

Research Article

Share the Crowdsensing Data with Local Crowd by V2V Communications

Chao Song,^{1,2} Ming Liu,^{1,2} and Xili Dai^{1,2}

¹School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

²Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, China

Correspondence should be addressed to Chao Song; chaosong@uestc.edu.cn

Received 12 December 2015; Revised 25 February 2016; Accepted 31 March 2016

Academic Editor: Qirong Ho

Copyright © 2016 Chao Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With an increase in the number of mobile applications, the development of mobile crowdsensing systems has recently attracted significant attention from both academic researchers and industries. In mobile crowdsensing system, the remote cloud (or back-end server) harvests all the crowdsensing data from the mobile devices, and the crowdsensing data can be uploaded immediately via 3G/4G. To reduce the cost and energy consumption, many academic researchers and industries investigate the way of mobile data offloading. Due to the sparse distribution of the WiFi APs, offloading the crowdsensing data is often delayed. In this paper, compared with offloading data via WiFi APs, we investigate the communication and sharing of crowdsensing data by vehicles near the event (such as a pothole on the road), termed as a local crowd. In such crowd, a vehicle can transmit the data to each other by vehicle-to-vehicle (V2V) communication. The crowd-based approach has a lower delay than the offloading-based approach, by considering the quality of truth discovery. We define a utility function related to the crowdsensing data shared by the local crowd in order to quantify the trade-off between the quality of the truth discovery and the user satisfaction. Our extensional simulations verify the effectiveness of our proposed schemes.

1. Introduction

Over 6.8 billion mobile phones were in use all over the world in 2013 [1]. With an increase in the number of mobile applications, many outdoor mobile applications are getting a lot of attention from both academic researchers and industries [2], such as Waze and pothole detection [3]. The ubiquity of smartphones has led to the emergence of mobile crowdsensing (MCS) tasks such as the detection of spatial events when smartphone users move around in their daily lives [4].

The cloud-based architecture has been widely used in the mobile crowdsensing system. A mobile device senses an event and then generates the crowdsensing data to describe it, termed as *report*. The mobile device communicates with the remote cloud (or back-end server) to upload its report. The back-end server harvests all the reports from these mobile devices and aggregates them to discover the truth of the event. Finally, the back-end server publishes the feedback result to

the querying users, who are interested in the event. The communication between the mobile devices and the back-end server can be immediate via the cellular network (3G/4G).

However, the problem of crowdsensing with cellular network is the increase in traffic demand and high energy consumption. Furthermore, the tasks of crowdsensing, which include gathering raw data and querying the services, can severely increase the overhead of the cloud, such as the computation, the storage, and the bandwidth. For the users, mobile data offloading through WiFi APs has demonstrated its feasibility in reducing the data burden on the cellular networks. More recently, delayed offloading has been proposed [5]: if there is currently no WiFi availability, (some) traffic can be delayed instead of being sent/received immediately over the cellular interface. Although the delayed offloading can reduce the cost of mobile data communication, this approach obviously increases the data delivery delay, so as to decrease the user satisfaction.

In this paper, we investigate the communication and sharing of the mobile crowdsensing data under the traditional cloud-based architecture. We propose a distributed architecture, which does not require communicating with the remote cloud. The crowdsensing data is disseminated to all the vehicles near the event in a predefined range, and we term these vehicles as the *local crowd* of this event. In such crowd, a vehicle can transmit the data to each other by vehicle-to-vehicle (V2V) communication. Like the remote cloud, the local crowd harvests the crowdsensing data of this event and then aggregates them to discover the truth of the event. Finally, the crowd sends the feedback results to the querying users, who move in this crowd. Thus, the communication and sharing of the crowdsensing data all happen in the local crowd, in order to reduce the data delivery delay. We define a utility function related to the crowdsensing data shared by the local crowd, in order to quantify the trade-off between the quality of the truth discovery and the user satisfaction. Compared with the cloud-based approach, we model the quality of the truth discovery by Kullback-Leibler divergence (or relative entropy). Our extensional simulations verify the effectiveness of our proposed schemes.

The remainder of this paper is organized as follows: Section 2 surveys the related work; Section 3 introduces our crowdsensing-based system and discusses the different approaches for communicating and sharing the crowdsensing data; in Section 4, we model the problem and propose an algorithm to optimize it; Section 5 evaluates the performance of the proposed approach; and the last section concludes this paper.

2. Related Work

We present related work in the following two parts, which are mobile crowdsensing and mobile data offloading.

2.1. Mobile Crowdsensing. Mobile phone sensing is a paradigm which takes advantage of the pervasive smartphones to collect and analyze data beyond the scale of what was previously possible. Yang et al. in [6] investigate novel sensors integrated in modern mobile phones and leverage user motions to construct the radio map of a floor plan, which was previously obtained only by site survey. Zhou et al. in [7] investigate the application of the prediction for the bus arrival time. They do not require the absolute physical location reference, and they mainly wardrive the bus routes and record the sequences of observed cell-tower IDs, which reduces the initial construction overhead. Yang et al. in [8] design incentive schemes for mobile phone sensing, with two system models: the platform-centric model, where the platform provides a reward shared by participating users, and the user-centric model, where users have more control over the payment they will receive. He et al. in [9] investigate the optimal task allocation and show that the allocation problem is NP hard. They also discuss how to decide fair prices of sensing tasks to provide incentives, since mobile users tend to decline the tasks with low incentives.

2.2. Mobile Data Offloading. This increase in traffic demand is overloading cellular networks, forcing them to operate

close to (and often beyond) their capacity limits [5]. A more cost-effective way to cope with the problem of highly congested mobile networks is by offloading some of the traffic through Femtocells and the use of WiFi. There exist two types of WiFi offloading. The usual way of offloading is on-the-spot offloading: when there is WiFi available, all traffic is sent over the WiFi network; otherwise, all traffic is sent over the cellular interface. More recently, delayed offloading has been proposed: if there is currently no WiFi availability, (some) traffic can be delayed instead of being sent/received immediately over the cellular interface. In the simplest case, traffic is delayed until WiFi connectivity becomes available. A more interesting case is when the user (or the device on her behalf) can choose a deadline (e.g., per application and per file). If up to that point no AP is detected, the data are transmitted through the cellular network [10, 11]. The authors in [12] define a utility function related to delayed offloading to quantitatively describe the trade-offs between user satisfaction in terms of the price and the experienced delay of waiting for WiFi connectivity. Ristanovic et al. in [13] propose two algorithms for delay-tolerant offloading of bulky, socially recommended content from 3G networks. The first one (called MixZones) uses opportunistic, ad hoc transfers between the users, and the second one (called HotZones) exploits delay tolerance and tries to download contents when users are close to WiFi access points.

3. Problem Statement

In this section, we first take our crowdsensing-based system, called Follow Us (FU), as an example to demonstrate the typical cloud-based architecture. The system FU is a road traffic condition monitoring and alerting application. Then, we discuss the communication and sharing of the crowdsensing data with the local crowd, which is a distributed approach.

3.1. Remote Cloud. For processing the crowdsensing data, one effective approach is to utilize a centralized architecture. A typical architecture of crowdsensing-based system consists of a centralized back-end server and a collection of mobile users. The mobile users have two responsibilities: (1) *probing user*, who uses mobile phones as well as the sensors to sense and report the events to the back-end server, and (2) *querying user*, who queries the information of the nearby events. The back-end server is responsible for collecting the reports of the events from the probing vehicles and intellectually aggregates such information. The smartphones in the system are assumed to be cooperative, which belong to or are affiliated to the system, willing to take sensing tasks and provide sensing services to the system. The issue of participation incentive [14, 15] of a rational or even strategic smartphone user is out of the scope of the paper.

In order to demonstrate the communication and sharing in crowdsensing-based system, we take a system as an example. One attractive example is a traffic monitoring application that uses a number of active probing vehicles to sense the road condition. To monitor the road condition and alert the anomalies, we have developed a crowdsensing-based system, called Follow Us (FU). The basic idea of FU is that the vehicles

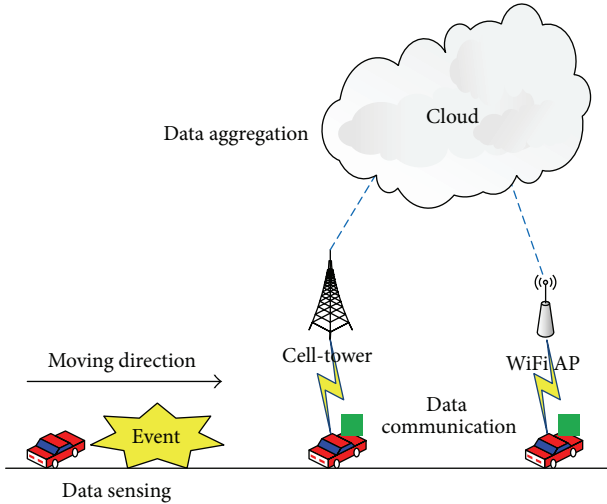


FIGURE 1: The cloud-based architecture of a mobile crowdsensing application.

moving ahead can report the road conditions to those behind. As shown in Figure 1, this application includes three primary parts as follows:

- (i) Data sensing: the smartphones in the probing vehicles sense the events on the road. As an example in Figure 1, when a vehicle meets an obstacle on the road, the driver drives the vehicle to avoid the obstacle. Thus, the smartphone mounted in the vehicle senses the anomaly of the motion of the vehicle by its accelerometer.
- (ii) Data communication: the sensing data are uploaded to the back-end server, and the feedback results will be downloaded from the server. In Figure 1, the smartphone on the vehicle will upload the report of this anomaly to the back-end server when it has an opportunity of connection to Internet (e.g., 3G/4G or WiFi AP).
- (iii) Data aggregation: the back-end server harvests and aggregates all the uploaded sensing data to obtain the feedback results to the querying users. The back-end server harvests the reports from all the probing vehicles and aggregates them by the algorithm of truth discovery. As a result, the back-end server publishes the report to the vehicles heading for the place of the anomaly.

