Pervasive and Mobile Computing 9 (2013) 881–891

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Analysis of smartphone user mobility traces for opportunistic data collection in wireless sensor networks☆

Xiuchao Wu, Kenneth N. Brown, Cormac J. Sreenan*

CTVR, Department of Computer Science, University College Cork, Ireland

ARTICLE INFO

Article history:
Available online 12 July 2013

Keywords:
Human mobility
Smartphone
Wireless sensor network
Opportunistic data collection

ABSTRACT

The increasing ubiquity of smartphones coupled with the mobility of their users will allow the use of smartphones to enhance the operation of wireless sensor networks. In addition to accessing data from a wireless sensor network for personal use, and the generation of data through participatory sensing, we propose the use of smartphones to collect data from sensor nodes opportunistically. For this to be feasible, the mobility patterns of smartphone users must support opportunistic use. We analyze the dataset from the Mobile Data Challenge by Nokia, and we identify the significant patterns, including strong spatial and temporal localities. These patterns should be exploited when designing protocols and algorithms, and their existence supports the proposal for opportunistic data collection through smartphones.

1. Introduction

As wireless sensor networks mature, we expect to see many long-term and large-scale deployments for various applications, such as environmental monitoring, domestic utility meter reading, and urban monitoring. Since the increasingly ubiquitous smartphones are tightly-coupled with their users, the interaction between smartphones and wireless sensor networks will play a very important role in future pervasive computing. For instance, a smartphone could get various information (temperature, air quality, etc.) from sensor nodes around its user and assist in making informed decisions. In such cases, it is normally assumed that smartphones and sensor nodes can communicate through some low power radios, such as Bluetooth and IEEE 802.15.4.[1] Also, smartphones have been proposed to act as sensor nodes in participatory sensing [1]. In this paper, instead of the above classical paradigms, we consider using smartphones to provide a service to wireless sensor networks, i.e., using smartphones to collect data from sensor nodes opportunistically (and relay to their corresponding servers).

As illustrated in Fig. 1, we have proposed to use smartphones carried by people in their daily life to collect sensor data opportunistically when their users pass by sensor nodes [2–5]. Under this scenario, smartphones will gather data from sensor nodes autonomously (without any user intervention or route change). To participate in opportunistic data collection, a smartphone user just needs to run a background application on the phone, and many users could be motivated with a very low reward. For instance, the owners of wireless sensor networks could reward these users by allowing them to access

☆ The preliminary results have been presented in The Mobile Data Challenge 2012 (by Nokia) Workshop, in conjunction with Pervasive 2012, Newcastle, UK.

* Corresponding author. Tel.: +353 (0)21 4205892.
E-mail address: cjs@cs.ucc.ie (C.J. Sreenan).

1 Bluetooth is distributed with almost all smartphones and it is also adopted by many sensor nodes, such as IMote and BTnode. IEEE 802.15.4 is the most widely used radio on sensor nodes and it starts to appear on smartphones. In Mobile World Congress 2012, TazTag released the first smartphone with both ZigBee (IEEE 802.15.4 radio, protocol stack, etc.) and Near-Field Communication (NFC) features (http://www.taztag.com/).

1574-1192/$ – see front matter
http://dx.doi.org/10.1016/j.pmcj.2013.07.003

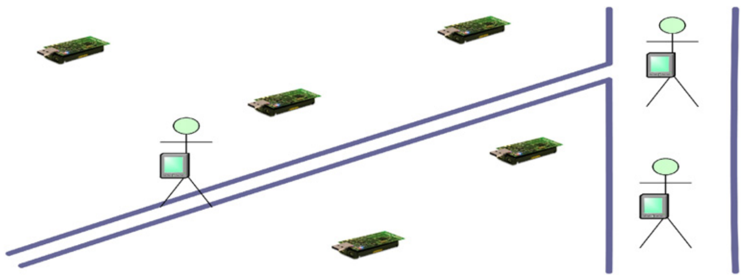


Fig. 1. Opportunistic data collection through smartphones.

the current sensor readings (temperature, humidity, etc.). In case that the sensor readings are not needed by smartphone users or the sensor readings cannot be publicized due to confidential and privacy reasons, these users could be rewarded by a small amount of virtual/real money through the cellular network system. Consequently, the cost of data collection can be reduced through exploiting the uncontrolled mobility of smartphone users. The incentive, security, and privacy issues that arise in opportunistic data collection through smartphones have been discussed further in [5] and they are beyond the scope of this paper.

Considering that the main point of opportunistic data collection is to exploit the *uncontrolled* mobility of smartphone users, we need to establish that the mobility of smartphone users could support this scheme, especially for wireless sensor networks in which sensor nodes are duty-cycled aggressively for longevity. The Mobile Data Challenge by Nokia [6] made available a large dataset consisting of, among other things, detailed mobility traces for smartphone users around the Lake Geneva Region, and we use the traces to establish whether or not the mobility of smartphone users is suitable. When analyzing the traces, we ask the following key questions:

1. In opportunistic data collection, is the smartphone's overhead (energy consumption, CPU, etc.) low enough so that the participation of smartphone users could be motivated with a very low reward?
2. For each encounter between a smartphone and a sensor node, does the smartphone stay in the communication range of the sensor node long enough to collect data opportunistically?
3. Could smartphone users visit a sensor node frequently enough to support a variety of applications?
4. How does the smartphone users' mobility distribute in time and space? How do these distributions influence the design and operation of the protocols and algorithms for opportunistic data collection?

