Efficient 3G Budget Utilization in Mobile Participatory Sensing Applications

Hengchang Liu[1], Shaohan Hu[1], Wei Zheng[2], Zhiheng Xie[3], Shiguang Wang[1], Pan Hui[4], and Tarek Abdelzaher[1]

[1]University of Illinois at Urbana-Champaign
[2]University of Wisconsin-Madison
[3]University of Virginia
[4]The Hong Kong University of Science and Technology, and Telekom Innovation Laboratories

***Abstract*— This paper explores efficient 3G budget utilization in mobile participatory sensing applications. [1] Distinct from previous research work that either rely on limited WiFi access points or assume the availability of unlimited 3G communication capability, we offer a more practical participatory sensing system that leverages potential 3G budgets that participants contribute at will, and uses it efficiently customized for the needs of multiple participatory sensing applications with heterogeneous sensitivity to environmental changes. We address the challenge that the information of data generation and WiFi encounters is not *a priori* knowledge, and propose an online decision making algorithm that takes advantage of participants' historical data. We also develop a heuristic algorithm to consume less energy and reduce the storage overhead while maintaining efficient 3G budget utilization. Experimental results from a 30-participant deployment demonstrate that, even when the budget is as small as 2.5% of a popular data plan, these two algorithms achieve higher utility of uploaded data compared to the baseline solution, especially, they increase the utility of received data by 151.4% and 137.8% for those sensitive applications.**

I. Introduction

In this paper, we develop a novel smart phone based vehicular participatory sensing system that achieves efficient utilization of limited 3G budgets to improve system performance. This work is motivated by the emergence of vehicular participatory sensing applications [1], [2], [3], where data are collected from vehicles, stored locally, then offloaded to backend servers via WiFi or 3G. We assume that users will typically not allow participatory sensing applications to use 3G communication without limitation, since unlimited data plans are no longer prevalent [10], [11]. The WiFi-based store-and-forward approach, on the other hand, may result in large latency motivating the work described in this paper.

Vehicles are becoming popular in sensor data collection. First, their natural mobility increases coverage for many participatory sensing applications [4]. Second, our daily commute itself has become a target of many research efforts, such as those that aim to save fuel consumption [3], find available parking positions [7], avoid traffic jams or routes in bad condition [1], [2], [8], or share general road-side events [6].

[1]We use *participatory sensing* in a broader sense to refer to any applications where data are collected with approval of participating volunteers, irrespectively of whether or not they are actively involved in making the measurements. We refer *mobile participatory sensing* to vehicular participatory sensing using smart phones.

Exploiting drivers' own phones avoids additional investment costs to participants. Compared to placing conventional PC-like devices in cars/buses [1], [5], [6], smart phones are more pervasive and easy to use, while meeting application requirements of sensing, computation, and storage.

The philosophy underlying our work is that we believe many participants are indeed capable of contributing a budget of 3G data. They either still use an unlimited data plan, or have a limited data plan but only use a small portion of it every month. The incentive for these participants to contribute a 3G budget in support of participatory sensing applications is that they want to have their own sensory data delivered and analyzed more reliably and quickly without extra cost, leading to improved feedback services for themselves.

Advances of technologies have made smart phones nowadays powerful enough to run multiple participatory sensing applications simultaneously. These applications typically have heterogeneous sensitivity/tolerance to environmental changes, for example, applications for traffic monitoring or parking lot availability are no doubt more sensitive than those for finding all Starbucks in a city or raising greener driving habits for aggressive drivers. This difference of properties among applications, typically represented by *utility* of received data, brings further opportunity to optimize the quality of information (QoI) during offloading process under the limited 3G budget constraint.

The main contribution in this paper lies in that we develop a novel communication framework in mobile participatory sensing applications, in which a decision making algorithm is designed to assign the limited 3G budget to the sensory data of multiple participatory sensing applications for better overall utility. The biggest challenge behind this problem is that the information of data generation and encounter of WiFi access points is not a *priori* knowledge, thus conventional deterministic resource allocation methods are not applicable here. Instead, our solution predicts these information by taking advantage of participants' historical data, and provides an online decision making algorithm to decide which sensory data should be offloaded via 3G communication while others wait for WiFi access points. In addition, we propose a heuristic algorithm that adopts a simple and effective fashion and maintains efficient 3G budget utilization.

978-1-4673-5946-7/13/$31.00

Our solution is evaluated by experimental results from a campus-wide deployment with 30 participants, each driving at least 100 miles. Simulation results of replaying the generated sensory data and WiFi encounters in the deployment demonstrate that our proposed solutions successfully improve the overall utility of received data, and can be tailored for heterogeneous needs of multiple participatory sensing applications.

The remainder of this paper is organized as follows. We compare our work with state of the art in Section II and present the system design in Section III. The evaluation for our proposed solution is discussed in Section IV. Finally, we conclude the paper in Section V.

II. State of the Art

Most prior mobile participatory sensing applications have focused on leveraging smart phones placed in vehicles. For example, the Nericell project [2] presents a system that performs rich sensing using smartphones that users carry with them in normal course, to monitor road and traffic conditions. The GreenGPS system [3] provides a service that computes fuel-efficient routes for vehicles between arbitrary end-points, by exploiting vehicular sensor measurements available through the On Board Diagnostic (OBD-II) interface of the car and GPS sensors on smart phones. SignalGuru [1] is a software service that relies solely on a collection of mobile phones to detect and predict the traffic signal schedule, producing a Green Light Optimal Speed Advisory (GLOSA). These systems rely on WiFi access points, and assume transmitting data through 3G networks is not desirable. However, open public WiFi is becoming less prevalent as more access points are becoming private or secure, resulting in a big delay time for generated sensory data to be delivered. Our paper aims to overcome this drawback by allowing participants who have many remaining 3G data every month to contribute a reasonable 3G budget without extra cost to help improve the system performance.

