Open access Journal International Journal of Emerging Trends in Science and Technology

Big-Data Gathering Using Mobile Collector in Densely Deployed Wireless Sensor Network

Authors

Amruta. S. Pattanshetti[1], Mr. N. D. Kale[2]

[1]ME –II Student, PVPIT College, Pune, Maharashtra

Email: *amrutacpatil20@gmail.com*

[2]Assistant Professor, Department of Computer Engineering, PVPIT College, Pune, Maharashtra

Email: *navnath1577@yahoo.co.in*

Abstract

With tremendous growth of Information and Communication Technology (ICT) Wireless sensor network has major contribution in big data gathering. Even if data generated by individual sensor is not significant, the overall data generated by all sensors in the network is contributed for generation of significant portion of big data. Hence, the energy efficient big data gathering is a very challenging task in the densely deployed wireless sensor network. Also cluster formation before data collection from sensors in the network is additional challenge. Recent research addressed these challenges with mobile sink, which in turn raise the challenge of determining the sink nodes trajectory. In this paper we have proposed new solution, M-mobile collector based data gathering with network clustering based on improved Expectation maximization technique. Mobile collectors traverse a fixed path to collect data from cluster centroids and sensors in the clusters. Finally all the collected data is transferred amongst M-collectors to reach to the static sink node. Also we derive optimal number of clusters to minimize the energy consumption.

Keywords: *Big data, Wireless Sensor Networks (WSNs), clustering, optimization, data gathering, and energy efficiency.*

1. Introduction

The progress in various areas of Information and Communication Technology (ICT) has contributed to excessive growth in the volume of data. As a result of this the big data has emerged as a widely recognised trend which is currently being attracted much attention from government, academia and industry. The big data consists of high velocity, high volume and high variety of information assets as shown in Fig. 1, which is very difficult to gather , store and process data by using various available technologies. The total volume of data generated by an individual sensor is not that significant; individual sensor requires a lot of energy to relay the data generated by surrounding sensor nodes.

In case of dense sensor networks, the life time of sensors will be very short because each sensor node relays a lot of data generated by neighbouring sensors. In case of densely and widely distributed WSNs (e.g. in schools, urban areas, mountains, and so forth), [1], [2] there are two problems in gathering the data sensed by millions of sensors.

1. The network is firstly divided to some sub networks because of limited wireless communication range.

2. The wireless transmission consumes lot of energy of the sensors. Even though the total volume of data generated by an individual sensor is not that significant, each sensor requires a lot of energy to relay the data generated by surrounding sensors.

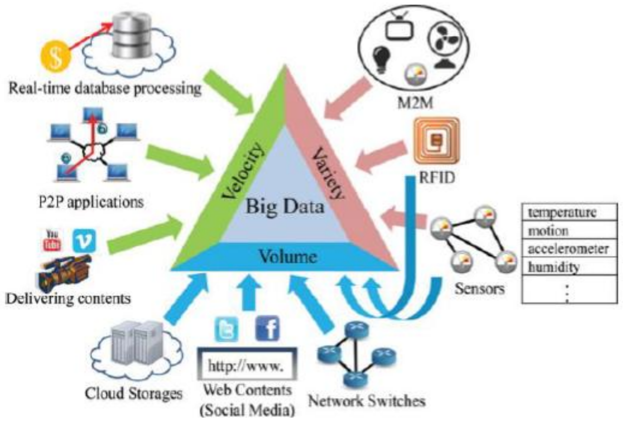


Figure. 1. Major trends of big data gathering.

The sensor nodes are resource controlled in term of energy, processor and memory and low range communication and bandwidth. Sensor nodes use their energy during receiving, transmitting and relaying the packets. So, designing routing algorithms that maximizes the life time until first battery expires is an important consideration. However, cluster heads will consume more energy than other sensor nodes. Due to continuous resource consumption problem failing cluster head and M-collector may arise. In order to tackle this problem some resource which reach sensor must be used which will make sensor network heterogeneous.

2. PROBLEM STATEMENT

Energy efficient big data gathering is a challenging task in the densely deployed wireless sensor network. Also cluster formation before data collection from sensors in the network is additional challenge. Previous researches addressed these challenges with mobile sink, which in turn raised the challenge of determining the sink nodes trajectory. Hence the problem to solve is to propose a method which addresses these challenges and helps in big data gathering in densely deployed wireless sensor networks.

