Information Sciences 329 (2016) 461–477

Contents lists available at ScienceDirect

Information Sciences

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

Energy-efficient node scheduling algorithms for wireless sensor networks using Markov Random Field model

Hongju Cheng [a], Zhihuang Su [a], Naixue Xiong [b,c,*], Yang Xiao [d]

[a] College of Mathematics and Computer Science, Fuzhou University, Fuzhou, FJ 350108, China
[b] Key Lab of Modern Optical System and Engineering Research Center of Optical Instrument and System, University of Shanghai for Science and Technology, Shanghai 200093, China
[c] Department of Business and Computer Science, Southwestern Oklahoma State University, Weatherford, OK 74074, USA
[d] Department of Computer Science, The University of Alabama, Tuscaloosa, AL 35487, USA

ARTICLE INFO

Article history:
Received 10 May 2014
Revised 9 September 2015
Accepted 15 September 2015
Available online 9 October 2015

Keywords:
Wireless sensor networks
Correlation optimization
Node scheduling
Markov Random Field
Energy efficiency

ABSTRACT

One important way to extend the lifetime of wireless sensor networks is to deploy the sensors in a dense manner. The redundancy among the sensed data demonstrates that it is not efficient to collect raw data from all nodes in the network if we further consider that the data is generally spatial-correlated. The node scheduling strategy aims at selecting a set of representative nodes to provide the required data service in a periodic manner with accuracy guarantee. This strategy can effectively reduce the number of active nodes and the amount of messages in the network, and extend the network lifetime accordingly. In this paper, we firstly introduce how to model the spatial correlation among sensed data by Markov Random Field (MRF) model. Secondly, we formulate the problem definitions, namely, the data amendment problem which maximizes the data coverage range for a given node by amending the raw noise-corrupted data from the neighbors, while the representative nodes selection problem focuses on reducing the number of representative nodes and covering all nodes in the network, and the node scheduling problem aims at maximizing the network lifetime. Thirdly, we propose a novel Data Amendment Procedure (DAP), Representative node Selection Procedure (RSP) and energy-efficient Node Scheduling Algorithm (NSA) respectively for these above problems. Finally, extensive experiments demonstrate that the proposed node scheduling algorithm can significantly improve the network lifetime compared with related works with an average increment of about 80%.

1. Introduction

A wireless sensor network generally consists of a large number of sensor nodes which are self-organized and connected by wireless communications [37]. The sensor node is equipped with the sensing devices, micro-processor, limited memory, wireless transmitter and batteries. It can collect information from the local environment and send to the application center, i.e., the sink. The wireless sensor networks can be widely used in many different fields including data gathering, remote monitoring, factory automation, smart home and security. For example, a classical application is to collect the environmental data [11], in which each node periodically collects local information within the sensing range, such as the temperature, light, humidity, etc. In

* Corresponding author. Tel.: +1 4046454067.
E-mail addresses: cscheng@fzu.edu.cn (H. Cheng), zhihuangsu@126.com (Z. Su), neal.xiong@swosu.edu, xiongnaixue@gmail.com (N. Xiong), yangxiao@cs.ua.edu (Y. Xiao).

http://dx.doi.org/10.1016/j.ins.2015.09.039

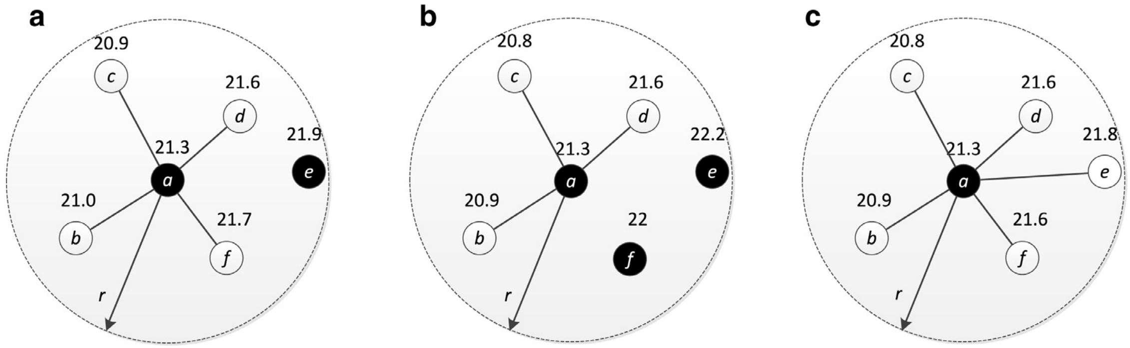

Fig. 1. An example for the representative nodes selection scheme: (a) using X, (b) using Y, (c) using Y'.

these applications, energy efficiency is always one of the most important criterions since nodes are generally powered by limited batteries and it is difficult to recharge them in some extreme environments. One important ways to improve the network lifetime is to deploy the network in a dense manner so that the monitored area is still covered even though some nodes run out of energy.