Consumption of mobile data by the pervasive usage of smart phones is forcing carriers to find ways to offload the network. Since the modern smart phones have been introduced worldwide, more and more users have become eager to engage with mobile applications and connected services. This eagerness has boosted up sales in the market more than 64% up annually worldwide in Q2 2010 [9]. Simultaneously, smartphone owners are using an increasing number of applications requiring the transfer of large amounts of data to/from mobile devices. As a consequence, the traffic generated by such devices has caused many problems to 3G network providers. AT&Ts subscribers in USA were getting extremely slow or no service at all because of network straining to meet iPhone users demand [10]. The company switched from unlimited traffic plans to tiered pricing for 3G data users in summer 2010. Similarly, Dutch T-Mobiles infrastructure has not been able to cope with intense 3G traffic, forcing the company to issue refunds for affected users [11]. Meanwhile, carriers are willing to use more pervasive technologies, such as Wi-Fi access points and hot spots. The proliferation of modern Wi-Fi enabled smartphones, together with the network providers tendency towards already existing technologies has turn Wi-Fi offloading into a reality. However, it is reported that the WiFi coverage is quite limited, usually under 20%, even in big cities [12].

Several existing work have investigated making use of different types of communications for data dissemination/collection purposes. Wiffler [12] is a system to augment access to 3G network through WiFi offloading, by leveraging delay tolerance and fast switching of devices. However, it focuses only on Internet access from moving vehicles. Han et al. proposed MoSoNet [14], the first work to exploit opportunistic communication to alleviate 3G traffic, achieved by using a target set and 3G recovery. However, this approach only works for data dissemination and is not applicable for data collection process as in our mobile participatory sensing applications. The VIP-delegation work [15] proposed a data dissemination/collection model based on social groups. Lee et al. measured the performance of 3G mobile data offloading through WiFi networks, and emphasized the incentive of delayed offloading to save traffic and energy [13]. The Multi-Nets [16] is a system capable of switching between wireless network interfaces (e.g., 3G and WiFi) on mobile devices in real-time, to achieve higher throughput and save energy. Besides the differences in both the nature of the problem and the application scenario compared to our work, these related work consider all data are of the same type and importance, while we take one step forward and address the scenarios of multiple applications running simultaneously, which is the trend as smart phones are more powerful nowadays. Moreover, all of them considering 3G communications assume there is no limitation for 3G usage, which is not reasonable in the near future; instead, we take advantage of those who are capable of contributing a 3G budget to the participatory sensing applications, and address the problem of efficient utilization of these precious 3G data.

III. System Design

In this section, we present the system design for efficient 3G budget utilization among multiple participatory sensing applications. We first describe the system model, then explain our proposed algorithms in detail.

A. System Model

Our system is designed to operate in a vehicular participatory sensing network of n mobile nodes (vehicles) that can generate sensor data via smart phones of participants. Each sensor node run M different types of participatory sensing applications simultaneously, and generate N types of data packets. The sampling rates of these packets are denoted by $\lambda_1, \lambda_2, \cdots, \lambda_N$, respectively. Some of these packets are more sensitive/tolerant to environmental changes than others, and the utility functions of these packets are $U_1, U_2, \cdots, U_N$. We assume that, for every data packet, its utility is a monotonically

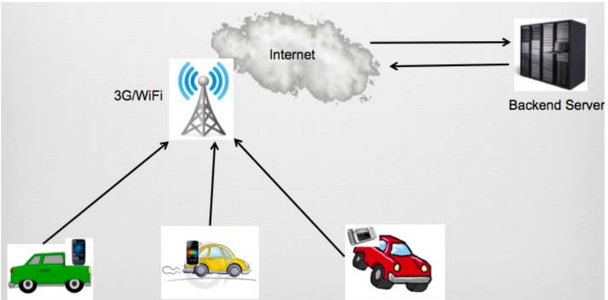


Fig. 1. Mobile participatory sensing system model.

decreasing function of time t. All nodes have buffers of fixed size C_0 that can be used to store packets.

The participatory sensing area is partially covered by WiFi access points. When a sensor node moves into the range of an access point, data packets in its buffer are offloaded to a backend server through WiFi communications. Without loss of generality, we assume that the WiFi capacity is large enough to offload all data packets in the buffer. This assumption is reasonable because participants usually can only access those WiFi networks at their homes, offices, and some public places like shopping malls, and spend quite a long time in these places. In addition, the size of participatory sensing data typically are small as long as no large amount of video clips are generated.

Each sensor node is associated with an amount of 3G budget. This budget is determined by the participant itself, based on his/her 3G data plan, living style, and willingness to contribute to the participatory sensing applications. For simplicity, the participatory sensing process is divided into multiple cycles, and the budget is equally assigned into each cycle. For instance, 600 bytes of budget per month can be transferred to 20 bytes per day. This assignment is additive, which means the remaining budget from previous cycles is accumulated into the new one.

Under this system model (as shown in Figure 1), we are interested in solving the following problem: At any time point in a cycle, whether packets in the buffer should be transmitted through 3G communications if there are still 3G budget remaining? If so, how much and which of them should be sent? This problem is challenging mainly because the information of when new sensor data are generated and the participants can encounter an access point is NOT *a priori* knowledge, and utility functions are different among multiple participatory sensing applications. In the rest of this section, we first describe the 3G-budget online algorithm to predict these information based on the participants' historical data, then present a heuristic algorithm running in a simple and effective fashion.

