ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 8, August 2015

An Efficient way to Gather Big Data in WSN using Mobile Sink Routing

Vijayalaxmi

M.tech[4th Semester], Computer Science and Engineering, Acharya Institute of Technology, Bangalore, India

Abstract: The big data has become a popular topic because of enormous growth in the field of information technology. The distributed sensor network is the production of big data. A single sensor in a network may not give a significant data, but the data sensed by millions of sensors produce a big data. Big data gathering in densely distributed sensor network with energy efficiently is a challenging task. One of the most efficient solutions for this challenge is the utilization of sink node. This method can determine the sink node's trajectory and data gathering method through clustering. In this paper the proposed method is establishing sink routing and clustering method based on k-medoid algorithm.

Keywords: Big Data, Wireless Sensor Network, Sink Node, Sink Routing, Routing, Clustering.

I. INTRODUCTION

Wireless Sensor network is a network of devices, indicated as sensor nodes, these nodes sense the environment and converse the data from the monitored field. The data is forwarded to a sink node via single or multiple hops. WSN has come up with an effective solution for many of the applications like military, healthcare, home, etc. Recent advances in wireless communications and electronics have enabled the low cost, low power and multi functional sensor nodes that are distributed densely within a network area. The unique feature of sensor network is the cooperative effort of sensor nodes.

As information technology develops rapidly, volume of the data also increasing concurrently. Big data is a buzzword used to describe a massive amount of both structured and unstructured data that is so large that it's difficult to process using traditional database and techniques. In most enterprise scenario the data that is too big or it moves too fast or it exceeds current dealing capacity. Big data has the potential to guide the companies improve operations and make faster, more intelligent divisions. Collecting large amount data from sensor nodes is the major concern in the field of information and communication technology. Individual sensor nodes may not provide accurate information. Therefore collecting data from multiple sensor nodes is very essential.

In order to gather these data, the WSN is constructed in such a way the sensors relay their data to the "sink". However, in case of widely and densely distributed WSNs, there are two problems in gathering the data sensed by millions of sensors, first, the network is divided to some sub-networks because of the limited wireless communication range. Second, the wireless transmission consumes the energy of the sensors. The data generated by an individual sensor is not significant; each sensor requires a lot of energy to relay the data generated by surrounding sensors. The life time of sensors in dense WSNs will be very short because each sensor node relays a lot of data generated by tremendous number of surrounding sensors.

II. RELATED WORK

In recent literature survey many techniques have been proposed to achieve efficient data aggregation in WSNs namely LEACH (Low-Energy Adaptive Clustering Hierarchy), EM (Expectation-Maximization). All these approaches have tried their level best for lesser energy consumption and prolong network lifetime.

LEACH is a self-organizing, adaptive clustering protocol. To have minimum energy consumption, nodes in LEACH are grouped into a number of clusters based on their battery usage. Each cluster has a cluster head, which communicates with every node of that cluster. The sink aggregates data, transmitted by cluster head, from other nodes. Since a cluster head loses energy due to repeated transmissions, the cluster head is re-selected based on the residual energy, as a consequence it prolongs the network lifetime.

Directed Diffusion [6] is a data centric protocols commonly used in wireless sensor networks. It consists of several elements: interests, gradients, data messages and reinforcements. To aggregate data by using Directed Diffusion, the sink node broadcasts an "Interest" message from the current source then the current destination becomes the source node by changes its address, reduces the time-to-live value and rebroadcasts the "interest" message.

The EM algorithm is a classical clustering algorithm, where nodes are distributed according to Gaussian mixture distribution to find the maximum like hood estimation of the marginal like hood by iteratively applying the E-step an d M-step.

The equation to EM can be expressed as:

$$P(x) = \sum_{k=1}^{k} \pi k N(x|\mu k, \sum k))$$

Where k and π indicate the total number of clusters and the mixing coefficient of the kth cluster, respectively.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 8, August 2015

III. PROPOSED METHODOLOGY

A. System Architecture

The architectural design process is concerned with establishing a basic structural framework for a system. It involves identifying the major components of the system and communication between these components. This paper focused on efficient big data gathering from heterogeneous wireless sensor networks. When deals with the heterogeneous sensor nodes, there will be a problem of data in accuracy. To overcome all the problems, here proposing a effective cluster based technique. Figure 1 depicts the block diagram of proposed system. Network initialization phase contains a network terrain with specific area and finite number of sensor nodes. K-medoid clustering algorithm is used for cluster formation and it creates the cluster head.

