Computer Networks 123 (2017) 184–199

Contents lists available at ScienceDirect

Computer Networks

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

Many-to-many data aggregation scheduling in wireless sensor networks with two sinks☆

Sain Saginbekov[a,*], Arshad Jhumka[b]

[a] Department of Computer Science, Nazarbayev University, Astana, Kazakhstan
[b] Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK

ARTICLE INFO

Article history:
Received 8 August 2016
Revised 21 May 2017
Accepted 23 May 2017
Available online 24 May 2017

Keywords:
Wireless sensor networks
Data aggregation scheduling
Two sinks
Many-to-many communication
Medium access control

ABSTRACT

Traditionally, wireless sensor networks (WSNs) have been deployed with a single sink. Due to the emergence of sophisticated applications, WSNs may require more than one sink. Moreover, deploying more than one sink may prolong the network lifetime and address fault tolerance issues. Several protocols have been proposed for WSNs with multiple sinks. However, most of them are routing protocols. Differently, our main contribution, in this paper, is the development of a distributed data aggregation scheduling (DAS) algorithm for WSNs with two sinks. We also propose a distributed energy-balancing algorithm to balance the energy consumption for the aggregators. The energy-balancing algorithm first forms trees rooted at nodes which are termed *virtual sinks* and then balances the number of children at a given level to level the energy consumption. Subsequently, the DAS algorithm takes the resulting balanced tree and assigns contiguous slots to sibling nodes, to avoid unnecessary energy waste due to frequent active-sleep transitions. We prove a number of theoretical results and the correctness of the algorithms. Through simulation and testbed experiments, we show the correctness and performance of our algorithms.

1. Introduction

A wireless sensor network (WSN) consists of a set of resource-constrained nodes, that communicate wirelessly. These nodes sense the environment for events of interest and subsequently relay the information to a dedicated device called *sink*, with data from several nodes aggregated along the way for energy efficiency reasons. The data can then later be analysed offline.

Traditionally, WSNs have been deployed with a single sink [1]. However, there are several reasons that limit the usefulness of a single sink. The emergence of more sophisticated applications, such as Heating, Ventilation, and Air Conditioning (HVAC) systems [28], requires WSNs with more than one sink. Moreover, the deployment of more than one sink may improve the network throughput and prolong network lifetime by balancing energy consumption, and may address fault tolerance issues [25,36,42].

WSNs are typically resource-constrained networks, with nodes having limited computational and energy resources. To reduce energy consumption, various approaches exist, such as duty-cycling and the use of appropriate medium access control (MAC) protocols. Major sources of energy waste at the MAC layer are:

- Message collisions, which require the retransmission of the collided packets,
- Message overhearing, where a node receives a message meant for another node, and
- Idle listening, which means that a node keeps on listening for messages on an otherwise idle channel [44].

To address the above problems, a typical solution is to use a Time Division Multiple Access (TDMA) based MAC protocol. TDMA MAC protocols work by dividing time into slots and assigning those slots to nodes. Each node can then only transmit in a slot to which it has been assigned. Several TDMA-based MAC protocols have been proposed for WSNs, e.g., [31,32,37]. However, most of them have been developed for a WSN with a single sink. Thus, there is a need for TDMA-based MAC protocols specifically designed for WSNs with multiple sinks. However, there is a dearth of work in this area. Data aggregation scheduling (DAS) algorithms in a WSN with *multiple sinks* have been presented in [3,21]. However, in these works, the data aggregation scheduling is done from

☆ This is an extended version of the paper [33] that was published in the Proceedings of the 5th International Conference on Sensor Networks (SENSORNETS 2016), Rome, Italy, February 2016.

* Corresponding author.
E-mail addresses: sain.saginbekov@nu.edu.kz (S. Saginbekov), H.A.Jhumka@warwick.com (A. Jhumka).

http://dx.doi.org/10.1016/j.comnet.2017.05.022
1389-1286/

many nodes to *one* sink which collects all the messages, whereas this work considers data aggregation scheduling from *many* nodes to *many* sinks, where every sink has to collect all messages.

The way we propose to solve the data aggregation scheduling problem in WSNs with two sinks is to first develop a backbone that connects the two sinks and then allocate slots to nodes that connect to the backbone. The problem of developing the backbone, i.e., a path connecting the sinks, is directly related to the problem of developing a (minimum) Steiner tree [11], which is a well-studied combinatorial optimization problem. Given a graph $G=(V,E)$ with weighted edges and a set of nodes $S\subseteq V$, a Steiner tree T interconnects all elements in S such that the sum of the edge weights in T is minimal. In the general case, this is an NP-complete problem [20]. The minimum Steiner tree problem can be seen as a combination of computing shortest paths and of computing spanning trees. If the number of vertices to be joined is two, then we only need to compute the shortest path. On the other hand, if all the vertices are involved, then a spanning tree is computed. In this paper, as we focus on two sinks, we will construct a shortest path, which can be achieved in polynomial-time.

