Data Aggregation Protocols to Prolong Network Lifetime in Wireless Sensor Network

Vishal Garg¹, Shikha Mangla²
¹²Department Of Computer Science, JMIT Radaur, Haryana India

Abstract—Wireless sensor network is made up of low-cost sensor nodes. Sensor nodes are deployed over a geographic region to sense the region. In wireless sensor network base station queries the sensor nodes, sensor nodes send the sensed data to base station. Sensed data contains redundant data because collected by various nodes in same region. If sensor nodes send the data directly to base station, nodes lose their energy quickly and network become unusable. So, data will be aggregated at aggregator nodes. Data aggregation in wireless sensor network removes redundant data collected by various sensor nodes. Various data aggregation protocols like LEACH, HEED, PEGASIS, EEBCDA etc. are used. Through these protocols network lifetime and energy efficiency is improved.

Keywords—Wireless Sensor Network, Data Aggregation, Energy Efficiency,

I. INTRODUCTION

In wireless sensor network, sensor nodes collect target specific information from the environment and this information is transferred to a central base station where it is processed, analyzed, and used by the application. Forwarding of entire information directly to base station increase redundancy and load on the network. So, instead of directly transfer of data to base station, some intermediate nodes called aggregator nodes between sensor nodes and base station are used. These nodes fuse the data and removes redundancy incurred by various sensor nodes. Such processing of data is generally referred as data aggregation. The main objective of data aggregation is to increase the network lifetime by reducing the resource consumption of sensor nodes (such as battery energy and bandwidth).

In Data Aggregation protocols group of sensor nodes sense the network and collect information from target region. When the base station queries the network, instead of sending each sensor node’s data to base station, one of the sensor nodes, called data aggregator, collects the information from its neighboring nodes, aggregates them (e.g., computes the average). This continuous data aggregation operation improves the bandwidth and energy utilization.

II. DATA AGGREGATION PROTOCOLS

A. Tree Based Data Aggregation

In [4,6] A Tree Based Data aggregation protocol proposed framework called Tiny Aggregation Service (TAG). TAG performs data aggregation in two phases:

• In the first phase, called distribution phase, base station queries are disseminated to the sensor nodes. In the data distribution phase a message is broadcasted by the base station so that sensor nodes may organize their routing tree and the base station can send its queries(message specify the level or distance from the root of the sensing node, level of root is equal to zero).

• In the second phase, called collection phase, the aggregated sensor readings are routed up the aggregation tree. Sensor nodes answer the base station queries through their parents.

B. Directed Diffusion

In [4,7] Directed diffusion takes places in three phases: (i) interest dissemination (ii) gradient set up (iii) path reinforcement and forwarding.

In the first phase, the base station sends an interest message to its neighbours, that describes the type of data that needs to be collected and the operational mode for the collection. Upon reception of this message, each sensor node rebroadcasts the message to its neighbors. Sensor nodes also prepare interest gradients which are basically the vectors containing the next hop that has to be used to propagate the result of the query back to the base station (gradient setup). For each type of data a different gradient may be set up. At end of gradient setup phase for a certain type of data. Data forwarding is performed to the Base Station. Data aggregation is performed during data forwarding phase.

C. Hierarchical (Clustered) Based Data Aggregation

In [1,2,4] cluster-based data aggregation protocols is defined where sensor nodes are divided into clusters. In each cluster, a cluster head is selected. Cluster head aggregate the data and send the aggregated data to the Base Station.
Cluster heads can communicate with the Base Station through the single hop or via long range radio transmission.

In the cluster based data aggregation nodes are organized into clusters as shown in fig.1 and sensor nodes sends their data to cluster heads through either single hop or multi hop and cluster heads send their aggregated data to the Base Station through single hop. When the sensor nodes use single hop to reach the cluster head, the nodes closer to a cluster head have a higher load of relaying packets as compared to other nodes. When the nodes use multi hop to send their data to the cluster head nodes that are farther away from cluster head have the highest energy due to long energy communication. In [10] When the sensor nodes are within one hop from base station, these nodes have extra burden of relaying packet and when these nodes expires network becomes unusable, So instead of minimizing the total energy expenditure in the network, the goal should be to minimize the energy expenditure of the sensor nodes around the base station, since these nodes determine the lifetime of the system.

