Computer Networks 65 (2014) 73–83

Contents lists available at ScienceDirect

Computer Networks

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

Efficient distributed data scheduling algorithm for data aggregation in wireless sensor networks

Bing-Hong Liu *, Jyun-Yu Jhang

Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, 415, Chien Kung Rd., Kaohsiung 80778, Taiwan, ROC

ARTICLE INFO

Article history:
Received 30 September 2013
Received in revised form 13 February 2014
Accepted 10 March 2014
Available online 17 March 2014

Keywords:
Wireless sensor network
Data aggregation
NP-complete problem

ABSTRACT

With the rapid development of applications for wireless sensor networks, efficient data aggregation methods are becoming increasingly emphasized. Many researchers have studied the problem of reporting data with minimum energy cost when data is allowed to be aggregated many times. However, some aggregation functions used to aggregate multiple data into one packet are unrepeatable; that is, every data is aggregated only at most once. This problem motivated us to study reporting data with minimum energy cost subject to that a fixed number of data are allowed to be aggregated into one packet and every data is aggregated at most once. In this paper, we propose novel data aggregation and routing structures for reporting generated data. With the structures, we study the problem of scheduling data to nodes in the networks for data aggregation such that the energy cost of reporting data is minimized, termed MINIMUM ENERGY-COST DATA-AGGREGATION SCHEDULING. In addition, we show that MINIMUM ENERGY-COST DATA-AGGREGATION SCHEDULING is NP-complete. Furthermore, a distributed data scheduling algorithm is proposed accordingly. Simulations show that the proposed algorithm provides a good solution for MINIMUM ENERGY-COST DATA-AGGREGATION SCHEDULING.

1. Introduction

In recent years, many applications of wireless sensor networks have been developed [1,2], such as environmental monitoring, battlefield surveillance, and seismic monitoring. The most important function in these applications involves periodically collecting and reporting data generated by sensors to a specific node, called a sink, for further analysis and monitoring. Due to the sensors' limited power, collecting data efficiently is an important issue in wireless sensor networks. In this paper, we develop data collection in wireless sensor networks.

Data aggregation is an efficient method for data collection in wireless sensor networks. A well-known data aggregation is to use symmetric function [3,4], such as max, min, mean, and range, to aggregate data and save significant energy consumption of reporting data. Because data aggregation using symmetric function often retrieves and reports interests, the sink has only partial information generated in the wireless sensor network. However, in many applications, such as battlefield surveillance and seismic monitoring, collecting all the information generated by sensors is necessary for accurate and precise analysis.

Data compression is another useful method for data aggregation to reduce the energy consumption of reporting data. In [5], Marcelloni and Vecchio propose an efficient data compression algorithm for wireless sensor networks with very limited energy, memory, and computational re-

* Corresponding author. Tel.: +886 7 3814526x5610; fax: +886 7 3811182.

E-mail addresses: bhliu@kuas.edu.tw (B.-H. Liu), 1099305151@kuas.edu.tw (J.-Y. Jhang).

http://dx.doi.org/10.1016/j.comnet.2014.03.003
1389-1286/

sources. In [6], Chen et al. propose a nonthreshold-based node-level algorithm for data compression to decrease energy consumption and maximize the network lifetime. In [7], Kimura and Latifi survey and discuss a number of compression schemes for wireless sensor networks.

How to efficiently aggregate and report data has received much attention recently. Clustering is a well-known structure for data aggregation. Each cluster head is responsible for aggregating data from sensors in the corresponding cluster. The data aggregated in the cluster head is then sent to the sink. In [8], Gionis et al. propose an algorithm to optimize and solve the clustering aggregation problem in large data sets. In [9], Kamimura et al. propose a clustering algorithm for decreasing the loading of cluster heads to prolong the lifetime of the sensor networks. In [10], Pattem et al. propose a data compression measure with spatial correlations to quantify the compressed information and get near-optimal performance for static clustering.

A tree is another important structure for data aggregation. Every node is responsible for aggregating the generated or received data and forwarding to the parent node in the tree. In [11], Lee and Wong propose an overlay tree structure to facilitate data aggregation. In addition, the tree structure can be re-constructed when sensors are no longer functional or links are broken. In [12], an approximation algorithm of constructing a spanning tree is proposed to prolong the network lifetime when a single sink exists. In [13], Luo et al. study the problem of constructing a maximum lifetime routing tree that is subject to a shortest path tree, and propose a distributed algorithm accordingly.

