Energy-Efficiency based Clustering and Data Aggregation for Wireless Sensor Networks

Priti K. Hirani
M-Tech CS (Student)
Jayoti Vidhyapeeth Women's
University, Jaipur, India

ABSTRACT

In wireless sensor network is maximizing the lifetime of the network. All sensor nodes directly send information to the base station so the energy requirement is very high. The data collection mechanism comes at the cost of power consumption and packet collision. We introduce automatic time series modeling based data aggregation scheme in wireless sensor networks. The decrease in the number of transmitting data values between sensor node and aggregation by using the time series predication model. We proposed in the paper energy-efficient cluster for wireless sensor network which is better suited periodical data gathering from the application. We proposed Energy –Efficiency Clustering and Data aggregation for sensor Network. There are four phases: cluster head, cluster head selection, data aggregation and maintenance. Important issued in wireless sensor network is minimize the total energy consumption requires to collect data.

Keywords

Wireless sensor network, data aggregation, clustering, energy –efficiency

1. INTRODUCTION

Wireless sensor network broadly used in military, environmental and medical system. Sensor node is small, expensive, self contained and battery –power device. Sensor node includes the sensing node to communication module data processing module. The large network filed is divided into a set of cluster. Each cluster has cluster header. All clusters are communicated with its cluster header. Cluster aggregation data transfer it to the base station. Sensor network is equipped with sensing and radio transmission units, while the power is highly limited. Due to the sensors network limited power and innovative techniques that improve energy efficiency to prolong the network lifetime are highly required. The advantages of these technologies reduced the cost and deployment of the small intelligent sensor node at homes in work place, super market, street, plantations and ocean to monitor the environment.

Clustering algorithm is very important for data mining and pattern reorganization and plays an especially important role in the reorganization data structure. An important application of clustering is voice reorganization character reorganization, etc. Clustering algorithm also used in machine vision, image vision, data compression and message retrieval. The cluster header performs the data aggregation there remove the redundancy and forward the aggregated packet to BS by using other CH as the intermediate nodes.

1.1 Network Model

In network model a few reasonable assumptions are followed:

- Numbers of N sensor nodes are uniformly dispersed within its field.

- All sensor nodes and the BS are stationary after deployment.

- If we want to communicate between two nodes the communication based on the single hop.

- Communication is symmetric and sensor can compute the approximate distance based on the received signal strength.

- All sensor nodes are equal significance.

We simplified the model for radio hardware energy dissipation. To transmit a l-bit of data to distance d radio expands:

$$E_{tx}(l,d) = l \times E_{ele} + l \times emp\ d^2 \ , \ d < d_{crossover}$$
$$l \times E_{ele} + l \times emp\ d^2 \ , d \leq d_{crossover}$$

When first item present in the energy consumption of radio dissipation, while the second item presents the energy consumption for amplifier radio.

1.2 Energy Model

Energy model is used for radio frequency signal.

- Radio frequency signal transmit and receive trough the radio transceiver.
- Radio transceivers have four states: transmit state, receive state, ideal state and sleep states.

The energy required to transmit b bits of data from node i to node j

$$E_{tx}(b_{i,j}) = (E_{ele} \times b) + (d(i,j)^2 \times E_{amp} \times b)$$

Where E_{ele} is power for transmitter and receiver circuit. d (i,j) is the distance between node i and j. E_{amp} is the power required for the amplifier circuit.

$$E_{amp} = \alpha_{amp} + \beta_{amp}$$

Where $\alpha_{amp} + \beta_{amp}$ is a constant which is depend on the amplifier architecture.

2. RELATED WORKS

There have been extensive researches in the data aggregation wireless sensor network. Clustering is one of the solutions of energy efficiency in wireless sensor network. Sensor nodes are organized into clusters, each cluster has a cluster head. The communication travel through the cluster head, which is travel forward to a neighbor cluster head unlit it reaches its destination. Heinzelman et al. proposed LEACH (Low Energy Adaptive Clustering Hierarchy) scheme to cluster sensor nodes and let the cluster head to aggregate data. The cluster head then, transmit the aggregated result directly to nodes. LEACH is one of the popular clustering algorithms. The LEACH operation is divided into round and each round starts with a set-up phase when the cluster is organized. All cluster members use the CSMA MAC protocol to transfer the data to its CH. Non-cluster head node turn off their RF transceiver completely until the arrival of their pre-allocated time slot. The drawback of LEACH is clusters are not evenly distributed due to its randomization of the local cluster head.

