Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 73 (2015) 460 – 467
Procedia Computer Science

The International Conference on Advanced Wireless, Information, and Communication Technologies (AWICT 2015)

Improvements in LEACH based on K-means and Gauss algorithms

Elkamel Rabiaa[a], Baccar Noura[b], Cherif Adnene[c]

[a]*National Engineering School of Carthage, Innov'Com laboratory*
[b]*National Engineering School Of Tunis, Innov'Com laboratory*
[c] *Faculty of science of tunis ,Signal Processing Laboratory*

Abstract

Several methods have been developed for routing in wireless sensor networks; the most popular is LEACH protocol. It owes its popularity to its simplicity and ability to handle large data sets. However, the main limitation of this method is its dependency to probabilistic calculations. In this paper, we propose a new improved version of the LEACH protocol. Our main contribution consists on the integration of an improved unsupervised algorithm which is the K-means on clusters and the use of the Gaussian elimination algorithm for choosing the Cluster Head. Based on Davies–Bouldin index (DBI), the K-means finds the optimal number of clusters K. The Davies-Bouldin index makes calculations much less complex, which is a potential purpose in clustering. Thus, DBI based K-means offers more simple parallel distributed computing of K and allows faster auto-classification of sensor nodes. In this manner a balanced energy clusters is created and distributed energy consumption is guaranteed. Simulations in a wireless sensor network show that our new approach greatly reduces the computation time thanks to K-means and minimize energy consumption thanks to Gauss.

Peer-review under responsibility of organizing committee of the International Conference on Advanced Wireless, Information, and Communication Technologies (AWICT 2015)

Keywords: WSN; Clustering, K-mean algorithm, optimal number of cluster, Energy efficiency;

1. Introduction

LEACH is considered as being the first routing protocol [1]. It combines the efficiency of energy consumption and the quality of access to the media, and it is based on the division into groups, with a view to enable the use of the concept of data aggregation for better performance in terms of lifetime. The communication architecture of

Peer-review under responsibility of organizing committee of the International Conference on Advanced Wireless, Information, and Communication Technologies (AWICT 2015)
doi:10.1016/j.procs.2015.12.046

LEACH consists, similarly to cellular networks, to form cells based on the signal amplitude, and to use the cell headers to the routers as well node. These cells are called groups (clusters) and the tees are group leaders (cluster-heads). The group's leaders are selected randomly according to a specific election algorithm based on a probability function that takes into account various criteria are based on the desired percentage of CHs and the number of iterations during which a node took the role of CH. The algorithm proceeds in turns (round), which have approximately the same pre-determined time interval. Each round consists of an initialization phase and a transmission phase.

- Initialization phase: In this phase, the nodes proceed to self-elect of CHs. They are based on the desired percentage of CHs and the number of iterations in which a node has taken the role of CH. Thus, a node n is a random value between 0 and 1. If this value is less than the threshold T (n), the node expresses CH.

$$T_n = \begin{cases} \dfrac{P}{1 - P \times (r \bmod \frac{1}{P})} ; n \in G \\ 0 \end{cases} \quad (1)$$

With: P: desired percentage of CHs, r: current iteration, G: set of nodes that have been selected as CH in the last (1 / P) iterations.

The CHs inform their neighbors of their election. Each remaining node decides to choose the nearest CH. After the formation of clusters, each CH programs the nodes belonging to the cluster by sending the codes and frequencies communication. Each node other than CH turn its antenna off when not communicating its data. The system sets the optimal number of clusters based on a few parameters such as network topology, the cost of communication and computation operations (Generally CHs represent 5% of network nodes).

-Transmission Phase: this phase is longer than the previous phase, and enables the collection of sensor data. Using the TDMA multiplexing scheme, members transmit their captured data for their own slots. This allows them to turn off their communication interfaces outside their slots to save energy. This data is then aggregated by merging the CH and compress and send the final result to the right node. After a predetermined time, the network will move to a new round. This process is repeated until all network nodes elect a CH along the preceding rounds. In this case, the round is reset to 0.

