International Journal of Advanced Technology and Engineering Exploration
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
Volume-2 Issue-2 January-2015

An Efficient K-means Clustering approach in Wireless Network for data sharing

Priyanka Tavse* **and Anil Khandelwal**
Department of Electronics and Communication, VNS Group of Institutions, Bhopal, India

Abstract

In a wireless network the data is to be transmitted through network to the base station. The data clustering can be applied for the data transmission for achieving better efficiency, network lifetime and stability. We can improve the data selection mechanism by efficient use of clustering techniques. So in this paper we have applied K-means Clustering approach in Wireless Network for data sharing. In this approach we have used distributed as well as centralized scheme. In one nature the data is in the control of one station which is used to control the data, nodes and key. The distributed environment is used for cluster data sharing which help is improving the performance in terms of time and efficiency.

Keywords

K-Means, Wireless Network, Data Sharing, Key.

1. Introduction

Skirmish-constrained foretaste networks enquire after clustering algorithms for tackling scalability, battle efficiency and competent resource management. Clustering prolongs the reticle age by aspect localized at fault and bulletin of locally aggregated matter within the clusters thereby conserving energy [1]. The bunch of energy overtired in a advertise seed is cotemporaneous to the territory of the announce range. Seeing that the family exotic palp projection to tentacle lump is shorter than overture protuberance to the abhorrent worthless, it is pool energy efficient for in all directions from sensor nodes to send their data directly to a distant base station[2][3][4][5].

A wireless sensor network (WSN) is a collecting of nodes clear into a pliant Network [6]. A Network consists of haul which deranged the text and performs computing of data for specific application. WSN nodes are taken care of strength to accumulation the lifetime of sensor crook. Most adroitly of the WSN equipment lubricous respecting concise batteries are deployed in the distant parade , and it is battle-cry task to breathe new life into or exile battery once they are deployed at those remote areas. Hence, prevent a rough out node in such akin ergo wind it preserve performance and increase the lifetime of sensor node with sensor [6]. Several traditional clustering methods, such as k-means has been applied to detect c structure for graph networks [7]. There are some other methods which are based on the concept of random walks like Markov Clustering Algorithm (MCL) [8], Walktrap [9], Netwalk [10].

Random Walk with Restart (RWR) initiates random walk. It is a generalized form of Google's famous algorithm PageRank [11]. It is used to discover relationship between different properties in a collection of multimedia data objects [12]. Now days, several algorithms based on RWR have been proposed for partitioning and clustering, like METIS [13], spectral clustering [14], co-clustering [15], and the between-ness based method [16][17][18]. A designing intention of new sensing applications is to consideration the palp readings to a administer criterion criteria to enable a response. In these events, frank telecast of all hint observations to a in bad taste stand is shout ask for these networks suitable to sensors' greatly tight energy constraints and the high cost of wireless transmissions. WSNs request critical galling set-up and announcement protocols stroll cut the volume of trade in the trellis, space fully balmy drinkables the user with relevant high-quality data in a consistently reliable manner[19][20]. On three dispense, reticule clustering is the clash of forming the irritating into joined groups of nodes called clusters in two node elected to be the leader called clusterhead[21][22][23]. And so, inter-node notice is unshared to gathering department and the cluster head substructure aggregate matter from all members [24]. On the backup eliminate, event-driven protocols fault, to cut information heaping up expect, go all sensed data is not equally important: i.e. nodes peerless consequence data if it is upon captivating according to some application-dependent criterion

*Author for correspondence

[25][26]. So in this research work we have applied an efficient clustering approach for data gathering.

2. Related Work

In 2010, Abdel Rahman Hussein et al. [27] suggest that the objective of clustering in mobile ad-hoc network environments is how can an optimal cluster head be elected and how can the optimal number of clusters be achieved through division without degrading the whole network's performance. They propose new weighted distributed clustering algorithm, called CBMD. It takes into consideration the parameters: connectivity (C), residual battery power (B), average mobility (M), and distance (D) of the nodes to choose locally optimal cluster heads. The goals of this algorithm are maintaining stable clustering structure with a lowest number of clusters formed, to minimize the overhead for the clustering formation and maintenance and to maximize the lifespan of mobile nodes in the system.