Two Protocols related to cluster Based Data Aggregation are as follows:

1) LEACH: In[2,4] A self-organizing and adaptive clustering protocol, called Low-Energy Adaptive Hierarchy (LEACH) is proposed. In LEACH type of nodes depends on the network. In [2,8] LEACH protocol network is homogeneous means only one type of nodes are used and the nodes communicate with their elected cluster heads (sensor nodes are cluster heads) using single hop communication. The cluster heads aggregate the received data, and transmit it to a base station using a single hop transmission. LEACH also uses rotation of the cluster heads for load balancing, since the cluster heads have the extra burden of performing the long range transmissions to the distant base station. Thus LEACH faces the problem of non-uniform energy drainage by role rotation. In [2,11] two type of nodes are use (i) Sensor Nodes (ii) Cluster Head Nodes. In this network there is no need to elect the sensor nodes as the cluster head. Cluster head nodes acts as data fusion points within the network and send the aggregated data to Base Station. Sensor nodes performs simple functionality, they performs only sensing, cluster heads are complex since they perform routing with in cluster and perform data fusion and perform long range communication. This paper measures the energy efficiency of single hop or multi hop mode.

The protocol consists of two phases:

- **Set Up Phase:** In this cluster structures are formed.
- **Data Distribution Phase:** In this cluster heads aggregate and transmit the data to the base station.

**Cluster Structure:** The region of cluster is circular and cluster head is located at the center of region. There are N sensor nodes uniformly distributed over the cluster area. The amount of energy required to transmit the packet over distance ‘r’ is given by:

\[ P = T(E + \mu r^k) \]

\( E = \text{Amount of energy spent in the transmitter electronic circuitry} \)

\( \mu r^k = \text{Amount of energy spent in RF Amplifier} \)

When receiving a packet receiver circuitry is required, the energy spent on receiving a packet is ‘L’. So total energy over distance x is \((2E + \mu r^k)\).

In cluster sensor nodes can communicate with cluster head via single hop or multi hop. In[2] another mode i.e. hybrid mode combination of single hop or multi hop is defined. Various modes of communication are as follows:

**One Hop Mode:** In[2] single hop mode each node directly transmit packet to another node. In single hop mode the nodes located farthest from cluster head have to spend maximum amount of energy. The battery energy of sensor nodes in the single hop communication system \( E_s \) be :

\[ P_s = T(E + \mu r^k) \]

**Multi Hop Mode:** In[2] multi hop network is divide into circular rings. The sensor nodes communicate with multi hop cluster head. Each sensor node has communication radius ‘R’. It is also assume that R<a means the communication area of sensor node is smaller than the total area of cluster otherwise the cluster is same as the single hop cluster. In order to determine the worst case energy drainage in the network, we divide the circle into concentric rings of thickness T. We note that with a multi-hop communication radius of R, if a packet is generated in the nth ring, during its journey to the cluster head, the packet has to travel through each of the inner rings. For each data gathering cycle, we determine the average energy expenditure of a sensor node in the nth ring, where n varies from 1 to R.
Since the nodes are uniformly distributed, the average number of sensor nodes which lie outside the nth ring is $N(n-1)^2/\pi^2$. Hence $N(n-1)^2/\pi^2$ no. of packets have to be relayed by the nodes in the nth ring into (n-1)th ring. There are $N(nT)^2-\pi(n-1)T^2)/\pi^2$ nodes in the nth ring that have to relay the packets coming from the nodes outside the ring. If we denote the average number of packets that a typical node in the nth ring has to relay by $P_n$, then we obtain:

$$S_n = \frac{N(nT)^2-\pi(nT)^2)/\pi^2}{N(nT)^2-\pi(n-1)T^2)/\pi^2} = \frac{r_2-nT^2}{T(2n-1)}$$

Therefore energy spent during one cycle by a node in the nth ring $e_n$ is:

$$p_n = (2E+\mu R_k \lambda_k + (E+\mu R_k)$$

Battery energy for multi hop scenario $P_m$ for n=1 is given by:

$$P_m = N(2E+\mu R_k \lambda_k + (E+\mu R_k) = T (E+\mu R_k) \times (\frac{22}{r_2} - 1) + (E+\mu R_k)$$

**Hybrid mode:** In[2] A hybrid mode for communication between the sensor nodes and the cluster heads is defined. Hybrid mode is the combination of single mode and multi mode. In single hop mode the sensor nodes which are farthest from the cluster head have the highest energy drainage. In multi-hop mode the sensor nodes that are closest to the cluster head have the highest energy drainage due to packet relaying. This section presents that the sensor nodes alternate between single hop mode and multi-hop mode periodically. When single hop mode is used the nodes near the cluster head are relieved of their relaying burden, and when multi hop mode is used the nodes which are farthest from the cluster head are relieved of their burden of long range transmissions to the cluster head.