B. Online 3G-budget Algorithm

As detailed above, in practice the global information about data generation and WiFi access point encounters is unknown in advance. In this section, we develop an online algorithm that does not assume the availability of these information and estimate them based on historical data of participants.

input : A smart phone with multiple applications running.
output: Communication decisions.

initialization;
while *system is running* **do**
  if *a data plan period starts* **then**
    budget B is reset to a pre-specified number;
  end
  if *new data x is generated* **then**
    update historical data generation information (m and $(Q_1, ..., Q_m)$);
    add newly generated data into queue;
  end
  if *wireless is connected* **then**
    update the mean time to wireless connection; upload all data in queue via wireless;
  else
    sort the utility loss of data in current queue $u_i - \hat{v}_i$ and future utility loss $\hat{U}_i - \hat{V}_i$ from historical data together in a decreasing order;
    determine which data should be uploaded via 3G, given the current budget and the order derived above;
    update the queue of generated data;
    update the budget B;
  end
end

Algorithm 1: Online 3G-budget algorithm.

Let each data plan period have K time cycles and each time point t_0 in a data plan period can be expressed as time t of cycle c_t. At time t of c_t within a data plan period, we assume there are n data clusters of size $q_1, ..., q_n$ in queue. Each of the data clusters is consisted of data packets generated at the same time and from the same type of application. These n data clusters have utilities $u_1, ..., u_n$ at time t_0. The objective of this online algorithm is to decide at this time point, which of the n data clusters should be uploaded and how much of them should be uploaded using 3G communications. Assume that when the wireless is connected for the first time after time t_0, the utilities of them become $v_1, ..., v_n$. After time t_0 and within the current data plan period, there are m data clusters generated with size $Q_1, ..., Q_m$ and utilities $U_1, ..., U_m$. These data clusters have utilities $V_1, ..., V_m$ when the wireless is first connected after the data are generated. All of m, $v_1, ..., v_n$, $Q_1, ..., Q_m$, $U_1, ..., U_m$ and $V_1, ..., V_m$ are unknown.

Assume that x_i out of q_i is uploaded at time t using 3G network and the rest of i^{th} data cluster will be uploaded via wireless when an access point is encountered. The reason that besides x_i, no data is uploaded using 3G network is, given the fixed budget, the uploaded data will cost the same amount of budget and an earlier upload would have larger utility.

Therefore, for the data in queue at time t_0, the total uploaded utility is,

$$\begin{aligned} F_t(x_1,...,x_n) &= \sum_{i=1}^{n} x_i u_i + \sum_{i=1}^{n} (q_i - x_i) v_i \\ &= \sum_{i=1}^{n} v_i q_i + \sum_{i=1}^{n} (u_i - v_i) x_i \end{aligned}$$

Then after time t_0, the total budget left is $B - \sum_{i=1}^{n} x_i$. Therefore, the maximum uploaded utility for the data packets generated after time t_0 is,

$$\begin{aligned} & G_t(x_1,...,x_n) = \\ & \max_{y_1,...,y_m} \quad \sum_{i=1}^{m} V_i Q_i + \sum_{i=1}^{m} (U_i - V_i) y_i \\ & s.t. \quad 0 \le y_i \le Q_i, i = 1,...,m \\ & \qquad \sum_{i=1}^{m} y_i \le B - \sum_{i=1}^{n} x_i \end{aligned}$$

The solution for $x_1,...,x_n$ is to maximize

$$E[F_t(x_1,...,x_n) + G_t(x_1,...,x_n)]$$

where the expectation is over $v_1,...,v_n$, m, $Q_1,...,Q_m$, $U_1,...,U_m$ and $V_1,...,V_m$.

In practice, the optimization is computationally expensive, so instead of maximizing the expectation of the total uploaded utility, we estimate the unknown m, $v_1,...,v_n$, $Q_1,...,Q_m$ and $V_1,...,V_m$ using the historical data and then solve for $(x_1,...,x_n)$. Here the estimations from the historical data are based on individual cycles. To estimate m and Q_i's, the mean cluster size generated from each type of application at every time point within a cycle using historical data is computed. Then the estimator of m, $\hat{m}$ is simply the number of data clusters with non-zero mean size. The estimators for Q_i's, $\hat{Q}_i$'s are the mean cluster sizes multiplied by $(K-c_t)$ for the time points prior to or at time t and multiplied by $(K-c_t+1)$ for the time points after t. To estimate $v_1,...,v_n$ and $V_1,...,V_{\hat{m}}$, the mean time to wireless connection for each time point within a cycle is computed from historical data and the estimators $\hat{v}_1,...,\hat{v}_n$, $\hat{V}_1,...,\hat{V}_{\hat{m}}$ are obtained using utility functions.

