Clustering Approach in Wireless Sensor Networks Based on K-means: limitations and Recommendations

Ali Abdul-hussian Hassan, WahidahMd Shah, Ali Mohamed Husien, Mohammed Saad Talib,Ali Abdul-Jabbar Mohammed,MohdFairuz Iskandar

Abstract: Clustering approach in wireless sensor network is very important, the structure of cluster and how to improve it is a first challenge that faced the developers, because of it represent as a base for design the cluster-based routing protocol. One of most popular cluster algorithms that utilizing into organize sensor nodes is K-means algorithm. This algorithm has beneficial in construct the clusters for various real-world applications of WSN.K-means algorithm suffering from many drawbacks that hampering his work.The lack of adequate studies that investigates in the limitations of this algorithm and seek to propose the solutions motivated us to do this study. In this paper the limitations of K-means and some suggestions are proposed. These suggestions can improve the performance of K-means, which will be reflected on saving the energy forsensor nodes and consequently maximize the lifetime of the wireless sensor networks.

I. INTRODUCTION

Wireless Sensor Networks WSN sis consisting on a big number of sensors, which has a finite battery power. Due to it is working in the risk and harsh environments, so that impossible or very hard for battery replacing or recharging[1][3].

Consequently, conserve the energy is very significant for this network. Routing protocols has an immense influences on the energy consumption[4][5], where the energy consumption is considering a major factor in routing protocol design.

A cluster base routing protocol is considering the best type of routing protocols in concept of energy saving for sensors and prolong network lifetime. In this protocol, sensor nodes is organize in specific groups called clusters. Each cluster has member nodes called ordinary nodes (ON) and special node called cluster head (CH). The CH has higher energy and utilized to collect and transmit data from ONs to base station

(BS)[6][7][8]. In this approach of the routing protocols, the messages that convey through network can be reduce.

Revised Manuscript Received on December 22, 2018.

Ali Abdul-hussian Hassan Faculty of Information and Communications Technology,UniversitiTeknikal Malaysia **Melaka,Hang Tuah Jaya**, 76100 Durian Tunggal, Melaka, Malaysia

WahidahMd Shah College of Education for Pure Sciences, University of Kerbala, Iraq

Ali Mohamed Husien College of Administration and Economics, University of Babylon, IraqEmail: ali.ah@uokerbala.edu.iq Orcid :0000-0002-0933-9807

Consequently, the network lifetime will be maximize. One of challenges that faced clustering method is how to form the nodes in specific clusters i.e. choosing clustering method faces with problems and considerations that must be considered in order to reduce the energy consumption and maximum lifetime in the network [9][10].Furthermore, do not decided the optimal clusters number in the cluster formation is producing some issues. K-means is considering one of algorithms that has been wide used for organizing the nodes in the clusters at wireless sensor networks. In this paper we investigate in K-means algorithm in terms the limitations of this algorithm and what the suggestion solutions that will be used for overcome these drawbacks. The rest of this study is arranged as follows:

In section 2 clustering process based on K-means algorithm. In section 3 clustering protocols based on k-means method will be studied. Limitations of K-means will illustrated in section 4. And in section 5, various Approaches for calculate optimum number of cluster will proposed. Finally, conclusion illustrated in section 6.

II. CLUSTERING PROCESS BASED ON K-MEANS ALGORITHM

The clustering technique is considering a widespread method which utilize to minimize the energy consumption in WSNs .Clustering procedure are organizing the sensor nodes into specific setsknown as a clusters. Single node in every cluster is selected as a major of cluster known as a cluster head (CH).The CH has many function in addition to sensing the environment such as; data gathering form all cluster member and convey it to BS, convey other CHs data to next hop, make fusion cluster data, and some time control the cluster based on clustering method[11][12][13][14]. The Key advantage of clustering is to minimizing the energy consumption and thereby prolong network lifetime. Figure 1 is illustrating a humble example of a clusters in WSN. Considered the structure of cluster and how to improve it is a first challenge that faced the developers, because of it represent as a base for design the cluster-based routing protocol.

