Maximizing Lifetime Data Aggregation in Multi-sink Wireless Sensor Networks with Unreliable Vehicle Communications

Zhihuang Su[1], Yongzao Chen[1], Hongju Cheng[1], and Naixue Xiong[2,*]

[1] College of Mathematics and Computer Science, Fuzhou University, Fuzhou, FJ 350108, China
zhihuangsu@126.com, chenyongzao178@163.com, cscheng@fzu.edu.cn
[2] Department of Computer Science, Colorado Technical University, CO 80907, USA
nxiong@coloradotech.edu

Abstract. In this paper, we study the problem of maximizing lifetime data aggregation in multi-sink wireless sensor networks with unreliable vehicle communication environment. Firstly, we analyze the communication between adjacent nodes, and present the optimal emission radius that can guarantee the minimum expected energy consumption. Secondly, we discuss the problem that how sensor nodes choose the sink node to send message. Thirdly, we propose the Tree-based topology Data Aggregation algorithm (TDA) based on the energy consumption balancing and the Directed Acyclic Graph based Data Aggregation algorithm (DAGDA) to improve the data acceptance probability. The simulation results show that our algorithms can extend network lifetime effectively.

Keywords: Wireless Sensor Network, Maximizing Lifetime, Data Aggregation, Unreliable Vehicle Communication Environment.

1 Introduction

Wireless sensor network is consisting of numerous micro-sensor nodes which are inexpensive and low energy consumption [1]. Sink nodes in network are responsible for accepting and processing data collected by sensor nodes. But, with the increasing network size, deploying a single sink node in network results in some nodes transmitting data to the sink node over a longer distance, which shorts network lifetime. In this case, deploying multiple sink nodes in network can be considered [2-4].

The in-network data aggregation [5-6] has been proposed and has become an important technology. The data collection operations from all nodes in network can be performed with the data aggregation tree, where sink node is a root node. A problem faced by the construction of data aggregation tree is how to construct the tree in an energy efficient way.

* Corresponding author.

R.C.-H. Hsu and W. Shangguang (Eds.): IOV 2014, LNCS 8662, pp. 180–189, 2014.

In this paper, we study the problem of maximizing lifetime data aggregation problem in multi-sink wireless sensor network with unreliable vehicle communication environment. To resolve the problem, we firstly propose a Tree-based topology Data Aggregation algorithm (TDA) to extend network lifetime, then propose the Directed Acyclic Graph based Data Aggregation algorithm (DAGDA) to improve the data acceptance probability.

2 System Model and Preliminary Work

2.1 Probability Model and Energy Cost Model of Accepted Signal

Stojmenovic et al. [7] apply the log-normal shadow fading model to derive the probability $p(d)$ for accepting a packet successfully as a function of distance d.

$$P(d) = \begin{cases} 1 - 0.5(d / R)^{qa}, & d < R \\ 0.5(2 - d / R)^{qa}, & R \le d < 2R \\ 0, & d \ge 2R \end{cases}, \tag{1}$$

where qa is the power attenuation factor, the value of q associated with the size of a data packet.

n this paper, we also adopt the model to quantify the acceptance probability [7], where $q = 2$ and $a \in [2, 6]$.

The energy of a node mainly consumed in the wireless communication during data aggregation, and we analyze the energy cost model for the aggregation operation of a node as follows:

$$\begin{aligned} E'_{Tx}(k,R) &= k * E_{amp} * R^a + m \\ E_{Rx}(k) &= k * Eelec \end{aligned}, \tag{2}$$

where k denotes the size of a data packet, $E_{amp} = 100pJ/bit/m^2$, $E_{elec} = 50nJ/bit$, R is the emission radius, a is the path loss exponent, m denotes the energy consumed in signal processing, $E_{Tx}(k,R)$ and $E_{Rx}(k)$ respectively denote the energy consumed in sending and accepting data packet. The energy cost model for the aggregation operation of a node can be simplified as follows:

$$\begin{aligned} E_{Tx}(k,R) &= k * E_{amp} * R^a \\ E_{Rx}(k) &= k * Eelec \end{aligned}. \tag{3}$$