Then $x_1,...,x_n$ can be estimated by solving,

$$\max_{x_1,...,x_n,y_1,...,y_{\hat{m}}} \quad \sum_{i=1}^{n} v_i q_i + \sum_{i=1}^{n} (q_i - x_i)\hat{v}_i \tag{1}$$

$$+ \sum_{i=1}^{\hat{m}} \hat{V}_i \hat{Q}_i + \sum_{i=1}^{\hat{m}} (\hat{U}_i - \hat{V}_i) y_i \tag{2}$$

$$s.t. \quad 0 \le x_i \le u_i, i = 1,...,n \tag{3}$$

$$0 \le y_i \le \hat{Q}_i, i = 1,...,\hat{m} \tag{4}$$

$$\sum_{i=1}^{n} x_i + \sum_{i=1}^{\hat{m}} y_i \le B \tag{5}$$

We note that this is a standard linear programming problem. In addition, with the assumption that all data packets have approximately the same size, the number of data packets that should be uploaded via 3G at time t_0 can be solved using following approach,

1) sort the all data clusters, including data clusters in current queue and future data clusters estimated from historical data, in the decreasing order of utility loss, i.e. $u_i - \hat{v}_i$ and $\hat{U}_i - \hat{V}_i$
2) if the budget B allows, the data packets with larger $u_i - \hat{v}_i$ or $\hat{U}_i - \hat{V}_i$ will have higher priorities of being uploaded via 3G network, according to the order derived above. If a future data cluster has priority and the budget also allows, even though it is not generated at time t_0, budget will be reserved for it. Thus there will be less budget for the data behind it. A data packet in queue at time t_0 will be uploaded at time t_0 via 3G network if all data clusters in front of it have not taken up all available budget.

The online 3G-budget algorithm needs to keep the following historical data to process: assuming that there are n time points in a cycle, and the total number of data types is m. First, a $n \times m$ matrix is used to record average amount of data generated for each data type at any time point in previous cycles. Then another $n \times m$ matrix to keep the number of cycles for the first matrix is required. Similarly, two vectors with length n are needed to record the average time to meet WiFi access point at each time point in previous cycles and the number of cycles for these records. Therefore, totally $2n(m+1)$ records need to be kept, and the total storage overhead highly depends on n indicating the frequency of running the algorithm.

In summary, the online 3G-budget algorithm is described in Algorithm 1.

C. Heuristic Algorithm

One drawback of the online 3G-budget algorithm is that it requires a large amount of computation to update the matrixes for historical data and run the algorithm at all time points, therefore, it may consume a big amount of storage and energy for resource-constraint mobile phone platforms.

Based on this observation, we propose a heuristic algorithm to provide a simple and effective solution. The idea is to split the overall 3G budget in each cycle into two pieces: reserved budget B_1 and flexible budget B_2. Namely, B_1 is reserved for those applications that are sensitive to environmental changes, denoted by *SENSITIVE*. applications in *SENSITIVE* can be selected by setting a threshold by the application provides about when they think the data is not interested any more, and a predefined threshold can be used to differentiate *SENSITIVE*. Based on the historical information, the average amount of data generated for *SENSITIVE* in a cycle can be calculated, denoted by n'. To make the reservation more conservative, we also set a balance coefficient, denoted by α, to n'. Thus, B_1 can be obtained by:

$$B_1 = \alpha \cdot n' \tag{6}$$

The flexible budget, B_2, can be used by those applications that are not sensitive to environmental changes, denoted by

Fig. 2. The hardware set.

Non-SENSITIVE. Moreover, applications in *SENSITIVE* can use this flexible budget as well if B_1 runs out and there are still remaining B_2.

When a new cycle starts and there are remaining data in the queue from the last cycle, these data will be uploaded using the new budget following the same rule as above. The order of upload in *SENSITIVE* and *Non-SENSITIVE* is based on the application thresholds set by different providers, and those with more sensitive properties have higher priority.

Within a cycle, the algorithm only runs at time points when new data are generated and the budget is not empty. Decisions are made for data in the queue in a greedy fashion: If there is budget in the right category, send that data via 3G; otherwise, wait for WiFi communications.

In summary, the details of the heuristic algorithm are described in Algorithm 2.

IV. Evaluation

In this section, we evaluate the performance of our approaches in terms of data utility based on an outdoor deployment of human subjects study in 2012. [2] For better understanding of the impact of the 3G budget, we investigate the performance trend with regard to varying amount of 3G budget.

A. Experimental Setting

We now give detailed information on the experiments and dataset we use in the evaluation of our solutions. To evaluate the proposed communication scheme, we conducted a human subjects study in 2012. The study involved 30 participants (drivers), and each of them was required to drive at least 100 miles. While we expect a participatory sensing application to run on participants' own phones, in this study we gave our subjects phones pre-loaded with our software (of which 15 were Galaxy Nexus phones [18], shown at the bottom right in Figure 2, and 15 were Nexus S phones [17], shown at the top right). These phones were placed under the windshield of their vehicles, as shown in circle A in Figure 3. Participants included students from various departments. To emphasize the performance in vehicular environments, they were asked to keep our phone in their car when driving, but otherwise carry on their daily routines as usual. They were not restricted to any specific routes and were not asked to change their normal driving habits. The phones were kept charged through a lighter-to-USB charger (Circle C in Figure 3) to support a long-term experiment and reduce the odds of individuals otherwise forgetting to charge a phone they do not use. When the phone was not charging (i.e., the engine is off), sensor data collection stopped but the communication component

[2] This study was conducted with approval of the Institutional Review Board as IRB protocol 10092.

```
input : A smart phone with multiple applications
        running.
output: Communication decisions.
initialization;
while system is running do
    B1 = current budget for SENSITIVE;
    B2 = current budget for Non-SENSITIVE;
    monitor the budget values and data queue;
    if data queue is not empty at current time point
    then
        update historical information;
        find the most sensitive data x;
        if x ∈ A then
            if reserved budget B1 > 0 then
                send x via 3G;
                update B1;
            else
                if flexible budget B2 > 0 then
                    send x via 3G;
                    update B2;
                end
            end
        else
            if flexible budget B2 > 0 then
                send x via 3G;
                update B2;
            end
        end
        find next x, repeat;
    end
    if new cycle starts then
        update B1, B2;
        check the data queue;
        while there are remaining data do
            find the most sensitive data x;
            send x via 3G following the same rule as
            above;
            update B1, B2;
            if no budget left then
                break;
            end
        end
    end
end
```

Algorithm 2: Heuristic algorithm.