Lindsey et al. [proposed PEGASIS scheme which is organized all sensor nodes into a chain structure and rotates each node to communicate with sink. Both LEACH and PEGASIS schemes assume that each sensor node in the network can reach sink directly in one hop, which is limiting the size of the network for which they are applicable.

Sunhee Yoon, proposed Clustered Aggregation (CAG) technique that forms clusters of node sensing similar value within a given threshold (spatial) and their remaining cluster unchanged as long as the sensor values within a threshold over time (temporal core relation).CAG work into two modes: interactive and streaming. In interactive mode of CAG generates a single set of response for queries. In streaming mode, periodic response is generated in response to a query. In interactive mode of CAG only the spatial correlated data send. The streaming mode, of CAG advantage of both spatial and temporal data. These two techniques enable CAG to compute result efficiency with high accuracy and generate that the result within the user provided threshold regardless of data distribution.

Energy –Efficiency based Clustering and Data Aggregation

The operation of Energy –Efficiency based Clustering and Data Aggregation is divided into round and every round consists set-up phase and Steady state phase. The network is divided into rectangular region and it's called swim lanes, then each swim lane partitioned into smaller region is called grids. The grids further away from BS are bigger and have more nodes participated in Cluster Heads. Swim lane is partitioned into several rectangular grids along Y- axis. Grid of each swim lane is assigned a level of bottom grid is 1. Both numbers of grids and length of each grid in a swim lane are related with the distance from the swim lane to BS. For different swim lanes, the further swim lane is away from BS. For same swim lane, the grid further away from BS has longer length. EEBCDA make sure that the grid further away from BS has larger size.
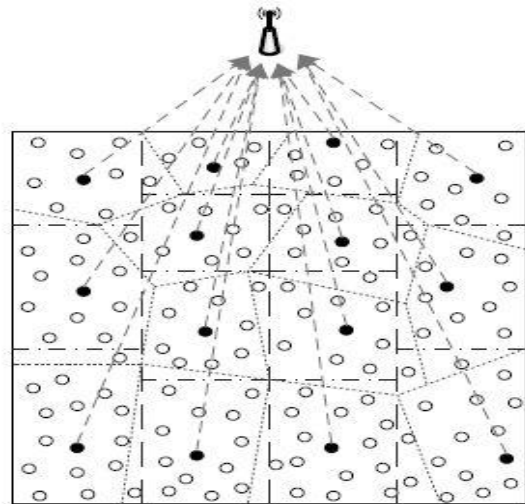

Fig. 1. Clustering formation

3. PROCESS OF DATA AGGREGATION

There are three processes of data aggregation: (1) how to find a clustering in data; (2) how to find a better way to cluster; (3) how to find a test method to prove that partition are correct and secure.

3.1 Divided k-average method

A clustering algorithm classified the data object into k part (k<=n) each part represent one cluster in given dataset that includes n data object and k- generated clustered. The k-average algorithm is used for the unsupervised learning algorithm. The process of clustering, first randomly select k object which is every object represent the average value of the core cluster and all other objects with found nearest distance. Ad process of calculation repeats until the creation function convergent.

The Following function is used in clustering method

$$E = \sum_{i=1}^{k} \sum_{p \subset C_i} |p - m_i|^2$$

In these formulate is the square error sum of all objects in the dataset, P is the space node which is representing the appointed data object, and m_i is the average value of the cluster c_i.

k- Average cluster algorithm formulated

Input: k average which is the number of clusters and n which is the number of objects in the data.

Output: k which is the number of objects which makes the square of error criteria to smallest.

Method:

Step 1: choose k cluster object for initial cluster.

Step 2: calculate distance square sum of any two nodes.

Step 3: First we calculated result of step (2) $z_1(i+1), z_2(i+2)\ldots\ldots z_k(i+1)$ the we calculate average of cluster object.