Figure 2 is an example of sensing data in FU. When the vehicle drives through the road with speed bumps, it will be shaken. The smartphone senses the shake by the accelerometers. The FU system in the smartphone records the changes of the accelerations and maps them into three accelerations: total acceleration, horizontal acceleration, and vertical acceleration. In order to describe the intensity of the accelerations, FU generates a report of this sensing event as follows:

$$r_i = (l_i, t_i, a_i^t, a_i^h, a_i^v), \quad (1)$$

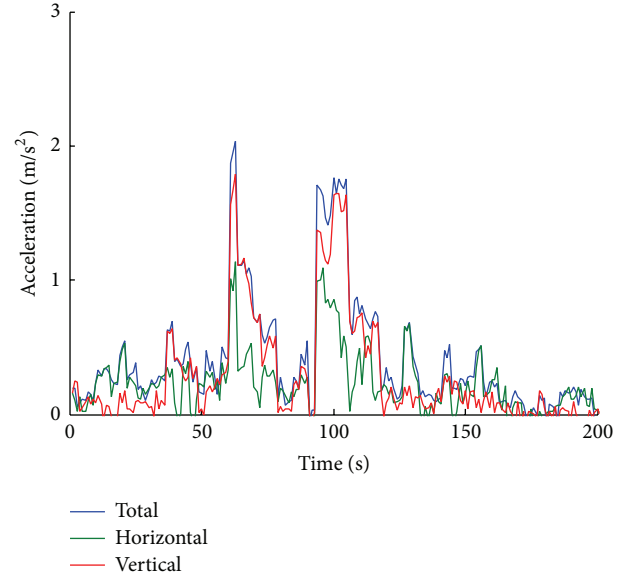


FIGURE 2: An example of sensing event.

where l_i denotes the location of the sensing event e_i , which includes the latitude, longitude, and altitude of its center by GPS. t_i denotes the sensing time of this event e_i . a_i^t , a_i^h , and a_i^v denote the total acceleration, horizontal acceleration, and vertical acceleration, respectively. They contain the mean and the standard variation of the accelerations during the sensing window.

For the centralized data aggregation, there are two possible ways for data communication as shown in Figure 1: (1) the vehicles can immediately communicate the data via cellular network (3G/4G); otherwise, (2) the vehicles can offload the data via the WiFi APs. The problem of crowdsensing with cellular network is the increase in traffic demand and high energy consumption. Furthermore, the tasks of crowdsensing, which include gathering raw data and querying the services, can severely increase the overhead of the cloud, such as the computation, the storage, and the bandwidth. For vehicular users, mobile data offloading through WiFi has demonstrated its feasibility in reducing the data burden on the cellular networks. More recently, delayed offloading has been proposed [5]: if there is currently no WiFi availability, the traffic can be delayed instead of being sent/received immediately over the cellular interface. Although the delayed offloading can reduce the costs of mobile data communication, this approach obviously increases the data delivery delay.

Definition 1 (delivery delay of a single sensing data). We define the delivery delay of a single sensing data (denoted by d) in such crowdsensing-based system as the duration from the time when a probing vehicle senses the data and generates a report to the time when the feedback result from the SAME report is received by another querying vehicle.

Thus, in the cloud-based architecture, the data delivery includes three parts, which are the data sensing, the data

communication, and the data aggregation. The data delivery delay with remote cloud (denoted by D_{RC}) via immediate cellular network can be calculated as follows:

$$D_{RC} = D_{sen} + D_{cell}^{up} + D_{agg} + D_{cell}^{down}, \quad (2)$$

where D_{sen} denotes the delay of sensing data, which includes the delays of sensing the event and generating its report. D_{agg} denotes the delay of data aggregation in the remote cloud. D_{cell}^{up} and D_{cell}^{down} denote the delay of uploading to the remote cloud via cellular network and that of downloading from the remote cloud via cellular network, respectively.

By way of delayed data offloading, the vehicle will carry the sensing data to move until it meets with an AP. The data delivery delay by mobile data offloading can be calculated as follows:

$$D_{RC} = D_{sen} + D_{carry} + D_{AP}^{up} + D_{agg} + D_{AP}^{down} + D_{wait}, \quad (3)$$

where D_{carry} denotes the delay of carrying the sensing data to an AP and D_{wait} denotes the delay of waiting for a querying user to download the feedback result from an AP. D_{AP}^{up} and D_{AP}^{down} denote the delay of uploading to the remote cloud via AP and that of downloading from the remote cloud via AP, respectively. Due to the delays of D_{carry} and D_{wait} , the data delivery delay via mobile data offloading is much longer than that via the cellular network.

3.2. Local Crowd. Distinct from the centralized data aggregation, the authors in [16] introduce a distributed road information sharing architecture with rumor and report. The basic idea behind this mechanism is that each vehicle that hears a rumor about the event maintains a time decaying belief about it. Rumors from multiple vehicles are combined additively until they exceed a prescribed threshold, at which point they are converted to confirm event reports. This threshold is committed to effect a desired trade-off in information reliability, between the rate of false negatives and the rate of false positives. Both rumors and reports are distributed through the network in an epidemic gossip spread fashion. As time goes, the belief value of the rumor is changed following a predetermined *decay function* in order to discount the aged information. Although the decay function may be any nonincreasing function of the elapsed time from the creation of the rumor, we focus on the exponentially decreasing function.

In this paper, we investigate the distributed approach for the crowdsensing data. We suppose that the vehicles can communicate with each other by vehicle-to-vehicle (V2V) communication, such as vehicular ad hoc networks (VANET) [17] or Device-to-Device (D2D) communication [18]. The sensing data are aggregated by the vehicles near the related event. We define these vehicles near the event in a predefined range (denoted by r) as the *local crowd* of this event. Consider

$$c_i = \{v \mid \|v.l - e.l\| \leq r, v \in \mathcal{V}\}, \quad (4)$$

where $v.l$ and $e.l$ denote the location of the vehicle and the event, respectively. \mathcal{V} denotes the set of the vehicles. The size

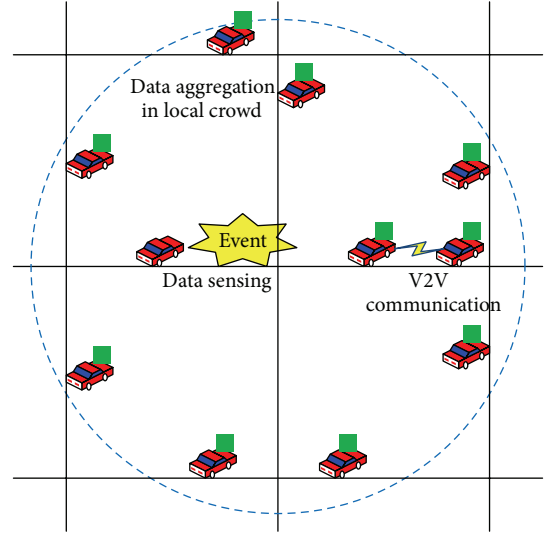


FIGURE 3: Crowdsensing data with local crowd.

of the local crowd is related to the requirement of notification for the feedback results.

As shown in Figure 3, a moving vehicle senses an event on the road, and then it generates a report to describe the event. When the vehicle meets another one in the local crowd of this event, it will send a copy of the report to another one. Thus, all the reports of this event will be disseminated to all the vehicles in the local crowd of this event by the way of epidemic routing [19]. When a vehicle moves out of the crowd, it will handle all the reports of the event in two different ways, according to the traffic density: (1) it will drop all the reports of the event under high traffic density; (2) under low traffic density, it cannot transmit the reports of the event with other vehicles, but the reports in its buffer will not be dropped. The vehicles in the crowd maintain the crowdsensing data from the event and calculate the feedback results by the algorithm of truth discovery. When a querying vehicle moves into this local crowd and meets another vehicle, it will receive a feedback result of this event. Therefore, the crowdsensing data are communicated and shared in this local crowd.

In such crowd-based architecture, the data delivery delay with local crowd (denoted by D_{LC}) can be calculated as follows:

$$D_{LC} = D_{sen} + D_{diss} + D_{agg} + D_{back}, \quad (5)$$

where D_{sen} denotes the delay of sensing data and D_{agg} denotes the delay of data aggregation in the vehicles of the local crowd. D_{diss} denotes the average delay of disseminating the reports to the vehicles in the crowd. D_{back} denotes the delay from the time when the feedback result is calculated to the time when the querying vehicle receives the result.

