Using Location Based Social Networks for Quality-Aware Participatory Data Transfer

Houtan Shirani-Mehr
University of Southern California, Los Angeles, CA
hshirani@usc.edu

Farnoush Banaei-Kashani
University of Southern California, Los Angeles, CA
banaeika@usc.edu

Cyrus Shahabi
University of Southern California, Los Angeles, CA
shahabi@usc.edu

ABSTRACT

The sensing systems that monitor physical environments rely on communication infrastructures (wired or wireless) to collect data from the sensors embedded in the environment. However, in many urban environments pre-existing communication infrastructures are not available, and installing and using new infrastructures is unjustifiably expensive and/or technically infeasible. For such environments, we envision *Participatory Data Transfer* (*PDT*) as an alternative communication medium that leverages LBSN (Location Based Social Networks) for data collection. With PDT, LBSN users use their mobile devices to receive data from sensors, and forward the sensed data through the physical network of their mobile devices as well as their connections in the online/virtual social network until the data is received by the data aggregators (data collectors). In this paper, we elaborate on this vision in the context of *Quality-aware Participatory Data Transfer* (*Q-PDT*), where PDT must be designed such that it ensures quality guarantees for the sensed data (e.g., sufficiently covering and accurately sensing, timely delivery). In particular, we define and discuss variations of the Q-PDT problem and study its computational complexity.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Spatial databases and GIS*

General Terms

Algorithms

Keywords

Social Network, Location-based Service, Sensor Placement, Participatory Data Transfer

1. INTRODUCTION

With the recent technological advancements in developing low-power and inexpensive sensing devices, sensing systems that allow for real-time and accurate monitoring of physical environments are becoming prevalent [2, 11, 5]. In particular, there are numerous sensing applications for *urban* environment monitoring, such as exposure analysis (pollution, noise, etc.), hazard detection (e.g., chemical contamination, fire, flood), and urban traffic analysis (for vehicles and people). Such sensing systems often consist of a set of sensors that sense the urban environment, a set of aggregators (or sinks) where the sensed data is collected for further processing, and a communication infrastructure that enables data transfer between sensors and aggregators. For example, with a noise-exposure monitoring application a set of sound sensors are placed in an environment. The measurements from the sensors are transferred through a wireless network to a set of processing nodes (i.e., aggregators). At the aggregators, the noise exposure for the entire environment is estimated based on the measurements obtained from the sensors.

While in some urban environments sensing systems can use the existing wired or wireless (e.g., Wi-Fi) infrastructures for communication, in many other environments pre-existing communication infrastructures are not available and installing new infrastructures is unjustifiably expensive or infeasible. For the latter case, we envision *participatory data transfer* (*PDT*) as an alternative communication medium based on data propagation through the urban population that is equipped with mobile devices. In particular, we consider users of the location-based social networks (LBSN) that use their mobile devices to receive data from sensors, and forward the sensed data through the physical network of their mobile devices as well as their connections in the online/virtual social network until the data is received by the aggregators of the sensing system. The LBSN users who participate in PDT only exchange data with their "friends", i.e., other users to whom they are connected through LBSN.

To ensure sufficiently accurate monitoring of the urban environment, sensing applications must guarantee the "quality" of the sensed data in terms of its coverage of the urban environment. Accordingly, with PDT-based sensing applications a *quality-aware* implementation of the PDT process is required. Arguably, developing quality-aware PDT approaches translates to solving two optimization problems conjointly: 1) placement/deployment of the sensors and aggregators given the distribution and movement patterns of the urban population, and 2) routing of the data through the real and virtual network of the users participating in PDT. The quality-aware PDT approach must also satisfy a set of constraints (e.g., energy efficient routing to guarantee minimum power demand on mobile devices) to ensure applicability.

While sensor deployment and ad hoc routing/networking (or opportunistic networking) are two problems that are both extensively studied in the literature, with the problem of quality-aware PDT one needs to consider the requirements of the two aforementioned problems in conjunction. Moreover, the specific instances of sensor deployment and ad hoc networking with quality-aware PDT are different from their corresponding traditional cases. In particular, with sensor placement for quality-aware PDT, one needs to consider two often competing optimization criteria, i.e., coverage of the environment and access to the participating urban population. Also, an optimal routing approach for quality-aware PDT must consider specific movement patterns of the urban population.

