

ENERGY EFFICIENT AND SECURED WAY OF BIG DATA GATHERING IN WSN USING K-MEDOIDS ALGORITHM

Ashwini.S¹, Srinivasa.H.P²

¹MTech student, Department of CSE, TJohn Institute of Technology, Bangalore, Karnataka, India

²Associate Professor, Department of CSE, TJohn Institute of Technology, Bangalore, Karnataka, India

Abstract

Big data has considered being promising application in the field of data and correspondence innovation (ICT). WSNs are used for gather big data to communication technology. In most of case the data from single sensor is not enough informative to user. To gather significant data, it is necessary to gather information from distributed sensors. Because of limited power in WSNs big data gathering is challenging task. To increase network lifetime in WSNs the ways for information exchange are chosen in a manner that the aggregate vitality expended along the way is decreased. To bolster enhanced information accumulation and high versatility, sensor hubs are regularly gathered into non-covering, disjoint subsets called groups. Bunches make various levelled WSNs which consolidate effective use of restricted assets of sensor hubs and thus augments lifetime of system. Here we proposed an efficient way for data gathering using clustering based technique. Here we used k-medoids clustering technique for the ideal number of bunches and made sink hub trajectory, used as efficient data collector.

Keywords: Wireless Sensor Network (WSNS), Big Data, Clustering, Streamlining, Information Gathering, and Vitality Proficiency

1. INTRODUCTION

Wireless sensor networks (WSNs) have come up with a powerful answer for a considerable lot of the applications like military, healthcare, home, etc. recent progresses in remote interchanges and hardware have empowered the improvement of ease, low power and multi practical sensor hubs[3]. A remote sensor system comprises of a few number of sensor hubs that are distributed densely within a network area. The position of sensor hubs are need not be foreordained or built. This permits arbitrary dissemination in out of reach landscapes or catastrophe alleviation operations. Then again, this likewise implies sensor system conventions and calculations must have self-arranging capacities. Another extraordinary element of sensor systems is the agreeable exertion of sensor hubs. Sensor hubs are fitted with an installed processor. Rather than sending the crude information to the hubs in charge of the combination, they utilize their preparing capacities to locally do basic calculations and transmit just the required and halfway handled information. The depicted components guarantee an extensive variety of uses for sensor systems. A percentage of the application regions are wellbeing, military and home. In military, for instance, the fast arrangement, self-association, and adaptation to non-critical failure attributes of sensor systems make them extremely encouraging detecting strategies for military order, control, interchanges, figuring, knowledge, observation, surveillance, and focusing on frameworks. Acknowledgment of these and other sensor system applications require remote impromptu systems administration procedures. Conventions used in wireless ad hoc networks were not well suited the unique features and application requirements of sensor networks. The major limitations of remote sensor arrange their restricted force

stockpiling limits and inclined to disappointments. Thus vitality productivity is the real issue in sensor system. As information technology grows rapidly, volume of the data also increasing simultaneously. Accessing the retrieving big data is crucial for the user in many applications[1][5]. remote sensor arranges their restricted force stockpiling limits and inclined to disappointments. Thus vitality productivity is the real issue in sensor system. the information is too enormous or it moves too quick or it surpasses current preparing limit. Enormous information can possibly help organizations enhance operations and make speedier, wiser choices. Collecting large amount data from sensor nodes is the major concern in the field of ICT. Individual sensor nodes may not provide accurate information. Therefore collecting data from multiple sensor nodes is very essential[2].

So as to assemble these information, the WSNs are built in such a way the sensors transfer their information to the "sink". Be that as it may, in the event of generally and thickly disseminated WSNs there are two issues in social event the information detected by a huge amount of sensors. Initially, the system is isolated to some sub-systems due to the constrained remote correspondence range. Second, the remote transposal devours the vitality of the sensors. Despite the fact that the volume of information produced by an unitary sensor is not huge, every sensor wants a considerable measure of vitality to transfer the information created by encompassing sensors. Particularly in thick WSNs, the life span of sensors will be precise in light of the fact that every sensor hub transfers a considerable measure of information created by colossal number of encompassing sensors. Keeping in mind the end goal to take care of these

issues, we require a vitality productive technique to assemble enormous intensity of information from a substantial number of sensors in the thickly conveyed WSNs[4].

