A Case Study of Participatory Data Transfer for Urban Temperature Monitoring

Houtan Shirani-Mehr[1], Farnoush Banaei-Kashani[1], Cyrus Shahabi[1], and Lin Zhang[2]

[1] University of Southern California,
Los Angeles, CA
{hshirani,banaeika,shahabi}@usc.edu
[2] Tsinghua University, Beijing, China
linzhang@tsinghua.edu.cn

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 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 users participation for data transfer. With PDT, users use mobile devices to receive data from sensors, and forward the sensed data through the ad hoc network of the mobile devices until the data is received by the data aggregators (i.e., data sinks). Sensor deployment and ad hoc routing/networking are two related problems that are both extensively studied in the literature. However, to enable efficient deployment of PDT for sensing applications one needs to consider the requirements of the two aforementioned problems in conjunction. In this paper, we present a case study of PDT with which we explore the performance of PDT-based data transfer with a sample urban sensing application, namely, an urban temperature monitoring application. Our experimental case study is by simulation based on real datasets including GPS track data for more than 2000 vehicles in the city of Beijing. We discuss our observations based on this case study which can serve as directions to design application-specific optimal PDT mechanisms.

Keywords: Participatory Data Transfer, Sensor Placement, Aggregator Placement, Urban Temperature Monitoring

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 [3, 7]. 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 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 an urban temperature monitoring application a set of temperature 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 temperature 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 such environments, we envision *participatory data transfer* (PDT) [15] as an alternative communication medium with which data is transferred by objects (individuals, vehicles, etc.) that are moving through the environment and are equipped with mobile devices such as cell phones, laptops and PDAs. With PDT, participating objects use their mobile devices to receive data from sensors. Thereafter, the data is forwarded through the ad hoc network of objects' mobile devices until the data is received by the aggregators of the sensing system.

The design objectives for *Quality-aware* PDT (Q-PDT) are to maximize the coverage of the area monitored by the sensing application while minimizing the delay in reporting the sensed phenomenon over time. Therefore, one needs to consider solving two optimization problems conjointly to develop an efficient Q-PDT solution: 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 network of the users participating in PDT. Although sensor deployment and ad hoc routing/networking (or opportunistic networking) are both extensively studied in the literature, a Q-PDT solution based on isolated solutions for each of these problems may be poor and ineffective.

In this paper, we present a case study of Q-PDT with which we develop and try various solutions for the two problems of placement and routing in combination to make observations on their joint performance for future design. The experimental case study is by simulation based on real datasets including the GPS tracks of more than 2000 vehicles in the city of Beijing. The vehicles' GPS tracks cover an area of approximately $600km^2$.

The organization of the paper is as follows. In Section 2 we present a model for PDT based sensing applications. Section 3 formally defines the Q-PDT problem. We study the computational complexity of the problem in Section 4. In Section 5, we provide our case study. Section 6 reviews the related work and finally we conclude the paper and discuss our future directions in Section 7.

2 System Model

A PDT based sensing system comprises of a set of sensors and aggregators, and leverages PDT to transfer data from sensors to aggregators to be processed or further forwarded to a central processing unit via some existing communication infrastructure (wired or wireless). With such a system, participating objects collectively serve as a data transfer medium to transfer data from sensors to aggregators. In this section, we present our models for the sensing application as well as the PDT process in a PDT-based sensing system.

2.1 Application Model

The sensing application tracks a phenomenon such as temperature in an environment E based on the measurements collected from the sensors placed in E. Assume the set of points in E is denoted by V. We assume the application monitors the phenomenon in E during the time interval T. T is discretized into equal length intervals, i.e., T=$(I_1, I_2, \ldots, I_n)$, and the application reports the phenomenon variations for every point in V at the end of each I_i. For example, with $|I_i| = 1$ hour for a temperature monitoring application, the application reports hourly the temperature at each point in V. The phenomenon variations during each interval I_i is assumed to be negligible and therefore any reading for a point q during the interval I_i, is a sufficiently accurate representation for the phenomenon during the entire I_i.