3. RELATED WORK AND MOTIVATION

The analysis conducted by Sagiroglu et al. [3] gave special attention that big data and its analysis are at the centre of modern science and business. Sagiroglu et al. identified a number of sources of big data such as , emails, videos, online transactions, images, logs, audios, search queries, health information, social networking interactions, mobile phones and applications, scientific equipment, and the sensors.

R. C. Shah et. al. [4] suggested that Data MULEs follow the basic steps for all the mobile sink schemes. In this scheme, the sensor nodes are divided into many clusters. Then, a route for patrolling the cluster is decided. The work in [4] assumes a simple data collection scheme whereby the mobile sink node divides sensor nodes into grids not considering of the sensor nodes location, and patrols the grids by using random walk between the neighbouring grids. On the other hand, this type of clustering, which does not take into account the nodes location, may result in inefficient data gathering.

Heinzelman et. al.[5] suggested the Low-Energy Adaptive Clustering Hierarchy (LEACH) [5] which is one of the most famous clustering algorithms in WSNs which uses the static sink node. In this scheme, the clustering algorithm is executed by each of the sensor nodes. All the sensor nodes exchange information about their residual energies, and the nodes that have higher residual energies are given a higher probability to become the cluster head.

Daisuke Takaishi et. al. [6] suggested that The big data is very difficult to capture, form, store, manage, share, analyse, and visualize by the conventional database tools. Moreover, the main characteristics of big data are namely variety, volume, and velocity. The main intention of the work was to enable seamless exchange of feeds from large numbers of heterogeneous sensors. He proposed a solution where the clusters are formed and the sink will be mobile and which will collect all the collected data from cluster centroids. This solution reduced the energy constrain but additional challenge was to determine the sink node trajectory.

As proposed by S. Katti [7] the energy efficient data collection can be achieved by data compression technique especially in densely distributed wireless sensor networks. Here the compression technology will compress the data by shrinking the volume of the data to be transmitted. But this technology requires the nodes to have good computational power and also high volume of storage.

4. PROPOSED MODEL

The Fig. 2, shows the proposed model for big data gathering in densely deployed wireless sensor network. Once the sensor nodes are deployed, firstly the clustering formation is done using modified expectation maximization technique. Secondly, after the formation of cluster, one sensor is selected as cluster head. Cluster head will be responsible for data collection from all the sensor nodes in the cluster. To save the battery power of cluster heads, cluster heads are chosen rotationally among all nodes in the same cluster. In order to collect the data from all cluster heads a resource rich Mobile collector is added to each cluster. M-Collector traverses a fixed path from one cluster to another cluster to send collected data to static sink via intermediate M- Collector.

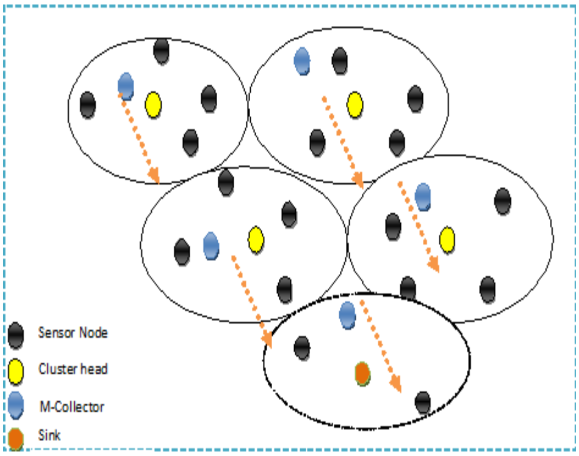


Figure 2 : Proposed Model –M-Collector Based Big-Data gathering.

4.1 Overview Of The Modified EM-Algorithm

EM algorithm assumes that nodes are distributed according to Gaussian mixture distribution,

$$p(x) = \sum_{k=1}^{k} \prod_k N(x|\boldsymbol{\mu} k, \boldsymbol{\Sigma} k) \qquad (1)$$

Where, k : indicates the total number of clusters.

Πk: indicate the total number of clusters and the mixing coefficient of the *k*th cluster.

