Journal of information and
communication convergence engineering
http://jicce.org

J. Inf. Commun. Converg. Eng. 20(4): 259-264, Dec. 2022 Regular paper

Performance Evaluation of k-means and k-medoids in WSN Routing Protocols

SeaYoung Park[1], Dai Yeol Yun[2], Chi-Gon Hwang[2], and Daesung Lee[3]*, *Member, KIICE*

[1]Department of Immersive Content Convergence, KwangWoon University Graduate School, Seoul 01897, Korea
[2]Department of information and communication Engineering, Institute of Information Technology, Kwangwoon University, Seoul 01897, Korea
[2]Department of Computer Engineering, Institute of Information Technology, Kwangwoon University, Seoul 01897, Korea
[3]Department of Computer Engineering, Catholic University of Pusan, Busan 46252, Korea

Abstract

In wireless sensor networks, sensor nodes are often deployed in large numbers in places that are difficult for humans to access. However, the energy of the sensor node is limited. Therefore, one of the most important considerations when designing routing protocols in wireless sensor networks is minimizing the energy consumption of each sensor node. When the energy of a wireless sensor node is exhausted, the node can no longer be used. Various protocols are being designed to minimize energy consumption and maintain long-term network life. Therefore, we proposed KOCED, an optimal cluster K-means algorithm that considers the distances between cluster centers, nodes, and residual energies. I would like to perform a performance evaluation on the KOCED protocol. This is a study for energy efficiency and validation. The purpose of this study is to present performance evaluation factors by comparing the K-means algorithm and the K-medoids algorithm, one of the recently introduced machine learning techniques, with the KOCED protocol.

Index Terms: Performance Evaluation, WSN, KOCED, K-means, K-medoids

I. INTRODUCTION

The Information Technology (IT) is widely applied to business, production, defense, public, and education. By embedding micro-computing devices into objects and providing ubiquitous computing, life is changing in a convenient and prosperous way. By making things intelligent by remote computers, users can more accurately perceive the actual situation. A wireless sensor network (WSN) is a small wireless transceiver network system that processes information collected by sensors and transmits them to a processor in combination with a ubiquitous computing technique. That is, it is a network composed of a base station for collecting data sensed by a sensor node and transmitting data outside the sensor space [1-2].

Wireless sensor protocols using K-means clustering do not form a cluster after selecting a cluster head, but construct a cluster first. This technique has the advantage that the cluster is formed evenly, so that most of the member nodes belonging to the cluster are uniformly present. In this method, after cluster configuration, nodes with a large amount of residual energy or nodes close to the cluster center point were selected as cluster heads. However, there is a disadvantage in that a node far from the base station becomes a cluster head or the same node continuously becomes a cluster head, resulting in rapid FND generation [3-6].

Received 03 December 2022, **Revised** 07 December 2022, **Accepted** 18 December 2022
***Corresponding Author** Daesung Lee (E-mail: dslee@cup.ac.kr)
Department of Computer Engineering, Catholic University of Pusan, Busan 46252, Korea

Open Access **https://doi.org/10.56977/jicce.2022.20.4.259** print ISSN: 2234-8255 online ISSN: 2234-8883

In this paper, we compare the performance evaluation with K-means algorithm and K-medoids algorithm for KOCED protocol, which improved the problem of K-means clustering. KOCED protocol is a clustering algorithm that applies the K-means algorithm for energy efficiency, and the cluster head election is made by optimizing ($k = \sqrt{(N/2\pi)}\sqrt{(\in\_fs/\in\_mp)}\, M/[d\_toBS]^2$) election. In addition, when clustering is actually performed, if the cluster has too many or too few nodes, the cluster head selects a node with less residual energy, which quickly fails to transfer data as well as the first node dead (FND). We improve the problem of selecting a node with little residual energy as a cluster head in consideration of the energy term in the election probability threshold equation. For the KOCED protocol, which improves the K-average clustering problem, we want to find out whether there are advantages in a specific environment by comparing performance evaluations with the K-average algorithm and the K-medoid algorithm. We also present the advantages and disadvantages of the K-means and K-medoid algorithms. And I would like to present the performance evaluation index.