In this paper, we elaborate on our vision of quality-aware PDT, and discuss different variations of this process given various PDT features and application settings. Moreover, we study the computational complexity of implementing the quality-aware PDT process and prove that it is NP-hard.

2. QUALITY-AWARE PARTICIPATORY DATA TRANSFER (PDT)

A sensing application which deploys PDT, includes a set of sensors and aggregators. It leverages PDT to transfer data from sensors to aggregators, where they are processed or further forwarded to a central processing unit via some existing communication infrastructure (wired or wireless). With such applications, participants collectively serve as a single data transfer medium to transfer data from sensors to aggregators. In this section, we formalize the problem of *Quality*-aware PDT for sensing applications that are designed to collect hight quality data.

2.1 Application Model

Imagine an application whose service depends on the quality of data (defined bellow) collected by sensors in an environment E during a time interval T. We assume the application can place at most N_S sensors and N_A aggregators in E. The set of sensors are denoted by $S{=}\{s_1, s_2, \ldots, s_{N_S}\}$, and the set of aggregators by $A{=}\{a_1, a_2, \ldots, a_{N_A}\}$. Each aggregator is either functional or non-functional (e.g., due to failure) at each instance of time where only functional aggregators can process packets. Figure 1 shows three sensors s_1, s_2 and s_3 and two aggregators a_1 and a_2 that are placed in the environment E. Each sensor $s_i \in S$ generates a data packet denoted by $p_{s_i \to D}(t)$ where $D{=}\{a_{j_1}, a_{j_2}, \ldots, a_{j_m}\}$ is the set of destination aggregators which can receive and process the packet, and t is the time at which the packet is generated. If $D{=}A$ then all the aggregators can receive the packet. The packet p is considered received if at least one functional aggregator in D receives it. We denote the earliest time at which $p_{s_i \to D}(t)$ reaches an aggregator in D by t_{r_p}. The packet also has a lifetime of T_p during which it should be received, otherwise the information carried by the packet becomes worthless to the application. For example, with a temperature monitoring application, the temperature readings carried by packets become stale for real-time monitoring purposes if received after some limited time period (e.g., an hour).

The quality of the data collected (and received) over T is denoted by $Q(P)$, where P is the set of all packets which are

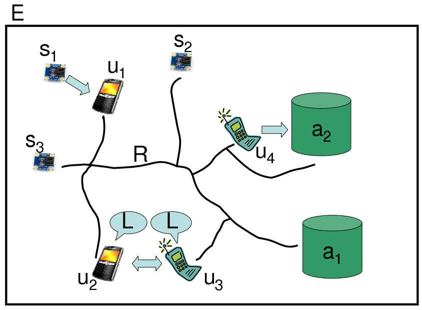

Figure 1: Sensors, aggregators and individuals who contribute to PDT

received by the aggregators within their lifetime and during T, i.e.,

$$P = \left\{ p_{s \to D}(t) | t_{r_{p_{s \to D}}} - t \leq T_{p_{s \to D}} \quad \& \quad t_{r_{p_{s \to D}}} \leq T \right\}. \quad (1)$$

The value of $Q(P)$ depends on the type of service provided by the application (we discuss specific possibilities in Section 4).