This paper is organized as follows. The analysis methodology is first introduced in Section 2. We also describe how the dataset is trimmed. The results of analysis are then presented and discussed in Section 3. Finally, Section 4 discusses related work and Section 5 concludes this paper with several key findings, such as the feasibility of opportunistic data collection through smartphones and the strong spatial and temporal localities that should be considered when designing the protocols and algorithms for opportunistic data collection.

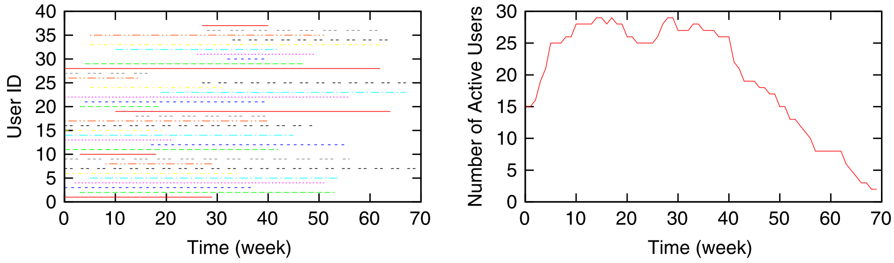
2. Data preparation

In this paper, the mobility of smartphone users is studied through analyzing the dataset from the Mobile Data Challenge by Nokia. Although a wide variety of information was collected for each smartphone user, we are mainly interested in the GPS readings recorded when a user was moving around outside. More specifically, we only use the following information of a GPS reading, $\langle time, latitude \text{ and } longitude, speed\rangle$, i.e., the time, the location, and the movement speed when this GPS reading was logged.

For opportunistic data collection, we are interested in how the encounters between smartphones and sensor nodes distribute in both space and time. Hence, the area visited by smartphone users is divided into grid cells with a size of one thousandth of a degree in both directions.[2] Approximately, a cell in the Lake Geneva Region is a rectangle with a size of 111 m $*$ 77 m and it matches well with the outdoor communication range of the current sensor node platform [7]. The duration of the Data Collection Campaign by Nokia is also divided into slots in units of hour, day, or week based on the analysis to be carried out. The distributions of GPS readings in time and space are then calculated and analyzed in this paper.

Before carrying out analysis, the dataset is first trimmed. We have removed a few GPS readings that are far away from the Lake Geneva region so that the number of cells to be considered can be reduced significantly. For reducing the number of time slots to be considered, the GPS readings which were logged when most of users had quit the Data Collection Campaign by Nokia are also removed. By trimming the dataset in this way, we are able to speed up the computation significantly without losing important data. Through removing these GPS readings, we can also avoid that the conclusions are skewed

2 Note that a cell here is totally different from the cell in cellular networks (i.e., the area covered by a base station). Instead of a circle/hexagon/square, it is just a small rectangle for simplifying our analysis. The two-dimensional GPS position of the northeast corner is $(x + 0.001°, y + 0.001°)$, assuming that (x, y) is the GPS coordinates of the southwest corner.



(a) Users' active periods.
(b) Active users vs. time.

Fig. 2. The participation of 37 smartphone users.

by the large areas and long periods in which the level of user participation is very low. Any GPS readings which have been truncated for user anonymity are also removed since we cannot associate such a reading to a specific cell.

Consequently, 893,920 GPS readings from 37 smartphone users are used in our analysis.[3] The latitude range is [46.1, 46.8], the longitude range is [6.4, 7.4], there are 700,000 cells, and the whole area is referred as the Lake Geneva Region. Sometimes, we only analyze the cells of the Lausanne Urban Area (one major city of the Lake Geneva Region), in which the latitude range is [46.50 46.55], the longitude range is [6.54, 6.66], and there are 6000 cells. The duration is from 05/09/2009 to 07/01/2011 and the time span is 70 weeks. Considering that smartphone users may not have participated for the whole period, based on the timestamps in their GPS readings, Fig. 2(a) plots the periods that these 37 users did actually participate in the Data Collection Campaign by Nokia. The level of user participation, i.e., the number of active smartphone users, is also plotted in Fig. 2(b).

3. Results of analysis

3.1. Percentage of movement time

Considering that a sensor node is normally powered by un-rechargeable battery, its radio must be duty-cycled for longevity. Hence, it is preferred to let a smartphone, with its rechargeable battery, always keep its radio on so that they can discover each other in a timely manner [3]. However, the energy consumed by a smartphone's radio for opportunistic data collection might become a serious concern.

Fortunately, we can reduce its energy consumption based on context information. A smartphone can deduce whether it is moving using its accelerometer [8,9]. It can then keep its radio on only when its user is moving around. In cases when its user is static, the smartphone can turn on its radio occasionally for collecting data and turn off its radio for most of the time to save energy. To study the energy overhead with this scheme, we need to know the percentage of time that a smartphone user is moving around.

In the dataset, a GPS reading is recorded every 10 s only when a user is moving around outside. Hence, if the interval between two consecutive GPS readings is too long (>300 s), we assume that the user is static and the radio can be turned off during that interval.[4] We then calculate the percentage of movement time for each smartphone user. Fig. 3(a) plots the cumulative distribution function (CDF) of the percentage of movement time across 37 users. It shows that for most smartphone users, the movement time is less than 10%. Hence, smartphone users are usually static and the radio for opportunistic data collection can be turned off most of the time. The overhead of opportunistic data collection in terms of energy consumption could therefore be low, thus encouraging user participation.

