WILEY | Hindawi

Research Article

Quality Utilization Aware Based Data Gathering for Vehicular Communication Networks

Yingying Ren,[1] Anfeng Liu [ID],[1] Ming Zhao [ID],[2] Changqin Huang,[2] and Tian Wang[3]

[1]*School of Information Science and Engineering, Central South University, Changsha 410083, China*
[2]*School of Information Technology in Education, South China Normal University, Guangzhou 510631, China*
[3]*College of Computer Science and Technology, Huaqiao University, Xiamen, Fujian 361021, China*

Correspondence should be addressed to Anfeng Liu; afengliu@mail.csu.edu.cn

The vehicular communication networks, which can employ mobile, intelligent sensing devices with participatory sensing to gather data, could be an efficient and economical way to build various applications based on big data. However, high quality data gathering for vehicular communication networks which is urgently needed faces a lot of challenges. So, in this paper, a fine-grained data collection framework is proposed to cope with these new challenges. Different from classical data gathering which concentrates on how to collect enough data to satisfy the requirements of applications, a Quality Utilization Aware Data Gathering (QUADG) scheme is proposed for vehicular communication networks to collect the most appropriate data and to best satisfy the multidimensional requirements (mainly including data gathering quantity, quality, and cost) of application. In QUADG scheme, the data sensing is fine-grained in which the data gathering time and data gathering area are divided into very fine granularity. A metric named "Quality Utilization" (QU) is to quantify the ratio of quality of the collected sensing data to the cost of the system. Three data collection algorithms are proposed. The first algorithm is to ensure that the application which has obtained the specified quantity of sensing data can minimize the cost and maximize data quality by maximizing QU. The second algorithm is to ensure that the application which has obtained two requests of application (the quantity and quality of data collection, or the quantity and cost of data collection) could maximize the QU. The third algorithm is to ensure that the application which aims to satisfy the requirements of quantity, quality, and cost of collected data simultaneously could maximize the QU. Finally, we compare our proposed scheme with the existing schemes via extensive simulations which well justify the effectiveness of our scheme.

1. Introduction

Vehicles which equipped various types of sensing devices can sense all kinds of information in the surrounding environment [1–3]. At the same time, due to its strong mobility, the use of participatory sensing can obtain a large amount of sensing data in a very economical way [4–9]. These data could build various applications based on bid data. The examples of the classical use of such applications are as follows: VTrack [10] and Waze [11] is a system which can provide omnipresent traffic information by collecting vehicle operating status information; WeatherLah [12] is for giving fine-grained situation on the ground; and NoiseTube [13] is for making noise maps. In such applications, the time and space for data gathering are very large which leads to the general use of

participatory sensing to get the data. This type of network is called crowdsourcing network or participatory network [14–19]. Apparently, vehicular communication networks (VCNs) have a greater advantage in this regard [20–22]. First of all, it is a good participatory network for the reason that it has big number of vehicles which equipped various sensing devices which can sense different data efficiently; second, vehicles generally have strong communication skills [22, 23], which enable vehicles to exchange information with other vehicles, data center, and the external environments. They also play a significant role in supporting various applications such as road safety, traffic management, and a wide range of applications based on perceptual data [10–13]. On the other way, vehicular communication networks are also regarded as vital key technologies for next-generation

intelligent transportation systems (ITS) which are envisioned to greatly improve the transportation safety and efficiency by incorporating wireless communication and informatics technologies in the transportation system. In the end, vehicles flow in the city with greater mobility, which can bring great convenience to data collection. Vehicular communication networks, together with the cloud computing network [4, 7, 18, 24–29] as a key component of fog computing [7, 24, 25], provide the basis for the development of new network architectures.

Data is the basic resource of the current big data network [30–32]. Thus, efficient data gathering is the basis for effective work on big data networks, crowdsourcing networks, or participatory networks [33–36]. Many effective data gathering efforts have been proposed. Generally speaking [37–40], in the applications based on the data gathering, the system consists of the following three parts. (1) Data collectors: it generally refers to people or facilities equipped with sensing equipment, such as vehicles, smartphones, and monitor sensors equipped in the industrial field. There is a large number of sensing devices and the wide spread of their distribution could provide continuous, sufficient data for big data networks. Sometimes data collector is called data reporter or reporter [41–43]. (2) Data demander (DD): data demander needs data, sometimes called applications (or task publisher). They publish the specified need of data gathering and afford a certain payment to data reporter to encourage them to collect data [41–43]. Reporters collect data and report data to the data demander. Data demander processes the collected data and forms the application which can satisfy the requirement of customers, such as VTrack [10], Waze [11], WeatherLah [12], and Noise Tube [13], which afford different information services to customers. Data demander charges customers a fee to cover their expenses. (3) Customers: they consume applications' services and pay applications a fee.