We presume 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=$\{a_1, a_2, \ldots, a_{N_A}\}$. Accordingly, each sensor $s_i \in S$ generates a series of packets, $P_{s_i}(t)$, where each packet contains a sensor measurement at t. A packet p is considered received if at least one aggregator receives it. We denote the earliest time at which p reaches an aggregator by t_{r_p}. The packet also has a lifetime of T_p which means that the sensors readings carried by packets become stale for real-time monitoring purposes if received after T_p. If p is generated at time $t_p \in I_i$, T_p=$|t_{e_{I_i}} - t_p|$ where $I_i = [t_{b_{I_i}} \;\; t_{e_{I_i}}]$. We assume a sensor s_i generates a packet p whenever an object o can communicate with s_i (while respecting the communication constraints such as distance with s_i). However, if o carries a non-stale packet already generated at s_i, no new packet will be generated to avoid unnecessary data transfer and reduce the communication overhead. Imagine an urban temperature monitoring as the running example. Figure 1 shows three temperature sensors s_1, s_2 and s_3 and two aggregators a_1 and a_2 deployed by this application that are placed in the environment E. If this application reports the hourly temperature of the urban environment (starting at 12:00am), the packets generated at 12:30pm and 13pm have the life times of 30 minutes and one hour, respectively.

2.2 PDT Model

PDT is the process in which a set of objects $O = \{o_1, o_2 \ldots, o_{N_O}\}$ participate in data transfer. The objects may be constrained in their movement in the environ-

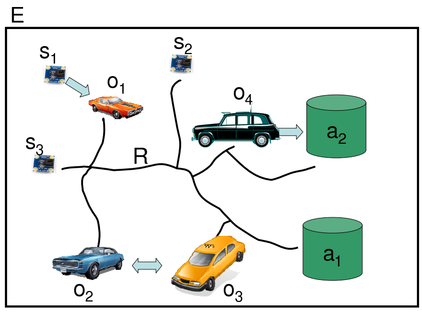


Fig. 1. Sensors, aggregators and objects contribute to PDT

ment. For example, in Figure 1 four vehicles o_1, o_2, o_3 and o_4 are participating in PDT. The vehicles are moving on a road network R where the road network segments are highlighted in the figure. The vehicles use their embedded mobile devices to exchange packets among each other (For example, o_2 and o_3 exchange packets collected on their way in Figure 1. Later, each of these vehicles may meet an aggregator and deliver the exchanged packets) and exchange packets with sensors (e.g., o_1 and s_1 in Figure 1) and aggregators (e.g., o_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. Finally, we assume all communicating devices which are used to exchange data are programmed to participate in PDT automatically; hence, participation does not require active intervention of the users.

3 Problem Definition

In this section, we define the problem of Quality-Aware PDT (Q-PDT). The objective with the Q-PDT is to consider movement patterns of the participating objects and accordingly place a set of sensors and aggregators in an environment such that we can maximize the quality of the packets received by the aggregators during the time interval T. Next, we first formally define the quality of a set of transferred packets, and afterward formalize the Q-PDT problem.

Data Quality: The application reports the phenomenon in the environment E at the end of each $I_i \in T = (I_1, I_2, \ldots, I_n)$. Therefore, we define the quality $Q(P)$ of transferred (and collected) packets P over the entire T as the summation of the quality of packets collected during each $I_i \in T$ (we exclude the packets from P which are not received by aggregators during their life-time). In other words,

