Evaluation Of Optimal Cluster Head Placement For Sensor Network Using Lloyd's K-Means Technique

Maduka N.C. [1]

Department 0f Physics
Federal University Gusau,
Zamfara State Nigeria
madukanosike@fugusau.edu.ng

Ezuruike Okafor S. F. [2]

Department of Electrical/Electronic Engineering
Imo State Polytechnic,
Omuma, Imo State, Nigeria

Nna Ernest Maagnes[3]
Department of Computer Engineering
Captain Elechi Amadi Polytechnic,
Port Harcourt, Rivers State Nigeria
maagenijchrist@gmail.com

***Abstract*— In this paper, the evaluation of optimal cluster head placement for sensor network using Lloyd's k-means technique is presented. The essence is to visualize the variation in the optimal cluster head placement when the Lloyd's K-Means clustering (LKMC) algorithm is implemented repeatedly for the same set of dataset points. A case study dataset consisting of 2500 data points (sensor nodes location coordinates) randomly distributed in an area of 900 m by 900 m is clustered into 5 clusters using the K-Means clustering (LKMC) algorithm. The clustering was repeated 4 times and each of the results were captured and the mean Euclidian distance of data points in all the clusters in the network is determined for each implementation. The results show that the Average Euclidian Distance per implementation varied from 874.1 m to 900 m which has the range of 25.9 m. There is about 2.8 % difference between the minimum averages Euclidian Distance among the different implementations of the LKMC algorithm on the same dataset. Hence, when using the LKMC algorithm, it is recommended that the algorithm be applied several times and the implementation with the minimum Average Euclidian Distance can be adopted.**

***Keywords*— Cluster Head, Clustering Algorithm, Sensor Node, Cluster Head Placement, Sensor Network, Lloyd's K-Means Technique**

1. INTRODUCTION

Over the years, clustering has been used as a means of optimizing the spatial placement of cluster head in clustered network of nodes [1,2,3]. The cluster head placement is normally determined using clustering algorithm [3,4,5]. Such algorithms use one or more approaches to determine the spatial location of the cluster head which will minimize the specified distance metric used in the algorithm [6,7,9]. By doing so, it is expected that the mean of the distance of each of the nodes in a given cluster is minimal.

Several clustering algorithms have been developed. Among them is the Lloyd's k-means technique which usually employ Euclidian distance for its operation [10,11,12]. The Lloyd's k-means requires the number of clusters, k to be specified and it will iteratively assigned and re-assign cluster heads among the nodes until the optimal k cluster heads are determined based on the Euclidian distance [13,14]. In practice, repeated execution of the Lloyd's k-means algorithm may result is a different set of cluster heads and hence different mean Euclidian distance per cluster. As such, in this paper, repeated execution and comparative evaluation of the resultant cluster heads placements for a given dataset is conducted. The essence of the study is to provide requisite idea that can help in the application of the Lloyd's k-means technique for optimal cluster head placement for sensor network. In this case, repeated implementation of the Lloyd's k-means algorithm may be required and the optimal solution is selected from the various results obtained from the different instances of implementation of the Lloyd's k-means algorithm on the same dataset.

2. METHODOLOGY

2.1 The Lloyd's K-Means clustering (LKMC) technique

The focus in this paper is to evaluate variability of the optimal cluster head placement for a given set of sensor nodes based on the Lloyd's K-Means clustering (LKMC) technique. The essence is to visualize the variation in the optimal cluster head placement when the LKMC algorithm is implemented repeatedly for the same set of dataset. Notably, when the LKMC algorithm is implemented repeatedly on the same set of data points, particularly where there is densely packed data points within the coverage area. It is found that the cluster head selected in the different implementations of the LKMC algorithm give rise to different set of cluster heads. The variation gives rise to different mean optimal Euclidian distance for the LKMC algorithm implementation. The LKMC algorithm is given in Algorithm 1.

Algorithm 1: The Lloyd's K-Means algorithm [15]

Step 1: Set the number of lusters as K

Step 2: Read in the data points $x_i \in X$ ori= 1, 2, 3… n where n is the number of data points

Step 3: Randomly select the initial K number of centroids $c_i \in C$ for i= 1, 2, 3… k

Step 4: Assign each of the $x_i \in X$ to the nearest centroid $c_i \in C$ for i= 1, 2, 3… k using the Euclidian distance

Step 5: Update each of the cluster centroids $c_i \in C$ for i= 1, 2, 3… k

Step 6: Repeat step 4 and step 5 until convergence occurs (when there is no further changes in the points in each of the k clusters

