Sensor Networks Data Gathering Scheme for Deterministic and Probabilistic Network Models

Mr.G.Ravichandran¹ Dr.A.Viswanathan²
¹Student of M.E ²Assistant Professor
¹,²K.S.R College of Engineering, Tiruchengode, Tamilnadu, India

Abstract— Data Aggregation Trees (DATs) are used to gather and aggregate data values from the sensor nodes. DATs are constructed to fulfill different user requirements under the Deterministic Network Model (DNM). Probabilistic Network Model (PNM) is built with probabilistic lossy links. Load-Balanced Data Aggregation Tree (LBDAT) is constructed under the PNM. Load-Balanced Maximal Independent Set (LMBIS), Connected MIS (CMIS) and LBDAT construction problems are solved in the data gathering scheme. Approximation algorithm is used to search for an LBMIS. Load Balanced Parent Node Assignment (LBPNA) is performed using the randomized approximation algorithm. LBDAT is constructed by assigning a direction of each link in the tree from the children node to the parent node. The LBDAT construction scheme is enhanced for the distributed model. The system is improved to support DNM and PNM based WSN. The distributed LBDAT scheme is tuned to manage monitoring area environment. Coverage relationship is integrated in the LBDAT construction process.

Key words: LBDAT, DNM, PNM

I. INTRODUCTION

The design of wireless sensor networks depends on the application requirements. Environmental monitoring is an application where a region is sensed by numerous sensor nodes and the sensed data are gathered at a base station where further processing can be performed. The sensor nodes for such applications are usually designed to work in conditions where it may not be possible to recharge or replace the batteries of the nodes. This means that energy is a very precious resource for sensor nodes and communication overhead is to be minimized. These constraints make the design of data communication protocols a challenging task.

A common scenario of sensor networks involves deployment of hundreds or thousands of low-cost, low-power sensor nodes to a region from where information will be collected periodically. Hence, sensor nodes will periodically sense their nearby environment and send the information to a sink which is not energy limited. The collected information can be further processed at the sink for end-user queries. In order to reduce the communication overhead and energy consumption of sensors while gathering, the received data can be combined to reduce message size. A simple way of doing that is aggregating the data. A different way is data fusion which can be defined as producing a more accurate signal by combining several unreliable data measurements. In this paper, we focus on scenarios where perfect aggregation is used while gathering data, meaning that all forwarded messages are of the same size.

II. RELATED WORK

Due to the dense sensor deployment, many different DATs can be constructed to relay data from the monitored area to the sink. According to the diverse requirements of different applications, the DAT related works can be roughly classified into three categories: Energy-Efficient Aggregation Scheduling [7], Minimum-Latency Aggregation Scheduling [10], [1], [2] and Maximum-Lifetime Aggregation Scheduling [6], [9], [8]. It is worth mentioning that aggregation scheduling attracts a lot of interests in the current literatures. Unlike most of the existing works which spend lots of efforts on aggregation scheduling, we mainly focus on the DAT construction problem.

Most of the existing DAT construction works are based on the ideal Deterministic Network Model (DNM), where any pair of nodes in a WSN is either connected or disconnected. Under this model, any specific pair of nodes are neighbors if their physical distance is less than the transmission range, while the rest of the pairs are always disconnected. In most real applications, the DNM cannot fully characterize the behaviors of wireless links due to the existence of the transitional region phenomenon. It is revealed by many empirical studies that, beyond the “always connected” region, there is a transitional region where a pair of nodes are probabilistically connected via the so called lossy links. Even without collisions, data transmissions over lossy links cannot be guaranteed. Usually there are much more lossy links than fully connected links in a WSN.

Therefore, in order to well characterize WSNs with lossy links, a more practical network model is the Probabilistic Network Model (PNM) [3], [4], [5]. Under this model, there is a transmission success ratio \( \ell_{ij} \) associated with each link connecting a pair of nodes \( v_i \) and \( v_j \), which is used to indicate the probability that a node can successfully deliver a package to another. An example is the number over each link represents its corresponding transmission success ratio and \( v_0 \) is the sink. For convenience, the WSNs considered under the DNM are called Deterministic WSNs, whereas the WSNs considered under the PNM are called Probabilistic WSNs. When \( \ell_{ij} = 1 \), DNM can be viewed as a special case of PNM.