$$Q(P) = \sum_{i=1}^{n} Q_i(P_i),$$

where P_i is the set of packets received during I_i and $Q_i(P_i)$ represents the quality of packets in P_i. We define $Q_i(P_i)$ as the amount of uncertainty reduction in predicting the phenomenon at unsensed locations (locations where no sensor is placed) during I_i, after receiving packets in P_i. We assume that the sensors report the phenomenon at the sensed locations (locations where sensors are placed) accurately and with no uncertainty. We use Entropy function [12] to measure the uncertainly in predicting the phenomenon at the environment points. Consider a random vector with values in $\mathbb{R}^m$ and probability density function $f(x)$. The entropy of X is defined by

$$H(X) = -\int_{\mathbb{R}^m} \log f(x) dx$$

Assume the set of environment points is represented by V and the set of points where sensors in S_i are located is denoted by V_i. Therefore, to calculate $Q_i(P_i)$ we only need to measure the uncertainty in predicting the phenomenon at unsensed locations. Accordingly, we define

$$Q_i(P_i) = H(V - V_{S_i}) - H(V - V_{S_i}|V_{S_i}), \tag{1}$$

where $H(V - V_{S_i})$ represents the entropy in predicting the phenomenon at points in $V - V_{S_i}$ before placing the sensors. Correspondingly, $H(V - V_{S_i}|V_{S_i})$ is entropy at the same locations after placing sensors at points in V_{S_i}.

However, we need to consider a model for the sensed phenomenon in order to calculate the entropy of predicted phenomenon. A conventional approach for modeling continuous physical phenomenon such as temperature is to assume that sensor readings have (multivariate) Gaussian joint distribution [9, 4]. In particular for the sensors in S_i, the set of $n = |S_i|$ corresponding random variables X have Gaussian joint distribution if

$$P(X = x) = \frac{1}{(2\pi)^{n/2}|\Sigma|} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)},$$

where μ, Σ and $|\Sigma|$ are the mean vector, covariance matrix and determinant of the covariance matrix, respectively. If we index each random variable $X_i \in X$ by i, we end up with a set of indices $\mathcal{V}$. Any subset of random variables with index set $\mathcal{A} \in \mathcal{V}$ has a Gaussian joint distribution as well. In the temperature monitoring application, the goal is to monitor temperature for every point in V and not just points in V_{S_i}. Fortunately, *Gaussian Process* (GP), which is a generalization of multivariate Gaussian, considers an infinite number of random variables and hence can be leveraged to model temperature for all the points in V. A GP is associated with a mean function $\mathcal{M}(.,.)$ and a symmetric covariance function $\mathcal{K}(.,.)$. For each random variable with index $u \in \mathcal{V}$, its mean μ_u is given by $\mathcal{M}(u)$. Accordingly, for each pair of indices $u, v \in \mathcal{V}$, their covariance σ_{uv} is $\mathcal{K}(u, v)$. With GP, given a set of sensor readings $x_{\mathcal{A}}$ corresponding to set $\mathcal{A} \in \mathcal{V}$, we can predict temperature at any point $y \in \mathcal{V}$ conditioned on $x_{\mathcal{A}}$, i.e., $P(X_y|x_{\mathcal{A}})$. The distribution of X_y is Gaussian whose conditional variance $\sigma^2_{y|\mathcal{A}}$ is

$$\sigma^2_{y|\mathcal{A}} = \mathcal{K}(y, y) - \Sigma_{y\mathcal{A}} \Sigma^{-1}_{\mathcal{A}\mathcal{A}} \Sigma_{\mathcal{A}y},$$

where $\Sigma_{y\mathcal{A}}$ is a covariance vector with one entry for each $u \in \mathcal{A}$ with value $\mathcal{K}(y, u)$. The entropy of a Gaussian random variable Z conditioned on a set of variables X is a function of its variance,

$$H(Z|X) = \frac{1}{2} \log(2\pi e \sigma^2_{Z|X}).$$

Consequently, this enables us to calculate $Q_i(P_i)$ based on Equation (1).