Figure 1 clusters in WSN [15].

$$E(\Gamma,V)=\sum_{i=1}^{k}\sum_{j=1}^{n}\left\|\bar{x}_j-\bar{v}_i\right\|^2$$

k	CH_1	CH_2	.....	CH_k

Then comes the most common challenge of this type of protocol, which is how to select the best node to be cluster head. One of most popular algorithms that used to gathers the nodes in WSNs is K-means (KM) algorithm. This algorithm is very beneficial in construct the clusters for various real-world applications of WSN [16].This algorithm is one of unsupervised clustering methods, which efficiently utilized to form spherical shapes clusters [17].Stuart Lloyd in 1982, was firstly researcher suggested this algorithm[18]. It divided points of data into specific number of clusters[19]. It mostly increase the distances between the clusters along wither duce distance inside the cluster. The goal of this algorithm is seek to find best cluster centric when diminishing the objective function based on a Squared-Error-Function (SEF). We can defined the objective function of KM as:

Where $\|\overline{x_j}\text{-}v_i\|^2$represent TheEuclidean distance that used to determine the distance between npoints of data x_{ji}withits cluster center v_i.

The processes of this algorithm is including the following phases[20]:

Phase 1:Locate the k centroids points in the space which is representing by the data set, where K is predefined number.

Phase 2:Allocateeverypoint of datato the specificcluster, which has nearest centroid distance.

Phase 3:Once all point of data have been clustered, re-determine the locations of the k centroids.

Phase 4:reiterate the Phase 2 and Phase 3tillno shown change inthe location of centroids.

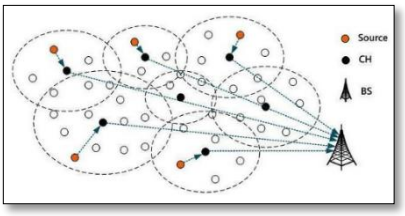
III. CLUSTERING PROTOCOLSBASED ON K-MEANS METHOD

In this section, we introduced some K-means algorithms that used for the nodes gathering in cluster-based routing protocols:

A. KPSO/KGA[21]

In this work, the authors extendedtheir previous workthat existing in[22].Wherein,theyimplement a two-phase hybrid K-Means (KM) with Particular Swarm Optimization (PSO) was called (KPSO),and Genetic Algorithm (GA) clustering algorithmswas called(KGA),for clusteringsensor nodes ofWSNs into clusters and selection of CH for each cluster. They improved their work throughincrease an extrastep for selection the optimum cluster members, which called it KPSO-PSO and KGA-GA.

This study was developed to presents an improved of the clustering algorithm into[22]. The algorithm ofPSO was appliedforWSN in order toorganizethe nodes in the specific clusters. Itpointed to determine the optimumclusters'number,and optimal CHs along with thebeststrategy of the clusters (i.e. the number of members and location of cluster centroid in each cluster). Herein,the algorithm ofPSO has molecules which refer to the clusters



number,then theCH index followed for the each cluster. As showed in the Figure 2.

Figure 2 PSO particle.

So, (k) is mean the number of clusters, as well as CH_i is the indicator of the CH for each cluster "*i*". For that reasons, this study haspromising results were compared to other traditional studies**.**

B. EECPK-means[16]

In this work, Energy -Efficient Clustering Protocol based on K-means midpoint algorithm (EECPK-means) for WSN has been presented. Wherein the midpoint algorithm is utilized for enhance the initial selection of cluster centroid process. Because of the initial cancroids selection was achieved by random way in K-means algorithm, which resulted unbalanced in the clusters, the presented work resulted a balanced clusters and consequently load-balance for CHS has achieved, and maximize the network lifetime. Moreover, the CH selection procedure will be optimization through depended on residual energy as a one of parameter in selection method as well as Euclidean distance that already utilized in KM. It determines the optimal number of required clusters based mathematical format, which includes the sensing area size along with the sensor number. Assume that N is meaning the sensors number which regularly distributed in the square area M×M. The optimal clusters number (kopt) can be determined as[23]:

$$k_{\text{opt}}=\frac{\sqrt{N}}{\sqrt{2\pi}}\sqrt{\frac{\varepsilon_{\text{fs}}}{\varepsilon_{\text{mp}}}}\frac{M}{d_{\text{BS}}^2}$$

