OppSense: Information Sharing for Mobile Phones in Sensing Field with Data Repositories

Edith C.-H. Ngai[1], He Huang[1], Jiangchuan Liu[2] and Mani B. Srivastava[3]

Abstract— **With the popularity and advancements of smart phones, mobile users can interact with the sensing facilities and exchange information with other wireless devices in the environment by short range communications. Opportunistic exchange has recently been suggested in similar contexts; yet we show strong evidence that, in our application, opportunistic exchange would lead to insufficient data availability and extremely high communication overheads due to inadequate or excessive human contacts in the environment.**

In this paper, we present *OppSense*, a novel design to provide efficient opportunistic information exchange for mobile phone users in sensing field with data repositories that tackles the fundamental availability and overhead issues. Our design differs from conventional opportunistic information exchange in that it can provide mobile phone users guaranteed opportunities for information exchange regardless the number of users and contacts in different environments. Through both analysis and simulations, we show that the deployment of data repositories plays a key role in the overall system optimization. We demonstrate that the placement of data repositories is equivalent to a connected K-coverage problem, and an elegant heuristic solution considering the mobility of users exists. We evaluate our proposed framework and algorithm with real mobile traces. Extensive simulations demonstrate that data repositories can effectively enhance the data availability up to 41% in low contact environment and significantly reduce the communication overheads to only 28% compared to opportunistic information exchange in high contact environment.

I. Introduction

Wireless Sensor Networks (WSNs) have been widely investigated for monitoring temperature, sound, motion, pollutants, etc. from the environment for many scientific and emergency applications [1]. Wireless sensors are resource-limited devices that do not provide user interfaces for to read the data, so that sensing data are usually reported by multi-hop communications to the sink(s) for processing [2]. With the popularity of mobile phones, innovative sensing applications are emerging and moving beyond scientific domain into broader and human interactive sensing for personal, social, and urban usages [3]. Mobile phones provide user-friendly graphical interfaces for visualizing and processing sensing data [4], [5]. Naturally, human beings can collect data from the nearby sensors using their mobile phones and benefit from the data for their own activities. More than that, opportunistic information exchange among the mobile phones becomes possible through intermittent connections through short-range communications, like bluetooth and WiFi, which offers another way to collect and share data with their surroundings. Compared with cellular networks (e.g. 3G mobile networks), opportunistic information exchange is more energy-efficient and less costly, especially when the data sources and the users are physically close to each other. This opportunistic network setting does not require any centralized server or infrastructure for communication and management. It works well even in areas with poor or no cellular coverage, and reduces the workload of cellular networks in dense areas. To fully explore these potentials, new system design and algorithms are necessary to provide efficient data collection and opportunistic information exchange for mobile phone users.

[1]Dept. of Information Technology, Uppsala University, Sweden
[2]School of Computing Science, Simon Fraser University, Canada
[3]Dept. of Electrical Engineering, University of California, Los Angeles, USA

In this paper, we consider innovative and human interactive sensing applications that integrate collection and sharing of sensing data among mobile phone users. The mobile phone users can collect data from the wireless sensors along their walks by direct communications. They also indicate their interests of data in the sensing field, so that they can exchange the sensing data with other mobile phone users. Potential applications include data collection and information exchange among mobile phone users in tourist attractions, hiking trails, universities and urban areas, etc. For example, a visitor of a national park, who is allergic to pollen, can collect real-time sensing data of the pollen level in his surroundings before deciding on his following paths. Other than that, users can monitor the environment such as pollution level in cities, make recommendations of shops and restaurants, or exchange information on road traffic and public transportations in urban areas.

Although there are existing work on opportunistic content sharing between mobile users [6], [7], the locality of data and the network performance have not been throughly investigated. For instance, the songs or videos that the mobile users are interested in may not have any direct relationship with their locations. On the contrary, the sensing data and the interest of mobile phone users are closely related to their locations and activities in our sensing applications. We believe that the efficiency of information exchange and the network performance could be greatly improved considering the locality of data and mobile phone users. In addition, the performances of existing opportunistic data sharing mechanisms depend heavily on the contact rates between mobile phone users. It gives no guarantee on the data availability and the communication overheads. If the contact rates between mobile phone users are too low, mobile phone users may obtain inadequate data. On the other hand, the communication overheads could be extremely high and inefficient if there are too frequent contacts

978-1-4577-0093-4/11/$26.00 ©2011 IEEE

between the users.

To address the above problems, we propose *OppSense*, a novel data collection and information exchange framework that can provide efficient information sharing with guaranteed data availability at reduced communication overhead by using data repositories. The objective of OppSense is: considering the paths of mobile phone users, we minimize the number of data repositories in the field, while guaranteeing satisfactory data availability at reduced communication overheads. We exploit the mobility patterns of users and deploy data repositories at optimal locations to provide sufficient cover to the paths of mobile phone users. Data repositories allow information exchange between users arriving at different times. The mobile phone users can upload and download sensing data at the data repositories independent of the number of users and their contact rates in the field. In addition, the data repositories can remove redundant data from the same sensors to improve the efficiency of information exchange. We formulate the placement of data repositories as a connected K-coverage problem and present a heuristic algorithm by exploiting the mobility of users. Finally, we analyze the system performance and evaluate our proposed framework based on real mobile traces in terms of data availability and communication overheads.

In this paper, we investigate on data collection and information exchange among mobile phone users for human interactive sensing environments. As a summary, the contributions of this paper are:

1) We propose a novel data collection and information exchange framework for mobile phone users in a sensing field. Different from traditional opportunistic exchange, our approach can provide users guaranteed opportunities for information exchange at minimized communication overheads in environments with any contact rates.
2) We formulate the placement of data repositories as a connected K-coverage problem and provide a heuristic algorithm by exploiting the mobility of mobile phone users. The impact of K on data availability and communication overheads are analyzed.
3) Extensive simulations are conducted based on real mobile traces [8], [9], which demonstrate that data repositories can effectively enhance data availability up to 41% in low contact environment and signficantely reduces the communication overheads to only 28% compared to opportunistic exchange in high contact environment.