Definition 1 (Problem of Quality-Aware PDT (Q-PDT)). *We define the problem of Q-PDT as follow: Given a set of sensors S and a set of aggregators A, place all the sensors and aggregators in an environment E, such that $Q(P)$ is maximized.*

4 Complexity

In this section, we prove that Q-PDT is an NP-hard problem by reduction from the *optimal coverage* problem [8] (which is also an NP-hard problem). The optimal coverage problem can be formalized as follows.

Definition 2 (Optimal Coverage). *Assume an environment E is represented by a discrete set of points V. Given k sensors to monitor a phenomenon in E, optimal coverage places sensors in a set of points $F \subseteq V, |F| = k$, such that the prediction accuracy of the phenomenon throughout E is optimized:*

$$\arg\max \; H(V - F) - H(V - F|F) = F, \tag{2}$$

where 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 optimal coverage problem. Towards that end, we prove that given an instance of optimal coverage, denoted by I_C, 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_C in polynomial time. Assume a given instance I_C whose goal is to place k sensors to monitor a phenomenon during the time interval T'. We propose the following mapping from I_C to I_Q to reduce I_C to I_Q. The environments for both instances are the same, i.e., the set of environment points in I_C 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 pairs of sensors and aggregators in I_Q where the ith pair consists of a sensor s_i and an aggregator a_i such that a_i is placed in the environment with a negligible distance of $\epsilon \approx 0$ with s_i. Due to the small distance between sensors and aggregators in each pair, there is no need for objects to transfer packets from sensors to aggregators. 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_C.

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.

5 Case Study

In this section, we focus on a special use case of Q-PDT in which PDT is deployed in a temperature monitoring application for an urban environment. With this case study, participating objects in PDT are a set of moving vehicles which are constrained to move on a road network R embedded in the environment E. Below, we first explain our experimental setup for this case study and thereafter provide our results and observations.

5.1 Experimental Setup

In this section, we first describe our PDT specifications and our assumed model for distribution of temperature in the environment. Thereafter, as Q-PDT is an NP-hard problem, we propose various heuristics for sensors and aggregators placement. Finally, we describe data routing techniques we used for our case study.

5.1.1 PDT Specifications We conducted our experiments using real data capturing the movements of vehicles in the city of Beijing. This dataset covers the GPS tracks of more than 2000 distinct vehicles collected during a day. The vehicles' GPS tracks cover an area of approximately $600km^2$. This area comprises the environment E. The vehicles locations are recorded every minute (a total of more than $1,400,000$ records). We assume the vehicles use Dedicated Short-Range Communications (DSCR) channels to communicate with each other, as well as with the sensors and aggregators. The DSCR [18] standard is developed to support low-latency wireless data communications between vehicles and from vehicles to roadside units. The effective range for communication over DSRC is up to 1km. Therefore, two vehicles (or a vehicle and an aggregator/sensor) are reachable if they are at most 1km apart.

5.1.2 Temperature Modeling Since we did not have access to the temperature data for the city of Beijing (the environment E), instead, we acquired the temperature readings recorded every second at EPFL campus [1] from 210 sensors to simulate temperature variations over E. The extent of the area for E is different from that of EPFL campus and hence we had to spatially rescale the temperature readings. To this end, first since the sensors at EPFL campus are not uniformly located, we interpolated the temperature measurements at EPFL campus to find the temperature at uniformly distributed locations (we selected 600 locations). Afterward, we divided the space of E into 600 grid blocks with the block size of $1km \times 1km$ and assigned each EPFL temperature measurement to the corresponding grid block while preserving the locality of the readings. We denote the set of these grid blocks by G. The temperature variations within each grid block $g \in G$ is assumed to be negligible and therefore any point in g is a sufficiently accurate representation of the temperature of all the point in g.