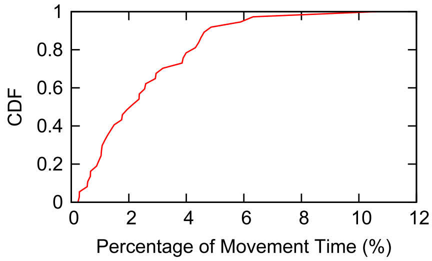
3.2. Movement speed

Since a sensor node is normally duty-cycled, a smartphone still needs to take time to discover a sensor node even when they are in close proximity. Furthermore, a smartphone and a sensor node normally belong to different authorities, and authentication must be carried out before collecting data. Hence, for opportunistic data collection, it is desired that a smartphone could stay in the communication range of a sensor node for a period that is sufficient for discovery, authentication, and data collection.

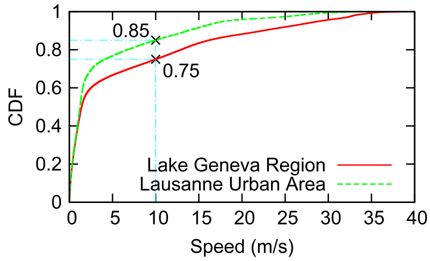
To check this issue, the cumulative distribution function of the movement speed in these smartphone users' GPS readings is plotted in Fig. 3(b). This plot indicates that the movement speed is quite low in many cases. In the Lausanne Urban Area,

[3] In the dataset obtained from Nokia, there are in total 1,553,154 GPS readings from 38 smartphone users. 491,566 GPS readings are purged because they have been truncated for user anonymity. Since only GPS readings in a few sensitive locations are truncated, these purged GPS readings do not affect the analysis results in this paper.

[4] Note that GPS readings could be absent due to many reasons. Here, we assume the dominant reason is that a smartphone user stops moving.



(a) Movement time.



(b) Movement speed.

Fig. 3. CDFs of movement time and movement speed.

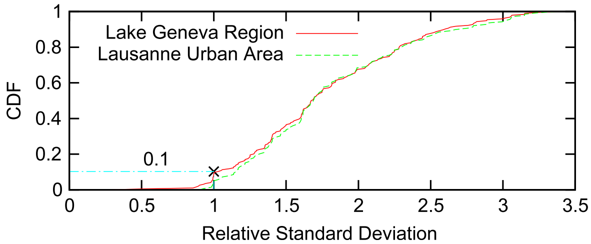


Fig. 4. Analysis of per-cell visits by smartphone users.

the speed for 85% of GPS readings is less than 10 m/s. Even for the much larger Lake Geneva Region with many roads, there are still 75% GPS readings whose speed is less than 10 m/s. Considering that the outdoor communication range of a sensor node is around 100 m, a lot of data could be collected during the encounter between a smartphone and a sensor node. With the assumptions that IEEE 802.15.4 radio is used (the data rate is 250 kbps) and the duration for data collection is 10 s, 312 K bytes can be collected per visit. Considering that the size of a sensor reading is normally small, thousands of sensor readings can be collected per visit. Fig. 3(b) also indicates that the movement speed may be high with non-negligible probability, even when only the Lausanne Urban Area is considered. This fact justifies our sensor node-initiated probing mechanism for timely discovery of these short encounters between sensor node and the fast-moving smartphone [3].

3.3. *Per-cell visits*

As mentioned earlier, a smartphone and a sensor node normally belong to different authorities, and some authentication scheme based on public key cryptography is needed for secure data collection. Hence, a smartphone and a sensor node may consume significant CPU, time, and energy for carrying out the related public key cryptography operations. In cases where a sensor node is repeatedly visited by a few smartphones, a hash-chain-based authentication scheme could be used by them to avoid carrying out the expensive public key cryptography operations during each encounter [10]. To verify whether a hash-chain-based authentication scheme could be applied, we quantified the cells that are visited at least once per day and by more than one user.

We investigated the visits that occur within each cell. For each smartphone user we counted the number of times that this user passed through this cell, by studying the relevant GPS readings. We then calculated the average and the standard deviation of these visit frequencies across all the users. Consequently, the relative standard deviation is calculated through dividing the standard deviation by the average.

