ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

An Efficient Data Aggregation Scheme and Cluster Optimization in Wireless Sensor Networks

R.Mehala[1], Dr.A.Balamurugan[2]

PG Scholar, Department of CSE, Sri Krishna College of Technology, Coimbatore, Tamilnadu, India [1]

Professor, Department of CSE, Sri Krishna College of Technology, Coimbatore, Tamilnadu, India[2]

ABSTRACT: Wireless Sensor Networks is a collection of sensor nodes with sensing, computation and communication capabilities. The sensor nodes are capable of receiving data from the monitoring environment. The sensed data should be gathered efficiently from the sensor nodes and is routed to base station through fixed path. The QoS metrics such as Bandwidth, Latency, Data accuracy, Throughput can also affect the performance of the network lifetime. The effective solution to address this challenge is to utilize the sink node's mobility to facilitate the data gathering .This technique can reduce energy consumption of the sensor nodes. The usage of mobile sink presents additional challenges such as determining the sink node's trajectory and cluster formation prior to data collection. Network lifetime is an important issue in an energy constrained sensor networks. The most widely used Expectation maximization (EM) algorithm is used for clustering process and to compute the optimal number of clusters in the network. One of the main problem is the implementation of flooding diffusion which is used to forward interest and discover the routing map that reduces network lifetime and data latency through high energy consumption by the use of AODV(Ad hoc on demand distance vector)routing .

KEYWORDS: Expectation Maximization, AODV, Gaussian mixture model

I. INTRODUCTION

With the recent technological advances in wireless communications, processor, memory, radio, highly integrated digital electronics, and micro electro mechanical systems (MEMS), it becomes possible to significantly develop tiny and small size, low power, and low cost multi-functional sensor nodes. Figure 1 describes that Big data comprises high volume, high velocity, and high variety of information assets [6][7], which is to be gathered, stored and processed by using the available technologies Wireless sensor networks (WSNs) have been used for numerous applications and deployed in many areas like monitoring the conditions of a bridge after its construction, monitoring the historic buildings and Surveillance in Defence organizations including military surveillance, facility monitoring and environment monitoring[1,2,3]. Typically WSNs have large number of sensor nodes with the ability to communicate among themselves and also to an external sink or a base-station. In this paper, we are considering base station as static one where the mobile sink moves across each clusters to collect the sensed data. The sensors coordinate among themselves to form a communication network such as a single multi-hop network or a hierarchical organization with several clusters and cluster heads. The sensors periodically sense the data, process it and transmit it to the sink. These sensor nodes can communicate within clusters and collaborate to accomplish a specific task. The sensor nodes are energy constrained, therefore it is inefficient for all the sensor nodes to transmit the sensed data directly to the mobile sink. Data sensed by the sensor nodes which are nearer to each other is redundant. Hence, there is a need for a method which combines the data from different sensor nodes and reduces the number of packets to be transmitted to the sink node. Wireless sensor nodes require less power for processing the data than compared to transmitting data. This results in the saving of energy and increase in the network lifetime. Data aggregation usually involves the fusion of data from multiple sensors at intermediate nodes and transmission of the aggregated data to the mobile sink. Since sensor nodes may generate contain redundancy of data, packets from multiple nodes can be aggregated so that the number of transmissions is reduced. This can be accomplished by data aggregation techniques. The cluster head attempts to collect

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

the critical data from the neighboring and intermediate sensor nodes and make it available to the mobile sink in an energy efficient manner with minimum data latency.

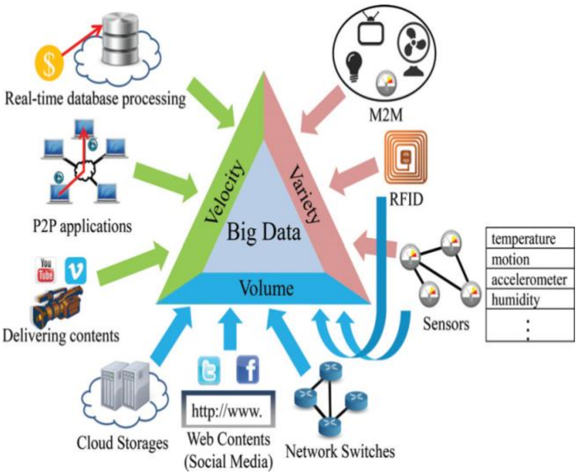

Fig 1. Various data in Sensor networks

The remainder of the paper is organized as follows.Section II reviews some related works and its problems.Section III presents the clustering based data aggregation and routing algorithm in WSN. Section IV presents the proposed clustering method based on the EM technique.Performance evaluation is presented in Section V.Finally Section VI concludes the paper.