The remainder of this paper is organized as follows: Section II presents related work. Section III describes the network architecture of opportunistic data sharing in human interactive sensing applications. In Section IV, we present the system design of OppSense that supports information exchange among mobile phone users by using data repositories. In Section V, we present the problem and solution for the placement of data repositories. We analyze the system performance in Section VI. In Section VII, we conduct extensive simulations to evaluate our solution based on real mobile traces. Finally, we conclude the paper in Section VIII.

II. Related Work

Mobile sinks and relays have been suggested for assisting data delivery in diverse wireless networks. Shah et al. [10] presented an architecture using moving entities (*Data Mules*) to collect sensing data. There have also been studies on mobile sinks with predictable and controllable movement patterns [11], [12], and the optimal time schedule for locating sojourn points [13]. Advanced algorithms [14], [15], [16] have been proposed to optimize the movement of mobile elements for better performance. The above work consider many-to-one communications from the sensors to the sink(s) and focus on controlling and determining the movement of the mobile nodes for data collection. None of them has considered that individual mobile users, other than the sinks, could also be the consumers of the sensing data. In this work, we support mobile phone users who want to collect and share sensing data according to their own interests and activities. Different from previous work, mobile phone users have their own moving patterns which could only be observed but not actively controlled in our applications.

Opportunistic networks [17] and delay tolerant networks (DTNs) [18] have been proposed as an interesting evolutions of Mobile Ad Hoc Networks (MANETs). Messages are routed through any possible node opportunistically as next hop, provided that it is likely to bring the message closer to the final destination. A number of routing and forwarding protocols have been proposed for opportunistic networks and DTNs [19], [20]. Burns et al. [21] proposed the MV routing protocol which learns the movement pattern of network participants and uses it to enable informed message passing. A similar approach is followed in the PROPHET protocol [22] to improve the delivery rate. Zhao et al. [23] proposed a *message ferrying* approach to address the network partition problem in sparse ad hoc networks. All these work focus on routing between two selected nodes through their intermediate peers opportunistically. On the contrary, our work does not specify any source and destination pairs, but it allows mobile phone users to share information with any intermittently connected peers freely according to their interests.

Social-aware content sharing have been studied for opportunistic networks [6]. Boldrini et al. [7] proposed a middleware that autonomically learns context and social information on the users of the network to predict users' future movements. Yoneki et al. [24] proposed a socio-aware overlay over detected communities for publish/subscribe communication. Jaho et al. [25] divided users into different interest-induced social groups and locality-induced social groups to improve information dissemination in social networks. Our work shares a similar view of considering human beings as communities in opportunistic content sharing. Different from them, our system guarantees satisfactory information exchange opportunities independent from the contact rates of users by applying data repositories.

Data repositories can improve the efficiency in information exchange and reduce the communication overheads of mobile phones. *Infostations* was proposed as a concept to provide isolated pockets of high bandwidth connectivity for data and messaging services in wireless systems, which complements to the relatively high cost and low bit-rate cellular systems [26]. Similarly, Ditto was proposed as a system to improve transfer performance by opportunistic caching data at nodes on the multi-hop route of data for mesh networks [27]. Although our work also leverages data caches to improve the system performance, OppSense considers a drastically different environment for data collection and information exchange with wireless sensors and mobile phones. Sensing data could be collected, carried, and shared among the mobile phones opportunistically, rather than through multi-hop communications. Most importantly, the locality of sensing data and the mobility pattern of mobile phone users have to be considered jointly to achieve satisfactory data availability at reduced communication cost.

III. Opportunistic Information Exchange for Mobile Phone Users in Sensing Field

A. Network Architecture

We consider a network which involves both wireless sensors and mobile phone users in a sensing field. The wireless sensors are sparsely deployed in a sensing field for monitoring temperature, humidity, noise level, pollen level, etc. from the environment. Users are carrying mobile phones that have the capability to request, store, process and share information. As they move in the sensing field, temporary occasions arise for short range communications with wireless sensors and other mobile phones, which open up the opportunity to collect data and share information. Short-range opportunistic communication between mobile phones (e.g. bluetooth) does not rely on any network infrastructure which is suitable for areas with poor or no cellular coverage. It can save the bandwidth and reduce the workload of cellular networks, and it is more energy-efficient. Potential human interactive sensing applications include sensing and information sharing among mobile phone users in national parks, tourist attractions, hiking trails, urban areas, universities, factories, etc. With the popularity of smart phones and PDAs, more innovative applications are emerging which can cover broader areas of our daily lives, social activities and public welfares.

Mobile phone users are direct consumers of the sensing data in our system. Depending on their activities and own preferences, they can collect data of their interests directly from the nearby sensors. Figure 1 shows an application of data collection and information exchange of mobile phone users in a national park. Each mobile phone can collect data from the wireless sensors within the communication range. Apart from that, a user may also want to receive information about other parts of the sensing field, e.g. a nearby area in the same national park where he has not visited yet. Information exchange allows him to obtain new information from other mobile phone users in the sensing field (see Figure 2). In this example, users 1 and 2 have collected and buffered different data $D1$ and $D2$ respectively. $D1$ from sensor 1 indicates a

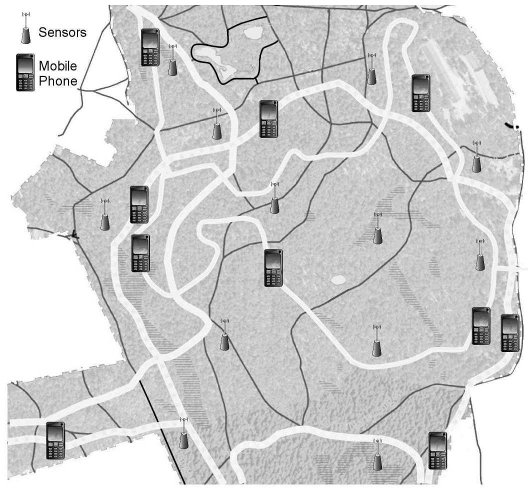

Fig. 1. Visitors of a national park can collect sensing data along their walks using their mobile phones.