2.2 Analysis of Energy Cost in Point to Point Communication

We use the average number of transmissions to quantify the number of a data packet need to be retransmitted, which can efficiently indicate the degree of randomness in a real network [8 - 9]. For a link (u, v) with reliability P ($P \neq 0$), node u need to transmit a data $1 / P$ times to ensure that node v can accept the data. Assuming that the emission radius radio is $\delta = R / d$, $q = 2$ and combing formula (1) and formula (3), it was found that the average energy cost of a data packet is

$$E_{Tx}=\begin{cases} kE_{amp}\delta^a d^a/(1-0.5\delta^{-2a}), & \delta>1 \\ 2kE_{amp}\delta^a d^a/(2-\delta^{-1})^{2a}, & 0.5<\delta\le 1 \end{cases} . \quad (4)$$

For the two cases of $\delta > 1$ and $\delta \le 1$ as show in formula (5), according to the function monotonic, the minimal average transmission cost can be described in formula (6)

$$\delta=\begin{cases} (3/2)^{1/(2a)}, & \delta>1 \\ 1, & 0.5<\delta\le 1 \end{cases} . \quad (5)$$

$$E_{\min Tx}=\begin{cases} (3/2)^{3/2}kE_{amp}d^a, & \delta>1 \\ 2kE_{amp}d^a, & 0.5<\delta\le 1 \end{cases} . \quad (6)$$

From formula (5) and formula (6), we can see that the minimal average transmission cost between any two node is $(3/2)^{3/2}\, k\, E_{amp}\, d^a$ when the emission radius is $(3/2)^{1/(2a)}\, d$. The minimal average transmission cost can be described in formula (7).

$$E_{\min Tx}(k,d)=(3/2)^{3/2}kE_{amp}d^a,\ \overline{R}=(3/2)^{1/(2a)}d\ . \quad (7)$$

From the above analysis, we can know that there always exits a minimal average transmission cost between any two nodes in network. Without confusion, in this paper we use the transmission cost to represent the minimal average transmission cost.

2.3 Problem Formulation

We consider a wireless sensor network with sensor nodes equipped with an omnidirectional radio transmission device. Each node in network can receive data from any direction and send data to any direction. The emission radius of nodes is adjustable. The position of nodes and sink nodes are stationary and each node can obtain its own position information. In this paper, we use V to denote the set of all nodes which consists of the set of sink nodes and the set of sensor nodes $V - S$.

The wireless sensor network collects data in a periodical manner. Each node has a finite initial energy and Sink nodes are powered nodes. A node dead when its energy is depleted. Each node in network should be assigned a corresponding receiving node and this process is called aggregation scheduling. An aggregation scheduling scheme may contains more than one scheduling. In this paper, we study the problem of how to optimize the data scheduling scheme to maximize network lifetime. The network lifetime defined in this paper is the round of aggregations during which the fraction of survived nodes c remains above a given threshold [10 - 11].

2.4 Sink Node Selecting Problem

For the minimum-time aggregation scheduling problem in multi-sink sensor networks, each node just cares the correctly received of data by sink nodes. The node that communicates directly with the sink node should send data to the nearest sink node to ensure the minimal transmission cost. For those nodes that need to transmit data via multi-hop sensor to the sink node, the data may not be sent to the nearest sink node which can increase the choice of the receiving node.

3 Tree-Based Topology Data Aggregation Algorithm (TDA)

This section firstly describes the steps of constructing data aggregation trees and the reconstructing procedures, and then introduces the complete process of TDA algorithm.

3.1 Construction of Data Aggregation Trees

The process of constructing data aggregation trees is to select nodes from N_2 and add them into N_1. We can enable a node in N_2 selects a nearest node in N_1 as parent node to ensure the minimal transmission cost, calculate the value of residual aggregation rounds r_{ij} according to formula (8), and let the nodes with large r_{ij} have a priority to join the data aggregation trees, which is contributed to the nodes with low residual energy and large transmission cost become leaf nodes, thus perform more aggregation rounds.

$$r_{ij} = e_i / E_{Txij}. \tag{8}$$

The main idea of constructing a data aggregation trees is as follows: the initial T is empty and the initial N_1 is the set of sink nodes S, we search the parent node and calculate the residual aggregation rounds of each node in N_2 during current round; each time we select a node with maximum aggregation rounds from N_2 and its corresponding parent node to join the data aggregation trees, while updating the maximum aggregation rounds of nodes in N_2 and their corresponding parent node, repeat this process until the set N_2 is empty. The detail process is as follows.