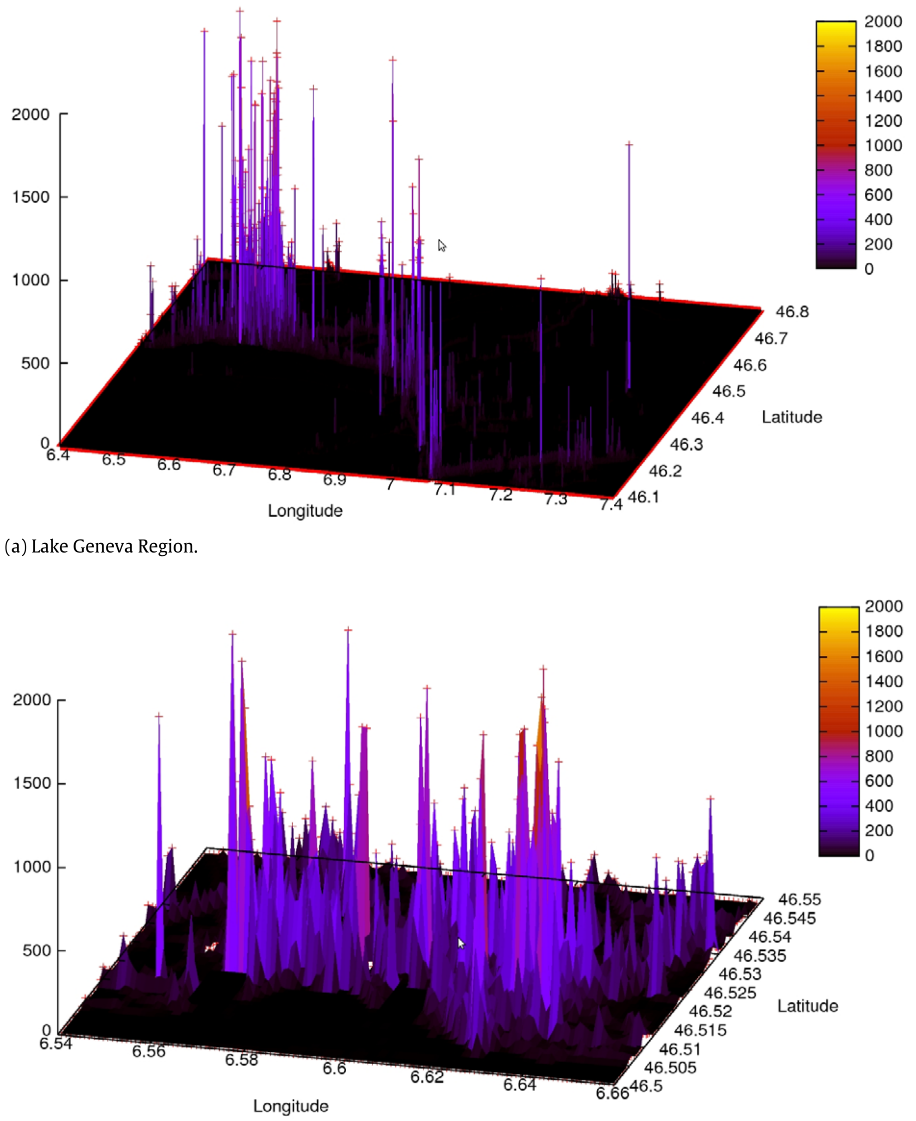
Fig. 4 plots the CDF of the relative standard deviation across these cells. It indicates that for most cells, the distribution of per-cell visits among smartphones has a large relative standard deviation (>1.0), and thus we observe that the visits to a cell are mainly contributed by a few users. We conclude that a hash-chain-based authentication scheme would be viable and the overhead of authentication in opportunistic data collection would be reasonable.

3.4. *Spatial analysis*

3.4.1. *Spatial distribution*

In the following analysis, we first calculate the number of GPS readings in each cell. We then plot the spatial distribution of GPS readings among all cells of the Lake Geneva Region in Fig. 5(a). The spatial distribution among cells of the Lausanne Urban Area is also plotted in Fig. 5(b).

Fig. 5(a) shows that the mobility traces of just 37 smartphone users still cover a large area, while Fig. 5(b) indicates that many cells in an urban area are visited frequently. Our analysis shows that 19% of cells in the Lausanne Urban Area are visited at least once per week and 2.466% of cells are visited at least once per day. Hence, we can expect that opportunistic data



(a) Lake Geneva Region.

(b) Lausanne Urban Area.

Fig. 5. Spatial distributions of GPS readings.

collection through smartphones can support many applications, especially when sensor nodes are deployed in urban areas where we live in most of the time.

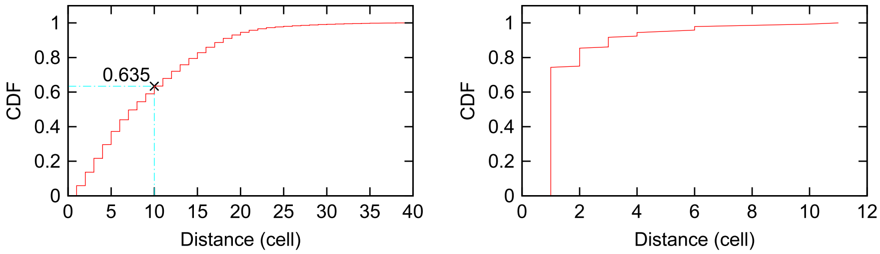
3.4.2. Spatial locality

Fig. 5(a) and (b) also indicate that a strong spatial locality exists in these distributions of GPS readings and different cells are visited by smartphone users with different frequencies. Through checking the map of Lake Geneva Region shown in Fig. 6, we find that Fig. 5(a) clearly illustrates that most of these GPS readings are within the towns alongside the A9 motorway of Switzerland. Fig. 5(b) indicates that even in the urban area, there are still some cells that have never been visited. There are also some *hot* cells that are visited much more frequently than other *cold* cells.

To study the spatial locality quantitatively, we have calculated the relative standard deviation of the spatial distribution of GPS readings in the Lausanne Urban Area. For each cell, we have the times that it is visited by smartphone users. We first calculate the average and the standard deviation of these numbers across all cells. The relative standard deviation is then calculated through dividing the standard deviation and the average, and its value is as high as 5.23. Hence, a strong spatial locality is identified and sensor data should flow among sensor nodes to improve the performance of opportunistic data collection through exploiting this spatial locality [4].

To study the feasibility of exploiting spatial locality, for the Lausanne Urban Area, a cell is marked as a hot cell if it is visited at least once per day. Otherwise, the cell is marked as cold cell. We then calculate the distance between a cold cell

Fig. 6. The map of Lake Geneva Region (captured from Google map).



(a) Cold cell.

(b) Hot cell.

Fig. 7. CDFs of the distance between a cold (hot) cell and its nearest hot cell.