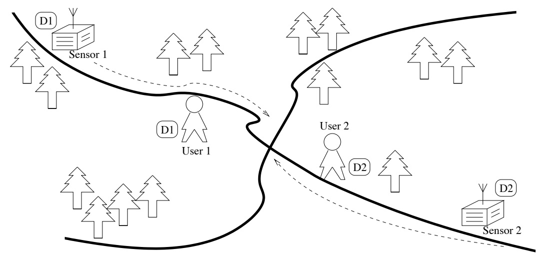

Fig. 2. Users 1 and 2 can exchange data when they meet at the cross.

high reading of the pollen level in that area. If user 2 is allergic to pollen, he may avoid going to that area.

B. Data Collection and Information Exchange

More formally, we divide the sensing field into a number of grid cells. There are I sensors, $S_1, S_2, ..., S_I$ deployed at grid cells $g_1, g_2, ..., g_I$ to monitor the environment. They can communicate with any mobile phones within their communication range F. The mobile phones are carried around by human beings who are walking in the sensing field. When a mobile phone user approaches a sensor, he can collect data from the sensor and visualize the data on his mobile phone. We consider a slotted time $t = 0, 1, ...t_n$ in our system. Each mobile phone user i moves along an independent path $P_i = (g_i^{t_0}, g_i^{t_1}, ...g_i^{t_n})$ in the sensing field, where g_i^t is the grid cell that i is located at time t. Table I lists the parameters that we use in this paper.

We define Q_i^t as the data of interest from user i at time t and $C_i^t = \bigcup_{t'=t_0}^{t} c_i^{t'}$ as the set of data collected by i from the grid cells that he has visited, where $c_i^{t'}$ is the data collected by i at time t' from the sensor in grid cell $g_i^{t'}$. If he is interested in other data in the sensing field, he can exchange data with other mobile phone users within the communication range. We also define D_i^t as the total data obtained by i directly communicating with the sensors or exchanging information with other mobile phone users up to time t, where $C_i^t \subseteq D_i^t$. For a given pair of mobile users (i, j), we adopts the contact

TABLE I

NOTATIONS

g_s	grid cell s
P_i	path of mobile phone user i
N	number of paths
M	number of data repositories
K	average number of data repositories on a path
R	set of data repositories
F	communication range
g_i^t	grid cell that user i located at time t
Q_i^t	data interest of i at time t
C_i^t	grid cells visited by i up to time t
D_i^t	data received by i up to time t
$U_t^{i,j}$	contact status between i and j at time t
T_k	contact time for information exchange
b_i	collection time of the data stored by i
$E_i^{T_k}$	data obtained by i after exchange at time T_k
R_m	data repository m
L_m	location of data repository R_m
$CG = (V, E)$	communication graph
KB_v	K-benefit of adding v to R
H	average path length
p_c	average contact probability
p_{new}	average probability of new data in exchange
δd	average increase of data in one time slot

model defined by [28] as

$$U_t^{i,j} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are in contact during time slot t,} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

For the pair (i, j) we consider the sequence of the time slots $T_0 < T_1 < ... < T_k < ...$ that describes all the values of $t \in \mathbf{N}$ such that $U_t^{i,j} = 1$. If both users i and j are interested in the data of the whole sensing field. They can exchange their collected data, $D_i^{T_k}$ and $D_j^{T_k}$ when they are in contact at time T_k.

IV. OppSense: Information Exchange for Mobile Phones Using Data Repositories

A. System Design

We propose a novel and efficient data collection and information exchange framework for mobile phone users in a sensing field using data repositories. OppSense can ensure guaranteed data availability with minimized communication overheads for mobile phone users independent of their contact rates. It overcomes a major problem of existing opportunistic information exchange in which the system performance depends heavily on the contact rates of the users that may lead to insufficient data availability and extremely high communication overheads. The main idea is that mobile phone users will be provided with guaranteed opportunities to exchange information with the data repositories. Different from opportunistic exchange, users only need to exchange information with the deployed data repositories, rather than with any other users that they meet. We achieve the assured system performance by deploying data repositories at optimal locations considering

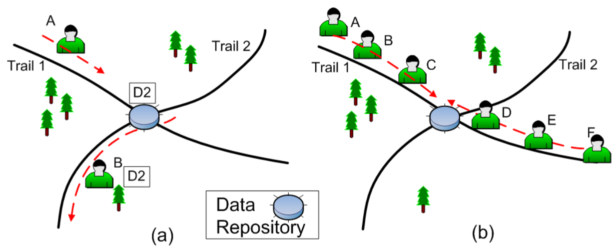

Fig. 3. (a) User A can download $D2$ at the repository, even though he did not meet user B directly. (b) The communication overheads are greatly reduced if the users only exchange information with the repositories, rather than with all users.

the mobility of mobile phone users. We observed that mobile phone users usually walk along common paths in many applications, e.g. hiking trails, walkways, etc. By exploiting the paths of users, data repositories could be carefully deployed to provide mobile phone users guaranteed opportunities to exchange information. Since the mobile phone users only exchange information with the data repositories, their data availability and communication overhead will not be affected by the excessive or inadequate contacts with the other users. In our system, the number of data repositories are minimized to reduce the deployment cost and communication overheads. In addition, data repositories will refresh and remove old data in their buffers to further improve the efficiency of information exchange.

Two major benefits of our system are illustrated as follows. Figures 3(a) and (b) show the same sensing environment with low and high contact rates respectively. In Figure 3(a), user B uploaded data $D2$ about trail 2 when he passed by the repository. User A can then download the data $D2$ when he arrives at the repository later. Our system can provide each mobile phone users opportunities to exchange information even though they do not meet each other. Consider another case in Figure 3(b), the high contact rates in opportunistic information exchange can bring very high communication overheads. The communication overheads are greatly reduced in our approach as the users only exchange information with the data repository. The data repository can buffer the data and reduce the communication overheads significantly without degrading the data availability.

