Information Sciences 181 (2011) 869–882

Contents lists available at ScienceDirect

Information Sciences

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

Top-k query evaluation in sensor networks under query response time constraint

Weifa Liang [a,*], Baichen Chen [a], Jeffrey Xu Yu [b]

[a] School of Computer Science, Australian National University, Canberra, ACT 0200, Australia
[b] Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong, Shatin, NT, Hong Kong

ARTICLE INFO

Article history:
Received 17 June 2008
Received in revised form 14 May 2009
Accepted 8 October 2010

Keywords:
Top-k queries
Wireless sensor network
QoS constraint
Routing trees
Query optimization
Energy conservation
Query response time
Network lifetime

ABSTRACT

Top-k query in a wireless sensor network is to find the k sensor nodes with the highest sensing values. To evaluate the top-k query in such an energy-constrained network poses great challenges, due to the unique characteristics imposed on its sensors. Existing solutions for top-k query in the literature mainly focused on energy efficiency but little attention has been paid to the query response time and its effect on the network lifetime. In this paper we address the query response time and its effect on the network lifetime through the study of the top-k query problem in sensor networks with the response time constraint. We aim at finding an energy-efficient routing tree and evaluating top-k queries on the tree such that the network lifetime is significantly prolonged, provided that the query response time constraint is met too. To do so, we first present a cost model of energy consumption for answering top-k queries and introduce the query response time definition. We then propose a novel joint query optimization framework, which consists of finding a routing tree in the network and devising a filter-based evaluation algorithm for top-k query evaluation on the tree. We finally conduct extensive experiments by simulation to evaluate the performance of the proposed algorithms, in terms of the total energy consumption, the maximum energy consumption among nodes, the query response time, and the network lifetime. The experimental results showed that there is a non-trivial tradeoff between the query response time and the network lifetime, and the joint query optimization framework can prolong the network lifetime significantly under a specified query response time constraint.

1. Introduction

Technological advances in recent years have enabled the deployment of large-scalable sensor networks consisting of hundreds or thousands of inexpensive sensors in an ad hoc fashion for a variety of environmental monitoring and surveillance applications including measurement of meteorological data (e.g. temperature, pressure, humidity), noise levels, chemicals, etc. [1,17]. During the course, a large volume of sensed data generated by sensors is needed to be aggregated within the network, and the result is needed to be relayed to the base station to respond to users queries. The sensor network thus is treated as a *virtual database* by the database community [14]. This type of database, however, is essentially different from the traditional database. For example, it is impossible to store all sensed data in a central site (the base station) due to the large volume of continuing sensing data generated by sensors. On the other hand, the battery-powered sensors will quickly become inoperative due to the large quantity energy consumption if all sensed data were sent to the base station for storage

* Corresponding author. Tel.: +61 2 6125 3019; fax: +61 2 6125 0010.
E-mail address: wliang@cs.anu.edu.au (W. Liang).

0020-0255/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2010.10.006

or processing. Furthermore, the lifetime of the network is closely tied to energy consumption rates at the sensors. Hence, how to query the network effectively and efficiently is an important and challenging issue. In particular, radio communication (transmission and reception) is the primary source of energy consumption in sensor networks [19], minimizing radio communication in query execution can save large amounts of energy and thereby prolonging the network lifetime significantly. Several studies on different query optimization in sensor networks have been conducted recently [21,10,2].

Query optimization in traditional databases has been extensively studied, and typical optimization metrics are the query response time and space required. Unlike their wired counterparts, query optimization in wireless sensor networks focused on energy efficiency in order to prolong the network lifetime. Since energy conservation has dominated most of the research in energy-constrained sensor networks, the concepts of query response time, end-to-end delay and jitters have not been taken into account in most published works. However, along with the introduction of imaging and video sensors, the increasing interest in many real-time applications of wireless sensor networks, including disaster management, combat field and security surveillance, has posed additional challenges. That is, in these applications our optimization objective is not only to prolong network lifetime but also to meet the certain end-to-end delay constraint (e.g. the query response time).

