Journal of Network and Computer Applications 38 (2014) 150–162

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

An efficient algorithm for scheduling sensor data collection through multi-path routing structures ☆

Hai Van Luu *, Xueyan Tang

School of Computer Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 17 September 2012
Received in revised form
16 February 2013
Accepted 22 March 2013
Available online 31 March 2013

Keywords:
Sensor network
Scheduling
Data collection
Multi-path routing

ABSTRACT

Multi-path routing is essential to improve the robustness of sensor data collection in error-prone wireless communication environments. Besides the robustness against communication failures, both the energy efficiency and time efficiency are also of primary importance in sensor data collection due to the limited energy supplies of sensor nodes and the real-time nature of sensor network applications. Contention-free time division multiple access (TDMA) protocols have the potential to reduce the energy consumption and the latency of sensor data collection. To collect sensor data using a TDMA protocol, sensor nodes need to be assigned appropriate time slots for transmitting and receiving data prior to the data collection process. We note that the distributed TDMA scheduling process for sensor data collection incurs overhead costs of energy consumption and time latency. However, these overhead costs are usually overlooked, especially when multi-path routing is used to collect sensor data. In this paper, we propose an efficient scheduling algorithm for data collection through multi-path routing structures in wireless sensor networks. The objective of our scheduling algorithm is to reduce both the message complexity and running time of the scheduling process as much as possible. In addition, we also develop a method for deriving a lower bound on the shortest possible length of the data collection schedule that can be generated by any algorithm. The lower bound latency estimation offers a practical method to evaluate the efficiency of data collection schedules produced by scheduling algorithms. Extensive experimental results show that the proposed scheduling algorithm significantly reduces the number of messages transmitted during the scheduling process and the running time compared to an existing scheduling algorithm. The length of the data collection schedule produced by our algorithm is normally within 1.9 times of the lower bound estimate across a wide range of network settings.

1. Introduction

Wireless sensor networks have been employed to collect data related to physical phenomena in various applications such as habitat monitoring, ocean monitoring and surveillance (Mainwaring et al., 2002; Huddlestone-Holmes et al., 2007; Chen et al., 2006). Sensor data are usually transported through multi-path routing structures toward the base station (Ganesan et al., 2001; Considine et al., 2004; Nath et al., 2004; Ye et al., 2005) to overcome the high loss rates of wireless communication links such that the base station can maintain accurate results of data collection. In addition, in-network data aggregation techniques (Nath et al., 2004) are often employed during the data collection process to conserve the limited energy supplies of sensor nodes. To facilitate communication among sensor nodes, their transmissions are usually coordinated by a Medium Access Control (MAC) protocol. Contention-based MAC protocols for coordinating communication among sensor nodes (Hill et al., 2000) inherently cost a large amount of energy consumption and time latency due to control messages, idle listening and back off time (Demirkol et al., 2006). In contrast, contention-free MAC protocols such as time division multiple access (TDMA) protocols have the potential to improve the energy efficiency and reduce the latency of sensor data collection since communication among sensor nodes can proceed in synchronous time slots without contention time, collisions or idle listening.

To collect sensor data using a TDMA protocol, sensor nodes need to be assigned appropriate time slots for transmitting and receiving data prior to the data collection process. Recently, there have been several studies on TDMA scheduling for sensor data collection through single-path routing structures (Yu et al., 2009; Li et al., 2010). We note that distributed TDMA scheduling algorithms for sensor data collection incur overhead costs of energy consumption and time latency. However, these overhead

* Corresponding author. Tel.: +65 8224 9482; fax: +65 6792 6559.
E-mail addresses: lu0002ai@e.ntu.edu.sg (H. Van Luu). asxytang@ntu.edu.sg (X. Tang).

☆ An earlier version of this paper was presented at the IEEE MSN'2010 conference (Luu and Tang, 2010a).

1084-8045/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jnca.2013.03.013

costs are usually overlooked. Moreover, single-path routing structures are highly susceptible to communication failures since a single failure would result in the loss of data acquired by sensor nodes in an entire subtree. In contrast, multi-path routing structures have the potential to significantly improve the robustness of data collection against communication failures (Nath et al., 2004; Luu and Tang, 2010b). Nevertheless, little work has been studied on TDMA scheduling for sensor data collection through multi-path routing structures. In this paper, we design an efficient distributed scheduling algorithm for constructing TDMA schedules for sensor data collection through multi-path routing structures.