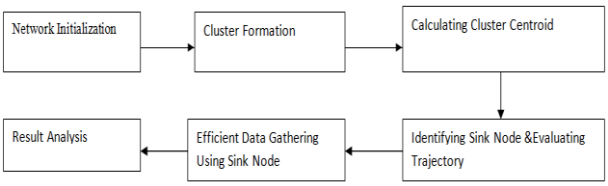

Fig.1 System Architecture

B. The trajectory of the mobile sink

After clustering of WSN nodes, here determine the actual trajectory of the mobile sink. The mobile sink traverses through clusters and aggregates data from various nodes. Since it possible to increase efficiency by reducing the travelling time, it is preferable that the mobile sink traces the shortest path among the cluster centroids.

C. Sink Node data Collection

After arriving at the centroid, the mobile sink will broadcast request message to nodes, which will in turn try to send the data collected to the mobile either directly or via other nodes in a multi-hop fashion. In addition to sending the collected data, the nodes will also rebroadcast the data request message to their neighbour nodes.

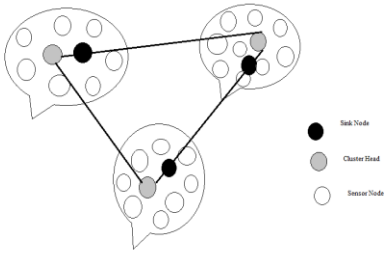

Fig.2 Data Gathering Using Sink

D. K-medoid Clustering

In traditional clustering algorithm at the first step cluster heads are selected and then clusters are formed. But in proposed algorithm at the first phase the nodes are divide into distinct clusters with k-medoid clustering scheme and then one node inside each cluster will be selected as a cluster head. Proposed method of constructing clusters in distributed fashion performs over traditional LEACH-like algorithms in terms of traffic balance and execution cost, thereby leading to a longer system lifetime. K-medoid algorithm divides the network nodes into k distinct cluster. Algorithm 1 presents the flow of constructing clusters. The basic idea of the proposed clustering algorithm resides in continuous iteration until a set of heads is figured out such that these heads distribute evenly within the network area and all clusters are about the same in the number of member nodes.

The k-medoid algorithm steps explained as follows

i. Choose k nodes arbitrarily as the initial medoids
ii. Repeat the first step
iii. Arrange every remaining node to the nearest medoid.
iv. Randomly choose a non medoid node.
v. Calculate the total rate of swapping old medoid node with newly chosen non-me of k-medoid
vi. If the total cost of swapping is less than zero, then execute that swapping procedure to outline the new set of k-medoid.
vii. Until no alter.

E. Performance Evaluation

The proposed clustering algorithm is simulated by MATLAB. Environment of the experiment is given below.

Node distribution	Uniformly random
Number of Clusters k	3
Number of Node N	30
Length of one side of the field	150

Table.1 Setup for the experiment

In this experiment the energy and delay is measured by the proposed clustering algorithm by varying the number of nodes. Table.1 shows the parameter used in this. Sensors are uniformly deployed in a 150 square metre area. After deploying the nodes, clusters are formed. The cluster head is selected based on Euclidian distance within the cluster, which is the centre of the medoid is selected as a cluster. After cluster formation phase the cluster head starts collecting the data from every node. Then Sink node will move around the sink trajectory and gather the information from cluster head. Here transmission distance is reduced hence energy consumption also reduced. This method takes less time to collect data. The graphs are plotted based on the experiment results compared with existing system.

The number of nodes vs. energy consumption is plotted.AS the number of nodes increases the energy consumption also increases. The energy consumption is less in proposed system as comparative to existing system.

The second graph is delay vs. number of nodes is plotted. The delay is less in proposed system as comparative to existing system. The nodes are divided into clusters, cluster head is selected based on distance parameter, which is most centre of the cluster selected a medoid or cluster head. The sink node moves directly to cluster head and collect the data. So here data transmission range is reduced, automatically delay will be reduced.

DOI 10.17148/IJARCCE.2015.4809

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 8, August 2015

The third graph is packet delivery and number of nodes. Packet delivery ratio is the total number of packet received into total number of packet sent. As the number of nodes increases packet delivery ratio also increases. The packet delivery ratio is less as comparative to existing system.

