

Energy Efficient Wireless Sensor Network For Big Data Management

#1 Deepali R.Tambekar

¹diptambekar@gmail.com

#1 Computer Engineering Department, PVPIT, Bavdhan, Pune
Savitribai Phule University, India



ABSTRACT

With tremendous growth of selective information and communication technology (ICT) Radio set sensor engagement has major contribution in big information gather. Even if data generated by individual sensor is not significant, the overall data generated by all sensors in the electronic network is contributed for generation of significant component of big data. So free energy efficient big data gathering is challenging task in the densely deployed wireless sensor network. Also clustering formation before data collection from sensing element in the network is additional challenge. Recent research addressed these challenges with mobile sinkhole, which in turn wage increase the challenge of determining the sink guest 's 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 route to collect data from cluster centroids and sensors in the clusters. Collected data is transferred amongst M-collectors to reach to the static sink node. Also we derive optimal no of clusters to minimize the energy consumption..

Keywords— Big data, Wireless Sensor Networks (WSNs), clustering, optimization, data aggregation, energy efficiency, EM algorithm.

ARTICLE INFO

Article History

Received : 6th April, 2015

Received in revised form :

8th April, 2015

Accepted : 13th April, 2015

Published online :

14th April 2015

I. INTRODUCTION

Recent years, with better foreground of practical application, inquiry study on wireless detector meshing (WSNs) has attracted more and more attention. In general, a WSN is made up of low-cost and low-energy sensor knob s that can communicate with each other by wireless connexion. Since sensor hubs are regularly conveyed in remote control or distant situations, reviving sensor "threatening behavior is typically unthinkable. In the event that the battery of one sensor hub runs out, the hub will dice. Hence, how to spare sensor vitality and expand system lifetime is a principle test of WSNs. Existing exploration show that the correspondence expense is the fundamental character of sensor hubs' vitality utilization. Furthermore, a large portion of correspondence expense is delivered during the time spent get together recognition data from sensor hubs. The employment experienced in the later past was of the more battery intensity utilization as real process builds;

require more productive information accumulation and gathering strategy with right choice making limit.

Therefore, the literature propose the efficient and effective architecture and mechanism of energy efficient techniques, M-Mobile River collector based data gathering with network clustering based on improved Expectation maximization technique. Mobile gatherer crossbeam a fixed path to collect data from cluster centroids and sensors in the bunch. Collected data is transferred amongst M-collectors to reach to the atmospheric static cesspool node. Also in the literature, derive optimal no of clusters to minimize the energy consumption.

2.5quintillion byte of data is created every day — so much that 90 % of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, Emily Post to social media sites, digital pictures and videos, purchase transaction criminal record , and electric cell phone GPS signals to name a few. This data is big data. Big data spans three dimensions: Volume, Velocity and Variety.

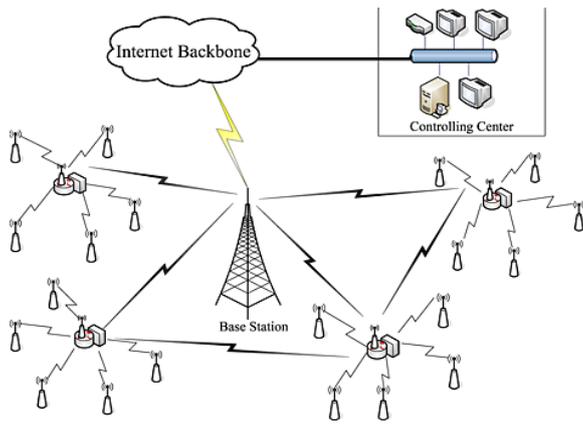


Fig 1 –The model for data gathering in wireless sensor network.[4]

Fig. 1 shows the model for big data gathering in wireless sensor network. Once the sensors are deployed, firstly the clustering formation is done using modified expectation maximization technique. After formation of cluster, one sensor is elected as cluster head. Cluster head will be responsible for data collection from all nodes in the cluster. In order to save the battery power of cluster heads, cluster heads are chosen rotationally from all nodes in the same cluster. In order to collect data from all cluster heads a resource rich Mobile collector is added to each cluster. M-Collector traverse a fixed path from one cluster to another cluster to send collected data to static sink via intermediate M- Collector.

In this literature energy minimized clustering calculation by utilizing the Expectation-Maximization (EM) algorithm for 2-dimensional Gaussian mixture distribution. The recommendation intends to minimize the total of square of remote correspondence separation since the vitality utilization is relative to the square of the remote correspondence separation.

To begin with spotlight on the "information appeal flooding issue" to choose the ideal number of clusters this refers to the energy inefficiency.

An excessive number of clusters can result in performance degradation and therefore, an adequate method for deriving the optimal number of clusters

A regular application in sensor system is to collect the sensed information at a far off base station. Every sensor hub has force control and the capacity to transmit information to whatever other sensor hub or direct to the BS. In each round of this information gathering application, all information from everything hubs need to be gathered and transmitted to the BS, where the end client can get to the information.

