Computer Networks 91 (2015) 425–437

Contents lists available at ScienceDirect

Computer Networks

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

Using central nodes for efficient data collection in wireless sensor networks☆

Vitaly Milyeykovski, Michael Segal*, Vladimir Katz

Communication Systems Engineering Department, Ben-Gurion University of the Negev, Israel

ARTICLE INFO

Article history:
Received 20 February 2015
Revised 3 August 2015
Accepted 21 August 2015
Available online 10 September 2015

Keywords:
Data gathering
Optimization
Sensor networks

ABSTRACT

We study the problem of data collection in Wireless Sensor Networks (WSN). A typical WSN is composed of wireless sensor nodes that periodically sense data and forward it to the base station in a multi-hop fashion. We are interested in designing an efficient data collection tree routing, focusing on three optimization objectives: energy efficiency, transport capacity, and hop-diameter (delay).

In this paper we develop single- and multi-hop data collection, which are based on the definition of node centrality: centroid nodes. We provide theoretical performance analysis for our approach, present its distributed implementation and discuss the different aspects of using it. Most of our results are for two-dimensional WSNs, however we also show that the centroid-based approach is asymptotically optimal in three-dimensional random node deployments. In addition, we present new construction for arbitrary network deployment based on central nodes selection. We also present extensive simulation results that support our theoretical findings.

1. Introduction

A wireless sensor network (WSN) consists of small autonomous low-cost low-power devices that carry out monitoring tasks. Initially developed for military use, WSNs can nowadays be found in many civil applications, such as environmental monitoring, biomedical research, seismic monitoring, and precision agriculture [1]. The devices are called *sensor nodes* and the monitored data is typically collected at a *base station*, following a specific collection pattern of activated wireless links.

As these networks have no hard-wired underlying topology, one of the most fundamental issues when a WSN is deployed is the formation of an efficient communication backbone, or in other words, answering the question *which links to use in order to collect the data from the sensor nodes?*

Efficiency can be defined in many ways, for example it can be maximizing the rate at which data is collected [23,45,47] from the sensor nodes, prolonging the network lifetime by reducing the energy consumption [6,9,34,36,40], minimizing the number of hops from the sensor nodes to the collecting base station [16,22], and other optimization objectives. It is apparent that the *topological structure of the communication backbone* plays a vital role in its efficiency. However, it is also important to note that a communication backbone which has good performance in some of the criteria can have a bad one in others. For example, using the minimum spanning tree (MST) as the backbone provides an optimal network lifetime performance for same initial battery charges [5], however it

☆ A preliminary version of this paper has appeared in [31].

* Corresponding author. Tel.: +972-8-6477234; fax: +972-8-6472883.
E-mail addresses: milyeyko@bgu.ac.il (V. Milyeykovski), segal@cse.bgu.ac.il, segal@bgu.ac.il, michael99segal@gmail.com (M. Segal).

http://dx.doi.org/10.1016/j.comnet.2015.08.041
1389-1286/

can have a very poor hop-diameter, which is critical for delay minimization.[1] Thus, the network designer has to take special care when deciding which links to activate for the purpose of data collection, as different optimization objectives may be have a negative effect on each other.

The problem of data collection can be divided into two major paradigms. Data collection *with aggregation* [25,43] allows each sensor node to accumulate the messages of its descendants and then pass only one fixed-size message towards the base station. The second paradigm, is data gathering *without aggregation* [29,30] which requires that *all* messages initiated by the sensors will eventually reach the base station.

Our main objective in this paper is to construct efficient communication backbones for single- and multi-hop data collection with aggregation in WSNs for random sensor node deployments, while measuring the efficiency based on the next three metrics.

- The *transport capacity* metric represents the sum of rate-distance products over all the active links. It is measured in *bit-meters* and was first introduced by [20]. The idea behind this measure is to capture both the notion of the overall rate and distance that the information travels in a network.
- *Hop-diameter* is another important metric which reflects the depth of the data gathering tree, i.e., the maximum number of hops from any of the sensor nodes to the base station.
- *Total energy consumption* is probably one of the most important parameters of a WSN as the sensor nodes are often deployed in areas where battery replacement is infeasible [8]. Wireless communication is a major contributor to the energy budget of a node. In this paper we focus on minimize the total energy consumed by all nodes for communication purposes.

