Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 7, July 2013, pg.240 – 246

RESEARCH ARTICLE

CLUSTERING APPROACH FOR FAST ENERGY EFFICIENT DATA COLLECTION IN WIRELESS SENSOR NETWORK

Mr. Sandeep.C[1], Mrs. Manimozhi I[2], Dr. Jitendranath Mungara[3]
[1]Student, Department of CSE, CMRIT, Bangalore, India
[2]Associate Professor, Department of CSE, CMRIT, Bangalore, India
[3]Dean and Professor, Department of CSE and ISE, CMRIT, Bangalore, India

[1] sandeep281989@gmail.com; [2] srimanisen@gmail.com; [3] jmungara@yahoo.com

Abstract— Most of the wireless sensor networks are deployed for data forwarding to the sink node. The main problem in wireless sensor network is the lifetime of the network and the data collection rate. The sensor node resources are limited that depends on the battery. So it is important to utilize its power efficiently. The data collection rate of a wireless sensor network is important for safety mission control applications where sensor nodes are deployed to detect oil/gas leak or structural damage. In this paper, we study efficient data collection in wireless sensor networks. We propose a method on cluster based routing and adaptive two-level scheduling for the data collection in wireless sensor network. We first from the clusters based on the node density then assign BFS timeslot scheduling within cluster and multichannel assignment between the sink nodes and clusterheads. Our design exploits the hierarchical structure of powerful cluster heads and the optimized multiple paths along with the adaptive scheduling to support reliable, high-throughput, and energy-efficient data transmission in wireless sensor networks. The simulation results shows that our proposed fast energy efficient data collection method improves the lifetime and the data collection rate of the network.

Key Terms: - Cluster; Two level scheduling; BFS Timeslot scheduling; Multichannel assignment

I. INTRODUCTION

A wireless sensor networks is usually made up of large number of sensor nodes, which are equipped with sensors, microprocessors and battery etc. Data collection from a set of sensors toward a common sink is a fundamental operation in wireless sensor networks (WSN). In many applications, it is essential to conserve the energy of sensor nodes and improve the lifetime of the network as well as increase the rate of data collection. Most of the wireless sensor network technologies are applied to various application areas such as the military, environment, medical service and business. For instance, one major application of wireless sensor network is to collect information periodically from a remote location where each node continually sense the environment and sends back this data to the base station which is usually located at some distance from the target field.

A Wireless Sensor Network (WSN) contains a large number of sensor nodes in the order of hundreds or thousands sensor nodes. The sensor nodes have the ability to communicate either among each other or directly to an external sink node. Sensor nodes are usually scattered in a region, which is an area where the sensor nodes are deployed. Sensor nodes coordinate among themselves to collect high quality information about the physical environment. Each sensor node makes its own decisions on the information it currently holds, and its knowledge of its computing, communication, and energy resources.

The problem in wireless sensor network is the data collection rate and extending the lifetime of the network, which mainly depends on the battery life of the sensor nodes. The data collection rate is affected due to the interference among the nodes and the link quality between the nodes in the network. Since the sensor nodes are battery powered, it is often very difficult to change or recharge the batteries. So to prolong the lifetime of the network is a critical issue. In order to achieve high energy efficiency, improve data collection rate and network scalability the sensor nodes are organised into clusters.

Within the cluster organisation, an intra-cluster communication can be single hop and multihop, as well as inter cluster communication. The previous works shows that the multihop between the nodes and sink nodes is more efficient than direct transmission because of the characteristics of wireless sensor networks. But, the hot-spots problem arises when using the multihop forwarding model in inter-cluster communication. Since the cluster heads closer to the sink node is burdened with heavy traffic, hence they will die much faster than the other cluster heads, reducing coverage area. Although many protocols are previously proposed to reduce energy consumption on routing paths to increase energy efficiency, which do not extend network lifetime due to the continuous many-to-one traffic pattern.

Sensor nodes in the network detect events and then communicate the collected information to a base station present at central location. The cost of transmitting a bit is higher than a computation cost and hence it would be advantageous to organize the sensor nodes into clusters. So in our proposed method we make use of clustered topology. In the clustered topology, the data gathered by the sensor node is communicated to the base station through a hierarchy of cluster heads. In each cluster, cluster heads have the capability to do some processing and aggregation on the collected data in order to remove redundancy in the collected data and improve the bandwidth usage. With the clustered topology, we propose two scheduling mechanism namely BFS timeslot assignment approach and Multichannel frequency assignment. The sensor nodes forwards the data to the cluster head using the using the proposed BFS timeslot assignment technique and the cluster heads forwards the data to the base station using the multichannel assignment technique.