5.1.3 Sensor Placement We assume there is no direct communication between sensors and aggregators, and hence, the sensors and aggregators should be placed on R to leverage PDT for data transfer. Only a subset of grid blocks in G overlap with R, and therefore, as a filtering step we select a set of grid blocks $G' \subseteq G$ which overlap with G. As the temperature variations within a grid block is assumed to be negligible, a sensor placed anywhere within a grid block $g \in G'$ can capture the temperature for all the points in g. We adapt the greedy approach proposed in [9] for choosing the sensors locations. Given the set of initial grid block G', this algorithm chooses the next location for sensor placement which provides the maximum reduction in the entropy of predicting temperature at unsensed locations (locations where still no sensor is placed). More formally, assume the currently selected sensors locations is represented by S_C (initially $S_C = \phi$). At each iteration the algorithm selects a grid block $g \in G'$ and add its center g_c to S_C which maximizes

$$\left[H(S_C \cup g_c) - H(S_C \cup |\overline{S_C})\right] - \left[H(S_C) - H(S_C|\overline{S_C} \cup g_c)\right].$$

To simplify notation we use $\overline{S_C}$ to mean $V - (S_C \cup g_c)$. After $|S|$ iterations, the set S_C contains the sensor locations.

5.1.4 Aggregator Placement Similar to sensors placement, we select a subset of grid blocks in G' to place the aggregators. We consider two different approaches to select aggregators locations. Next, we explain each approach in details.

RAND Approach: With the first approach, we randomly select $|A|$ grid blocks from G' where the set of selected grid blocks centers comprise the aggregators locations. We call this approach $RAND$.

FREQ Approach: With the second approach, we consider the frequency with which the vehicles visit the grid blocks in G'. Intuitively, the points which are visited more often over time are better candidates to place the aggregators. We denote this approach by $FREQ$. With FREQ, we select $|A|$ grid cells from G' which are visited the most. To find the frequency of visit for each grid cell, we used the vehicles positions recorded during the first 12 hours of the day. To prevent the aggregators to be placed very close to each other we consider a constraint such that the aggregators locations should be at least $d$km apart. The selected grid cells centers comprises the aggregators locations. We experimentally find the reasonable value for d in Section 5.2.1.

5.1.5 Routing Protocols Variety of routing protocols are proposed to transmit a packet from a source to destination node in ad hoc/oppurtunitic networks. In this work, we study the performance of two of those protocols [16] in the context of PDT. We next explain each protocol in detail.
Epidemic Routing: The first protocol is *Epidemic* routing which floods the packets into the network. With epidemic routing, whenever two vehicles o_1 and o_2 meet all the packets carried by o_1 (o_2) are transferred to o_2 (o_1). Clearly, with this approach packets are propagating fast between the vehicles but at the same time this approach consumes a huge amount of communication resources.
Random Routing: The second approach is *Random* routing. With random routing a transfer probability p_r is selected and a packet is transmitted with the probability of p_r at each contact. By changing the value of p_r one can control the amount of packet transfer in the network.

We selected these two algorithms as they are easy to implement and at the same time they enable us to change the amount of data transfer in the network and observe its consequences on the effectiveness of Q-PDT.

5.2 Results

In this section, we present the results of our experimental case study.

5.2.1 Aggregator Placement In this experiment, we evaluate the efficiency of different aggregator placement approaches. We first study the efficiency of FREQ algorithm and afterward compare the effectiveness of FREQ and RAND approaches.

5.2.1.1 Efficiency of FREQ: With the FREQ approach, the selected aggregator locations should be at least $d$km apart. We varied d from 1km to 7km and measured the quality of data during a time interval $|I_i| = 1$ hour in order to study the impact of the selected d. Our results are shown in Figure 2 for $|S| = 60$ and $|A| = 15$. As the figure illustrates, there is a trade-off in selecting d. Increasing d results in selecting aggregator locations which are more widely distributed in the space, and hence, the aggregators are reachable from vehicles moving in various directions. On the other hand, by increasing d we are selecting points which are less visited by vehicles over time. Given our experimental settings, the best value for d is 3km for both routing protocols (see the figure).