Here d_{BS} is indicator to the distance between's and CHs, $\varepsilon_{fs}$is indicator toffee space model as well as ε_{mp} is indicator to multipath model.

$$DB=\sum_{i=1}^{c}\max\frac{S_c(Q_k)+S_c(Q_l)}{d_k(Q_k,Q_l)}$$

$$S_c=\frac{\sum_i\|x_i-C_k\|}{N_k}$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

C. MRRCE[24]

The authors in this work proposed a Multi- Hop Routing Energy Efficient Scheme (MRRCE)which depended on parameters such as distance residual and energy to getcluster heads and optimum clusters for energy efficiency objective, and accordingly the network lifetime is prolonging. In this work, they utilized the Steiner Points(SPs)concept[25]. Where after the network is gridding, draw a sequence of vertical and horizontal lines from each location node coordinates. It causes generating a lot of crossings among all sensors.Ininstance, each 2-sensors which are not placed in the equal direction, they have four joint crossings points, while the two of them are their coordinate's locations[26]. The SPs are utilized as an alternative way for depending on random selection of initial cluster centroid step in K-means. Due to, SPs are the Joints all the nodes in the cluster, CHs find in best location for initial CH while in the traditional K-means algorithm, random nodes were utilized as initial CH, that mean these location could be in any place, even this place do not has any node around. Concluding, K-means algorithm has big dependency on the initial selection of CH, by choosing the correct location as initial CH, the number of iteration will be decrease.

D. K means-Davies Bouldin index[27]

Autho $c_p = (1/|C_p|)\sum_{h;x_h \in C_p} x_h$ an improved K means-Davies Bouldin index in order to generate a balanced in energy consumption on the clusters. In addition to utilized Gaussian elimination for select the best node as a CH which caused the distribution in the energy consumption. To solve the problem of determine the optimum number of clusters; they utilized the Davies Bould in Index (DBI) for this objective, in order to maximize the network lifetime. The DBI address each class separately and strive for determine how similar it is to the class which is nearest to it. It assess the intracluster objects similarity and intercluster dissimilarity. As a result, it generates a well clustering. The DBI is formed as:

Where $S_c(Q)$ represent the average distance between a cluster nodes Q_k and its center.

Where N_k: the number of nodes in the cluster k, and d_{ce}: the distance between the clusters centers Q_k and Q_l. where:

$$d_{ce} = \|C_k - C_l\|$$

Then the centroid of each cluster is selecting as a first CH based on K-means. Next CH will be selection trough the Gaussian elimination algorithm based on residual energy by BS.

E. EBRP[28]

In this protocol, authors' proposed an Energy Balanced Routing Protocol (EBRP) for wireless sensor networks. In this protocol, theydistribute the sensor network into various clusters based on K-means++ algorithm:
Random selection for node in the set X as cluster center c_1. The distances is calculated between the new center of cluster and other nodes by BS. Anew node x_m is selected as a new center of cluster c_p, while the probability of selection.

$$d^2(x_m, c_p)/\sum_{j=1}^{n} d^2(x_j, c_p)$$

Where p ∈ 1… k. d (xm,cj) is the Euclidean distance between xm and cj, and the p is the cluster number. In the second stage, reiterated this stage till all the centers of clusters K are determined. In 3th stage, the distance between every node to cluster center is determined, and the node is calculated to the adjacent center. It is indicated by xi ∈Cp, Cprepresent the nodes set cluster p. While in the 4th stage, a new center for every cluster is represented by:

Where p ∈ {1,…, k}.Lastly, reiterated stages3 and this step till getting the number iterations setting. The K clusters are found by the algorithm through BS. After that, BS is doing the broadcast message contain the coordinate of center and the cluster number for all clusters.