II. RELATED WORK

The data routing in the wireless sensor networks can be done in different ways namely 1. Flat routing, 2. Hierarchical routing and 3. Location based routing. In flat routing, the sensor nodes use multihop to forward the data to sink. This approach is also called as data centric routing, where the base station sends queries to a region of sensor nodes and waits for data from the sensor nodes. In this approach the data is collected based on the attributes provided by the base station. In hierarchical or cluster based routing, as advantages such as scalability and energy efficient communication. This means that creation of clusters and assigning special tasks to cluster heads such as data aggregation and fusion, can greatly contribute to overall system scalability, lifetime, and energy efficiency. In location based routing, the sensor nodes are addressed by their locations. The distance between the sensor nodes can be determined by the received signal strength.

In most of the previous works on data collection in wireless sensor networks are based on the many to one model called the Convergecast, where many sensor nodes sense the environment and forward the data to a single sink or base station which is at some distance from network.

The LEACH hierarchical protocol was one the best clustering protocol which saved a lot of energy of the network when compared with the direct communication. In leach protocol, the cluster head is selected randomly that does not take into account energy consumption and the cluster setup time is more when compared with direct communication. The improvement of leach protocol is PEGASIS protocol, which performs better than the leach protocol. But the pegasis protocol introduces delay for the distant nodes and single leader becomes a bottleneck.

III. METHODOLOGY

1. Network model

Let us consider a wireless sensor network consisting of N nodes uniformly distributed over vast region to continuously monitor the environment. We denote the i^{th} sensor node by N_i and the corresponding sensor node set Node = $\{n_1, n_2, n_3,..., n_N\}$, where $|Node| = N$. We make some assumptions about the sensor nodes and the network model:

1. The sink node is placed far away from the sensing region. Sensor nodes and sink node are all stationary after the deployment.

2. All the sensor nodes are homogenous and have the same capabilities. Each sensor node is given a unique identifier.

3. The sensor nodes can use the transmission power control based on the distance to the receiver.

4. Sensor nodes needn't have the location information.
5. The links between the sensor nodes are symmetric. A sensor node can compute the distance to the other nodes based on the nodes received signal strength, if transmission power is given.

2. Cluster Head Selection

At first all the sensor nodes in the network determines a random number between 0 and 1 and then calculates a threshold value(T(n)) based on the energy of the sensor nodes, if the calculated threshold value is less the nodes random value, that node becomes the cluster head.

$$T(n) = \frac{P}{1-P \times \left(r \bmod \frac{1}{P}\right)} * \frac{E\ residual}{E\ initial} \quad \forall n \in G$$

With P as the cluster-head probability, r as the number of the current round and G as the set of nodes that have not been cluster-heads in the last 1/P rounds. Each of the sensor nodes in the network will be capable of being the cluster heads and after each few rounds of data collection a new cluster head can be selected based on the user requirements. The sensor nodes in the network will be having cluster head probability value used in the calculation of threshold value, for it become the cluster head. The sensor nodes satisfying the threshold value advertise themselves as the cluster heads and wait for other sensor nodes to join them. The cluster heads are elected based the on energy of the nodes i.e. the high energy nodes are first elected as cluster head.

3. Cluster Formation

In this paper, we form the clusters based on the node density. Since inter cluster routing requires more energy to forward the data to the sink node, due to hotspots created near the sink node. Hence we form the clusters based on the node density and forward the data to sink node in a single hop. The sensor nodes after they receive the cluster join request from the elected cluster head, they join the cluster head. The sensor nodes which are assigned to the clusters are named as the allocated set. The other left sensor nodes are named as unallocated set. While forming the clusters some of the sensor nodes might be intersected into more than one cluster that is a node might join more than one cluster group. Now, the intersected sensor nodes will be pushed to the cluster having more number of sensor nodes. Next, if the clusters with few sensor nodes are formed, they will be removed and those sensor nodes of the cluster will pushed to the neighbouring cluster using the distance between the cluster heads. Later, the cluster head having more energy between the two merged clusters will be chosen as the cluster head. Forming clusters with more number of nodes will increase the intra cluster communication costs but reduces the communication costs between the cluster heads and the base station that is a few numbers of channels would be enough to achieve data transmission. On the other hand if there are clusters having few number of sensor nodes reduces the intra cluster communication cost whereas it increases the communication cost between the cluster heads and the base station as it requires more number of channels to achieve data transmission. So it would be better to have clusters with more number of sensor nodes, as it saves more energy. Now, the nodes left out in the unallocated set will be pushed to the nearby clusters based on the received signal strengths between the nodes and the cluster heads.