II. RELATED WORKS

Daisuke Takaishi et. al. [1]. "The data gathering from densely distributed sensor networks is, difficult using traditional database tools. To address this challenge, utilize the sink node's mobility to facilitate the data gathering which helps to reduce the energy consumption of the sensor nodes. The main objective of the work was to enable seamless exchange of feeds from large numbers of heterogeneous sensors."

In sensor network minimizing of data transmission is difficult using distributed clustering algorithm. If the

WSN is divided into sub networks, a node may not be able to have all the information about all the nodes, thus Optimization cannot be achieved. The centralized clustering algorithm, which is supervised by a super node, is much suitable for the mobile sink scheme.

R.C. Shah et. Al. [2] "The basic steps of all mobile sink schemes are followed by data MULEs. First it divides sensor nodes into clusters. Then it decides the route through a specified area in order to maintain order and security of each cluster. The work in [2] assumes that a simple data collection scheme whereby the mobile sink node divides sensor nodes into grids regardless of the sensor nodes' location, and patrols the grids by using random walk between the neighbouring grids. However, this type of clustering, which does not take into account the nodes' location, may result in inefficient data gathering."

Heinzelman et. al. [3] "Low-Energy Adaptive Clustering Hierarchy (LEACH) is one of the most famous clustering algorithms in WSNs which uses the static sink node. In this type, the clustering algorithm is executed by each of the sensor nodes. 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. So by periodical re-clustering energy consumption of each node will be equal. LEACH assumes that each node can communicate with each other but when the WSN's deployed in wide areas are not able to use this algorithm. So there will be a limitation of node's communication range."

III. WIRELESS SENSOR NETWORK FOR BIG DATA

With the firehouse of information enabled by Facebook, Twitter, location-based services, and other forms of social media. Wireless sensors are (almost) everywhere, fueling a boom in machine-to-machine communications and generating a ton of data. They're in vending machines, parking meters, home security systems, and even healthcare devices for the elderly. They are wireless sensors, a key component of the burgeoning machine-to-machine (M2M) industry where devices use wired and wireless connections to communicate with each other. While the management of large numbers of wireless sensors is certainly a Big Data issue, Horn says most M2M devices generate relatively small quantities of information.

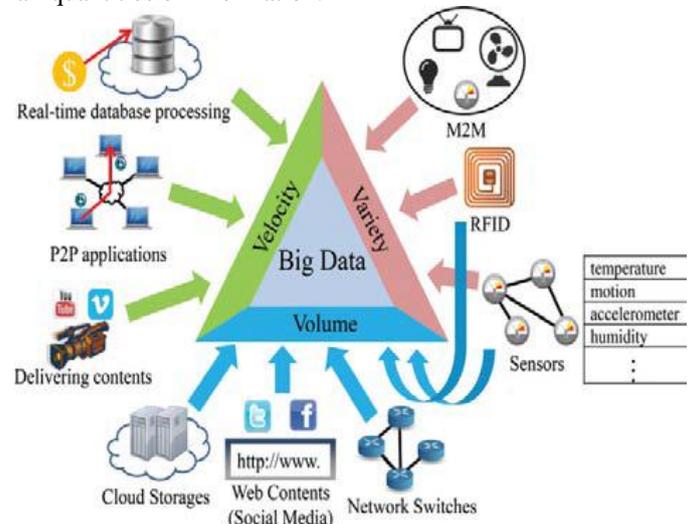


Fig 2 : Major Trends of Big Data Gathering[1]

IV. EXISTING SYSTEM

Overview of EM Algorithm

EM algorithm assumes that nodes are distributed as per the Gaussian mixture distribution,

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

Where,

K : Total number of clusters

π_k : Indicate the total number of clusters and mixing the coefficient of k th cluster.

$N(x|\mu, \Sigma) =$

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

Where,

X : Position vector of all nodes

μ_k : Position vector of centroids of cluster k .

Σ_k : 2x2 covariance matrix of the k th cluster.

Equation 1 calculates each node's value of degree of dependence that is referred to as responsibility. The responsibility shows how much a node depends on a cluster. The n th node's value of degree of dependence on k th cluster is given by following equation:

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

Because of its definition, the responsibility takes the values between 0 and 1.

At second step, EM algorithm evaluates K weighted center of gravity of 2-dimensional location vector of nodes.

At the third step, location of cluster centroids are changed to the weighted centers of gravity evaluated in the second step. EM algorithm evaluates the value of the log likelihood as shown below.

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

EM algorithm repeats all the steps until the value of log likelihood converge. The EM algorithm always terminates when this value of log likelihood is monotonously decreasing, and because the EM algorithm repeatedly updates cluster centroids' position vector, μ_k , and nodes' responsibility to k th cluster, γ_{nk} , the sum of square of distances of each node to cluster gradually decreases and finally becomes optimal.

Clustering Method

WSNs, which are having high volume and high variety of data, we need to consider "groups", which refer to sets of nodes that can communicate with each other. To collect data from all nodes, the number of clusters must be set to more

than the number of groups. Our proposed clustering method can be summarized in below clustering Algorithm 1.