Our process of developing the solution is guided by a simple and useful questionnaire consisting of the following questions:

What is a good combination?

- A good clustering method ensures
  - A large intra-group similarity.
  - Low inter-group similarity.
- The quality of a group depends on the similarity measure used by the method and its implementation.

What type of Cluster Head Election Algorithm use?

The cluster head election (CH) is the process used to select a node in the cluster as a leader node. The CH maintains information related to his group. This information includes a list of the cluster nodes and the path of each node. The election of a specific node as a cluster head is a very important job, but sophisticated. Various factors may be considered for election of the best node as a cluster head. Some of these factors include the location of the node relative to other nodes, mobility, energy, confidence, and throughput of the node.

The consolidation of clustering sensor nodes has been widely followed by the research community to achieve the goal of scalability (scalability) of the network. There are two main clustering techniques: hierarchical clustering and clustering partitioned. According to a study done by D.Sonagara and all [2] showed that the technique of clustering partitioned is more efficient and faster than the hierarchical classification and grouping is partitioned based on strong assumptions. There is a wide variety of partitioned techniques, but here we took the K-Means algorithm that is widely used in the context of clustering document.

The rest of the paper is organized as follows. Section II will present a background discussion of art's state in k-means algorithms. The following section presents k-means clustering approach. Finally section intends to detail simulation scenarios and to discuss the obtained results. This paper will be closed by a conclusion that summarizes this work and suggests new perspectives.

2. Related Works

D. Mechta and all proposed in [3] a new approach called LEACH-CKM that hybridizes two algorithms: the classification algorithm K-means clustering to group the nodes and the routing protocol Minimum Transmission Energy (MTE) to carry information from remote nodes. They used the k-means algorithm instead of the algorithm "annealing" as used LEACH-C for cluster formation. This strategy also allows all nodes in the network to communicate their information. For CH election a node must meet the following criteria: Eligibility and Belonging to the first zone: the first zone that has a radius of 145m of the base station.

In [4] another extension of LEACH is proposed based on the K-means algorithm followed by Elbow method. In the proposed method, the acquisition is performed using the k-means algorithm has and to find an optimal value "k" it uses the Elbow method. It produces a new cluster system for WSN with dynamic selection of the number of clusters automatically. The algorithm begins by executing Elbow to find the optimal number then it starts Cluster formation phase using the K-means. At the beginning of a cycle each sensor itself elects to be the head cluster. In the following cycles the probability of becoming a group leader is set according to nodes energy level relative to the total remaining energy into the .A.Sheta and all [5] proposed another method based on the development of a hybrid clustering algorithm which has two folds:

1) Use the unsupervised learning algorithms k-means to select the sensors in each group using any number of clusters

2) Using Particle Swarm Optimization (PSO) and Genetic Algorithms (GAs) individually to select the best CHs.

These two algorithms are appointed KPSO and KGAs. KPSO provided better results against KGAs.

In [6] K-means algorithm will be applied to another protocol which is SEP 'Stability Election Protocol. The architecture of SEP is characterized by heterogeniste nodes and integrating new types of nodes called Advanced node. The decision to become a CH is made at the beginning of each turn by each node independently by choosing a random number between [0, 1]. If the random number is below the threshold then the node becomes a CH in the current cycle. The proposed protocol improves SEP the level of training the cluster by applying the K-means for the formation of clusters, which ensures uniform combination and for the choice of the cluster head, it is based on the Euclidean distance from the cluster center and maximum residual energy in the cluster that provides a more efficient solution of energy in WSN. For the transmission of data each CH sends their data in direct mode to sink. Following the same axis, in [7] an improvement on the technical data aggregation prefix frequency filtering (PFF) is added. The improvement is the integration of a clustering algorithm which is K-means before applying the PFF on data from each cluster level Aggregator. In this paper each aggregator calculates K as follows: The algorithm PFF works in two phases, the first at the nodes called local aggregation and the second-level aggregators. The k-means algorithm is applied only to the high level of the network also the idea of applying the low level can give more results. In addition the document did not use decision feedback filter system to cancel the interference between residual Inter symbols Interference (ISI). New DV-hop [8] is a new clustering algorithm is recently proposed which is based in the first stage on the conventional DV-hop, while the second step starts by using K-means to find the cluster center. The principle of DV-Hop is the most fundamental is to calculate the minimum number of hops between a node and each remote Anchor using vector exchange protocol.