The algorithm for Cluster formation

input: *A network with one or more CHs*
CH sends cluster join request to neighbouring nodes
Sensor nodes join the CHs after receiving the cluster join request
if *sensor nodes intersect with two or more clusters*
  move the intersected nodes to clusters having more nodes
end if
if *cluster contains few number of sensor nodes*
  merge those nodes into the neighbouring cluster
  elect the new CH for the merged clusters
end if
if *sensor nodes are left out from the clusters*
  move those nodes neighbouring clusters using the RSSI value
end if
output: *A network with one or more clusters*

4. Coverage reduction

When the sensor nodes are scattered randomly in the network during the deployment some of the sensor nodes might be close to one another and collect the same data. So we can make use of sleep mode to reduce the data amount and energy consumption. After the cluster formation the sensor nodes within cluster are further divided into groups based on their locations and sensing area of the nodes. So that sensor nodes with higher energy can be active and the other sensor nodes can go to sleep mode. If the sensing area of the sensor nodes within the cluster overlaps then the node having more residual energy will be active and other node will be put to sleep. Even if the nodes are present in the same regions then the node with more residual energy will be active. The reason for grouping sensor nodes is to distribute the active nodes evenly in the cluster to reduce the data similarity. Furthermore, the number of active nodes can adjusted based on the data precision required. After each round of data collection the entire process above is repeated that is a new cluster head selected after each round of data collection.

The algorithm for Coverage reduction
input: *A clustered network*
group the sensor nodes within the clusters based on their locations
if *sensing area of the nodes overlap*
  put the nodes with less energy to sleep mode
end if
if *the nodes are in same region*
  put few nodes with less energy to sleep mode
  put the other nodes to active mode
end if
output: *A clustered network with coverage reduction done*

5. Two level scheduling

After the cluster formation, we propose a two level scheduling for the data collection from the network to the base station. The two level scheduling are one between the cluster head and the sensor nodes within the cluster and the other between the base station and the cluster head.

5.1 BFS timeslot assignment

After the cluster formation, first, the cluster head construct the BFS tree within the cluster. Next, respecting the interfering constraint and adjacency constraints the timeslots are assigned to the sensor nodes within the cluster. Then the sensors nodes collect the data and forward it to the cluster head during its timeslot assigned by the cluster head.

The algorithm for BFS timeslot assignment
input: *A cluster with CH*
construct a BFS tree within each cluster with CH as root node
assign timeslots to nodes respecting the interfering and adjacency constraints
nodes forward the data to CH or parent nodes in their assigned timeslots
output: *A time schedule cluster for data forwarding to CH*

5.2 Multichannel assignment

The cluster head collects the data and aggregates it, so that data required only is sent to base station. Next, multiple frequency channels are assigned between the cluster heads and the base station. The links are scheduled based on the load that a link carries between the cluster head and the base station. The link which carries more number of packets (or load) is scheduled first and the remaining links are scheduled based on their loads. If there are few number of frequency channels, then timeslots can be assigned to the remaining cluster heads.

The algorithm for multichannel assignment
input: *A base station and all CHs*
assign the frequency channels between BS and CHs
if *frequency channels available are less than CHs*
  assign timeslots to the remaining CHs
end if
forward the data to the BS in the assigned channel or timeslot
output: *A network for forwarding data to the BS*