The rest of this paper is as follows. Section 2 introduces related research. Details of the K-means algorithm, K-medoids algorithm clustering methods, and their pros and cons are provided in Section 3. Section 4 evaluates the performance of KOCED, the K-means algorithm, and the K-medoids algorithm. Finally, the thesis is concluded in section 5.

II. RELATED METHOD

A. K-means Algorithms

The K-means clustering [7-8] is one of the representative segregated clustering algorithms, and although its principle is simple, it has good performance. Each cluster of K-means clustering has one center. Each member node belongs to a central point whose distance is close. Member nodes allocated to the same central point are gathered to form a cluster. For K-mean clustering, you must determine the number of clustering in advance. In general, the larger the value, the greater the number of clusters. And the smaller the value, the smaller the number of clusters. Therefore, the determination of the number of clusters is very important. The typical method is empirical rule. The number of sensor nodes is calculated as the number of clusters required as shown in the following equation (1).

$$k \cong \sqrt{n/2} \tag{1}$$

1) KC Protocols

The K-means Centrality (KC) protocol is one of the wireless sensor network protocols that use the K-means clustering algorithm. After dividing the sensor space into clusters, a node near the center of the cluster is selected as the cluster head. The KC protocol selects the node closest to the cluster center point among the nodes with the smallest number of cluster head elections as the final cluster head. Therefore, all nodes in the cluster can rotate once and be elected as the cluster head, and energy consumption can be distributed accordingly, thereby increasing the energy efficiency of the network.

2) KCE Protocols

The case protocol KC protocol concentration is similar to the energy KCE protocol, as well as the remaining energy of the node selected by the cluster head. The node is considered residual energy after considers. The advantages of the Casey protocol and the KCE protocol are the same size, and the energy efficiency of the cluster is the same as that of the cluster head height by selecting a node at the cluster center point. The disadvantage is the numerical optimization of groups due to cluster heads .It does not consider distance to base stations that consume a lot of energy to consider.

3) KOCED Protocols

The KOCED protocol compensates for the shortcomings that clustering takes a long time. It was limited to the time of occurrence This compensates for the disadvantage of increasing the required time because it is not necessary to undergo a clustering process every round.

B. K-medoids Algorithms

The k-medoids [9-11] problem is a clustering problem similar to the k-means. The k-means and k-medoid algorithms attempt to construct clusters by minimizing the distance between points designated as the center of the cluster. Unlike k-means, the k-medoid algorithm chooses around the actual data points, so it can interpret cluster centroids better than k-means, where the centroid of the cluster doesn't have to be one of the input data points. Also k-means can be used with any difference measure, although it usually requires a Euclidean distance for an efficient solution. k-medoid is more robust against noise and outliers than k-means because it minimizes the sum of pairwise discrepancies instead of the Euclidean sum of squares. k-medoids is a classic partitioning technique in clustering that divides a data set of n objects into k clusters. Here, the number of clusters k known a priori (programmers must specify k before running the k-medoid algorithm. Meaning.) The "lead" of a given value of k can be evaluated in the same way as the silhouette method. The media in a cluster is defined as the object in the cluster that has the smallest mean difference from all objects in the cluster. That is, it is the most central point of the cluster.

1) KCA Protocols

The KCA Algorithm proposed a K-medoid based Clustering Algorithm (KC) to obtain a universal clustering method to reduce energy consumption and extend network lifespan. In the proposed method, all node sensing data is collected through a cluster head node using a base station. First, we collect the node coordinates and residual energy information, and then compute the cluster number k. Optimize the K-medoids algorithm to shorten the iteration time by calculating the mean point and residual energy of the central circle. In the proposed scheme, the algorithm includes two steps: a setup phase and a communication phase.