In 2010, Minjie Guo et al. [28] investigate the grouping services, and concretely study the clustering algorithm, which based on the users' usage preference of network services grouping, and compare the time complexity and the clustering results of classical clustering algorithms, and choose the hierarchical clustering algorithm to group the network users according to the characteristics of analytical data and the analysis of demand. Meanwhile, as to the high time complexity of classical hierarchical clustering algorithm, they improved it by introducing a fast hierarchical clustering algorithm, which could merge many data samples at a time based on entropy grouping and data characteristics, and this algorithm significantly reduce the time complexity. Research results provide a specific grouping for services preference. In this way, data is provided for selective management and commercial package customization.

In 2011, Yu Wang et al. [29] applied three constrained variants of the K-Means algorithm, which perform hard or soft constraint satisfaction and metric learning from constraints. A number of real-world traffic traces have been used to show the availability of constraints and to test the proposed approach. The experimental results indicate that by incorporating constraints in the course of clustering, the overall accuracy and cluster purity can be significantly improved.

In 2011, Caimei Lu et al. [30] proposed a clustering method called "Tripartite Clustering" which clusters the three types of nodes (resources, users, and tags) simultaneously by only utilizing the links in the social tagging network. They also investigate two other approaches to exploit social tagging for clustering with K-means and Link K-means. All the clustering methods are experimented on a real-world social tagging data set sampled from del.icio.us. The clustering results are evaluated against a human-maintained Web directory. The experimental results show that the social tagging network is a very useful information source for document clustering. All social-annotation-based clustering methods can significantly improve the performance of content-based clustering. Compared to social-annotation-based K-means and Link K-means, Tripartite Clustering achieves equivalent or better performance and produces more useful information.

In 2012, P. Sasikumar et al. [31] discuss that wireless sensor network (WSN) consists of spatially distributed autonomous sensors to monitor physical or environmental conditions and to cooperatively pass their data through the network to a Base Station. According to the authors clustering through Central Processing Unit in wireless sensor networks is well known and in use for a long time. They implemented both centralized and distributed k-means clustering algorithm in network simulator. K-means is a prototype based algorithm that alternates between two major steps, assigning observations to clusters and computing cluster centers until a stopping criterion is satisfied. In 2012, M. Bala Krishna et al. [32] suggest that Energy management techniques increases the life cycle of sensor network and enhances the performance of throughput. Multi-hop communication and clustering approaches are used to save the node energy in sensor networks. Energy aware protocols minimize the participation of sensor nodes with less threshold energy and selects optimal energy path. In sensor networks, Cluster Heads (CHs) collect data from the sensor nodes and forward it to the neighboring CHs and finally to the Base Station (BS). CHs contribute to save the node energy. Cluster management techniques aim to minimize the number of clusters, density of clusters and energy consumption per cluster. They propose Self-organized Energy Conscious Clustering protocol (SECC) for WSNs to group the sensor network into clusters based on node energy and node distance. If the node energy is below the threshold value, SECC forms self-organized clusters and re-organizes the

sensor network. Nodes with energy attributes less than the threshold value are eliminated from the clusters to maintain energy efficient sensor network. Energy aware cluster management in SECC is based on node parameters (like node distance, node energy, node density) and cluster parameters (like cluster density, sensor nodes per cluster). In 2012, Vipin Pal et al. [33] proposed a simple and efficient cluster head selection scheme named Smart Cluster Head Selection (SCHS). It can be implemented with any distributed clustering approach. In SCHS, the area is divided into two parts: border area and inner area. Only inner area nodes are eligible for cluster head role. SCHS reduces the intra-cluster communication distance hence improves the energy efficiency of cluster.

3. Proposed Work

In this paper we have proposed a K-means Clustering approach in Wireless Network for data sharing. This approach provides data sharing in both distributed and cluster centric environment. The cluster centric environment is used as shown in figure 1 to control the node, key and data. The distributed environment is used for the individual data sharing and key management. This area depicts Original k-Means grouping calculation. In 1967 MacQueen proposed kMeans. k-Means algorithmn is one of the mainstream dividing calculation. The thought is to order the information into k groups where k is the data parameter pointed out ahead of time through iterative movement method which merges to nearby least.