The main content of this paper is to study how data demander chooses suitable data reporters to collect high quality data to create high quality applications and maximize the profits of the system. Generally, the data collection process is as follows: the data demander publishes the data gathering area, time, other attributes such as data quality, and the reward for reporting data samples. Data reporter senses and collects data with participatory sensing and reports the sensing data to data demander and gets the corresponding reward. How data demander selects right data reporter to collect data is an important issue. Some corresponding strategies and algorithms have been proposed. In summary, the following categories are mainly included.

(1) The data gathering strategy which aims to collect the specified number of data samples: the main goal in this type of strategy is to obtain the specified amount of data samples. These incentives mechanism of these strategies is generally based on the market supply and demand. If the reward provided by the data demander is enough to get adequate data samples, the system would try to cut the reward to reduce the cost. But the enthusiasm of data reporter will decrease if the reward is reduced. Thus, the amount of data samples collected by the data demander will decline too. In such a process of adjusting the reward, the system will reach a balanced stage. At this point, data demander collects the expected amount of data samples and the payment of demander is the least. However, the main drawbacks of this method are as follows. (a) First, only considering the amount of data obviously does not necessarily meet the needs of demander. The data collection should be based on a certain degree of data quality. (b) This strategy does not consider the time and space of data distribution and it will seriously affect the practical application. For example, in applications such as VTrack [10], Waze [11], WeatherLah [12], and NoiseTube [13], if the collected data samples are concentrated in a certain local area and many other areas do not have enough data samples, the lack of data in these areas will make the function of these applications lose their due role.

(2) Quality-based data gathering mechanism: such studies consider not only the amount and the cost of collected data but also the quality of the collected data so that the quality of the application could be improved. In this type of research, the criteria for evaluating data quality are generally defined first, which is followed by the selection to get high quality data based on the proposed criteria. For example, Quality of Contributed Service (QCS) is proposed by Tham and Luo [44] to evaluate the contribution of collected data to the application synthesized by DD. The metric named Quality-of-Information (QoI) is also proposed by Song et al. [3] to evaluate the quality degree of data samples. Thus, in this type of data collection strategy, DD maximizes the quality of composite applications by selecting the data of high QCS or QoI reporters. Data coverage which is proposed by Reddy et al. [31] is another measure of the quality of data collected. It mainly refers to whether the collected data could cover the perceived area well. Thus, when considering data coverage, the main problem is which data should be chosen to maximize the number of covered areas and balance the amount of data collected in different areas. However, due to the following three reasons, these strategies still have some shortcomings. (a) First, choosing high quality data does not definitely make applications have high quality. Even if the data of high Quality-of-Information (QoI) is preferentially selected every time, it is impossible to obtain high quality applications overall every time. Although the quality of a single data may be high, the quality of the application will not improve if data samples in this area are much enough which make the quality in this area high enough already. At this point, even the data with high quality cannot improve the quality of the application on the whole. At this moment, in the grids which have few data samples the quality will be efficiently improved even with the data which has low quality. Therefore, data collection with QoI only has local optimization but not overall optimization. (b) In the strategy which selects high QoI data, it is often assumed that the cost of each collected data is the same. Therefore, selecting high QoI data is good for the system. It must be pointed out that the cost of data with high QoI is different. And it is difficult to maintain data of high QoI with the same reward. On the other hand, the reporter in different areas collects data with different cost, so the quality of sensing data is not the same with the same cost, which also leads to deficiencies of the strategy which selects high quality data

