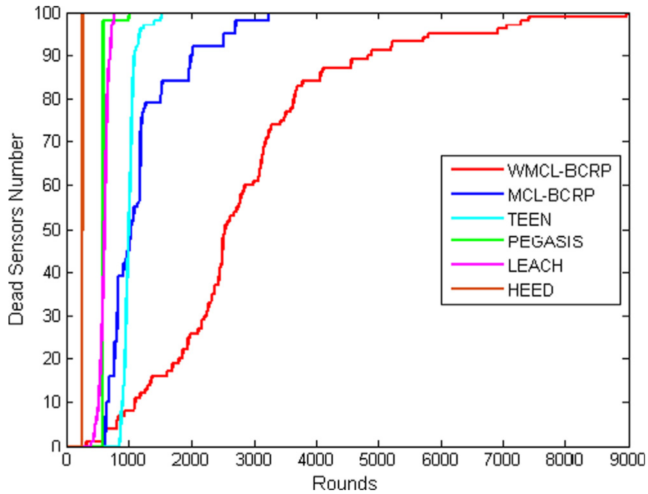

Fig. 6. WSN lifetime under the weighted Markov-clustering-based (WMCL- BCRP), unweighted Markov-clustering-based (MCL- BCRP), TEEN, PEGASIS, LEACH and HEED routing protocols.

SIS or LEACH, and ten times compared to HEED. Moreover, the weighted Markov-clustering-based routing (WMCL-BCRP) triples the network lifetime (i.e., last sensor expiration) compared to its unweighted ancestor (MCL-BCRP) sixfold, ninefold, 11-fold, and 33-fold compared respectively to the TEEN, PEGASIS, LEACH and HEED protocols. However, the latest first sensor expiry delay was obtained with TEEN, followed by the unweightedancestor (MCL-BCRP), thenPEGASIS, LEACH, and finally the weighted Markov-clustering-based routing (WMCL-BCRP). Only HEED had a precocious first sensor expiry than the weighted Markov-clustering-based routing (WMCL-BCRP).

The best performances recorded during the simulations in terms of average sensor lifetime and last sensor expiry are clearly in favor of the proposed protocol compared to its unweighted ancestor and the other protocols. This is mainly due to the efficient management of energy inside clusters by using a cluster-head selection without additional overheads and by limiting the number of queried sensors in each cluster. However, the poor performance of the proposed protocol in terms of first sensor expiry is due to the selection of queried sensors with the lowest weight factor, that is, with weak batteries in high-density and low-density regions. Despite this disadvantage, the restriction on queried sensors contributes significantly to the conservation of residual energy within clusters.

5.3. Average dissipated energy

Based on simulations, the weighted Markov-clustering-based protocol (WMCL-BCRP)can be expected to provide more efficient management of energy consumption by WSNs compared to the unweighted Markov-clustering-based (MCL-BCRP), TEEN, PEGASIS, LEACH and HEED protocols. It appears to keep energy dissipation very gradual and the average dissipation minimal (refer to Fig. 5 and Table 4). The colored lines in Table 4 correspond to the consumption of 50 % and 100 % of the initial battery energy, for example, red corresponds to the WMCL-BCRP protocol, and blue is associated to the MCL-BCRP protocol. The same colors are used to distinguish the different curves of simulated protocols in Fig. 5. LEACH and PEGASIS have similar energy consumption curves. From Table 4, the round at which energy consumption reaches 50 % ranges from the round133 for HEED to the round 1574 for weighted Markov-clustering-based protocol (WMCL-BCRP), with PEGASIS, LEACH, TEEN and MCL-BCRP respectively at rounds 299, 304, 499 and 525, at which points the WMCL-BCRP has consumed only 4.2 %, 9.4 %, 9.6 %, 15.6 % and 16.4 % of the initial energy. Furthermore, when sensor energy is 100 % spent, at round 265 with HEED, 772 with LEACH, 1014 with PEGASIS, 1542 with TEEN, and 3238 with MCL-BCRP, WMCL-BCRP-routed networks are still alive, having consumed respectively 8.4 %, 24.4 %, 32.6 %, 49 % and 85 % of the initial energy, and are not exhausted until round 8976, when all sensor batteries are exhausted.

