New Smart Nodes Distribution using Kmeans Approach to Enhance Routing in Wsn

Hanane Aznaoui*, Said Raghay and Layla Aziz

Laboratoire de Mathematiques Appliquees et Informatique. Faculte des Sciences et Techniques Gueliz, Universite Cadi Ayyad, 40 000 Marrakech, Maroc; h.aznaouii@gmail.com, aghay@gmail.com, Laylaa.az@gmail.com

Abstract

Objectives: Wireless Sensor Network (WSN) is an advanced technology, applied to many fields of research. However, it still limited due to some drawbacks, the energy consumed is one of them, which presents a critical issue. So that, our objective in this paper is to decrease the energy consumed during communication and prolong the network lifetime in Geographic Adaptive Fidelity (GAF) protocol. Methods/Statistical analysis: The Kmeans method was been exploited for improving the energy consumed in the network during routing data, which permit extending the network lifetime. It aims at distributing the sensor nodes. Where, the gravity center is determined as an active node, considering the least distance to the center of gravity and respecting the multihop communication between Active nodes. Findings: Simulation results confirmed that our new improved protocol reduces significantly nodes energy, which improves the network lifetime. Application/Improvements: By introducing our enhanced version Kmeans GAF, we can improve localization systems.

Keywords: Center of Gravity, GAF, Grid, Kmeans Algorithm, Location-Based, Routing Protocols, WSN

1. Introduction

Nowadays, WSN is among the technologies that attracted the whole world, it could be used in all field such as monitoring of sound, temperature or pressure. It is a set of huge number of nodes scattered on the network, aimed at collecting incident and send it through one or more node to the base station. The aggregated data are transferred to the treatment center by the sink in order to treat or archive it. This-operating process showed at Figure 1.

Figure 1. WSN architecture.
Sensor nodes are arbitrarily scattered and organized by wireless communication, aimed at work for a long period and powered by battery. In the majority of cases, it is hard and even impossible to recharge or change batteries. In spite of that and in order to prolong the lifetime of each sensor node. It is necessary to implement new schemes of routing protocols considering energy constrains, such as aggregation data, localization position, synchronization during communicating and network security. When applying traditional routing protocols in WSNs, different shortcomings are consumed. However, several routing are designed and categorized to three families: data-centric routing, hierarchical routing and location based-routing protocols.

1.1 Data-Centric (DC) Routing
Where the operating is based in to sending questions to spaces of interest by sink and waiting for request from responsible sensor node, considering into account properties and type of data, several routing based on data centric type are created Directed Diffusion (DD), Rumor Routing (RR) and Sensor Protocol for Information via Negotiation Protocol (SPIN), Cougar, Active Query Forwarding in Sensor Networks (AQUIRE).

1.2 Hierarchical
It aims at dividing the network into clusters and in each one, one sensor node is elected as manager, which called cluster head, responsible for sending data received from all members node to the BS after applying aggregation, which aims at conserving more energy and prolonging the network lifetime comparatively to the other kind of routing protocols. Low-Energy Adaptive Clustering Hierarchy (LEACH), Power-Efficient Gathering in Sensor Information Systems (PEGASIS), Threshold Sensitive Energy Efficient Sensor Network Protocol (TEEN), Adaptive Periodic Threshold Sensitive Energy Efficient Sensor Network Protocol (APTEEN) are designed as hierarchical routing.

1.3 Location-Based
Nodes sensor are randomly located in area of interest, habitually known through geographic position, the separating distance between node to another is considered by the signal received from nodes. In addition, the transfer of data are only between neighboring nodes, which targets at optimizing the energy consumed and extending the network lifetime, routing based on location are Minimum Energy Communication Network (MECN), Small Minimum-Energy Communication Network (SMECN), Geographic Adaptive Fidelity (GAF), Geographic and Energy-Aware Routing (GEAR).

Our paper is organized as two sections, the first one contains the related work especially GAF protocol and its improved versions, and the second one contains the proposed work.