In Wu et al. (2006), Gandham et al. (2005), Gandham et al. (2006), Yu et al. (2009) and Li et al. (2010), it is assumed that a sensor node can only successfully receive a message without collisions if there is exactly one node within its communication range performing data transmission at a time. Thus, a scheduled transmission of a node in data collection must not interfere with the reception of other nodes within its communication range. To allow each node to determine its transmission schedule locally in a distributed manner, the node would normally need to know the transmission schedules of other nodes within its two-hop neighborhood (Yu et al., 2009; Li et al., 2010). The number of nodes within the two-hop neighborhood of a node is usually on the order of $O(\Delta^2)$, where Δ is the number of nodes within the node's communication range. As a result, the number of messages transmitted and received by the sensor nodes in a distributed scheduling process would increase rapidly when the node density increases. We define the number of messages sent by all nodes in a scheduling algorithm as the *message complexity* of that algorithm. Due to the limited communication bandwidth of sensor nodes, higher message complexity would also result in longer running time of the scheduling process. If the time and the traffic costs of the scheduling process become significant, the benefit of TDMA data collection would be compromised. In this paper, we aim to design a distributed scheduling algorithm for constructing TDMA schedules for wireless sensor data collection through multi-path routing structures such that both the message complexity and running time of the scheduling process are reduced as much as possible.

The remainder of this paper is organized as follows. Section 2 summarizes the related work. Section 3 outlines the major contributions of our paper. Section 4 introduces some preliminaries on collecting sensor data through multi-path routing structures and the requirements of a valid schedule for data collection. Our distributed scheduling algorithm for data collection is described in Section 5 and the lower bound latency analysis is presented in Section 6. Experimental evaluation is described in Section 7. Finally, Section 8 concludes the paper.

2. Related work

Early work of link scheduling in wireless sensor networks has focused on building schedules for every pair of neighboring nodes to communicate once. Gandham et al. (2005) introduced a distributed link scheduling algorithm for undirected communication graphs that consists of two phases. The first phase is to find a valid edge coloring of all links using $(\Delta+1)$ colors, where Δ is the maximum degree of sensor nodes in the network. Then, the transmission direction of each link is determined in the second phase to construct a collision-free schedule. It was proved that reversing the transmission directions of all links results in another collision-free schedule. Thus, two-way communication among nodes is supported with at most $2(\Delta+1)$ time slots. Cheng and Yin (2007) considered directed communication graphs and proposed a scheduling algorithm that directly assigns time slots to directed links. Ma et al. (2009) introduced link scheduling algorithms that assign consecutive time slots to links destined for the same node. The resultant link schedule guarantees that every node switches to the listening mode only once for receiving data from all its neighbors. However, these studies did not consider in-network data aggregation in scheduling.

To facilitate in-network data aggregation, the child nodes have to be scheduled for transmission before their parent nodes in the routing structure. In this way, sensor nodes can aggregate data received from their children before transmitting the data to their parents. Wu et al. (2006) proposed a distributed scheduling algorithm to assign time slots to the links of a tree-based routing structure rooted at the base station. This algorithm requires each node to have a transmission slot different from all the nodes within its communication range. As a result, this algorithm may suffer from the exposed station problem because two neighbor nodes may be able to transmit data in the same time slot without collisions at their parents if they have different parents in the tree-based routing structure. Paradis and Han (2008) also considered the link scheduling problem for collecting data from sensor nodes to a base station through a tree-based routing structure. This study exploited in-network processing for collecting raw sensor data readings. It was assumed that each packet transmitted can accommodate a certain number of sensor readings. A two-phase algorithm called TIGRA was proposed to schedule links for data collection. TIGRA provides a collision-free schedule for sensor data collection and energy consumption is optimized by grouping raw data readings together in the packets transmitted. However, since no data aggregation is exploited, the energy consumption of the nodes closer to the base station would normally be much higher than that of the nodes further away from the base station because the nodes closer to the base station would have to forward many data packets from their descendants to the base station. As a result, the sensor nodes closer to the base station would deplete their energy sources more rapidly than other nodes.

Many aggregation scheduling algorithms employ a simple greedy strategy to select the transmission slot of a sensor node as the earliest collision-free time slot after the transmission slots of all the children of the node (Chen et al., 2005; Huang et al., 2007; Yu et al., 2009; Li et al., 2010). Chen et al. (2005) designed an approximate algorithm for scheduling data collection and aggregation through tree-based routing structures. It was also shown that the scheduling problem for data aggregation with minimum latency through tree-based routing structures is NP-hard. Huang et al. (2007) proposed a scheduling algorithm for data collection through an improved tree structure with a latency bound of $23R+\Delta-18$, where R is the network radius and Δ is the maximum degree of sensor nodes in the network. Wan et al. (2009) further improved the tree structure for data collection and designed scheduling algorithms with lower latency bounds. All the above scheduling algorithms are centralized in nature. Yu et al. (2009) developed a distributed scheduling algorithm for data collection and aggregation through tree-based routing structures and proposed a new tree structure to further improve the latency bound. The same scheduling algorithm was also used by Li et al. (2010). Xu et al. (2009) designed a scheduling algorithm for tree-based routing structures in which the communication range and the interference range of sensor nodes are assumed to be different. None of the above work has considered scheduling data collection through multi-path routing structures.

3. Contributions

In this paper, we propose a novel distributed scheduling algorithm for TDMA-based sensor data collection using a predetermined scheduling order of sensor nodes. Our proposed scheduling algorithm is generic and is applicable to any single-path