B. Information Exchange in Data Repositories

In our system, stationary data repositories are deployed in the sensing field to provide efficient information exchange considering the mobility of users. Data repositories and mobile phone users can exchange information when they are within their communication range. A mobile phone user will upload his collected sensing data to the data repository and download the sensing data of his interests from there. Data repository will store the sensing data uploaded by the mobile phone users. If a data repository has received multiple readings from the same sensor, only the most recent reading will be stored.

Let T_i and T_j be the contact times of users i and j with the data repository R_m, where $T_i < T_j$. Supposed that the two users have both visited grid cell g_s at different times b_i and b_j with readings $r1$ and $r2$ respectively, where $b_i < b_j$. The data repository stored the reading $r1$ when user i arrived at T_i in the format of $< g_s, b_i, r1 >$. When user j arrives at time T_j later, the data repository finds that the sensor reading $r2$ from j is more recent than $r1$, give that $b_i < b_j$. Then, it will update the stored data as $< g_s, b_j, r2 >$. Based on this refreshing mechanism, the old data from the same locations could be removed. Thus, both the communication and storage overheads will be reduced.

C. Information Exchange in Mobile Phones

A mobile phone user will upload his collected data and download new data from the data repositories in an information exchange. We observe that the interests of data from individual users are spatially and temporally related to their current locations and activities. In general, a user i will be mostly interested in the recent data about his surroundings. He can gain new information from the repository R_m in contact at time T_k, $E_i^{T_k}$, which is represented by

$$E_i^{T_k} = \{Q_i^{T_k} \bigcap D_m^{T_k}\}, \tag{2}$$

where $Q_i^{T_k}$ is the data of interest from i, $D_m^{T_k}$ is the data gathered by R_m up to time T_k.

The total amount of data obtained by i after the exchange becomes

$$D_i^{T_k'} = E_i^{T_k} \bigcup D_i^{T_k} \tag{3}$$

$$= \left\{Q_i^{T_k} \bigcap D_m^{T_k}\right\} \bigcup D_i^{T_k}, \tag{4}$$

where $D_i^{T_k}$ is the data collected by i before the exchange.

V. Deployment of Data Repositories

The placement of data repositories is critical for efficient information exchange. A random or arbitrary placement may lead to inadequate data availability, high deployment cost and high communication overheads. In the worst case, a poor deployment would even partition the network and block the information flows between different areas. In our system, we aim at minimizing the number of data repositories deployed, while providing satisfactory data availability with reduced communication overheads for the mobile phone users. We observe that human beings usually walk along pre-defined or common paths, like the hiking trails in country parks, or walkways in campuses or cities. Our problem formulation aims at providing sufficient connected K-cover to the paths of the mobile phone users. The parameter K in the connected cover controls the minimum number of repositories along each path, which can be determined by the application requirements.

A. Formulation of Placement Problem

We consider a sensing field being divided into a number of grid cells of size F x F. We assume that the sensors are deployed at the center of the grid cells. Mobile phones and sensors in the same grid cell can communicate with each other. Consider a network with N paths $P_1, P_2, ..., P_N$ as shown in

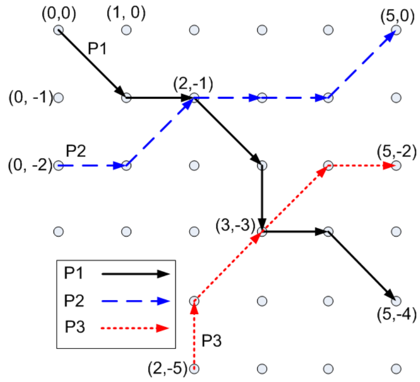

Fig. 4. Three paths $P1, P2$ and $P3$ are shown in a sensing field. Only the center of the grid cells and their (x,y) coordinations are indicated here.

Figure 4, a set of data repositories $R \in \{R_1, R_2, ...R_M\}$ will be deployed at the selected grid cells along these paths, where M is the maximum number of available data repositories.

Definition 1. **(Communication Graph)** Given a number of grid cells, V, the communication graph of the network is the undirected graph CG with V as the set of vertices and an edge between any two vertices if there is a path connecting the two vertices. Let $L \in \{L_1, L_2, ..., L_M\}$ be the set of grid cells where the set of data repositories $R \in \{R_1, R_2, ..., R_M\}$ will be deployed. The communication subgraph induced by the locations of repositories is the subgraph of CG involving only the vertices in L.

Definition 2. **(Connected K-Cover of Data Repositories)** Consider a network consisting of N users with a set of paths $P = \{P_1, P_2, ..., P_N\}$. Given a set of paths P of the mobile phone users, a set of grid cells L is said to be a connected K-cover of data repositories if the following two conditions hold:

1) For each $P_i = (g_i^{t_0}, g_i^{t_1}, ..., g_i^{t_{n_i}})$, there exists at least K data repositories $R_m \in R$ located along P_i, i.e. $\exists L_m = g_i^t$ for $t \leq t_{n_i}$, where t_{n_i} is the end of P_i. This condition implies that each path in P is covered by at least K data repositories.
2) The data repositories must be connected by the paths of mobile phone users. The subgraph induced by L in CG is connected, where CG is the communication graph of the grid cells. In other words, any data repository R_m located at L_m can communicate with any other data repositories in the cover, through the paths of mobile phone users.

The parameter K in our problem formulation indicates the minimum number of data repositories along each path in P. An analogy is similar to how many notice boards a hiker would see along his walk in a national park. A larger K provides the mobile phone users more opportunities to exchange information. However, it also requires greater number of data repositories which are constrained by the deployment cost in practice. Unreasonably high K would increase the communication overheads and degrade the efficiency of our system.

