Journal of Network and Computer Applications 58 (2015) 208–216

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Smartphone-based data collection from wireless sensor networks in an urban environment

Zuhal Can*, Murat Demirbas

Department of Computer Science and Engineering, University at Buffalo, SUNY, Buffalo, NY 14260, United States

ARTICLE INFO

Available online 11 September 2015

Keywords:
Multi-hop data collection
Human mobility
Smartphone
Wireless sensor network

ABSTRACT

Using smartphones as mobile basestations and leveraging human mobility is a promising approach for urban data collection from Wireless Sensor Networks (WSNs). In this paper, we evaluate the feasibility of this approach applying analyses on a city-wide mobility dataset. Our spatial analysis shows that popular locations cluster close to each other and sensor nodes located in rarely visited locations can transmit their data in a few hops to smartphones visiting these popular locations. Our energy-efficiency analysis indicates the feasibility of employing energy-conserving approaches on both smartphones and WSN nodes based on mobility behavior of smartphone users. We evaluated and compared on-demand and continuous data collection protocols on several WSN islands with different size and connectivity regarding to data collection efficiency. Our simulation results show that continuous data collection protocols surpass on-demand data collection protocols in terms of data delivery ratio and latency. We found that data collection protocols run more efficiently in many-connected small islands compared to fewer connected large islands.

1. Introduction

Achieving data collection from Wireless Sensor Network (WSN) has been traditionally performed by placement of static pre-constructed devices, such as basestations (Can and Demirbas, 2013). However, in addition to having a cost of construction and disposal, an increase in the amount of static infrastructures increases the global energy consumption (Fettweis and Zimmermann, 2008). As a solution to this problem, smartphones pose a significant interest for the WSN research community due to their potential to serve as mobile basestations in an urban environment. Allowing nodes to communicate with smartphones without resorting to a static infrastructure enhances the energy conservation and expands the sustainable management of natural resources. For example, 1096 relay nodes are deployed to collect CO_2 data from 100 sensor nodes scattered in Mao et al. (2012). Smartphones-based data collection from these sensor nodes would decrease the cost of static relay node deployment and also alleviate the connectivity problems caused by failure-prone relay nodes.

Owing to their ubiquitousness and portability, smartphones are suitable for mobile data collection from WSNs in an urban environment. Smartphones as mobile data collectors can collect WSN data and transmit data to a centralized controller. Even though data collection by smartphones causes a delay due to the human mobility, there are some applications that can tolerate this delay, such as habitat, garbage bins, and power monitoring applications (Mainwaring et al., 2002; http://research.cens.ucla.edu/urban/; Katz et al., 2011).

While there have been studies on data collection by mobile data collectors in WSN (Can and Demirbas, 2013; Di Francesco et al., 2011; Soysal and Demirbas, 2010) and on mobile sensing by smartphones (Lane et al., 2010; Macias et al., 2013; Khan and Khan, 2010; Macías et al., 2011; Sendra et al., 2014), there are very few efforts for evaluating smartphones as mobile data collectors from WSNs (Wu et al., 2014). Mobility datasets have been studied for understanding collaborative social behaviors of people (Laurila et al., 2013; Zheng et al., 2009, 2008, 2010). While human mobility accumulates on special locations, there can be regions that lack regular visitors; however, studies discussing the impacts of the human mobility on the smartphone-based data collection from WSNs have not investigated data collection from unvisited regions (Wu et al., 2013; Park and Heidemann, 2011; Shepard et al., 2011). Therefore, a comprehensive analysis is required to highlight the correlation between the human mobility and the smartphone-based data collection from WSNs in an urban environment including unvisited regions. This analysis should also address the energy-related concerns on smartphones and WSN nodes.

To address this problem, we analyzed a GPS trajectory dataset collected from Beijing, China (http://research.microsoft.com/en/us/

---

* Corresponding author.
E-mail addresses: zuhaltep@buffalo.edu (Z. Can), demirbas@buffalo.edu (M. Demirbas).

http://dx.doi.org/10.1016/j.jnca.2015.08.013

projects/GeoLife/). For our analyses, we divided the target region into geographical cells assuming that each cell has a stationary sensor node that is monitoring the corresponding area. We examined this trajectory dataset under spatial and energy-efficiency analyses. In the *spatial analysis*, the spatial distribution of frequently and infrequently visited cells is explored to understand the feasibility of smartphone-based data collection in infrequently visited cells and also in isolated areas, which we call islands. Our analysis shows that islands can be managed as a WSN for transmitting data to smartphones. The *energy-efficiency analysis* discusses the energy-related concerns on smartphones and WSNs. Our analysis investigates several duty cycling schedules based on smartphone user visits for energy management on WSN islands. We evaluate the smartphone-based data collection efficiency on islands of different size and connectivity with a comparison of on-demand and continuous data collection protocols.

