Ad Hoc Networks 12 (2014) 52–68

Contents lists available at SciVerse ScienceDirect

Ad Hoc Networks

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

Latency-minimizing data aggregation in wireless sensor networks under physical interference model

Hongxing Li [a,*], Chuan Wu [a], Qiang-Sheng Hua [b], Francis C.M. Lau [a]

[a] *Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong*
[b] *Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China*

ARTICLE INFO

Article history:
Received 20 May 2011
Received in revised form 20 September 2011
Accepted 6 December 2011
Available online 20 December 2011

Keywords:
Data aggregation
Wireless sensor networks
Physical interference model
Minimum-latency

ABSTRACT

Minimizing latency is of primary importance for data aggregation which is an essential application in wireless sensor networks. Many fast data aggregation algorithms under the protocol interference model have been proposed, but the model falls short of being an accurate abstraction of wireless interferences in reality. In contrast, the physical interference model has been shown to be more realistic and has the potential to increase the network capacity when adopted in a design. It is a challenge to derive a distributed solution to latency-minimizing data aggregation under the physical interference model because of the simple fact that global-scale information to compute the cumulative interference is needed at any node. In this paper, we propose a distributed algorithm that aims to minimize aggregation latency under the physical interference model in wireless sensor networks of arbitrary topologies. The algorithm uses $O(K)$ time slots to complete the aggregation task, where K is the logarithm of the ratio between the lengths of the longest and shortest links in the network. The key idea of our distributed algorithm is to partition the network into cells according to the value K, thus obviating the need for global information. We also give a centralized algorithm which can serve as a benchmark for comparison purposes. It constructs the aggregation tree following the nearest-neighbor criterion. The centralized algorithm takes $O(\log n)$ and $O(\log^3 n)$ time slots when coupled with two existing link scheduling strategies, respectively (where n is the total number of nodes), which represents the current best algorithm for the problem in the literature. We prove the correctness and efficiency of our algorithms, and conduct empirical studies under realistic settings to validate our analytical results.

1. Introduction

Data aggregation is a habitual operation of many wireless sensor networks, which transfers data (*e.g.*, temperature) collected by individual sensor nodes to a sink node. The aggregation typically follows a tree topology rooted at the sink. Each leaf node would deliver its collected data to its parent node. Intermediate sensor nodes of the tree may optionally perform certain operations (*e.g.*, sum, maximum, minimum, mean, etc.) on the received data and forward the result. Because the wireless medium is shared, transmissions to forward the data need to be coordinated in order to reduce interference and avoid collision. The fundamental challenge can be stated as: How can the aggregation transmissions be scheduled in a wireless sensor network such that no collision may occur and the total number of time slots used (referred to as *aggregation latency*) is minimized? This is known as the *Minimum-Latency Aggregation Scheduling* (*MLAS*) problem in the literature [1–5].

The *MLAS* problem is typically approached in two steps: (i) data aggregation tree construction and (ii) link transmission scheduling. For (ii), we assume the simplest mode

* Corresponding author. Tel.: +852 96594974; fax: +852 25598447.
E-mail addresses: hxli@cs.hku.hk (H. Li), cwu@cs.hku.hk (C. Wu), qshua@mail.tsinghua.edu.cn (Q.-S. Hua), fcmlau@cs.hku.hk (F.C.M. Lau).

1570-8705/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.adhoc.2011.12.004

in which every non-leaf node in the tree will make only one transmission, after all the data from its child nodes have been received. A correct solution to the *MLAS* problem requires that no concurrent transmissions interfering with each other should take place. If steps (i) and (ii) are carried out simultaneously in a solution, we have a "joint" design.

To model wireless interference, existing literature mostly assume the *protocol interference model*, in which a transmission is successful if and only if its receiver is within the transmission range of its transmitter *and* outside the interference range of any other concurrent transmitters. The best results known for the *MLAS* problem or similar problems ([2–5]) under the protocol interference model bound the aggregation latency in $O(\Delta + R)$ time slots, where R is the radius of the sensor network in hops and Δ is the maximal node degree (*i.e.*, the maximum number of nodes in any node's transmission range). The protocol interference model however has been found to be too simplistic and cannot serve as an accurate abstraction of wireless interferences. Instead, the *physical interference model* [6], which captures the reality more accurately, is becoming more popular. Little research however has so far been done to address the *MLAS* problem under the physical interference model.