3.3. Comparison. We have discussed three different approaches for the crowdsensing-based system: (1) to immediately communicate with remote cloud via cellular network (such as 3G/4G); (2) to delay offloading to the remote cloud

TABLE 1: Comparison among the three approaches.

	Delay	Cost	Energy	Quality
3G/4G to remote cloud	Low	High	High	High
Offload to remote cloud	High	Low	Low	High
V2V to local crowd	Median	Low	Low	Median

via roadside APs; (3) to share the crowdsensing data in a local crowd by V2V communications. Table 1 shows the comparison among the three approaches for crowdsensing data by four different metrics, which are data delivery delay, cost, energy consumption, and quality of truth discovery.

3.3.1. 3G/4G to Remote Cloud. The crowdsensing data are aggregated by the back-end server (or cloud) and are immediately communicated by the cellular networks (3G/4G). The data delivery delay of this approach is the lowest. It consumes the highest cost and energy for the users, which has been discussed in [5]. The back-end server can harvest all the crowdsensing data, so the quality of the feedback results is high.

3.3.2. Offload to Remote Cloud. Due to the sparse distribution of the access points, the crowdsensing data is delayed in offloading to the back-end server (or cloud). Thus, the data delivery delay of this approach is the highest. It consumes the lowest cost and energy for the users, which has been discussed in [5]. Like the 3G/4G-based approach, the quality of the feedback results is high.

3.3.3. V2V to Local Crowd. The crowdsensing data are communicated and shared in this local crowd by V2V communication, such as MANET or Device-to-Device (D2D) communication. The data delivery delay is less than that of offloading-based approach under a high traffic density, which has also been discussed in [13]. It consumes much less cost and energy than 3G/4G-based approach. The local crowd may harvest part of the crowdsensing data; the quality of the feedback results is not higher than 3G/4G-based and offloading-based approaches, and we will discuss it in the next section.

4. Truth Discovery in Local Crowd

In this section, we analyze the performance of the scheme with the local crowd. First, we discuss the trade-off between the quality of the truth discovery and the user satisfaction in the local crowd. Then, we define a utility function related to the crowdsensing data shared by the local crowd, by considering both of the quality of the truth discovery and the user satisfaction. Compared with the cloud-based approach, we model the quality of the truth discovery by Kullback-Leibler divergence (or relative entropy). Last, we formulate the sharing crowdsensing data in local crowd as an optimization problem. The notations used in this paper are given in Notations Section.

4.1. Trade-Off between the Quality and the Satisfaction. As our discussion in the previous section, in such crowdsensing-based system, when a vehicle moves through an object (e.g., an obstacle) or an event (e.g., an accident), the smartphone in it can sense it. Then, the mobile device generates a report

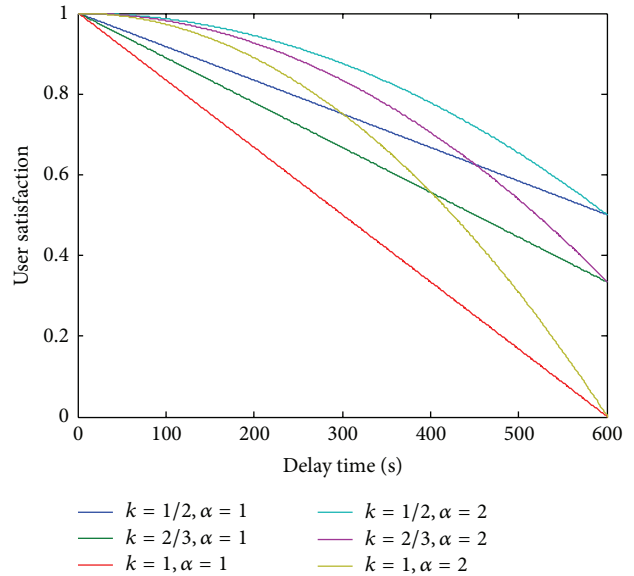


FIGURE 4: User satisfaction.

to describe the anomaly and disseminates the report to the vehicles in the local crowd. Thus, all the reports of the event will be shared by all the vehicles in the local crowd.

With the increase of time, the vehicles in the local crowd will receive more and more reports about the event, so the quality of the truth discovery is getting better. Meanwhile, the mobile users' satisfaction will be reduced. Thus, in the crowdsensing data shared by the local crowd, there is a trade-off between the quality of the truth discovery and the user satisfaction.

Definition 2 (feedback delay for a querying user). We define the feedback delay for a querying user (denoted by t) in such crowdsensing-based system as the duration from the time when the first probing user senses the event and generates a report, to the time when any feedback results are received by the querying users.

Here, we consider two metrics to evaluate the performance of the scheme with the local crowd, which are the quality of the truth discovery and the user satisfaction. Inspired by the definition of the utility function in [12], we define a utility function related to the crowdsensing data shared by the local crowd, in order to quantify the trade-off between the quality of the truth discovery and the user satisfaction. The utility function $U(t)$ of delay period t starting from the time of event detection by the probing user ($t = 0$) until the time of the feedback to the querying user is as follows:

$$U(t) = Q(t) \cdot S(t), \tag{6}$$

where $Q(t)$ denotes the quality of the truth discovery by the local crowd, compared with the cloud-based approach. $S(t)$ is a function measuring the degree of user satisfaction based on the delay time t .

Thus, the higher quality of the truth discovery and the higher user satisfaction can increase the utility of crowdsensing-based system with local crowd. Next, we will discuss

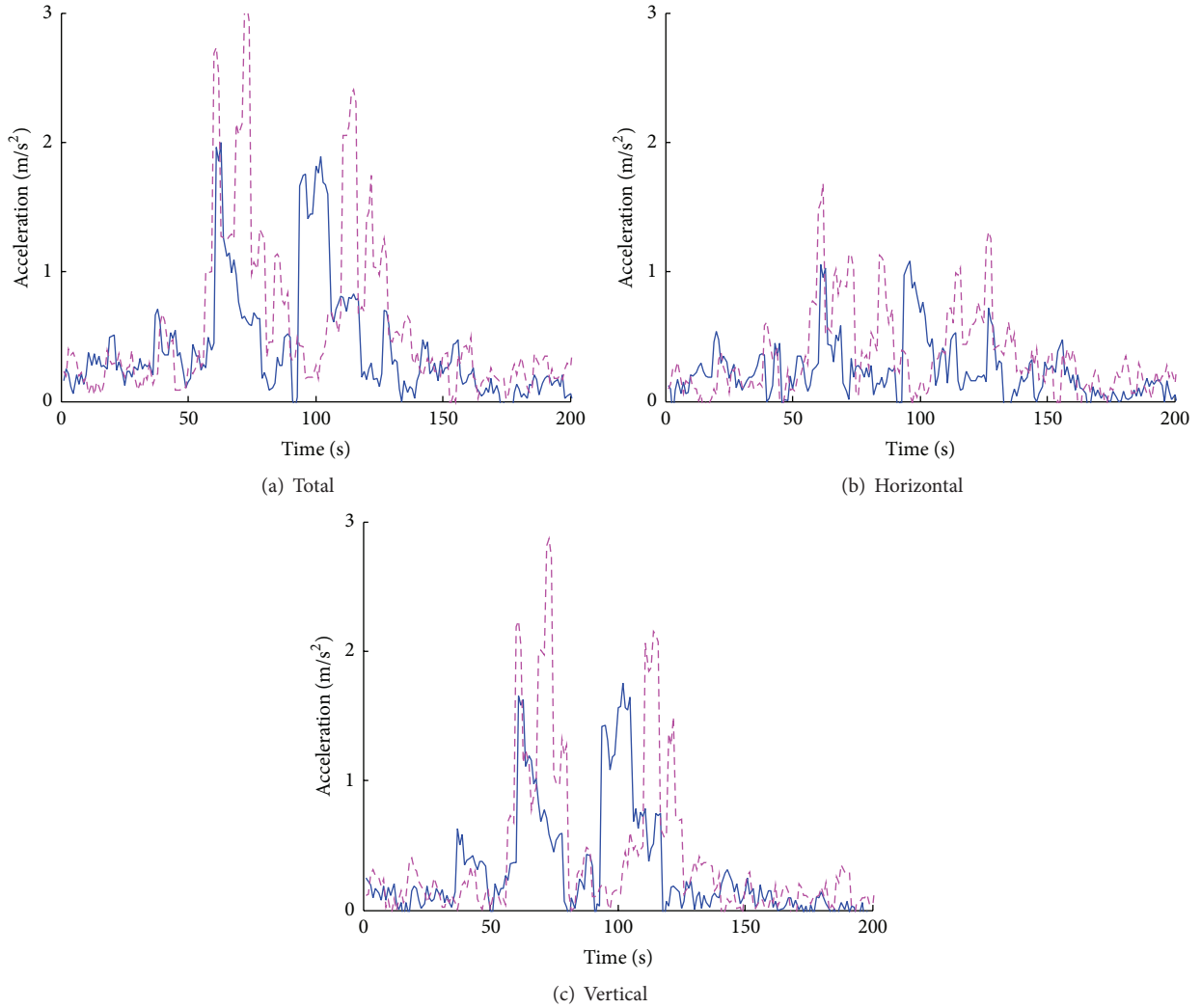


FIGURE 5: Compare the sensing data for the same event.

our models of quality of the truth discovery and user satisfaction, respectively.

