Computer Networks 105 (2016) 99–110

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

# On maximizing the lifetime for data aggregation in wireless sensor networks using virtual data aggregation trees

Ngoc-Tu Nguyen[a], Bing-Hong Liu[a,*], Van-Trung Pham[b], Yi-Sheng Luo[a]

[a] Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, 415, Chien Kung Rd., Kaohsiung 80778, Taiwan
[b] Department of Information Technology, Pham Van Dong University, 99, Hung Vuong Rd., Quang Ngai 570000, Viet Nam, Vietnam

ARTICLE INFO

*Article history:*
Received 16 November 2015
Revised 22 April 2016
Accepted 30 May 2016
Available online 30 May 2016

*Keywords:*
Wireless sensor network
Data aggregation
NP-complete

ABSTRACT

Data gathering is a basic requirement in many applications of wireless sensor networks (WSNs). Because the energy of sensors is limited, the data-gathering mechanism must be carefully designed to save the energy consumption of sensors to prolong the network lifetime. Recently, many researchers have studied gathering data efficiently in WSNs to minimize the total energy consumption when a fixed number of data are allowed to be aggregated into one packet. However, when the total energy consumption is minimized, the energy consumption of sensors for data gathering cannot be guaranteed to be balanced, and thus, the network lifetime cannot be guaranteed to be maximized. This motivates us to study the problem of scheduling virtual data aggregation trees to maximize the network lifetime when a fixed number of data are allowed to be aggregated into one packet, termed the Maximum Lifetime Data Aggregation Tree Scheduling (MLDATS) problem. The MLDATS problem is shown to be NP-complete in the paper. In addition, a local-tree-reconstruction-based scheduling algorithm (LTRBSA) is proposed for the MLDATS problem. We use simulations to evaluate and demonstrate the performance of the LTRBSA when the sink has 2-hop, 3-hop, and all information in the networks. Simulation results show that the LTRBSA of using sink's 3-hop information provides comparable performances to that of using all information in the networks, and outperforms other methods proposed in the simulation.



## 1. Introduction

Wireless sensor networks (WSNs) are composed of many wireless sensors that are deployed in a wide range of field. Recently, many applications of WSNs have been developed, such as environmental monitoring, surveillance, fire detection, and seismic monitoring [1–5]. One of the most important operations in these applications of WSNs is data gathering. The data-gathering mechanism collects sensing data from sensors and reports to a specific node, called a sink. Because the energies of sensors are limited, efficiently gathering data is a major challenge in WSNs. In this paper, we study efficient data gathering to maximize the lifetime of WSNs.

In some WSNs [6–8], some sensors, termed source nodes, are responsible for sensing the environment and generating sensing data, and some sensors, termed relay nodes, are allowed to forward received data to others. When any source node runs out of the energy, the area covered by the source node cannot be monitored anymore, and thus, the end of the lifetime for the WSN is assumed to be reached because the network cannot provide adequate quality of service for applications [9–11]. To efficiently prolong the network lifetime, our idea is to use a set of data aggregation trees that work in turn for different time periods such that the network lifetime is maximized. Hereafter, the data aggregation trees used in different time periods are called virtual data aggregation trees. Take Fig. 1, for example. In Fig. 1, node $s$ is the sink, and the number close to a node denotes the node's remaining energy. For simplicity, we assume that the energy consumption for any non-leaf node is 2 per round to receive and forward data, where data are assumed to be generated and reported to the sink in one round. In addition, we also assume that the energy consumption for the node that is a source node and also a leaf in the tree is 1 per round to transmit data. While we are given a WSN and a tree $T_1$ as shown in Fig. 1(a), $T_1$ can survive at most 3 rounds because non-leaf node $b$ consumes energy $2 \times 3 = 6$ for the three rounds and has remaining energy 1 after the rounds. The remaining energies of nodes after the three rounds can be seen in Fig. 1(b). If we use the tree $T_2$, as shown in Fig. 1(b), to be the data aggregation tree in the network, $T_2$ can survive another 2 rounds. Note that

* Corresponding author. Fax: 886 7 3811182.
*E-mail addresses:* kuastu2011@gmail.com (N.-T. Nguyen), bhliu@kuas.edu.tw, bhliu@cc.kuas.edu.tw (B.-H. Liu), pvtrung@pdu.edu.vn (V.-T. Pham), a15641854@hotmail.com (Y.-S. Luo).

http://dx.doi.org/10.1016/j.comnet.2016.05.022
1389-1286/

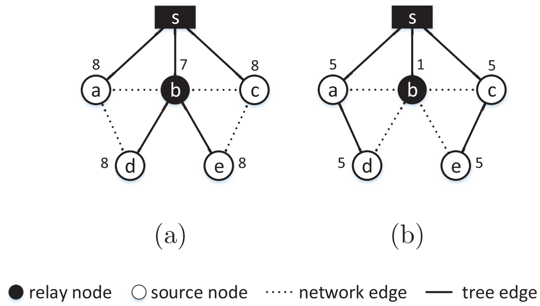


**Fig. 1.** Brief example of the network lifetime while using virtual data aggregation trees, where the energy consumption per round for the node, which is a source node and also a leaf in the tree, is 1, that for any non-leaf node is 2, and that for others is 0. (a) and (b) show the data aggregation trees $T_1$ and $T_2$, respectively.

node $b$ in Fig. 1(b) does not consume any energy because it is a relay node, and has no data to receive or transmit. Also note that node $a$ has remaining energy $5 - 2 \times 2 = 1$ after the two rounds, and cannot complete the next round. Therefore, we have that the network can survive 5 rounds when $T_1$ and $T_2$ are used in turn to be the data aggregation trees, which is better than that using only one data aggregation tree. This motivates us to study the problem of scheduling virtual data aggregation trees to maximize the network lifetime when a fixed number of data are allowed to be aggregated into one packet, termed the Maximum Lifetime Data Aggregation Tree Scheduling (MLDATS) problem.