F. GAK-means[29]

Authors in this work presented anovelmixture of K-means and improved Genetic Algorithm (GA) in order to decrease energy consumption and prolong network lifetime. GA-K means, seek to minimize the energy consumption through determining the optimal number of cluster heads (CHs)by utilized improved of **Genetic Algorithm** (GA), where the number of CHs also point to the clusters number in the network. Due to GA have some issues like it is very slowly in the converging for huge data, hung in local optimum solution that mean not global optimum. Consequently, GA is employed along with K-Means algorithm. So in this protocol, GA with a few number of iterations is employed to realize the optimal solution.
The main point is the set of points that found by utilizing GA is employed as **initial points** for the K-Means algorithm. After that every datum according to its symmetry is assigned to single cluster. In case, when depend on the K-means only in order to form the

clusters, it cannot find the optimum solution, because of K-means has sensitivity for data center, and failure to identify for the noise data and failure to discover non spherical cluster and last lyitmay hung in local optimum solution.

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

Authors	Year	Protocol name	Cluster formation			CH selection	
			Clusters number	Initial selection point	Cluster formation method/parameters	Method	parameters
Sheta, Alaa F	2015	KPSO/KGA	PSO	×	K-means/Euclidean distance	PSO/GA	Residual energy and Euclidean distance
Ray, Anindita De, Debashis	2016	EECPK-means	mathematical format	midpoint algorithm	K-means/Euclidean distance	K-means	Unique nodes ID, refers tothe Euclidian distance between nodes and its own cluster centroid
Rezaei, Elham Baradaran, Amir Abbas Heydariyan, Atefeh	2016	MRRCE	×	Steiner Points	K-means/Euclidean distance	K-means	Euclidean distance
Elkamel, RabiaaCherif, Adnane	2017	K means-Davies Bouldinindex	Davies Bouldin Index	×	K-means/Euclidean distance	Gaussian elimination	residual energy
Li, LinLi, Donghui	2018	EBRP	×	×	K-means++/Euclidean distance	FLS-GA	Distance between node to BS d_{BS}, center d_{center} and Energy E_r
PrabhuThiyagarajan	2016	GA-Kmeans	GA	GA	K-means/Euclidean distance	Improved GA	Residual energy

IV.LIMITATIONS OF K-MEANS

- The initial cancroids are selected by random way for the input data set. Ate very iteration, the K-means construct different clusters depending on the different random selected for initial cancroids. As a result the initial cancroids in random selection cause the local optima.
- No guarantee for K-means will converge into optimal or better solution. Due to it run in many iteration, and no any analysis that doing to find the better result, consequently sometime produce unbalanced cluster size.
- Cluster number is not determined accurately and automatically, accordingly it needed to be set cluster number according to user input.
- K-means has very high complexity oftime, where the points of data must re-assigned to number of iterations times for every loop run.
- In some rare cases, K-means may produce an empty cluster, because of the random initial centroids selection.

As mention above, K-means algorithm suffering from many drawbacks that hampering his work. Among these problems, determine the clusters number which is considered the most influential problem on performance of K-means in WSNs. Because of the nature of the CH activities, it expend extra energy to perform these various functions[30]. So, balancing the energy consumption of CHs is very significant problematic to long term for functioning of WSNs.Consequently, in the next section we present with some methods that will determine the value of K.

V.VARIOUS APPROACHES FOR CALCULATE OPTIMUM NUMBER OF CLUSTER

There are many methods in literature that used to selecting cluster number[20].Authors focused on the specific approaches, where they are characterized as a fast convergence, simplicity, and it can performs well in large scale of WSNs, they are:

1. By rule of thumb
2. Information Criterion Approach
3. Choosing k Using the Silhouette
4. Elbow method
5. Cross-validation
6. An Information Theoretic Approach

By rule of thumb

It is very simple method[31]. This method can by apply to any type of data.

$K \cong \sqrt{N/2}$

Where *N*is the number of sensor nodes (data points).