4.2. User Satisfaction. With the increase of delay, the mobile users become impatient and hence their satisfaction will be greatly reduced [5]. In (6), $S(t)$ is a general definition of the user's perception of the delay of crowdsensing. We assume that the querying user will usually become more and more impatient while waiting for these feedback results. For simplicity without losing generality, we define the satisfaction function as follows:

$$S(t) = \begin{cases} 1 - k \cdot \left(\frac{t}{T_{\max}}\right)^\alpha, & t \leq T_{\max}, \\ 0, & t > T_{\max}, \end{cases} \quad (7)$$

where T_{\max} is the maximum delay tolerance of the user with respect to the requested content and is used to normalize the function. k is the decreasing rate of user satisfaction as time

elapses, to avoid a zero utility value at $t = T_{\max}$, and α is a decay factor for the delay time t .

Hence, the function $S(t)$ for the user would be a monotonically decreasing function between 0 and 1, after T_{\max} , meaning that the reception afterwards is useless. Figure 4 shows the user satisfaction as a function of the delay time, where the parameter T_{\max} is set as 10 minutes. In this paper, the parameter k is set as $1/2$, and α is set as 1.

4.3. Quality of Truth Discovery. The quality of the truth discovery is related to the number of crowdsensing data. We take the centralized approach as the benchmark, because the back-end server can harvest all the crowdsensing data. Thus, the quality of the truth discovery by the local crowd is evaluated by the comparison with the centralized approach.

However, the crowdsensing data are stochastic. The same vehicle meets the same event at different times; the sensing data are different. We do an experiment where the identical vehicle moves through the same dump twice and record the changes of its accelerations. Figure 5 shows the results to compare two tests, which include the total acceleration,

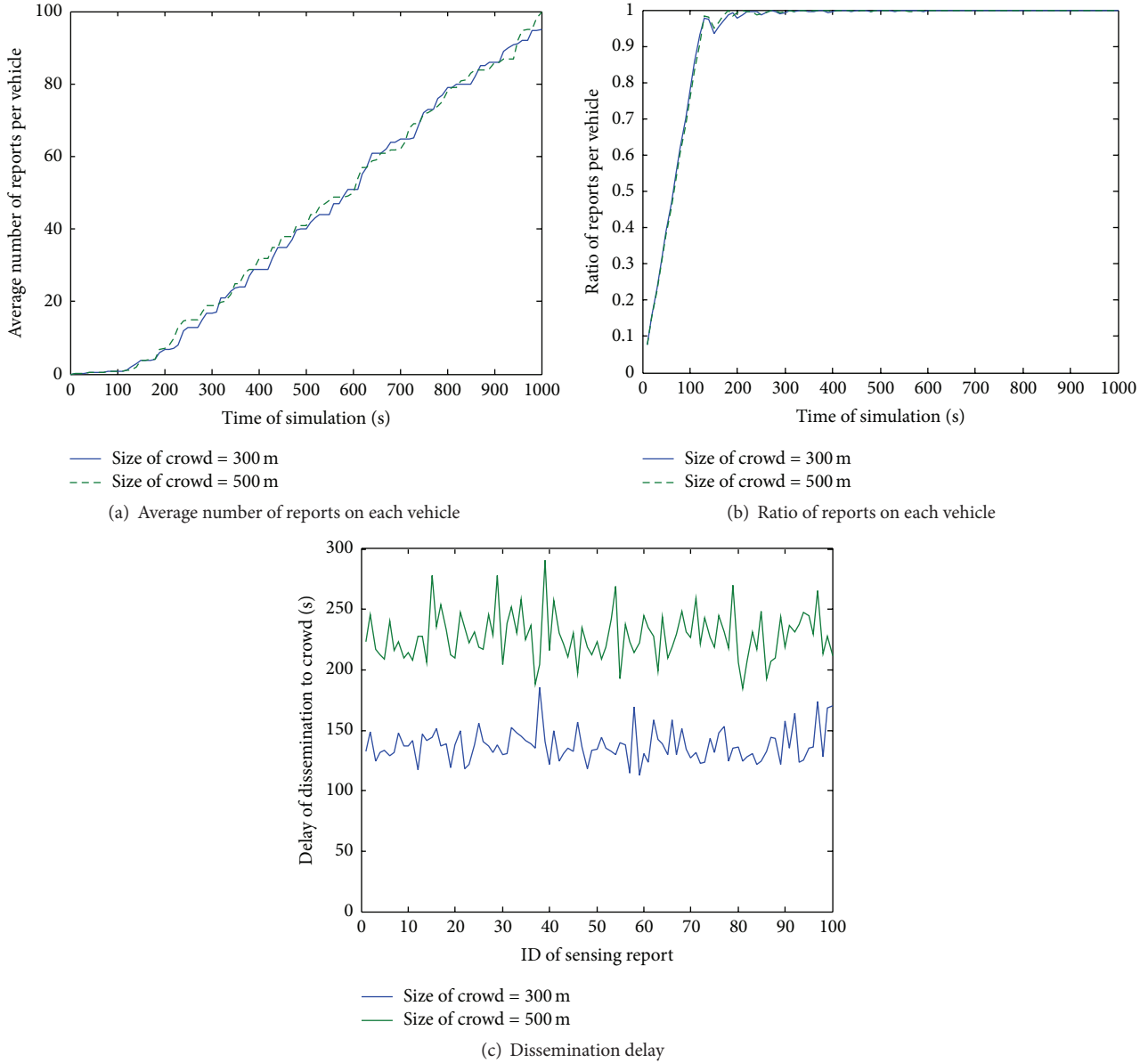


FIGURE 6: Dissemination of the crowdsensing data in a local crowd.

horizontal acceleration, and vertical acceleration. We notice that the results of the two tests are different. The uncertainty of the crowdsensing data could be affected by many factors, such as the mobile devices, the drivers, the vehicle, and the environment.

Due to the uncertainty of the crowdsensing data, we utilize the Kullback-Leibler divergence (or relative entropy) to evaluate the quality of the local crowd. In probability theory and information theory, the Kullback-Leibler divergence (also information divergence, relative entropy, or KL divergence) is a nonsymmetric measure of the difference between two probability distributions [20].

Let $R(t)$ denote the discrete probability distributions of the crowdsensing data in the remote cloud at time t , and let $L(t)$ denote the discrete probability distributions of the crowdsensing data in the local crowd at time t . Thus,

the quality of the truth discovery by the local crowd is evaluated by the Kullback-Leibler divergence of R from L as follows:

$$\begin{aligned} Q(t) &= Q_{\max} - D_{\text{KL}}(L(t) \parallel R(t)) \\ &= Q_{\max} - \sum_i L(t, i) \ln \frac{L(t, i)}{R(t, i)}, \end{aligned} \quad (8)$$

where Q_{\max} denotes the maximal quality. D_{KL} denotes the function of the Kullback-Leibler divergence, and it is equal to 0 when L and R have the same distribution. $L(t, i)$ denotes the probability of the crowdsensing data by the local crowd during the period t , whose value is equal to i . Likewise, $R(t, i)$ denotes the probability of the crowdsensing data by the remote cloud during the period t , whose value is equal to i . Thus, with the growth of time, the vehicles in the local crowd

TABLE 2: Parameters of simulation.

Parameter	Value
Size of scenario	4000 m × 3000 m
Traffic density (k)	50 or 10 vehicles per km
Average speed of the vehicles (v)	40 km/h
Size of the local crowd	500 m
Transmission range of V2V communication	50 m

will receive more and more reports about the event, so the quality of the truth discovery is getting better.

We use our customized simulator to evaluate the performance of communication and sharing of the crowdsensing data with the local crowd. The scenario of our simulation is 4000 m × 3000 m. The traffic density (k) is defined as the number of vehicles per unit length of the roadway. The average speed of the vehicles is 40 km/h. The transmission range of V2V communication is 50 m. All the parameters used in our simulation are listed in Table 2.

We simulate the dissemination of the crowdsensing data in the local crowd. During the period of the simulation, more and more vehicles sense the event and disseminate the reports to the crowd. Thus, the vehicles in the crowd will receive more and more reports, to improve the quality of the truth discovery. Figure 6(a) shows the average number of the reports received by the vehicles in the crowd. We notice that the number of the reports received by the vehicles in the crowd is increasing, because more vehicles sense the event and disseminate to the crowd. The dissemination under the smaller size of the crowd is faster than that under the bigger crowd. Figure 6(b) indicates the ratio between the average number of reports received by each vehicle and the total number of the reports. We find that the ratio is sharply increasing before about 150 seconds, and then remains at about 1.

When a probing vehicle meets and senses an abnormal event, it will generate a sensing report and disseminate it to all the vehicles in the crowd of this event. Thus, we evaluate the delay of this dissemination from the first vehicle to all the vehicles in the crowd. Figure 6(c) shows the delay of the dissemination to the crowd. The IDs of the sensing reports denote the different reports generated by separate vehicles. Due to the mobility of the vehicles, the delays of the different reports are varied. We notice that the smaller size of the crowd (300 m) has lower delay than the bigger size (500 m).