Connected K-Coverage Problem for Data Repositories: Given a set of paths P of mobile phone users in a sensing field, the connected K-coverage problem for data repositories is to deploy minimum data repositories that are connected and provide K-cover to all paths. The connected K-coverage problem is NP-hard as it is a generalization of the connected 1-coverage problem which is already known to be NP-hard [29], [30].

B. Placement Algorithm for Data Repositories

We present a greedy algorithm to solve the connected K-coverage problem for the placement of data repositories. The greedy algorithm works by selecting, at each stage, a path that connects an already selected grid cell to a grid cell that covers the insufficiently covered paths the most. In other words, the selected grid cell gives the greatest benefit, named as K-benefit, to achieve the K-cover. The algorithm terminates when the selected set of grid cells for data repository deployment provides K-cover to all paths of mobile phone users.

Intuitively, only grid cells along the paths of mobile phone users would be considered for data repository deployment. Let $CG = (V, E)$ be the graph formed by these grid cells $v \in V$. Any pair of nodes in V is connected by an edge $e \in E$ if they are along the same path. We define KB_v as K-benefit which indicates the benefits of adding a new data repository at location v.

$$KB_v = B(R \bigcup v, K) - B(R, K), \tag{5}$$

where $B(R, K) = \sum_{P_i \in P}(min(K, \sum_{R_m \in R} z(P_i, R_m)))$ and $z(P_i, R_m)$ is the number of times that path P_i is covered by data repository R_m. $B(R, K)$ denotes the K-value of a set of data repositories R. It is defined as the sum of the total number of times (bounded by K) each path is covered by the data repositories in set R.

Consider Figure 4 with $K = 2$ as an example, $B(\emptyset, 2) = 0$ initially. The K-benefit of grid cell $g_{(2,-1)}$ is $KB_{g_{(2,-1)}} = B(\{g_{(2,-1)}\}, 2) - B(\emptyset, 2) = 2$. Similarly, the K-benefits of grid cells $g_{(3,-3)}$ and $g_{(5,2)}$ are $KB_{g_{(3,-3)}} = 2$ and $KB_{g_{(5,2)}} = 1$ respectively.

The algorithm for the placement of data repositories is presented in Algorithm 1. First, we set the remaining data repositories DR_{num} as the maximum number of available data repositories M. R represents the set of already included data repositories which is initially empty. The algorithm starts with calculating the K-benefit KB_v for all grid cells $v \in V$. Then, it includes the grid cell $v*$ with the maximum K-benefit KB_{v*} in graph V. If multiple grid cells have the same K-benefit, the one with the maximum visiting frequency from the mobile phone users will be selected. The reason is that more popular grid cell will provide mobile phone users more opportunities to exchange information. By selecting the node with maximum K-benefit, the algorithm will provide the best cover to the insufficiently covered paths. After the selection, R and DR_{num} will be updated. At each stage, the algorithm adds in one data repository that is connected to the already installed data repositories R and covers the insufficiently covered paths the most. The algorithm continues as there are paths not satisfied K-cover and more data repositories are available.

Algorithm 1 Deployment of data repositories

```
{Step 1: Initialization}
Initialize K;
DR_num = M;
P = {P_1, P_2, ..., P_N};
R = ∅;

{Step 2: Adding data repositories}
while (P is not K-covered) and (DR_num > 0) do
   Find all nodes v ∈ V which are connected to subgraph induced
   by R in CG;
   Calculate the K-benefit, KB_v for all v;
   Find v* ∈ V with the maximum KB_v*;
   if Multiple nodes v* have the same KB_v* then
      Choose the v* with maximum visiting frequency;
   end if
   R = R ⋃ v*;
   DR_num = DR_num − 1;
end while
Return DR_num, R;
```

VI. Analysis

A. Data Availability

We define data availability as the data coverage received by a user of the sensing field. A user can collect the data either by himself or by exchanging information with the data repositories. The data availability is high if a user receives data from more grid cells in the field. We quantify the data availability as $|D_j^t|$, which is the cardinality of the set of collected data D_j^t in mobile phone j at time t.

The data availability considering data exchange between mobile phone j and data repository R_m in contact is modeled by

$$|D_{R_m}^t| = \begin{cases} |D_{R_m}^{t-1}| & \text{if } U_t^{R_m,j} = 0, \\ |D_{R_m}^{t-1}| + |D_j^{t-1}| p_{new} & \text{if } U_t^{R_m,j} = 1, \end{cases} \tag{6}$$

where $p_{new} = \frac{|D_j^{t-1} - D_{R_m}^{t-1}|}{|D_j^{t-1}|}$ is the proportion of new data to R_m from j. p_{new} is larger if the path of j has less overlapping with the paths of other mobile phone users.

$$|D_j^t| = \begin{cases} |D_j^{t-1}| + \delta d & \text{if } U_t^{R_m,j} = 0, \\ |D_{R_m}^{t-1}| + |D_j^{t-1}| p_{new} & \text{if } U_t^{R_m,j} = 1, \end{cases} \tag{7}$$

where δd is the average newly collected data by j in one time slot.

We let p_c be the average contact probability between R_m and j, i.e. $p_c = P[U_t^{R_m,j} = 1]$ for each time slot, where $p_c = K/H$, K is the average number of repositories on a path and H is the average path length. Figure 5 shows the average data availability, which is measured in number of grid cells varying K with $H = 1000$, $\delta d = 0.1$ and $p_{new} = 0.3$. Data availability increases quickly with K at the beginning and becomes steady

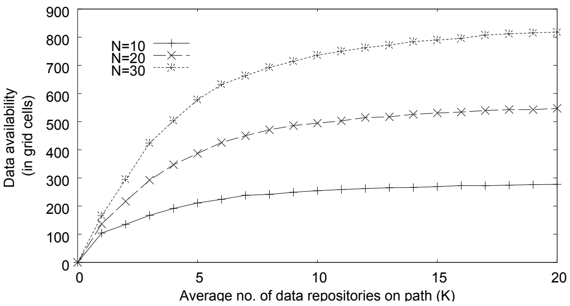

Fig. 5. Data availability varying K with $H = 1000$, $p_{new} = 0.3$, $\delta d = 0.1$, $N = 10, 20$ and 30.