Information Criterion Approach

The clusters number in the combination paradigm increase results in a raise in the dimensionality of the paradigm, producing a drab raise in its likelihood.If user was need to finding the max likelihood model with any value of clusters number, he would finally end up by getting result that each data point is the unique member in the cluster.

Clearly, he wishes to avoid this problem, where he must select some criteria which do not based likelihood as a singular factor.The Information Criteria Parameter is utilized for choosing amongst models along with different parameters number. It tries to balance the increasing in the likelihood because of extra parameters through setting a threshold value for each parameter. This process of selection consist from 2-stages of clustering phase, wherein the cluster number can selected as automatic choosing by user, which depended on any of information criteria. The techniques of model selection are depend to calculate the clusters number using mixture models[32].

Traditionally, selection of model is executed in two-phases. Firstly, the candidate set of models produce through some principles of learning (execute by max. likelihood learning (ML)) for a models scope.

Secondary, the appropriate model will be selected which depend on criterion of model selection. The Minimum Description Length (MDL) criterion[33], the Consistent Akaike'sInformation criterion (CAIC)[34],and Akaike's Information criterion (AIC)[35], are considering as famous examples for model selection criteria, that actually synchronize with the Bayesian inference criterion (BIC)[36],

AIC and BIC are defined as

- BIC= -2Ln(likelihood) + k Ln(N)
- AIC= -2Ln (likelihood) + 2k

Where K is representing the degrees of model for freedom calculated as the rank of Variance Covariance Matrix of the parameter e(V) and N is the number of objects thatutilized in the estimation, for more specifically, it represent the number of the likelihood independent terms. Practically, N is defined as e(N).

Choosing k Using the Silhouette

A various approaches are using the Indexes that making comparison between intra-cluster distances and enter-cluster distances: when the distance is greater that mean the well result getting. Several of them are stated in [11]. The most popular indexes are: Milligan and Cooper [37].

A. The first one was called a Correlation of Point bacterial. Which is, the factor of correlation among the enter-to-enter matrix distance and the matrix of binary partition set every pair of the objects which locates into same cluster to1, otherwise sets to 0.
B. The second one is ordinal version which presented in[38].
C. A good method for get balanced clusters. In the experiments, the width of silhouette has appeared good performance, was presented in the[39]. The silhouette width concept is involving the variance between the tightness in intra-cluster and secession from the others. More precisely, the silhouette width is calculate as follow:

$$s(i)=\frac{b(i)-a(i)}{\max(a(i),b(i))}$$

The *a(i)* represent the average distance between object (*i)* and other cluster objects, *b(i)* represent the min. average distances between object (*i)* and all the objects in the all other clusters. Range values of Silhouette Width *s(i)*is between -1 to 1. If all the silhouette width values are close to 1, it means that the set I is well clustered.

If S(i) ≈ 1, that mean is a good clustered, and vice versa. But when the S(i) =0, that mean this object can be re-assign to any other side. The clusters can be categorized by the Average Silhouette Width (ASW) for singular objects. The max of ASW for various clusters number is considering the correct number of clusters.

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

Elbow Method

The earlier way for estimation the right number of clusters in a data set (N)[40]. It is a visual method. The idea behind this method is that initiate the cluster number equal 2, then

$$d(i, c_k)=(y_i-c_k)^T\Gamma_k^{-1}(y_i-c_k)$$

increase the cluster number one by one until reach to number less than 20% from N. In each step, the cluster cost must be calculating. At certain value of K, the cluster cost change will significantly, and then curve of cluster cost goes down gradually after that. This inflection point is representing the optimum K value. Disadvantage of elbow method is: This "elbow point" cannot at all times be clearly identified. Occasionally, there is no elbow point, or more than one elbows point as shown in Fig. 2

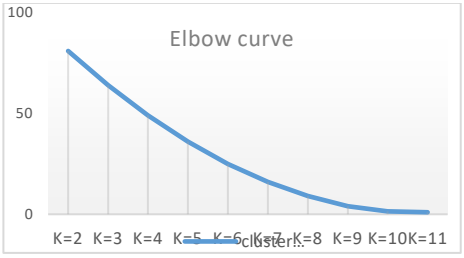


Figure 2 Elbow curve (no elbow point).