At first EM algorithm calculates each nodes value of degree of dependence that is referred to as responsibility. The value of responsibility shows how much a node depends on a particular cluster. Following equation gives nth nodes value of degree of dependence on kth cluster:

$$\gamma nk = \frac{\prod_k N(x_n | \mu_k, \Sigma_k)}{\sum_{j=1}^{k} \pi_j N(x_n | \mu_j, \Sigma_j)} \qquad (2)$$

The responsibility takes values between 0 and 1. In the second step, the EM algorithm evaluates K weighted centre of gravity of a 2-dimensional location vector of nodes. This evaluation uses the responsibility value as weight of the nodes.

In the third step, the locations of the cluster centroids are changed to the weighted centres of gravity evaluated in the second step. The log likelihood can be calculated from EM algorithm as:

$$P = \ln p(X|\mu, \Sigma, \prod)$$

$$= \sum_{n=1}^{N} \ln\{\sum_{k=1}^{k} \pi_k N(x_n|\mu_k, \Sigma_k)\} \qquad (3)$$

4.2 Proposed clustering algorithm

```
Initialize cluster centroids μ to random locations.
Calculate cluster's parameters ∏ and ∑.
Calculate Dnk and P.
While |P – P^new| < ϵ do
 Select a group g which has the biggest value v_g.
            for k ϵ K_g do
            for n ϵ N_g do
Calculate nth node's responsibility value γ_nk.
end for
Calculate number of nodes belong to cluster, N_k.
Update the cluster's parameters ∏, μ and ∑ by using N_k.
end for
Evaluate the log likelihood P^new.
End while
Return cluster centroids, μ, covariance matrix, ∑ and the number of nodes that belongs to each cluster.
```

5. DATA GATHERING PROCEDURE USING THE PROPOSED CLUSTERING TECHNIQUE

5.1 Data gathering procedure using the proposed clustering technique:

After clustering, the mobile collectors patrols the particular cluster and collects the data from sensors and centroids of the cluster. Decreasing the delay generated by the mobile sink is the main purpose behind introducing the Mobile collector in wireless network. Mobile collectors collect the data from centroids and sensor nodes in the network by traversing a fixed length path. Directed diffusion [8] is one of the well known data collection method. Also One phase pull [9] where the mobile collector sends data request messages at the cluster centroids and when a sensor node receives a data request message from say cluster K then that re-broadcasts that data request message, then the centroid sends the data to the collector. As mobile collectors traverse fixed length path which reduces the complexity generated by mobility of collectors. Mobile collectors travel from one cluster to another cluster to transfer the collected data from one mobile collector to another to reach to the static sink.

Steps for gathering of the data:

1. Sensor nodes will forward data to the cluster head.
2. Mobile Collector belonging to each cluster will travel its fixed path, when there is connectivity between cluster head and M-collector, cluster head transfer data to the M-Collector.
3. When M-Collector gets connectivity with another mobile collector in the direction of sink, it transmits collected data to that M-Collector.
4. Likewise all M-Collector transfers data towards sink node.
5. Sink receives data from nearby M-collectors.

5.2 Computing the optimal number of clusters

To obtain the optimal number of clusters, we need to define objective function, W(K), which can be defined as the sum of energy consumption in one cycle of M-Collector patrol as follows.

$$W\,(K) = D_{Req}\,E_{Req}\,(K) + D_{Dat}\,E_{Dat}\,(K) \qquad (4)$$

Where,

E_{Req} (K): The sums of the square of transmission distance of data requests.

E_{Dat} (K): The sums of the square of transmission distance of data messages.

D_{Req}: The data size of data request messages.

D_{Dat}: The data size of data message.

*E*Dat (*K*) is evaluated according to the following equation:

$$\mathrm{EDat} = \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{h=1}^{H_{nk}} \gamma_{nk} \,.\, l_h^2 \qquad (5)$$

Hnk is the hop count from the nth node to the kth cluster centroid and lh is communication distance of each hop.The optimal number of clusters Kopt is defined by the following equation.

$$K_{opt} = \max\,(G, \arg{}_{K} \min\,(W(K))) \qquad (6)$$

G: Group.

We consider group of node that has Ng nodes and Kg cluster centroids. Data request message is sent from every cluster, and further each and every node re-broadcasts it one time. The total required energy to transmit data request message is formulated as follows:

$$E_{Req} = \sum_{g=1}^{G} K_g\, N_g R^2 \qquad (7)$$

R is the maximum transmission range of the sensor nodes.