The previous simulation results show that the performance of the proposed protocol in terms of energy management is significantly improved than other protocols. The main reasons that allow the gradual exhaustion of the energy and the minimal average dissipation with WMCL-BCRP, can be summarized as follows:

- The cluster formation process is run by the base station only after expiry of an attractor, which conserves energy during this process.
- The restriction of sensors that are allowed to send data to their cluster-heads reduces significantly the average dissipated energy within clusters at each round.
- Cluster-head selection reduces overhead by queried only sensor weights at the beginning of each round.

5.4. Number of expired sensors

The cumulative number of expired sensors recorded during simulations are graphed in Fig. 6 and summarized in Table 5. From Fig. 6, it can be observed that the weighted Markov-clustering-based protocol clearly performed better than the others at extending network lifetime in simulations longer than its nearest competitor.

Table 5 illustrates the rate of sensor expiration in networks managed by the proposed protocol, its unweighted ancestor, HEED, LEACH, PEGASIS and TEEN protocols. First, Last expirations and the rate of sensor expiration were recorded for each protocol as follows:

- First expirations (7 % of sensors) were recorded at round 255 under the HEED protocol. The network died only 10 rounds later, due to sensor expiration accelerated by the resulting energy-consuming cluster formation process.
- First sensor expirations (1 %) occurred at round 401 under the LEACH protocol. The network lifetime was 772 rounds, which was predictable in view of the resulting lower energy consumption compared to HEED.
- PEGASIS kept the network alive longer than LEACH, but sensor expiry was massive (98 %) at round 593 and the network was thereafter non-functional with only two (02) sensors until its expiry at round 1014. The sudden massive expiry is due to the uniform energy dissipation by all sensors in the PEGASIS chain.
- The TEEN protocol delayed the first sensor expirations until round 852, which was the longest among the protocols tested. However, the entire network was dead by round 1690.
- The unweighted ancestor (MCL-BCRP) kept all sensors alive until round 628, after which expiry was rapid but slowed down, the loss being held at 0 several times for hundreds of rounds and the network staying technically alive until round 3238, due to its low energy dissipation.
- The weighted Markov-clustering-based protocol (WMCL-BCRP) allowed its first sensor expiry at round 328, which was the second earliest. This is due likely to choosing the lowest-weighted sensors in the queried sensor selection process. However, sub-

Table 5
Accumulation of expired sensors in WSNs under the weighted Markov-clustering-based (WMCL-BCRP), unweighted Markov-clustering-based (MCL-BCRP), TEEN, PEGASIS, LEACH and HEED routing protocols.

Round	WMCL-BCRP	MCL-BCRP	TEEN	PEGASIS	LEACH	HEED
255	0	0	0	0	0	7
257	0	0	0	0	0	38
259	0	0	0	0	0	83
261	0	0	0	0	0	95
263	0	0	0	0	0	99
265	0	0	0	0	0	100
328	1	0	0	0	0	-
401	1	0	0	0	1	-
593	2	0	0	98	38	-
628	3	1	0	98	59	-
772	4	19	0	98	100	-
852	7	39	1	98	-	-
900	7	39	9	98	-	
950	8	42	21	98	-	-
1000	8	45	49	98	-	-
1013	8	45	54	99	-	-
1014	8	45	54	100	-	-
1200	12	73	96	-	-	-
1277	13	79	97	-	-	-
1300	13	79	97	-	-	-
1400	16	79	97	-	-	-
1500	16	79	98	-	-	-

(continued on next page)

Table 5 *(continued)*

1513	16	79	98	-	-	-
1542	16	83	100	-	-	-
1800	20	84	-	-	-	-
2000	26	90	-	-	-	-
2500	44	92	-	-	-	-
3000	61	98	-	-	-	-
3238	72	100	-	-	-	-
4000	84	-	-	-	-	-
5000	91	-	-	-	-	-
5796	95	-	-	-	-	-
6000	95	-	-	-	-	-
6911	95	-	-	-	-	-
7000	96	-	-	-	-	-
8000	99	-	-	-	-	-
8500	99	-	-	-	-	-
8976	100	-	-	-	-	-