2.2 PDT Model

PDT (short for Participatory Data Transfer) is the process in which a set of individuals $U = \{u_1, u_2 \ldots, u_{N_U}\}$ contribute to data transfer. For example, in Figure 1 four individuals u_1, u_2, u_3 and u_4 are participating in PDT . In this case, we assume the individuals are restricted to move on a road network R (highlighted line segments in Figure 1) in E, and they can use their cell phones to exchange packets among each other (e.g., u_2 and u_3 in Figure 1) and exchange packets with sensors (e.g., u_1 and s_1 in Figure 1) and aggregators (e.g., u_4 and a_2 in Figure 1). A data packet can be transmitted from a device to another one if the devices are *reachable* from each other. Two devices are reachable if they are able to communicate data given the existing communication media. For example, if no wireless or wired connection exists, two cell phones are reachable when they are close enough to exchange data via Bluetooth or infrared. Moreover, we assume the cell phones of two individuals communicate if the individuals allow data exchange. In particular, we consider an LBSN L which captures friendship relationship between its users. We assume two users u_i and u_j can exchange packets on their cell phones at time t if u_i and u_j are friends of each other, i.e., $u_i \in L$ and $u_j \in L$ and they are in the friends lists of each other (this constraint can be generalized as needed). For example, u_2 and u_3 in Figure 1 are exchanging data packets as they are friends of each other according to L. When exchanging packets, individuals can transmit any subset of data packets stored on their cell phones. Finally, we assume an individual cannot transmit more than T_P packets over time T because of cell phone power limitations.

2.3 Quality-Aware PDT Problem

Given the application and PDT models discussed above, we define the problem of Quality-aware PDT (Q-PDT) as follow: Given a set of sensors S and a set of aggregators A, place all the sensors and aggregators, such that $Q(P)$ is maximized while the following constraints are satisfied:

1. PDT devices can transmit data when they are reachable from each other and
2. PDT devices do not transmit more than T_p packets.

3. COMPLEXITY

In this section, we prove that Q-PDT is an NP-hard problem by reduction from the optimal *sensor placement* problem [6] (which is also an NP-hard problem). The sensor placement problem can be formalized as follows.

DEFINITION 1 (SENSOR PLACEMENT). *Assume an environment E is represented by a discrete set of points V. Given k sensors to monitor a phenomenon in E, sensor placement places sensors in a set of points $F \subseteq V, |F| = k$, such that the prediction accuracy of the phenomenon throughout E is optimized:*

$$\arg\max \ H(V - F) - H(V - F|F) = F, \qquad (2)$$

where $H(.)$ represents the entropy function and therefore the set F is a set of points which maximally reduces the entropy over the rest of the space $V - F$.

The following theorem proves that the Q-PDT problem is NP-hard.

THEOREM 1. *Q-PDT is an NP-hard problem.*

PROOF. We prove the theorem by providing a polynomial time reduction from the sensor placement problem. Towards that end, we prove that given an instance of sensor placement, denoted by I_S, there exists an instance of Q-PDT, denoted by I_Q, such that the solution to I_Q can be converted to the solution of I_S in polynomial time. Assume a given instance I_S whose goal is to place k sensors to monitor a phenomenon during the time interval T'. We propose the following mapping from I_S to I_Q to reduce I_S to I_Q. The environments for both instances are the same, i.e., the set of environment points in I_S are mapped to environment points in I_Q. Furthermore, for both instances we monitor the phenomenon within the same time interval, i.e., T' is mapped to T. We assume there are k sensors and also k aggregators in I_Q and each aggregator can be placed with a short distance of $\epsilon \approx 0$ from a sensor. Each sensor s_i generates a packet $p_{s_i \rightarrow \{a_i\}}(t)$, where a_i is the closest aggregator to s_i. In this case, a sensor s_i can communicate directly with a_i due to small distance between them which eliminates the need for individuals to transfer packets to a_i. Finally, the quality of the transferred packets ($Q(P)$) is defined as the amount of reduction of entropy in predicting the phenomenon at unsensed locations after receiving the packets in P (this definition is based on the optimization criteria in Equation (2)). Given this mapping, it is easy to observe that if the answer to I_Q is the set of environment points F, F is also the answer to I_S. $\square$

We conclude from Theorem 1 that optimal solution for the Q-PDT problem is rendered unscalable as the spatial extent of the environment grows large. Hence, heuristic solutions should be developed to solve Q-PDT for large environments.

4. VARIATIONS OF Q-PDT PROBLEM

Given various settings of the PDT-based sensing applications as well as different features and limitations with PDT execution, the Q-PDT problem assumes more specific variants. In this section, we discuss such application and PDT settings/characteristics.

4.1 Application Settings

Here, we first focus on different services provided by the application (i.e., data propagation and data collection) and their corresponding measures for the quality of the sensed data. Thereafter, we study the application settings related to properties of the sensors and aggregators which includes the dynamism of sensors and aggregators, lifetime of the packets generated by sensors, and existence of sensors and aggregators in real and virtual worlds.