The remaining sections of the paper are organized as follows. In Section 2, related work is studied. The definition and the hardness of the MLDATS are formally illustrated in Section 3. In addition, a local-tree-reconstruction-based scheduling algorithm is proposed in Section 4 to find virtual data aggregation trees for the WSN. The analysis is provided in Section 5. The performance of our proposed method is evaluated in Section 6. Finally, the paper is concluded in Section 7.

## 2. Related work

In WSNs, some studies address on transmitting data simultaneously without data collision [12,13]. In [12], a localized time division multiple access (TDMA) medium access control (MAC) protocol is proposed for high-data-rate sensor networks. In the proposed method, each sensor in networks is assigned time slot for data transmission such that no confliction will be occurred between sensors. In [13], a distributed link scheduling algorithm is proposed such that all nodes in networks can be concurrently involved in slot reservation for data transmission without data collision.

The tree is a common structure that leads the data generated in WSNs to the sink. By using the tree structure, each node in the tree is responsible for forwarding its generated data and the data received from its child nodes to its parent node. In [14], Lee and Wong propose an overlay tree structure to prolong the lifetime of sensors in the network. In the tree, the nodes that have higher residual energies are selected as the parents to facilitate data collection. In addition, when sensors are no longer functional or network links are broken, the tree can be reconstructed by their proposed method. In [15], Luo et al. study the problem of selecting a maximum lifetime tree from a set of shortest path trees. In [16], Dasgupta et al. propose an approximation method that uses intelligent selection of trees to solve the maximum lifetime data collection problem in sensor networks. In [17], Wu et al. propose an approximation algorithm of constructing a spanning tree to prolong the network lifetime when a single sink exists. In [18], Chen et al. propose an adjustable load-balancing convergecast tree protocol for WSNs to prolong the network lifetime. In these studies, the data is directly forwarded via the tree without minimizing the size of the aggregated data, and therefore, it spends lots of energy for data gathering.

Because data aggregation techniques [19,20] have been widely applied in WSNs to minimize the size of the aggregated data, recently, researchers have studied efficiently gathering data when multiple data are allowed to be aggregated into one packet [21,22]. In [23], Madden et al. propose methods for improving the reliability and performance of retrieving data via a tree when basic database aggregates, including min, max, sum, average, and count, are used with grouping. In [24], Kalpakis et al. propose scheduling methods based on admission flow networks to maximize the network lifetime. In [25], Krishnamachari et al. use the data aggregation tree to model data-centric routing to yield energy-efficient dissemination. In [26], Wu et al. study the construction of a data-gathering tree to maximize the network lifetime.

Many researchers have studied efficiently gathering data in WSNs when a fixed number of data are allowed to be aggregated into one packet [27–29]. In [27], Liu and Cao propose solutions to design a fault-tolerant and energy-efficient distributed monitoring system in WSNs. In [28], Kuo and Tsai propose methods of constructing data aggregation trees such that the total energy cost for gathering data is minimized. In [29], Liu and Jhang propose novel data aggregation and routing structures for gathering data. With the structures, a distributed data scheduling algorithm is proposed to schedule data to nodes such that the energy cost for gathering data is minimized, when all data are aggregated at most once. However, minimizing the total energy cost for gathering data does not imply that the network lifetime is maximized.

## 3. Problem definition and its hardness

In this section, we first describe the network model for WSN that can be represented by a connected graph in Section 3.1. Based on the network model, we present the structure of a data aggregation tree used for data aggregation in WSNs and its lifetime in Section 3.2. Finally, the Maximum Lifetime Data Aggregation Tree Scheduling (MLDATS) problem and its hardness are developed in Section 3.3.

### 3.1. Network model

In this paper, the communication model in the wireless sensor network is assumed to be a unit disk graph model [30]. In the model, a sensor $u$ can receive messages from sensor $v$ if $u$ is within the transmission range of $v$. Hereafter, $u$ is said to be $v$'s neighboring node in the network if $u$ can receive messages from $v$. When all sensors have the same transmission range, the WSN can be represented as a connected weighted graph $G(V_G, E_G, w_G, \rho_G)$, where node $v \in V_G$ denotes a sensor in the networks, edge $(u, v) \in E_G$ represents that $u$ and $v$ can communicate with each other, $w_G(v)$ denotes the energy of $v$, and $\rho_G(v)$ is the number of units of raw data generated by $v$ to report to sink $s \in V_G$ per unit time, where the sink $s$ is a special node in the network that is responsible for data collecting, processing, and analysis. The data generated in a unit time have to be reported to the sink in the same unit time [31,32], which is called a working round in this paper. In addition, the nodes $v$ with $\rho_G(v) > 0$ are called source nodes; and the nodes $v$ with $\rho_G(v) = 0$ are called relay nodes hereafter. The relay node can receive data from other nodes and forward the data to the next node for reporting data to the sink. The source node can generate its own raw data for each working round and works like a relay node to help relay data. Fig. 2(a) shows an example of the connected graph representing a WSN, where the number of units of raw data generated by a node is shown as the right number in