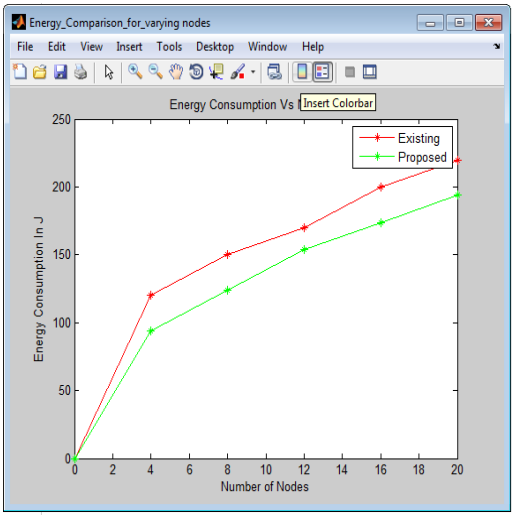

Fig.3 Energy consumption for varying nodes

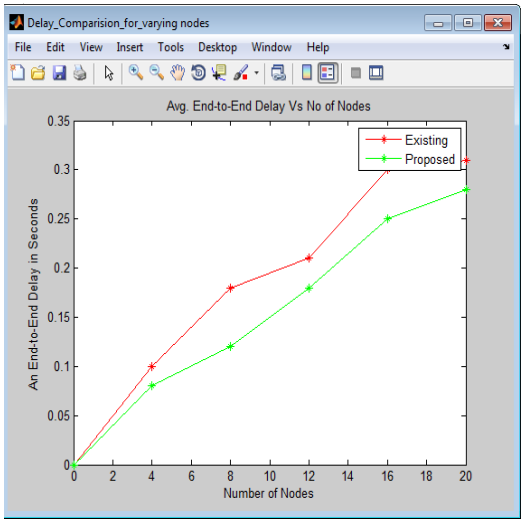

Fig.4 delay comparison for varying nodes

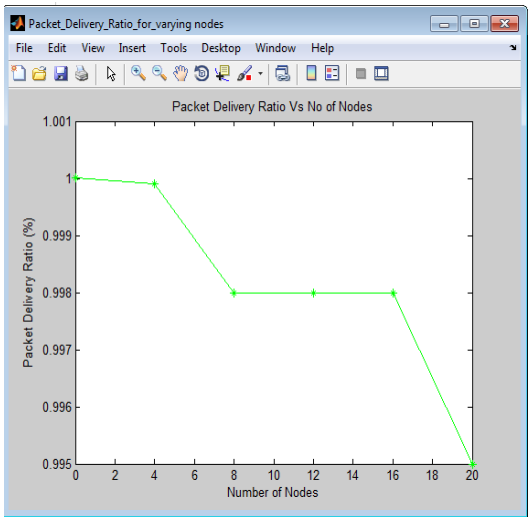

Fig.5 Packet delivery ratio for varying nodes

IV. CONCLUSION

The paper introduces a robust algorithm for big data gathering using k-medoids 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 a effective algorithm in selecting centroid. sink node trajectory is used to reduce the delay and improve the performance. The results show the effectiveness and also the performance of wsns magnificently improved.

REFERENCES

[1] IBM, Armonk, NY, USA, (2012, Jan). Four Vendor Views on Big Data and Big Analytics; IBM[Online]. Available:http//www/01.inm.com/software/in/data/big

[2] Priyanka Devi,Khushneetkaur,"A Robust Cluster Head Selection Method Based on K-medoid Algorithm to maximize Network Life Time and Energy Efficiency for Large WSNs",International Journal of Engineering Research &Technology,Volume 3,Issue 5,2014

[3] Li Peng,Guo-Yong Dong, Fei-Fei Dai and Guo-ping Liu," A new Clustering Algorithm based on ACO and k-medoidsOptimization Methods",The International Federation of Automatic control,2014.

[4] ChunyaoFu,Zhifang Jiang and Wei Wei2and Ang Wei,"An Energy Balanced Algorithm of LEACH protocol in WSN", International Journal of Computer Science Issues,Volume 10, Issue 1,January 2013.

[5] DeepaliVirmani and SatbirJain,"decentralized Lifetime Maximizing Tree with Clustering for Data delivery in Wireless Sensor Networks".

[6] ChalemekIstanagonwiwat,Ramesh Govindan,Deborah Estrin,John Heidemann,and Fabio Silva,"Directed Diffusion for Wireless Sensor Networking".

[7] R.Mehala and A.Balamurugan,"A survey on Enhanced Energy Efficient Data Gathering Scheme in Wireless Sensor Networks",International Journal of Science and Research,Volume 10.Issue 1,January 2013.

[8] Lin Zhao Zhibo Chen Guodong Sun School of Information Science and Technology,Beijing Forestry University,Beijing 100083,China "Dynamic Cluster-based Routing for Wireless Sensor Networks",Journal of International,Vol.11,November 2014

[9] Daisuke Takaishi,Hiroki Nishiyama,Nei Kato,and Ryu Miura, Graduate School of Information and Communications Technology,Tokyo Japan ,"Towards Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks", IEEE Transaction Volume 2,No,3,September 2014

[10] Priyanka Devi, Khushneet Kaur, "A balanced Cluster Head Selection Based on k-medoids to Enhance Wireless Sensor Network Life Time".

DOI 10.17148/IJARCCE.2015.4809