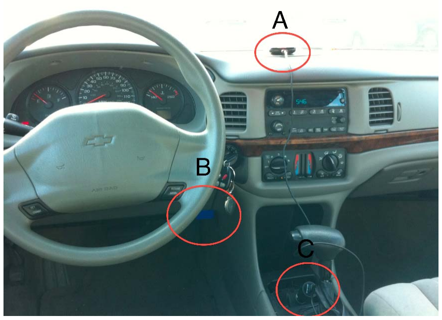


Fig. 3. System in deployment.

continued to run. The participatory sensing application in this deployment was to collect various types of sensory data about the vehicles while participants were driving, including GPS locations (GPS), on-board diagnostics (OBD) data [19], acceleration values (ACC), and gyros (GYRO) data. Collection of this data set was a good experimental methodology because these data can be used for multiple purposes, such as enhancing greener driving habits and analyzing traffic conditions.

Of the two types of phones used, the Galaxy Nexus is newer and much more powerful. The Nexus S phones have a 1 GHz Samsung Exynos CPU, 512 MB RAM, 16 GB iNAND storage partitioned into 1 GB internal storage and 15 GB USB storage, and uses a 1,500 mAh rechargeable Li-ion battery. The Nexus S supports GPS, Bluetooth 2.1 + EDR, and Wi-Fi 802.11b/g/n, accelerometer input, and 3-axis gyroscope input. The Galaxy Nexus phones have a 1.2 GHz TI OMAP 4460 ARM Cortex-A8 dual core CPU, 1 GB RAM, 16 GB built in storage, and uses a 1,750 mAh rechargeable Li-ion battery. The Galaxy Nexus supports GPS, Bluetooth v3.0 + HS, and Wi-Fi 802.11a/b/g/n, accelerometer input, and 3-axis gyroscope input.

Engine data were collected by the ELM 327 OBD-to-bluetooth adaptor (bottom center in Figure 2), connected to the OBD-II port, which is a standard feature in all vehicles sold in the US since 1996, and usually located under the dashboard (as shown in Circle B in Figure 3). Collected engine information included engine speed, RPM, mass air flow, and throttle position. This information was transmitted to the smart phone via bluetooth. The GPS, acceleration, and gyro data were collected via corresponding sensors on the phone respectively. GPS samples were obtained at 1 sample/second for both types of phones. The sampling rate of the OBD data, accelerometer, and gyroscope sensors were all set to be as fast as the hardware allowed. We observed that the actual data rate varied depending on car engine and phone. These generated data were also displayed on a logging interface screen so as to notify the participants that the system was running. All data were saved in a local database on the phone, and offloaded eventually to our backend server, when an available WiFi access point was present. We used the UIUC-campus-wide IllinoisNet as the only authentication-required WiFi that was available to the phones, which implies that participants can upload their data when they drive to office but not when they are at home. All open public WiFi access points could be used as well.

To record each meaningful event and generate a complete dataset, a log file was stored locally on each smart phone. It included the timestamp and content of GPS, OBD, ACC, and GYRO data, the start and end timestamps of each WiFi access point encounter, as well as which sensory data packets were offloaded to the backend server during each encounter. In addition, the timestamps when the car engine was ON/OFF were also recorded for statistics purposes. Finally, we noticed that ten participants finished this deployment by taking one long-distance trip instead of driving locally for a long period of time, thus we excluded them out and only focus on the other 20 participants.

B. Methodology

We proceed to describe the methodology adopted in our evaluation. We first obtain the whole dataset by combining the local log files of 20 participants. Then we divide each log file into multiple cycles for 3G budget assignment. In our evaluation we set the length of a cycle to 24 hours. In each cycle, the detailed information of generated data and WiFi encounters is available. We define the GPS data as the *SENSITIVE* category to simulate real-time traffic monitoring applications, in which the data is not fresh any more after 30 minutes; and the other three types of data as the *NON-SENSITIVE* category to simulate fuel consumption analysis and driving habit improvement applications, in which the data can last as long as a month. For simplicity of calculation, we define the utility function, donated by $f(t)$, for a piece of data in *SENSITIVE* (*NON-SENSITIVE*) as: $f(\cdot)$ starts from 1 when the data is generated, and decreases linearly as time continues, finally achieves a predefined threshold at time of 30 minutes (30 days). We set the threshold to 0.001 in our experiment.

We compare the system performance of three candidate solutions: *Baseline*, *3G-Budget*, and *Heuristic*. *Baseline* represents the default case in which 3G communications are not applicable and data can only be offloaded by WiFi access points. The results for *Baseline* can be calculated directly from the dataset. The other two solutions, on the other hand, require the replay of the dataset by running their own algorithms as described in Section III-B and III-C, respectively. Varying amounts of 3G budget are adopted to evaluate its impact on system performance for applications in different categories.

C. Results

1) Statistics: In this section, we present some statistics from our deployment, and analyze how they are related to the performance of our proposed solutions.

Figure 4 presents the number of active trips and offloading events (where by offloading, we mean upload data to server) for the 20 participants. Here active trips are defined by the intervals in which sensory data are generated during driving. We set the threshold between two consecutive active trips to 30 minutes. Note that offloading events refer to those really have

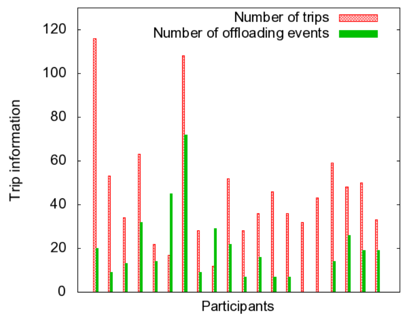


Fig. 4. Trip information for participants.