2.2 Computation of the mean Euclidian distance of data points in all the clusters in the network

In this study, the mean optimal Euclidian distance for the LKMC algorithm implementation is computed as using the centroid coordinates (Cx_k, Cy_k) and the data points coordinates for each cluster denoted as $(x_{k,i}, y_{k,i})$ where, Cx_k denote the x coordinate of the centroid in cluster k and Cy_k denote the y coordinate of the centroid in cluster k, where k = 1,2,3…, nk where nk is the number of clusters in the network. Again, $x_{k,i}$ denote the i^{th} x coordinate of the data point in cluster k and $y_{k,i}$ denote the i^{th} y coordinate of the data point in cluster k , where i = 1,2,3…, mk where mk is the number of data points in clusters k. The Euclidian distance of data point i in cluster k is denoted as $d_{k,i}$ where;

$$d_{k,i} = \sqrt{\left(Cx_k - x_{k,i}\right)^2 + \left(Cy_k - y_{k,i}\right)^2} \quad (1)$$

The mean Euclidian distance of data points in cluster k is denoted as d_k where;

$$d_k = \left(\frac{1}{mK}\right)\left(\sum_{i=1}^{i=mK}\left(d_{k,i}\right)\right) \quad (2)$$

The mean Euclidian distance of data points in all the clusters in the network is denoted as d_{AVG} where;

$$d_{AVG} = \left(\frac{1}{nK}\right)\left(\sum_{i=1}^{i=K}\left(d_k\right)\right) \quad (3)$$

2.3 The simulation of the clustering with the case study dataset using the K-Means clustering (LKMC) algorithm

A case study dataset consisting of 2500 data points (sensor nodes location coordinates) randomly distributed in an area of 900 m by 900 m, as shown in Figure 1. The K-Means clustering (LKMC) algorithm was used to determine the cluster heads where the number of clusters is set as 5. The clustering was repeated 4 times and each of the results were captured and the mean Euclidian distance of data points in all the clusters in the network is determined for each implementation using the set of Equation 1 to Equation 3.