II. RELATED WORK

Kenji Miyao et al. [5] proposed that broadcasting in wireless sensor networks is a costly operation and thus topology control achieve efficient broadcasting with low interference and low energy consumption. By topology control, each node minimizes its transmission power by maintaining network connectivity in a localized manner.Local Tree-based Reliable Topology (LTRT) algorithm is motivated by (Local Minimum Spanning Tree) LMST and the Tree-based Reliable Topology (TRT). LTRT is a localized version of TRT and combined with LMST to guarantee k-edge connectivity of the resulting topology. Improving reliability leads to an increase of energy consumption and can maintain the extent of such an increase within a tolerable limit. LTRT can achieve nearly good performance at a much lower computational cost. The topology construction and link deletion phases are repeated until the topology achieves k-edge connectivity.

Liang He et al.[11] suggested that Mobility-assisted data collection in sensors creates a new dimension to reduce and balance the energy consumption for sensor nodes .The on-demand scenario where data collection requests arrive at the mobile element progressively is investigated and the data aggregation process is modelled as an M/G/1/c-NJN queuing system with an intuitive service discipline of nearest-job-next (NJN).Data collection in sensor networks relies on the wireless communications between sensor nodes and the sink, which may consume the limited energy supply of sensor nodes due to the super- linear path loss exponents. The sensor nodes near the sink tend to deplete their energy much faster than other nodes due to the data aggregation towards the sink, which imposes heavier volume of data to forward and leads to a unbalanced energy consumption in the entire network.The mobile elements travel around in the sensing field to collect data from sensor nodes with wireless communication techniques.

Lindsey et al. [13] presented a chain based data aggregation protocol called power efficient data gathering Protocol for sensor information systems (PEGASIS). In PEGASIS, nodes are organized into a linear chain for data aggregation. In cluster-based sensor networks, nodes transmit their data to the cluster head where data aggregation is performed. When the cluster head is away from the sensors, they might spend more energy in communication. The improvements in energy efficiency can be obtained if sensors transmit only to close neighbors. The main idea behind chain based data aggregation is that each sensor transmits only to its closest neighbor. The nodes can form a chain by employing a

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

greedy algorithm or the sink can determine the chain in a centralized manner. Greedy chain formation of nodes assumes that all nodes have global knowledge of the network. The nodes away from the sink initiates chain formation and during each step, the closest neighbor of a node is selected as its successor in the chain. In each data gathering round, a node receives data from one of its neighbors, stores the data with its own and transmits the fused data to its other neighbor along the chain. The leader node which is similar to cluster head transmits the aggregated data to the sink. *Wendi Rabiner Heinzelman et al*.[12] approached that LEACH performs local data fusion to "compress" the amount of data being sent from the clusters to the sink which reduces energy dissipation and enhancing system lifetime. Sensors elect themselves to be local cluster-heads at any given time with a threshold probability. These cluster head nodes broadcast their status to the other sensors in the network. Each sensor node determines a group,it wants to belong by choosing the cluster-head that requires the minimum communication energy. Once all the nodes are grouped into clusters, each cluster-head creates a timestamp for the nodes in its cluster. This allows the radio components for each non-cluster-head node to be turned off at all times except during its transmit time, thus minimizing the energy wastage in the individual sensors. Once the cluster-head contains all the data from the nodes , the cluster-head node aggregates the data and then transmits the compressed data to the base station.

III. CLUSTERING BASED DATA AGGREGATION AND ROUTING IN WSN

In this section, the outline of clustering and data aggregation problem and network model in WSN is considered. The overview of EM algorithm for clustering and optimal number of clusters is derived using hop count and transmission distance. Based on EM algorithm, clustering method and the procedure to gather data in WSN is proposed.