first. For example, for vehicular communication networks, in urban centers, the number of vehicles is very large and the sensing data environment (such as wireless communication infrastructure) is also more advanced. As a result, the quality and the quantity of data collected in these areas are also very high. Thus, according to the high quality data-first strategy, there will be a lot of data collected in urban centers, whereas the majority of nonurban areas will collect very little data. This will seriously affect the performance of the application. (c) On the other hand, the problems in coverage are the simplification of the actual situation. There are also some studies that divide the area of data collection into smaller grids, and the coverage is used to equalize the amount of data collected in these grids. In practice, however, the amount of data needed by DD varies from region to region. It will increase the cost of the system unnecessarily if these regions are designated same amount of data. For example, data collection for many applications is related to location and time. For the situation that the needed data amount changes over time, the amount D of the data which the system needs to collect also changes constantly. To reduce the cost of data collection, the consideration of D is to ensure that the data collected could reflect the actual situation. And therefore, under different circumstances, D of different application often changes. For example, in the haze monitoring, if the weather and the haze are stable in the whole city, the cost can be lowered by reducing D. And the high quality services provided by application will not be affected. If there are large changes in the weather, it is necessary to collect more data. And the higher frequency of data acquisition should be applied to guarantee that the requirements of customers could be satisfied by applications. Obviously, this situation also exists in other applications. And these cases have not been considered in the previous studies. In addition, data collection varies from region to region. In general, urban centers or densely populated areas require a high level of data collection and therefore require a large amount of data to be collected. Relatively speaking, collecting a small amount of data in the remote area does not affect the performance of the application. For example, in the traffic information applications, the city center area is more congested, which requires more detailed data. Relatively speaking, the traffic congestion in the remote areas is much less than that in the urban centers, so it is not necessary to acquire the same amount of data.

The last downside is that the overall cost is seldom considered in previous studies. In fact, the system cost is a key factor which has a significant impact on the application. Therefore, DD needs to reduce cost based on improving the quality. This is rarely considered in previous studies. Many game-based or incentive-based mechanisms focus on the cost of the system. However, the affection of data quality, fine-grained time, and space is ignored. For example, to save costs, the collection of data is reduced during periods of steady climate, while different amounts of data are collected in different areas.

However, at present, there is a lack of such a framework that can support data collection in a fine-grained manner in time and space. Therefore, this paper fundamentally solves this problem by proposing a fine-grained data collection strategy.

(1) A fine-grained data gathering framework is proposed for vehicular communication networks to collect the most appropriate data and to best satisfy the multidimensional requirements (which mainly include data gathering quantity, quality, and cost) of application.

(2) A metric named "Quality Utilization" (QU) is to quantify the ratio of quality of the collected sensing data to the cost of the system. The application of "Quality Utilization" to our proposed algorithm shows that it is more effective in evaluating the overall performance of data collection strategy.

(3) Under the proposed data framework, a Quality Utilization Aware Data Gathering (QUADG) scheme is proposed to realize efficient, high quality data collection, in which the data collection time and area are divided into very fine granularity to meet the application of data collection quantity, quality, cost, and other multidimensional fine-grained requirements. Three data collection algorithms are proposed to maximize QU under different application requirements. The first algorithm is to ensure that the application which has obtained the specified quantity of sensing data can minimize the cost and maximize data quality by maximizing the QU. The second algorithm is to ensure that the application which has obtained two requests of application (the quantity and quality of data collection, or the quantity and cost of data collection) could maximize the QU. The third algorithm is to ensure that the application which aims to satisfy the requirements of quantity, quality, and cost of collected data simultaneously could maximize the QU.

(4) Finally, we compare our proposed scheme with existing schemes via extensive simulations. Extensive simulation results well justify the effectiveness of our scheme.

The rest of the paper is organized as follows. We review related work in Section 2. In Section 3, we describe the system model and formulate the problem of our data collection strategy. Section 4 presents the details of the fine-grained data collection framework as well as our Quality Utilization Aware Data Gathering (QUADG) scheme. We evaluate the proposed QUADG scheme via simulations in Section 5. We conclude the paper in Section 6.

2. Related Work

Mobile crowdsourcing [14–19, 23, 40–44] is a new data acquisition pattern that uses mobile devices (such as mobile smart phones and mobile automotive devices) to sense and acquire large amounts of data. It is a manifestation of the Internet of Things. It refers to the interactive, participatory perception networks [45–48] which are formed by mobile device. In crowdsourcing, the tasks will be published to the individuals or groups to help the professionals or the public to collect data and analyze information or to complete large-scale computing tasks and to share knowledge. Due to the rapid development of intelligent terminal equipment technology, a great variety of new intelligent terminal equipment has appeared which are widely used in various fields and applications such as industry, agriculture, transportation, market operation, environmental protection, and health care