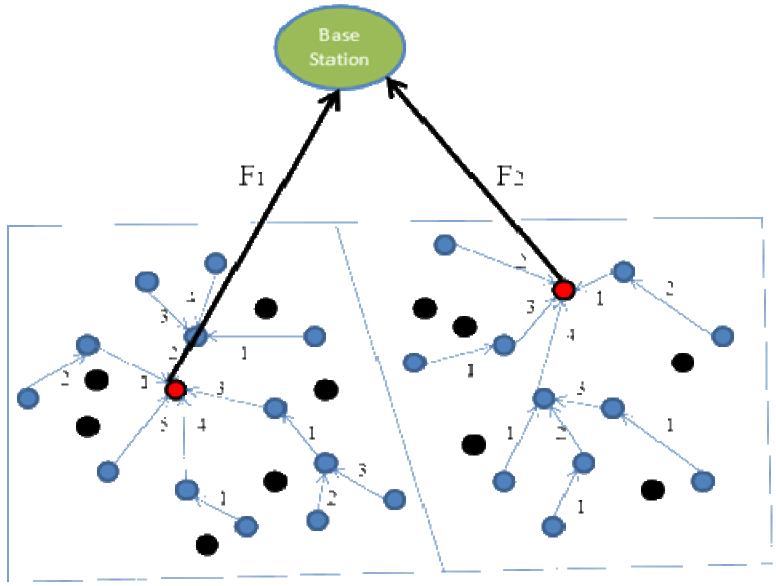


Figure: Illustration of the proposed method

IV. SIMULATION

The simulation experiments is done to analyse whether the proposed energy efficient clustering scheme improves the life time of the network and data collection rate achieved.

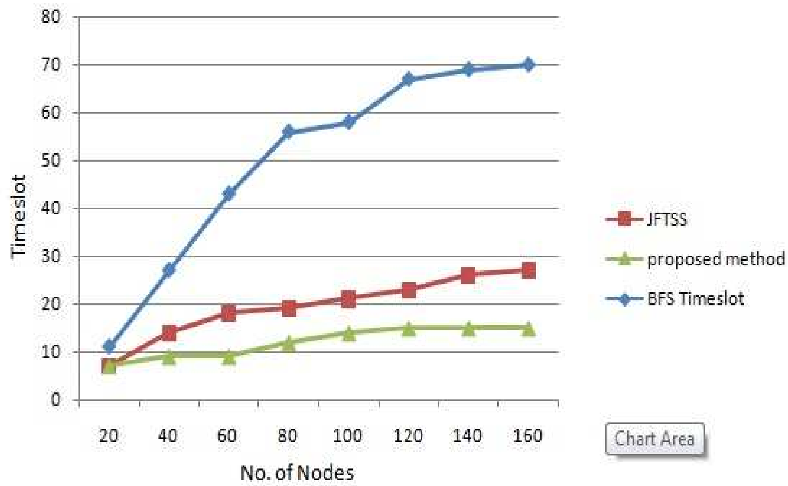
Before the simulation experiments, we first analyse the power consumption by the sensor nodes. The energy required by a sensor node for transmitting a signal is highly related to the distance. The following equation shows the energy consumed when sending a signal to a distance d by an amplifier.

Energy consumption = $E_{fs} * d^2$, if $d \le d_0$
$E_{tr} * d^4$, if $d > d_0$

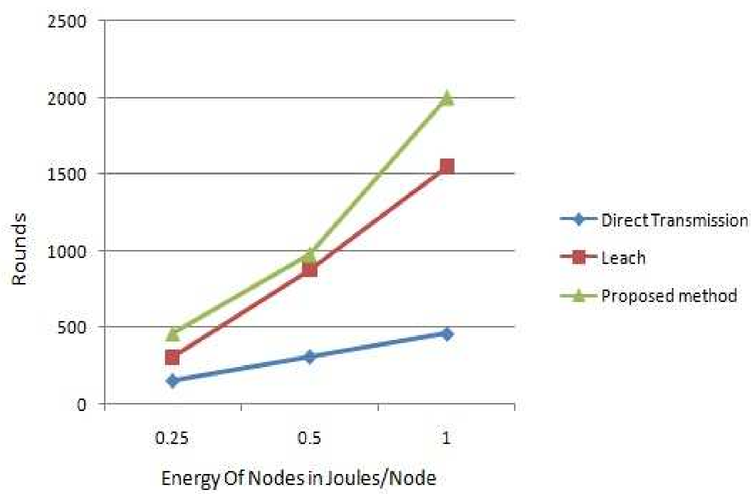
Using d_0 as a threshold, if the transmission distance is shorter than d_0, a free-space propagation model is used to calculate the consumed energy, which is proportional to the square of distance. If transmission distance is longer than d, the two-ray ground propagation model is used for calculation and the consumed energy is proportional to the fourth power of distance. In the above equation, the parameters fs and tr represents the free-space propagation model and two ray ground propagation model with their values equal to 10 pJ/bit/m^2 and 0.0013 pJ/bit/m^4, respectively.