If there is no imbalance of location of cluster centroids, the number of nodes that belongs to each cluster is the same.

$$\frac{K_g}{N_g} = \frac{K}{N} \qquad (8)$$

Here, if the number of nodes is larger than 1, the connectivity, *C*, can be approximated as follows:

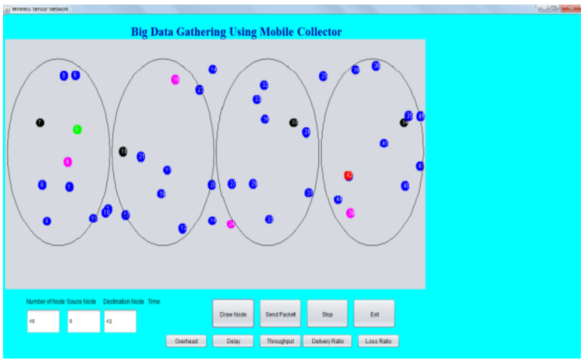
$$C = \frac{\sum_{g=1}^{G} N_g(N_g-1)}{N(N-1)} \div \frac{\sum_{g=1}^{G} N_g^2}{N^2} \qquad (9)$$

From all above equation we can calculate the energy required.

$$E_{Req} = K\,N\,R^2\,C \qquad (10)$$

Thus, it can be shown that the number of clusters has a major effect on connectivity. By calculating the required energy for data transmission as shown in equation (10), and data request transmission as shown in equation (7), the optimal number of clusters as shown in the above equation (6).

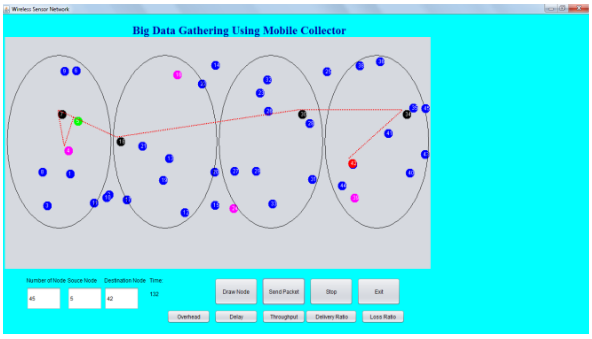
6. RESULT



Number of Node: 45
Souce Node: 5
Destination Node: 42
Time:

Figure 3: Snapshot 1.

The Fig. 3 shows the snapshot in which the user enters the no. of nodes, source node and destination node and then the user clicks draw node button which draws the nodes on the canvas. Then clustering is done.



Number of Node: 45
Souce Node: 5
Destination Node: 42
Time: 132

Figure 4: Snapshot 2.

The Fig. 4 shows the snapshot in which when the user clicks on send packet then the data packet is sent from source node to destination node via. the Mobile collectors in the intermediate clusters. The above snapshots show the results of the project.

7. RESULT ANALYSIS

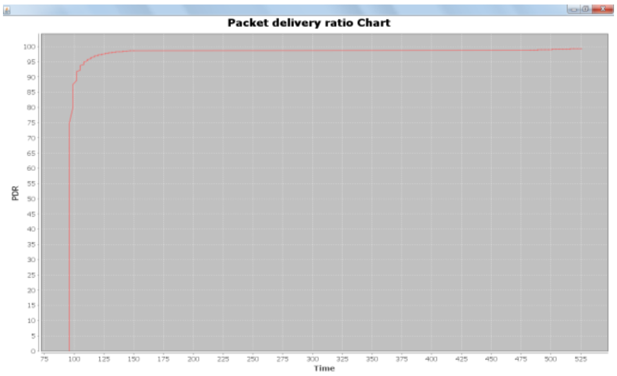


Figure 5 : Snapshot 3(Packet delivery ratio).

The above Fig. 5 shows the packet delivery ratio chart of the project. The packet delivery for the particular nodes shows that the delivery of the packets is almost maximum. The delivery ratio is calculated with respect to time on X-axis.