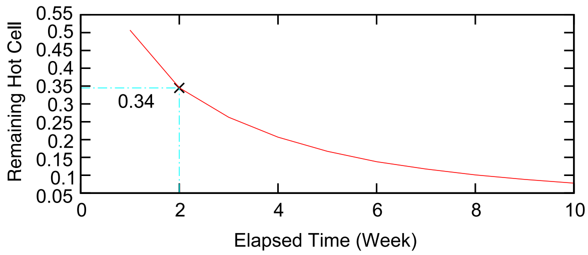


Fig. 8. The invariability of hot cells.

and its nearest hot cell. The cumulative distribution function of these distances is plotted in Fig. 7(a) and this plot shows that for 63.5% cold cells, the distance is less than ten cells. The distance could be reduced if the mobility traces of more users are considered. However, considering that human mobility is normally constrained by roads and streets, cold cells should continue to exist. Hence, sensor data should be exchanged among sensor nodes for exploiting spatial locality and the data could reach a hot cell through a few hops. We have also calculated the distance between a hot cell and its nearest hot cell. The result in Fig. 7(b) shows that for most of hot cells, one of its direct neighbors is also a hot cell. Hence, opportunistic data collection through smartphones is robust to the failure of sensor nodes in a hot cell. It also indicates that the neighboring hot cells tend to be visited sequentially and this characteristic should be exploited if the duty cycle of sensor nodes is not too low.

3.4.3. Seasonal changes

To exploit the spatial locality for opportunistic data collection, a hot cell should continue to be a hot cell for a long time so that sensor data will not chase the moving hot cells and consume too much energy to arrive at a current hot cell and be collected by a smartphone. Hence, for each week, we calculate the number of GPS readings for each cell and these numbers have been plotted into a 3-D figure. Several animations are then produced based on these figures to demonstrate the changes of the spatial distribution as time elapses. These animations are available at the official webpage of the Mobile Data Challenge by Nokia [11].

To study the seasonal changes of hot cells quantitatively, for each week, a cell in the Lausanne Urban Area is first marked as a hot cell according to the same criteria used in Section 3.4.2 (i.e., a hot cell is visited at least once per day). We then plot the percentage of hot cells that continue to be hot cells with the elapse of time. Fig. 8 shows that 34% of hot cells are still hot cells after two weeks. Hence, spatial locality is quite steady and it could be exploited in opportunistic data collection.

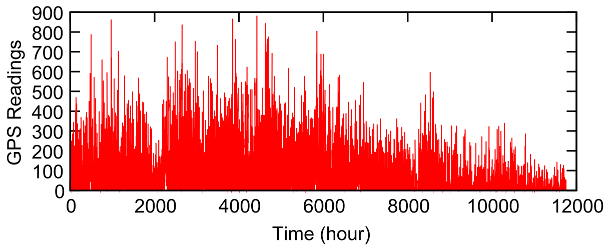
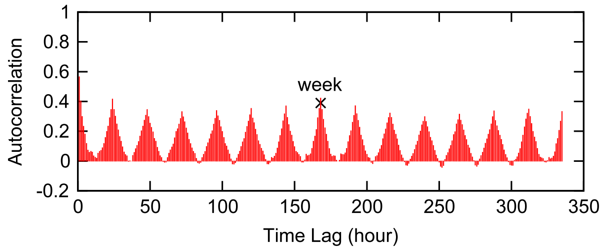
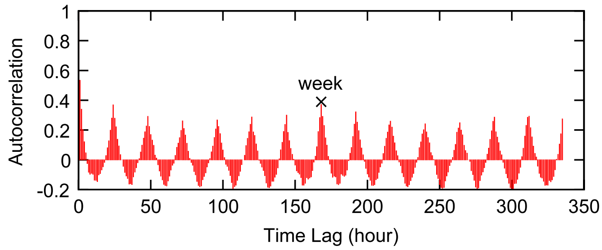


Fig. 9. Temporal distribution.



(a) The whole duration.



(b) Between the 15th and the 35th weeks.

Fig. 10. Autocorrelations with different time lags.

However, it also indicates that some changes do exist in long term and sensor nodes must learn and exploit the spatial locality online. These long-term changes may be caused by our choice of a low value for defining when a cell is hot. Some cells may be classified as hot or cold cell alternatively in different weeks. This issue merits further investigation as part of a broader user mobility study.

3.5. Temporal analysis

To carry out temporal analysis, the whole duration is divided into time slots of one-hour length. The number of GPS readings in each time slot is then counted and this temporal distribution is plotted in Fig. 9.

3.5.1. Period analysis

Previous studies find that human mobility normally follows some repeated patterns (diurnal, etc.) [12]. To check whether repeated patterns exist in smartphone users' mobility, autocorrelations of the above time series are calculated with different time lags and the results are plotted in Fig. 10(a). This plot indicates that the mobility of smartphone users does have a repeated pattern whose epoch length is 24 h.

However, the diurnal pattern is not obvious since there is no negative autocorrelation at a 12 h lag.[5] As illustrated in Figs. 2(b) and 9, one potential reason is that the number of active users and the number of GPS readings are reduced significantly in the late phase of the Data Collection Campaign by Nokia. Hence, period analysis is carried out again for the GPS readings between the 15th and the 35th week (2520–5880 h) during which the number of active users and the number of GPS readings are stable. The corresponding results of period analysis are then plotted in Fig. 10(b), which demonstrates the existence of the diurnal pattern clearly.

Furthermore, neither Fig. 10(a) nor (b) shows the common weekly pattern in human mobility. When the time lag is one week ($7 * 24 = 168$ h), the autocorrelation is only slightly higher than other time lags that are multiples of 24 h. This issue will be discussed later when we carry out per-cell analysis.