2) PAM Protocols

The PAM (Partitioning Around Medoids) is a segmentation method that is mainly used for areas that require robustness for outlier data, random distance measurement criteria, or areas where there is no clear definition of median or median. This is the same as the k-means, with the two approach goals subdividing the measurement set/description in k-subset/cluster to subset the sum of the distance between the measurement and the measurement cluster center. In a k-means algorithm, the subset centroid is the average of measurements over a subset, often called centroid. In the k-medoids algorithm, the subset centroid is one that does not have a subset called medoid. The k-medoids algorithm returns medoids, which are the actual data points in a data set. This allows the algorithm to be used in conditions where the data mean does not cover the data set. This is at the heart of k-medoids and k-means algorithms where centroids are returned via k-means algorithms that cannot contain data sets. Therefore, the k-medoids algorithm is useful for clustering unambiguous data when the mean is difficult to explain/understand.

3) CODS-KM Protocols

The CODS-KM (Collection Oriented Distributed Scheme for WSN-K-Medoids) moves through the cluster head (CH) to an access point (RP) called a collection point (CP) for all communications, and then directly to the sink (receive node). Move. Here, the sink will act as a mobile sink that collects information at the collection point within a short period of time. The proposed system suggests mitigation of travel time during transmission and finally collects all the information at the collection point and continues until the end of the simulation process.

The proposed approach to solving the described problem works by dividing the system in each sector and mobile component. This allocation takes into account the propagation of nodes to keep a strategic distance from long separations. The proposed approach is used to create mobile element visits by classifying the node array. We propose the use of a collection point-based algorithm to obtain this set. Once the visiting node is recognized, the proposed approach begins by dividing the system into two parts.

III. PERFORMANCE COMPARISON EVALUATION OF K-MEANS AND K-MEDOIDS ALGORITHMS

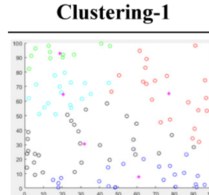
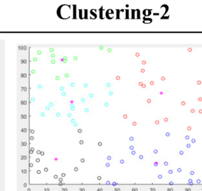
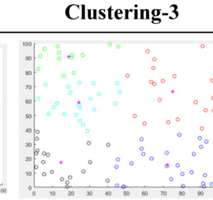
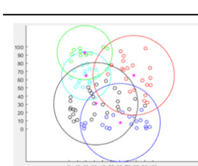
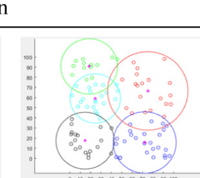
A. Performance of K-means and K-medoids

Compare k-means algorithm and k-medoids algorithm. First, we analyze the basics of k-means algorithm and applied k-means algorithms. In this regard, the advantages and disadvantages and performance comparison factors are to be identified. And we analyze the basics of k-medoids algorithm and applied k-medoids algorithms. In this regard, the advantages and disadvantages and performance comparison factors are to be identified. In the k-means algorithm, we look at the clustering configuration steps of k-mean and understand how clustering is formed. In addition, performance evaluation is conducted on KC, KCE, and KOCED algorithms, which are algorithms applying k-mean. In the k-medoids algorithm, we look at the steps of configuring the clustering of k-medoids to understand how clustering is formed. In addition, performance evaluation is performed on KCA, PAM, and CODS-KM algorithms, which are algorithms using k-medoid.

1) K-means Algorithm

The clustering construction step of the k-means algorithm is constructed using the EM algorithm. It consists of an expectation phase and a maximize phase. K-means clustering works iteratively until the EM algorithm converges. In K-means clustering, the EM algorithm is applied because it is necessary to simultaneously find the location of the center point of each cluster and the cluster to which each individual

Table 1. K-means Algorithm: Clustering Works

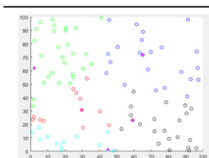
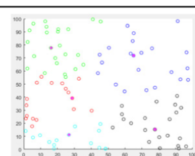
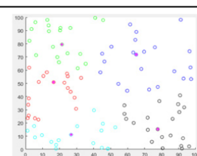
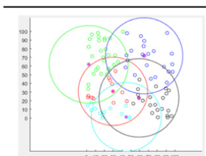
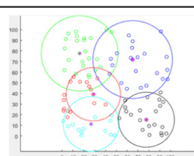
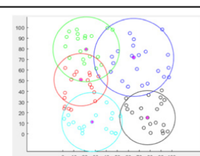
First : Clustering-1	Second : Clustering-2	Third : Clustering-3
		