[49, 50]. The rapid development of mobile sensing device makes mobile crowdsourcing reach a new height in both quantity and quality of data [14–19]. This is due to the following aspects. (1) First, there is a huge increase in the number of mobile intelligent sensing devices. Take smart-phones [10–13, 39, 43] as an example. By 2011, the global mobile smartphones have been sold more than PCs. By 2016, the number of global mobile subscribers has reached 7 billion [51]. China's mobile subscribers exceeded 900 million in 2016 [52]. The number of vehicles with more perceptual devices, stronger perception ability, and mobility also maintained a sustained and rapid growth. According to statistics, just in China, by the end of June 2017, there were 304 million motor vehicles in China and 328 million car drivers. There are 23 cities which have more than 2 million cars and 6 cities which have more than 6 million cars. At the same time, the advent of various new types of sensing devices has led to a geometric increase in the number of sensing devices. For example, the emergence of smart watches, smart bracelets, and smart health monitoring devices for human characteristics enables people to have 24-hour health check-ups and measurements without any uncomfortable feeling [53]. The widespread use of new radar, camera, water, wind, and geo-hazard sensing devices makes the data obtained by human beings not only limited to the surrounding environment of human beings, but also extended to unmanned rivers, lakes, and seas. The ability to sense data has reached an unprecedented height. (2) Second, the ability of these sensing devices to sense and acquire data has also been greatly improved. The processing power of latest Huawei P10 smart phone, whose CPU has eight cores: four 2.4 GHZ core and four 1.8 GHZ core, has exceeded some PCs. (3) The rapid development of wireless network technology makes mobile-aware devices seamlessly integrate into the existing basic network, thus forming a new type of ubiquitous network. Wireless broadband technologies such as 4G-LTE, LTE-A, and Wi-Fi have been widely used [2]. 5G and D2D technology is also witnessing rapid development. The development of these wireless broadband technologies has pushed pervasive perceptual computing to a new level. This allows users to use high quality network anytime and anywhere, which further boosts the growth of mobile crowdsourcing network [40–44]. This has aroused widespread concern in industry and academia, and thus there are more and more researches on it. Because the mobile crowdsourcing network can realize a wide range of data perception, it has been widely used in a variety of applications at the beginning of its development. Common Sense [54], an air quality testing application developed by the University of California, Berkeley (UCB), Creek Watch [55], a project conducted by IBM that investigates water pollution, and PEIR [56], a project utilizing the location information to study environmental impacts, are all the research in the field of crowdsourcing. In addition, Ear-Phone [57], noise level monitoring project, BikeNet [58], the health service project, CenceMe [59], the social awareness combined with social networking project, Waze [11], the well-known commercial mapping service company [57], and so on are classical application of crowdsourcing in social life. But these studies are still in progress. To summarize,
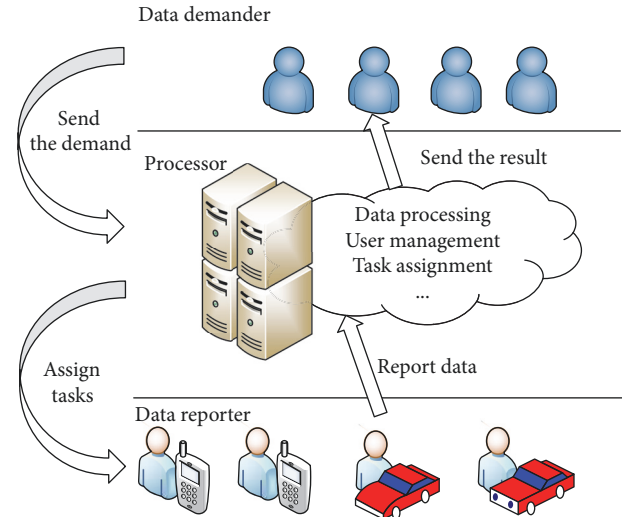

FIGURE 1: The classical architecture of crowdsourcing.

the relevant research is mainly focused on the following aspects: (1) the study on crowdsourcing network architecture; (2) strategies and methods used in data collection strategy; and (3) quality assurance strategies and methods for data collection strategies.

(1) The Framework of the Crowdsourcing Network. A typical system architecture of crowdsourcing is shown in Figure 1. The system architecture consists of three components. (1) Data demander (DD): DD usually is a service builder and service provider of an application, or service platform. In general, DD is to adapt to the needs of the market or to meet the application needs of a group of service consumers in the market. These applications are always based on big data and therefore require crowdsourcing to collect large amounts of data. If the data has been collected, DD will do the data extraction, cleaning, and refining and synthesize advanced services with collected data. The service is then provided to the service consumer who will be charged for a fee to cover the expense of collecting data, composing the service, and so on. (2) Data collector: it generally refers to people, vehicles, mobile phones, and others which have or are equipped with sensing devices. They are also called data reporter. These large amounts of perceptual data devices could provide real-time, large quantities of data for the application. (3) Service consumers: they consume the services provided by the application and pay providers a fee.