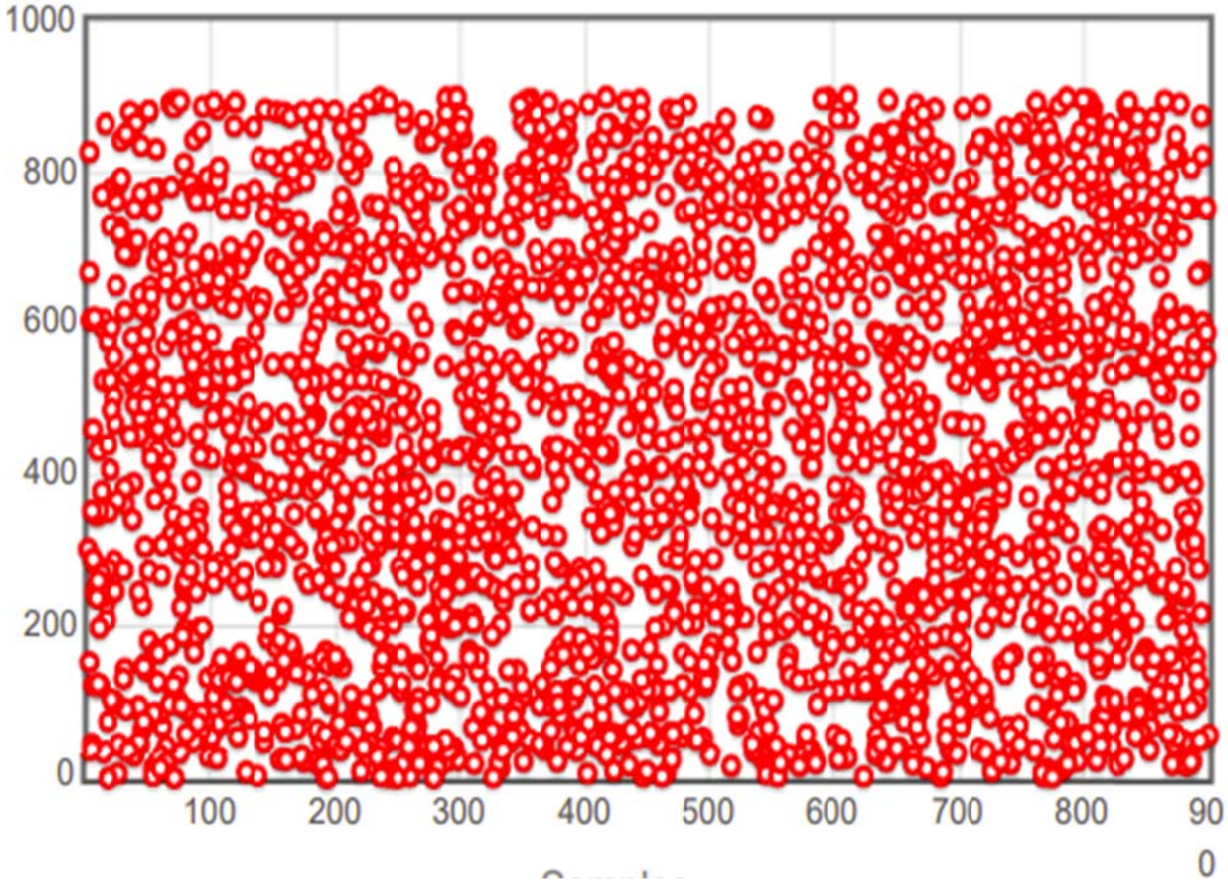

Figure 1 The scatter plot of the case study dataset consisting of 2500 data points randomly distributed in an area of 900 m by 900 m

3. RESULTS AND DISCUSSIONS

The results of the four different implementation of the (LKMC) algorithm on the case study dataset consisting of 5 clusters are shown in the scatter plots in Figure 2, Figure 3, Figure 4 and Figure 5. The results of the centroid (cluster head placement) obtained in the first implementation of the LKMC algorithm is shown in Table 1 while the summary of the number of sensor nodes in each cluster and the mean Euclidian distance for each cluster in the first implementation of the LKMC algorithm is presented in Table 2.

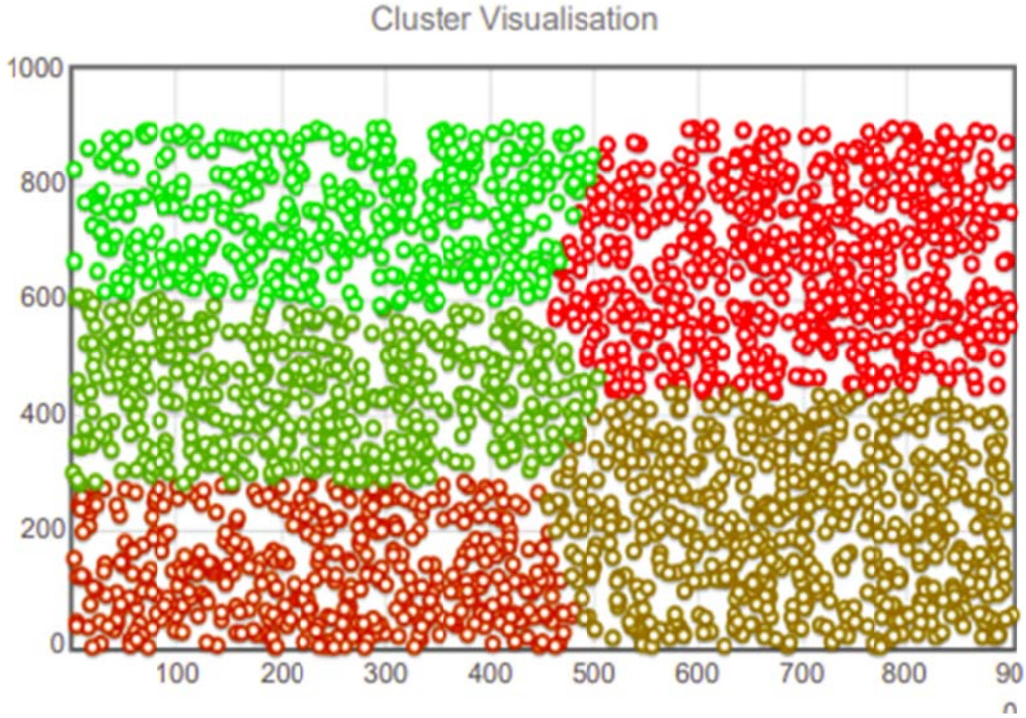

Figure 2 The scatter plot of the first implementation of the (LKMC) algorithm on the case study dataset consisting of 5 clusters