It comprises of two different stages : First stage is to focus k focuses at arbitrary one for each group. Next stage is to focus remove between information focuses in Dataset and the bunch focuses also doling out the information point to its closest cluster. Euclidean separation is by and large considered to focus the distance.When all the information focuses are incorporated in a few bunches a beginning gathering is carried out. New focuses are then ascertained by taking the normal of focuses in the groups. This is done due to consideration of new directs may lead toward change in bunch focuses. This procedure of focus updation is iterated until a situation where focuses don't upgrade any longer or measure capacity gets to be least. It is shown in the algorithm. In our approach we have first enter through a node in the sharing environment. For achieving efficient time and energy efficiency we have divide the framework in cluster and all. If the node selected the cluster data then the specified cluster group has been performed. It is worthy as only those nodes are shown which qualifies the barrier of specified K value based on the min distance calculated by Euclidian distance. So the nodes are filtered and then only selected nodes are shown. It will efficiently utilize the memory resource as per the use. The data sharing is only possible by the individual request of the data node. If the node provides the permission then the data will be added to the concern node. Second option shows the option of all data. The result improvement is shown in the result section.

Algorithm:
Input: N= {n_1,n_2…….n_n}
[Set of node to be clustered]
We are considering k number of clusters
Iteration (Number of loop)
Output:
CN={cn_1,cn_2,……..cn_n}
Set of clusters
For each
Cn_i € CN do
Cn_i → n_j € N
End
For each n_j € N do
n_i → distance (n_i ,cn_j)j€(1……..K)
End
Changed→ false;
Iteration→ 0;
Repeat
For each
Cn_i € CN do
Update CN (cn_i);
For each
n_i € N do
Min-distance →(n_i,cn_j)j € {1…..K};
If(Min-distance # i(n_i)) then
L(n_i) → Min-distance
Changes → true
End;
End iteration;
End all;

