Computer Communications 36 (2013) 1054–1066

Contents lists available at SciVerse ScienceDirect

Computer Communications

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

An energy-aware spatio-temporal correlation mechanism to perform efficient data collection in wireless sensor networks

Leandro A. Villas [a,b,e,*], Azzedine Boukerche [a], Daniel L. Guidoni [c], Horacio A.B.F. de Oliveira [d], Regina Borges de Araujo [e], 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] Department of Computer Science, Federal University of São João del-Rei, São João del-Rei, Brazil
[d] LAMP – Research Laboratory, Federal University of Amazonas, Manaus, Brazil
[e] WINDIS – Research Laboratory, Federal University of São Carlos, São Carlos, Brazil

ARTICLE INFO

Article history:
Available online 16 April 2012

Keywords:
Wireless sensor networks
Spatial correlation
Temporal correlation
Data collection

ABSTRACT

Large scale dense wireless sensor networks (WSNs) will be increasingly deployed in different classes of applications for accurate monitoring. Due to their high density of nodes, it is very likely that information that is both spatially and temporally correlated can be detected by several nodes what can be exploited to save energy, a key aspect on these networks. Furthermore, it is important to take advantage of these correlations to decrease communication and data exchange. However, current proposals usually result in high delays and outdated data arriving at the sink node. In this work, we go further and propose a new algorithm, called Efficient Data Collection Aware of Spatio-Temporal Correlation (EAST), which uses shortest routes for forwarding the gathered data toward the sink node and fully exploit both spatial and temporal correlations to perform near real-time data collection in WSNs. Simulation results clearly indicate that our proposal can sense an event with a high accuracy of more than 99.7% while still saving the residual energy of the nodes in more than 14 times when compared to the accurate data collection strategy reported in the literature.

1. Introduction

Wireless sensor networks (WSNs) [1–4] can be defined as a cooperative network of small, battery-operated, wireless sensor nodes whose main goal is twofold: to monitor their surroundings for local data and to forward the gathered data toward a sink node using typically multihop communication. This sink node will then be responsible for processing all of the received data from several source nodes and reporting them to a monitoring facility. This network architecture allows a number of novel monitoring-based applications in several areas such as environmental, medical, industrial, and military.

One of the main limitations of the WSNs is the battery-operated nature of their sensor nodes, which makes this kind of network highly energy-constrained. Thus, saving energy should be one of the main concerns of protocols and applications in WSNs. Since communication among the nodes is one of the main sources of energy consumption, most protocols in WSNs try to avoid or delay communication until it is really required [5–11]. However, by doing so, outdated and/or incomplete information is usually obtained by the sink node making the underlying application neither reliable nor useful.

In our work we go further from only delaying or avoiding communication by proposing a solution that gets the most out of each required communication. We do so by exploiting both the high density of WSN nodes and the observed similarity of nearby gathered data. This not only extends the lifetime of a WSN but also provides near real-time information about the monitored area. Our proposal extends some current published studies [12–14], which show that in several WSN applications, nearby nodes data tend to be correlated in both time and space:

- *Temporal correlation*: the change pattern of current sensor readings is equal or similar to the readings observed at previous times.
- *Spatial correlation*: the change pattern of the data sensed by nearby nodes is expected to be the same or similar.

These correlations have been exploited by some current techniques such as spatial suppression [15–19] and temporal

* 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), guidoni@dcc.ufmg.br (D.L. Guidoni), 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).

0140-3664/$ - see front matter Crown Copyright © 2012 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comcom.2012.04.007

suppression [12,20]. However, as mentioned before, these solutions not only introduce delays in data transmissions but also lead to the reception of outdated information by the sink node [21,22,20].

In our solution, we save energy not by delaying or suppressing messages, but by combining correlated information to make a better use of the data communication. In our proposed algorithm, named EAST (Efficient Data Collection Aware of Spatio-Temporal Correlation), sensor nodes are clustered under a spatial correlation approach while the leader and representative nodes execute a temporal suppression technique. The leader node generates a representative value for the whole cluster based on data received by the representative nodes, which form a subset of all nodes sensing the same event.

One of the key aspects of our solution is that it dynamically adjusts itself according to both the event characteristics and the residual energy of the sensing nodes, which are commonly ignored by current proposed solutions. Thus, in the EAST algorithm (i) residual energy of sensor nodes is balanced, (ii) energy consumption is reduced by eliminating redundant notifications, and (iii) both the number of representative nodes and error threshold are dynamically adjusted according to the event characteristics and accuracy requirements. In order to provide a better understanding of the behavior of our proposed algorithm in different scenarios, we present an extensive set of experiments that show the need for new algorithms (specially for real-time applications) and also clearly indicate the good performance of the proposed EAST algorithm when compared to related approaches.