4.4. Optimization Problem. The feedback delay for a querying user (t) should be satisfied by the condition that the utility is no less than the predefined threshold δ , as follows:

$$U(t) \geq \delta. \quad (9)$$

Depending on the demand of the quality (denoted by Q_D), we formulate the optimal feedback time for a querying user (t) with the maximal utility as an optimization problem as follows:

$$\begin{aligned} \max_t \quad & U(t) = Q(t) \cdot S(t) \\ \text{s.t.} \quad & t \leq T_{\max} \\ & Q(t) \geq Q_D \\ & U(t) \geq \delta. \end{aligned} \quad (10)$$

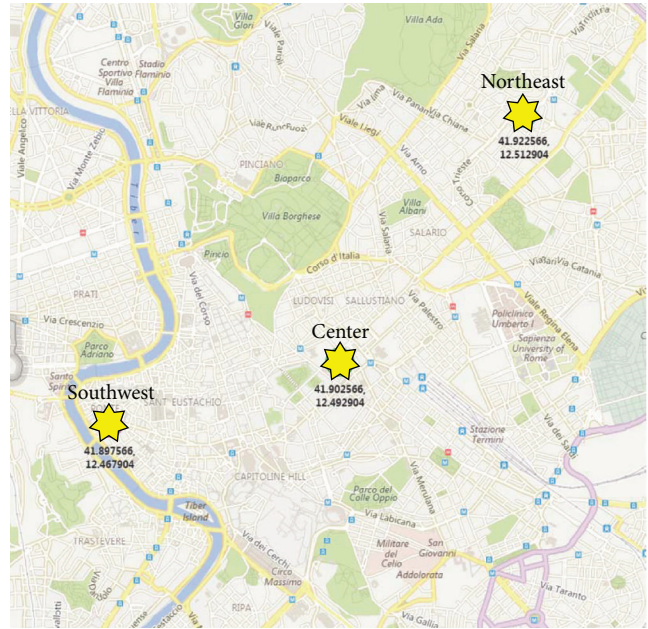


FIGURE 7: The scenario of Roma taxi.

In such optimization problem, $Q(t)$ can be estimated by the historical statistics with the help of (8). The system harvests the local crowd information and the remote cloud information and calculates $Q(t)$ as the historical records. For a new local crowd, the system chooses the record which is the geographically nearest during the same period of a day as a reference to estimate its $Q(t)$. $S(t)$ is a linear function of the time t . By considering the requirements on the data delivery delay (T_{\max}) and the quality of the truth discovery (Q_D), we can find the optimal time (t) with the maximal utility, which will be further discussed in the next section.

5. Simulation Results and Discussions

In this section, we evaluate the performance of the local crowd. Our experiments are based on the dataset consisting of real vehicular traces. We evaluate the number of vehicles in the local crowd, the number of sensing reports, the ratio of the vehicles with reports, and the average delay of the reports. All the parameters are listed in Table 3.

5.1. Taxi-ROMA Dataset. When a vehicle moves into a local crowd, it will join in this crowd. Oppositely, when a vehicle moves out of a local crowd, it will leave this crowd. We do experiments on the Taxi-ROMA dataset [21]. This dataset contains real mobility traces of taxi cabs in Rome, Italy. It contains GPS coordinates of approximately 320 taxis collected over 30 days. We select the dataset of the traces collected on February 5, 2014, which contains 172 taxis. The traces cover the area with the range of 66 km × 59 km.

As shown in Figure 7, we set three events happening at different places with the yellow marks, which are termed as the northeast event, the southwest event, and the center event.

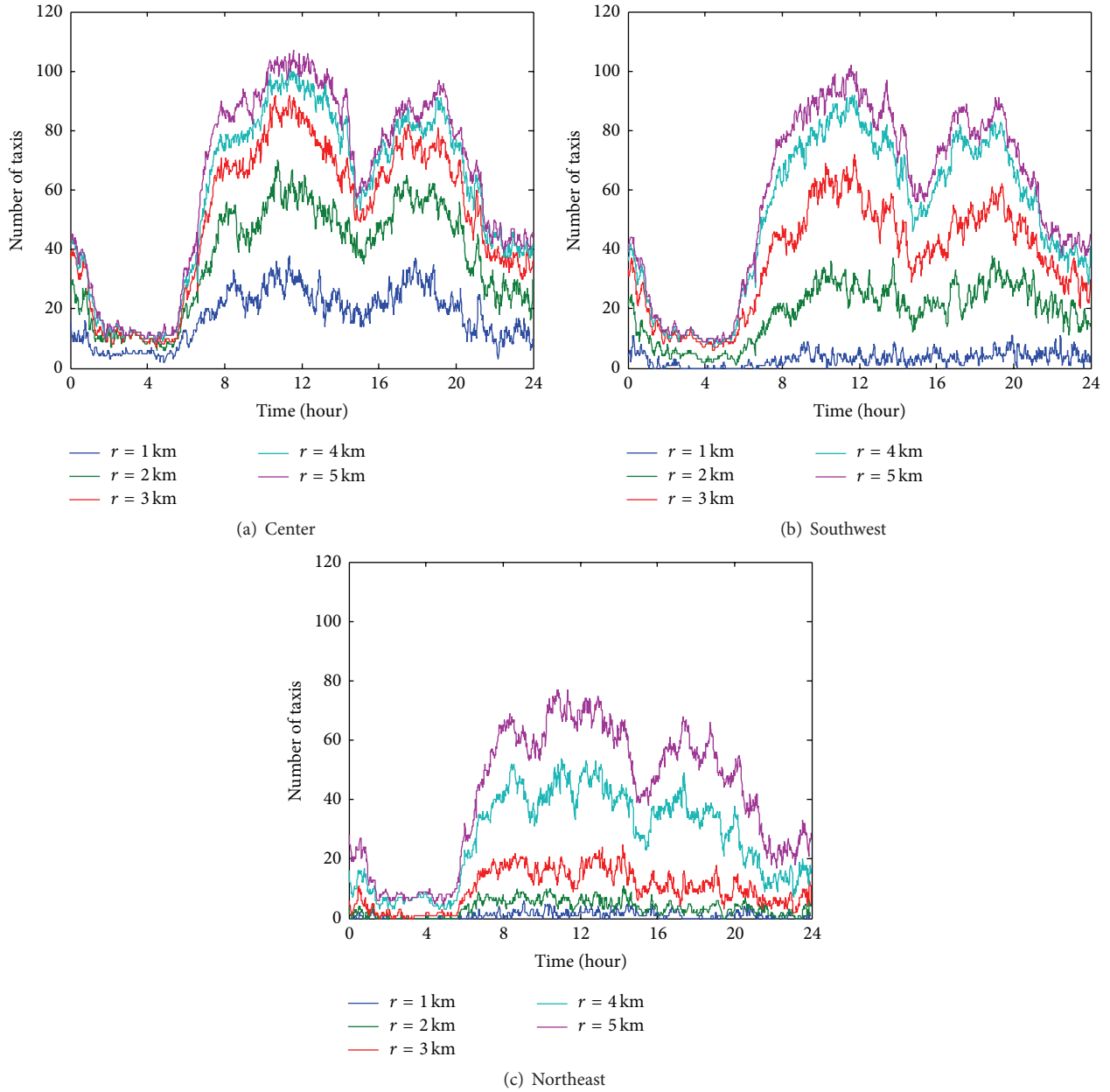


FIGURE 8: Number of the vehicles in a local crowd with different predefined ranges.

TABLE 3: Parameters of Roma taxi.

Parameter	Value
Day of the dataset	February 5, 2014
Number of the taxis	172
Range of the area (km ²)	66 × 59
Number of the events	3
Communication range (m)	300
Sensing range (m)	200
Range of the local crowd (km)	1, 2, 3, 4, and 5

The traffic density near the center event is the highest and that near the northeast event is the lowest.

The communication range of each vehicle is 300 m. In our experiments, the vehicles can communicate with each other or sense the events, only in the local crowd of each event. When the vehicle moves out of the crowd, it cannot transmit the reports of the event with other vehicles, but the reports in its buffer will not be dropped. The sensing range of an event is defined as the range from this event that a vehicle can sense it. In our experiment, we set the sensing range as 200 m.