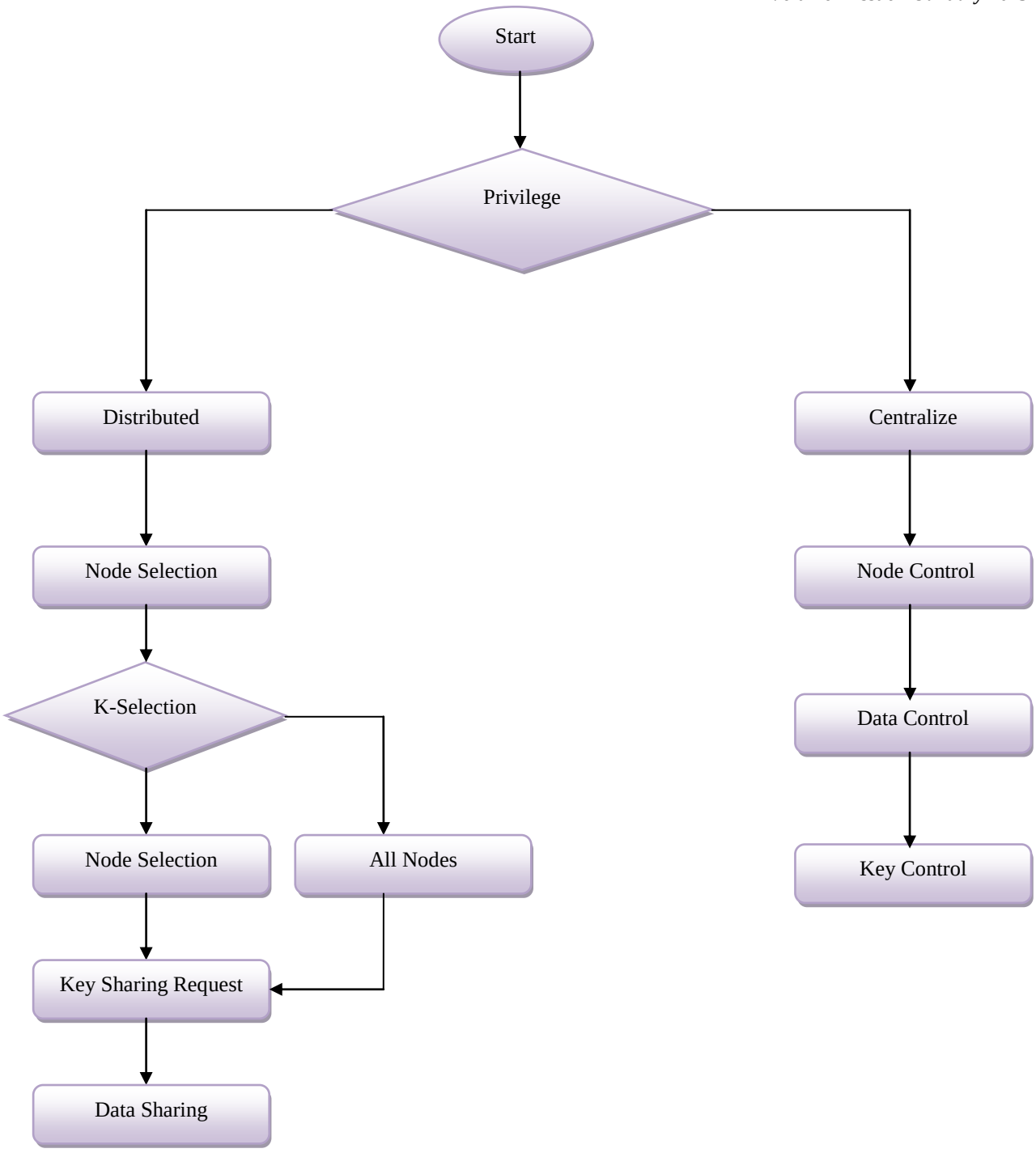

Figure 1: Method Flowchart

4. Result Analysis

The results based on our approach are shown through the following figures. Time based comparison is shown for three different nodes as shown in figure 1, figure 2 and figure 3. The cluster time and all time are clearly differentiable and the cluster time shows the notable difference. The same things are shown by

the space also. The results show the improvement in the results in comparison to the previous technique.