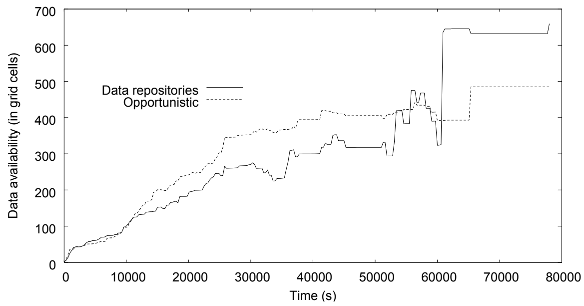

Fig. 6. The average data availability of mobile phone users in environment with high contact rates at $N = 35$, $K = 3$ and $M = 10$.

gradually. It indicates that excessive data repositories mainly bring redundant data after saturation. We plot the curves with $N = 10, 20$ and 30 respectively. The results show that higher number of users will increase the overall data availability of the system as they can collect more data from different paths.

B. Communication Overheads

When a mobile phone user reaches a data repository, they first exchange information (or data list) of their buffered data. Then, they can select and exchange data of their interests. We calculate the communication overheads, Msg_j in number of messages for information exchanges between the mobile phone j and data repositories.

$$Msg_j = 2H_j p_c, \tag{8}$$

where H_j is the path length of j and p_c is the average contact probability between j and the data repositories. $H_j p_c$ also implies the average number of contacts between j and the data repositories. Consider our deployment algorithm, given K as the minimum number of data repositories on each path, the average number of messages for information exchanges in each user, $\overline{Msg}$, is at least $2K$.

VII. Performance Evaluation

We evaluate the performance of information exchange with data repositories among the mobile phone participants based on the traces collected from NCSU [8], [9]. The human mobility traces are collected with GPS receivers carried by 35 participants at every 10 seconds. These traces are mapped into a two dimensional area and recomputed to a position at every 30 seconds by averaging three samples over that 30 second period to account for GPS errors [8]. All mobile traces start at the same time, but they may end at different times.

We divide the sensing field into 100x150 grid cells with sensors deployed at the center of each of them. We assume that the mobile phone users are interested in the sensing data of the whole sensing field. They can collect data from the sensors in the same grid cells. In addition, they can exchange information with the data repositories along their walks. We also compare our system performance with opportunistic exchange in which mobile phone users exchange information with any other users within their communication range.

A. Environment with High Contact Rates

We first evaluate the data availability and communication overheads of mobile phone users in an environment with high contact rates. We set $K = 3$ and run our deployment algorithm which results in $M = 10$ data repositories for the 35 users.

1) Data Availability: Data availability is measured by the coverage of available data received by the users. Data from the same grid cells are counted only once as the buffered data will be refreshed from time to time. Figure 6 shows that average data availability of all users measured in number of grid cells. The users start their walks at the same time in our traces, but they may finish their walks at different times depending on the paths they chose. This explains why the average data availability of the users may drop sometimes. Both the results of information exchange with data repositories and opportunistic information exchange are plotted. After the system is stabilized at 60000s, the data availability provided by data repositories is much higher than opportunistic exchange. It takes some time for our system to start up as the data repositories are empty at the beginning of the experiment. The results show that the data availability with data repositories out performs opportunistic information exchange, even though the mobile phone users have high contact rates in this environment. We also expect that the data availability of the network with data repositories will further increase as more data are accumulated in the data repositories in the long run.

2) Communication Overheads: Figure 7 shows the communication overheads for information exchange of individual mobile phone users. The number of messages in opportunistic exchange is very high with an average of 118 messages per mobile phone user. It is because the number of contacts between mobile phone users are very high in this experiment. The communication overheads of information exchange with data repositories has an average of only 34 messages per user. The results demonstrate that data repositories can reduce the communication overheads significantly without degrading the data availability.

B. Environment with Low Contact Rates

We evaluate our system performance in environment with low contact rates here. To simulate the low contact environment, we select 10 traces randomly from the original set of

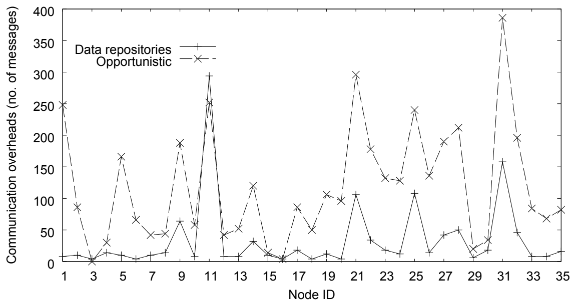

Fig. 7. The communication overheads of individual users in environment with high contact rates at $N = 35$, $K = 3$ and $M = 10$.

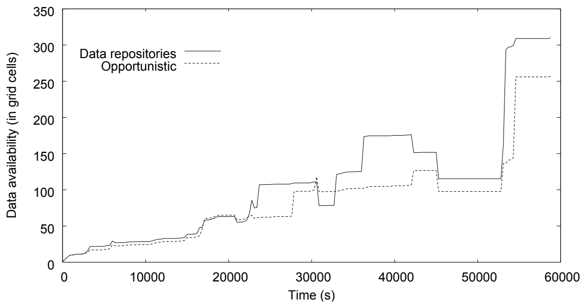

Fig. 8. The average data availability of mobile phone users in environment with low contact rates at $N = 10$, $K = 6$ and $M = 7$.

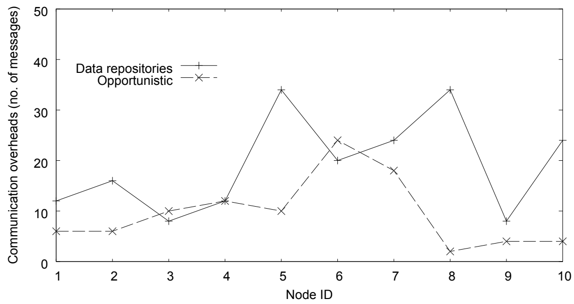

Fig. 9. The communication overheads of individual users in environment with low contact rate at $N = 10$, $K = 6$ and $M = 7$.