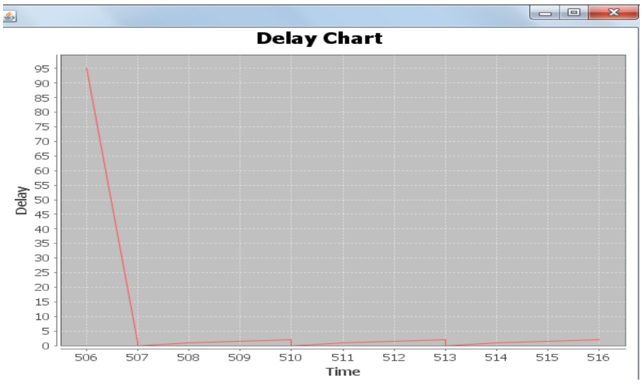


Figure 6: Snapshot 5(Packet delay)

The above Fig. 6 shows the delay generated during sending the packets from source to destination which is nearly zero.

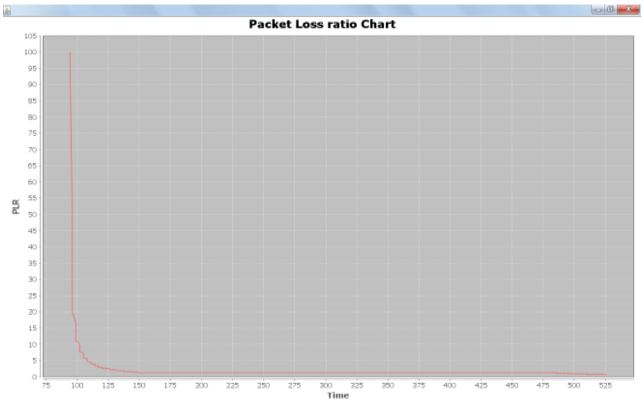


Figure 7: Snapshot 4(Packet loss ratio).

The above Fig. 7 shows the packet loss ratio chart where the loss of packets is almost zero.

8. CONCLUSION

In this paper, we carried out the survey to investigate the challenging and demanding issues pertaining to collection of the Big Data generated by densely deployed wireless sensor networks. We also investigated the issue related to the mobile sink nodes trajectory used in the mobile sink scheme and the cluster formation. To address these challenges we proposed and designed a new scheme where we used Mobile collector based data gathering with network clustering based on improved Expectation

Maximization technique. Mobile collectors traverse a fixed path to collect data from cluster head and sensors in the clusters. Collected data is transferred amongst mobile collectors to reach to the static sink node.

ACKNOWLEDGEMENT

I wish to thank N D.Kale Department of Computer Engineering for his guidance and timely help given to me in completion of my paper.

REFERENCES

1. Daisuke Takaishi , Hiroki Nishiyama Towards Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks DOI 10.1109/ TETC.2014.2318177, IEEE Transactions on Emerging Topics in Computing.
2. Bisio and M. Marchese, Efficient satellite-based sensor networks for information retrieval, IEEE Systems Journal, vol. 2, no. 4, pp. 464475, Dec 2008.
3. S. Sagiroglu and D. Sinanc, Big data: A review, in International Conference on Collaboration Technologies and Systems (CTS), 2013.
4. R. C. Shah, S. Roy, S. Jain, and W. Brunette, Data MULEs: modeling and analysis of a three-tier architecture for sensor networks, Ad Hoc Networks, vol. 1, no. 2-3, pp. 215 - 233, 2003.
5. W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, Balakrishnan, Energy efficient communication protocol for wireless microsensor networks, in Annual Hawaii International Conference on sytem Sciences, vol. 2, Jan. 2000.
6. Daisuke Takaishi , Hiroki Nishiyama Towards Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks DOI 10.1109/TETC.2014. 2318177, IEEE Transactions on Emerging Topics in Computing, October 2014.
7. S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, XORs in the air: Practical wireless network coding, IEEE/ACM Transactions on Networking, vol. 16, no. 3, pp. 497510, Jun. 2008.
8. C. Intanagonwiwat, R. Govindan, and D. Estrin, Directed Diffusion: a scalable and robust communication paradigm for sensor networks, in MobiCom00 Proceedings of the 6th annual international conference on Mobile computing and networking, 2000.
9. M. Chen, T. Kwon, and Y. Choi, Energy-efficient differentiated directed diffusion (eddd) in wireless sensor networks, Computer Communications, vol. 29, no. 2, pp. 231245, 2006.

.