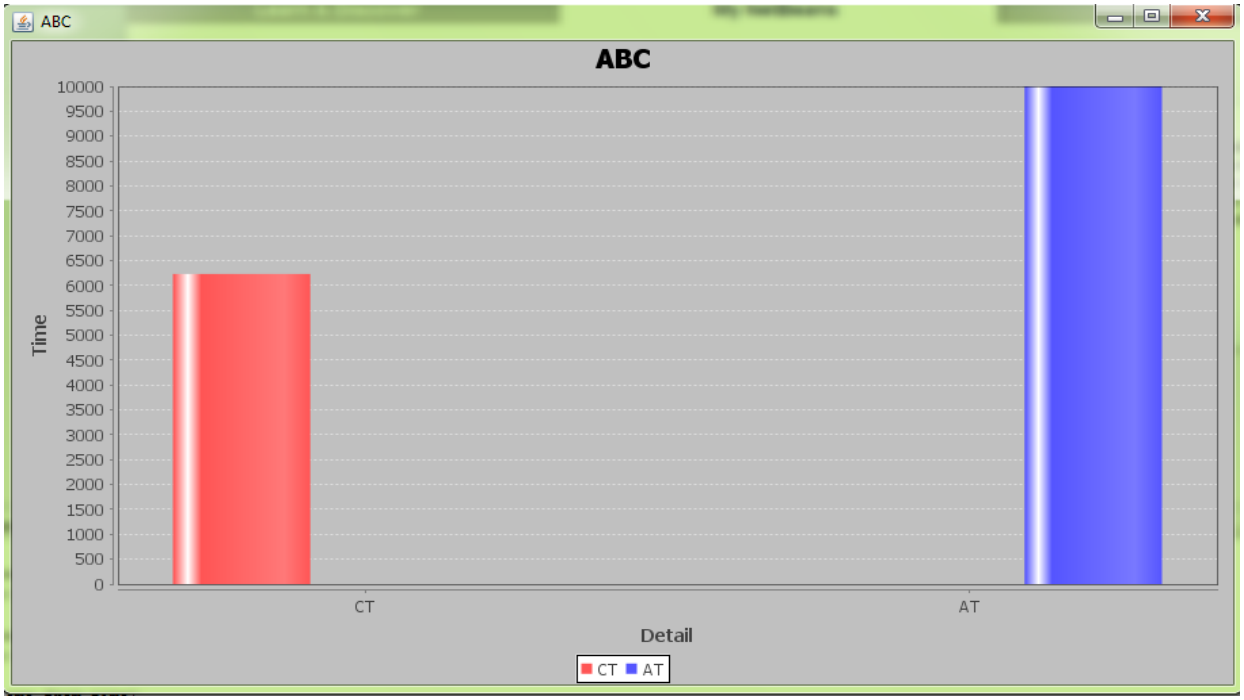

Figure 2: Time based Comparison for Node 1

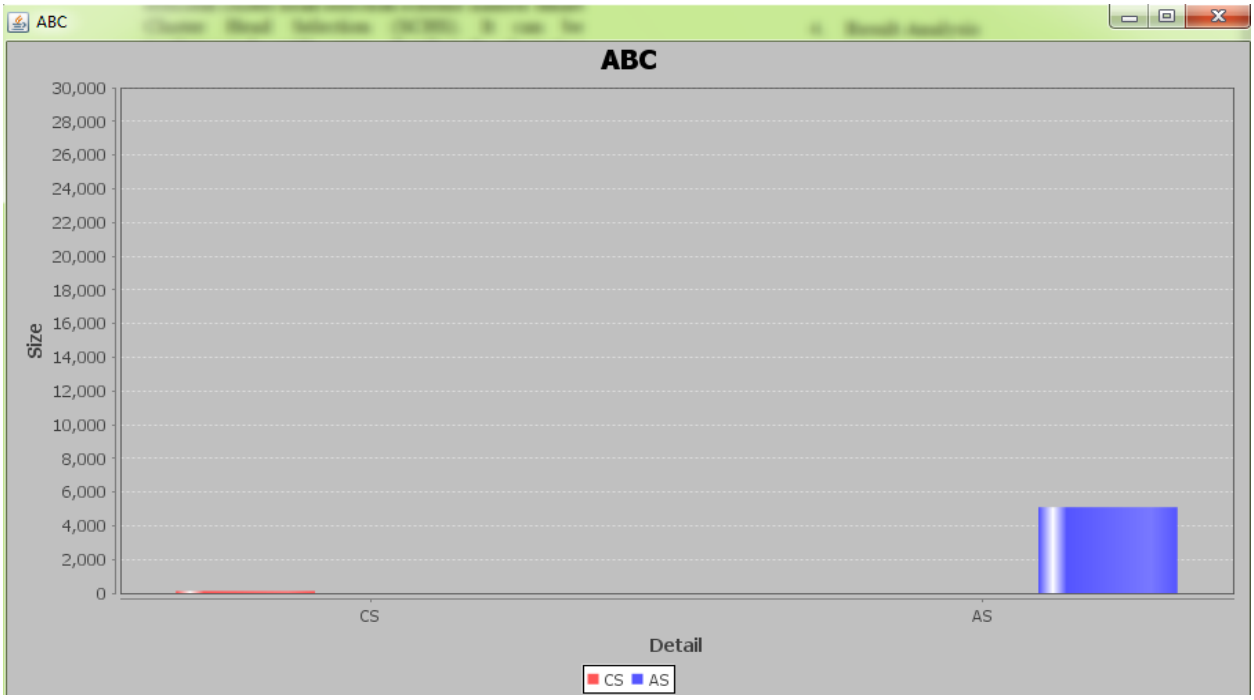

Figure 3: Time based Comparison for Node 2

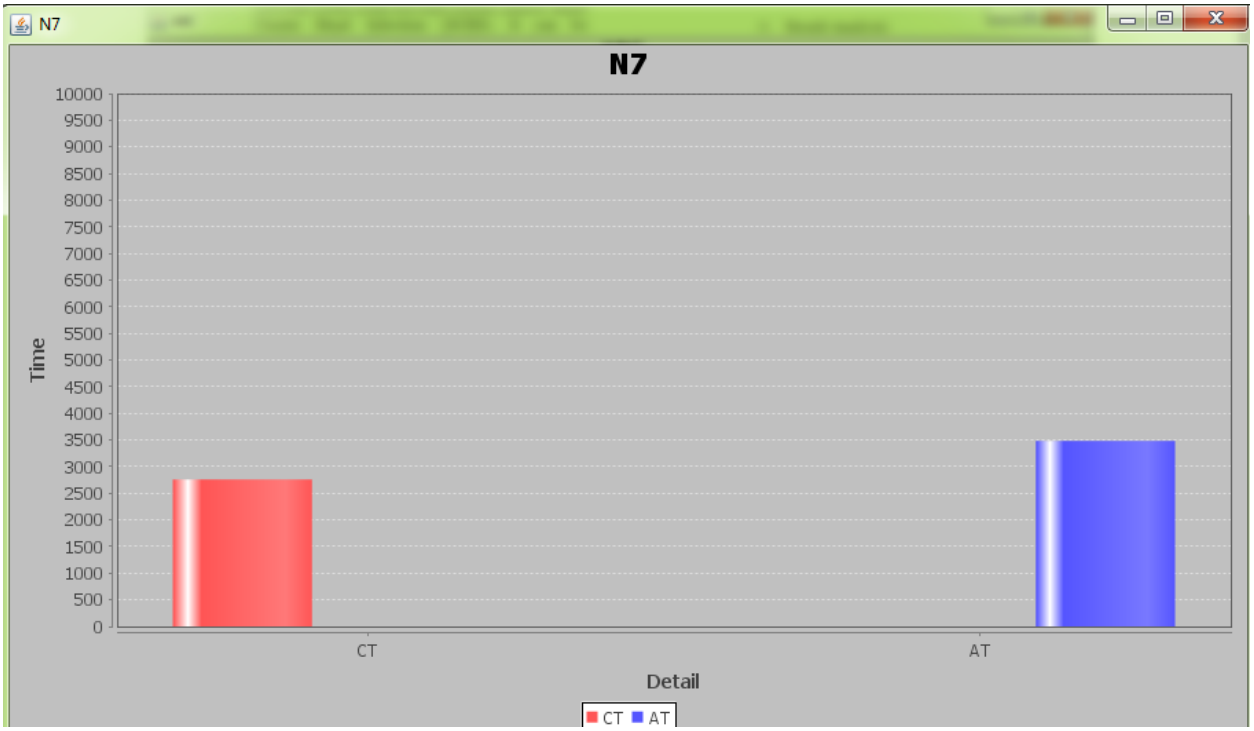

Figure 4: Time based Comparison for Node 3

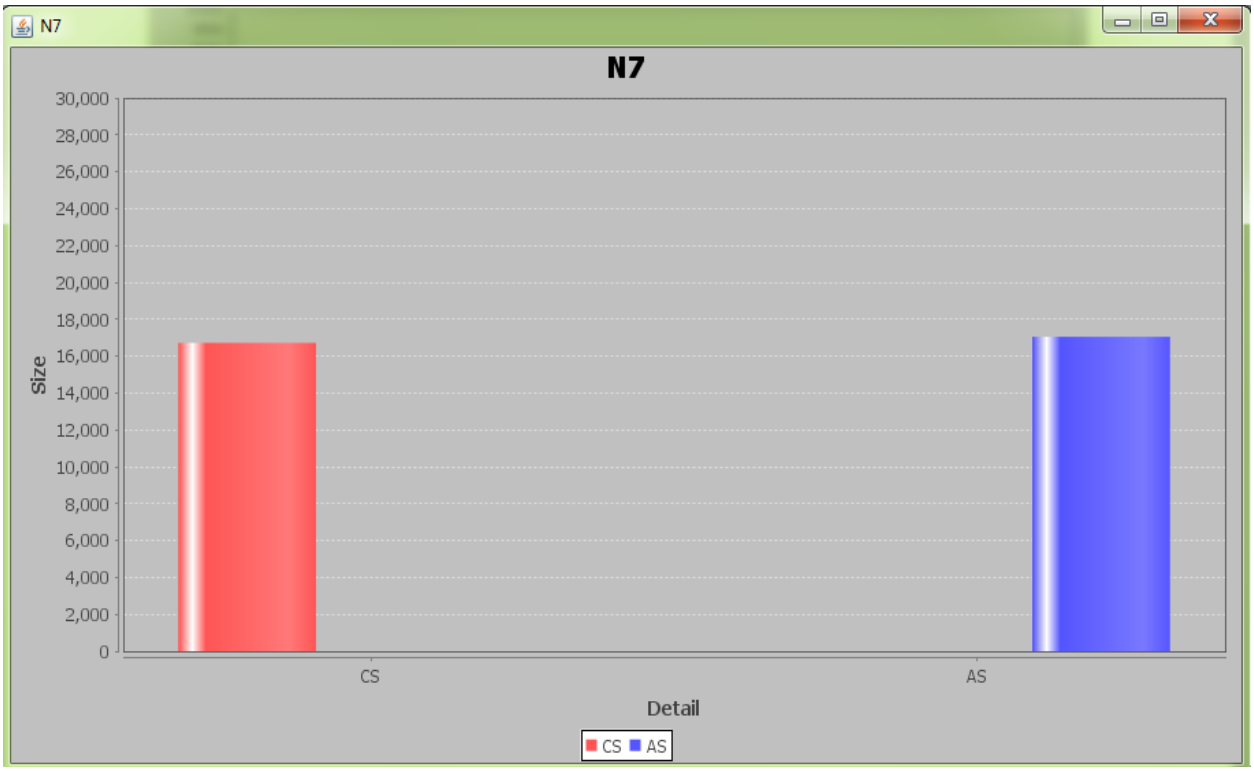

Figure 5: Time based Comparison based on space

5. Conclusions

In this paper we have presented an efficient K-means Clustering approach in Wireless Network for data sharing. This scheme provide the facility of distributed and centralize system. So that our framework will get the benefits of both of the two. In this way we will maintain better flexibility in terms of node selection, time and space management.

References

[1] Younis O., Fahmy S.: 'Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks', IEEE Trans. Mobile Compu!., 2004, 3, (4),pp. 366-379.

[2] L. Paradis and Q. Han, "A Survey of Fault Management in Wireless Sensor Networks", Journal of Network and Systems Management, vol. 15, no. 2, pp. 171-190, 2007.