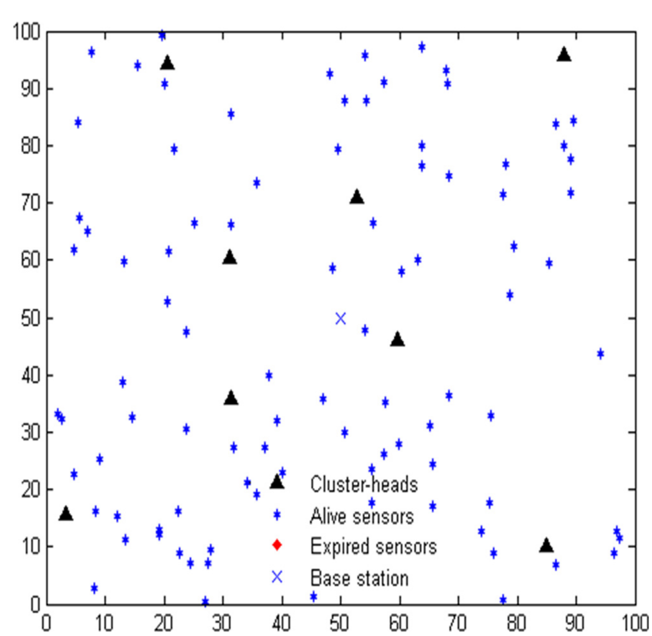

Fig. 7. The state of a WMCL-BCRP-routed network when the same one is dead under the HEED protocol (at round = 265, with 0 expired sensors).

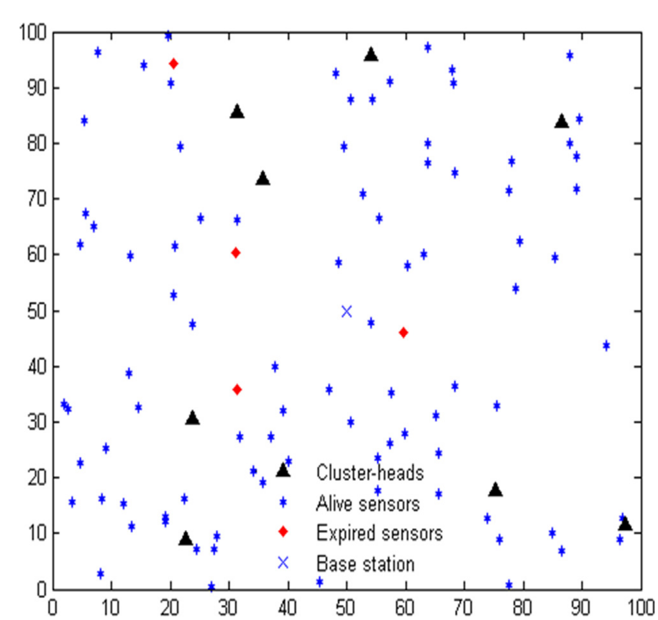

Fig. 8. The state of a WMCL-BCRP-routed network when the same one is dead under the LEACH protocol (at round = 772, with just 4 expired sensors).

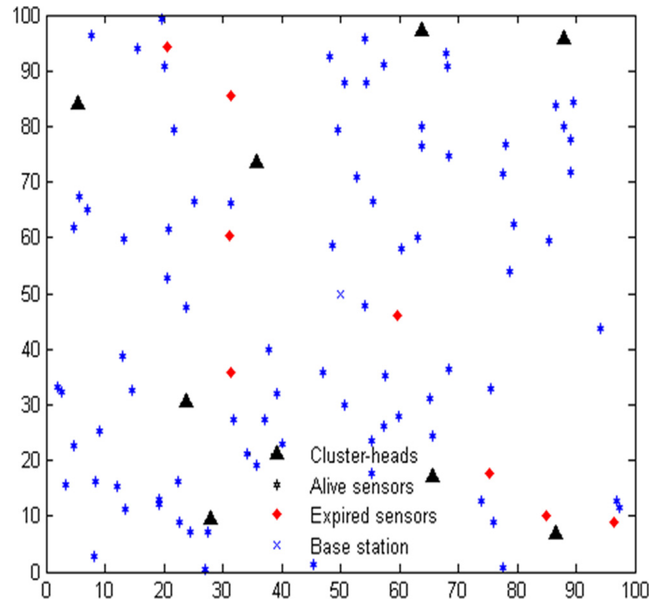

Fig. 9. The state of a WMCL-BCRP-routed network when the same one is dead under the PEGASIS protocol (at round = 1014, with just 8 expired sensors).