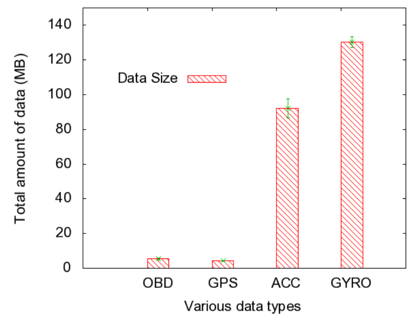


Fig. 5. Comparison of total data amount among various data types.

data uploaded to backend server, not including when the WiFi connection is established but there is not enough time to finish the socket connection and transfer data. We observe that both numbers vary a lot for different users. Some users take more trips than others in the deployment, within the range from 22 to 116 trips. This result indicates that some participants drive more frequently than others. The frequency of offloading events also varies for different users. For instance, we found that two participants were never able to finish an offloading event, because there are no WiFi coverage on their normal driving routines. On the other hand, one participant finished 45 offloading events in 17 active trips, which implies that his/her frequent driving route is well covered by WiFi. Overall, the number of offloading events is small compared to the number of active trips for most participants, this is reasonable because most students go to office/class by walk or school bus, and there are no IllinoisNet coverage when they drive to other off-campus places such as grocery stores or shopping malls. These results indicate that the utility for uploaded sensory data can be very low due to the fact that open wireless access points are not widespread. Our proposed solutions take advantage of the quick delivery of 3G communications and improve the systems performance in terms of utility of uploaded data.

We also recorded the total amount of data in different types for each participant in the deployment. This gives us the comparison of data size in the *SENSITIVE* and *NON-SENSITIVE* categories. The results of total amount of data are shown in Figure 5. We can see that the average total amount of data for OBD, GPS, ACC, and GYRO are 5.38, 4.26, 92.3, and 130.22 mega bytes (MB), respectively. These results indicate that there are 53.5 times of data generated for *Non-SENSITIVE* than the *SENSITIVE*, which implies that the *Non-SENSITIVE* will be dominating in the overall performance.

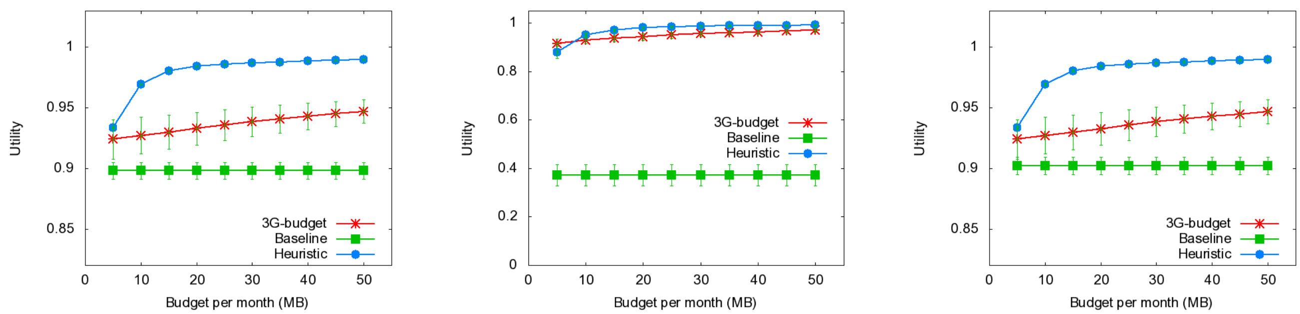
2) Performance: We now compare the performance of the candidate solutions, *Baseline*, *3G-Budget*, and *Heuristic* (with balance coefficient set to 1), that we described in Section IV-B. The metrics are the utility of data received at the backend server. The amount of 3G budget every month ranges from 5 MB to 50 MB, stepping by 5 MB. We choose this range mainly based on two reasons. First, the total amount of GPS data was around 5 MB on average. Therefore, with this smaller budget, we can investigate the scenario in which some *SENSITIVE* data are on hold in queue for the next cycle when new budget is available. Second, participants may not want to reserve too much budget from their limited data plan. Taking the popular AT&T 200 MB/month data plan [20] for example, one quarter of the limit, which is 50 MB, may be the largest possible budget contributed by participants. The budget is reset to 0 after each 30 cycles to reflect that data plan starts for a new month. The dataset for each participant is applied to the three candidate solutions, and statistics such as mean and standard deviation are calculated.

Figure 6 shows the results of data utility with error bars using different candidate solutions under varying amount of 3G budget. First, we notice that the overall utility for *Baseline* is 0.89, and only 0.37 for *SENSITIVE*. These results indicate that data offloading only by WiFi access points is not sufficient enough, especially for those applications that are quite sensitive to environmental changes.

Second, we observe that both *3G-budget* and *Heuristic* bring a performance gain to the *Baseline* solution, even when the 3G budget is 5 MB. The average utility increases from 0.37 to 0.93 using *3G-budget* and 0.88 using *Heuristic* for *SENSITIVE*, which implies 151.4% and 137.8% improvements, and from 0.90 to 0.96 using *3G-budget* and 0.93 using *Heuristic* for *NON-SENSITIVE*. These results demonstrate that 3G communications are indispensable in mobile participatory sensing applications, and a small portion of data plan (e.g., 2.5% of 200 MB in our case) can lead to a boost of system performance especially for those applications that are more sensitive to environment changes.