Cross-validation

Another method for calculating the clusters number is Cross Validation (CV), this method presented by[41]. It is depend on stability of cluster. This approach divided the objects for two parts or more. One them is utilized for clustering the data and for validation can utilized the other(s)part. The major concept is tending to frequently configure a similar clusters of data that generating from similar source. Where, this algorithm will be stable for randomization input. But when the data size be bigger and especially with the greater correlation, this method not has accurate to determine the suitable clusters number. Wang [42] was improve a cross validation by introduced some new parameters in order to calculating the clusters number. This novel election standard in order to determine the quality of clusters via utilized the instability among the samples. Where the instability for cluster process is evaluated via CV, in order to minimalist the instability. All data is segmented for three sets; two of them is training-sets and the rest set is for validation in order to be similar to original concept of stability. After that, depended on the algorithm that used for clustering, a distance is measured for the independent training sets and use the validation set for evaluated the inconsistencies. It has been confirmed to be dynamic and powerful on a several examples.

An Information Theoretic Approach

An anther method for estimating the cluster numbers, that it sets limited to the parametric assumptions, may be strictly theory propose utilizing concepts from the rate distortion theory, it is easy to understanding and computing, and it has high effective for problems in the various range. It depend on "distortion" method that which calculates the dispersion in each cluster.

Sugar and James make improvement in the statistic by using jump-statistic, which they use the factor W in equation below to extend depending on Gaussian distribution method[43].

Precisely, the distance between an entity and centroid in equation below is calculated as:

The symbol Γk represent the covariance matrix of cluster, and the jump-statistic is calculated as:

JS (K) = WK -M/2 - WK-1-M/2, where W0-M/2 ≡ 0

The max.of JS (K) is corresponding to the correct clusters number. This is reinforced by a derivation of mathematical indicating that if the data can be regarded a standard sample from the Gaussian distributions mixture, where the distances among cancroids are almost sufficiently great. Thereafter, the max-jump would really happen at K equivalent the Gaussian components number in the mixture.

IV.CONCLUSION

In this work, we examine one of the most clustering algorithms that used in the WSNs, which is known as K-means, and its controversial issues. The problem in estimation the optimum clusters number has a great influences on balanced clusters in the network. This problem caused unbalanced energy consumption in network, consequently reduce the network lifetime. Occasionally, researchers are usingmathematical formula and heuristic algorithms along with K-means to solve this problem and find optimal clusters number. In mathematical formula is not appropriate for all distributed methods, especially the random distributed. Also the heuristic algorithms is not suitable because it has a slow convergence, complex calculations and complex time processing. In order to overcome these problems, we proposed aspecific methods can be utilize with K-means algorithm to determine the appropriate clusters number.

ACKNOWLEDGMENT

This work is fully supported by Universiti Teknikal Malaysia Melaka-Zamalah Scheme. So, the authors would like to thank Zamalah Scheme for providing the facilities and financial support for this research.