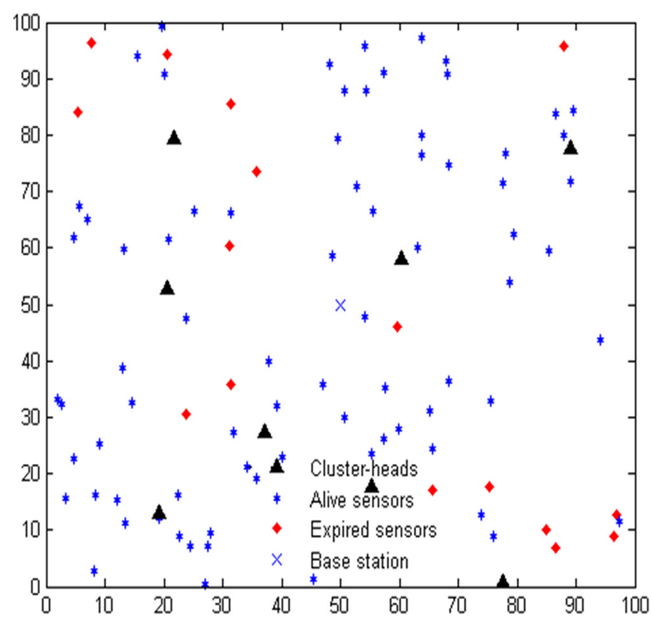

Fig. 10. The state of a WMCL-BCRP-routed network when the same one is dead under the TEEN protocol (at round = 1542, with just 16 expired sensors).

sequent sensor expiry was the slowest among the tested protocols. The proposed protocol was better than its unweighted ancestor at keeping the remaining sensors alive, holding the loss to 95 for over 1100 rounds (5796–6911). The network stayed technically alive until round 8976, which is two and half times.

The results of the previous simulations are also in favor of the proposed protocol which can maintain a very moderate rate of expiration compared to the other protocols. Indeed, when the

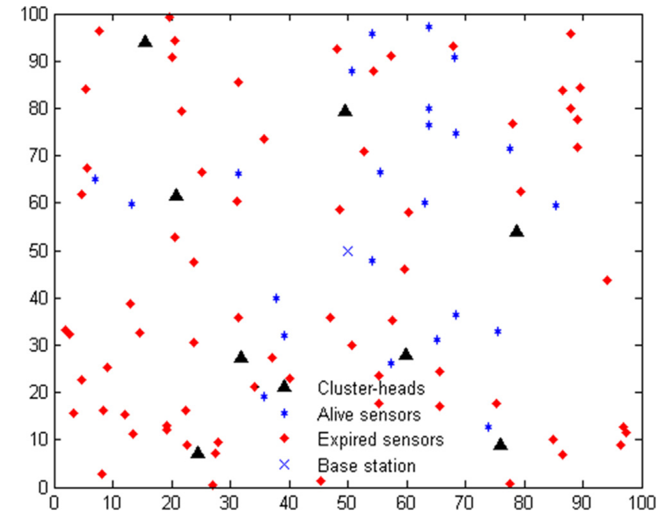

Fig. 11. The state of a WMCL-BCRP-routed network when the same one is dead under the MCL-BCRP protocol (at round = 3238, with just 72 expired sensors).

entire network was dead under the HEED, LEACH, PEGASIS, TEEN protocols, and the unweighted ancestor, it had respectively 0, 4, 8, 16, and 72 expired sensors under the weighted Markov-clustering-based protocol (WMCL-BCRP) protocol and was still functional. Figs. 7 to 11 illustrate respectively the state of a WMCL-BCRP-routed network when the same network is dead under HEED, LEACH, PEGASIS, TEEN protocols and the unweighted ancestor. More precisely, when the HEED-routed network dies, 100 % of the sensors are still alive in the weighted Markov-clustering-routed network, and 96 %, 92 %, 84 % and 28 % of sensors are when all have expired in respectively the LEACH, PEGASIS, TEEN or unweighted Markov-clustering-routed networks.