Cluster Head Selection		
		
Clustering Region		

belongs. The clustering process of the k-means algorithm is as follows. As shown in Table 1 below, when the clustering operation is performed in step 1, the cluster head selection mark is displayed in purple. A clustering area is also displayed around the selected cluster head.

2) K-medoids Algorithm

The k-medoid algorithm is all partial (it divides the data set into groups) and tries to minimize the distance between a point in a cluster labeled as and a point designated as the center of that cluster. Unlike k-means algorithms, k-medoid chooses around the actual data points, so it can interpret cluster centroids better than k-means, where the centroid of the cluster doesn't have to be one of the input data points. The clustering process of the k-medoids algorithm is as follows. As shown in Table 2 below, when the clustering operation is performed in step 1, the cluster head selection mark is displayed in purple. A clustering area is also displayed around the selected cluster head.

Table 2. K-medoids Algorithm: Clustering Works

First : Clustering-1	**Second : Clustering-2**	**Third : Clustering-3**
		
Cluster Head Selection		
		
Clustering Region		

3) Simulations

We will try to derive the performance of k-means algorithm and k-medoids algorithm using MATLAB simulator. The values of the simulation parameters set for the simulation are shown in Table 3 below. In the case of a base station in a network environment, it was placed outside the sensor field. In the case of 100 sensors, it was assumed that their positions did not change after they were randomly arranged, and that their initial energy values were set to be the same. The size of the sensor fields is 100 × 100 m / 200 × 200 m / 400 × 400 m. And the positions of the base stations were arranged as 50, 150 / 100, 300 / 200, and 600.

Table 3. Simulation energy model

Parameter	Setting Value
Number of sensor nodes	100
E_{DA}	5 nJ/bit/signal
E_{elec}	50 nJ/bit
Initial energy of sensor node	0.5 J
ε_{fs}	10 pJ/bit/m^2
ε_{mp}	0.0013 pJ/bit/m^4
Size of the sensor field	100 × 100 / 200 × 200 / 400 × 400 M
Location of base station	50,150 / 100, 300 / 200, 600 point

B. Performance of KC & KCE & KOCED Protocols

We analyze the basics of k-medoids algorithm and applied k-medoids algorithms. In this regard, the advantages and disadvantages and performance comparison factors are to be identified. In the k-means algorithm, we look at the clustering configuration steps of k-mean and understand how clustering is formed. In addition, performance evaluation is conducted on KC, KCE, and KOCED algorithms, which are algorithms applying k-mean.

- KC (K-means Centrality) Protocol
- KCE (K-means Centrality with Energy) Protocol
- KOCED (K-means Centrality with considering, Energy, and Distance) Protocol

The following Table 4 and Table 5 are the results of comparing the KC, KCE, and KOCED protocols. In Table 1, each field was set to (100 * 100), (200 * 200), and (400 * 400) for comparison.

Table 4. Performance of KC, KCE, and KOCED protocols by field

Field 100 * 100	**Field 200 * 200**	**Field 400 * 400**

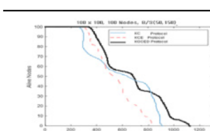
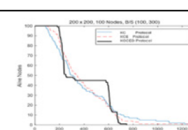
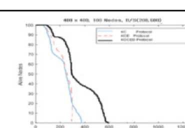

		

Figure 1-(a)	Figure 1-(b)	Figure 1-(c)

In Table 5, based on Figure 4, the time point rounds when FND (First Node Dead), HND (Half Node Dead), and LND (Last Node Dead) occur.