REFERENCES

1. M. Hammoudeh and R. Newman, "Adaptive routing in wireless sensor networks: QoS optimisation for enhanced application performance," *Inf. Fusion*, vol. 22, pp. 3–15, 2015.
2. M. S. Talib, A. Hassan, B. Hussin, Z. S. Talib, Z. S. Rasoul, and M. Sammour, "Data dissemination based clustering techniques for VANETs: A review," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 10, no. 4 Special Issue. pp. 596–604, 2018.
3. A. Abdul-Jabbar Mohammed, M. A. Burhanuddin, and H. Basiron, "Key Enablers of IoT Strategies in the Context of Smart City Innovation," *Jour Adv Res. Dyn. Control Syst.*, vol. 10, no. 04, pp. 582–589, 2018.
4. A. A.-H. Hassan, W. M. Shah, M. F. Iskandar, and A. A.-J. Mohammed, "Clustering methods for cluster-based routing protocols in wireless sensor networks: Comparative study," *Int. J. Appl. Eng. Res.*, vol. 12, no. 21, 2017.
5. M. Sabet and H. R. Naji, "A decentralized energy efficient hierarchical cluster-based routing algorithm for wireless sensor networks," *AEU - Int. J. Electron. Commun.*, vol. 69, no. 5, pp. 790–799, 2015.
6. K. M. Awan, Ahmad Ali, Farhan Aadil, and Kashif Naseer Quresh, "Energy efficient cluster based routing protocol (EECBRP) for Wireless Sensor Network," in *International Conference on Advancements in Computational Sciences, ICACS*, 2018, pp. 1–6.
7. M. Faeq Ali, N. Haryati Harum, and N. Azman Abu, "An Intelligent Radio Access Technology Selection For Vehicular Communications," 2017.
8. M. N. Al-Mhiqani, R. Ahmad, K. H. Abdulkareem, and N. S. Ali, "Investigation study of Cyber-Physical Systems: Characteristics, application domains, and security challenges," *ARPN J. Eng. Appl. Sci.*, vol. 12, no. 22, pp. 6557–6567, 2017.
9. P. Neamatollahi, H. Taheri, M. Naghibzadeh, and S. Abrishami, "A distributed clustering scheme for wireless sensor networks," in *2014 6th Conference on Information and Knowledge Technology, IKT 2014*, 2014, pp. 20–24.
10. S. Dehghani and B. Barekatain, "An Enhanced Energy-Aware Cluster-Based Routing Algorithm in Wireless Sensor Networks," *Wirel. Pers. Commun.*, vol. 98, no. 1, pp. 1605–1635, 2018.
11. A. A. H. Hassan, W. M. Shah, M. F. Iskandar, M. S. Talib, and A. Abdul-Jabbar Mohammed, "K nearest neighbor joins and mapreduce process enforcement for the cluster of data sets in bigdata," *J. Adv. Res. Dyn. Control Syst.*, 2018.
12. S. T. Hasson and Z. Y. Hasan, "Roads clustering approach's in VANET models," in *2017 Annual Conference on New Trends in Information and Communications Technology Applications, NTICT 2017*, 2017, no. March, pp. 316–321.
13. A. Abdul-Jabbar Mohammed, M. A. Burhanuddin, A. Alkhazraji, and H. Basiron, "IoT Devices and Sensors Management Framework for Mobile E-health Applications," *J. Adv. Res. Dyn. Control Syst.*, vol. 10, no. 02–Special Issue, pp. 2157–2161, 2018.
14. H. A. Khudhair and S. T. Hasson, "A Monte Carlo approach to estimate the coverage overlapping areas in WSNs," *J. Eng. Appl. Sci.*, vol. 13, no. 3, pp. 624–628, 2018.
15. M. Yang, C. L., Tarng, W., Hsieh, K. R., & Chen, "A security mechanism for clustered wireless sensor networks based on elliptic curve cryptography," *IEEE Int. Conf. Syst. man, Cybern. (IEEE SMC)*.
16. A. Ray and D. De, "Energy efficient clustering protocol based on K-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network," *IET Wirel. Sens. Syst.*, vol. 6, no. 6, pp. 181–191, 2016.
17. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithms: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, 2002.
18. S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982.
19. G. A. Wilkin and H. Xiuzhen, "K-means clustering algorithms: Implementation and comparison," in *Proceedings - 2nd International Multi-Symposiums on Computer and Computational Sciences, IMSCCS'07*, 2007, pp. 133–136.
20. T. M. Kodinariya, "Review on determining number of Cluster in K-Means Clustering," vol. 1, no. 6, pp. 90–95, 2013.