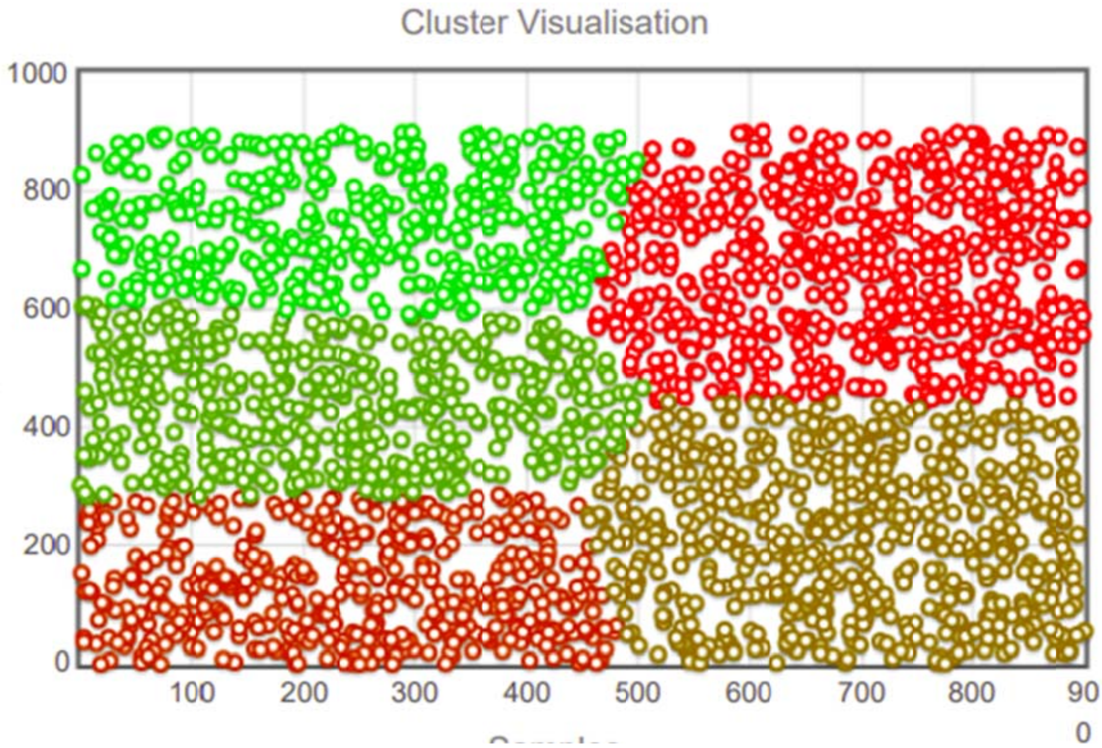

Figure 3 The scatter plot of the second implementation of the (LKMC) algorithm on the case study dataset consisting of 5 clusters

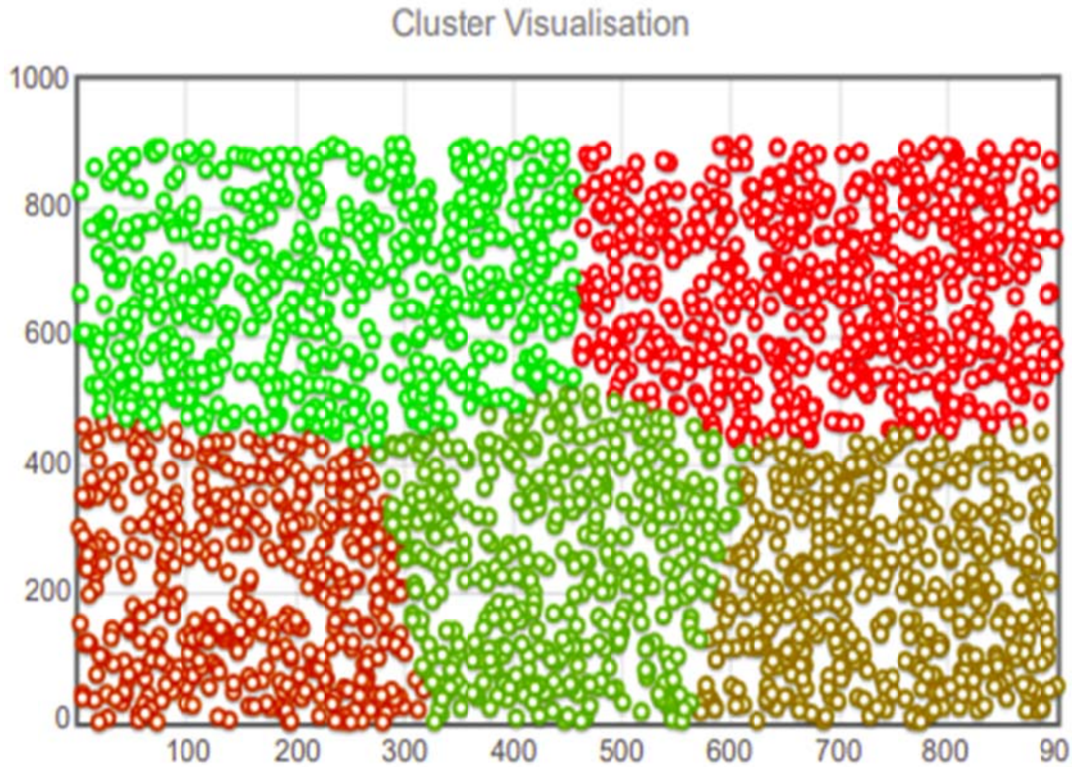

Figure 4 The scatter plot of the third implementation of the (LKMC) algorithm on the case study dataset consisting of 5 clusters