Based on previous simulations, weighted Markov-clustering-based routing allows a significant reduction in cumulative number of expired sensors. This advantage is a direct consequence of the reduction in energy consumption, and thus prolonging the lifetime of WSNs compared to next best protocol, its unweighted ancestor, and other protocols. However, the results also show that the difference in performance between weighting and not weighting the sensor selection is less apparent when networks with uniform sensor distribution are considered. In other words, when sensor density varies widely from one network region to another, the proposed sensor weighting method becomes advantageous in terms of network lifetime, average sensor lifetime, average dissipated energy, and the number of expired sensors per round.

6. Conclusion

Battery energy places major limitations on the deployment of wireless sensor networks (WSNs). Routing protocols have been developed to optimize sensor use and hence energy efficiency during data transmission and thereby prolong network lifetime. This article proposes an improvement of the Markov-clustering-based routing protocol, in which a clustering algorithm is combined with a strategy based on location, abundance and residual battery energy to select sensors as cluster-heads. Called weighted Markov-clustering-based routing, the proposed protocol selects cluster-heads and queried sensors. Cluster-head selection is based on the best weighting factor within sensor-dense regions to reduce intra-cluster communications with the base station. In contrast, the selection of queried sensors is limited to the candidates with the lowest weights both in dense and sparse regions of the network to conserve energy within the clusters. These selection strategies aim to restrict the sending of redundant data in dense regions, maintain sparse regions for as long as possible, and thereby reduce network energy consumption. Simulations show that the new protocol represents an improvement in terms of network lifetime; average sensor lifetime, average dissipated energy per round, and expired sensors per round. The applicability of the protocol to other fields will be investigated in future work.