3.5.2. Temporal locality

In opportunistic data collection, if there are *rush* hours in which a sensor node is visited by smartphones much more frequently, a sensor node can discover smartphones mainly during rush hours so that it can upload the same amount of data with much less energy consumption [2]. Hence, we will check the existence of rush hours, i.e., temporal locality, in the mobility of smartphone users. Considering that the mobility of smartphone users has a strong diurnal pattern, the distribution of all GPS readings among 24 h of a day is then calculated and plotted in Fig. 11. This plot indicates that rush hours do exist in the morning (8 am) and evening (4–6 pm).

5 Due to the diurnal pattern followed by human mobility, for two time slots that are separated by 12 h, the mobility level during the time slot in daytime should be higher than the average and the mobility level during the other time slot in nighttime should be lower than the average. Thus, the autocorrelation at a 12 h lag should be negative.

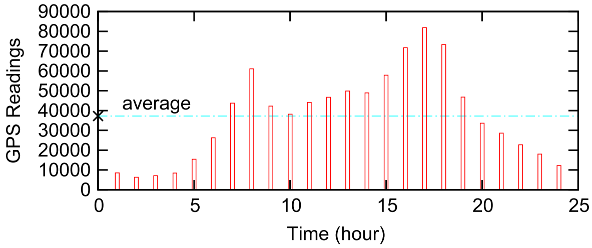


Fig. 11. The existence of temporal locality.

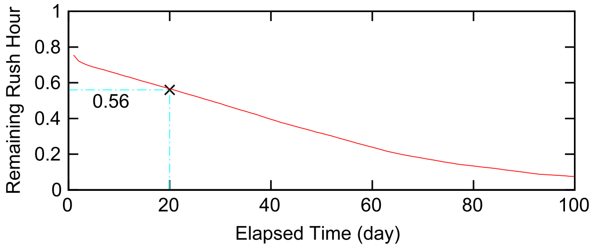
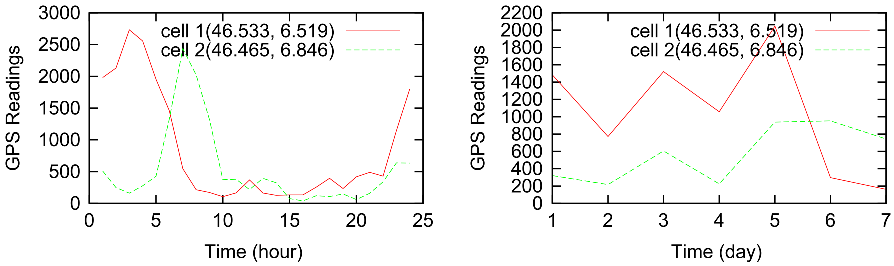


Fig. 12. The invariability of rush hours.



(a) Daily distribution. (b) Hot cell.

Fig. 13. Temporal distributions of two cells.

If an hour continues to be a rush hour for many days, a sensor node can learn and exploit the temporal locality easily. To study this issue quantitatively, for each day, an hour is marked as a rush hour if its number of GPS readings is larger than two times the average across 24 h. This large threshold is used to avoid too many hours being marked as rush hours. Fig. 12 then plots the percentage of rush hours that continue to be rush hours with the elapse of time. It shows that 56% of rush hours are still rush hours even after 20 days. Hence, temporal locality is quite steady and it could and should be exploited. However, Fig. 12 also indicates that rush hours stop being rush hours after a long period, seasonal changes do exist, and a sensor node should learn and exploit rush hours autonomously. These seasonal changes may be caused by our strict standard that one hour is a rush hour continuously. Even though one hour is a rush hour in most days, a few exceptional days will affect the results significantly. This issue merits further investigation as part of a broader user mobility study.

3.5.3. Per-cell analysis

We notice that in Fig. 11, the number of GPS readings in a rush hour is not much higher than the average. The possible reason is that the rush hours of various cells are different. They will cancel each other since we study the temporal locality for the whole area. To validate this conjecture, we carried out the following per-cell temporal analysis.

For two cells that are visited frequently, their distributions of GPS readings among 24 h of a day are calculated and plotted in Fig. 13(a). This plot clearly validates the above conjecture since these cells have different rush hours.

In the above period analysis in Section 3.5.1, we noticed that no weekly pattern existed in Fig. 10(a) or (b). This issue might be caused by the same reason, i.e., the period analysis is carried out for the whole area. Hence, for the above two cells, their distributions of GPS readings over 7 days of a week are plotted in Fig. 13(b). This plot shows that cell 1 is visited more frequently in weekdays and cell 2 is visited more frequently in weekends. Hence, weekly pattern may exist for some cells. However, due to the small numbers of GPS readings per cell, per-cell period analysis does not produce any meaningful results and these results are not reported here.

We note that cell 1 is around the A1 motorway between Geneva and Lausanne. Its rush hours are in late night and it is visited more frequently in weekdays. One possible explanation is that smartphone users visited cell 1 may be the drivers of

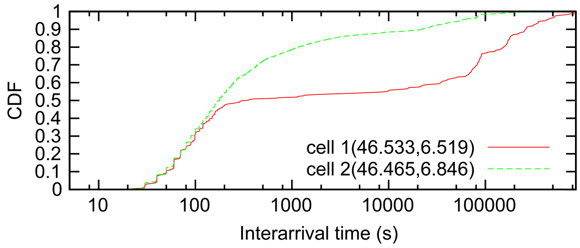


Fig. 14. CDF of the inter-arrival time.

heavy good vehicles who work at night. Cell 2 is in the city center of Vevey, Switzerland and it is much easier to understand its temporal distributions. The rush hours are in the morning since commuters arrive to offices during that period. It is visited more frequently in weekends because more persons come to city center for shopping and entertainment. Hence, different locations may have different facilities and they may be visited by different kinds of persons. It is hard for engineers to figure out the temporal distribution of smartphone users' visits for each sensor node and it should be worthwhile to let a sensor node learn its own situation autonomously.