The protocol interference model considers only interferences within a limited region, whereas the physical interference model tries to capture the cumulative interference due to all other concurrently transmitting nodes in the entire network. More precisely, in the physical interference model, the transmission of link e_{ij} can be successful if the following condition regarding the Signal-to-Interference-Noise-Ratio (*SINR*) is satisfied:

$$\frac{P_{ij}/d_{ij}^{\alpha}}{N_0 + \sum_{e_{gh} \in \Lambda_{ij} - \{e_{ij}\}} P_{gh}/d_{gj}^{\alpha}} \geqslant \beta. \tag{1}$$

Here Λ_{ij} denotes the set of links that transmit simultaneously with e_{ij}. P_{ij} and P_{gh} denote the transmission power at the transmitter of link e_{ij} and that of link e_{gh}, respectively. d_{ij} (d_{gj}) is the distance between the transmitter of link e_{ij} (e_{gh}) and the receiver of link e_{ij}. α is the path loss ratio, whose value is normally between 2 and 6. N_0 is the ambient noise. β is the *SINR* threshold for a successful transmission, which is at least 1.

We give an example, in Fig. 1, to demonstrate the advantage of the physical interference model over the traditional protocol interference model, with which the network capacity is underestimated (data aggregation time is longer). In the figure, six nodes are located on a line, where sink a aggregates data from the other five nodes, *b–f*. The number on a link is the distance between the two nodes joined by the link. Under the protocol interference model, any two concurrent transmissions conflict with each other, and therefore five time slots are needed to aggregate all the data to the sink a, such as by the sequence $f \to e \to d \to c \to b \to a$. On the other hand, with

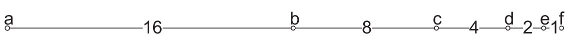

Fig. 1. A data aggregation example.

the physical interference model, three time slots are enough: at time slot 1, the transmissions $b \to a$, $d \to c$, and $f \to e$ can be scheduled concurrently, using transmission power $2N_0\beta\ 16^{\alpha}$. At time slots 2 and 3, $e \to c$ and $c \to a$ can be scheduled consecutively with transmission power $N_0\beta 6^{\alpha}$ and $N_0\beta 24^{\alpha}$, respectively. It can be easily verified that the above link scheduling and power assignment satisfy the *SINR* condition (1) at each receiver under typical network settings, *e.g.*, $\alpha = 4$ and $\beta = 1$. In this paper, we investigate the *MLAS* problem under the physical interference model.

A solution to the *MLAS* problem can be a centralized one, a distributed one, or mixed. For a large sensor network, a distributed solution is certainly the desired choice. Distributed scheduling algorithm design is significantly more challenging with the physical interference model, as "global" information in principle is needed by each node to compute the cumulative interference at the node. We are only aware of one study [7] which presents a distributed solution to the *MLAS* problem under the physical interference model; they derived a latency bound of $O(\Delta + R)$ in a network, where sensors are uniformly randomly deployed. One of the drawbacks of this work is that the efficiency guarantee is not provided for arbitrary topologies.

In this paper, we tackle the minimum-latency aggregation scheduling problem under the physical interference model by designing both a centralized and a distributed scheduling algorithm. Our algorithms are applicable to arbitrary topologies. The distributed algorithm we propose, *Cell-AS*, circumvents the need to collect global interference information by partitioning the network into cells according to a parameter called the link length diversity (K), which is the logarithm of the ratio between the lengths of the longest and the shortest links. Our centralized algorithm, *NN-AS*, combines our aggregation tree construction algorithm with either one of the link scheduling strategies proposed in [8,9] to achieve the best aggregation performance in the current literature. Our main focus in this paper is on the distributed algorithm; the centralized algorithm is included for completeness and to serve as a benchmark in the performance comparison. For situations in practice, where centralization is not a problem, the centralized algorithm may be a useful choice.

We conduct theoretical analysis to prove the correctness and efficiency of our algorithms. We show that the distributed algorithm *Cell-AS* achieves a worst-case aggregation latency bound of $O(K)$ (where K is the link length diversity), and the centralized algorithm *NN-AS* achieves worst-case bounds of $O(\log n)$ and $O(\log^3 n)$ when coupled with the link scheduling strategies in [8,9], respectively (where n is the total number of sensor nodes). In addition, we derive a theoretically optimal lower bound for the *MLAS* problem under any interference model—log (n). Given this optimal bound, the approximation ratios are $O(K/\log n)$ with *Cell-AS*, $O(1)$ with *NN-AS* and the link scheduling in [8], and $O(\log^2 n)$ with *NN-AS* and the link scheduling in [9]. We also compare our distributed algorithm with Li et al.'s algorithm in [7] both analytically and experimentally. We show that both algorithms have an $O(n)$ latency upper bound in their respective worst cases, while *Cell-AS*