5.2. *Number of Vehicles in the Local Crowd.* We evaluate the number of the vehicles in a local crowd with different predefined ranges (r) from 1 km to 5 km as a function of time during a day. Figure 8 shows the results of the events

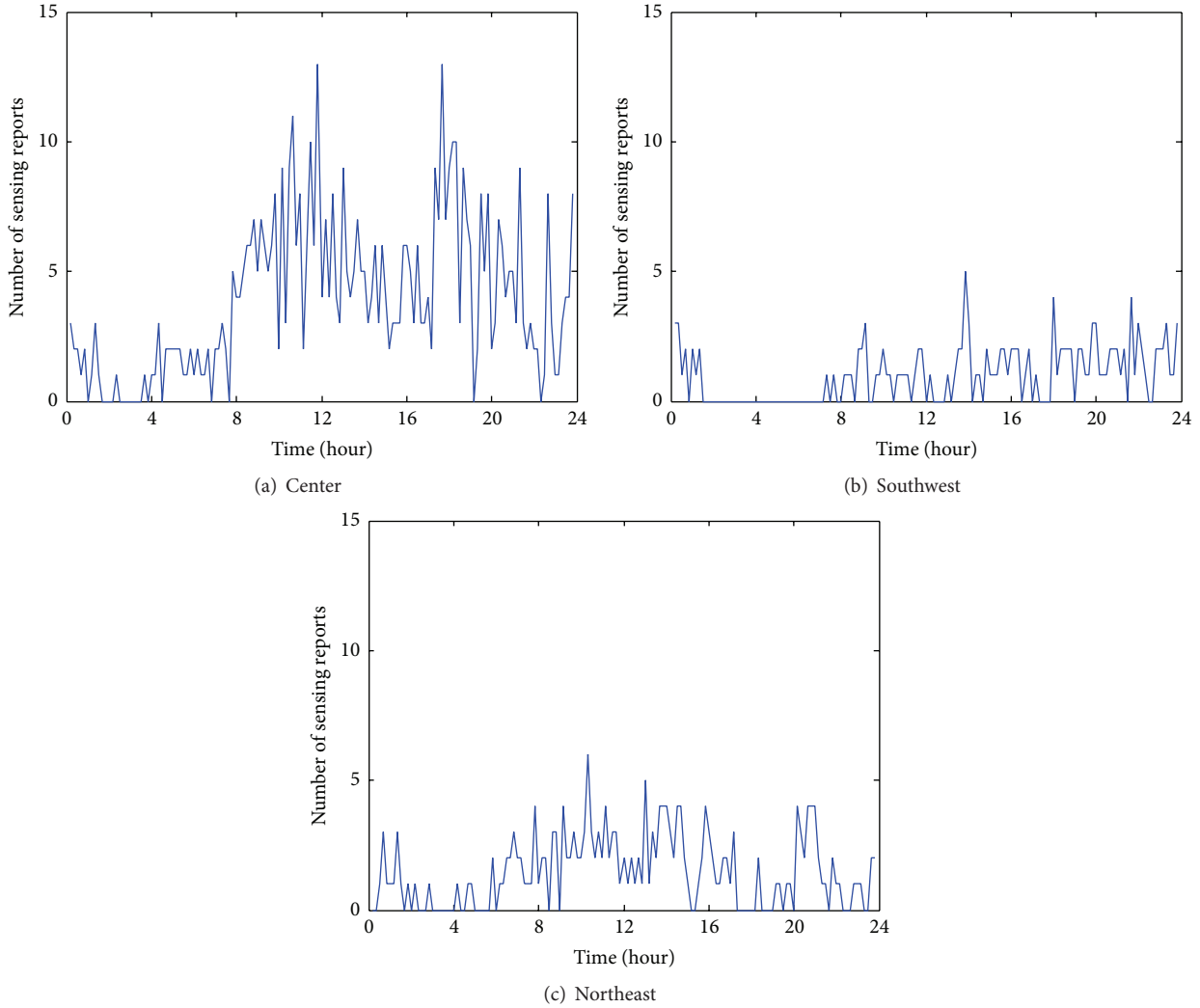


FIGURE 9: Number of the sensing reports in a local crowd.

which happen in different places. The size of the window for sampling is 60 seconds.

We notice that the number of vehicles in the local crowd of the center event is the largest, due to the high traffic density. In contrast, the number of vehicles in the local crowd of the northeast event is the smallest, due to the low traffic density.

While the range of the local crowd is increasing, the number of vehicles in the local crowd is also increasing. The local crowd with the range of 5 km has the largest number of vehicles for each event at different places.

The number of vehicles in the local crowd is changed at different times during the whole day. In particular, the number of vehicles between 2 am and 5 am is the smallest and that between 10 am and 12 am is the largest. That is also caused by the traffic density at different times.

5.3. Number of the Sensing Reports. When a vehicle moves into the sensing range of an event, it will generate a report for it. We evaluate the number of the sensing reports in a local

crowd as a function of time during a day. Figure 9 shows the results of the events which happen in different places. The range of the local crowd is set as 3 km.

We notice that the number of the sensing reports from the center event is the largest, due to the high traffic density. In contrast, the number of the sensing reports from the northeast event is the smallest, due to the low traffic density.

The number of the sensing reports is changed at different times during the whole day. In particular, the number of the sensing reports from the center event between 2 am and 4 am is the smallest. The number of the sensing reports from the southwest event between 2 am and 7 am is the smallest. The number of the sensing reports from the northeast event between 2 am and 5 am is the smallest.

The numbers of the sensing reports from the center event at about 12 am and 5 pm are the largest. The number of the sensing reports from the southwest event at about 1p is the largest. The number of the sensing reports from the northeast event at about 10 am is the largest. That is also caused by the traffic density at different times.

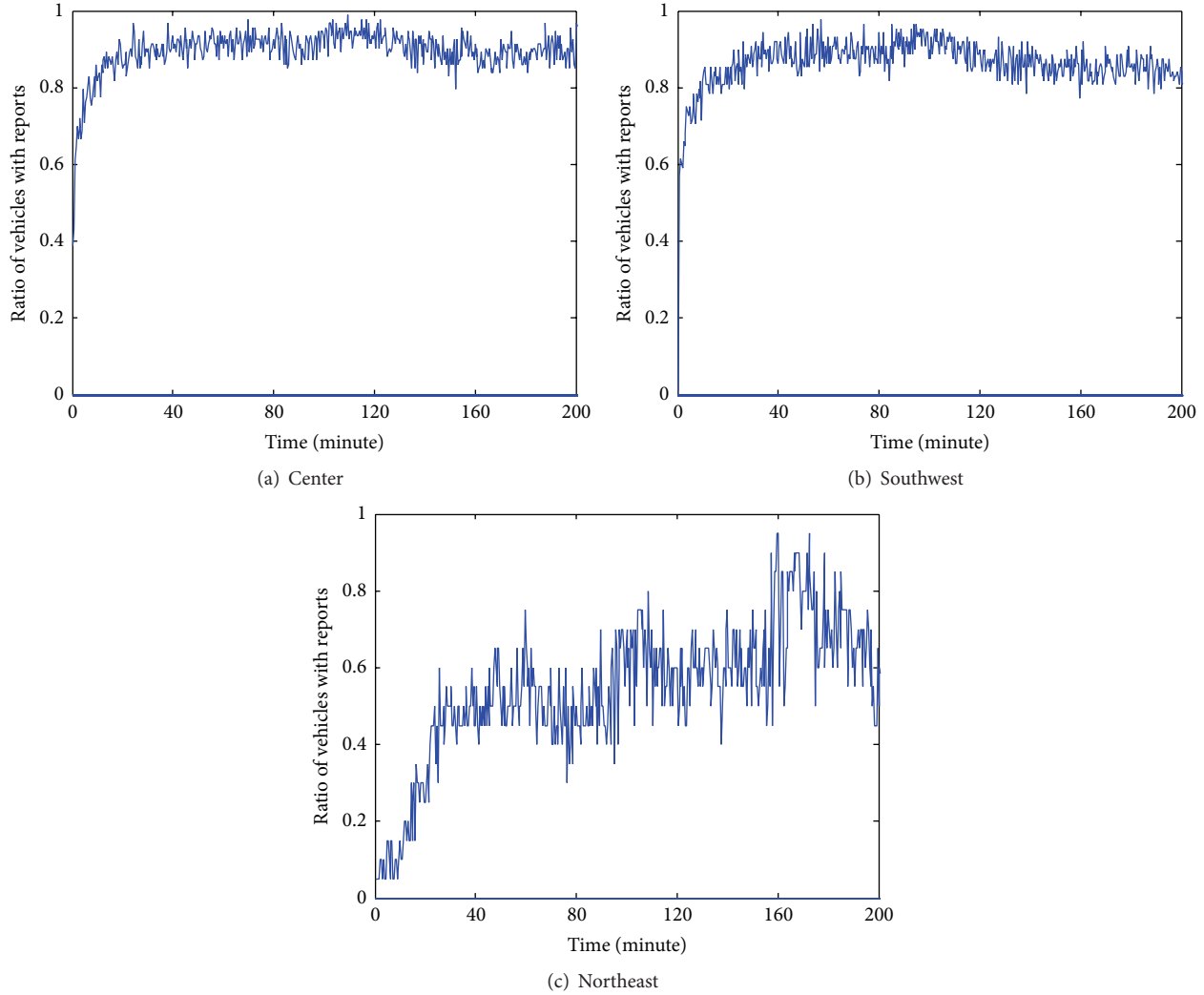


FIGURE 10: Ratio of the vehicles with the reports.

5.4. Ratio of the Vehicles with Reports. After generating a report by a vehicle, this report will be disseminated in the local crowd of this event. The ratio of the vehicles with the reports is defined as the ratio of the number of vehicles which has the reports to that of vehicles in the local crowd. We evaluate the changes of the ratio of the vehicles with the reports during 200 minutes after the first report is generated. Figure 10 shows the results of the events which happen in different places. The range of the local crowd is set as 3 km. Initially, the ratio is the lowest, since only the first vehicle in the local crowd has the report. We notice that with the increasing time, the ratio is also increasing, because more and more vehicles will receive the reports. The local crowd of the center event has the highest ratio, due to the high traffic density. In contrast, the local crowd of the northeast event has the lowest ratio, due to the low traffic density. Because there are some new vehicles without any reports moving into the crowd, the ratio cannot reach 1.