References

[1] Jondhale SR, Maheswar R, Lloret J. Fundamentals of wireless sensor networks. Springer; 2022.
[2] Singh PK, Bhargava BK, Paprzycki M, Kaushal NC. Handbook of wireless sensor networks: issues and challenges in current scenarios. Springer; 2020.
[3] El Emary IMM, Ramakrishnan S. Wireless sensor networks: from theory to applications. CRC Press; 2019.
[4] Vijayalakshmi SR, Muruganand S. Wireless sensor networks: architecture - applications - advancements. Mercury Learning and Information; 2018.
[5] Selmic RR, Phoha VV, Serwadda A. Wireless sensor networks: security, coverage, and localization. Springer International Publishing; 2016.
[6] Fahmy HMA. Wireless sensor networks: concepts, applications, experimentation and analysis. Springer Singapore; 2016. https://doi.org/10.1007/978-981-10-0412-4.
[7] Forster A. Introduction to Wireless Sensor Networks. (1st edition). Wiley-IEEE Press; 2016.
[8] Priyantha, N. B. (2005) *The Cricket indoor location system* [Thesis, Massachusetts Institute of Technology] https://dspace.mit.edu/handle/1721.1/33924..
[9] Stehlik, M. (2011) *Comparison of Simulators for Wireless Sensor Networks* [Masarykovauniverzita, Fakultainformatiky] https://is.muni.cz/th/u6wrk/review_advisor.pdf..
[10] Levis P, Lee N, Welsh M, Culler D. TOSSIM: Accurate and scalable simulation of entire TinyOS applications. In: Proceedings of the 1st international conference on embedded networked sensor systems. p. 126–37. doi: https://doi.org/10.1145/958491.958506.
[11] Ilakkiyadeepaa, B. S., Vidhni, S. V. V., &Ranganathan, S. (2015) Wireless sensor networks based smart home control system. 10, 17567–17571.
[12] Cetinkaya, O., & Akan, O. B. (2016) *Use of wireless sensor networks in smart homes* (pp. 233–257) http://publications.eng.cam.ac.uk/1024026/..
[13] Gnanavel R, Anjana P, Nappinnai KS, Sahari N. Smart home system using a Wireless Sensor Network for elderly care. In: Second international conference on science technology engineering and management (ICONSTEM). doi: https://doi.org/10.1109/ICONSTEM.2016.7560922.
[14] Dishongh TJ, McGrath M. Wireless sensor networks for healthcare applications. Artech House; 2010.
[15] Julien-Vergonjanne A, Sahuguède S, Chevalier L. Optical wireless body area networks for healthcare applications. In: Uysal M, Capsoni C, Ghassemlooy Z, Boucouvalas A, Udvary E, editors. Optical Wireless Communications: An Emerging Technology. Springer International Publishing; 2016. p. 569–87. doi: https://doi.org/10.1007/978-3-319-30201-0_26.
[16] Gulati, K., Kumar Boddu, R. S., Kapila, D., Bangare, S. L., Chandnani, N., &Saravanan, G. (2021) A review paper on wireless sensor network techniques in Internet of Things (IoT). *Materials Today: Proceedings*https://doi.org/10.1016/j.matpr.2021.05.067..
[17] Gopika D, Panjanathan R. Energy efficient routing protocols for WSN based IoT applications: A review. Mater Today: Proc 2020. doi: https://doi.org/10.1016/j.matpr.2020.10.137.
[18] Sharma, R., Prakash, S., & Roy, P. (2020) Methodology, Applications, and Challenges of WSN-IoT. *2020 International Conference on Electrical and Electronics Engineering (ICE3)*, 502–507 https://doi.org/10.1109/ICE348803.2020.9122891..
[19] Touati Y, Ali-Chérif A, Daachi B. Gestion énergétique dans les réseaux de capteurs sans fil. ISTE Group; 2017.
[20] Liu X. A Survey on Clustering Routing Protocols in Wireless Sensor Networks. Sensors 2012;12(8):11113–53. doi: https://doi.org/10.3390/s120811113.
[21] Gupta SK, Jain N, Sinha P. Clustering protocols in wireless sensor networks: A survey. Internat J Appl Inform Syst (IJAIS) 2013;5(2).
[22] Sirsikar S, Wankhede K. Comparison of clustering algorithms to design new clustering approach. Procedia Computer Science 2015;49:147–54. doi: https://doi.org/10.1016/j.procs.2015.04.238.
[23] Xu L, Collier R, O'Hare GMP. A Survey of clustering techniques in wsns and consideration of the challenges of applying such to 5G IoT scenarios. IEEE Internet Things J 2017;4(5):1229–49. doi: https://doi.org/10.1109/JIOT.2017.2726014.
[24] Heinzelman, W. R., Chandrakasan, A., &Balakrishnan, H. (2000) Energy-efficient communication protocol for wireless microsensor networks. Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, vol.2..
[25] Fang W, Zhang W, Yang W, Li Z, Gao W, Yang Y. (2021) Trust management-based and energy efficient hierarchical routing protocol in wireless sensor networks. Digital Commun Networks 2021;7(4):470–8.