Top-k query is one of very popular queries in wireless sensor networks, which is to find the k nodes with the highest values (readings) among the sensor nodes. One such an example is a sensor network deployed for monitoring the air pollution index of a region of interest. A typical top-k query issued to the network is to find the k sensors with the highest pollution index readings at any given time point. Although several studies on top-k query evaluation and maintenance in sensor networks have been conducted recently [23–25], none of them takes the query response time into consideration while focusing on energy efficiency only. However, in some real-time applications, the query response time is very critical. For example, consider a sensor network used to detect the forest fires. When a forest fire happens, the forest management authority may issue a top-k query to request the vicinity images of the k sensors with the highest temperature readings. Such a query has a stringent query response time imposed, because timing is very crucial for fire fighters to distinguish the fires. The faster the system responds to the query, the sooner the fire fighters are able to figure out the whereabout of fires accurately and timely.

1.1. Related work

Top-k query evaluation in distributed environments has been extensively studied in the literature [7,3,4,15,29]. Typically, there are two types of top-k queries. One is the distributed top-k query which aims to find the k highest ranked objects, where the ranking score of an object is an aggregated value from a number of attribute values stored at distributed sources. Many algorithms for this type of top-k query have been developed, which includes the Threshold Algorithm (TA) [7], the Three-Phase Uniform Threshold algorithm (TPUT) [4], the KLEE algorithm [15] and the distributed Threshold Join Algorithm (TJA) [29]. Another is to find the k nodes with the highest readings in a sensor network, assuming that each node generates a sensing reading. For this latter one, Wu et al. [24,25] exploited the semantics of top-k query and proposed a novel Filter-based algorithm (FILA) for monitoring the top-k results. Formally speaking, their algorithm maintains the top-k readings continuously by assigning a specific interval of key values for each sensor node, according to the current top-k key values, and this interval serves as the filter for the node to suppress its unnecessary updates. Meanwhile, the base station keeps a copy of the filter of each sensor node to maintain an (error bound) approximate view of the node's reading. A sensor node reports its updated reading to the base station only when the reading passes its filter. Otherwise, the updated reading will be ignored. Consequently, the total energy consumption for top-k query evaluation can be reduced significantly through the reduction of unnecessary data transmission within the network. The energy savings delivered by their solution however is based on the assumption that each sensor node is within the transmission range of the base station, each updating probe broadcast by the base station can be heard by all sensor nodes, and the total reception energy consumption of all sensor nodes was not taken into account. In real life, this assumption is too restrictive, the total reception energy consumption of all sensor nodes by receiving the base station's probing message cannot be ignored, because it is well known that the reception energy consumption of a sensor node is about one third of its transmission energy consumption in most short-distance wireless communications. For example, the reception energy consumption on MICA2 mote is 14.4 mJ/s, while its transmission energy consumption is only 36 mJ/s [6,23]. Therefore, if the sensing readings among sensor nodes are updated frequently, the probing cost will become prohibitively expensive. Silberstein et al. [23] considered the top-k query problem in sensor networks by providing several approximate solutions with high probability, based on top-k sampling of the past readings. They demonstrated the power and flexibility of sampling-based approach by formulating the problem under an energy constraint as linear programming and developed a series of top-k query planning algorithms, assuming that the sensor network is a tree network.

A closely related problem that considered both the query response time and the network lifetime is the data gathering problem subject to the end-to-end delay constraint. In fact, the top-k query with the response time constraint is a special case of this general setting. Despite that there are several studies on data gathering that tradeoff the end-to-end delay and the network lifetime [20,11,28], they are either inapplicable or have their limitations on this special case, because they assume that either the message length is fixed and given in prior or the data transmission rate at each node is dynamically adjustable. For example, Lindsey et al. [11] proposed an optimization metric *energy* $\times$ *delay* for data gathering and demonstrated that a chain-based routing tree delivers the longest network lifetime. However, finding such a chain in sensor