To let smartphones and sensor nodes find each other efficiently, it could be helpful to design the probing scheme based on the distribution of the inter-arrival times of their encounters [13]. Hence, for each of the above two cells, we also calculate the intervals between the consecutive visits of smartphone users. Fig. 14 plots the cumulative distribution functions of their inter-arrival times. It indicates that the smartphone arrival patterns observed by cells are location-dependent. Instead of designing a probing scheme for all sensor nodes, it is better to let each sensor node adapt to its own situation.

In summary, the results of per-cell analysis indicate that there are no common repeated patterns, temporal locality, or inter-arrival times across all cells and a sensor node must autonomously learn and exploit the temporal distribution of its own location.

4. Related work

4.1. Mobile data collection

Due to the limited computing capability and storage size of sensor nodes, these nodes normally send their data to an application server through some dedicated static sink nodes with the aim of further processing [14]. However, due to environmental constraints and/or cost issues, sensor nodes tend to be deployed sparsely and these networks tend to be partitioned. Consequently, deploying large numbers of static sink nodes for collecting sensor data from these sensor nodes would incur prohibitive costs in terms of deployment, maintenance, and data backhaul. The cost of equipping each sensor node with a cellular network interface is even higher.

In [15–20], the use of mobile nodes has been proposed to move around in the deployed area and collect data from sensor nodes. Depending on the applications, their mobility can be either controlled or not, and these mobile nodes may collect data from sensor nodes within the range of one or multiple hops. In [21], the use of mobile phones had also been proposed to collect data from static sensor nodes purposely or opportunistically. However, none of them studied the scenario when the uncontrolled mobility of the public is considered.

Apart from the low data collection cost discussed in Section 1, opportunistic data collection through smartphones also has some benefits of adopting mobile sinks, such as the increased network reliability through removing the dependency on static sink nodes and the extended network lifetime through removing hot-spots near the static sink nodes [18,19]. Although data delivery latency could be high in opportunistic data collection, there are many promising wireless sensor network applications which are delay-tolerant. Thus, we have proposed opportunistic data collection through smartphones in [5]. Several protocols [2–4] are also designed for efficient data collection through exploiting the temporal and spatial localities of human mobility reported in [12,22,23]. The findings in this paper validate these human mobility patterns in a more appropriate spatial granularity and provide more directions to improve the performance of opportunistic data collection through smartphones.

One might consider using buses and/or postal carriers to collect data from sensor nodes, exploiting their regular mobility. This represents a special case of our proposal, one in which mobility patterns are somewhat more deterministic. Of course this approach is only suitable if the routes that are traversed by the bus/postal carrier provide sufficient overlap with the area of interest.

4.2. Human mobility analysis

Based on Wi-Fi users observed by two Wi-Fi access points (one is deployed in a residence building and the other is deployed in an academic building) in a one year period, human mobility was studied in [24]. It is confirmed that rush hours, i.e., temporal locality, does exist in human mobility. As for seasonal changes of rush hours, the existence depends on the locations of access points.

The mobility datasets of phone users have also been studied by the research community [12,22,23], and it has been pointed out that their mobility follows some repeated patterns and demonstrates strong temporal and spatial localities. However, in these datasets, only the current base station is recorded when a phone user communicates through a cellular network (call, short message, etc.). Hence, the phone user's location accuracy is as coarse as several kilometers or even tens of kilometers due to the large communication range of a cellular base station. Although the mobility analysis based on these datasets is valuable for urban planning, the location accuracy is too coarse for opportunistic data collection since the communication range of a sensor node is normally less than 100 m [7].

We believe that our study based on the dataset from the Mobile Data Challenge by Nokia is extremely valuable to opportunistic data collection through smartphones. It is the mobility traces of *smartphone users* that are analyzed in this paper and the location accuracy of GPS readings could be tens of meters, which is sufficient for opportunistic data collection.

5. Conclusion and future work

For the purpose of opportunistic data collection through smartphones, the smartphone users' mobility traces from the Mobile Data Challenge by Nokia are analyzed in this paper and our findings are summarized below.

1. Opportunistic data collection through smartphones should be a very promising solution. The overhead for the smartphone in terms of energy consumption and CPU can be very low and the mobility of smartphone users could provide a performance level that is sufficient for many wireless sensor network applications, especially when sensor nodes are deployed in urban areas.
2. The mobility of smartphone users follows some repeated patterns (diurnal, etc.) and the distributions in time and space have strong localities. When designing the protocols and algorithms for opportunistic data collection, these localities should be considered and exploited. For instance, a sensor node should try to discover smartphones mainly during *rush* hours [2], and sensor data should also be exchanged among sensor nodes to exploit the spatial locality of smartphone users' mobility [4]. Due to the existence of seasonal changes and the location-dependent mobility patterns observed by sensor nodes, sensor nodes should learn and exploit these localities autonomously.

In this paper, the used dataset only includes the mobility traces of 37 smartphone users. Some planned analysis (per-cell period analysis, etc.) cannot produce any meaningful results since there is insufficient data. In the case that a larger dataset becomes available, we will carry out this analysis to get more extensive results. Based on the above findings, we will refine our protocols proposed for opportunistic data collection through smartphones [2–4]. With the dataset from the Mobile Data Challenge by Nokia, these proposals will also be re-evaluated through trace-based simulations.