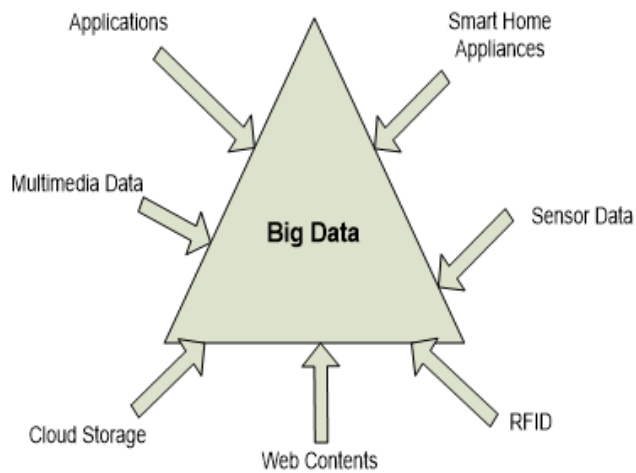


Figure 1. An Overview of Big data Gathering

To accomplish vitality proficient information accumulation in thickly dispersed WSNs, there have been numerous current methodologies. Grouping is a standout amongst the most broadly utilized systems to make WSNs vitality productive. Low vitality versatile grouping various levelled (LEACH) bunching calculation is acquainted with expand the life time of remote sensor system. In utilized desire augmentation bunching method. In this paper we proposed an enhanced strategy to get-together huge information in thickly circulated sensor system utilizing k-medoid grouping calculation. K-medoid calculation is utilized as grouping of WSNs. K-medoid is not just viable in bunching additionally powerful in selecting group heads in huge WSNs. Later we presented a successful sink hub steering for gathering information from every bunch heads.

2. Related work

In recent literature many techniques have been proposed to achieve efficient data aggregation in WSNs namely LEACH (Low-Energy Adaptive Clustering Hierarchy), EM (Expectation Maximization) [7]. All these methodologies have attempted their level best for decrease vitality utilization and draw out system lifetime. Drain is a self-sorting out, versatile grouping convention. To have least vitality utilization, hubs in LEACH are assembled into various groups in view of their battery use. Every bunch has a group head, which speaks with each hub of that group. The sink totals information, transmitted by bunch heads, from different hubs. Since a bunch head loses vitality because of rehashed transmissions, the group head is re-chosen in view of the leftover vitality, as an outcome it drags out the system lifetime[10].

Coordinated Diffusion is an information driven conventions regularly utilized as a part of remote sensor systems. It comprises of a few components: interests, slopes, information messages and fortifications (positive and negative). To total information by utilizing Directed Diffusion, the sink hub telecasts an "Interest" message that comprises of a period to-live esteem, furthermore the locations of the source and destination hubs. The destination hub on getting the solicitation transmits fitting information message to the source having the detected information. In the event that the downstream hubs can't be come to by the "interest" message from the present source then the present destination turns into the source hub by changes its location, decreases the time-to-live esteem and rebroadcasts the "Interest" message. The EM calculation is an established bunching calculation, where hubs are conveyed by blend dispersion to locate the most extreme probability estimation of the peripheral probability by iteratively applying the E-step and M-step. The equation to EM can be expressed as:

$$p(x) = \sum_{k=1}^K (\pi_k N(x|\mu_k, \Sigma_k)) \quad (1)$$

Where K and π_k show the aggregate number of bunches and the blending coefficient of the k th group, separately. $N(x|\mu, \sigma)$ is characterized as takes after,

$$N(x|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} \quad (2)$$

where x is the position vectors of all hubs. Bunch parameters, μ_k and Σ_k , are the position vector of centroid of group k and 2×2 covariance grid of the k th group, separately. At the underlying step, EM figuring learns each center point's estimation of level of dependence that is insinuated as commitment. The commitment demonstrates how much a hub relies on upon a bunch. The n th hub's estimation of level of reliance on k th group is given by taking after comparison.

$$Y_{nk} = \frac{\pi_k N(x_n | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j N(x_n | \mu_j, \Sigma_j)} \quad (3)$$