Table 5. FND, HND, LND for each field of KC, KCE, and KOCED protocols

FND	100 * 100	200 * 200	400 * 400
KC Protocol	346	161	98
KCE Protocol	374	173	101
KOCED Protocol	412	214	134
HND	**100 * 100**	**200 * 200**	**400 * 400**
KC Protocol	612	256	243
KCE Protocol	587	287	249
KOCED Protocol	734	243	311
LND	**100 * 100**	**200 * 200**	**400 * 400**
KC Protocol	912	1231	392
KCE Protocol	804	1124	305
KOCED Protocol	1123	789	598

We analyze the basics of k-medoids algorithm and applied k-medoids algorithms. In this regard, the advantages and disadvantages and performance comparison factors are to be identified. In the k-medoids algorithm, we look at the steps of configuring the clustering of k-medoids to understand how clustering is formed. In addition, performance evaluation is performed on KCA, PAM, and CODS-KM algorithms, which are algorithms using k-medoid.

- KCA (K-medoids based Clustering Algorithm) Protocol
- PAM (Partitioning Around Medoids) Protocol
- CODS-KM (Collection Oriented Distributed Scheme for WSN - K-Medoids) Protocol

The following Table 6 and Table 7 are the results of comparing the KCA, PAM, and CODS-KM protocols. In Figure 6, each field was set to (100 * 100), (200 * 200), and (400 * 400) for comparison.

Table 6. Performance of KCA, PAM, and CODS-KM protocols by field

Field 100 * 100	Field 200 * 200	Field 400 * 400

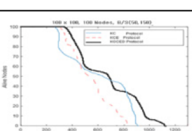
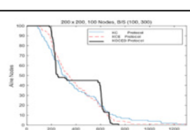
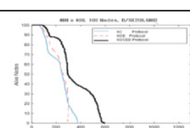

		

Figure 1-(a)	Figure 1-(b)	Figure 1-(c)

In Table 7, based on Table 6, the time point rounds when FND, HND and LND occur.

Table 7. FND, HND, and LND for each field of KCA, PAM, and CODS-KM protocols.

FND	100 * 100	200 * 200	400 * 400
KCA Protocol	772	95	9
PAM Protocol	845	161	13
CODS-KM Protocol	832	194	7
HND	**100 * 100**	**200 * 200**	**400 * 400**
KCA Protocol	1078	171	17
PAM Protocol	1132	321	19
CODS-KM Protocol	1115	398	24
LND	**100 * 100**	**200 * 200**	**400 * 400**
KCA Protocol	1122	284	1671
PAM Protocol	1190	371	601
CODS-KM Protocol	1192	481	709

IV. CONCLUSIONS

In wireless sensor networks, sensor nodes are often deployed in large quantities in places that are difficult to access by humans. It is difficult to supply power such as battery replacement or charging. Efficient energy use of sensor nodes is very important. Therefore, one of the most important considerations when designing a routing protocol in a wireless sensor network is to minimize the energy consumption of each sensor node. Many routing protocols using K-means and K-medoids, which are representative machine learning techniques, have been proposed. Accordingly, the performance of the KC, KCE and KOCED protocols and the performance of KCA, PAM and CODS-KM protocols are compared. The results of the performance comparison are as follows. For the K-means protocols, the KC, KCE and KOCED protocols occurred in rounds 346, 374 and 412 in the FND (100 * 100) field. And in the (200 * 200) field, it occurred in rounds 161, 173 and 214. And in the (400 * 400) field, it occurred in rounds 98, 101 and 134. For K-medoids protocols, the KCA, PAM and CODS-KM protocols occurred in rounds 772, 845 and 832 in the FND (100 * 100) field. And in the (200 * 200) field, it occurred in rounds 95, 161 and 194. And in the (400 * 400) field, it occurred in rounds 9, 13 and 7. Accordingly, among the K-means algorithms, the KOCED protocol appears to have the highest efficiency in the wide field. This is the result of (K_opt) optimization and energy consideration in the clustering process. As a future study, we will present items on the evaluation elements of the KOCED protocol and conduct research to increase energy efficiency.