5.5. Quality and Utility of Local Crowd. The quality of the truth discovery at local crowd is evaluated by (8), which

compares the sensing data harvested by the local crowd with those harvested by the remote cloud. We evaluate the quality of the local crowd in different places, and the results are shown in Figure 11. The parameter Q_{\max} is set by 1. The range of the local crowd is set as 3 km. As in the aforementioned introduction, the sensing data from the same event are stochastic. In this simulation, the sensing data from the event follow a Poisson distribution. At the beginning, the quality is the lowest, since the vehicles in the local crowd have few sensing reports. We notice that with the increasing time the quality is also increasing, because the vehicles will receive more and more sensing reports. Finally, the quality will be approximated to Q_{\max} . Among the three places, the local crowd of the northeast event has the lowest quality, due to the low traffic density. Moreover, we notice that, for center and southwest regions, the quality of the local crowd is reaching Q_{\max} after about 20 minutes. In Figure 10, we also notice that, after about 20 minutes, the ratio of the vehicles with the reports for center and southwest regions is about 90%. This is because when the average number of the received reports in the local crowd is close to the total number at remote cloud,

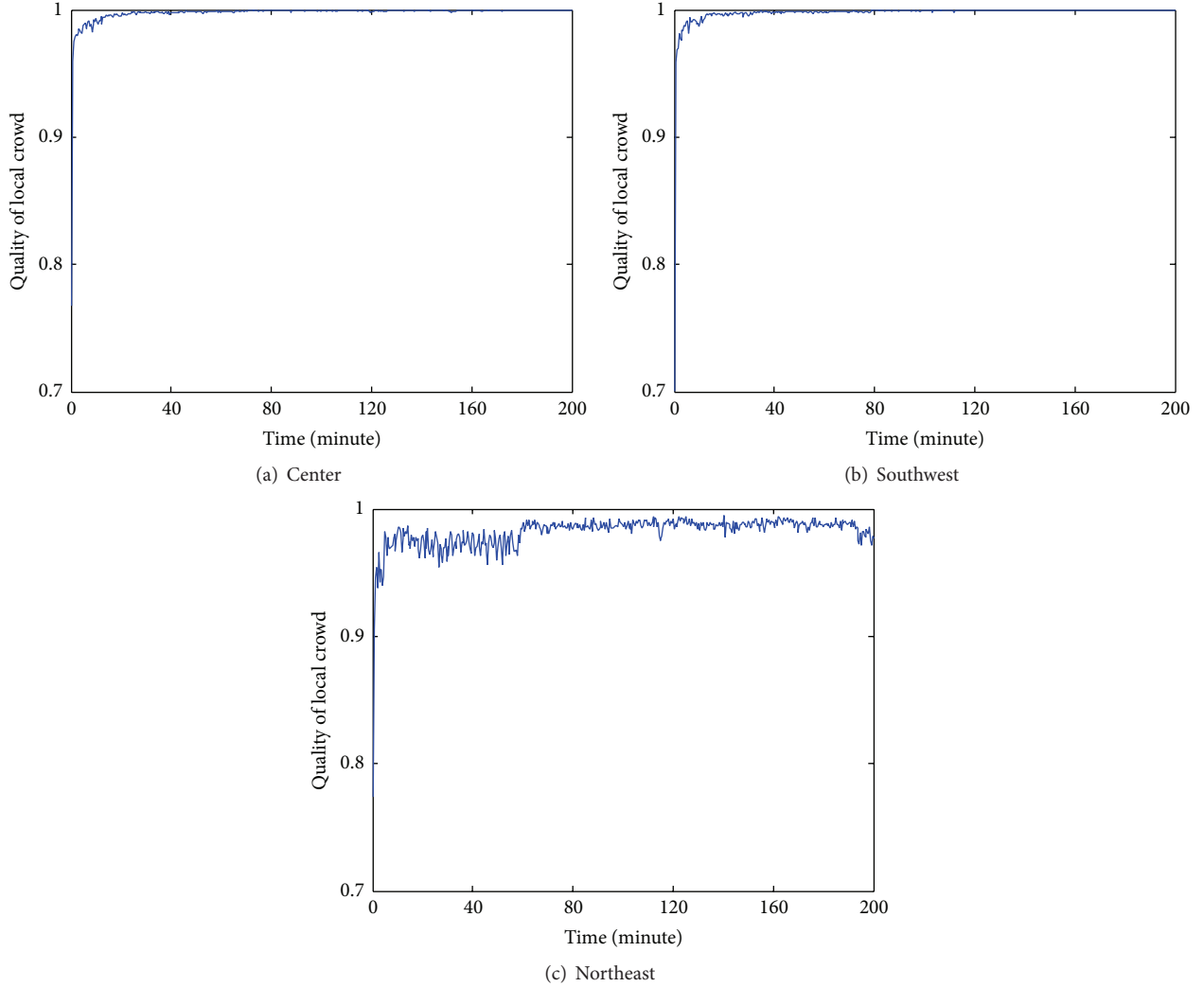


FIGURE 11: Quality of the local crowd.

there is much less difference between the performances of the local crowd and the remote cloud.

With the help of (10), we can find the optimal time with the maximal utility by considering the requirements on the data delivery delay and the quality of the truth discovery. We evaluate the utility of the local crowd in different places during the period of 10 minutes (T_{\max}) from the beginning, and the results are shown in Figure 12. The range of the local crowd is set as 3 km. Among the three places, the local crowd of the northeast event has the lowest utility, due to the low traffic density. We notice that when the time is equal to 2 minutes, the query user has the maximal utility.

5.6. Average Delay of Disseminating the Reports. We define the delay of disseminating a report from the time when the report is generated to the time when its copy is received by another vehicle. We evaluate the average delays of disseminating the reports from the events at different places as a function of different ranges of the local crowds from 1 km to 5 km. The lifetime of the report is 800 seconds, so the report

will be dropped by the mobile device after that time. Figure 13 shows the average delays of disseminating the reports from the events which happen in different places.

We find that the average delay of disseminating the reports from the center event is the lowest, due to the high traffic density. In contrast, the average delay of disseminating the reports from the northeast event is the highest, due to the low traffic density.

However, in the local crowd with the shortest range, the average delay is high, because the number of the vehicles is not enough to disseminate the reports. In the local crowd with the longest range, the average delay is also high, because of the long distance among the vehicles.

6. Conclusion

In mobile crowdsensing system, the cloud (or back-end server) harvests the crowdsensing data from the mobile devices and then aggregates the data to the feedback results for the querying users. Offloading the crowdsensing data

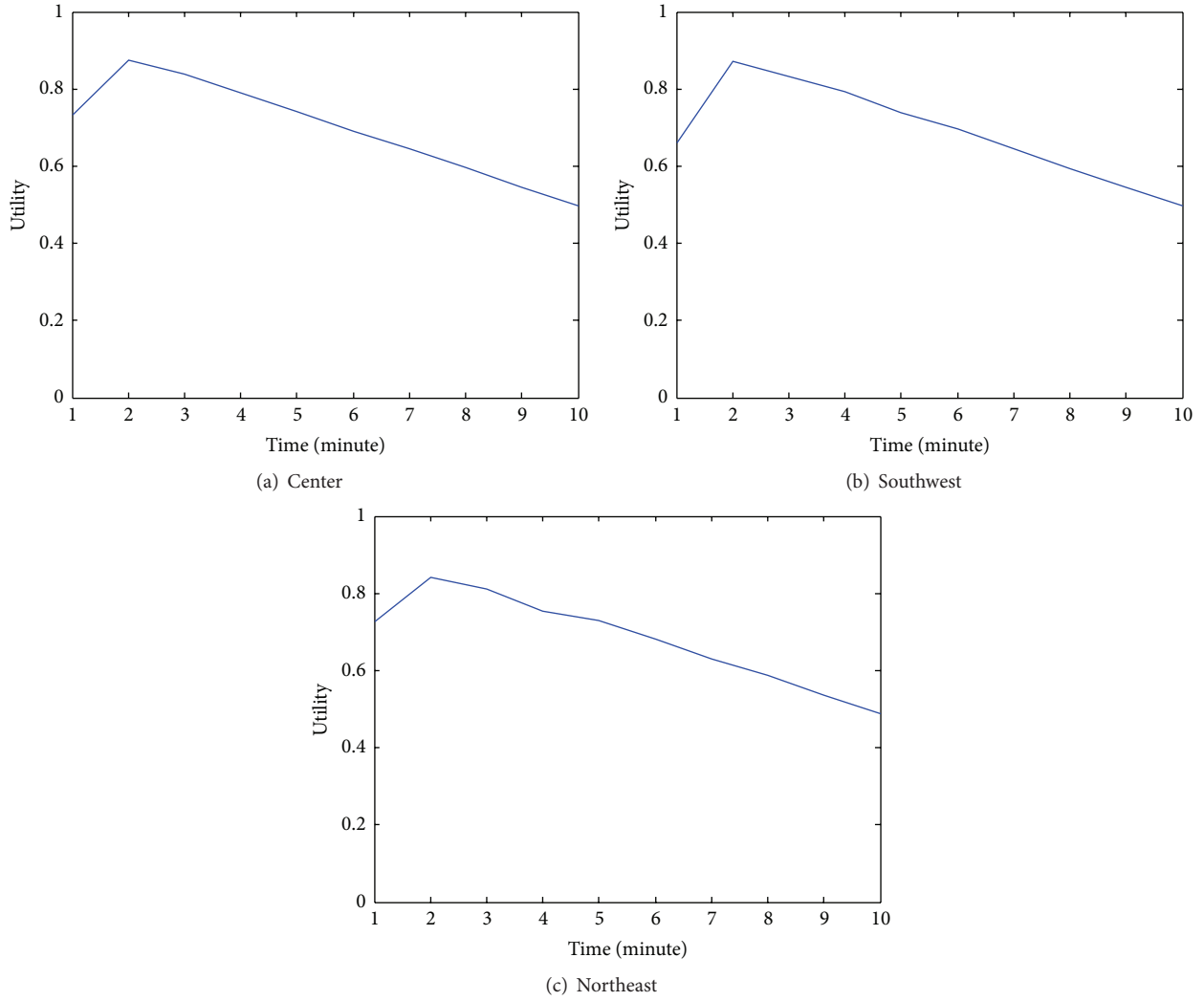


FIGURE 12: Utility of the local crowd.