Process 1. Construction of data aggregation trees T.

```
T = ∅, N1 = S, N2 = V − S;
Selecting a parent node pi for each i in N2, the resi-
dual aggregation rounds of i is ri = ripi;
while (N2 ≠ ∅) do
  Selecting a node i ∈ N2 with maximum ri;
  T = T ∪ (i, pi), N2 = N2 − {i}, N1 = N1 + {i};
  for (j1 = 1 to |N2|)
       The j1-th node is node j;
    if (dji < djpj) then pj = i, rj = rji;
  end for
end while
```

3.2 Reconstruction of Data Aggregation Trees

To let the nodes that are easy to die have enough residual energy in later data aggregations and avoid the frequent of reconstructing data aggregation trees, the constructed data aggregation trees calculate the stage aggregation rounds according to formula (9).

$$F(T) = \max\{freq * \min(e_i / (E_{Txi} + cn_i * E_{Rxi})), \mathrm{LR}\}, \tag{9}$$

where $freq$ is the coefficient of the stage aggregation rounds which in the range of (0,1], cn_i is the number of child node of node i, $E_{Txi} + cn_i E_{Rxi}$ is the energy cost E_i of node i. LR denotes the minimum stage aggregation rounds. In addition, it also needs to reconstruct the data aggregation trees if there are nodes dead.

The information, such as residual energy of nodes, should be updated before performs the reconstruction of data aggregation trees.

Process 2. Updating Information before Reconstruction.

```
Calculate the stage aggregation rounds F(T) of T;
for (each sending node i in T)
  ar_i = ar_i + F(T), e_i = e_i - F(T) * E_i;
  if (node i dead), then V = V - {i}, dn = dn + 1, set
   the state of i as dead;
end for
```

3.3 Algorithm Process

The TDA algorithm performs the aggregation operation through data aggregation trees, and reconstructs the data aggregation trees after a certain aggregation rounds. The process will continue until the network dead, i.e. the fraction of dead nodes c larger than a given threshold. The algorithm is described as follows.

Algorithm 1. Data Aggregation algorithm (TDA).

```
Input: V, S, The position information of sensor nodes
Output: Scheduling scheme and network lifetime
  dn = 0, e_i = E_0, for i ∈ V - S, j ∈ V, calculate the
  E_ij with formula (7), all nodes' state is Un-dead;
  while (dn < n * c) do
    Construct trees T by calling the Process 1;
    Calling the Process 2 and the network perform F(T)
     rounds data aggregation according to T;
  end while
  Obtain the scheduling scheme and network lifetime.
end
```

4 Directed Acyclic Graph Based Data Aggregation Algorithm (DAGDA)

We can improve the acceptance probability by the way that multiple nodes simultaneously receiving the data. Based on this, we propose the Directed Acyclic Graph based Data Aggregation algorithm (DAGDA) to improve the data acceptance probability.

4.1 Analysis of Data Transmission Probability in Directed Acyclic Graph

For a node i, the probability of the currently transmitting data be accepted is P_i. We can update the accept probability by formula (10) when assigning a new node j to receive the data sent by node i.

$$P_i = 1 - (1 - P_i)(1 - P(d_{ij} \tag{10}$$

4.2 Strategies to Avoid Loops

In order to avoid the presence of loops, we use the strategy in opportunistic routing: nodes in network send forward the data they received [12 - 13]. Specifically as follows: we establish a transmission link for each node in network according to a certain rules in sequence; the corresponding graph *G*' corresponding to the established links is no loop; in order to judge whether link (*i*, *j*) is a viable link, we can firstly assume that join the link into *G*' and the new graph can be denoted as *G*'', then start from *j*, depth-first search in *G*'', and the link (*i*, *j*) is not a viable link if we can traversal node *i* which means that there is a loop in *G*'', otherwise we establish a new link (*i*, *j*) and update *G*' into *G*''. The time complexity of the operation that judge whether there is a loop for a given node is $O(n)$.