For most sensor nodes in wireless sensor network, the consumed energy is proportional to the square of distance when collecting and forwarding data to their cluster head. The cluster allocation algorithm can increase the node density in a cluster and thus reduce the distance and energy consumed in forwarding data. As the operation continues, the sensor nodes near the base station may exhaust their electricity and thus cannot forward data for the outer sensor nodes. Consequently, the outer sensor nodes may need to transmit data directly to the base station at a longer distance, so the energy consumed is proportional to the fourth power of distance.

In this study, the simulation programs were written by Java. To obtain more accurate results, all experiments were performed 100 times to compute the average results. The parameters for the experiments are described in the following. The size of the application area is 700m * 700m and there are 2000 sensor nodes deployed. The initial energy in each sensor node equals 1J, and the energy for sending and receiving data are the same, 50nJ/bit. Each data aggregation takes 5nJ/bit with the compression rate = 70%. The amplifier parameters for the two models are fs = 10pJ/bit/m^2 and tr = 0.0013pJ/bit/m^4, respectively.



First, we compare the proposed method with the JFTSS and the BFS timeslot for the data collection rate (or the number of timeslots required). The timeslots required for the proposed method is 7, for JFTSS it is 7 and for BFS timeslot it is 11, when there are 20 sensor nodes in the networks. But when the number of sensor nodes is increased the proposed method performs better that is for 160 sensor nodes the proposed method requires only 15 timeslots for data collection and the other two methods JFTSS requires 27 and BFS timeslot requires 70 timeslots for data collection. Hence, the proposed method as better data collection rate than the JFTSS and the BFS timeslot.



Next, we compare the energy efficiency of our proposed method with LEACH and direct communication method. If the clustering method proposed is used then for 0.25J of energy of each node in the network we can achieve 492 rounds of data collection whereas in the leach clustering method only 312 rounds of data collection can be achieved and in direct communication (if the nodes individual collect and forward the data to the base station directly) only 155 rounds is achieved. For 1J of energy the proposed method can do 1998 rounds, leach can do 1548 rounds and direct communication can do only 462 rounds of data collection. Hence the proposed clustering method is more energy efficient than the leach method.

V. CONCLUSION

In this paper, we studied the fast energy efficient data collection in wireless sensor network. We made use of the clustering of network to extend the life time of the network and improve the data collection rate of the network. The nodes within the cluster are further grouped and some nodes are put to sleep to save the energy of the nodes and avoid redundancy of the data. For routing of data between the nodes and cluster head and between the cluster heads and base station we use the BFS timeslot and multichannel assignment which reduces the interference between the nodes and improves the data collection rate. The simulation results shows that the proposed methods increases the lifetime (rounds) and the data collection rate significantly when compared with the LEACH and JFTSS method.

REFERENCES

[1] S. Gandham, Y. Zhang, and Q. Huang, "Distributed time-optimal scheduling for convergecast in wireless sensor networks," Computer Networks, vol. 52, no. 3, pp. 610–629, 2008.

[2] Islam T. Almalkawi, Manel Guerrero Zapata, and Jamal N. Al-Karaki, "A Cross-Layer-Based Clustered Multipath Routing with QoS-Aware Scheduling for Wireless Multimedia Sensor Networks" Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks, 2012.

[3] Ö. D. Incel, A. Ghosh, B. Krishnamachari, and K. K. Chintalapudi, "Fast data collection in tree-based wireless sensor networks," USC CENG, Los Angeles, CA, Tech. Rep. CENG-2010-8, 2010.

[4] N. Lai, C. King, and C. Lin, "On maximizing the throughput of convergecast in wireless sensor networks," in Proc. GPC, 2008, pp. 396–408.

[5] A. Ghosh, O. D. Incel, V. A. Kumar, and B. Krishnamachari, "Multi-channel scheduling algorithms for fast aggregated convergecast in sensor networks," in MASS '09, Macau, China.

[6] S. Hussain and A.W. Matin, "Hierarchical Cluster-Based Routing in Wireless Sensor Networks," Proc. Int'l Conf. Information Processing in Sensor Networks (IPSN), 2006.

[7] http://en.wikipedia.org/wiki/Low_Energy_Adaptive_Clustering_Hierarchy.