Thus by alternating between the two modes of communication load is uniformly distributed. The cluster head can broadcast a packet periodically to all the nodes in its cluster and packet contains the information asking them to switch between the two communication modes. The idea is to use different paths (not necessarily using the nearest node as the next hop node) for relaying of packets, and to determine the ratio of time for which each of the paths should be sustained so as to minimize the overall energy expenditure.

2) **HEED:** In[1,12] presents HEED (Hybrid Energy Efficient Distributed Clustering) algorithm for data aggregation to save energy. HEED periodically selects cluster head according to hybrid of the node residual energy and secondary parameter such as node proximity to its neighbor such as node degree. HEED terminates in O(1) iterations incurs low message overhead and achieves uniform cluster head distributed across the network. HEED surety connectivity of clustered network.

Network lifetime can be define as time elapsed until the first node dies means it loss its entire energy. Energy consumption is characterized by (i) useful sources (ii) wasteful sources. Useful energy consumption is due to (i) transmitting/receiving data, (ii) processing query requests, and (iii) forwarding queries/data to neighboring nodes. Wasteful energy consumption can be due to (i) idle listening to the media, (ii) retransmitting due to packet collisions, (iii) overhearing, and (iv) generating/handling control packets. Various clustering protocol has been proposed to prolong the network lifetime. Network lifetime is prolonged through (i) reducing the number of nodes contending for channel access, (ii) summarizing information and updates at the cluster heads, and (iii) routing through an overlay among cluster heads, which has a relatively small network diameter. This paper presents a HEED protocol: a hybrid of residual energy and communication cost. HEED has four primary objectives: (i) prolonging network lifetime by distributing energy consumption, (ii) terminating the clustering process within a constant number of iterations, (iii) minimizing control overhead (to be linear in the number of nodes), and (iv) producing well-distributed cluster heads. A node may fail if it loss its entire energy. This paper presents that HEED protocol increase network lifetime through proper distributed cluster head in the network.

**Network Model Requirements** It has following requirements:

- Nodes are not completely stationary means nodes move at very slow speed.
Each node has same transmission power levels.
- Energy consumption is not uniform for all nodes.
- Nodes are not equipped with GPS antennae. So they cannot determine their position.
- Nodes are left unattended after deployment. Battery recharge is not possible.
- Links are asymmetric. Asymmetric means that two nodes can communicate with different transmission power levels.

**Clustering Objective** Consider that \( n \) nodes are distributed in a field and we have to identify set of cluster head which cover the entire field.

Following requirement must be met:
- Clustering should terminate in fixed no of iterations.
- Cluster head has high residual energy than regular nodes.
- Clustering is completely distributed. Each node makes decision independently.

**HEED Protocol** In [2] Main Objective of HEED protocol is to prolong network lifetime. This is achieved by proper distribution of cluster head. It takes two parameters

First parameter is residual energy of node. Node with highest residual energy is selected as cluster head. This makes the set of tentative cluster head. To select the cluster head among set of tentative cluster head secondary parameter i.e. "intra cluster communication cost" is used. The secondary parameter is function of (i) cluster size (ii) whether or not variable transmission power levels are permissible for intra cluster communication. If the fixed transmission power level is used then cost is proportional to node degree otherwise to the 1/node degree. It means that nodes join the cluster head with minimum degree to distribute the load.

**HEED Protocol Operation** In HEED protocol we have to determine the proper distribution of cluster head. For this we have to determine node is cluster head or not? This is done by firstly calculating the \( \text{CH}_{\text{prob}} \) ( \( \text{CH}_{\text{prob}} \) is set by node, it is the probability of becoming a cluster head). It is defined as follows:

\[
\text{CH}_{\text{prob}} = \text{C}_{\text{prob}} \times \frac{E_{\text{resi}}}{E_{\text{max}}}
\]

\( \text{C}_{\text{prob}} \) is used to limit the initial cluster head announcement.
\( E_{\text{resi}} \) = Estimated current residual energy of node.
\( E_{\text{max}} \) = maximum energy(Highest energy among all nodes)

In HEED protocol a set of tentative cluster head(TCH) is selected by a parameter i.e. residual energy of each node. Nodes with maximum residual energy come under this set.