We propose a novel approach for the construction of communication backbones by identifying *central locations* in the deployment area and routing all data through these regions. The general idea is that these locations would serve as aggregation points both on a local and global scale. In particular, we use an interesting geometrical notion of *centroids*, which is defined as the central geometrical position of a collection of nodes, which are used as a guide for the construction of hierarchical aggregation trees.

The rest of this paper is organized as follows. In Section 2 we present our system settings and state our objective. Related works are surveyed in Section 3. Sections 4 and 5 are the technical sections of the paper and show the construction of data collection communication backbones for three scenarios, single-hop general network and multiple-hop random network. We present additional construction for arbitrary network deployment based on central nodes selection in Section 6. Simulation results for various types of networks are presented in Section 7 and compared to the results of similar spirit obtained in [14]. Finally, we conclude and discuss future work in Section 8.

[1] Imagine n sensor nodes located on a straight horizontal segment, with the base station being to the right of the right-most sensor. It is easy to show that the hop-diameter of MST in this case is n.

2. System settings

A WSN consists of n wireless sensor nodes, $S = \{s_1, \ldots, s_n\}$, distributed in some area A. These nodes perform monitoring tasks and periodically report to a base station r which is located somewhere within the area A (we consider different locations throughout the paper). During the report phase, the sensor nodes propagate a message to the base station through a *data collection tree*, $T_S = (S \cup \{r\}, E_S)$, rooted at r. We consider *data collection with aggregation*, where every node $s \in S$ forwards a single unit size *report message* to its parent. The message holds an accumulated information collected from a subtree of T_S rooted at s. An example of this scenario can be found in temperature monitoring systems for fire prevention, intrusion detection, seismic readings, etc.

We assume the use of *frame-based* MAC protocols (see [10]) which divide the time into frames, containing a fixed number of slots. The main difference from the classic TDMA is that instead of having one access point which controls transmission slot assignments, there is a localized distributed protocol mimicking the behavior of TDMA. The advantage of a frame-based (TDMA-like) approach compared to the traditional IEEE 802.11 (CSMA/CA) protocol for a Wireless LAN is that collisions do not occur, and that idle listening and overhearing can be drastically reduced. When scheduling communication links, that is, specifying the sender–receiver pair per slot, nodes only need to listen to those slots in which they are the intended receiver – eliminating all overhearing. When scheduling senders only, nodes must listen in to all occupied slots, but can still avoid most overhearing by shutting down the radio after the MAC (slot) header has been received. In both variants (link and sender-based scheduling) idle listening can be reduced to a simple check if the slot is used or not. Several MAC protocols were developed to adapt classical TDMA solutions which use access points to ad-hoc settings that have no infrastructure; these protocols employ a distributed slot-selection mechanism that self-organizes a multi-hop network into a conflict-free schedule (see [35,46]).

Let $d(u, v)$ be the Euclidean distance between two sensor nodes $u, v \in S$. It is customary to estimate that the energy required to transmit from u to v is proportional to $d(u, v)^\alpha$, where α is the *path-loss coefficient*. In perfect conditions $\alpha = 2$, however in more realistic settings (in presence of obstructions or noisy environment) it can have a value between 2 and 4 (see [33]). In this paper we assume $\alpha = 2$ for simplicity. However, it is possible to extend our results for other values of α which are greater than 2.

Let $E(T_S)$ be the **energy requirement** to execute a single report phase. Note that every sensor performs a single transmission, during which it sends a single message to its parent in T_S. Thus, the energy requirement is proportional to the sum of squares of lengths of the edges $E(T_S)$. The focus of this paper is to study the asymptotic performance of data collection trees, thus we can express $E(T_S)$ as follows, $E(T_S) = \sum_{(u,v) \in E_S} d(u, v)^2$.

Minimizing the energy requirement is one of the primary optimization objectives when deploying a WSN due to the very low battery reserves at the sensor nodes and the high costs that are associated with replacing these batteries (if at all possible).