This paper is organized as follows. In the next section, we discuss some proposed solutions that exploit spatial or temporal correlation in WSNs. In Section 3, we introduce our proposed solution, the energy-aware spatio-temporal correlation mechanism (EAST algorithm). In Section 4 we present a detailed performance evaluation of our solution by analyzing the obtained simulation results. Finally, in Section 5, we make some final discussions and present our conclusions.

2. Related work

In the current literature, we can find three main categories of data correlation protocols: (i) spatial correlation; (ii) temporal correlation, and (iii) spatio-temporal correlation. Table 1 presents a summary of the basic characteristics of the main proposed spatial and/or temporal data correlations algorithms for WSNs. In the following, we present some of these studies as well as the benefits of exploiting spatial/temporal data correlation in WSNs.

2.1. Spatial correlation

The spatial correlation of sensory information among the nodes that detect an event exists when those nodes are geographically close, i.e., they have similar information. In this case, instead of having all sensor nodes reporting the same data, it is more efficient to choose a few representative nodes to notify the sink node about the detected event. A representative node reports the event information of a given area on behalf of a group of nodes that collects similar information in the same area.

Sensor readings about the environment are typically periodic; consequently, the time-ordered sequence of sensed data constitutes a time series. Due to the nature of the physical phenomenon, there is a significant temporal correlation among each consecutive observation of a sensor node and gathered data is usually similar over a short-time period. Thus, in these cases, sensor nodes do not need to transmit their readings if the current reading is within an acceptable error threshold regarding the last reported reading. The sink node can just assume that any unreported data is unchanged from the previously received ones. The degree of correlation between consecutive sensor measurements might vary according to the characteristics of the phenomenon.

Akyildiz et al. [15] studied the relation between reliability of event detection and spatial location of the sensor nodes in the event area. Their solution estimates the number of sensor nodes (representative nodes) required to send the detected event to the sink in order to have reliable event information. Each representative node represents a spatially correlated group of nodes. Although their solution achieves overall energy gain, it fails to consider the remaining energy during the selection of the representative nodes – an assumption that should not be neglected in a WSN because of hardware constraints. Thus, if a representative node works in the correlation region for a long period of time, it will spend more energy due to the number of transmitted messages compared to the other nodes.

Yoon and Shahabi [18] proposed a new mechanism for spatial correlation in WSNs. The proposed mechanism, called Clustered Aggregation Technique (CAG), creates clusters of nodes with similar sensing values and only a node inside the cluster notifies its reading to the Sink node whereas the other nodes ignore their readings. The CAG algorithm is divided into two phases: query and response. In the query phase, the data-centric clusters are created according to a user-specified error threshold τ. Nodes that have sensed values smaller than this threshold belong to the same cluster. In the second phase (response phase), just one node per cluster (cluster-head) sends its sensed value to the sink node notifying the detected event. The authors showed that the proposed mechanism can reduce significantly the number of transmitted messages during the data collection. However, during the first phase, the CAG algorithm uses a flooding-based protocol to disseminate the query to all sensor nodes, which is not needed in most scenarios. Moreover, the maintenance of the data-centric clusters remains a difficult problem [23].

Liu et al. [24] proposed another clustering algorithm, named Energy-Efficient Data Collection framework (EEDC), to exploit spatial data correlation. They consider that nodes collect data

Table 1
Summary of the basic characteristics of the main proposed spatial and/or temporal data correlations algorithms for WSNs.

Scheme	Route structure	Objective	Spatial correlation	Temporal correlation	Overhead	Scalability	Drawback
EEDC	Single hop	Eliminate control overhead	Yes	No	Very Low	Very Low	Centralized and single-hop network
CAG	Tree-based cluster	Eliminate data redundancy	Yes	No	Very High	Medium	Maintenance data-centric
GSC	Tree-based cluster	Eliminate data redundancy	Yes	No	High	Low	Is not applied to multi-hop members
SBR	Tree-based	Eliminate data redundancy	No	Yes	Medium	High	Sink node can receive outdated information
SCCS	Tree-based cluster	Eliminate data redundancy	Yes	Yes	Medium	High	Sink node can receive outdated information