On the other hand, all the aforementioned works did not consider the load-balance factor when they construct a DAT. Without considering balancing the traffic load among the nodes in a DAT, some heavy-loaded nodes may quickly exhaust their energy, which might cause network partitions or malfunctions. For instance, for aggregating the sensing data from 8 different nodes to the sink node \( v_0 \), a...
shortest-path-based DAT for the probabilistic WSN. The intermediate node $v_4$ aggregates the sensing data from four different nodes, whereas $v_7$ only aggregates one sensing data from $v_6$. For simplicity, if every link is always there and every node has the same amount of data to be transferred through the intermediate nodes with a fixed data rate, heavily-loaded $v_4$ must deplete its energy much faster than $v_7$.

From the intermediate nodes usually aggregate the sensing data from neighboring nodes in a shortest-path-based DAT. Actually, the number of neighboring nodes of an intermediate node is a potential indicator of the traffic load on each intermediate node. It is not the only factor to impact the traffic load on each intermediate node. The criterion to assign a parent node, to which data is aggregated for each node on a DAT, is also critical to balance traffic load on each intermediate node. We refer the procedure that assigns a unique parent node for each node in the network, as the Parent Node Assignment (PNA) in this paper. Evidently, with respect to load-balance, the PNA is the best, which also implies the LBDAT can extend network lifetime notably compared with the DATs, since the traffic load is evenly distributed over all the intermediate nodes.

In summary, the investigated problem in this paper is distinguished from all the prior works in three aspects. First, most of the current literatures investigate the DAT construction problem under the DNM, whereas our work is suitable for both DNM and PNM. Second, the load-balance factor is not considered when constructing a DAT in most of the aforementioned works. Finally, the DAT construction problem is our major concern, whereas the prior works focus on the aggregation scheduling problem. Therefore, in this paper, we explore the DAT construction problem under the PNM considering balancing the traffic load among all the nodes in a DAT.

III. NETWORK MODELS AND DATA AGGREGATION TREES

Under the Probabilistic Network Model (PNM), we model a WSN as an undirected graph $G=(V,E,P(E))$, where $V = V_s \cup \{v_0\}$ is the set of $n+1$ nodes, denoted by $v_i$, where $0 \leq i \leq n$. i is called the node ID of $v_i$ in the paper. E is the set of lossy links. $\forall v_i,v_j \in V$, there exists a link $(v_i, v_j)$ in G if and only if: 1) $v_j$ and $v_j$ are in each other’s transmission range and 2) $l_{ij} > 0$. For each link $(v_i, v_j) \in E$, $l_{ij}$ indicates the probability that node $v_i$ can successfully directly deliver a packet to node $v_j$; and $P(E) = \{ l_{ij} \mid (v_i,v_j) \in E, 0 \leq l_{ij} \leq 1 \}$. We assume the links are undirected, which means two linked nodes are able to transmit and receive information from each other with the same ij value. Because of the introduction of ij, we define the 1-Hop neighborhood and the h-Hop neighborhood. The physical meaning of 1-Hop Neighborhood is the set of the nodes that can be directly reached from node $v_i$.

Since load-balance is the major concern of this work, the measurement of the traffic load balance under the PNM is critical to solve the LBDAT construction problem.

We first define a novel metric called potential load to measure the potential traffic load on each node. The number of neighboring nodes of a node is a potential indicator of the traffic load on each node. It is not the only factor to indicate the potential traffic load on each node in probabilistic WSNs. For example, if $v_j = 0.5$, then the expected number of transmissions to guarantee $v_i$ to deliver one packet to $v_j$ is $1 \cdot 0.5 = 2$. The less the $l_{ij}$ value, the more potential traffic load on $v_j$ from $v_i$.

We solve the LBDAT construction problem in three phases in this paper. First, we construct a Load-Balanced Maximal Independent Set (LBMIS) and then we select additional nodes to connect the nodes in LBMIS, denoted by the Connected MIS (CMIS) problem. Finally, we acquire a Load-Balanced Parent Node Assignment (LBPNA). After LBPNA is determined, by assigning a direction of each link in the constructed tree structure, we obtain an LBDAT. We formally define the LBMIS, CMIS, LBPNA and LBDAT construction problems sequentially.