4.1.1 Applications Services

Two generic types of services provided by sensing applications are data collection and data propagation. With data collection, the goal is to transfer data generated at sensors to aggregators to reconstruct/monitor a phenomenon (such as temperature or air pollution) during the time interval T. The sensed data may be used to predict the evolution of the phenomenon in the future as well. The quality of the collected data packets P, $Q(P)$, is defined to be proportional to the amount of uncertainty reduction in predicting the phenomenon at unsensed locations (i.e., locations where no sensor is placed) during T. More formally, consider the set of environment points and the set of points where sensors are placed, denoted by V and F, respectively. The quality of data $Q(P)$ is defined as follows:

$$Q(P) = U(V - F) - U(V - F|F)$$

where $U(x)$ represents the uncertainty in predicting the phenomenon at x.

On the other hand, with data propagation the goal is to maximize the number of distinct aggregators which receive the data packets from the sensors. For example, assume a weather pollution warning application. This application consists of a set of towers which are equipped with sensors to measure the pollution. The towers show warnings on digital boards attached to them when the weather pollution is excessively high. The goal is to place a limited number of such towers throughout the city to maximize the number of individuals who receive the warnings (directly or indirectly via PDT). In this case, the sensors are attached to the towers, the packets are the warnings, and the aggregators are all the people living in the city. With such applications, the quality of the collected data is defined as the number of individuals who receive the warnings, i.e., the number of aggregators which receive a data packet. Accordingly, we can define the quality $Q(P)$ for data propagation applications as follows:

$$Q(P) = |A_f|$$

where A_f represents the set of aggregators which receive a packet in P.

4.1.2 Effect of Sensors and Aggregators Properties

Here, we study the application settings related to characteristics of the sensors and aggregators. The first important application setting is concerning dynamism of the sensors/aggregators. For example, consider a case where sensors are mounted on top of moving cars or buses. In this case, quality-aware PDT still involved placing sensors and aggregators to maximize the quality of transferred data. However, the placement becomes the problem of finding optimal trajectories (paths) for each sensor and aggregator. Similarly, sensors and aggregators can be attached to the individuals who contribute to data transfer as well.

Another application dependent characteristic that can affect optimization of the PDT process is the lifetime of the packets. The lifetime of a packet can be short (e.g., information related to noise or temperature) or long (e.g., an advertisement which is valid for months). Within the same application, longer lifetime packets need less transmission to provide the same service.

Finally, sensors and aggregators can exist both in virtual world and real world. For example, an advertisement may be generated in an LBSN or transmitted to the virtual entity of an individual through LBSN (examples of sensors and aggregators in real world were already provided in the previous sections). As opposed to physical world, geographical distance between individuals is not the dominating factor in deciding reachability of an individual from the other one.

4.2 PDT Characteristics

Two of the most important user characteristics are how users moving in the environment and how they exchange data between each other. We discuss these characteristics in this section. Assume a limited set of pathways R in E, the individuals can take either known paths or any random paths on R. For example, an individual who takes a bus to go to work, takes a known path on her way to work. On the other hand, a student who is moving on a university campus may take very different paths each day. The uncertainty in individual's location is much higher when paths are not known in advance. This uncertainty greatly affects how we can predict and model users' movement in developing optimal PDT solutions.

Another PDT characteristic is concerning how two individuals u and v (who are also friends in L) can reach each other for data exchange. Two individual can be connected either via physical connection in real world (e.g., wireless connection or Bluetooth), and/or via virtual connection in virtual world. LBSN is an example of virtual world with user IDs as virtual entities. Each virtual world entity has a corresponding individual in the real world. Two individuals are connected in virtual world if their virtual entities are connected in the virtual world. Accordingly, with PDT individuals communicate through the real world or the virtual world and hence, a data packet generated at a sensor, travels through a series of virtual and real world communications to reach to its destination.