A.Clustering problem

When considering data gathering scheme in WSN using mobile sink, the biggest challenge is used to reduce the energy consumed by the sensor nodes. The energy depletion can be reduced by forming clusters of sensor nodes and then to perform data gathering. When forming clusters it should be ensure that no same node should be present in one or more clusters. EM algorithm is powerful and well-known method to solve the clustering problem by iteratively calculate the sum of square of distance between every node and cluster centroid, we adopt EM algorithm over the 2-dimensional Gaussian mixture distribution.

B.Data Aggregation problem

Data aggregation is a process of aggregating the sensor data using aggregation approaches. These approaches uses the sensor data from the sensor nodes and then aggregates the data by using some aggregation algorithms such as centralized approaches like LEACH(low energy adaptive clustering hierarchy),TAG,DirectedDiffusion[8][9] .The data aggregated is transfer to the mobile sink node by selecting the efficient path. The most popular data aggregation algorithms are cluster-based data aggregation algorithms, in which the nodes are grouped into clusters and each cluster consists of a cluster head (CH) and some members. Each member transmits the sensed data to its CH, then each CH aggregates the collected data and transmits the data to BS. The cluster- based WSNs have an inherent problem of unbalanced energy dissipation. Some nodes drain their energy faster than others and result in earlier failure of network.

C.Data request flooding problem

The sink node sends data request message to invoke data transmission from sensor nodes when it arrives at the cluster centroids. The nodes that receive data request message send the data to the sink node and broadcast data request message to their neighboring nodes. The data request messages are broadcasted repeatedly until all nodes in the same group receives the message. Some nodes will receive data request message more than one time, but they send data only at the first time of receiving the message. The broadcasts of data request message cause high energy consumption because the network will be flooded with redundant wireless communication. Thus, reducing data request message is also important for mobile sink in data gathering scheme.

D.Overview of Expectation Maximization algorithm

The EM algorithm is a classical clustering algorithm, where nodes are distributed according to Gaussian mixture distribution [8],

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \qquad (1)$$

where K and π_k indicate the total number of clusters and the mixing co-efficient of the k^{th} cluster. The EM algorithm seeks to find the maximum likelihood estimation of the marginal likelihood by iteratively applying the E-step and M-step. Every nodes do not belong to any cluster and centroids of clusters which represented by the cross are randomly decided. EM algorithm calculates each node's degree of dependence that is referred to as responsibility. The responsibility value shows how much a node depends on a cluster. This responsibility value is calculated by the nodes location and centroid location of cluster. Normally, each node depends only on one cluster. However, it is possible for nodes to depend on more than one cluster so that those nodes will not focus their energy in a single cluster. After that EM algorithm calculates the responsibility value, it calculate the centroid of the cluster by using node's location. The centroid of each cluster is calculated to minimize the distance between each nodes which belong to the cluster and the centroid.

IV. PROPOSED CLUSTER-BASED AGGREGATION SYSTEM

A .Considered network model

The system consists of large number of sensor nodes.The sensor nodes are of equal energy at the initial phase.There are two sinks in the proposed system one is the static sink node and other is mobile sink node[4].The static sink node is placed in the center of the network and in this system there are two mobile sinks. These mobile sinks travels around the cluster centroid to gather data. The cluster centroid gets data request message from the sensor nodes and forward the message to one of the mobile sink. When the mobile sink gets message from more than one cluster centroid it forwards to other mobile sink.

B. Clustering phase

The sensor nodes are deployed in a physical environment for monitoring and sensing the data. A large number of sensor nodes of equal energy and mobile sink of higher energy is deployed in the sensing area. At the initial stage, the nodes are scattered and cluster centroid is placed at random location by using EM algorithm[10].Then the clusters of nodes are formed by calculating the shortest distance between cluster centroid and senor nodes. The responsibility value for each node is measured. The responsibility value refers to the degree of dependence of sensor nodes. Some sensor nodes have lower degree and some nodes have higher degree value.The next step is to calculate the centroid location for each cluster. The centroid is calculated based on location information of cluster members. The centroid location of each cluster is the data collection points in the network.