Taking the load-balance factor into consideration, we are seeking an MIS in which the minimum potential load of the nodes in the constructed LBMIS is maximized. In other words, the potential traffic load on each node in the LBMIS is as balance as possible. Now, we are ready to define the CMIS problem. For convenience, the nodes in set $M$ are called independent nodes, whereas, the nodes in set $C$ are called LBMIS connectors. The nodes in the set $G \setminus (M U C)$ are called leaf nodes. Furthermore $G \setminus (M U C)$, $v_i$ is also called a non-leaf node. Hence, the set of non-leaf nodes are denoted by $D = M U C$.

Constructing a load-balanced connected topology is just one part of the work to build an LBDAT. In order to measure the actual traffic load, one more important task needed to be resolved is how to do parent node assignment for leaf nodes in the network. Since the actual traffic load of each node in a DAT is depended on the number of its children, which are composed of leaf nodes and non-leaf nodes.

IV. LOAD-BALANCED DATA AGGREGATION TREE (LBDAT) FOR PNM

A tree structure is decided after the Load-Balanced Parent Node Assignment (LBPNA) A is produced, which includes LBPNA for non-leaf nodes and leaf nodes. By assigning a direction of each link in the constructed tree from the children node to the parent node, we obtain an LBDAT. We already illustrate how to find an parent node assignment for non-leaf nodes. We first formulate the LBPNA for leaf nodes as an Integer Linear Programming (ILP). Then, we present an approximation algorithm by applying the linear relaxation and random rounding technique. Finally, we exploit an example to illustrate how to build an LBDAT. We already illustrate how to find a parent node assignment for non-leaf nodes. We study the LBPNA for leaf nodes.

A. LBPA for Leaf Nodes:

As we have already known, constructing an arbitrary aggregation tree with the maximum lifetime is NP complete. Through similar proving procedure, it can be shown that LBPNA is also an NP-complete problem. In this subsection,
we first model LBPNA as an ILP. We define a binary variable $\beta_i$ to indicate whether the sensor $v_i$ is a non-leaf node or not. $\beta_i$ sets to be 1 if the sensor $v_i$ is a non-leaf node. Otherwise, $i$ sets to be 0. Additionally, we assign a random variable $ij$ for each link connecting a non-leaf node $v_i$ and a leaf node $v_j$ on the graph $G$ modeled from a probabilistic WSN, i.e.,

$$\xi_{ij} = \begin{cases} 1, & \text{if non-leaf node } v_i \text{ is assigned to } v_j \\ 0, & \text{otherwise} \end{cases}$$

Consequently, LBPNA can be formulated as an Integer Linear Programming (ILP) as follows:

$$\max \mathcal{G} = \min \left\{ \alpha \mid \sum_{v_j \in N(v_i)} \left[ \frac{B_i}{T} \cdot \frac{1}{l_{v_j}} \cdot \mathcal{G} \right] \forall v_i \in D \right\}$$

s.t.

$$\sum_{v_j \in N(v_i)} \beta_i \mathcal{G} = 1, \forall v_j \notin D$$

$$\xi_{ij} \in [0,1]$$

The objective function $\mathcal{G}$ is the minimum actual load among all the non-leaf nodes. The first constraint states that each leaf node can be allocated to only one non-leaf node, whereas the second constraint indicates that $\xi_{ij}$ is a binary variable. The number of leaf child nodes and the number of non-leaf child nodes both contribute to the actual load of a non-leaf node. The leaf child nodes of parent node $v_i$ can be represented by $v_j : \beta_i e_{ij} > 0$. The traffic load introduced by non-leaf children to $v_i$ is denoted by $\varphi_i$. The number of non-leaf child nodes of an independentparent node $v_i \in M$ is no more than 12. Whereas, the number of non-leaf child nodes of an LBMIS connector parent node $v_j \in C$ is no more than 4. Therefore, for simplicity, we assume that the total actual load of leaf children nodes is approximated to $12 \left[ \frac{B_i}{T} \right]$ i.e.

$$\sum_{v_j \in N(v_i)} \left( \frac{B_i}{T} \cdot \frac{1}{l_{v_j}} \right) \approx 12 \left[ \frac{B_i}{T} \right]$$

By relaxing variable, we get the relaxed formulation which falls into a standard Linear Programming (LP) problem, denoted by $LBPNA^{LP}$ as follows:

$$\max \mathcal{G} = \min \left\{ \alpha \mid \forall v_i \in D \right\}$$

s.t.