3. K-means Clustering Approach

Clustering of nodes is one of the techniques that can increase the lifetime of the entire network by aggregating data to the cluster head. The latter is based on an algorithm to combine two classes of a partition for a more aggregate score and more exactly on the K-means algorithm. The K-means algorithm is an Automatic algorithm clustering by Geon [9]. From a number of initial partitions, seeking to improve iteratively partitioning. Each score is represented by a kernel, so there kernel k.

3.1. Principe of K-means

The algorithm k-means is to group the items according to a specific criterion the input of the algorithm is the number k groups (cluster). Once the number of groups before the algorithm is chosen arbitrarily as points k "initial" centers of k groups. The next step is to calculate the distance between each individual (point) and the k centers, the smallest distance is chosen to include this person in the group with the closest center. After all individuals grouped, there will be sub-k clouds (cluster) of the total cloud disjoint. For each group, the algorithm calculates the new center of gravity. The algorithm stops when the groups become stable built.

Select k points that represent the average position of the partitions

Repeat until convergence:

-Assigning each observation to the nearest partition (ie perform a tessellation using average).

$$S_t^{(t)} = \left\{X_j : \left\|X_j - m_i^{(t)}\right\| \leq \left\|X_j - m_{i^*}^{(t)}\right\| \forall i^* = 1..k\right\} \quad (2)$$

-Updating the mean of each cluster

$$m_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{X_j \in S_i^{(t)}} X_j \quad (3)$$

- Convergence is reached when there is no change.

3.2. Davies-Bouldin indes (DBI)

The use of K-means algorithm allows the classification in a large network and saves more time to calculate. However it requires from the number of classes to solve this problem we would use a tactic to set the number of cluster initially and essentially find clusters of equal size to facilitate the task for the next step ie d. the clusters election phase we introduce Gaussian elimination to the choice of the Cluster Head which aims to gain more time and minimize communication between nodes. To remedy the problem of optimal choice of k we will use a tactic that is Davies–Bouldin index (DBI) has a great advantage [10].

Let $R_{i,j}$ be a measure of how good the clustering scheme is. This measure, by definition has to account for $M_{i,j}$ the separation between the i^{th} and the j^{th} cluster, which ideally has to be as large as possible, and Si, the within cluster scatter for cluster i, which has to be as low as possible. Hence the Davies–Bouldin index is defined as the ratio of Si and $M_{i,j}$ such that these properties are conserved:

$$R_{j,j} \geq 0$$
$$then R \succ R$$
$$When S_i = S_k and M_{i,j} \leq M_{i,k}$$
$$then R_{i,j} \geq R_{i,k}$$
$$When S_i = S_k and M_{i,j} \leq M_{i,k}$$
$$then R_{i,j} \geq R_{i,k}$$

With this formulation, the lower the value, the better the separation of the clusters and the 'tightness' inside the clusters A solution that satisfies these properties is:

$$R_{i,j} = \frac{S_i + S_j}{M_{i,j}} \tag{4}$$

This is used to define Di:

$$D = \max_{i \neq j} R_{i,j} \tag{5}$$

If N is the number of clusters:

$$DB \equiv \frac{1}{N} \sum_{i=1}^{N} D_i \tag{6}$$

Davies–Bouldin index (DBI) is used to find the number of clusters. This process is applied to determine the greatest number of grapes to ensure the best quality grapes. The quality grouping (especially many groups) is effective in improving energy efficiency.