Algorithm 1 Clustering Algorithm

```

Initialize cluster centroids,  $\mu$ , to random locations.
Calculate clusters' parameters,  $\pi$  and  $\Sigma$ .
Calculate  $D_{nk}$  and  $P$ .
while  $|P - P^{new}| < \epsilon$  do
  Select a group  $g$  which has the biggest value  $v_g$ .
  for  $k \in K_g$  do
    for  $n \in N_g$  do
      Calculate  $n$ th node's responsibility value,  $\gamma_{nk}$ .
    end for
    Calculate number of nodes belong to cluster,  $N_k$ .
    Update the clusters' parameters,  $\pi$ ,  $\mu$  and  $\Sigma$ , by using  $N_k$ .
  end for
  Evaluate the log likelihood  $P^{new}$ .
end while
Return cluster centroids,  $\mu$ , covariance matrix,  $\Sigma$ , and the number of nodes that belongs to each cluster.

```

As mentioned in the above algorithm, first the mobile sink sets the cluster centroids, μ , to random locations. Using a random position vector of cluster centroids, the communication distances of each of the node in the cluster to the respective cluster centroids, D_{nk} is calculated. The mixing coefficient Π and the covariance matrix Σ is calculated. After completion of the cluster initialization phase, the proposed method selects a group g that has the largest value of proportion of number of nodes to the number of clusters in a group g , shown as:

$$v_g = \frac{K_g}{N_g} \quad (5)$$

Group that has the highest value of v_g , our method picks up all nodes that belong to group g and updates this node's responsibility value γ_{nk} . The responsibility value reflects how much node n belongs to cluster k . With the help of values of updated responsibility γ_{nk} , cluster centroids μ , the covariance matrix Σ are re-calculated and the number of nodes which belongs to k th cluster is calculated as shown in the following equation,

$$N_k = \sum_{x_n \in X} \gamma_{nk} \quad (6)$$

These calculations are executed again and again until the difference between the newly calculated P and previously calculated P becomes smaller than small number ϵ .

Creating Optimal Number of Clusters

To find out the optimal number of clusters, first we have to find out objective function $W(K)$ where the objective function is the sum of required energy of data and data request message transmission. So objective function can be defined as the sum of energy consumption in one cycle of mobile sink patrol as follows.

V. CONCLUSION

The mobile data-gathering scheme for large-scale sensor networks increases the performance of the network. The systematic study of the challenging and required issues related to collection of the Big Data generated by densely deployed wireless sensor networks. We also investigated the issue related to the mobile sink node's trajectory used in the mobile sink scheme and the cluster formation. To address these challenges we proposed 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 centroids and sensors in the clusters. Collected data is transferred amongst M-collectors to reach to the static sink node. Also we derive optimal no of clusters to minimize the energy consumption.

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] R. C. Shah, S. Roy, S. Jain, and W. Brunette, "Data MULEs: MULEs: modeling and analysis of a three-tier architecture for sensor networks," *Ad Hoc Networks*, vol. 1, no. 2-3, pp. 215 -233.2003.
- [3] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, Balakrishnan, "Energyefficient communication protocol for wireless microsensor networks," in *Annual Hawaii International Conference on system Sciences*, vol. 2, Jan. 2000.
- [4] <http://onlinelibrary.wiley.com/doi/10.1002/wcm.v15.1/issuetoc>.
- [5] <http://www.informationweek.com/mobile/when-wireless-sensor-meet-big-data/d/d-id/1105913>.

$$W(K) = D_{Req} E_{Req}(K) + D_{Dat} E_{Dat}(K) \quad (7)$$

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:

$$E_{Dat} = \sum_{n=1}^N \sum_{k=1}^K \sum_{h=1}^{H_{nk}} \gamma_{nk} \cdot l_h^2 \quad (8)$$

H_{nk} is the hop count from the n th node to the k th cluster centroid and l_h is communication distance of each hop The optimal number of clusters K_{opt} is defined by the following equation.

$$K_{opt} = \max(G, \arg \min_K (W(K))) \quad (9)$$

G : Number of Groups

Here consider group of node that has N_g nodes and K_g cluster centroids. Data request message is sent from every cluster and every node re-broadcasts it one time. The total required energy to transmit data request message is formulated as follows:

$$E_{Req} = \sum_{g=1}^G K_g N_g R^2 \quad (10)$$

R : Maximum transmission range of sensor nodes.

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

$$\frac{K_g}{N_g} = \frac{K}{N} \quad (11)$$

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

$$C = \frac{\sum_{g=1}^G N_g (N_g - 1)}{N(N-1)} \div \frac{\sum_{g=1}^G N_g^2}{N^2} \quad (12)$$

Using all of the above equations we can calculate the energy required.

$$E_{Req} = K N R^2 C \quad (13)$$

Thus, it can be seen that the number of clusters has a significant effect on connectivity. By calculating the required energy for data transmission as in (13), and data request transmission as in (8), the optimal number of clusters, (9), can be calculated by using (8) and (9)