Data & Knowledge Engineering 102 (2016) 78–95

Contents lists available at ScienceDirect

Data & Knowledge Engineering

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

An efficient top-*k* query processing framework in mobile sensor networks

Heejung Yang[a], Chin-Wan Chung[b,a,*], Myoung Ho Kim[a]

[a] The School of Computing, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea
[b] Chongqing Liangjiang KAIST International Program, Chongqing University of Technology, Chongqing, China

ARTICLE INFO

Article history:
Received 8 December 2014
Received in revised form 28 September 2015
Accepted 5 February 2016
Available online 22 February 2016

Keywords:
Distributed databases
Database applications
Mobile sensor networks
Top-*k* query processing

ABSTRACT

Mobile sensor networks consist of a number of sensor nodes which are capable of sensing, processing, communicating and moving. These mobile sensor nodes move around and explore their surrounding areas. Top-*k* queries are useful in many mobile sensor network applications. However, the mobility of sensor nodes incurs new challenges in addition to the problems of static sensor networks (i.e., resource constraints). Since mobile sensor nodes tend to move continuously, the network condition changes frequently and they consume considerably more energy than static sensor nodes. In this paper, we propose an efficient top-*k* query processing framework in a mobile sensor network environment called mSensor. To construct an efficient routing topology, we devise a mobility-aware routing method. Using the semantics of the top-*k* query, we develop a filter-based data collection method which can save the energy consumption and provide more accurate query results. We also devise a data compression method for disconnected sensor nodes to deal with the problem of limited memory space of sensor nodes. The performance of our proposed approach is extensively evaluated using synthetic data sets and real data sets. The results show the effectiveness of our approach.

1. Introduction

Recent advances in robotics and low power embedded systems have enabled the deployment of mobile sensor networks for many applications. Mobile sensor networks implant mobility into static sensor networks. In a static sensor network, static sensor nodes are deployed in a region and collect data periodically in their static locations. Static sensor nodes have capabilities of sensing, processing, and communication. On the other hand, mobile sensor networks are composed of mobile sensor nodes and each mobile sensor node not only has capabilities of the static sensor node, but also moves around the neighborhood. This mobile sensor node explores its surrounding and exchanges information with its peers through wireless communication.

A top-*k* query in sensor networks returns the *k* nodes with the highest (or lowest) sensor readings.Top-*k* queries are widely used for retrieving the k most interesting data in many mobile sensor network applications such as environmental monitoring, disaster and emergency management, and military surveillance. They can be used to not only monitor the data generated by sensors in real time but also perform data analysis for archival purposes or research study. Consider an environmental monitoring example. Environmental engineers want to know about contaminated areas and to take appropriate actions for the contamination. However, these areas might be too dangerous to access in person. In this case, mobile sensors can be used. These

* Corresponding author.
E-mail addresses: heejung@islab.kaist.ac.kr (H. Yang), chungcw@kaist.edu (C.W. Chung), mhkim@kaist.ac.kr (M.H. Kim).

http://dx.doi.org/10.1016/j.datak.2016.02.001
0169-023X/

mobile sensor nodes are equipped with chemical or pollution sensors to detect the amount of toxic chemicals and GPS sensors to obtain location information. Each mobile sensor node moves around and generates sensor data (*nodeID, GPS data, sensing value*). A top-*k* query is issued to find out the locations with the highest sensing values.