Third, it is obvious to see that, the performance goes up as the amount of 3G budget increases for *3G-budget* and *Heuristic* in all cases. This is because more packets are offloaded faster through 3G communications. Especially, the utility for *SENSITIVE* increases from 0.934 at 5 MB to 0.961 at 50 MB using *3G-budget* and from 0.879 at 5 MB to 0.991 at 50 MB using *Heuristic*. The variance of data utility decreases too as the amount of budget increases. These results indicate that the more budget participants are willing to contribute to the participatory sensing applications, the better system performance it will lead to.

Fourth, the performance of *3G-budget* is slightly better than *Heuristic* for *SENSITIVE* when the budget is very small (5



(a) Overall. (b) *SENSITIVE*. (c) *Non-SENSITIVE*.

Fig. 6. Utility Comparison: *Baseline*, *3G-budget*, and *Heuristic*.

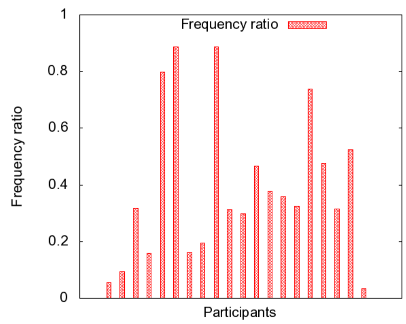


Fig. 7. Frequency ratio of *Heuristic* to *3G-budget* on active cycles.

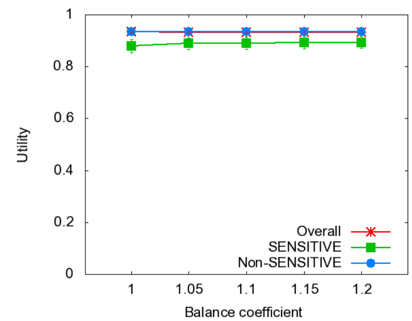


Fig. 8. Utility performance under varying balance coefficient.

MB) compared to the amount of generated data. The utility is 0.93 for *3G-budget* and 0.88 for *Heuristic*. This is mainly because only a portion of budget is reserved to *SENSITIVE* using *Heuristic*, and the other part is quickly consumed by data in *Non-SENSITIVE*, while *3G-budget* holds on the budget for a while to wait for the generation of new *SENSITIVE* data. When the budget increases from 10 to 50 MB, *Heuristic* outperforms *3G-budget* because more budget are assigned to *SENSITIVE* in *Heuristic*, but the data running *3G-budget* sometimes are on hold if the decision making algorithm estimates that a WiFi encounter is approaching. We also notice that *Heuristic* results in better utility than *3G-budget* for *Non-SENSITIVE*. The main reason is that the driving pattern for participants are not regular enough for accurate estimation in *3G-budget*. We noticed that most participants did not drive much on weekdays, probably because they went to office/class by walk or school bus. This makes it a common case that much more data are generated on weekends than weekdays, therefore, the estimation is affected by anticipating that more data will be generated later in the data plan period with larger utility loss and thus decides to hold the budget, but finally that does not happen.

Figure 7 depicts the average frequency ratio of *Heuristic* to *3G-budget* in cycles with new data generated for each participant. For instance, a value 0.36 for Participant 3 indicates that *Heuristic* runs 36% time of *3G-Budget* on average for this user. In general, we observe that *3G-budget* runs a lot more than *Heuristic* for most users. The ratio ranges from 0.03 to 0.88 and the average is 0.39, which implies that *Heuristic* only runs less than 40% than *3G-budget* on average.

In addition, both *3G-budget* and *Heuristic* perform well in terms of customized for the needs of different types of participatory sensing applications. The comparison of Figure 6(b) with 6(c) clearly shows that data in *SENSITIVE* obtained more performance gain than those in *Non-SENSITIVE*, based on their different utility functions. These results demonstrate that our proposed solutions are capable of dealing with multiple heterogeneous participatory sensing applications and provide tailored solution to improve the overall utility in the system.

Finally, we compare the utility performance using *Heuristic* under varying balance coefficients, ranging from 1 to 1.2 stepping with 0,05. The 3G budget is 5 MB. The results are shown in Figure 8. We can see that the utility for *SENSITIVE* increases from 0.879 to 0.893 while the utility for Non-SENSITIVE decreases from 0.933 to 0.932. These results indicate that, when reserving more budget for *SENSITIVE*, more data in *SENSITIVE* were offloaded faster through 3G communication and thus the improved performance; while on the other hand, data in *Non-SENSITIVE* have to wait longer on average, resulting in poorer performance. Due to the relatively smaller 3G budget compared to the total amount of data in *Non-SENSITIVE*, the decreasing ratio is very small.

D. Discussions

Experimental results show that our proposed solutions achieve better system performance in terms of data utility, and can be tailored for heterogeneous needs of multiple participatory sensing applications. The *3G-budget* and *Heuristic* solutions have their own strengths and drawbacks, and can be applied under different application scenarios. *3G-budget* targets at optimizing the global data utility by collecting information of historical data, however, its performance rely much on the accuracy of the estimation. When participants drive in an irregular pattern, it is hard to predict its overall performance. In addition, it needs to run continuously at all time points no matter whether new data are generated or not. *Heuristic* runs the decision making algorithm less frequently and saves more energy, but when the budget is very small compared to the amount of generated data, its performance is not as good as *3G-budget* for the *SENSITIVE* category, because only a portion of budget is reserved to *SENSITIVE*, and the other part is quickly consumed by data in *Non-SENSITIVE*, while *3G-budget* holds on the budget for a while to wait for the generation of new *SENSITIVE* data.

