Ad Hoc Networks 12 (2014) 69–85

Contents lists available at SciVerse ScienceDirect

Ad Hoc Networks

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

A spatial correlation aware algorithm to perform efficient data collection in wireless sensor networks

Leandro A. Villas [a,b,*], Azzedine Boukerche [a], Horacio A.B.F. de Oliveira [c], Regina B. de Araujo [d], Antonio A.F. Loureiro [b]

[a] Paradise Research Laboratory – SITE, University of Ottawa, Ottawa, Canada
[b] UWL – Research Laboratory, Federal University of Minas Gerais, Belo Horizonte, Brazil
[c] Dept. of Computer Science, Federal University of Amazonas, Manaus, Brazil
[d] WINDIS – Research Laboratory, Federal University of São Carlos, São Carlos, Brazil

ARTICLE INFO

Article history:
Received 16 May 2011
Accepted 24 August 2011
Available online 31 August 2011

Keywords:
Spatial correlation
Data aggregation
Wireless sensor networks
Routing protocol

ABSTRACT

Large scale dense wireless sensor networks (WSNs) will be increasingly deployed in different classes of applications for accurate monitoring. Due to this high density of nodes, it is very likely that both spatially correlated information and redundant data can be detected by several nearby nodes, which can be exploited to save energy. In this work we consider the problem of constructing a spatial correlation aware dynamic and scalable routing structure for data collection and aggregation in WSNs. Although there are some solutions for data aggregation in WSNs, most of them build their structures based on the order of event occurrence. This can lead to both low quality routing trees and a lack of load balancing support, since the same tree is used throughout the network lifetime. To tackle these challenges we propose a novel algorithm called dYnamic and scalablE tree Aware of Spatial correlaTion (YEAST). Results show that the routing tree built by YEAST provides the best aggregation quality compared with other evaluated algorithms. With YEAST an event can be sensed with 97% accuracy, and 75% of the nodes' residual energy can be saved within the phenomena area when compared with the classical approach for data collection.

1. Introduction

Wireless sensor networks (WSNs) [1,2] are composed of a large number of sensor nodes that cooperatively monitor physical or environmental conditions and transmit collected data to the sink node. This type of network has become popular due to its applicability that includes several areas such as environment, homeland security, industry, domestics, agriculture, meteorology, health, space, military and many other applications that can be critical to save lives and assets [3–5]. Several physical properties can be monitored, including temperature, humidity, pressure, ambient light, sound, vibration, and motion.

Sensor nodes are energy-constrained devices. Energy consumption is generally associated with communication, which is often the most expensive activity in terms of energy. A simple solution to this problem could be the periodic replacement of the node battery. However, this solution is not feasible due to the large number of nodes in the network or because the sensor nodes may be inaccessible in some applications such as monitoring volcanoes or space exploration. For that reason, algorithms and protocols designed for WSNs should consider the energy consumption in their conception [6,7]. The density of sensor nodes in the interested area may vary spatially and temporally depending on the application requirements [8]. In case of a high density of nodes, the sensing process is likely

* Corresponding author at: Paradise Research Laboratory – SITE, University of Ottawa, Ottawa, Canada

E-mail addresses: lvillas@site.uottawa.ca, leandro@dcc.ufmg.br (L.A. Villas), boukerch@site.uottawa.ca (A. Boukerche), horacio@dcc.ufam.edu.br (H.A.B.F. de Oliveira), regina@dc.ufscar.br (R.B. de Araujo), loureiro@dcc.ufmg.br (A.A.F. Loureiro).

1570-8705/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.adhoc.2011.08.005

to produce a large amount of redundant or spatially correlated data. In this case, nodes spatially close tend to collect similar information. Since the reduction of energy consumption is a key issue to increase the lifetime of the network, spatial correlation can be exploited to manage the set of nodes that collect redundant/similar data. The next issue is how to select a subset of these nodes to report the data without losing much accuracy.

For more efficient data gathering with a minimum use of limited resources, sensors should be configured to report data more intelligently by making local decisions. Data aggregation and spatial correlation are possible techniques for local decision-making. Such strategies help to maximize energy conservation in an application-specific sensor network [9–11].