2. Related work

2.1 GAF Protocol
In WSN, to decrease the consumed energy also to extend the network lifetime, several location-based protocols are suggested. Among them, GAF protocol, for GAF suggested first for MANETs, nevertheless, it also used for WSNs. It allows organizing the network structure into equal virtual grids created on their geographic positions using localization systems. Its permit communication between sensor nodes in neighboring grid. Each sensor node can be in one of the three transition states: Active, Discovery or Sleeping as shown in Figure 2. The conceptual model of GAF protocol shown at Figure 3, where just one node is in active state in each grid, which is responsible for transmitting data to another sensor located in the neighboring grid, however other sensors in the same grids are in the sleep state. This concept allows extending the lifetime of the network and resolves the problem of researching the equivalents sensor for transmitting data.

Figure 2. Transition state in GAF.
Even though GAF used one active node in each grid for optimizing the energy consumed, it still limited due to many drawbacks. GAF protocol does not consider the remaining energy during the selection of active node. It accepts only adjacent communication. Therefore, high numbers of active nodes contribute in routing data, which allow consuming more energy. In addition, several signal propagation problems exist due to the unreachability of the BS from nodes. For that, several improved version of GAF has been appeared in order to overcome their constraints, such as Diagonal Geographic Adaptive Fidelity (DGAF), Two Level - Geographic Adaptive Fidelity (T-GAF), B-GAF, Hierarchical Geographic Adaptive Fidelity (H-GAF), Hexagonal Geographic Adaptive Fidelity (HEX-GAF) and optimized GAF.

### 2.2 DGAF

Diagonal GAF\textsuperscript{16} in a direct way, it authorizes communication between two diagonal grids. It comes in order to elude the problem of communication in a basic GAF, which is only in horizontal and vertical direction.

The conceptual model of DGAF showed at Figure 4, where we found n0 and n1 deployed in a farthest away in two adjacent grids, r units is the size of each square grids and the transmission range is R units. For basic GAF, the distance between two adjacent grids must not be larger than R.

\[
\sqrt{r^2 + (2r)^2} \leq R \iff r \leq \frac{R}{2\sqrt{2}}
\]

For DGAF:

\[
\sqrt{r^2 + (2r)^2} \leq R \iff r \leq \frac{R}{\sqrt{5}}
\]

### 2.3 T-GAF (Two Level GAF)

A new improved version of GAF appeared with Authors in\textsuperscript{17} in order to decreases the number of sensors node, which participate in routing function by optimizing the hop count of the basic GAF. T-GAF allows direct communication between neighbor’s, source and destination, so that it uses two levels to route data. This version aims at minimizing the consumed energy comparatively to the basic one due to the reducing hop count using, and the way of selecting active nodes, which is based on its highest residual energy. The same notion and idea is also applied in the D-GAF. The conceptual models of T-GAF and T-DGAF are showed at Figure 5.

### 2.4 B-GAF\textsuperscript{18}

Founded on three-dimensional organization, aims at distributing the network into diverse number of cubes grids having the same size. Each cubes represents a cluster, and in each cluster a Cluster Head (CH) is defined, the CH is selected considering its highest residual energy and the
distance separates it from the sink. The probability for selecting the CHs combines both energy and distance parameters. Where only CHs are active and responsible for routing data while the remaining nodes are in sleep mode. This new version aims at avoiding the highest energy consumed by the CHs.

2.5 HEX-GAF

Proposed by Authors in\textsuperscript{14}, named as Hexagonal GAF and based on dividing the network on hexagonal grid. Therefore, the hexagon structure replaces the square grid in basic GAF. This structure mesh has six possibility of next hop, considering the maximum distance because of symmetry. The entire next hop cells for cell A are equally reachable by definition. Figure 6 exposed the hexagon architecture.