$$\sum_{v_j \in N(v_i)} \beta_i \mathcal{G} = 1, \forall v_j \notin D$$

$$\xi_{ij} \in [0,1]$$

In $LBPNA^{LP}$, $\alpha_i = \sum_{v_j \in N(v_i), v_j \notin D} \left( \frac{B_i}{T} \cdot \max \left( \frac{\beta_i e_{ij}}{l_{v_j}} \right) + \varphi_i \right)$ using max

$$\left\{ 1, \frac{B_i\xi_{ij}}{T} \right\}$$

is mainly because that if $v_j$ has some data been forwarded by $v_i$, $v_j$ must transmit at least one data packet to $v_i$ since data packets are the basic communication units in a WSN. Due to the relaxation enlarged the optimization space, the solution of LP LBPNA corresponds to an upper bound of the objective of ILPLBPNA.

### B. Approximation Schemes:

Given an instance of LBPNA modeled by the integer linear programming $LBPNA^{ILP}$, the sketch of the randomized approximation algorithm is shown in Algorithm 1. We summarize Algorithm 1 as follows: first, solve the relaxed linear programming $LBPNA^{LP}$ to get an optimal fractional solution, denoted by $\mathbf{s}^{*}$ (where $\mathbf{s}^{*} = (s_{ij1}^{*}, ..., s_{ijn}^{*}, ..., s_{m1}^{*}, ..., s_{mm}^{*})$) and then round to integers $\mathbf{\hat{s}}_{ij}$ by a random rounding procedure, which consists of five steps as shown in lines 2-13 of Algorithm 1.

1. Sort the $s_{ij}^{*}$ values in each row of $s_{ij}^{*}$ in the decreasing order and store the corresponding $j$ (v_j’s ID) in a two-dimensional array $A[n][m]$;
2. Set all $\mathbf{\hat{s}}_{ij}$ to be 0.
3. Start from the first row in the sorted array $A$. If there is no parent node assigned to $v_i$ in its 1-hop neighborhood, then let $\mathbf{\hat{s}}_{ij}$ with probability $s_{ij}^{*}$.
4. Then, go to the next row in $A$.
5. Repeat steps (3) till reach the end of array $A$.
6. Repeat steps (3) and (4) for times, where $k = \left( \delta \min e_{ij} \mid 1 \leq i \leq n, 1 \leq j \leq m, e_{ij} > 0 \right) + \delta$ is any constant satisfying $0 < \delta < 1$.

### C. Algorithm 1: Approximation Algorithm for LBPNA:

1. Solve $LBPNA^{LP}$. Let $\mathbf{s}^{*}$ be the optimum solution.
2. Sort the $s_{ij}^{*}$ values in each row of $s_{ij}^{*}$ in the decreasing order and then store the corresponding $j$ (v_j’s ID) in a two-dimensional array denoted by $A[n][m]$;
3. $\mathbf{\hat{s}}_{ij} = 0$;
4. while $k \leq k = 6 \log(n) / \delta^2 \min |e_{ij}^{*} | 1 \leq i \leq n, 1 \leq j \leq m, e_{ij}^{*} > 0 |$ do
5. $k=0, l=0$;
6. while $k < n$ do
7. $i = k$;
8. while $l < m$ do
9. $l = A[k][l]$;
10. if $v_j \in N_1(v_i)$ and $\mathbf{\hat{s}}_{ij} = 0$ then
11. $\mathbf{\hat{s}}_{ij} = 1$ with probability $s_{ij}^{8}$;
12. break;
13. $k = k + 1$;
14. return $\left( \mathbf{\hat{s}}, \delta \right) = \min \{ \alpha_i \mid \forall v_i \in D \}$;
The time complexity of Algorithm 1 is $Q(n \log(n))$. The complexity of the the first step (line 2 in Algorithm 1) is $Q(n \log(n))$, since it is a sorting process. The time complexity of the random rounding procedure (line 4 to line 13 in Algorithm 1) is $Q(n \log(n))$. This is because the procedure, setting $c_{ij} = 1$ with probability $ij$, is repeated times. Moreover,

$$k = \left\lceil \frac{6 \log(n)}{\delta^2 \min \{|\xi_{ij}| \mid 1 \leq i \leq n, 1 \leq j \leq m, \xi_{ij} \geq 0\} \right\rceil$$

Hence, the time complexity of Algorithm 1 is $Q(n \log(n))$. Next, the correctness of the proposed approximation algorithm 1 is proven. Now, the probability that a leaf node is not assigned a parent non-leaf node in its 1-hop neighborhood after the random rounding is $\frac{1}{n^2}$, which implies $Pr[a \text{ leaf node has no neighboring non-leaf node}] \leq n^2 \left(\frac{1}{n^2}\right) = \frac{1}{n^2}$.