In any iteration ‘i’ a node elects to become a cluster head with probability \( \text{CH}_{\text{prob}} \). A node ‘vi’ selects its cluster head to that node that has lowest cost in \( S_{\text{CH}} \). If a node elects itself as a cluster head it sends announcement message i.e. \( \text{cluster head message(Node id, Selection Status, Cost)} \) if \( \text{CH}_{\text{prob}} \) is less than 1 or reached to 1. If a node left uncovered and node completes HEED execution than the node announce itself as cluster head i.e. final \( \text{CH} \). A tentative cluster head is elected as regular node if there is a node with lower cost.

**D. Chain Based Data Aggregation**

In [3] sensor networks, data aggregation helps to reduce the amount of data transmitted between sensor nodes and the Base Station. Data aggregation combines one or more data packets from different sensor nodes by data aggregation nodes. The LEACH protocol is solution to this data collection problem, where a small number of clusters are formed in a self-organized manner. Data Aggregation node in each cluster collects and fuses data from nodes in its cluster and transmits the result to the BS. LEACH uses to rotate the cluster heads. Further improvements can be obtained if each node communicates only with close neighbors, and only one designated node sends the combined data to the BS in each round. An improved protocol called PEGASIS (Power-Efficient Gathering in Sensor Information Systems), which is near optimal for this data gathering application in sensor networks. The key idea in PEGASIS is to form a chain among the sensor nodes so that each node will receive from and transmit to a close neighbor. Gathered data moves from node to node, get fused, and eventually a designated node transmits to the BS. Nodes take turns transmitting to the BS so that the average energy spent by each node per round is reduced. Building a chain to minimize the total length is similar to the traveling salesman problem, which is known to be intractable. However, with the radio communication energy parameters, a simple chain built with a greedy approach performs quite well.

**1) Network Model for PEGASIS:** In this model radio consumes \( E_{\text{pow}} = 50\text{nj/bit} \) to run the transmitter/receiver circuitry and \( E_{\text{amp}} = 100\text{pj/bit/m}^2 \) for transmitter amplifier. We can calculate the transmission and receiving cost in terms of energy. Total energy consumed for transmitting a n-bit message at distance ‘r’ is given by:

\[
E_{\text{Trans}(n,r)} = E_{\text{pow}}*k + E_{\text{amp}}*n*r^2
\]

Total energy consumed for transmitting a k-bit message is given by:

\[
E_{\text{Recv}}(n) = E_{\text{pow}}*n
\]
2) PEGASIS (Power-Efficient Gathering in Sensor Information Systems): In this protocol a chain is formed of sensor nodes, only one type of node is used means all nodes are sensor nodes no cluster head is used. All the nodes can perform fusion. In this the sensor nodes transmit and receive the data from closest neighbor and may be a leader for transmission to the base station. So the load is distributed among the sensor nodes. The nodes are organized to form a chain. Chain is constructed using greedy algorithm by sensor node starting from some node. Alternatively, the BS can compute this chain and broadcast it to all the sensor nodes. For constructing the chain all the nodes have the global knowledge of the network. Chain is formed before the first round of communication. In this when a node dies, the chain is reconstructed to remove the dead node.

For gathering data in each round, each node receives data from closest neighbor, fuse with its own data and transmit the data to next closest neighbor. Node ‘i’ is at some random position ‘j’ on the chain. Nodes take turns transmitting to the BS, and we will use node number i mod N (N represents the number of nodes) to transmit to the BS in round i. Thus, the leader in each round of communication will be at a random position on the chain, which is important for nodes to die at random locations. In fig. 2, node N2 is the leader, and it will pass the token along the chain to node N0. Node N0 will pass its data towards node N2. After node N2 receives data from node N1, it will pass the token to node N4, and node N4 will pass its data towards node N2.

![Fig. 2 Information Passing Approach](Image)

PEGASIS performs data fusion at every node except the clustering node. In this a node receives the data from previous node, fuse its own data and form the packet of original length. In fig. 2 N1 receives data from N0, fuse it with its own data and pass it to N2, similarly N3 receives data from N4 and fuse it with its own data and pass the data to N2 and N2 fuse its data with the both neighbours. Finally, N2 transmits the data to Base Station. In PEGASIS, there is possibility that some nodes have neighbours at large distance, such nodes consume more energy than other nodes. We improve performance by not allowing such nodes as leaders. This is achieved by setting a threshold on neighbours distance to become leaders. Whenever a node dies chain will be reconstructed to bypass the died node.

E. Energy Efficient and balanced cluster based data aggregation

In[5] EEBCDA overcomes the problem of unbalanced energy dissipation. In this area is divide into unequal grids, the grid which is at small distance from base station has more sensor nodes than other grids and consumes more energy. Cluster head rotate among the unequal grid. By this method EEBCDA share energy load.