In perspective of its definition, the commitment takes values some place around 0 and 1. At the second step, the EM computation surveys K weighted centre of gravity of a 2-dimensional range vector of centres. This evaluation uses the commitment regard as weight of center points. At the third step, the ranges of the group centroid are changed to the weighted centers of gravity evaluated in the second step. Moreover, EM count evaluates the estimation of the log likelihood as showed as takes after.

$$P = \ln p(X|\mu, \Sigma, \pi) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k N(x_n | \mu_k, \Sigma_k) \right\} \quad (4)$$

Prior to the estimation of log probability joins, the EM calculation refreshes everystrides. This estimation of log probability is unvaryingly diminishing, and the EM calculation dependably ends. Since the EM calculation over and again redesigns bunch centroid' position vector, μ_k , and hubs' obligation to kth group, $[\gamma_n]_k$, the whole of square of separations of every hub to bunch step by step diminishes lastly gets to be ideal.

3. Proposed Algorithm

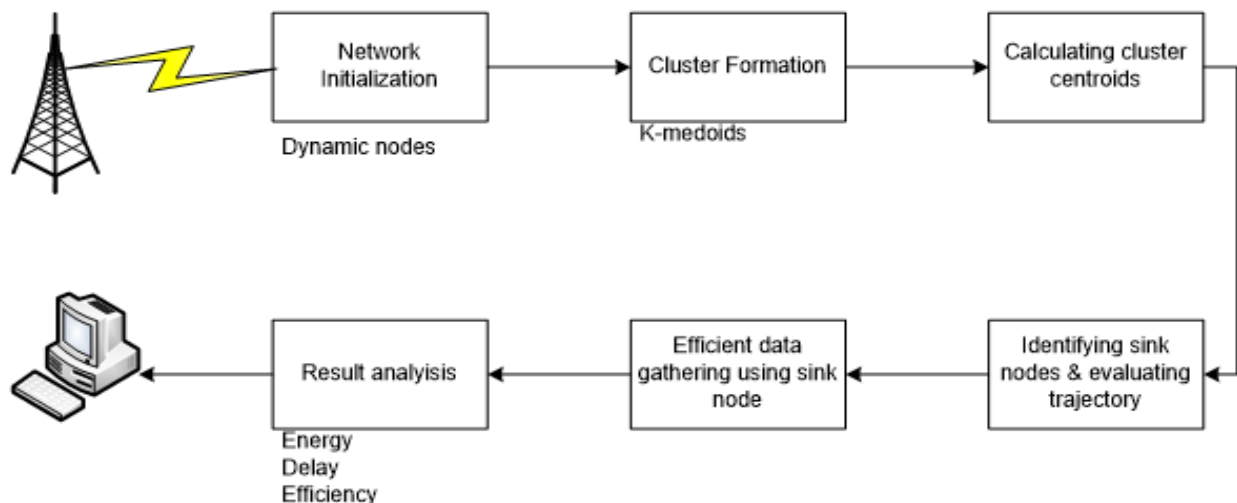


Figure 2. Block Diagram of Proposed Architecture

3.1 K-medoids Algorithm

Not exactly the same as routine gathering counts for remote sensor frameworks, in which bundle heads are picked firstly and gatherings are surrounded besides by allowing non to head centres interface with some head, the proposed bunching investigates the backwards method: firstly, the hubs are separated into unmistakable bunches with the medoid grouping plan, and afterward one hub inside every bunch will be chosen to be the head. Our strategy for building groups in dispersed style performs over conventional LEACH-like calculations regarding activity parity and execution cost, consequently prompting a more extended framework lifetime. We utilize the k-medoid calculation to isolate the system hubs into k unmistakable groups. Calculation 1 exhibits the stream of developing bunches. The fundamental thought of the proposed grouping calculation dwells in ceaseless cycles until an arrangement of heads is made sense of such that these heads convey uniformly inside the system zone and all bunches are about the same in the quantity of part hubs.

3.2 Algorithm for K-medoids

Require: the system graph $G(V,E)$; z , the quantity of groups to be framed; λ , the blurring element of channel

Guarantee: an arrangement of bunches including all hubs

In this paper we mainly focused on proficient enormous information gathering from heterogeneous remote sensor systems. At the point when manages the heterogeneous sensor hubs, there will be a problem of data inaccuracy. To overcome all the problems, we implemented effective cluster based technique.