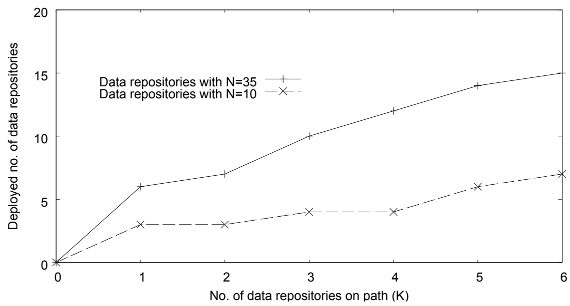

Fig. 10. Required number of data repositories for deployment with $N = 35$ and $N = 10$.

real mobile traces. We run our data repository deployment algorithm with $K = 6$ and find that $M = 7$ data repositories are required to fulfill the K-cover.

1) Data Availability: Figure 8 shows that our approach with data repositories can achieve much higher data availability than opportunistic information exchange from the very beginning. It demonstrates that data repositories can effectively improve the data availability in environment with insufficient contacts.

2) Communication Overheads: Figure 9 shows that the communication overheads of information exchange with data repositories are slightly higher than that in opportunistic information exchange. However, the number of messages in both schemes are very small with only 20 and 10 messages per user on average. The slightly higher communication overheads with data repositories are reasonable given the improved data availability for mobile phone users.

C. Impact of Number of Data Repositories on Path (K)

We repeat the experiments with the same traces as above varying the number of data repositories on path K. Figure 10 shows the required number of data repositories M for deployment in the environment with high contact rates ($N = 35$) and low contact rates ($N = 10$) respectively. We find that more data repositories are required to provide K-cover when N increases.

1) Data Availability: Figure 11 shows the data availability of the same experiments with $N = 35$ and $N = 10$ respectively. The data availability of opportunistic exchange is also plotted for comparison, which is a constant independent of the number of repositories. The data availability with repositories increases with K and becomes steady after saturation, which is consistent with our analytical results in Figure 5. There is a small drop on the curve with data repositories at $N = 35$ and $K = 4$ though, which is believed to be due to the randomness of our heuristic placement algorithm. In general, data repositories can achieve much higher data availability than opportunistic information exchange. The data availability with data repositories at $N = 10$ and $K = 6$ increases 41% compared with opportunistic exchange.

2) Communication Overheads: The corresponding communication overheads are shown in Figure 12. The communication overheads with data repositories reduces to only 28% of opportunistic exchange at $K = 3$ and $N = 35$, even giving its improved data availability as discussed in Figure 11. Similarly, the communication overheads with data repositories at $N = 10$ are reasonably low, which are around 10-25 messages. Given

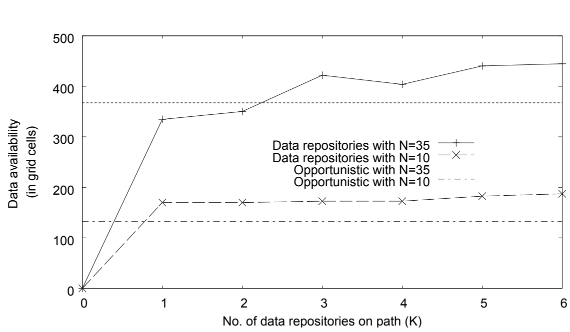

Fig. 11. The average data availability varying K with $N = 35$ and $N = 10$.

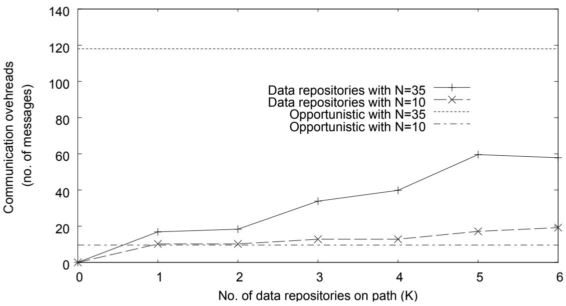

Fig. 12. The average communication overheads varying K with $N = 35$ and $N = 10$.

its greatly improved data availability (up to 41% as shown in Figure 11), the overheads with data repositories are well justified. We find that the communication overheads with data repositories increase with K quite linearly, which validates our analytical results in Section VI.B. It is interesting to see that the communication overheads with $N = 35$ is much higher than the lower bound $2K$ presented in our analysis. It is because many users visit the same repositories more than once along their paths in our traces.

VIII. Conclusions and Future Works

In this paper, we proposed OppSense, a novel data collection and information exchange framework for mobile phone users in a sensing field. It leverages data repositories to provide satisfactory data availability with minimized communication overheads for information sharing even in environments with inadequate or excessive contacts. It overcomes a major problem of existing opportunistic exchange in which the system performance depends heavily on the contact rates between the users. OppSense achieves the guaranteed system performance by exploiting the paths of mobile phone users and deploying data repositories at optimal locations. It provides assured opportunities of information exchange for all the mobile phone users with minimized data repositories. We formulated the placement of data repositories as a connected K-coverage problem and presented a heuristic algorithm for this NP-hard problem. We evaluated our proposed framework and algorithm using real mobile traces and compared its performance with existing opportunistic exchange mechanism. Extensive simulations demonstrated that data repositories can effectively increase the data availability 41% in low contact environment and significantly reduce the communication overheads to only 28% in high contact environment.

In the future, we are interested in having hybrid schemes using both data repositories and mobile-to-mobile information exchange. We also plan to build a real application and study the human and social behaviors in practice.

Acknowledgement

This work was supported by the VINNOVA VINNMER program funded by the Swedish Governmental Agency for Innovation Systems.

References

[1] I. F. Akyildiz, W. Su, and T. Sandarasubramaniam, “Wireless sensor networks: a survey,” *Computer Networks*, vol. 38, no. 5, pp. 393–422, 2002.