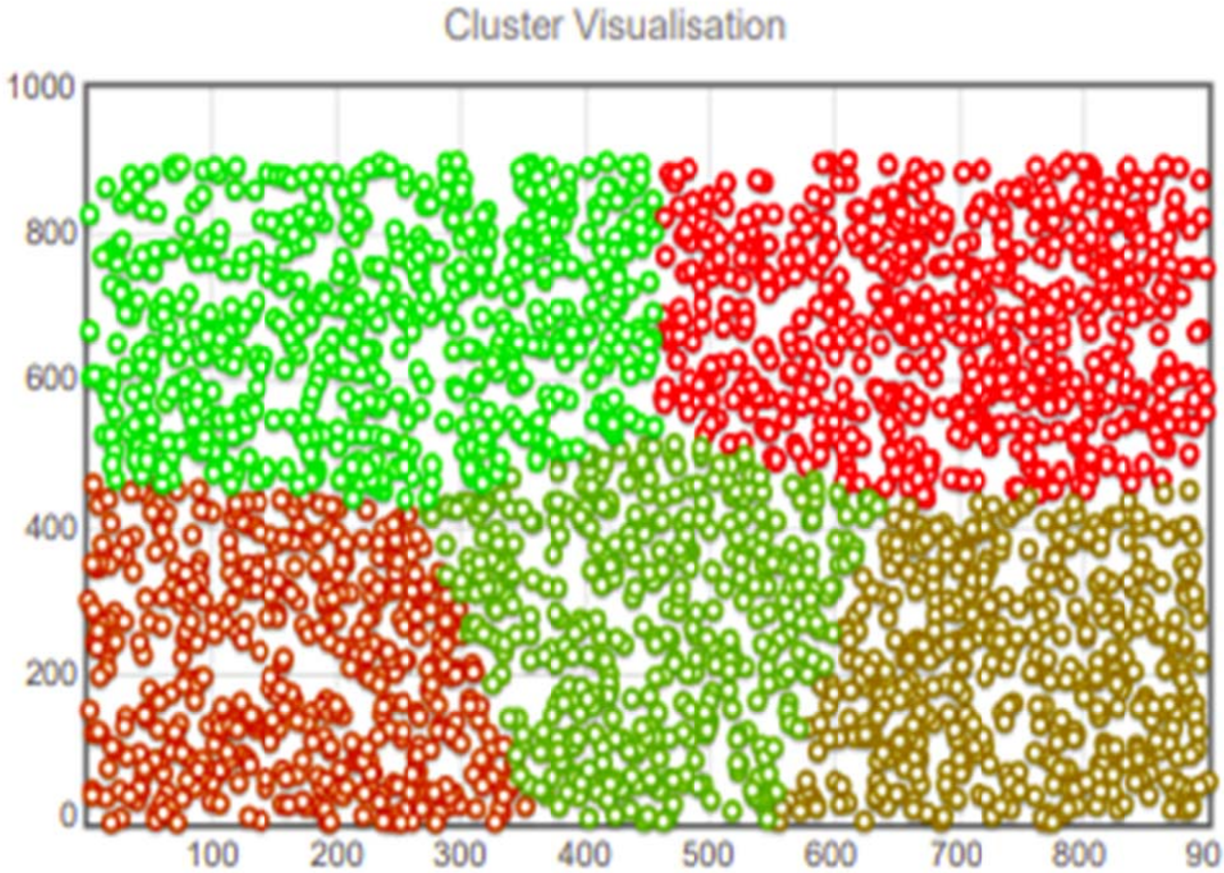

Figure 5 The scatter plot of the fourth implementation of the (LKMC) algorithm on the case study dataset consisting of 5 clusters

Table 1 The centroid (cluster head placement) obtained in the first implementation of the LKMC algorithm

	The centroid (cluster head placement)	
Cluster K	Cx_k	Cy_k
0	689.735	217.444
1	256.76	749.556
2	693.52	667.452
3	232.783	438.558
4	241.476	132.148

Table 2 The summary of the number of sensor nodes in each cluster and the mean Euclidian distance for each cluster in the first implementation of the LKMC algorithm