The EM algorithm is a clustering algorithm, which determines that nodes are distributed according to Gaussian mixture distribution,

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right\}$$

In (Eq.1) where x is the position vectors of all nodes K indicates the total number of clusters, π_k indicates the mixing coefficient of the k^{th} cluster, N represents the numbers of clusters respectively. At the initial step, the EM algorithm calculates each node's value of degree of dependence that is referred to as responsibility. The responsibility value calculated 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.

$$\text{Distance, } D_{nk} = \sqrt{(x2-x1)^2 + (y2-y1)^2} \qquad (2)$$

$$\gamma_{nk} = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}. \qquad (3)$$

From (Eq.2) & (Eq.3) the clusters of nodes are formed by calculating the shortest distance between cluster centroid and senor nodes. The distance is calculated by taking one-hop neighbor from each sensor node to cluster centroids. The next step is to calculate the centroid location for each cluster. The cluster centroid is calculated based

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

on location information of cluster members.Nodes are grouped based on fixed transmission distance from the centroid. After calculating the distance from each node, the centroid location is calculated and grouped into the cluster.

C.Routing and Data gathering using mobile sink

Mobile sink patrols around the cluster members using AODV(Ad hoc On-demand Distance Vector) Routing algorithm to indicate its occurence.When the mobile sink comes to the sensor node ,it sends a message to the cluster centroid with its unique identification number.After patrolling around all the sensor nodes the mobile sink moves to the cluster centroid to gather information.The sensor nodes sends the data to the mobile sink and the sensed data is uploaded to base station.

D.Computing the optimal number of clusters

To decide the optimal number of clusters, objective function is defined. The objective function is defined as the sum of required energy of data and data request message transmissions. Thus the objective function W(K), can be defined as the sum of energy consumption in one cycle of mobile sink patrol as follows.

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

(Eq.4) indicates $E_{Req}(K)$ and $E_{Dat}(K)$ are the sums of the square of transmission distance of data requests and data messages, respectively. D_{Req} and D_{Dat} indicate the data size and data request messages respectively.

V. SIMULATION RESULTS

Simulation process and results analysis are obtained by setting the topology,configuration of nodes properties, and also properties of MAC layer,protocol type, channel type, simulation time, modulation type and transmission way of sensor networks. Performance of the system can be evaluated by QOS metrics like bandwidth, throughput, number of nodes and data accuracy.Table 1 shows the simulation parameters

Table 1 Simulation Scenario Parameters

Parameters	Values
Area of Simulation	(1250X900)
Nodes number	38
Types of Routing protocol	AODV
Type of the MAC	802.11
Internet protocol type	UDP
Antenna Model	Omni directional
Mac package	50
Transmission speed	1.5Mbps
Energy	200 Joules

It is important to measure the performance of the network by considering the clustering algorithm and evaluating its parameters with other algorithms. In this section, the results are obtained by using NS2 simulator.The results of the proposed system is shown in Figure 2 shows the throughput measured from the system. Figure 3 shows the redundancy of packets which includes data packets. Figure 4 indicates the energy consumed by the individual nodes. Figure 5 indicates the average energy consumed by the nodes in the network