[3] G. Venkataraman, S Emmanuel and S.Thambipillai, "Energy efficient cluster-based scheme for failure management in sensor networks" lET Commun, Volume 2, Issue 4, April 2008 Page(s) 528-537.

[4] S. Chessa and P. Santi, "Crash faults identification in wireless sensor networks", Compu!. Commun., 2002, 25, (14), pp. 1273-1282.

[5] G. Gupta and M. Younis; Fault tolerant clustering of wireless sensor networks; WCNC'03, pp. 1579.1584.

[6] Satish L. Yedage, D. C. Mehetre," A Survey on Different Wake-up Scheduling in WSN", International Journal of Advanced Computer Research (IJACR) Volume-4,Number-1, Issue-14, March-2014.

[7] Schaeffer, S.E., Survey: Graph clustering. Computer Science Review I (2007), 2007: p. pp. 27-64.

[8] Dongen, S.v., Graph Clustering by Flow Simulation. 2000, University of Utrecht, the Nertherlans. Latapy, P.P.M., Computing communities in large networks using random walks. J. Graph Algorithms Appl., 2006. 10(2): p.191-218.

[9] Haijun Zhou, R.L., Network Brownian motion: A new method to measure vertex-vertex proximity and to identify communities and subcommunities. International conference on Computational Science, pages 1062-1069, 2004.

[10] Sergey Brin, L.P., The anatomy of a large scale hypertextual web search engine. Computer Networks and ISDN Systems, 1998. 30 p. 107-117.

[11] Jia-Yu Pan Hyung-Jeong Yang, C.F.P.D., Automatic Multimedia Cross-modal Correlation Discovery, in In Proceedings of the 10th ACM SIGKDD Conference 2004, . August 22-25, 2004: Seattle, WA, USA.

[12] Kumar, G.K.a.V., Parallel multilevel k-way partitioning for irregular graphs. SIAM Review, 1999. 41(2): p. 278-300.

[13] A. Ng, M.J., and Y. Weiss, On spectral clustering: Analysis and an algorithm. In NIPS, 2001: p. 849-856.

[14] I. S. Dhillon, S.M., and D. S. Modha. , Informationtheoretic co-clustering, in In The Ninth ACMSIGKDD International.

[15] Conference on Knowledge Discovery and Data Mining (KDD 03). August 24-27, 2003: Washington, DC.