5.2.1.2 FREQ vs. RAND: Using the same settings as those of the experiments in the previous section, we compare the efficiency of the FREQ (d=3) and RAND placement approaches. Figure 3 shows the quality of data for different aggregators placement approaches and different routing protocols. The results show that the FREQ placement approach outperforms RAND. As expected, with FREQ,

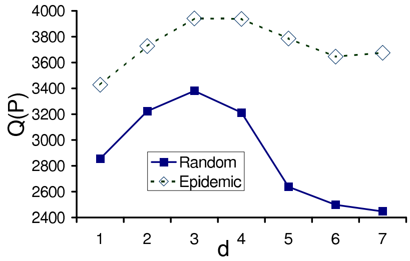
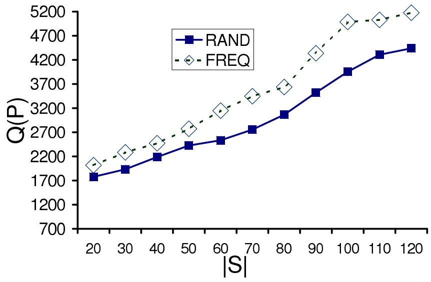
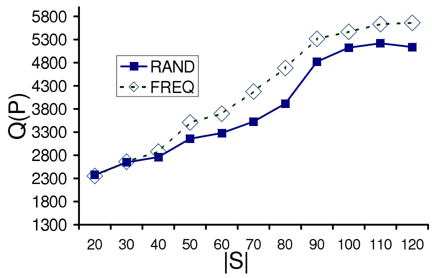


Fig. 2. Data quality vs. d.

the quality of data is higher than RAND because aggregators are placed in the locations which are visited more often as compared to the of rest of the locations. Moreover, with RAND, in some cases aggregators are placed in locations which are not visited at all which makes the performance of this approach very poor. In the rest of the experiments in Section 5.2, we assume FREQ is used for aggregator placement.



(a) Random routing



(b) Epidemic routing

Fig. 3. Data quality vs. $|S|$ for different aggregator placement approaches and routing protocols.

5.2.2 The Count of Generated Packets With this experiment, we study how the number of generated packets depends on the routing protocol and the number of sensors. The result is shown in Figure 4 for $p_r = 0.1$ and during a time interval $|I_i|$. As expected, as the number of sensors ($|S|$) increases, the number of generated packets grows. Moreover, the increase rate is much higher in two cases ($|S| = 20$ and $|S| = 100$). The reason is that for these two cases a set of sensors are placed at new locations in G' which are much more visited by cars than other locations and hence more packets are generated. On average over all the cases, the number of generated packets in epidemic routing is 90% more than random routing.

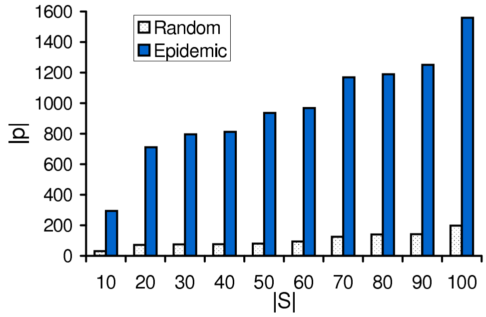


Fig. 4. Generated packets vs. number of placed sensors ($|S|$).

5.2.3 Data Quality Over Time With this experiment, we evaluate the quality of the transferred data over time. To this end, we set $T = (I_1, I_2, \ldots . I_8)$ where $|I_i| = 1$ hour and T spans the time interval [12 20], $p_r = 0.1$ for random routing, $|S| = 60$ and $|A| = 15$. The result is shown in Figure 5. Over all the cases, epidemic routing is collecting data which has 49% higher quality than random routing. The difference between data quality for these two routing protocols varies between 41% and 64% where the highest difference occurs when the dynamism of the vehicles is greater, i.e., I_1=[12 13] and I_5=[16 17]. The reason is that when the dynamism of the vehicles are higher, more vehicles become reachable to each other, to sensors and to aggregators. With epidemic routing, the packets can directly hop through the reachable devices to travel from a sensor to an aggregator. However, with random routing the probability that a packet p travels from one vehicle to another decreases as the number of vehicles which are traversed by p.