[2] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, “Wireless sensor networks for habitat monitoring,” in *Proc. of ACM International Workshop on Wireless Sensor Networks and Applications (WSNA)*, 2002, pp. 88–97.

[3] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A. Peterson, “People-centric urban sensing,” in *in Proc. of International Workshop on Wireless Internet (WICON)*, Aug 2006, pp. 18–31.

[4] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, “Participatory sensing,” in *Workshop on World-Sensor-Web at SenSys 2006*, Oct 2006.

[5] Z. Ruan, E. C.-H. Ngai, and J. Liu, “Wireless sensor network deployment in mobile phones assisted environment,” in *Proc. of IEEE IWQoS*, Jun 2010.

[6] M. Conti, F. Delmastro, and A. Passarella, “Social-aware content sharing in opportunistic networks,” in *Proc. of SECON*, Jun 2009, pp. 1–3.

[7] C. Boldrini, M.Conti, F.Delmastro, and A.Passarella, “Context- and social-aware middleware for opportunistic networks,” *Journal of Network and Computer Applications*, 2010.

[8] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, “On the levy-walk nature of human mobility,” in *Proc. of INFOCOM*, Apr 2008, pp. 924–932.

[9] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, “Slaw: A mobility model for human walks,” in *Proc. of INFOCOM*, Apr 2009, pp. 855–863.

[10] R. Shah, S. Roy, S. Jain, and W. Brunette, “Data mules: Modeling a three-tier architecture for sparse sensor networks,” in *Proc. of the IEEE Workshop on Senosr Network Protocols and Applications (SNPA)*, May 2003, pp. 30 – 41.

[11] A. Chakrabarti, A. Sabharwal, and B. Aazhang, “Using predictable observer mobility for power efficient design of sensor networks,” in *Proc. of IPSN*, Apr 2003, pp. 129–145.

[12] A. Kansal, A. Somasundara, D. Jea, M. Srivastava, and D. Estrin, “Intelligent fluid infrastructure for embedded networks,” in *Proc. of ACM MobiSys*, Jun 2004, pp. 111–124.

[13] Z. M. Wang, S. Basagni, E. Melachrinoudis, and C. Petrioli, “Exploiting sink mobility for maximizing sensor networks lifetime,” in *Proc. of Hawaii International Conference on System Sciences (HICSS)*, 2005.

[14] Y. Gu, D. Bozdag, E. Ekici, F. Ozguner, and C.-G. Lee, “Partitioning-based mobile element scheduling in wireless sensor networks,” in *Proc. of SECON*, Sep 2005, pp. 386–395.

[15] N. Bisnik, A. Abouzeid, and V. Isler, “Stochastic event capture using mobile sensors subject to a quality metric,” in *Proc. of ACM MobiCom*, Sep 2006, pp. 98 – 109.

[16] X. Xu, J. Luo, and Q. Zhang, “Delay tolerant event collection in sensor networks with mobile sink,” in *Proc. of INFOCOM*, Mar 2010, pp. 1–9.

[17] L. Pelusi, A. Passarella, and M. Conti, “Opportunistic networking: data forwarding in disconnected mobile ad hoc networks,” *IEEE Communications Magazine*, vol. 44, pp. 134–141, Nov 2006.

[18] Z. Zhang, “Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: Overview and challenges,” *IEEE Communications Surveys*, vol. 8, no. 1, 2006.

[19] J. Sushant, K. Fall, and R. Patra, “Routing in a delay tolerant network,” in *Proc. of SIGCOMM*, Aug 2004, pp. 145–158.

[20] A. Vahdat and D. Becker, “Epidemic routing for partially connected ad hoc networks,” *Tech. Rep. CS-2000-06, Department of Computer Science, Duke University, Durham, NC*, 2000.

[21] B. Burns, O. Brock, and B. N. Levine, “MV routing and capacity building in disruption tolerant network,” in *Proc. of INFOCOM*, Mar 2005, pp. 398–408.

[22] A. Lindgren, A. Doria, and O. Schelén, “Probabilistic routing in intermittently connected networks,” *Lecture Notes in Computer Science*, vol. 3126, pp. 239–254, Jan 2004.

[23] W. Zhao, M. Ammar, and E. Zegura, “A message ferrying approach for data delivery in sparse mobile ad hoc networks,” in *Proc. of ACM MobiHoc*, Mar 2005, pp. 187–198.

[24] E. Yoneki, P. Hui, S. Chan, and J. Crowcroft, “A socio-aware overlay for publish/subscribe communication in delay tolerant networks,” in *Proc. of ACM MSWiM*, Oct 2007, pp. 225–234.

[25] E. Jaho and I. Stavrakakis, “Joint interest- and locality-aware content dissemination in social networks,” in *Proc. of Conference on Wireless On demand Network Systems and Services (WONS)*, Feb 2009, pp. 173–180.

[26] D. Goodman, J. Borras, N. B. Mandayam, and R. D. Yates, “Infostations: a new system model for data and messaging services,” in *IEEE Vehicular Technology Conference*, May 1997.

[27] F. R. Dogar, A. Phanishayee, H. Pucha, O. Ruwase, and D. G. Andersen, “Ditto - a system for opportunistic caching in multi-hop wireless networks,” in *MobiCom*, Sep 2008.

[28] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, “Impact of human mobility on opportunistic forwarding algorithms,” *IEEE Trans. on Mobile Computing*, vol. 6, no. 6, pp. 606–620, 2007.

[29] H. Gupta, Z. Zhou, S. R. Das, and Q. Gu, “Connected sensor cover: self-organization of sensor networks for efficient query execution,” *IEEE/ACM Trans. Netw.*, vol. 14, no. 1, pp. 55–67, 2006.

[30] V. S. A. Kumar, S. Arya, and H. Ramesh, “Hardness of set cover with intersection 1,” in *Automata, Languages and Programming*, 2000, pp. 624–635.