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

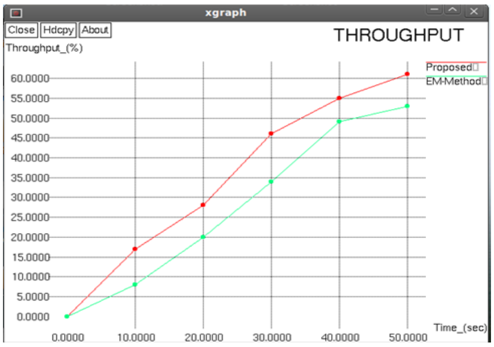

Fig 2. Throughput

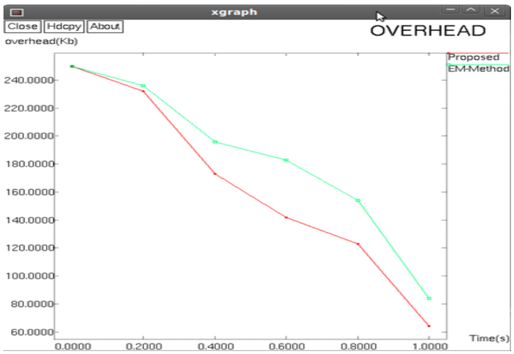

Fig 3. Overhead

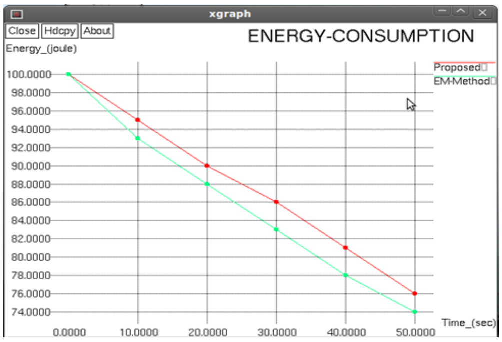

Fig 4. Energy Consumption

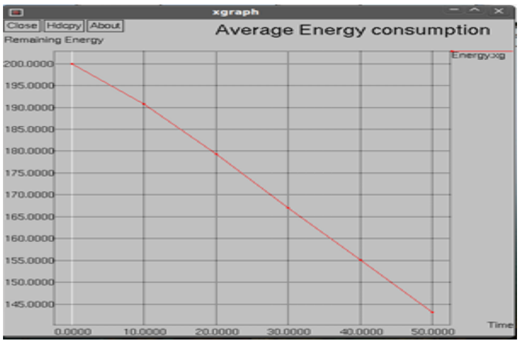

Fig 5. Average Energy Consumption

The network energy consumption is minimum with the mobile sink node. Comparing to the stationary sink, the mobile sink can balance the load and increase the network lifetime. Maintaining network reliability requires high energy, but EM algorithm can minimize the increase of energy consumption by 20% to 30%. This improvement occurs from considering network connectivity and communication distance. Thus the throughput, energy consumption of nodes and overhead of messages are calculated.

VI. Conclusion and Future Work

This paper presented Clustering-based Data aggregation algorithm for Wireless Sensor Networks by using two mobile sinks, which improves network lifetime and reduces the transmission distance between each nodes.By using this proposed method, the sensed data are collected efficiently which minimizes data latency. Wireless sensor networks are used for application oriented tasks. The common methods for routing and data gathering can be used to consume energy in an efficient manner. Some links in the network may get disconnected due to network connectivity.The mobile sink which sends the data in that particular link may suffer data latency.This problem can be avoided by routing the mobile sink in a fixed path of the network.In future,security can be provided to detect the malicious sensor node in the network.

References

1. D. Culler, D. Estrin, and M. Srivastava, "Overview of Sensor Networks," IEEE Computer, August 2004.
2. N. Xu, S. Rangwala, K. Chintalapudi, D. Ganesan, A. Broad, R. Govindan, and D. Estrin, "A Wireless Sensor Network for Structural Monitoring," Proceedings of the ACM Conference on Embedded Networked Sensor Systems, Baltimore, MD, November 2004.
3. A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson, "Wireless Sensor Networks for Habitat Monitoring," WSNA'02, Atlanta,Georgia, September 2002.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

4. Xiaobing Wu , Nanjing Univ., Nanjing Guihai Chen ,"Dual-Sink: Using Mobile and Static Sinks for Lifetime Improvement in Wireless Sensor Networks ",Computer Communications and Networks, ICCCN 2007, pp 1297 – 1302.,2007.
5. K. Miyao, H. Nakayama, N. Ansari, and N. Kato,' LTRT: An effcient and reliable topology control algorithm for ad-hoc networks',*IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 6050 to 6058,2009.
6. 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.[Online].Available:http://cra.org/ccc/docs/init/bigdata whitepaper.pdf.
7. S. Sagiroglu and D. Sinanc, "Big data: A review," in International Conference on Collaboration Technologies and Systems (CTS), 2013
8. Nandini. S. Patil, Prof. P. R. Patil, "Data Aggregation in Wireless Sensor Network", IEEE International Conference on Computational Intelligence and Computing Research, 2010.
9. Xuxun Liu, "A Survey on Clustering Routing Protocols in Wireless Sensor Networks,"Sensors 2012,doi:10.3390/s120811113,9 August 2012
10. Daisuke Takaishi, Hiroki Nishiyama, Nei Kato, and Ryu Miura, "On the Effect of Data Request Message Flooding in Dense Wireless Sensor Networks with a Mobile Sink," IEEE 78th Vehicular Technology Conference (VTC 2013 Fall), Las Vegas, Nevada, USA, Sep. 2013.
11. L. He, Z. Yang, J. Pan, L. Cai, J. Xu, and Y. Gu ,' Evaluating service disciplines for on-demand mobile data collection in sensor networks', vol. PP, no. 99,2009.