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Min	Max	Average
No. of nodes	**610**	**433**	**593**	**452**	**412**	**412**	**610**	**500**
Percentage of total nodes	**24.4**	**17.3**	**23.7**	**18.1**	**16.5**	**16.5**	**24.4**	**20.0**
Maximum Euclidian distance, d_k (m)	**230.3**	**467.6**	**226.4**	**14.1**	**463.0**	**14.1**	**467.6**	**280.3**
Minimum Euclidian distance, d_k (m)	**698.9**	**785.4**	**678.7**	**791.7**	**794.2**	**678.7**	**794.2**	**749.8**
Average Euclidian distance, d_{AVG} (m)	**465.7**	**632.0**	**466.2**	**2178.2**	**631.8**	**465.7**	**2178.2**	**874.8**

Table 3 The centroid (cluster head placement) obtained in the four different implementations of the LKMC algorithm

	The centroid obtained in the first implementation of the LKMC algorithm		The centroid obtained in the second implementation of the LKMC algorithm		The centroid obtained in the third implementation of the LKMC algorithm		The centroid obtained in the fourth implementation of the LKMC algorithm	
Cluster K	Cx_k	Cy_k	Cx_k	Cy_k	Cx_k	Cy_k	Cx_k	Cy_k
0	695.2	665.0	690.4	676.3	692.9	666.0	689.2	675.3
1	241.5	132.1	162.8	210.1	241.5	132.1	154.9	215.7
2	690.5	215.5	742.3	207.3	690.4	215.9	741.2	212.0
3	232.7	438.9	450.3	263.0	232.8	438.6	440.1	248.6
4	260.7	749.5	232.2	677.1	256.8	749.6	232.8	676.8

The results of the centroid (cluster head placement) obtained in the four different implementation of the LKMC algorithm are shown in table 1 while the scatter plot of the centroid (cluster head placement) obtained in the four different implementation of the LKMC algorithm is given in Figure 6. It is shown in the results in Table 3 and Figure 6 that each of the implementations has set of cluster head placements different from the ones obtained in the other implementations. The implication of the variation in the cluster head placements is that the mean Euclidian distance will differ in each implementation. This is captured in the results presented in Table 4, Figure 7 and Figure 8. The results in Table 4 , Figure 7 and Figure 8 show that the Average Euclidian Distance, d_{AVG} (m) per implementation varied from 874.1 m to 900 m which has the range of 25.9 m. There is about 2.8 % difference between the minimum averages Euclidian Distance among the different implementations of the LKMC algorithm on the same dataset. Hence, when using the LKMC algorithm, it is recommended that the algorithm be applied several times and the implementation with the minimum Average Euclidian Distance, d_{AVG} (m) can be adopted.

Figure 6 The scatter plot of the centroid (cluster head placement) obtained in the four different implementation of the LKMC algorithm

Table 4 The summary of the Average Euclidian Distance, d_k (m) for each of the 5 clusters in the 4 different implementations and also the Average Euclidian Distance, d_{AVG} (m) per implementation

	Average Euclidian distance, d_k (m) for Cluster 0	Average Euclidian distance , d_k (m) for Cluster 1	Average Euclidian distance, d_k (m) for Cluster 2	Average Euclidian distance, d_k (m) for Cluster 3	Average Euclidian distance, d_k (m) for Cluster 4	Minimum Euclidian distance, d_{kmin} (m)	Maximum Euclidian distance, d_{kmax} (m)	Average Euclidian distance, d_{AVG} (m)
Implementation 1	465.7	632.0	466.2	2178.2	631.8	465.7	2178.2	874.8
Implementation 2	466.9	656.6	416.9	2230.9	722.9	416.9	2230.9	898.9
Implementation 3	467.8	632.0	466.6	2177.6	626.5	466.6	2177.6	874.1
Implementation 4	468.3	648.1	418.2	2241.8	723.6	418.2	2241.8	900.0