In this context, we make the following novel contributions:

- In Section 4, we formalise the problem of DAS scheduling in a WSN with two sinks, where data needs to reach both sinks.
- In Section 5, we prove an impossibility[1] result, as well as derive a lower bound for solving weak DAS, a special case of DAS.
- In Section 6, we propose two algorithms which, taken together, solve weak DAS for two sinks. The output of the combined algorithms is a schedule that matches the predicted lower bound.
- Through both simulation and testbed experiments, we show the performance and correctness of the algorithms in Section 8.

The other parts are as follows: In Section 2, we present an overview of related work. We present the formal basis of our work in Section 3. In Section 7, we present the experimental setup. We conclude the paper in Section 9.

2. Related work

A data aggregation technique is used in data gathering in WSNs to reduce energy consumption, as it reduces the number of transmissions [22]. A number of data aggregation protocols have been proposed in the literature [12,14,23,27,34,38,45,47]. The routing structure of these protocols could be tree-based [12,23,34,38], cluster-based [14,45], chain-based [27], and grid-based [19,30]. However, these protocols have been proposed for WSNs with a single sink. Although our proposed protocol is developed for WSNs with two sinks, the routing structure of our proposed protocol is based on a tree structure.

Protocols that have been developed for communication in WSNs with multiple sinks can be found in [3,5,6,13,21,24,28,39,41,48]. The authors of [5] developed an algorithm where a node chooses a sink in a multi-sink WSN to send its data in such a way that it minimizes energy consumption. A scheme proposed in [28] performs data collection from many nodes to many sinks, i.e., many-to-many communication. The main idea of the protocol is to reduce the number of redundant transmissions by leveraging neighbourhood information. An algorithm that builds two node-disjoint paths from every node to two different sinks was proposed in [39]. If one of the two paths fails, the other path is used to route the data. In [41], the authors propose a routing protocol that is using a hexagon-based architecture. The nodes in the network are grouped into hexagons, based on their locations. The routing protocol proposed in [48], is based on trees. In the protocol, different trees rooted at different sinks are used to forward data. The authors of [6] have proposed an online algorithm for data collection in WSNs with multiple sinks, where sinks are deployed in a stepwise fashion during network operation. In [29], the authors present different routing schemes that are based on a logical tree structure, which is built based on the residual energy of each node. One of these schemes uses a secondary sink to maximize the network lifetime. A data reporting algorithm used for object tracking in multi-sinks WSNs was presented in [10]. The algorithm attempted to reduce energy consumption and balance the load among sinks and nodes.

The works presented in [3,13,21,24] are more closely related to the work presented in this paper. In [21], in addition to an algorithm that computes the shortest-path trees rooted at each sink in a multi-sinks WSN, the authors proposed a scheduling algorithm that uses a graph colouring technique. Then, a node will send messages to its nearest sink in a collision-free fashion. The authors of [3] proposed two algorithms for scheduling data aggregation in multi-sinks WSNs. One of the algorithms is a Voronoi-based scheduling algorithm, where the sensing area is divided into regions to form k forests, one forest for each sink. Subsequently, the algorithm computes the schedules for the nodes. In [13,24], the authors proposed cross-layer schemes that consider the routing as well as the MAC layer to maximize the network lifetime and reduce the latency respectively. Most of the above protocols either developed appropriate routing protocols for multi-sinks WSNs or the techniques involved a subset of the nodes forwarding messages to a single sink only. In contrast, we consider the case where *many* nodes send their data to *two* sinks, which is an instance of many-to-many communication.

3. System model

We provide the necessary formal background, including the models we use, the syntax we use to write the algorithms and the associated semantics, the computation and communication models.

3.1. Topology and processes

Communication in WSNs is typically modeled a circular communication range centered on a node, and it is typically assumed that all nodes have the same communication range. With this model, a node is thought to be able to directly exchange data with all devices within its communication range. In graph-theoretic terms, we represent a WSN as a *undirected* graph $G=(V,E)$ with a set V of vertices representing the nodes, and a set E of edges (or links) representing the communication links between pairs of nodes. A path between two nodes n_1 and n_i is a sequence of nodes $\gamma=n_1\cdot n_2\ldots n_i$ such that $\forall j, 1\leq j<i, (n_j,n_{j+1})\in E$. We assume a network where no node has both sinks as neighbours. We also assume that the network topology remains constant, i.e., there is no node crash and no link failure.

A program consists of a finite set of processes. Each process contains a finite set of variables, each taking values from a given finite domain, and a finite set of actions. An assignment of values to variables is called a state and the set of value assignments denote the state space of the program. A predicate defines a set of states, such that the predicate evaluates to true in these states.

3.2. Program syntax and semantics

We write programs in a style similar to the guarded command notation [7]. An action has the form

[1] This means that a given problem cannot be solved deterministically.