To compute a top-*k* query in a sensor network environment, it is necessary to retrieve sensor readings from multiple sensor nodes and compare their values to generate the results. A naive implementation of top-*k* query processing is to use a centralized approach in which all sensor readings are transmitted to the base station and then the base station computes the top-*k* result. If a certain sensor reading is smaller than the *k*-th sensor reading of the top-*k* result, the transmission of this data is unnecessary. However, this approach transmits all sensor readings regardless of their values, thus it consumes too much energy. Since sensor nodes are battery-powered and deployed in an unattended manner, it is not easy to replace their batteries. Therefore reducing the energy consumption is a major concern in sensor networks. Batteries of sensor nodes are depleted by sensing, computation, and communication. Among these tasks, communication is the primary source of the energy consumption [1-3]. Thus several techniques have been proposed to resolve the limited energy constraints by reducing the communication. An *in-network aggregation* technique, initiated in TAG [4], is one of them. In this approach, a routing tree rooted at the base station is first established, and the data are transmitted along the routing tree. An intermediate sensor node compares the received data with its own sensor reading and sends the top-*k* sensor readings to the parent. Therefore, this approach can reduce the energy consumption compared to the centralized approach. Since users are often not interested in exact answers and small errors can be accepted in many scenarios, approximate top-*k* query processing techniques are proposed. Silberstein et al. [3] developed a sampling-based approach. They use samples of the past sensor readings and formulate the problem of optimizing approximate top-*k* queries under the energy constraint as a linear program. FILA [5] is another approximation approach using a filter. The basic idea is to install a bounded filter at each sensor node to suppress unnecessary sensor updates. A sensor node transmits its sensor reading only when the sensor reading exceeds the bounded filter.

However, all of these approaches are based on the static sensor network environment. In mobile sensor networks, the mobility of sensor nodes incurs new challenges in addition to the problems of static sensor networks (i.e., resource constraints). Since mobile sensor nodes tend to continuously move around, the network conditions change frequently and they consume considerably more energy than static sensor nodes. In static sensor networks, once the routing topology is constructed, it can be used throughout the whole network lifetime, although some minor changes might be required because of node failures or bad link qualities, etc. However, in mobile sensor networks, the routing topology is changed frequently because of the movement of mobile sensor nodes. This incurs a significant routing overhead and increases the energy consumption. Therefore, we need a new routing method to deal with the network dynamics and the energy limitation of mobile sensor nodes. To reduce the energy consumption in data collection, it is important to avoid unnecessary transmissions. In top-*k* query processing, an unnecessary transmission means that the transmission of data which has a smaller sensing value than the *k*-th sensing value in the top-*k* result. Therefore, we need an energy-efficient data collection method while considering the semantics of the top-*k* query. Some sensor nodes might be disconnected from the network. In this case, a disconnected sensor node has to store its data until the network is connected. Since we cannot store all data in the disconnected sensor node because of the limited memory space, we need a space-efficient data storing method.

In this paper, we propose mSensor which is an efficient top-*k* query processing framework in mobile sensor networks. In order to deal with the network dynamics and the limited energy constraints of sensor nodes, a mobility-aware routing method is proposed together with relevant routing criteria. The *Analytical Hierarchy Process* (AHP) is used to construct the best route from all routing factors according to the user's preferences of the routing factors. The filter-based data collection method using the semantics of the top-*k* query is devised to reduce the energy consumption. Since sensor data whose sensing value is smaller than the *k*-th sensing value in the local top-*k* result cannot be part of global top-*k* result, it is not necessary to transmit that sensor data. That is, *k*-th sensing value in the local top-*k* result can be used as a filter to filter out unnecessary transmissions. In order to deal with the limited memory constraints of sensor nodes, a data compression method using the Haar wavelet transform and thresholding is proposed to store data that have not been transmitted. Since Haar wavelet transform reduces the amount of data very effectively, it can be adapted well in resource limited sensor networks [6]. Thresholding that discards some detail coefficients from the Haar wavelet transform is used for compact representation of the data. Our approach can save the energy in dealing with the problems caused by the mobility and resource constraints of mobile sensor nodes and provide a sufficiently accurate top-*k* result. The contributions of this paper are as follows:

- We propose mSensor which is an efficient top-*k* query processing framework in a mobile sensor network environment.
- We devise an efficient routing method to deal with the network dynamics and the energy limitation of sensor nodes. There are various factors that affect the quality of a route. We define routing criteria to represent important route characteristics for top-*k* query processing in mobile sensor networks. Based on the routing criteria, we provide an effective method to select the best route.
- We devise a filter-based data collection method to reduce the energy consumption which is more critical for mobile sensor nodes than static sensor nodes. We exploit the semantics of the top-*k* query to both increase the filtering efficiency and provide a more accurate result.
- We devise an effective data compression method for disconnected sensor nodes. Since sensor nodes have very limited memory constraints, we provide a data compression method using the Haar wavelet transform which can be performed with low computing power of sensor nodes, and thresholding in a way that is advantageous to top-*k* query processing.