4.3 Strategies to Avoid Loops

DAGDA starts from the point of balancing the energy consumption: we let these nodes die easily when they communicate directly with the sink node have a priority to establish scheduling thus extend the network lifetime. The DAGDA includes two processes: the establishment of aggregation scheduling and the reconstruction of aggregation scheduling.

(1) Establishment of aggregation scheduling. This paper select the node which nearest to the given node and its join would not result in the new graph *G*'' containing loops as one of the receiving nodes and determining the optimal emission radius. Then, we make the node that is equipped with two times of emission radius and would not form loop to become receiving nodes of sending nodes. We sort nodes in network with increasing order of the distance to the sending node, and select the receiving node that does not form a loop by a first adaptation method. In addition, we let the nodes die easily when communicate directly with the sink node have a priority to be selected as receiving node to die soon. The specific process is as follows:

Process 3. Process of constructing the aggregation scheduling *S* on the basis of directed acyclic graph.

```
S = Ø;
for (i1 = 1 to |V|)
  for the i1-th node i, calculate the distance between i
   and sink node dis' = min dij(j ∈ S) and corresponding
   Eis';
end for
Sort nodes in V with increasing order of aggregation
rounds;
for (i1 = 1 to |V|)
  The i1-th node i, search j which nearest to i and
   without loop, S = S ∪ (i, j);
  Calculating Rij;
  Sort nodes whose distance to node i is smaller than
2Rij;
```

```
    for (k1 = 1 to |Nhi|)
      The k1-th node is k;
      if join (i, k) and there is no loop in S, then S =
      S ∪ (i, k) and Update i's acceptance probability;
    end for
  end for
```

(2) Reconstruction of aggregation scheduling. We also consider the reconstruction of aggregation scheduling in DAGDA, where the stage aggregation rounds can be determined by formula (9). The DAGDA executed the two steps that establishment of aggregation scheduling and reconstruction of aggregation scheduling repeatedly, until the network dead. The algorithm is described as follows:

Algorithm 2. Directed Acyclic Graph based Data Aggregation algorithm (DAGDA).

```
Input: V, S, Position information of all nodes
Output: Scheduling scheme, network lifetime
  dn = 0, ei = E0, for i ∈ V - S, j ∈ V, calculate the
  Eij according to formula (7); set the state of all
  nodes as Un-dead;
  while (dn < n * c) do
    Calling the Process 3 to establish the scheduling;
    Calling the Process 3 to update the node information
    and the network perform F(T) rounds data aggregation
    according to the established scheduling;
  end while
  Obtain the network lifetime
end
```

5 Simulation Results and Analysis

We adopt MATLAB as the platform tool and demonstrate detailed simulation experiments to evaluate the performance of the above algorithms. In this paper, we assume that the distance between nodes is far away and the energy mainly consumed in communication. The target area size is 1000m × 1000m. The Size of data packet is 100bits. The value of c is E_0. The value of a is 2. All experimental results are obtained by 10 different network topologies and calculated with the average value.

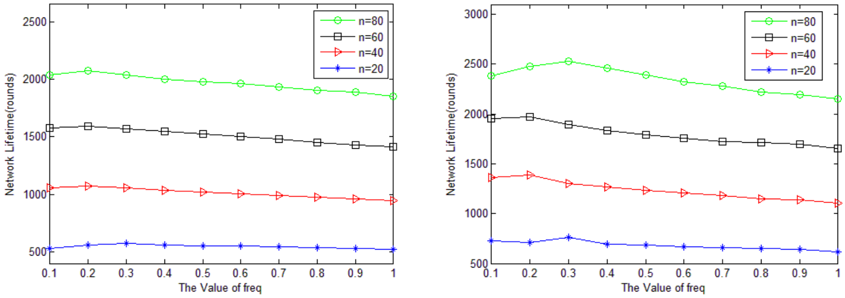
5.1 Analysis of the Value of *freq*

The value of *freq* determines when to reconstruct an aggregation scheduling. Its value affect the network lifetime and we analysis the value by experiments.