2.6 HGAF

proposed by Authors in\textsuperscript{20} called Hierarchical GAF due to using of a covered structure under selection of active nodes in the preformed cells. Based on the objective of keeping connectivity between active nodes as coordinators of the grids. This is done by limiting the active nodes positions in cells and synchronizes these positions using a sub-cells distribution. Selecting the active nodes hierarchically (cells and sub-cells) as shown in Figure 7 and that guaranties the communication between the adjacent cells.

2.7 GAF and CO Protocol

A new improved versions designed by Authors in\textsuperscript{21}. Named as GAF with Connectivity-Awareness, where the network is separated into hierarchical and hexagonal cells. As an alternative of mrectangular cells in basic GAF. The essential objective of this management architecture shown in Figure 8 is that, one node is kept as active node in every single hexagonal cell, in order to transfer information and sensing activities during time of routing which helps on saving energy consumed comparatively to basic GAF.

Due to this architecture, this protocol can be deployed as algorithm in several strategies, such as sleeping approaches and clustering.

2.8 Optimized GAF

A different new improved version developed by Authors in\textsuperscript{22}, founded on improving the discovery phase of states of transition as shown at Figure 9. Optimized GAF also based on three states of transition Discovery, Active and sleep, same as the basic version, however its process is different.

- Discovery phase: Where a sequence of nodes is selected to become active nodes assigned to the nodes having maximum remaining energy. This phase will be executed once time just for finding the sequence of actives nodes.
- Active Phase: After $T_a$ Node will become active without entering in discovery phase.
- Sleep Phase: After $T_s$, next node will become active node.
3. K-means Approach

The way of distribution nodes in network, can make difference and increase the lifetime of the total network. Clustering is one those method that allows dividing the network. Optimizing its energy consumed, by aggregating data before send it to the active node called cluster head. K- means approach is a popular clustering algorithm, designed by22. Aims at distributing the areas of interest into K clusters.

3.1 Principle of K-means:

The main objective of this algorithm is categorizing a set of data using a number of groups. K-means process is mainly based on a number of steps:
- The first step is defining the number of k groups (cluster) whose centers are chosen randomly.
- The second step is calculating the distance separating each point from the K predefined centers.
- The third step is including the point according to the minimum distance to the group center. The means are firstly defined using this formula:

\[ S_i^{(t)} = \{X_n : \|X_n - \text{mean}_i\| \leq \|X_n - \text{mean}_{n}^{(t)}\| \forall n^* = 1....K \} \]

The forth step is updating the predefined means using the following formula:

\[ \text{mean}_i^{(t+1)} = \frac{1}{S_i^{(t)}} \sum_{X_n \in S_i^{(t)}} X_n \] The convergence of this algorithm is defined when the K centers do not change their locations.

4. Proposed Algorithm

Our proposed idea aims at improving a new developed version of GAF. Based on K- means approach. Where, the sensors are distributed using K-means (Figure 10). We determine the gravity center and we attribute each sensor to its grid, considering the least distance to the active node. So that, Grid are formed and active nodes are selected, for transmitting their obtained data, from environment to the sink. As a first step, sensor node have been scattered, gridded, then in each cell we find its center of gravity. In a second step, Clusters are formed and active node are elected, considering two criteria of interest: maximum remaining energy and minimum distance to the base station.

4.1 Determination of Gravity Center

The Oghlidos distance form used in order to determine the center of gravity inside each grid, which is based on distance of each node from \((x,y)\) axis environment.

\[ \text{Gravity Center} = \sqrt{(x_2 - x_1) + (y_2 - y_1)} \]

4.2 Center of Gravity

In our case, the center of gravity is a selected related to the closest nodes to build a zone then, for selection of active node in each grid, there are different parameters that will be used in this election24. The active node selection in each grid based on three principal criteria:

- The node having the maximum energy level.
- The node having the minimum distance to the sink.
- The node having the minimum distance with the remaining normal nodes in the same grid.