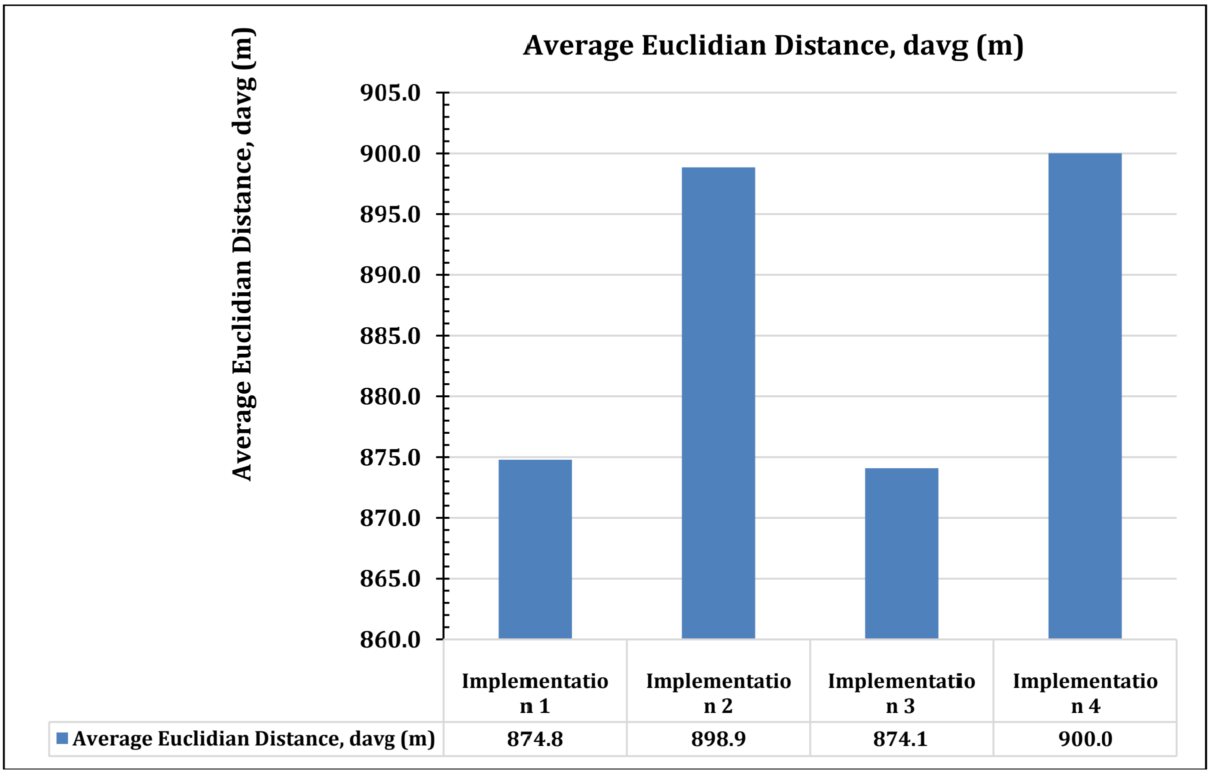

Figure 7 The Average Euclidian Distance, d_{AVG} (m) Per Implementation

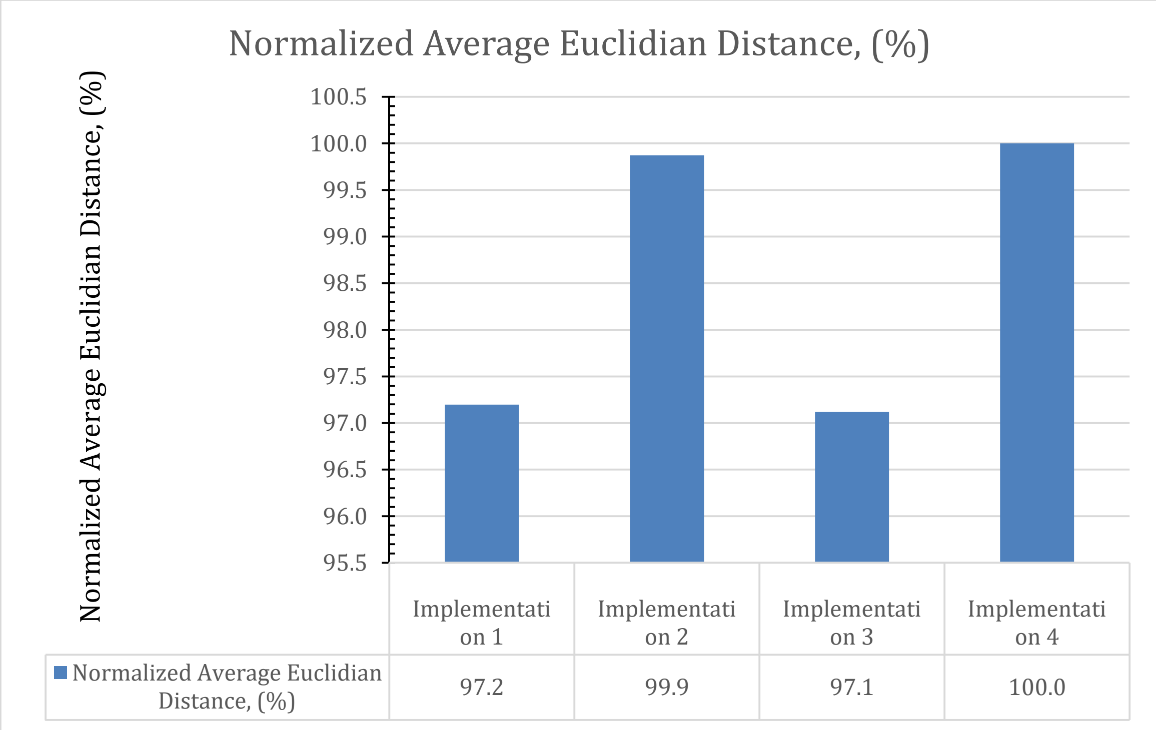

Figure 8 The Normalized Average Euclidian Distance, (%) Per Implementation

4. CONCLUSION

The Lloyd's K-Means clustering (LKMC) technique for optimal cluster head placement in a sensor network is presented. The LKMC algorithm is implemented on one dataset for four different times. The essence of the repeated implementation is to visualize the variation in the optimal cluster head placement from one implementation to the other. The Euclidian distance of each of the sensor node in each cluster is first computed, then the average Euclidian distance per cluster is computer and finally the average Euclidian distance for all the clusters in each implementation is computed. The results show that there is up to 2.8 % difference in the average Euclidian Distance among the different implementations of the LKMC algorithm on the same dataset. Therefore, in practice, the LKMC algorithm need to be repeatedly implemented and then the one that gives the minimum average Euclidian Distance is adopted.

REFERENCES

[1] Wohwe Sambo, D., Yenke, B. O., Förster, A., & Dayang, P. (2019). Optimized clustering algorithms for large wireless sensor networks: A review. *Sensors*, *19*(2), 322.

[2] Raj, B., Ahmedy, I., Idris, M. Y. I., & Md. Noor, R. (2022). A survey on cluster head selection and cluster formation methods in wireless sensor networks. *Wireless Communications and Mobile Computing*, *2022*(1), 5322649.

[3] Ahmad, M., Hameed, A., Ikram, A. A., & Wahid, I. (2019). State-of-the-art clustering schemes in mobile ad hoc networks: objectives, challenges, and future directions. *IEEE Access*, *7*, 17067-17081.

[4] Mehra, P. S., Doja, M. N., & Alam, B. (2020). Fuzzy based enhanced cluster head selection (FBECS) for WSN. *Journal of King Saud University-Science*, *32*(1), 390-401.