However, to collect all raw data to the destination is not efficient especially in cases that the sensed data is a noise version of the realistic observed phenomenon in the same area. Firstly, it is generally tolerant if the final collected data is within a given error threshold since there is a deviation between raw sensed data and physical value of the environment [38]. Secondly, the sensed data of adjacent nodes are generally correlated when they are located in the same geographical area [8,12]. Node i can *represent* another node, i.e., node j, if the sensed data y_i of node i is correlated with y_j of node j. The correlation between y_i and y_j can be denoted as $d(y_i, y_j)$ which is generally determined by the practical characteristics of different applications. There are several definitions proposed to describe the data correlation between the sensed data [15], such as: (1) relative error: $d(y_i, y_j) = \frac{|y_i - y_j|}{\max(y_i, c)}$, where $c > 0$ and c is a sanity bound in case that $y_i = 0$; (2) absolute error: $d(y_i, y_j) = |y_i - y_j|$; and (3) squared error: $d(y_i, y_j) = (y_i - y_j)^2$. By following the above observations, a subset of nodes can be selected as the representative to provide the required sensing service within the error guarantee and the rest nodes keep idle to conserve energy. This strategy can not only reduce the number of data messages sent to the sink, but also help to solve other issues in dense wireless sensor networks [12], such as the transmission conflicts, lower network throughput, etc.

Although there are several related works concerning with the representative node selection problem, they generally select the represent nodes by using the raw data directly, and simply reduce the node selection problem to the set cover problem (or variants) [4,12,15,16,18,29]. The simulation results show that these schemes can effectively reduce the number of representative nodes, and thus prolong the network lifetime. However, there is a deviation between the practical sensed data and real value of the environment [29,36] due to the effects of real environment, node parsimony and some other factors. It means that the selected node set with the above solutions might be incorrect since the data deviation is not considered at all.

Let $x_i(t)$ denote the noise-free data of node i at time t. Node i can represent the neighbor nodes when it is selected as the representative; otherwise, node i shall be represented by another node as mentioned in the related node selection problems [12,15,18]. Let $e_i(t)$ represent the noise for node i at time t. The noise-corrupted data for node i at time t can be formulated as $y_i(t) = x_i(t) + e_i(t)$. Here we demonstrate an example to illustrate the above issue. Let $Y = \{y_1, y_2, \ldots\}$ denote the set of raw noise-corrupted data, and $X = \{x_1, x_2, \ldots\}$ denote the set of noise-free data. Assume that it is tolerant to collect data with the error threshold ε of 0.5. Fig. 1(a) and (b) show the results for the node selection problem by using X and Y separately, where each circle denotes one node and the data value is marked above the circle. The sensing radius is denoted as r, and there is an edge between node a and other nodes if they can communicate directly and the data deviation is smaller than the given error threshold ε. The selected representative nodes are marked with black solid circle. As we can see from Fig. 1(a), the selected representative node set by using X is $\{a, e\}$, and $\{a, e, f\}$ by using Y. In this example, the number of representative nodes with Y is larger than that by using X.

It is obviously not proper to select the representative nodes directly by the noise version of sensed data with the above observation. In this paper, we first introduce how to describe the correlation among adjacent nodes with Markov Random Field (MRF) model. Then we describe how to amend the raw noise-corrupted data with a novel Data Amendment Procedure based on the above MRF model. We also propose a novel representative nodes selection procedure in which the representative are selected in a two-steps manner and the noise-corrupted data is amended accordingly.

Fig. 1(c) denotes the proposed data amendment scheme, in which $Y' = \{y'_1, y'_2, \ldots\}$ is the set of final amended data with the mentioned data amendment procedure. As shown in Fig. 1(b), node f and e are initially out of the data coverage range of node a. The Mean Square Error (MSE) is one efficient way to describe the noise value in the wireless sensor networks [19,39]. It is easy to know that MSE between the raw noise-corrupted data Y and X is 0.033 and 0.007 between the amended data Y' and X. Obviously, the noise of Y' is not larger than Y. The selected representative nodes set by using Y' is $\{a\}$. This example demonstrates that an efficient data amendment scheme is helpful to reduce the number of selected nodes by utilizing both the dense deployment manner and data correlation in the network.