We also investigated that different driving patterns by participants largely affect the system performance, and thus optimizing the system based on specific driving patterns of participants becomes an interesting future research direction. For example, as shown in our work, when a participant (e.g. graduate student) drives infrequently on weekdays, running *3G-budget* result in a big waste of energy for most of the time and inaccurate estimation because of this irregular driving pattern, therefore *Heuristic* is a better choice when the budget is not small.

Our communication framework can be further optimized by other potential research directions. For instance, there are potential opportunities to combine our work with the smartphone based vehicular networking techniques that have been emerging recently [21]. Due to the unknown *priori* information, opportunistic peer-to-peer sharing can help offloading those packets in *SENSITIVE*, and corresponding prioritization algorithm becomes an interesting research problem.

V. Conclusion

In this paper, we presented the design, implementation, and evaluation of a novel participatory sensing system that leverages the 3G budget that participants contribute and is customized to the heterogeneous needs of multiple participatory sensing applications. Our proposed two algorithms improve the utility of uploaded participatory sensing data and no user involvement is needed. Our results, from a 30-vehicle campus-wide deployment, show that even when the budget is as small as 2.5% of a popular data plan, these two algorithms achieve higher utility of uploaded data compared to the baseline solution, especially, they increase the utility of received data by 151.4% and 137.8% for those sensitive applications. Given the popularity and stronger power of smart phones and importance of vehicular sensor networks, we hope that this work will motivate further research on dealing with limited 3G data plan and running multiple participatory sensing applications simultaneously.

Acknowledgement

Research reported in this paper was sponsored by NSF under grants CNS 1040380 and CNS 10-59294, and the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] R. Balan, N. Khoa, and J. Lingxiao. Real-Time Trip Information Service for a Large Taxi Fleet. In MobiSys, 2011.
[2] P. Mohan, V. N. Padmanabhan, and R. Ramjee. Nericell: using mobile smartphones for rich monitoring of road and traffic conditions. In SenSys, 2008.
[3] R. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. Abdelzaher. GreenGPS: A Participatory Sensing Fuel-Efficient Maps Application. In Proceedings of ACM 8th Annual International Conference on Mobile Systems, Applications and Services (MobiSys), 2010.
[4] E. Miluzzo, N. D. Lane, K. Fodor, R. A. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A T. Campbell. Sensing Meets Mobile Social Networks: The Design, Implementation and Evaluation of the CenceMe Application. In Proc. of 6th ACM Conference on Embedded Networked Sensor Systems (SenSys '08), Raleigh, NC, USA, Nov. 5-7, 2008.
[5] N. Banerjee, M. D. Corner, and B. N. Levine. Design and Field Experimentation of an Energy-Efficient Architecture for DTN Throwboxes. IEEE/ACM Transactions on Networking, 18(2):554–567, April 2010.
[6] B. Hull, V. Bychkovsky, K. Chen, M. Goraczko, A. Miu, E. Shih, Y. Zhang, H. Balakrishnan, and S. Madden. CarTel: A Distributed Mobile Sensor Computing System. In Proc. ACM SenSys, 2006.
[7] S. Mathur, T. Jin, N. Kasturirangan, J. Chandrasekaran, W. Xue, M. Gruteser, and W. Trappe. Parknet: drive-by sensing of road-side parking statistics. In MobiSys, 2010.
[8] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson. VTrack: accurate, energy-aware road traffic delay estimation using mobile phones. In SenSys, 2009.
[9] Canalys 2010. URL: http://www.canalys.com/pr/2010/r2010081.html.
[10] Customers Angered as iPhones Overload AT&T. NY Times, September 2009. URL: http://www.nytimes.com/2009/09/03/technology/companies/03att.html.
[11] iPhone overload: Dutch T-Mobile issues refund after 3G issues. Ars Technica, July 2010. URL: http://arstechnica.com/tech-policy/news/2010/06/dutch-t-mobile-gives-some-cash-back-because-of-3g-issues.ars.
[12] A. Balasubramanian, R. Mahajan, and A. Venkataramani. Augmenting Mobile 3G Using Wifi. In ACM MobiSys, 2010.
[13] K. Lee, I. Rhee, J. Lee, Y. Yi, and S. Chong. Mobile data offloading: how much can wifi deliver? In ACM SIGCOMM 2010.
[14] B. Han, P. Hui, V. S. A. Kumar, V. M. Marathe, G. Peig, and A. Srinivasan. Cellular traffic offloading through opportunistic communications: a case study. In ACM CHANTS, 2010.
[15] Marco Valerio Barbera, Julinda Stefa, Aline Carneiro Viana, Marcelo Dias de Amorim, and Mathias Boc. VIP Delegation: Enabling VIPs to Offload Data in Wireless Social Mobile Networks. In DCOSS, 2011.
[16] Shahriar Nirjon, Angela Nicoara, Cheng-Hsin Hsu, Jatinder Singh, and John Stankovic. MultiNets: Policy Oriented Real-Time Switching of Wireless Interfaces on Mobile Devices. In IEEE RTAS, 2012.
[17] The Google Nexus S phone. URL: http://www.google.com/phone/detail/nexus-s.
[18] The Google Galaxy Nexus phone. URL: http://www.google.com/nexus/#/tech-specs.
[19] R. Birnbaum and J. Truglia. Getting to Know OBD II. New York, 2000. ISBN 0-9706711-0-5.
[20] The AT&T data plan. URL: www.att.com/shop/wireless/plans.html.
[21] Sacha Trifunovic, Bernhard Distl, Dominik Schatzmann, and Franck Legendre. WiFi-Opp: ad-hoc-less opportunistic networking. In ACM CHANTS, 2011.