The Figure 2 depicts the block diagram of proposed system. In network initialization we create a network terrain with specific area and finite number of sensor nodes. For cluster formation we used k-medoids clustering algorithm. K medoids forms clusters and gives cluster heads.

1: Select z hubs arbitrarily as starting bunch centre node, and utilize x to mean the arrangement of those z group centre node.

2: Constitute an unfilled set P_O for head Y_O of x , and introduce P_O by $P_O \cup Y_O$

3: Constitute a set E instated vacant

4: while any hub l of j/x do

5: l picks Y_O of x as its centre node such that

$$\arg(Y_O) \min\{d(l, Y_O)^\lambda \mid \forall Y_O \in x\}$$

Where $d(l, Y_O)$ is the Euclidean separation from l to Y_O .

6: $P_O \leftarrow P_O \cup l$

7: end while

8: while any P_O do

9: Select a hub l from P_O as new centre node such that $\arg_l \min\{\sum_{(q=1, q \neq l)} P_O \mid [d(l, q)^\lambda \mid \forall l \in P_O]\}$

```

10: E ← EU 1
11: end while
12: if E ≠ x then
13: x ← E
14: E ← ∅
15: go to step 4
16: end if
    
```

3.3 The Trajectory of Mobile Sink Node

In the wake of bunching of WSN hubs, we will decide the genuine direction of the portable sink. The portable sink navigates through bunches and totals information from different hubs. Since it is conceivable to build effectiveness by decreasing the voyaging time, it is ideal that the versatile sink follows the briefest way among the bunch centroid.

3.4 Mobile Sink Data Collection

Subsequent to touching base at the centroid, the versatile sink will telecast information demand message to hubs, which will thusly attempt to send the information gathered to the portable sink either straightforwardly or by means of different hubs in a multi-jump design. Notwithstanding sending the gathered information, the hubs will likewise rebroadcast the information demand message to their neighbour hubs.

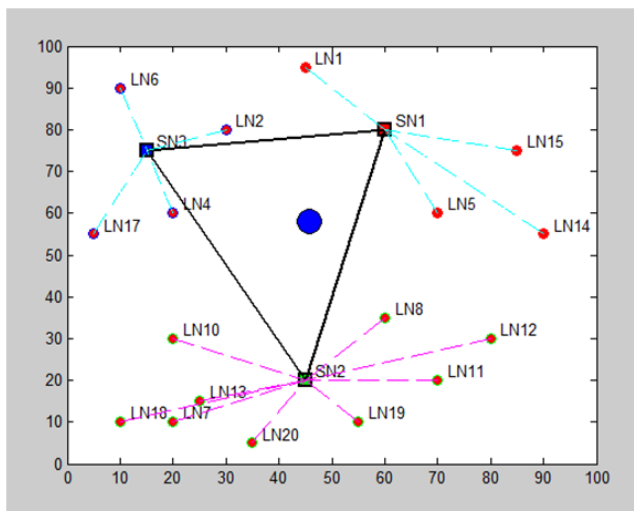
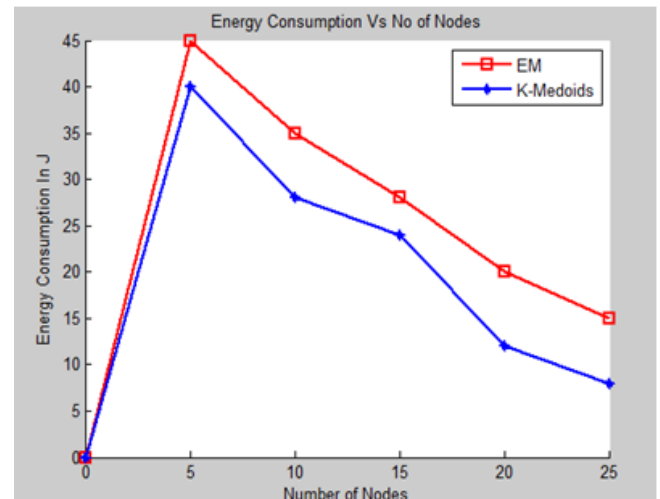