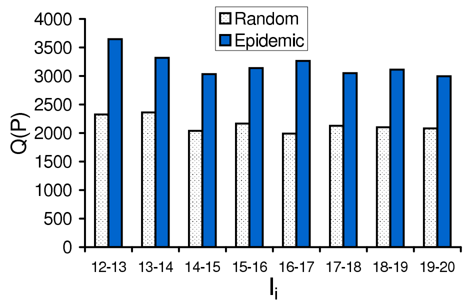


Fig. 5. Data quality vs. I_i.

5.2.4 Effect of p_r In this section, we study the effect of changing p_r for random routing. We set $|S| = 50$, $|A| = 13$, $T = I_1$ and varied p_r from 0.1 to

1 (with $p_r = 0$, no data transfer occurs and hence $Q(P) = 0$). Figure 6 shows the effect of changing p_r on $Q(P)$ for a time interval $|I_i| = 1$ hour. As expected, by increasing p_r, $Q(P)$ grows because there are more packet exchanges between vehicle and between vehicles and sensors/aggregators. The rate of increase for $Q(P)$ is generally decreasing when p_r grows and it is largest at $p_r = 0.2$.

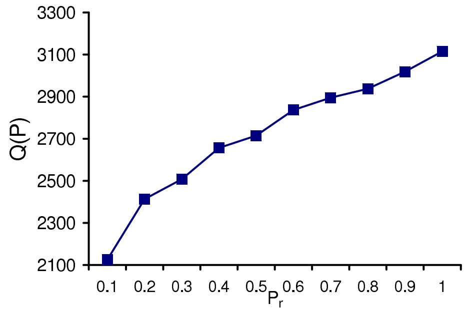


Fig. 6. Data quality vs. p_r.

5.3 Observations

We can summarize our observations based on the experimental case study as follows.

5.3.1 Sensor/Aggregtor Placement The approach to place the aggregators which considers the dynamism of the vehicle (FREQ), outperforms the random aggregators placement approach in terms of the quality of transferred data (by 15% and 25% for random and epidemic routing protocols, respectively). Moreover, the quality of the transferred data depends on the dynamism of the vehicles as well. The quality of the transferred data is higher when the dynamism of the vehicles is greater. With our experiments, we observed that 24% difference exists between quality of data for sub intervals in $T = [12 \quad 20]$. This suggests that an optimal solution for Q-PDT should consider the movement of the participating objects as well as the changes in their movement behavior over time.

5.3.2 Routing Protocols: The epidemic routing approach outperforms the random routing because of flooding the packets into the network of vehicles. However, the communication cost with epidemic routing is much higher. For example, with $p_r = 0.1$, the number of generated packets is 90% lower than that of epidemic routing where the data quality is lower by 49%. This shows that there is a trade-off in selecting p_r in terms of available communication resources and the desired sensing application quality. With our experiments, the difference between random and epidemic routing varies over time (see Section 5.2.3). Therefore, the value for p_r should change in optimal Q-PDT based on the movement behavior of the participating objects over time.

6 Related Work

PDT in the context of location-based social networks was proposed as a vision [15] to transfer packets for sensing systems which are deployed in the environments where no pre-existing communication infrastructure are available. In this work, we provided an experimental case study on using PDT in a more general setting in which a network of cars are participating in data transfer. With this case study, we studied various approaches for packets routing and sensors/aggregators placement by conducting experiments on real PDT data.