Considering the previous declared criteria, having the most energy level has an important role. Hence, from the relation based on criteria for selection of the best active node25 we get.

\[ A.N = \frac{RE}{(\sum_{j=1}^{i} Dst^2 + Nb + Tp) + (1 - \frac{Db}{100})^2} \]

Where AN= Active node, RE= Residual energy extent of VI node, Dst= distance separating node to node, Nb=Number of bits, Tp= Transmission power,
By the use of this relation, we can select the best active node.

Once we create grid and select active nodes of each grid, we must transmit data by active node, already selected to neighboring active node to send it to the base station.

5. Simulation and Results

5.1 Simulation Parameters

We present in this part our simulation results of our proposed algorithm K-means GAF compared to the Basic GAF protocol considering different metrics such as number of alive nodes and energy consumption. The initial distribution nodes in our network shown in Figure 10, also the simulation parameters are according to the following table 1.

| Parameter                  | Value          |
|----------------------------|----------------|
| Simulation area            | 100*100        |
| Number of nodes            | 100            |
| Base station location      | 100,100 m      |
| Number of active nodes     | 9              |
| Channel Type               | Wireless       |
| Node's initial energy      | 0.3            |

5.2 Simulation Results

Our Simulation results exposed below based on number of dead nodes and energy consumed.

We observe from graphs shown in Figure 11 and 12 that our Kmeans GAF is more performant than the original GAF. In addition, number of dead nodes is more in Basic version of GAF. Our improved version called Kmeans GAF is a modified and improved version of GAF, it has fewer numbers of Dead Nodes and it demonstrates well results comparatively to the Basic GAF as shown in Figure 11.

![Figure 11. Dead nodes comparison graphs.](image1)

As in Figure 12, the comparison graphs of energy consumed demonstrate that the energy consumed in Basic GAF is more comparatively to the Kmeans GAF.

![Figure 12. Energy consumption comparison.](image2)

The use of Kmeans approach for distribution nodes in our case allows reducing energy because it considers both energy and distance parameters.

In addition, our proposed protocol selects the active node using maximum remaining energy and minimum distance separating the member node and the Base Station (BS). This efficient Active node selection resolved the random selection procedure considered in the basic GAF. Consequently, the energy consumption is minimized even if the BS is located farthest than Active nodes.
6. Conclusion

In order to design a new enhanced version of Basic GAF, which is able to decrease energy consumed. Our application of Kmeans approach allows decreasing energy consumed due to the smart distribution of nodes in the areas of interest with the use of center of gravity. In addition, the selection of active node considering both criteria maximum energy and minimum distance permits the decrease of energy consumption, which allows prolonging the network lifetime.

Therefore, the comparison of our new improved algorithm Kmeans GAF with the basic GAF algorithm, assume that not only the number of dead node are decreased and network life time are prolonged very much, but also the all energy consumption of the total network will be also reduced.

7. Acknowledgements

Many thanks, and sincere gratitude expressed to all colleagues in Laboratory of LAMAI in FSTG Marrakech, Morocco.