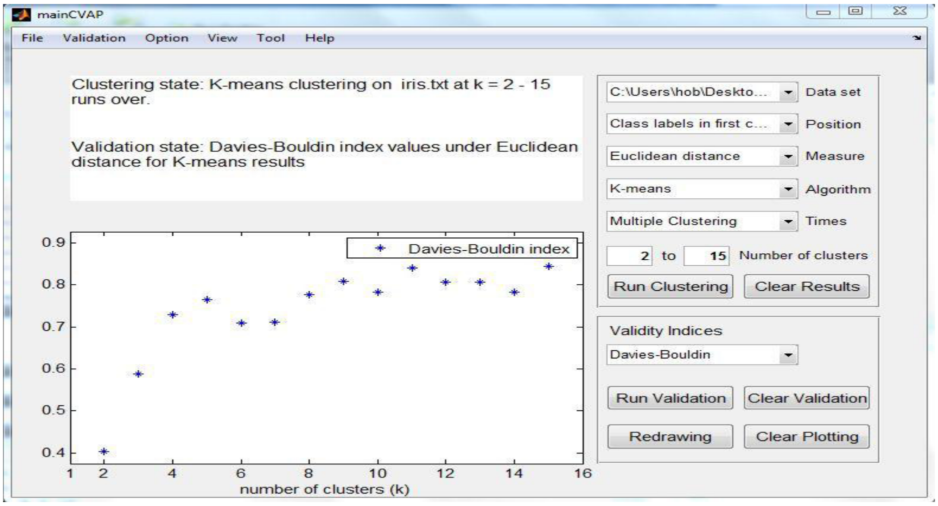

Fig1. Number of optimal of cluster.

4. Proposed Approach

Our proposed approach ameliorates the clustering procedure. K-Means clustering and LEACH has been linked to improve the clustering assignment to improve the cluster peculiarity and to make the intact process enough energy efficient to enlarge the WSN network lifetime. The process may be divided into three stages for the first round of communication, in setup phase we apply, K-means algorithm for cluster formation, which guarantee uniform groups. The cluster formation by K-means algorithm ensures perfect clustering. Proposed protocol is divided into many rounds, and each round contains cluster formation phase and Steady state phase. Using k-means as clustering technique reduces overhead during the re-election of the leader, as shown in the following figure.

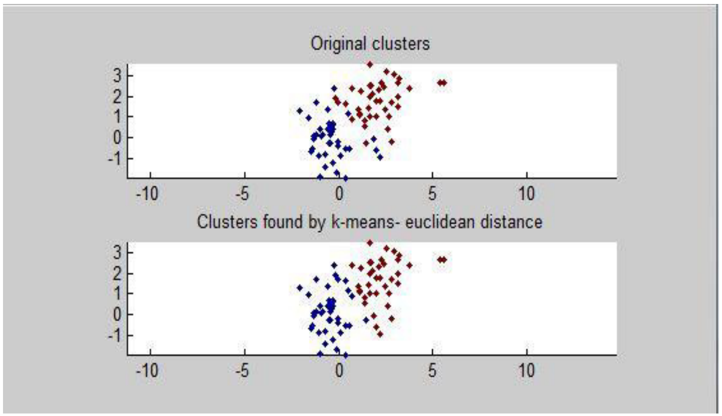

Fig.2. Comparaison beetwen original cluster and cluster by k-means

4.1. Cluster Formation phase

In the first step, clusters are formed by the K-Means method. Each cluster comprises a set of nodes K-means is an algorithm minimization who alternated as a whole K, will seek to separate a set of points in K clusters.

In the second stage, a CH is elected in each cluster in the first round the centroid are proposed as a cluster head in the other round we will used an effective technique that is the Gaussian elimination algorithm his is done by taking account of the residual energy of the estimated future nodes, and the number of revolutions which may be responsible for the cluster formation to maximize network lifetime. It models the network and the energy expended by the nodes as a linear system and this using the Gaussian elimination algorithm to select the CHs [11].

The matrix represents the energy consumption of each cluster node. The elements $a_{i,j}$ refers to the energy consumed by the node i j if the node is the CH. Furthermore, bi denotes the residual energy of node i, and xi expresses the time during which the node i is CH.