A body of relevant work is the literature on sensor deployment and sensing coverage in the field of sensor networks. In [6, 11], 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 [5, 17], whereas, these models are not optimal to model verity of phenomena. In [14] we introduced an approach for efficient placement of the visual sensors. Recently, in [8, 9] Gaussian processes are used to model the monitored phenomena and consequently sensor placement. Moving sensors are studied recently to improve the coverage compared to static sensors [19].

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 [2]. A variety of opportunistic routing algorithms are proposed to transfer a data packet from a source to a destination node in the network [13, 10].

The difference between this work and the aforementioned two categories of related problems (i.e., sensor deployment/coverage and ad hoc opportunistic networking) is that we consider the objectives and constraints of both problems (which are partially competing constraints) concurrently.

7 Conclusions and Future Directions

In this paper, we introduced the problem of quality-aware participatory data transfer (Q-PDT). We presented an experimental case study on the problem based on real datasets including the GPS track data for more than 2000 vehicles in the city of Beijing. As part of our future work, we plan to study the Q-PDT problem on more datasets and also design optimal Q-PDT solutions for various applications based on the observations made in this work.

Acknowledgments. This research has been funded in part by NSF grant CNS-0831505 (CyberTrust), the NSF Integrated Media Systems Center (IMSC), unrestricted cash and equipment 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.

References

1. `http://sensorscope.epfl.ch/`.
2. *Handbook of wireless networks and mobile computing.* John Wiley & Sons, Inc., New York, NY, USA, 2002.
3. U. Center for Embedded Networked Sensing (CENS). `http://urban.cens.ucla.edu/projects/`.
4. A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong. Model-driven data acquisition in sensor networks. In *VLDB '04: Proceedings of the Thirtieth international conference on Very large data bases*, pages 588–599. VLDB Endowment, 2004.
5. S. S. Dhillon and K. Chakrabarty. Sensor placement for effective coverage and surveillance in distributed sensor networks. In *Proc. of IEEE Wireless Communications and Networking Conference*, 2003.
6. C.-F. Huang and Y.-C. Tseng. The coverage problem in a wireless sensor network. In *WSNA '03*, New York, NY, USA, 2003. ACM.
7. 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.
8. 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.
9. A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *J. Mach. Learn. Res.*, 9:235–284, 2008.
10. A. Lindgren, A. Doria, and O. Schelén. Probabilistic routing in intermittently connected networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 2003.
11. S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. B. Srivastava. Coverage problems in wireless ad-hoc sensor networks. In *IEEE INFOCOM*, 2001.
12. A. Mokkadem. Estimation of the entropy and information of absolutely continuous random variables. *Information Theory, IEEE Transactions on*, 35(1):193 –196, jan. 1989.
13. 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.
14. H. Shirani-Mehr, F. Banaei-Kashani, and C. Shahabi. Efficient viewpoint selection for urban texture documentation. In *GSN '09*. Springer-Verlag, 2009.
15. H. Shirani-Mehr, F. B. Kashani, and C. Shahabi. Using location based social networks for quality-aware participatory data transfer. In *GIS-LBSN*, 2010.

16. L. Song and D. F. Kotz. Evaluating opportunistic routing protocols with large realistic contact traces. In *CHANTS '07: Proceedings of the second ACM workshop on Challenged networks*, pages 35–42, New York, NY, USA, 2007. ACM.
17. C.-H. Wu, K.-C. Lee, and Y.-C. Chung. A delaunay triangulation based method for wireless sensor network deployment. *Comput. Commun.*, 2007.
18. J. Yin, T. ElBatt, G. Yeung, B. Ryu, S. Habermas, H. Krishnan, and T. Talty. Performance evaluation of safety applications over dsrc vehicular ad hoc networks. In *VANET '04: Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*, pages 1–9, New York, NY, USA, 2004. ACM.
19. X. Yu, H. Zhao, L. Zhang, S. Wu, B. Krishnamachari, and V. O. K. Li. Cooperative sensing and compression in vehicular sensor networks for urban monitoring. In *IEEE International Conference on Communications*, 2010.