[5] Wohwe Sambo, D., Yenke, B. O., Förster, A., & Dayang, P. (2019). Optimized clustering algorithms for large wireless sensor networks: A review. *Sensors*, *19*(2), 322.

[6] Lin, D., & Wang, Q. (2019). An energy-efficient clustering algorithm combined game theory and dual-cluster-head mechanism for WSNs. *IEEE Access*, *7*, 49894-49905.

[7] Wohwe Sambo, D., Yenke, B. O., Förster, A., & Dayang, P. (2019). Optimized clustering algorithms for large wireless sensor networks: A review. *Sensors*, *19*(2), 322.

[8] Al-Sulaifanie, A. I., Al-Sulaifanie, B. K., & Biswas, S. (2022). Recent trends in clustering algorithms for wireless sensor networks: A comprehensive review. *Computer Communications*, *191*, 395-424.

[9] Li, Y., Cai, J., Yang, H., Zhang, J., & Zhao, X. (2019). A novel algorithm for initial cluster center selection. *ieee Access*, *7*, 74683-74693.

[10] Ostrovsky, R., Rabani, Y., Schulman, L. J., & Swamy, C. (2013). The effectiveness of Lloyd-type methods for the k-means problem. *Journal of the ACM (JACM)*, *59*(6), 1-22.

[11] Hamerly, G., & Drake, J. (2015). Accelerating Lloyd's algorithm for k-means clustering. *Partitional clustering algorithms*, 41-78.

[12] Slonim, N., Aharoni, E., & Crammer, K. (2013). Hartigan's K-means vs. Lloyd's K

means–is it time for a change?. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI).*

[13] Mohd, W. M. B. W., Beg, A. H., Herawan, T., & Rabbi, K. F. (2012). An improved parameter less data clustering technique based on maximum distance of data and lioyd k-means algorithm. *Procedia Technology*, *1*, 367-371.

[14] Wilkin, G. A., & Huang, X. (2008). A practical comparison of two K-Means clustering algorithms. *BMC bioinformatics*, *9*, 1-5.

[15] Abernathy, A., & Celebi, M. E. (2022). The incremental online k-means clustering algorithm and its application to color quantization. *Expert Systems with Applications*, *207*, 117927.