Recently, Kuo and Tsai [14] study how to efficiently report data by constructing a routing tree when a fixed number of data are allowed to be aggregated into one packet [14,15], where every sensor except for the sink and relay nodes in the networks has to periodically generate data. Two problems, termed MINIMUM ENERGY-COST AGGREGATION TREE and MINIMUM ENERGY-COST AGGREGATION TREE WITH RELAY NODES, are proposed. Both problems are shown to be NP-complete. In addition, approximation algorithms of constructing routing trees are proposed to minimize the total energy cost of data transmission in wireless sensor networks.

For data aggregation in wireless sensor networks, some aggregation functions used to combine multiple data into one packet are unrepeatable [16–18]; that is, every data is aggregated only at most once. Take the cross-correlation of waveforms, for example. Cross-correlating waveforms generates a number that cannot be cross-correlated with other waveform [16]. This motivates us to research the problem of how to report data with minimum total energy cost when a fixed number of data are allowed to be aggregated into one packet and every data is aggregated at most once. In this paper, we propose novel data aggregation and routing structures for reporting generated data. With the structures, we study the problem of scheduling generated data to nodes in networks for data aggregation such that the energy cost of reporting data is minimized, termed MINIMUM ENERGY-COST DATA-AGGREGATION SCHEDULING. The remainder of this paper is organized as follows. The network model and problem formulation are introduced in Section 2. The hardness of MINIMUM ENERGY-COST DATA-AGGREGATION SCHEDULING is given in Section 3. In Section 4, a distributed data scheduling algorithm, termed the DDSA, is proposed for MINIMUM ENERGY-COST DATA-AGGREGATION SCHEDULING. We evaluate, with simulations, the performance of the DDSA in Section 5. Finally, we conclude this paper in Section 6.

2. Network model and problem formulation

Our focus is on applications for wireless sensor networks, where sensors periodically generate sensing data and report the data to a sink for further analysis. Example scenarios in this study include civil structure maintenance [14] and continuous environmental condition monitoring [18] such as sound, vibration, humidity, or temperature monitoring. In the following, the wireless sensor network that can be represented by a connected weighted graph is first introduced in Section 2.1. Subsequently, we describe the data aggregation and routing structure used for data aggregation in Section 2.2. Finally, our problem, termed MINIMUM ENERGY-COST DATA-AGGREGATION SCHEDULING, is introduced in Section 2.3.

2.1. Connected weighted graph

In this paper, the communication model in the wireless sensor network is assumed to be a unit disk graph model [19], in which a sensor u can receive messages sent from sensor v if u is within the transmission range of v. When all sensors have the same transmission range, the wireless sensor network can be represented as a connected weighted graph $G = (V_G, E_G, w_G)$, where $v \in V_G$ denotes a sensor in the wireless sensor network, edge $(u, v) \in E_G$ indicates that u and v are neighbors that can communicate with each other, and $w_G(v) \in \mathbb{Z}_0^+$, the weight of node $v \in V_G$, denotes the total number of units of raw data periodically generated by v. In the wireless sensor network, a special node $s \in V_G$, termed a sink, works as a gateway between the wireless sensor network and the outside network. Fig. 1(a) shows an example of the connected weighted graph representing a wireless sensor network, where node s is the sink. Note that nodes a, b, c, and d generate 2, 1, 5, and 1 units of raw data, respectively.

2.2. Data aggregation and routing structures

Every node $v \in V_G$ is responsible for periodically reporting its generated data to sink s if $w_G(v) \neq 0$. Let $\alpha \in \mathbb{Z}^+$ ($\alpha \geqslant 2$) be the aggregation ratio, denoting the maximum number of units of raw data allowed to be aggregated into one unit-size packet [14,15]. To efficiently report the generated data, the raw data generated at node u is allowed to be aggregated at another node v through the shortest path from u to v. Hereafter, node v is called the aggregated node of raw data r if r is aggregated at node v. Note that sink s can be considered a special aggregated node that is able to aggregate unlimited raw data. In the networks, the raw data can be forwarded by an aggregation structure. An aggregation structure constructed for the network $G = (V_G, E_G, w_G)$ is a directed weighted mesh