Figure 1 study the impact of network size on network lifetime and we respectively give the experiment results of TDA and DAGDA. In simulation environment setup, the number of sink nodes is 3, the value of c is 0.2, the value of *freq* is select from [0.1, 1]. Experiment results show that, as *freq* increases, both the network lifetime of

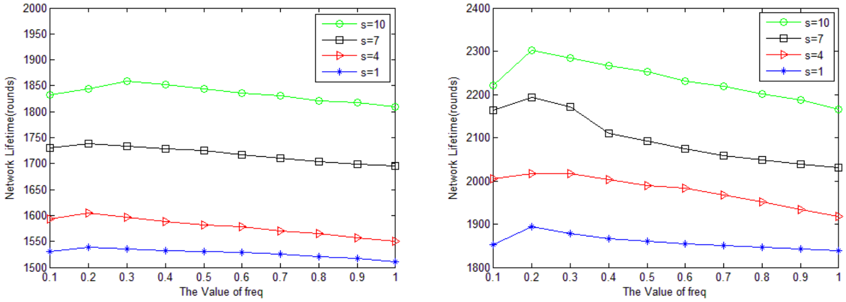
TDA and DAGDA will first increase and then decrease, and the network lifetime is large while the *freq* is small.

Fig. 2 study the impact of the number of sink nodes on network lifetime, where the number of nodes in network is 60, the value of *c* is 0.2. It is easy to know that, the trends of network lifetime in Fig. 2 are the same as Fig. 1. The network lifetime is large when the value of *freq* is 0.2, thus we set *freq* as 0.2 in the follow experiments.



(a) TDA (b) DAGDA

Fig. 1. Impaction of *freq* on network lifetime, where $n \in \{20, 40, 60, 80\}$



(a) TDA (b) DAGDA

Fig. 2. Impaction of *freq* on network lifetime, where $s \in \{1, 4, 7, 10\}$

5.2 Algorithms Performance Analysis

In this paper, to applied these two strategies in [12 - 13] to our problem, we modify them as follows: The VDA apply the way of our construction and reconstruction data aggregation tree; DDA uses the acceptance probability model of our proposed to designate the nodes within two times of emission radius, and to receive data sent by sending node after the emission radius of nearest node is determined.

The first set of experiment compare the network lifetime through the following parameters: the number of sensor nodes is 60, the number of sink node is 3 and the value of *c* varies from 0.05 to 0.5. As we can see from Fig. 3, the network lifetime increased with the value of *c* increasing. The performance of DAGDA and DDA are better than the tree-based topology algorithms and the performance of DAGDA is better than DDA regardless of the value of *c* varies. In those tree-based topology algorithms, the performance of TDA is better than the VDA.

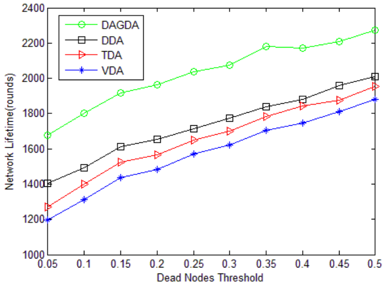


Fig. 3. Impact of *c* on network lifetime

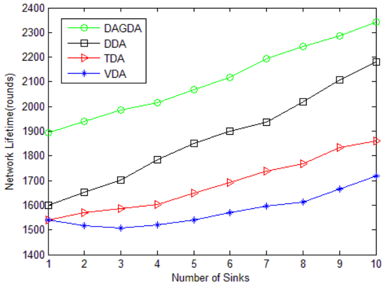


Fig. 4. Impact of number of sink nodes

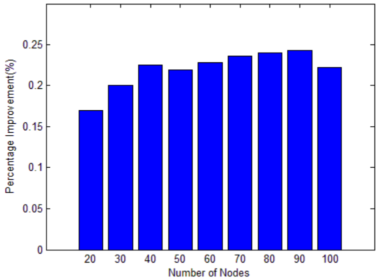


Fig. 5. Improvement of average acceptance probability with DAGDA

The third set of experiment compare the network lifetime through the following parameters: the number of nodes is 60, the value of *c* is 0.2 and the number of sink nodes in network varies from 1 to 10 and the result is shown in Fig. 4. We can see that, the network lifetime increased with the increasing of sink nodes in network, but the incensement of network lifetime of VDA is small.