Step 4: if new cluster core is same as previous one the step the calculating process will be continued.

3.2 Data aggregation tree construction

Data aggregation is a simple tree construction algorithm. First root broadcast a construction beacon message which includes its own ID and 0 as the values of depth. When a node X is receiving a broadcast message for the first time from a node Y, X makes its own depth if Y plus a parent to be Y. broadcast the message again. This process will continue until all the node had received this message. The BS can disseminate the query message through the tree. The aggregation features that BS request, a random number is added into the query which is generated by the BS as grouping and use probabilistic grouping as well as query identification the next phase.

The construction of the aggregation tree can finish lie on the selection of a group of the main node. A group is the topmost mode with a group that completed and submitted the aggregate value of the group. Group nodes are selected dynamically based on the count value and the group seed s_g received in the query dissemination phase. In tree construction there are two functions used for the group selection. one is cryptographically secure function that is uniformly mapped the input value phase[0,1] and other grouping function is f_g takes positive integer s input and output between[0,1].
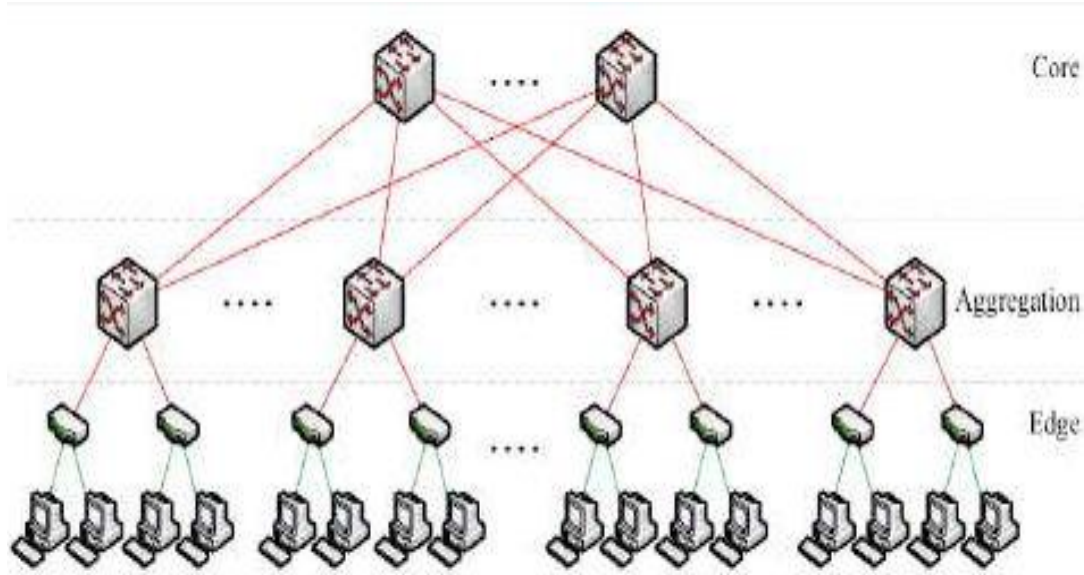

Fig. 2. Data aggregation tree construction

3.3 The aggregation process

First, we introduce the packet format used in aggregation. Each aggregation packet contains sender ID. An aggregated data values and count value to show how many nodes have contributed to the aggregated data. A flag field is used in each packet to specify what further aggregation is needed at the root. Flag value 1 means that no further aggregation is needed while 0 means this flag field initializes to 0. The process of aggregation is included leaf node aggregation, intermediate node aggregation and leader node aggregation.

3.3.1 Leaf node aggregation

Data aggregation starts from the leaf node in the aggregation tree toward the BS. Leaf node doesn't need to do any aggregation it simply sends its ID, data and count value to its parent.

$u \rightarrow v$: u, 0, E $(K_{u.v}, 1/R_u/S_g)$ /MAC_u

MAC_u = MAC $(K_u, 0/1/u/R_u/S_g)$

Which 0 is the aggregation flag, 1 is count values, R_u is read in node u, and MAC_u is the MAC value computed by node u with individual key shared with BS. s_q is identifying the query and prevented replay attacks.