References

[1] M. Inam, Z. Li, and Z. A. Zardari, "A novel improved energy-efficient cluster based routing protocol (IECRP) for wireless sensor networks," *Journal of Information and Communication Convergence Engineering*, vol. 19, no. 2, pp. 67-72, Jun. 2021. DOI: 10.6109/jicce.2021.19.2.67.

[2] D. Y. Yun and D. S. Lee, "Design of the fuzzy-based mobile model for energy efficiency within a wireless sensor network," *Journal of Information and Communication Convergence Engineering*, vol. 19, no. 3, pp. 136-141, Sep. 2021. DOI: 10.6109/jicce.2021.19.3.136.

[3] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, no. 4, pp. 393-422, Mar. 2002. DOI: 10.1016/S1389-1286(01)00302-4.

[4] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292-2330, Aug. 2008. DOI: 10.1016/j.comnet.2008.04.002.

[5] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, New York: NY, USA, pp. 88-97, 2002. DOI: 10.1145/570738.570751.

[6] A. Perrig, J. Stankovic, and D. Wagner, "Security in wireless sensor networks," *Communications of the ACM*, vol. 47, no. 6, pp. 53-57, Jun. 2004. DOI: 10.1145/990680.990707.

[7] P. Sasikumar and S. Khara, "K-means clustering in wireless sensor networks," in *2012 Fourth International Conference on Computational Intelligence and Communication Networks*, Mathura, India, pp. 140-144, 2012. DOI: 10.1109/CICN.2012.136.

[8] A. Sheta and B. Solaiman, "Evolving a hybrid K-means clustering algorithm for wireless sensor network using PSO and gas," *International Journal of Computer Science Issues (IJCSI)*, vol. 12, no. 1, pp. 23-32, Jan. 2015.

[9] H. S. Park and C. H. Jun, "A simple and fast algorithm for K-medoids clustering," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3336-3341, Mar. 2009. DOI: 10.1016/j.eswa.2008.01.039.

[10] J. Wang, K. Wang, J. Niu, and W. Liu, "A K-medoids based clustering algorithm for wireless sensor networks," *in Proceedings of 2018 International Workshop on Advanced Image Technology (IWAIT)*, Chiang Mai, Thailand, pp. 1-4, 2018. DOI: 10.1109/IWAIT.2018.8369769.

[11] S. Jain and N. Bharot, "K medoids based clustering algorithm with minimum spanning tree in wireless sensor network," in *Proceedings of 2019 International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, pp.1771-1776, 2019. DOI: 10.1109/ICCES45898.2019.9002548

SeaYoung Park

He receivced the B.S and the M.S degree and the Ph. D degree in Department of Computer Science from Kwangwoon University, Korea, in 2010, 2016, 2022. In Ph.D, the research topic is to optimize the routing protocol energy consumption of Wireless Sensor Network. His research interests include wireless sensor routing protocols, VR, AR, and artificial intelligence.

Dai Yeol Yun

He received the B.S. and the M.S. degree in Department of Telecommunications Engineering from HanYang University, Korea, in 1997,1999. Ph. D degree in Department of Plasma Bioscience and Display from KwangWoon University, Seoul, South Korea, in 2019. His current research interests include nonlinear system analysis and control, feedback linearization, computer aided control, computer network, image fusion and WSN, Sensor Network.

Chi-Gon Hwang

He was born in Masan, Korea, in 1970. He received the B.S. degree in Business Administration from Changwon National University, Korea, in 1995. the M.S. and Ph.D degree in computer software from Kwangwoon University, Seoul, Korea, in 2004 and 2012. His current research interests include distributed computing, cloud computing, ontology, A.I., NLP(Natural Language Processing).

Daesung Lee

He is a professor in the Department of Computer Engineering, Catholic University of Pusan, Korea. He received the B.S., M.S. and Ph.D. degrees from the Inha University, Korea, in 1999, 2001 and 2008, respectively, all in Electrical Engineering Computer Science & Engineering from Inha University. His research interests include security in network, convergence and operating system.