Figure3. Data Gathering using Sinknode Trajectory

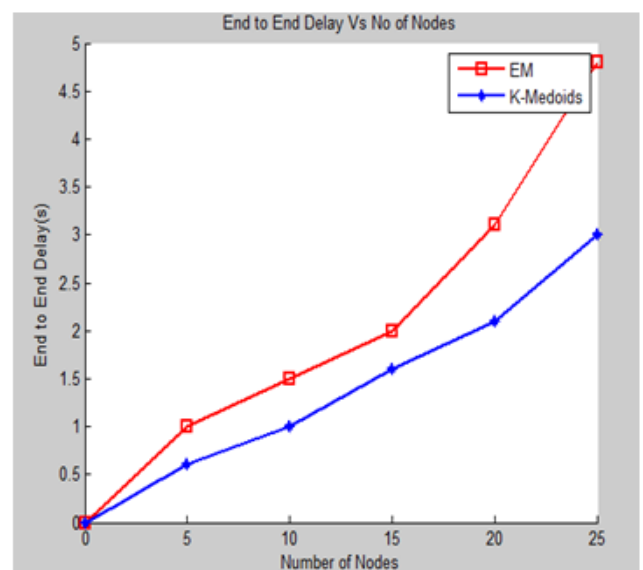
4. Performance Evaluation

The proposed clustering algorithm is simulated by MATLAB. First we evaluated efficiency in energy consumption, second we evaluated reduced delay in data communication and finally we calculated packet delivery ratio.



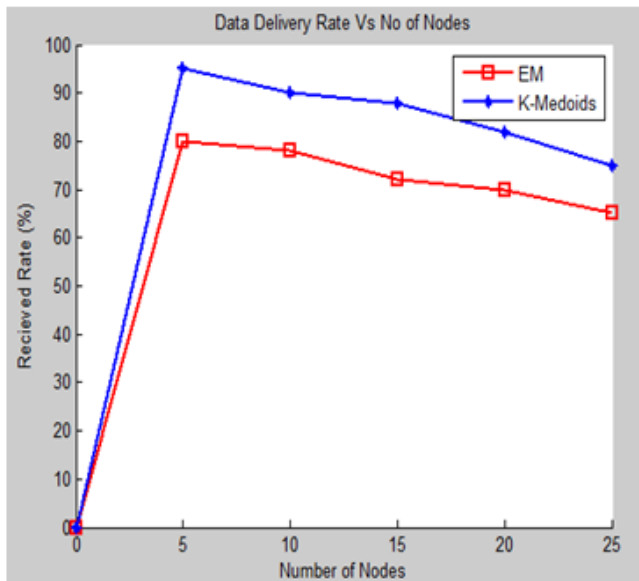
a. Energy Consumption of Sensor Nodes

In this above fig .we see that EM technique does not consume energy due to this drawback the data packets will be send to nodes and there may be delay with in packets .so to overcome the delay of packets .we use another algorithm called as K-Medoid it just Collect the nodes with neighbor nodes and form a groups so all the data information will be stored in the cluster so the data can be easily passed and can save energy .



b. Delay in Data Communication

In the above fig .The packets delay may occur while passing the data from nodes tobase station .Using EM technique. To overcome this We use K-medoid algorithm technique so the packets transfer from node to CH and from CH to BS the packets will be transferred without delay by using time slots to pass the packets.



c. Packet Delivery Ratio of Proposed method

In the above figure. The data packets delivery was not in a flow there was a packets loss while transferring the data in EM technique. To overcome the packets Delivery we use K-Mediod to Transfer the data without delay by using cluster method each data was transferred within the data rate given to the packets .as the data packets were transferring with in delivery time given to each data packets without any packets loss.

5. CONCLUSION

In this paper we actualized a vigorous calculation for big data gathering using k-medoid algorithm. K-medoid clustering is reliable clustering algorithm in reducing power consumption mainly for network with large number of sensor nodes. K-medoid is also effective algorithm in selecting centroid. For reducing delay and also improve the performance of the network we used sink node trajectory to gather data. The results show the effectiveness of proposed system. Our method is very robust in lessening delay, energy consumption and also the performance of WNSs magnificently improved.