3.3.2 Intermediate node aggregation

The intermediate node receives an aggregate from a child node, first checks the flag. If the value of the flag is 0, it keeps a local copy of the aggregate otherwise the node directly forward the packet to its parent node. For a packet received from child node with flag 0. The node first decrypts the data using its pair wise key shared with child nodes.

W is the parent of v, node v is node a leader; v sends the following packet to w.

$v \rightarrow$ w.v,o,E $(K_{v,w}, 3/Agg_v/S_g)/MAC_v$

$Agg_v = F_{agg} (R_v, R_u, R_u)$

MAC_v = MAC$(K_v, 0/3/Agg_v/MAC_u + MAC_u/S_g)$

3.3.3 Leader node aggregation

Intermediated node has processed the aggregation from its child node and find that group leader node based on the condition H $(s_g /x) < f_g(c)$. a regular intermediate node, it also computes a new aggregate, keep local copies of these packets with flag 0,and corresponding MAC uses its individual key. Regular intermediate node, it changes the flag to 1 in its aggregation packet and encrypt the new aggregated with its individual key shared with BS.

4. CONCLUSION

In this paper, we have studied the problem of energy efficiency data transmission in wireless sensor network and cluster based data aggregation methods. Four phases are consisting in the data aggregation method. Energy –Efficiency based Clustering and Data Aggregation is divided network into a grid with unequal size, the grid further away from BS has bigger size and more nodes. Also, we study data aggregation process which is in three parts: the first finds cluster tendency, second is to find cluster and group leader and third is made secure aggregation. Also introduces the divided k-average method, data aggregation tree construction, aggregation process.

5. REFERENCES

[1] Rajagopalan.R, Pramod K, V, "Data aggregation technique for sensor network: A survey", IEEE Communication Survey and Tutorials, vol-8, no.4,2006.

[2] W.Heinzelman, A. chandrakasan, H. Balakishan, "Energy-Efficient Communication Protocol for wireless Micro Sensor Network", processing of the 33rd International conference System Science, 2000.

[3] K-means Clustering over a Peer to Peer Network:, IEEE transaction on Knowledge and Data engineering, vol,21, no, 10.

[4] Masoume Jabbarifar, Alireza Shmeli Sendi, Alireza Sadighian,Naser Ezzati Jivan,Michel Dagenais, "A Reliable and Efficient Time Synchronization Heterogeneous Wireless Sensor Network", Wireless Sensor Network, Vol.2 No.12, PP.910-918, 2010.

[5] Jigui Sun, Jie Liu and Lianyu Zhao, "Research on Clustering" Journal of Software ,2008

[6] Intanagonwiwat, C., Estrin, D., Govindan, R., Heidemann, J: Impact of Network Density on Data Aggregation in Wireless Sensor Network.. In: Proceedings of 22nd International Conference on Distributed Computing System ,pp.2-5, 2002

[7] Xu,H., Huang, L., Zhang, Y., Huang, H., Jiang, S., Liu, G: Energy-efficient Cooperative Data Aggregation For Wireless Sensor Network. Journal of parallel and Distributed Computing 70 953-961. 2010

[8] Ding, M., Cheng, X., Xue, G.: Aggregation Tree Construction in Sensor network. In: Proceedings of IEEE 58th Vehicular Technology Conference pp. 2168-2172. IEEE Press, 2003

[9] Younis, O. and Fahmy, S, " A Hybrid Energy Efficiency distributed Clustering Approach For AD Hoc Sensor Networks" ,IEEE Transaction on Mobile Computing ,vol 3 No.4 pp.366-379, 2004

[10] Patel, R.B., Aseri, T.C., Kumar, D, "Energy Efficiency heterogeneous Clustered Scheme for Wireless Sensor Network", International Journal of Computer Communication Elsevier ,2009

[11] Jun Yue, Weiming Zhang,Weidong Xiao, Daquan Tang, Jiuyang Tang, "Energy Efficient and Balanced Cluster – Based Data Aggregation Algorithm for Wireless Sensor Network" international workshop on Information and Electronics Engineering