5. RELATED WORK

The first body of relevant work is the literature on sensor deployment and sensing coverage in the field of sensor networks. In [4, 8], the coverage problem is formulated as a decision problem to determine whether every point in the service area of the sensor network is covered by at least k sensors. However, with sensor deployment the goal is to maximize the coverage by proper sensor placement. Most proposed approaches for sensor deployment assume simple sensing models with circular (omnidirectional or unidirectional) coverage for sensors [3, 12]. However, visual sensors cannot be modeled as such; in [10] we introduce an approach for efficient placement of the visual sensors. The other category of related work is on opportunistic networking in mobile ad hoc networks. Mobile ad hoc networks are typically composed of mobile nodes that communicate over wireless links without any central control [1]. A variety of opportunistic routing algorithms are proposed to transfer a data packet from a source to a destination node in the network [9, 7].

However, the solutions proposed for these two categories of related problems (i.e., sensor deployment/coverage and opportunistic networking) are not readily applicable to the Q-PDT problem. With Q-PDT, one needs to consider the objectives and constraints of both aforementioned problems concurrently. More importantly, the optimization criteria with those problems are often competing, which makes existing algorithms for sensor deployment/coverage and opportunistic networking poor solutions for the Q-PDT problem.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we introduced the problem of quality-aware PDT and discussed its different variations given various PDT features and application settings. We also studied the complexity of the problem and proved that it is an NP-hard problem. Currently, we are studying a special case of the Q-PDT problem in which sensing application monitors temperature in an environment and individuals riding cars are forming the communication medium to transfer data packets. According to the complexity of Q-PDT problem, we are developing efficient heuristics to solve this special case of Q-PDT. As part of our future work, we plan to extend this special case to monitor phenomena different than temperature. In the long term, we plan to study other cases of Q-PDT problem.

7. ACKNOWLEDGMENTS

The vision presented in this paper is partly inspired by the vision defined for the MASON project at Tsinghua University (http://sensor.ee.tsinghua.edu.cn/). Many thanks to Dr. Lin Zhang and his group for helpful discussions. This research has been funded in part by NSF grant CNS-0831505 (CyberTrust), the NSF Integrated Media Systems Center (IMSC), unrestricted cash-gift from Google, and in part from the METRANS Transportation Center, under grants from USDOT and Caltrans. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

8. REFERENCES

[1] *Handbook of wireless networks and mobile computing.* John Wiley & Sons, Inc., New York, NY, USA, 2002.

[2] U. Center for Embedded Networked Sensing (CENS). `http://urban.cens.ucla.edu/projects/`.

[3] S. S. Dhillon and K. Chakrabarty. Sensor placement for effective coverage and surveillance in distributed sensor networks. In *Proc. of IEEE WCNC*, 2003.

[4] C.-F. Huang and Y.-C. Tseng. The coverage problem in a wireless sensor network. In *WSNA*, 2003.

[5] F. B. Kashani, H. Shirani-Mehr, B. Pan, N. Bopp, L. Nocera, and C. Shahabi. Geosim: A geospatial data collection system for participatory urban texture documentation. *IEEE Data Eng. Bull.*, 2010.

[6] A. Krause, R. Rajagopal, A. Gupta, and C. Guestrin. Simultaneous placement and scheduling of sensors. In *IPSN '09*, Washington, DC, USA, 2009. IEEE Computer Society.

[7] A. Lindgren, A. Doria, and O. Schelén. Probabilistic routing in intermittently connected networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 2003.

[8] S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. B. Srivastava. Coverage problems in wireless ad-hoc sensor networks. In *IEEE INFOCOM*, 2001.

[9] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain, and R. Krishnan. Prioritized epidemic routing for opportunistic networks. In *MobiOpp '07*, New York, NY, USA, 2007.

[10] H. Shirani-Mehr, F. Banaei-Kashani, and C. Shahabi. Efficient viewpoint selection for urban texture documentation. In *GSN '09*. Springer-Verlag, 2009.

[11] H. Shirani-Mehr, F. B. Kashani, and C. Shahabi. Efficient viewpoint assignment for urban texture documentation. In *GIS*, 2009.

[12] C.-H. Wu, K.-C. Lee, and Y.-C. Chung. A delaunay triangulation based method for wireless sensor network deployment. *Comput. Commun.*, 2007.