[26] Akyildiz IF, Vuran MC. Wireless sensor networks. John Wiley & Sons; 2010.
[27] Manjeshwar A, Agrawal DP. TEEN: A routing protocol for enhanced efficiency in wireless sensor networks. In: Proceedings 15th International Parallel and Distributed Processing Symposium. IPDPS 2001. p. 2009–15. doi: https://doi.org/10.1109/IPDPS.2001.925197.
[28] Manjeshwar A, Agrawal DP. APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless. In: Proceedings 16th international parallel and distributed processing symposium. doi: https://doi.org/10.1109/IPDPS.2002.1016600.
[29] Lindsey S, Raghavendra CS. PEGASIS: Power-efficient gathering in sensor information systems. Proc IEEE Aerospace Conf 2002;3:1125–30. doi: https://doi.org/10.1109/AERO.2002.1035242.
[30] Younis O, Fahmy S. HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans Mobile Comput 2004;3(4):366–79. doi: https://doi.org/10.1109/TMC.2004.41.
[31] Taieb Brahim M, Abbad H, Boukil-Hacene S. A low energy MCL-based clustering routing protocol for wireless sensor networks. Internat J Wireless Netwk Broadband Technol (IJWNBT) 2021;10(1):70–95.
[32] Van Dongen, S. M. (2000) *Graph clustering by flow simulation* [PhD Thesis] http://localhost/handle/1874/848..
[33] Fanian, F., Rafsanjani. M. K. (2019) Cluster-based routing protocols in wireless sensor networks: A survey based on methodology, Journal of Network and Computer Applications Volume 142, 15 September 2019, Pages 111-142, Elsevier..
[34] Nam, D. H (2020) Comparison Studies of Hierarchical Cluster Based Routing Protocols in Wireless Sensor Networks, EPiC Series in Computing Volume 69, 2020, Pages 334–344, Proceedings of 35th International Conference on Computers and Their Applications..
[35] Ramani KPL, Badholia A. Cluster based routing protocols in WSN sensor. IT in Industry 2021;9(1):2021.
[36] Yagoub MFS, Rodrigues JJPC, Khalifa OO, Mohammed AB, Korotaev V. Service redundancy and cluster-based routing protocols for wireless sensor and mobile ad hoc networks: A survey. Internat J Commun Syst Wiley Online Library, 2020. doi: https://doi.org/10.1002/dac.4471.
[37] Ujager, F. S., Mahmood, A., Usman, M., A., Rathore, M. S. (2022) A Hierarchical Energy Conservation Framework (HECF) of Wireless Sensor Network by Temporal Association Rule Mining for Smart Buildings, Egyptian Informatics Journal, Volume 23, Issue 1, March 2022, Pages 137-147..
[38] Saleh SS, Mabrouk TF, Tarabishi RA. (2021) An improved energy-efficient head election protocol for clustering techniques of wireless sensor networks. Egypt Inform J 2021;22(4):439–45.
[39] Heinzelman WB, Chandrakasan AP, Balakrishnan H. An application-specific protocol architecture for wireless microsensor networks. IEEE Trans Wireless Commun 2002;1(4):660–70. doi: https://doi.org/10.1109/TWC.2002.804190.
[40] Yassein MB, Al-zoubi A, Khamayseh Y, Mardini W. Improvement on LEACH protocol of wireless sensor network (VLEACH). Int J Digit Content Technol Appl 2009;3(2):132–6.
[41] Biradar, R. V., Sawant, S. R., Mudholkar, R. R., &Patil, V. C. (2011) Multihop routing in self-organizing wireless sensor networks. International Journal of Computer Science Issues (IJCSI), 8(1) 155https://doi.org/10.1109/HICSS.2000.926982..
[42] Loscri V, Morabito G, Marano S. A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH. IEEE Vehicular Technology Conference 2005;62(3):1809.
[43] Gaurav, K.N., Chetna, D., (2021)ESO-LEACH: PSO based energy efficient clustering in LEACH,Journal of King Saud University - Computer and Information Sciences,Vol. 33(8)947-954,ISSN 1319-1578,https://doi.org/10.1016/j.jksuci.2018.08.002..
[44] Radhika M, Sivakumar P. Energy optimized micro genetic algorithm based LEACH protocol for WSN. Wireless Networks 2021;27:27–40. doi: https://doi.org/10.1007/s11276-020-02435-8.
[45] Mukherjee, P., Pattnaik, P. K., & Panda, S. N. (2020) IoT and WSN Applications for Modern Agricultural Advancements: Emerging Research and Opportunities.145 p. IGI Global https://www.igi-global.com/gateway/book/218569..
[46] Sajedi SN, Maadani M, NesariMoghadam M. F-LEACH: a fuzzy-based data aggregation scheme for healthcare IoT systems. J Supercomput 2022;78:1030–47. doi: https://doi.org/10.1007/s11227-021-03890-6.
[47] MacQueen J. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Some methods for classification and analysis of multivariate observations. University of California Press; 1967.