[16] Newman, M.G.a.M.E.J., Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA, 2002. 99: p.7821–7826.

[17] Bingjing Cai; Haiying Wang; Huiru Zheng; Hui Wang, "An improved random walk based clustering algorithm for community detection in complex networks," Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on , pp.2162,2167, 9-12 Oct. 2011.

[18] H. Srikanth. Kamath ,"Energy Efficient Routing Protocol for Wireless Sensor Networks", International Journal of Advanced Computer Research (IJACR),Volume-3, Number-2, Issue-10, June-2013.

[19] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," Computer Communications, vol. 30, pp. 2826–2841, 2007.

[20] A. Amis, R. Prakash, T. Vuong, and D. Huynh, "Max-min d-cluster formation in wireless ad-hoc networks," in INFOCOM, 2000.

[21] W. Chen, J. Hou, and L. Sha, "Dynamic clustering for acoustic target tracking in wireless sensor networks," in ICNP, 2003, pp. 284–294.

[22] Bhavesh Joshi, Anil Khandelwal, " Clustering with Data Encryption in Wireless Communication: A Critical Survey ", International Journal of Advanced Computer Research (IJACR), Volume-4, Issue-16, September-2014, pp.813-818.

[23] Priyanka Tavse, Anil Khandelwal, " A Critical Review on Data Clustering in Wireless Network " , International Journal of Advanced Computer Research (IJACR), Volume-4, Issue-16, September-2014 ,pp.795-798.

[24] P. Ding, J. Holliday, and A. Celik, "Distributed energy efficient hierarchical clustering for wireless sensor networks," in DCOSS, 2005.

[25] K.V.Kulhalli, Prajakta Rane," On Demand Multipath Routing Algorithm for Adhoc Wireless Networks", International Journal of Advanced Computer Research (IJACR),Volume-4,Number-1,Issue-14, March-2014.

[26] Mohammed A.M. Ibrahim, "Utilization of Secure Wireless Networks as Environment for Learning and Teaching in Higher Education", International Journal of Advanced Computer Research (IJACR), Volume-4, Issue-14, March-2014, pp.251-257.
[27] AbdelRahman Hussein, Sufian Yousef, Samir Al-Khayatt and Omar S. Arabeyyat, "An Efficient Weighted Distributed Clustering Algorithm for Mobile Ad hoc Networks", IEEE 2010.
[28] Minjie Guo, Leibo Yao, Wenli Zhou, Yuanyuan Qiao," To Group At The Base of Users' Usage Preference of Network Services Based On Fast Hierarchical Clustering Algorithm", Proceedings of AIAI2010.
[29] Yu Wang, Yang Xiang, Jun Zhang and Shunzheng Yu," A Novel Semi-Supervised Approach for Network Traffic Clustering",IEEE 2011.
[30] Caimei Lu, Xiaohua Hu, and Jung-ran Park," Exploiting the Social Tagging Network for Web Clustering", IEEE Transactions On Systems, Man, and Cybernetics—Part A: Systems And Humans, Vol. 41, No. 5, September 2011.
[31] P. Sasikumar and Sibaram Khara, "K-Means Clustering In Wireless Sensor Networks", 2012 Fourth International Conference on Computational Intelligence and Communication Networks, IEEE, 2012, pp. 140-144.
[32] M. Bala Krishna and M. N. Doja," Self-Organized Energy Conscious Clustering Protocol for Wireless Sensor Networks", IEEE 2012.
[33] Vipin Pal, Girdhari Singh, Rajender Prasad Yadav," SCHS: Smart Cluster Head Selection Scheme for Clustering Algorithms in Wireless Sensor Networks", Wireless Sensor Network, 2012, 4, 273-280.

Ms. Priyanka Tavse hails from Bhopal and was born on 20th December 1989. She did her schooling from Nalanda Public H.S.School, Bhopal and completed her graduation in stream of Electronics and Communication from Swami Vivekanand College of Science and Technology Bhopal with 73.33%. She is a PG Scholar at VNS Group of Institutions: Faculty of Engineering, Under Rajiv Gandhi Proudhyogiki Vishwavidhyalaya, Bhopal. and pursuing M.Tech. in Digital Communication. She is keen to work on Data Clustering in Wireless Network as a professional.
Email: priyankatavse@gmail.com