21. A. F. Sheta, "Evolving Clustering Algorithms for Wireless Sensor Networks with Various Radiation Patterns to Reduce Energy Consumption," 2015.
22. A. Sheta and B. Solaiman, "Evolving a Hybrid K-Means Clustering Algorithm for Wireless Sensor Network Using PSO and GAs," *Int. J. Comput. Sci. Issues*, vol. 12, no. 1, pp. 23–32, 2015.
23. T. Mu and M. Tang, "LEACH-B: An improved LEACH protocol for wireless sensor network," in *2010 6th International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2010*, 2010.
24. E. Rezaei, A. A. Baradaran, and A. Heydariyan, "Multi-hop Routing Algorithm Using Steiner Points for Reducing Energy Consumption in Wireless Sensor Networks," *Wirel. Pers. Commun.*, vol. 86, no. 3, pp. 1557–1570, 2016.
25. H. Zhou, "Efficient Steiner tree construction based on spanning graphs," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 23, no. 5, pp. 704–710, 2004.
26. A. Kashaf, N. Javaid, Z. A. Khan, and I. A. Khan, "TSEP: Threshold-sensitive stable election protocol for WSNs," in *Proceedings - 10th International Conference on Frontiers of Information Technology, FIT 2012*, 2012, pp. 164–168.
27. R. ELkamel and A. Cherif, "Energy-efficient routing protocol to improve energy consumption in wireless sensors networks," *Int. J. Commun. Syst.*, vol. 30, no. 17, p. e3360, 2017.
28. L. Li and D. Li, "An Energy-Balanced Routing Protocol for a Wireless Sensor Network," *J. Sensors*, vol. 2018, pp. 1–12, 2018.
29. B. Barekatain, S. Dehghani, and M. Pourzaferani, "An Energy-Aware Routing Protocol for Wireless Sensor Networks Based on New Combination of Genetic Algorithm & k-means," *Procedia Comput. Sci.*, vol. 72, pp. 552–560, 2015.
30. A. A. H. Hassan, W. M. Shah, M. F. Iskandar, M. N. Al-Mhiqani, and Z. K. Naseer, "Unequal clustering routing algorithms in wireless sensor networks: A comparative study," *J. Adv. Res. Dyn. Control Syst.*, vol. 10, no. 2 Special Issue, pp. 2142–2156, 2018.
31. S. A. Leeper, "Wet cooling towers: rule-of-thumb design and simulation," *No. EGG-GTH-5775. EG G Idaho, Inc., Idaho Falls, ID (United States)*.
32. H. Bozdogan, "Mixture-model cluster analysis using model selection criteria and a new informational measure of complexity," *Proc. first US/Japan Conf. Front. Stat. Model. An informational approach. Springer, Dordrecht*, pp. 69–113, 1994.
33. A. Barron, J. Rissanen, and B. Yu, "The Minimum Description Length Principle in Coding and Modeling," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2743–2760, 1998.
34. H. Bozdogan, "Model selection and Akaike's Information Criterion (AIC): The general theory and its

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.

35. H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, 1974.
36. G. Schwarz, "Estimating the Dimension of a Model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.
37. G. W. Milligan and M. C. Cooper, "An examination of procedures for determining the number of clusters in a data set," *Psychometrika*, vol. 50, no. 2, pp. 159–179, 1985.
38. L. J. Hubert and J. R. Levin, "A general statistical framework for assessing categorical clustering in free recall," *Psychol. Bull.*, vol. 83, no. 6, pp. 1072–1080, 1976.
39. K. Pollard and M. van der Laan, "A method to identify significant clusters in gene expression data," *Proc. SCI*, vol. II, pp. 318–325, 2002.
40. P. Harrington, "The k-means clustering algorithm," in *Machine Learning in Action*, 2012.
41. P. Smyth, "Clustering using Monte Carlo cross-validation," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 126–133.
42. Y. Fang and J. Wang, "Selection of the number of clusters via the bootstrap method," *Comput. Stat. Data Anal.*, vol. 56, no. 3, pp. 468–477, 2012.
43. C. A. Sugar and G. M. James, "Finding the Number of Clusters in a Dataset: An Information-Theoretic Approach," *J. Am. Stat. Assoc.*, vol. 98, no. 463, pp. 750–763, 2003.