Some of our findings are as follows:

- We categorized geographical cells in the Beijing area according to the GPS readings falling in each cell. We call frequently visited cells as hot cells and infrequently visited cells as cold cells. Most of the cold cells have a distance of less than 5 hops to a hot cell. This is a convenient distance for a node in a cold cell for transmitting its data in multi-hop manner to a visiting smartphone user in a hot cell.
- As a subset of cold cells, we denote almost never visited cells as frozen cells. We noticed that frozen cells are not arbitrarily scattered in the Beijing area, but clustered tightly. We grouped the clustered frozen cells and call it an island. The existence of islands represents that there are regions that lack regular visitors. Since islands occur adjacent to city roads, multi-hop data collection is feasible from island nodes by smartphone users traveling on city roads.
- In order to investigate data collection from islands, we managed each island as a distinct WSN and run on-demand and continuous data collection protocols on each island. While an on-demand protocol provides energy-efficient data collection, if energy-efficiency is not a main requirement for a data collection protocol, continuous data collection protocols perform better with respect to data delivery ratio and latency.
- To achieve energy-efficiency at the smartphone side, we notice opportunities regarding data collection period. Most participants have a mobility period of less than 10% during their participation. If smartphone radio is only turned on when the smartphone user is mobile, smartphones can run a data collection application rarely and consume less energy.
- For conserving energy at WSN, nodes can run a duty cycling algorithm based on smartphone user visits. Since at 91.5% of cell visits, visitors are coming from a neighbor cell, a duty cycling schedule can be developed based on neighbor cell visits. Similarly, WSN nodes can run a duty cycling algorithm according to the smartphone users' visit frequency in their cells.

The rest of the paper is organized as follows: Section 2 describes the GPS trajectory dataset. In Sections 3 and 4, the spatial and energy-efficiency analyses are explained, respectively. In Section 5, we evaluate the data collection efficiency by smartphones on various WSN islands. Section 6 presents related efforts on the formation of islands. Section 7 concludes the paper and suggests directions for future work.

2. Dataset

For our analyses, we exploit the participants' mobility traces which are collected in the Geolife project between 04/12/2007 and 07/27/2012 [15]. For building this dataset, 182 people participated for the Geolife project (accumulating a total duration of more than 45,000 h) by recording their daily outdoor movements by GPS logger devices and GPS recording phones. Participants of the dataset are both pedestrians and vehicle passengers. This dataset consists of mobility trajectories including GPS readings. A GPS reading entry includes the latitude, longitude, and altitude information along with a time stamp as shown in Table 1. 91% of the trajectories are logged frequently, e.g. every 1–5 s or every 5–10 m per point (http://research.microsoft.com/en/us/projects/GeoLife/).

Table 1
Fields of a GPS entry with a sample.

Field	Value
Latitude	40.000031
Longitude	116.326378
Unoccupied flag	0
Altitude	492
Date (number of days)	39902.0362037037
Date as a string	2009-03-30
Time as a string	00:52:08

We focused on the GPS dataset collected within the Beijing city that contains 18,670 trajectories and 19,137,967 GPS readings. Due to the massiveness of the dataset, we split the dataset area into equal-sized cells in order to simplify the management. Splitting up a city's area into geographical cells is a popular method to work with massive dataset which is also used in several studies (Castro et al., 2013; Liu et al., 2010). For our analysis, we assume that a sensor node is deployed at each geographical cell. Our analyses are performed on geographical cells of 100 m^2 size.

Geolife trajectories may not be perfectly accurate. In an urban area, high buildings can split source signals of the GPS device and may cause erroneous readings with an accuracy of a few meters. The Geolife dataset is found to have an accuracy of 3 m in a previous study (Zhao et al., 2014).

In this dataset, we performed our analyses on a specific spatial portion and a temporal slot. We selected the central portion of the Beijing city for our analyses since the GPS readings are collected mostly from this region, as shown in Fig. 1a. We trimmed the Geolife dataset for extracting the GPS readings on this central portion, which is about a 10 km^2 area falling between the coordinates [40.02032658N, 116.292898703E] and [39.93041245N, 116.412874925E]. We are going to refer to this selected portion as *the Beijing area* in this paper. For the temporal slot, we selected the year 2009 since most of the GPS readings on the Beijing area are collected in this year, as demonstrated in Fig. 1b. During this time slot, 67 participants were active in the Beijing area. Participants' activities and the weekly GPS readings in the Beijing area during year 2009 are demonstrated by Fig. 1c and d.

3. The spatiality analysis

3.1. Hot and cold cells

In the Beijing area, cells are visited with varying frequencies. While some of the cells have no GPS readings, some of them can have thousands of GPS readings. The total GPS readings falling in a cell indicates how frequently this cell is visited. We categorized geographical cells in the Beijing area with respect to the total GPS readings falling in each cell. If the total GPS readings falling in a cell is less than a threshold value during a year, we call this cell a *cold cell*, otherwise, we call it a *hot cell*.

We select the threshold value of 365, representing the minimum total amount of daily visits per year. We assume that if at least 365