to the cloud has less cost and energy consumption than the way of 3G/4G, but it has longer data delivery delay. Compared with offloading data via WiFi APs, we investigate the communication and sharing of crowdsensing data by vehicles near the sensing event, termed as a local crowd. The crowd-based approach has a lower delay than the offloading-based approach, by considering the quality of truth discovery. We define a utility function related to the crowdsensing data shared by the local crowd, in order to quantify the trade-off between the quality of the truth discovery and the user satisfaction. Our extensional simulations verify the effectiveness of our proposed schemes.

Notations

- r : Predefined range of the local crowd
- d : Delivery delay of a single sensing data
- t : Feedback delay for a querying user
- $U(t)$: Utility of the crowdsensing data in the local crowd at time t

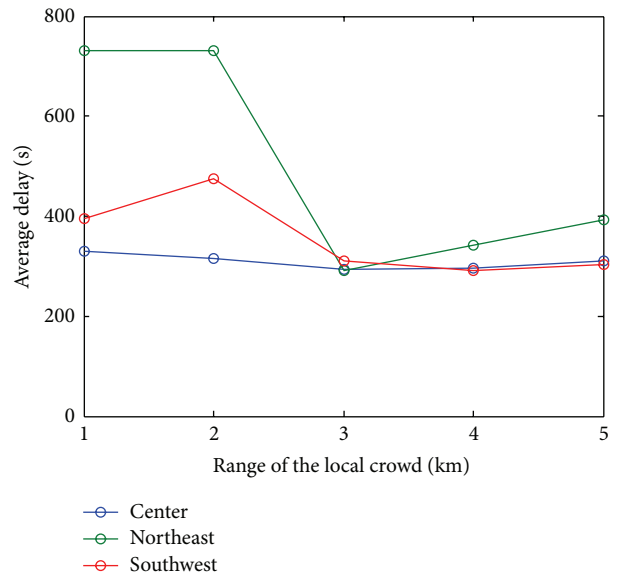


FIGURE 13: Average delay in the local crowd.

- $S(t)$: Satisfaction of feedback from the local crowd at time t
- $Q(t)$: Quality of the feedback results from the local crowd at time t
- T_{\max} : The maximum delay tolerance of the user
- $R(t)$: The discrete probability distributions of the crowdsensing data in the remote cloud at time t
- $L(t)$: The discrete probability distributions of the crowdsensing data in the local crowd at time t
- δ : Threshold of the utility
- Q_D : Demand of the quality of the truth discovery.

Competing Interests

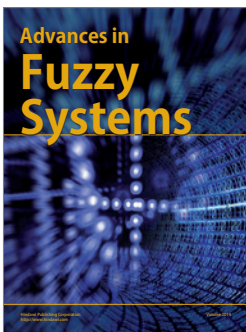
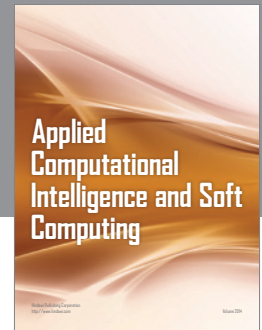
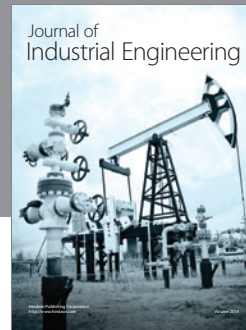
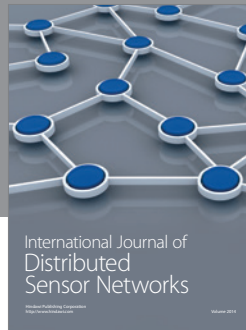
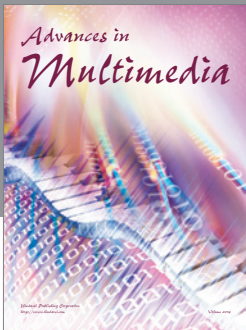
The authors declare that they have no competing interests.

Acknowledgments

This work is supported by NSFC Grants nos. 61170256, 61103226, 61173172, 61272526, and 61370204, the Fundamental Research Funds for the Central Universities (nos. ZYGX2013J077 and ZYGX2013J067), and the Applied Basic Program of Sichuan Province of China (no. 2014JY0192).

References

- [1] "List of countries by number of mobile phones in use," <http://en.wikipedia.org/wiki/>.
- [2] E. Miluzzo, N. D. Lane, K. Fodor et al., "Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application," in *Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems (SenSys '08)*, pp. 337–350, November 2008.
- [3] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: using a mobile sensor network for road surface monitoring," in *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services (ACM MobiSys '08)*, pp. 29–39, Breckenridge, Colo, USA, June 2008.
- [4] R. W. Ouyang, M. Srivastava, A. Toniolo, and T. J. Norman, "Truth discovery in crowdsourced detection of spatial events," in *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM '14)*, pp. 461–470, ACM, Shanghai, China, November 2014.
- [5] F. Mehmeti and T. Spyropoulos, "Is it worth to be patient? Analysis and optimization of delayed mobile data offloading," in *Proceedings of the IEEE Conference on Computer Communications (IEEE INFOCOM '14)*, pp. 2364–2372, Toronto, Canada, April 2014.
- [6] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: wireless indoor localization with little human intervention," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking (MobiCom '12)*, pp. 269–280, Istanbul, Turkey, August 2012.
- [7] P. Zhou, Y. Zheng, and M. Li, "Demo: how long to wait?: predicting bus arrival time with mobile phone based participatory sensing," in *Proceedings of the ACM 10th International Conference on Mobile Systems, Applications, and Services (MobiSys '12)*, pp. 459–460, 2012.
- [8] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking (MobiCom '12)*, pp. 173–184, Istanbul, Turkey, August 2012.
- [9] S. He, D.-H. Shin, J. Zhang, and J. Chen, "Toward optimal allocation of location dependent tasks in crowdsensing," in *Proceedings of the 33rd IEEE Conference on Computer Communications (IEEE INFOCOM '14)*, pp. 745–753, Ontario, Canada, May 2014.
- [10] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," *Proceedings of the ACM 8th International Conference on Mobile Systems, Applications, and Services (MobiSys '10)*, pp. 209–222, 2010.
- [11] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: how much can wifi deliver?" *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 536–550, 2013.
- [12] D. Zhang and C. K. Yeo, "Optimal handing-back point in mobile data offloading," in *Proceedings of the IEEE Vehicular Networking Conference (VNC '12)*, pp. 219–225, Seoul, South Korea, November 2012.
- [13] N. Ristanovic, J.-Y. Le Boudec, A. Chaintreau, and V. Erramilli, "Energy efficient offloading of 3G networks," in *Proceedings of the 8th IEEE International Conference on Mobile Ad-Hoc and Sensor Systems (MASS '11)*, pp. 202–211, Valencia, Spain, October 2011.
- [14] T. Luo and C.-K. Tham, "Fairness and social welfare in incentivizing participatory sensing," in *Proceedings of the 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '12)*, pp. 425–433, Seoul, Republic of Korea, June 2012.
- [15] Y. Zhang and M. van der Schaar, "Reputation-based incentive protocols in crowdsourcing applications," in *Proceedings of the 31st Annual IEEE International Conference on Computer Communications (IEEE INFOCOM '12)*, pp. 2140–2148, Orlando, Fla, USA, March 2012.
- [16] J. Ahn, Y. Wang, B. Yu, F. Bai, and B. Krishnamachari, "RISA: distributed road information sharing architecture," in *Proceedings of the IEEE INFOCOM*, pp. 1494–1502, Orlando, Fla, USA, March 2012.
- [17] J. Zhao and G. Cao, "Vadd: vehicle-assisted data delivery in vehicular ad hoc networks," in *Proceedings of the 25th IEEE International Conference on Computer Communications (IEEE INFOCOM '06)*, Barcelona, Spain, April 2006.
- [18] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 4, pp. 1801–1819, 2014.
- [19] A. Vahdat and D. Becker, *Epidemic routing for partially-connected ad hoc networks [M.S. thesis]*, 2000.
- [20] S. Kullback, "Information theory and statistics," *American Mathematical Monthly*, vol. 504, no. 3, p. 301, 1968.
- [21] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, and A. Rabuffi, "CRAWDAD data set roma/taxi (v. 2014-07-17)," July 2014, <http://crawdad.org/roma/taxi/>.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