The network is divide into rectangular regions called swim lanes and swim lane is further divide into smaller rectangular regions called grids. The node with maximum residual energy is selected as cluster head. The grids which are at smaller distance from base station has more nodes and more residual energy to participate in CH rotation. EEBCDA is divided into rounds and round is divide into three phases. These are:

- Network Division Phase
- Set-Up Phase
- Steady State Phase

1) Network Division Phase: In this base station is outside the deployment area along Y-axis. The deployment area is divide into S rectangular swim lanes along X-axis. All swim lanes are of equal width i.e. W, and the length of swim lane is equal to border deployment area. Then use a sequence of integers from 1 to S is used to assign the ID to swim lanes and ID of leftmost swim lane is 1. And further swim lane is divide into grids along Y-axis. Each grid of each swim lane is assigned a level and then assign a sequence if integers from 1 as the levels of grids in each swim lane and the level of bottommost grid is 1. Each grid has same width with swim lane. The number of grid and length of each grid are related with distance from swim lane to BS. A swim lane which is at small distance from base station, fewer grid it has. For same swim lane the grid farther away from base station has longer length. Each grid is assigned a tuple (i, j) where ‘i’ defines swim level and ‘j’ defines level of grid.

2) Set-Up Phase: In the set up phase cluster is formed and cluster head is selected in each round, the node is selected as cluster head that has maximum residual energy. If there are two nodes with same residual energy then cluster head is selected according to node ID. In the first round, the CH of a grid is selected by the cooperative work of nodes within the grid.

Firstly, each node sends a NODE_MESSAGE (id, (u, v), Er, (x, y)) message to other nodes in the same grid, where id is node ID, (u,v) is node’s grid ID, Er is node’s residual energy and (x, y) is the location of node.
The range of NODE_MESSAGE only has to cover the node’s grid to save energy. Every node collect the information of all nodes in the same grid and select the CH of its grid. In each grid after the first round, the CHs are selected by the cooperative work of the CHs of last round (LRCH). During data gathering in every round, every member transmits its residual energy information along with its data to cluster head. At the beginning of current round, each last round cluster head(LRCH) sorts all nodes in its cluster of last round in terms of grids, and chooses a tentative cluster head (TCH) for each grid, then, sends a LRCH_TCH_MESSAGE message which contains the information of TCHs to the LRCHs of corresponding grids, the transmission distance of LRCH_TCH_MESSAGE is the maximal distance between this last round cluster head (LRCH) and the vertexes of all corresponding grids. Finally, each LRCH chooses a cluster head(CH) from grid’s TCHs. Once the CHs have been selected, each CH broadcasts a CH_BRD_MESSAGE message to inform other nodes. The range of CH_BRD_MESSAGE of each CH only has to cover the CH’s grid, by this way, EEBCDA not only guarantees that every ordinary node is covered by at least one message, but also saves energy. Then, each ordinary node chooses the closest CH to join a cluster.

3) Steady State Phase: In this phase sensor node send data to cluster head and then cluster head aggregate the data and than sends the fused data to base station.

F. Hybrid Approach

In[4,9] Hybrid approach combines the clustered approach with directed fusion. In this approach during the first phase query definitions are include in the interest message. Interest message is broadcasted over the network. In this cluster head that has the knowledge of query definition can aggregate the data and the cluster head that performs inter cluster communication are involved in transmission of interest message. In this regular sensor nodes do not transmit data until they are capable of servicing a request.

III. CONCLUSION

This paper concludes that there are many protocols to achieve data aggregation. But hierarchical (clustering based) routing is widely used. LEACH protocol i.e. low energy adaptive clustering hierarchy is used.

In LEACH protocol a node is elected as cluster head that has maximum residual energy. In this cluster head rotate among nodes. PEGASIS: Power-Efficient Gathering in Sensor Information Systems is another protocol. Under this protocol a chain of sensor nodes is constructed. In PEGASIS, Node acts as regular node or also as cluster head. A node receive data from previous node, fuse it with its own data and transfer to to next node. Another protocol known as HEED (Hybrid Energy Efficient Distributed clustering) approach is used. HEED used another parameter i.e. intra cluster communication cost is used. But to achieve energy efficient data aggregation is a trade off. EEBCDA (Energy Efficient Based Cluster data Aggregation), under this network is divide into rectangular grids of unequal size. The grids which are at small distance from base station has more energy and more nodes. In LEACH cluster head will communicate with base station by single hop but sensor nodes communicate with cluster head through single hop or multi hop.

REFERENCES