In this way, the matrices B and X are formed, so that as shown in the following equation.

$$B = \begin{bmatrix} b_1 \\ b_2 \\ . \\ b_k \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,k} \\ a_{2,1} & a_{2,2} & \dots & a_{2,k} \\ . & . & \dots & . \\ a_{k,1} & a_{k,2} & \dots & a_{k,k} \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_k \end{bmatrix} \qquad (7)$$

Some nodes that turn into cluster heads as per above conditions send their cluster head announcement information to inform other nodes. The other nodes turn up as non-cluster head nodes send cluster joining information to cluster head.

4.2. Steady state phase

1. Nodes in a cluster, sends their data with transmission probability P according to TDMA schedule, and cluster head receives, and aggregates the data.

2. Cluster heads will send their data directly to the base station.

5. Simulation and Results

5.1. Evaluation Parameters

For the simulations and development of our algorithms, the network model Proposed Consists of 100 sensor nodes Deployed over a field of (M× M) square unit area. The scheme is developed using MATLAB because of its ease of interface and availability of pre-programmed functions Necessary. We compare the algorithms according to criteria: Average Energy and Number of Dead Nodes.

5.2. Simulated Results

First, we measured the energy consumed by the nodes in the different rounds and this by implementing the three protocols. We present in the figure below the rate of energy consumed at the nodes. The figure 3 shows that the average energy consumption of nodes with the LEACH and LEACH-C protocols is higher than that measured with our suggestion and this by a rate of 0.1J in the round 50. This level of energy consumption is due to the expensive high spots in terms of energy as carried by CH when he was elected since the choice of CH, as the Gaussian elimination criterion is significantly better than the choice according to the criterion of LEACH. At the LEACH-C protocol, the lifetime of the network is low, because the nodes are being depleted faster. Is because the choice of the cluster head will make each new round involving additional energy consumption.

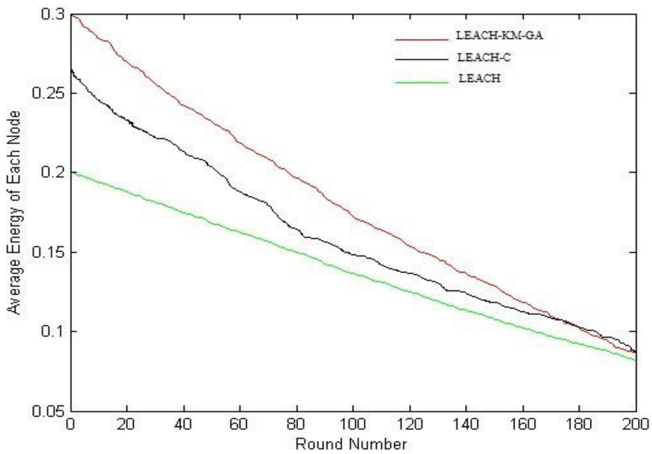

Fig.3.Average Energy vise number round.

As illustrated by the results of the figure 4, shows the evolution of the number of dead nodes based on the number of iterations. We observe that in LEACH, the first node exhausts its energy after 10 iterations while in LEACH-C and our proposed, the first node depletes its energy respectively after 95 and135 iterations. Also we note that in round 160 the number of dead nodes is 25, 5, and 3 respectively for LEACH, LEACH-C and our protocol. Therefore, we find that LEACH-KM-GA is more efficient in terms of network lifetime compared to other protocols.

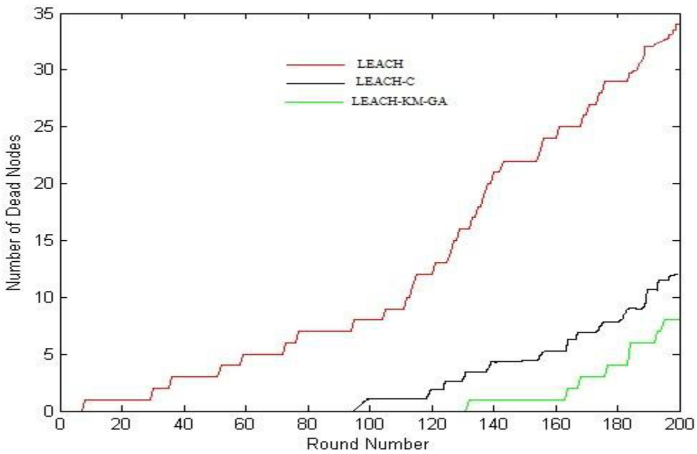

Fig.4.Number of Dead Nodes vise number round.