Once the cluster head gathers information from each node within a cluster, then that information should be passed to sink node in single-hop path. If the communication range is far than normal then multi-hop path is taken for gathering data. By using optimal routing mechanism we can gather big-data energy efficiently and also by using novel approaches it is possible to secure the data effectively from eavesdroppers.

REFERENCES

[1] Frank Armour, American University, fjarmour@gmail.com.” Big Data: Issues and Challenges Moving Forward”, 1530-1605/12 \$26.00 © 2012 IEEE.

- [2] Danah boyd a & Kate Crawford ba Microsoft Research, One Memorial Drive, Cambridge, MA, 02142, USA,” CRITICAL QUESTIONS FOR BIG DATA”, Information, Communication & Society Vol. 15, No. 5, June 2012, pp. 662–679 ISSN 1369-118X print/ISSN 1468-4462 online # 2012 Microsoft.
- [3] F. L. LEWIS Associate Director for Research ,Head, Advanced Controls, Sensors, and MEMS Group Automation and Robotics Research Institute ,The University of Texas at Arlington,”Wireless Sensor Networks”,ed. D.J. Cook and S.K. Das, John Wiley, New York, 2004.
- [4] Gregory J. Pottie Department of Electrical Engineering University of California, Los Angeles, California 90095-1594 pottie@icsl.ucla.edu, http://Ndantzig.ee.ucla.edu/oclab/Pottie.html “Wireless Sensor Networks”, -0-7803-4408-1/98/\$10.00 2010.
- [5] Big Data: The Management 60 Revolution by Andrew McAfee and Erik Brynjolfsson Data Scientist: The SexiestJob of the 70 21st Century by Thomas H. Davenport and D.J. Patil,” Big Data Businesses are collecting more data than they know what to do with. To turn all this information into competitive gold, they’ll need new skills and a new management style”, Artwork Tamar Cohen Doodle Dash Board (detail) 2011, silk screen on vintage catalog page.
- [6] Shuwei Chen¹ , Kathy Clawson² , Min Jing¹ , Jun Liu¹ , Hui Wang¹ , Bryan Scotney² ¹ School of Computing and Mathematics, University of Ulster Shore Road, Newtownabbey, BT37 0QB, U.K. ² School of Computing and Information Engineering, University,” Uncertainty Reasoning Based Formal Framework for Big Video Data Understanding”, 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT).
- [7] Shwet Ketu Indian Institute of Information Technology Allahabad, India shwetiiita@gmail.com,” Effect of Corpus Size Selection on Performance of Map Reduce Based Distributed K-Means for Big Textual Data Clustering”, and/or a fee. Request permissions from Permissions@acm.org. ICCCT '15, September 25-27, 2015, Allahabad, India © 2015 ACM. ISBN 978-1-4503-3552-2/15/09\$15.00.
- [8] Kazim Sari * Faculty of Economics and Administrative Sciences, Department of International Logistics and Transportation, Beykent University, Sisli Ayazaga Mah., Hadimkoru Yolu Mevkii, 34396 Sisli, Istanbul, Turkey,” Exploring the impacts of radio frequency identification (RFID) technology on supply chain performance”, European Journal of Operational Research 207 (2010) 174–183.
- [9] A. Divyakant, B. Philip, and et al., “Challenges and opportunities with Big Data,” 2012, a community white paper developed by leading researchers across the United States.
- [10] M. Youssef, A. Youssef, and M. Younis, “Overlapping multihop clustering for wireless sensor networks,” IEEE Transactions on Parallel and Distributed Systems, vol. 20, no. 12, pp. 1844–1856, Dec. 2009.

- [11] Parminder Kaur, Mamta Katiyar, "Energy-Efficient Chain-Based Data Gathering in WSN," International Journal of Science, Engineering and Technology Research (IJSETR) Volume 1, Issue 6, December 2012.
- [12] Hae-sang park, Chi-Hyuck jun, "A simple and fast algorithm for k-medoids clustering", department of Industrial and Management Engineering, POSTECH, San 31 Hoja-dong, Pohang 790-784, south Korea, available online 9 February 2008.