Compared with TDA algorithm, the DAGDA can extend the network lifetime. Thus, we compare the average data acceptance probability between DAGDA and TDA through the following parameters: the number of sink nodes is 3, the number of sensor nodes varies from 20 to 100, the results as are shown in Fig. 5. We can see that, the average data acceptance probability of DAGDA better than TDA.

6 Conclusion

In this paper, we study the problem of maximizing network lifetime data aggregation in multi-sink wireless sensor networks with unreliable vehicle communication environment. We propose the Tree-based topology Data Aggregation algorithm (TDA) and the Directed Acyclic Graph based Data Aggregation algorithm (DAGDA) to extend network lifetime. Experimental results on synthesized data sets show that our proposed algorithms can significantly extend the network lifetime compared with related works.

Acknowledgement. This work is supported by the National Science Foundation of China under Grand No. 61370210 and the Development Foundation of Educational Committee of Fujian Province under Grand No. 2012JA12027.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, Y.: A survey on sensor networks. IEEE Communications Magazine 40(8), 102–114 (2002)
2. Xue, Y., Cui, Y., Nahrstedt, K.: Maximizing lifetime for data aggregation in wireless sensor networks. Mobile Networks and Applications 10(6), 853–864 (2005)
3. Wu, Y., Mao, Z., Fahmy, S., Shroff, N.: Constructing maximum-lifetime data gathering forests in sensor networks. IEEE/ACM Transactions on Networking (TON) 18(5), 1571–1584 (2010)
4. Liu, S.Y., Huang, C.C., Huang, J.L., Hu, C.L.: Distributed and localized maximum-lifetime data aggregation forest construction in wireless sensor networks. In: IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 655–660. IEEE (2012)
5. Rajagopalan, R., Varshney, P.K.: Data aggregation techniques in sensor networks: A survey. IEEE Communications Surveys 6(4), 48–63 (2006)
6. Krishnamachari, L., Estrin, D., Wicker, S.: The impact of data aggregation in wireless sensor networks. In: Proceedings of 22nd International Conference on Distributed Computing Systems Workshops, pp. 575–578. IEEE (2002)
7. Tan, H.O., Korpeoglu, I., Stojmenovic, I.: Computing localized power-efficient data aggregation trees for sensor networks. IEEE Transactions on Parallel and Distributed Systems 22(3), 489–500 (2011)
8. Shi, L., Fapojuwo, A.O.: TDMA scheduling with optimized energy efficiency and minimum delay in clustered wireless sensor networks. IEEE Transactions on Mobile Computing 9(7), 927–940 (2010)
9. Xiong, N., Huang, X., Cheng, H., Zheng, W.: Energy-Efficient algorithm for broadcasting in Ad Hoc wireless sensor networks. Sensors 13(14), 4922–4946 (2013)
10. Aziz, A., Sekercioglu, Y.A., Fitzpatrick, P., Ivanovich, M.: A survey on distributed topology control techniques for extending the lifetime of battery powered wireless sensor networks. IEEE Communications Surveys and Tutorials 15(1), 121–144 (2013)
11. Cheng, H., Su, Z., Zhang, D., Lloret, J., Yu, Z.: Energy-Efficient Node Selection Algorithms with Correlation Optimization in Wireless Sensor Networks. International Journal of Distributed Sensor Networks 2014, Article ID 576573, 1–14 (2014)
12. Chachulski, S., Jennings, M., Katti, S., Katabi, D.: Trading structure for randomness in wireless opportunistic routing. In: Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp. 169–180 (2007)
13. Zeng, K., Yang, J., Lou, W.: On energy efficiency of geographic opportunistic routing in lossy multihop wireless networks. Wireless Networks 18(8), 967–983 (2012)