6. Conclusion

We summarize our article is to addresses the following points: Optimization of energy consumption of a network and the extension of its duration of life. First, we will address the problem of maximizing the network lifetime by planning optimal network partitioning under constraint cover, cluster formation. In this context we will use a powerful technique for clustering to enable global management of the network, a distribution of the burden and facilitate the management of resources which is the k-means algorithm. Then we tackled the same maximization problem of network life by integrating a tactic for choosing representative from each cluster will allows more energy since it will control the other nodes.

Several perspectives can be regarded for our research. We plan to study the performance of our routing protocol and this by changing the communication paradigm and taking into consideration other metrics such as quality of service or the throughput of transmission.

References

1. W.R. Heinzelman, A.Chandrakasan, and H.Balakrishnan "Energy-Efficient Communication Protocol for Wireless Microsensor Networks" Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000.
2. D.Sonagara, S.Badheka « Comparison of Basic Clustering Algorithms" International Journal of Computer Science and Mobile Computing Darshan Sonagara et al, International Journal of Computer Science and Mobile Computing, October- 2014, Online at www.ijcsmc.com.
3. D.Mechta, S.Harous, I.Alem, D. Khebbab "LEACH-CKM: Low Energy Adaptive Clustering Hierarchy protocol with K-means and MTE" 978-1-4799-7212-8/14/$31.00 ©2014 IEEE.
4. P.Bholowalia, A.Kumar "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN" International Journal of Computer Applications (0975 – 8887) Volume 105 – No. 9, November 2014.
5. A.Shetaand B.Solaiman"Evolving a Hybrid K-Means Clustering Algorithm for Wireless Sensor Network Using PSO and Gas" IJCSI International Journal of Computer Science Issues, Volume 12, Issue 1, No 1, January 2015 ISSN (Print): 1694-081 (Online): www.IJCSI.org.
6. B.Kandari and R. Singh"K-SEP: A more stable SEP using K-Means Clustering and Probabilistic Transmission in WSN" International Journal of Current Engineering and Technology ©2014 at http://inpressco.com/category/ijcet K. Elissa, "Title of paper if known," unpublished.
7. .Harb, A.Makhoul and D. Laiymani, A.Jaber and R.Tawil "K-Means Based Clustering Approach for Data Aggregation in Periodic Sensor Networks" 2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob).
8. A. R. Kulaib, R. M. Shubai, M. A. Al-Qutayri, and Jason W. P. Ng"Improved DV-hop Localization Using Node Repositioning and Clustering" 978-1-4799-6532-8/15/$31.00 ©2015 IEEE.
9. G.Yong Park, H. Kim, H.W.Jeong, and H.Y.Youn, "A Novel Cluster Head Selection Method based on K-Means Algorithm for Energy Efficient Wireless Sensor Network", 2013 27th International Conference on Advanced Information Networking and Applications Workshops,
10. S.Petrovi´c "A Comparison Between the Silhouette Index and the Davies-Bouldin Index in Labelling IDS Clusters" NISlab, Department of Computer Science and Media Technology, Gjøvik University College, P.O. box 191, 2802 Gjøvik, Norway
11. S. A. Nikolidakis, D.Kandris , D. Vergados and C.Douligeris"Energy Efficient Routing in Wireless Sensor Networks Through Balanced Clustering" Algorithms 2013, 6, 29-42; doi:10.3390/a6010029, algorithms ISSN 1999-4893 www.mdpi.com/journal/algorithms.