8. References

1. Akyildiz F, Su W, Sankarasubramaniam Y, Cayirci E. A survey on sensor networks in IEE Communications Magazine. 2002; 40(8):102–14.
2. Abdullah M. Issue 2 (2014) Routing protocols for wireless sensor networks: classifications and challenges. Quest Journals, Journal of Electronics and Communication Engineering Research. 2014; 2(2):5–15.
3. Intanagonwiwat C, Govindan R, Estrin D. Directed diffusion: a scalable and robust communication paradigm for sensor networks. Proceedings of ACM Mobi- Com; 2000, Boston, MA; 2000. p. 56–67.
4. Braginsky D, Estrin D. Rumor routing algorithm for sensor networks. Proceedings of the First Workshop on Sensor Networks and Applications (WSNA), Atlanta, Georgia, USA; 2002. p. 22–31.
5. Kulik J, Heinzelman WR, Balakrishnan H. Negotiation-based protocols for disseminating information in wireless sensor networks. Wireless Networks. 2002; 8:169–85.
6. Yao Y, Gehrke J. The cougar approach to in network query processing in sensor networks, Special Interest Group on Management of Data (SIGMOD) Record. 2002; 31(3): 9–18.
7. Sadagopan N et al. The acquire mechanism for efficient querying in sensor networks. Proceedings of the First International Workshop on Sensor Network Protocol and Applications, Anchorage, Alaska; 2003.
8. Aziz L, Raghay S, Aznaoui H, Jamali A. A new enhanced version of VLEACH protocol using a smart path selection. International Journal of GEOMATE. 2017 Feb; 12(30): 28 –34.
9. Aziz L, Raghay S, Jamali A. A New improved algorithm of LEACH protocol for wireless sensor networks. In Proceedings of 2014 Mediterranean Microwave Symposium (MMS), Marrakech, Morocco; 2014.
10. Aziz L, Raghay S, Aznaoui H, Jamali A. Comparative analysis of LEACH and V-LEACH protocols in wireless Sensor Networks, International Journal of Computer Science and Information Security. 2016; 14(4):112–9.
11. Rama Lakshmi G, Srikanth V. Location-Based Routing Protocol in Wireless Sensor Network-A Survey. 2015; 5(4).
12. Aziz L, Raghay S, Aznaoui H, Jamali A. A new approach based on a genetic algorithm and an agent cluster head to optimize energy in Wireless Sensor Networks. International Conference on Information Technology for Organizations Development (IT4OD), Fez, Morocco; 2016.
13. Aznaoui H, Raghay S, Aziz L, Ait-Mlouk A. A comparative study of routing protocols in WSN. 5th International Conference on Information and Communication Technology and Accessibility (ICTA), Marrakech, Morocco; 2015.
14. Li L, Halpern JY. Minimum-energy mobile wireless networks revisited. Proceedings IEEE ICC'01, Helsinki, Finland; 2001 Jun. p. 278–83.
15. Xu Y, Heidemann J, Estrin D. Geography-informed energy conservation for ad hoc routing. In Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom ’01), Rome, Italy; 2001 Jul. p. 70–84.
16. Shang F, Liu J. Multi-hop topology control algorithm for wireless sensor networks. Journal of Networks. 2012; 7(9):1407–14.
17. Chen Z, Ouyang Y, Liu Q. A GAF Algorithm of Dislocated Grid for Wireless Sensor Network, volume 163 of the series Advances in Intelligent and Soft Computing 2013. p. 721–8.
18. Soni V, Mallick K. A novel scheme to minimize hop count for GAF in wireless sensor networks: two-level GAF. Journal of Computer Networks and Communications. 2015; 2015:9.
19. Liu G, Wen W. A improved GAF clustering algorithm for three-dimensional underwater acoustic networks. In Computer Communication Control and Automation (3CA). International Symposium. 2010; 1:508–12.
20. Inagaki T, Ishihara S. HGAF:A power saving scheme for wireless sensor networks. Information Processing Society of Japan Journal. 2009 Oct; 50(10):2520–31.
21. You J, Lieckfeldt D, Salzmann J, Timmermann D. GAF and Co: Connectivity aware topology management for sensor networks. IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC); 2009. p. 2260–4.

22. Grover J, Shikha, Sharma M. Optimized GAF in wireless sensor network. Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions). 3rd International Conference on, Noida; 2014. p. 1–6.

23. Sasikumar P, Khara S. K-means clustering in wireless sensor networks. Fourth International Conference on Computational Intelligence and Communication Networks; 2012.

24. Elkamel R, Baccar N, Cherif A. Improvements in LEACH based on K-means and Gauss algorithms, Procedia Computer Science. 2015; 73.

25. Atefeh H, Amir Abbas B, Elham R. Improved k-means Algorithm Method for Reducing Energy Consumption in Wireless Sensor Networks. J. Basic. Appl. Sci. Res. 2013; 3(9):534–42,