A possible strategy to cope with data redundancy is to explore spatial correlation. This technique has been explored in the literature [12–16] but none of these studies focus on the selection of appropriate representative nodes according to their residual energy. A representative node reports event information of a given area on behalf of a group of nodes that collects similar readings in the same area. This is an interesting strategy since no node will be penalized by sending all notifications regarding the detected event. In doing so, the node consumes more energy than the other nodes, which also detected the same event. The Energy-Efficient Data Collection (EEDC) algorithm [17] proposes a similar strategy of considering the energy of the nodes but it is based on an impractical assumption that nodes can directly communicate with the sink in single hop.

In this work, we go further and present the dYnamic and scalablE tree Aware of Spatial correlaTion (YEAST) algorithm. Based on the notion of correlated region, which is the region sensor nodes collect similar readings, the proposed algorithm maximizes data aggregation along the communication route, and decreases the costs in the route discovery. A hybrid routing approach composed of proactive and reactive parts is proposed. The proactive part is responsible for discovering the positions of both the neighbors and the sink used later to construct the routing infrastructure and the routing itself. The reactive part is responsible for clustering nodes aware of spatial correlation and constructing data aggregation aware routes to transmit the collected data. Finally, the YEAST algorithm uses dynamic routes to ensure load balancing in the delivery of data.

The YEAST algorithm uses spatial correlation and a hybrid approach to overcome most of the drawbacks of the already proposed solutions. The main contributions of our solution are: (i) the created routes do not depend on the order of events and are not held fixed during the occurrence of events as in static routing approaches [18,19]; (ii) it has low communication overhead, which improves the quality of the routing tree and maximizes data aggregation along the communication route by creating a routing tree with a small number of nodes; (iii) it balances the residual energy of sensor nodes, reducing energy consumption by eliminating redundant notifications; and (iv) it dynamically adjusts the number of representative nodes, i.e., it adjusts the size of the correlated region in the event according to the application requirements and event characteristics.

In the YEAST approach, nodes that detected the same event are grouped in correlated regions and are responsible for applying the spatial correlation mechanism to select the representative node in each correlation region. The sink node, following the current application characteristics, defines the correlated region in which only one reading is sufficient and representative for the event reading precision. Note, however, that the size of the correlated region can be changed dynamically in order to achieve the required accuracy of the sensed information. The entire region of sensors per event is effectively a set of representative nodes performing the task of data collection and aggregation.

This paper is organized as follows. Section 2 provides an overview of the existing approaches exploiting spatial correlation. Section 3 discusses in-network data aggregation and presents the related work. Section 4 shows our newly proposed YEAST algorithm. The performance evaluation of the algorithm is described in Section 5. The applicability of the YEAST algorithm is discussed in Section 6 and, finally, Section 7 presents our conclusions and future work.

2. Exploiting spatial correlation in WSNs

In this section, we present the benefits of exploiting spatial data correlation in WSNs. We also discuss some of the existing approaches and algorithms that take advantage of spatial correlation in WSNs. As an initial motivation, Fig. 1 presents the energy consumption in the process of data collection in a WSN of two different routing approaches when the sink node, located at position (0,0), receives data from a detected event that has a radius of 70 m and is located at position (600,600). The first approach (Fig. 1a) is a simple method for data collection where all nodes that detected the event send the sensed data toward the sink node. The second approach (Fig. 1b) is a more sophisticated strategy that uses spatial correlation to save energy. In this case, only a subset of nodes that detected the event sends sensory data to the sink node. In both scenarios, the notification of the detected event was performed at each second, and the event duration was of only 10 m. The first approach (Fig. 1a) sends 32,157 notifications, whereas the second approach (Fig. 1b) sends only 5667 notifications.

The difference between the two approaches is notable and, by using spatial correlation, the second approach was able to save a large amount of energy, extending the overall network lifetime.

A WSN may remain idle until the occurrence of an event, when it reacts and builds a routing infrastructure to deliver the sensed data. This is known as a reactive routing strategy since the infrastructure is built only when necessary. Reactive routing can save energy during idle periods in an event-driven WSN [20]. In proactive routing, on the other hand, a routing infrastructure is built and maintained even in the absence of events, which can increase the energy consumption but has the advantage of decreasing latency. Since both approaches have their pros