Acknowledgments

This work is supported in part by Science Foundation Ireland through CTVR (SFI 10/CE/I 1853) and by HEA PRTLI-IV through NEMBES. The authors thank the reviewers for their insightful comments and all people involved with the Mobile Data Challenge by Nokia for the dataset used in this paper.

References

[1] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, M.B. Srivastava, Participatory sensing, in: The 1st Workshop on World-Sensor-Web: Mobile Device Centric Sensory Networks and Applications (in Conjunction with ACM SENSYS), ACM, 2006, pp. 117–134.
[2] X. Wu, K.N. Brown, C.J. Sreenan, Exploiting rush hours for energy-efficient contact probing in opportunistic data collection, in: Proceedings of the 31st International Conference on Distributed Computing Systems Workshops, IEEE, 2011, pp. 240–247.
[3] X. Wu, K.N. Brown, C.J. Sreenan, SNIP: a sensor node-initiated probing mechanism for opportunistic data collection in sparse wireless sensor networks, in: Proceedings 1st International Workshop on Cyber-Physical Networking Systems, IEEE, 2011, pp. 726–731.
[4] X. Wu, K.N. Brown, C.J. Sreenan, Data pre-forwarding for opportunistic data collection in wireless sensor networks, in: Proceedings of the Ninth International Conference on Networked Sensing Systems (INSS), IEEE, 2012, pp. 1–8.
[5] X. Wu, C.J. Sreenan, K.N. Brown, A shared opportunistic infrastructure for long-lived wireless sensor networks, in: Proceedings of the 2nd International ICST Conference on Mobile Lightweight Wireless Systems Workshops, 2011, pp. 330–337.
[6] J.K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, M. Miettinen, The mobile data challenge: big data for mobile computing research, in: Mobile Data Challenge by Nokia Workshop, in conjunction with Int. Conf. on Pervasive Computing, Newcastle, UK, 2012.
[7] J. Polastre, R. Szewczyk, D. Culler, Telos: enabling ultra-low power wireless research, in: Proceedings of the Fourth International Symposium on Information Processing in Sensor Networks, 2005, pp. 364–369.
[8] D.H. Kim, Y. Kim, D. Estrin, M.B. Srivastava, Sensloc: sensing everyday places and paths using less energy, in: Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, ACM, 2010, pp. 43–56.
[9] Z. Zhuang, K.-H. Kim, J.P. Singh, Improving energy efficiency of location sensing on smartphones, in: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, ACM, 2010, pp. 315–330.
[10] L. Lamport, Password authentication with insecure communication, Communications of the ACM 24 (11) (1981) 770–772.
[11] The mobile data challenge by nokia, http://research.nokia.com/mdc.
[12] M.C. Gonzalez, C.A. Hidalgo, A.-L. Barabasi, Understanding individual human mobility patterns, Nature 453 (2008) 779–782.
[13] W. Wang, V. Srinivasan, M. Motani, Adaptive contact probing mechanisms for delay tolerant applications, in: Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking, ACM, 2007, pp. 230–241.
[14] I. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a survey, Computer Networks 38 (2002) 393–422.
[15] V. Dyo, S.A. Ellwood, D.W. Macdonald, A. Markham, C. Mascolo, B. Pasztor, S. Scellato, N. Trigoni, R. Wohlers, K. Yousef, Evolution and sustainability of a wildlife monitoring sensor network, in: Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, ACM, 2010, pp. 127–140.

[16] Z. Li, M. Li, J. Wang, Z. Cao, Ubiquitous data collection for mobile users in wireless sensor networks, in: Proceedings of the IEEE International Conference on Computer Communications, IEEE, 2011, pp. 2246–2254.
[17] B. Pasztor, M. Musolesi, C. Mascolo, Opportunistic mobile sensor data collection with SCAR, in: Proceedings of the IEEE International Conference on Mobile Adhoc and Sensor Systems, IEEE, 2007, pp. 1–12.
[18] R.C. Shah, S. Roy, S. Jain, W. Brunette, Data mules: modeling a three-tier architecture for sparse sensor networks, in: Proceedings of the 1st IEEE International Workshop on Sensor Network Protocols and Applications, IEEE, 2003, pp. 30–41.
[19] A.A. Somasundara, A. Kansal, D.D. Jea, D. Estrin, M.B. Srivastava, Controllably mobile infrastructure for low energy embedded networks, IEEE Transactions on Mobile Computing 5 (8) (2006) 958–973.
[20] M. Zhao, Y. Yang, Bounded relay hop mobile data gathering in wireless sensor networks, in: Proceedings of the 6th IEEE International Conference on Mobile Adhoc and Sensor Systems, IEEE, 2009, pp. 373–382.
[21] U. Park, J. Heidemann, Data muling with mobile phones for sensornets, in: Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, ACM, 2011, pp. 162–175.
[22] R.A. Becker, R. Caceres, K. Hanson, J.M. Loh, S. Urbanek, A. Varshavsky, C. Volinsky, A tale of one city: using cellular network data for urban planning, IEEE Pervasive Computing 10 (2011) 18–26.
[23] F. Calabrese, G.D. Lorenzo, L. Liu, C. Ratti, Estimating origin–destination flows using mobile phone location data, IEEE Pervasive Computing 10 (4) (2011) 36–44.
[24] V. Dyo, C. Mascolo, A node discovery service for partially mobile sensor networks, in: MIDSENS (in conjunction with MIDDLEWARE), 2007, pp. 13–18.