Similarly, according to the Borel-Cantelli Lemma, the above probability is 0 almost surely, which implies that it is almost sure that every leaf node is assigned a parent nonleaf node in its 1-hop neighborhood after executing Algorithm 1.

V. ISSUES ON LBDAT

Data Aggregation Trees (DATs) are used to gather and aggregate data values from the sensor nodes. DATs are constructed to fulfill different user requirements under the Deterministic Network Model (DNM). Probabilistic Network Model (PNM) is built with probabilistic lossy links. Load-Balanced Data Aggregation Tree (LBDAT) is constructed under the PNM. Load-Balanced Maximal Independent Set (LB MIS), Connected MIS (CMIS) and LBDAT construction problems are solved in the data gathering scheme. Approximation algorithm is used to search for an LB MIS. Load Balanced Parent Node Assignment (LBPNA) is performed using the randomized approximation algorithm. LBDAT is constructed by assigning a direction of each link in the tree from the children node to the parent node. The following issues are identified from the LBDAT Model.

- Centralized LBDAT construction model
- Correlation between independent sets are not considered
- LBDAT is not adapted for Deterministic Network Model (DNM)
- Monitoring area factors are not considered

VI. DATA GATHERING SCHEME FOR DETERMINISTIC AND PROBABILISTIC NETWORK MODELS

The data aggregation tree construction process is designed to collect and aggregate data values from sensors. Deterministic and probabilistic network models are supported by the system. Distributed DAT construction scheme is used in the system. The system is divided into six major modules. They are WSN construction, coverage analysis, data capturing process, aggregation trees for PNM, aggregation trees for DNM and data request process.

Sensor network construction is performed in WSN construction module. Sensing and transmission coverage are analyzed in the coverage analysis module. Data capturing process module is designed to perform environment monitoring process. Load balanced aggregation tree is also constructed for probabilistic network model environment. Aggregation tree construction is performed with load balancing in deterministic network model. User data requests are processed in the data query request process module.

A. WSN Construction:

Sensor network and node properties are collected from the user. Sensor network is deployed with node count and network area details. Node name, coverage details and data capture type information are collected for each node. Node energy level and bandwidth status are separately maintained for each node.

B. Coverage Analysis:

Sensing and transmission coverage details are used in the coverage analysis. Nodes within the overlapped coverage are assigned as dominating sets. Nodes out of overlapped coverage are assigned as independent sets. Sink nodes are selected with the resource details.

C. Data Capturing Process:

Environment monitoring and data sensing is carried out under the data capture process. Captured data values are updated into the local storages. Captured data values are updated with time details. The data values are transferred to the sink node.

D. Aggregation Trees for PNM:

The Load Balanced DAT construction process is designed for the Probabilistic Network Model (PNM). Load balanced parent node assignment process selects the parent node for the DAT. Child nodes is assigned to the parent node with load level details. Centralized and distributed LBDAT construction methods are used in the system.
E. Aggregation Trees for DNM:
The LBDAT is constructed for the Deterministic Network Model (DNM). Centralized and distributed schemes are used for the DAT construction process. Data aggregation tree is constructed with parent and child nodes. Parent node is selected with reference to the coverage and energy level details.

F. Data Request Process:
Data collection is carried out in the data query process. User query values are transferred to the sink nodes. Sensed data and aggregated data values are produced as query response for the users. Network area monitoring process is analyzed with query history details.

VII. CONCLUSION
Data Aggregation Trees (DAT) is constructed to handle data collection in Wireless Sensor Networks. Load Balanced DAT construction model is designed for the Probabilistic Network Model (PNM). The system is enhanced to support Distributed LBDAT construction process. The system is adapted for both Deterministic and Probabilistic Network Models. Data transmission load is reduced by the system. The system improves the network lifetime for DNM and PNM. Network traffic is minimized by the system. Data collection latency is controlled by the system for different network coverage models.

REFERENCES