Although the above structure is one of the most commonly used crowdsourcing network architectures, it can vary from one application to another depending on the subject's role. At some point, the study does not emphasize service consumers, as in this article. At this time, the crowdsourcing network becomes a two-tier structure where data requesters and application services platforms are on the same aspect and data collectors are on the other. At this point the study focused on the interaction between data collection and DD, while in some studies, the emphasis is placed on the interaction and influence among the three roles. So, the interactive

structure has three components. It is obvious that the interaction in the three components will be more complicated. The interaction mechanism currently has the following aspects. (1) Interaction between DD (or service platform) and data collectors: to get a certain amount of data, DD need to take some incentives to motivate the data collector. (2) Interaction between DD and service consumer: with collected data, DD provides the corresponding service to the service consumer. And the service consumer pays a certain fee to enjoy the service. There exists the relationship of mutual restriction and mutual influence between the two kinds of interaction mechanisms. First, if the incentive payment for the data collector provided by DD is not high enough, the quantity and quality of the data collected by the data collector will not be good enough. As a result, the quality of service provided by the DD is unsatisfying. However, if the quality of service received by service consumers is poor, their consumption will be affected. Service consumers may choose other cost-effective service providers, which will lead to the decrease of DD's revenue. The decrease of revenue may further force DD to reduce the intensives of data collectors and thus the data collector cannot be effectively motivated. The reduction of incentive will further deteriorate the quality of service, which makes the system plunge into a vicious circle. So, the problem is how to formulate reasonable incentives for DD to attract data collector and to afford the high quality applications to service consumer so that data collector could obtain a certain income to get profits. However, how to optimize the incentive is not an easy task. Therefore, researchers put forward various optimization strategies and incentive strategies. Some incentive mechanisms and strategies closely related to this study are given below.

(2) Incentive Strategies and Methods Used in Data Collection. Although there are so many perceptual devices, or data reporters, it is still a major challenge for DD to collect the right data. First, data transmission and acquisition will require data collectors to pay a certain amount of vigor, time, energy, and other costs. As a result, data collectors, if not motivated enough to participate actively in data collection, will not be able to complete the task of collecting data easily. It has been pointed out by researchers [40–44] that the price paid by the data reporter mainly consists of two aspects: (1) the resources of perception such as battery power and (2) the computing resources and data traffic of participants' mobile devices. And in the process, participants also need to spend time and labor. Without proper rewards, participants are not interested in staying active in the long term. On the other hand, some sensitive data provided by the participants may reveal their privacy. The data perceived by participants includes many types of data, such as text, image, audio, and video. Most of these data are time and location sensitive, and users need to provide corresponding time and location besides data itself. The leakage of spatiotemporal information is another major factor preventing participants from joining.

To ensure that an adequate amount of data is collected, the incentive mechanism is an effective and commonly used method [19, 37, 40, 60]. It encourages and stimulates the participation of participants in sensing tasks via appropriate incentive mechanisms (models) and provides high quality, reliable sensing data. Different incentive mechanisms have different incentive effect for different groups of participants under different scenarios. From the point of reward towards data collection, the forms of intensives are usually as follows: money-based incentives, fun games incentives, social relationship incentives, and virtual integral incentives. (1) The most direct, effective, and most commonly used incentive form is the monetary incentive, which pays back participants' sensing data by means of money [61]. (2) Fun game incentive refers to the utilization of the game's entertainment and attractiveness to motivate data collectors to complete the perceived task. (3) Social relationship motivation is that, in a social networking relationship, participant could be motivated by maintaining a sense of belonging in a social relationship. (4) Virtual integral incentive means that participants get virtual integrals from perceived tasks in return. The real money converted by virtual credits, a certain kind of real thing, or the feeling of satisfaction brought by the mechanism will inspire users to participate in the sensing tasks.

There is a competition and game between DD and data collectors. Both sides hope to get the maximum benefit with minimum cost. From the view of DD, its main objective is to motivate more participants with minimal payment or controllable payment. It means DD need to improve the level of participation, while ensuring the high quality and reliability of data reported by participants. As a participant in completing the task, the main obstacle to actively participating in the perceived task is the problem of resource consumption and privacy security caused by perceived tasks.

According to the different incentive focus, the intensive mechanisms can be divided into two types: server-centric and user-centric. The server-centric approach first needs to be informed in advance about all information of data collectors such as quotes and data quality and then selects a subset which has the smallest payment and biggest utilization of data collectors as winner set. This server-centric approach to payment mainly uses game theory, the most important of which is the auction model. In the auction model, each participant has a bid b, a true bid v, and b ⩾ v to ensure that participants could receive nonnegative returns. The server learns the bidding information b in advance and then chooses the participants with the lowest bid rather than all the participants to reduce the payment. Obviously, DD tends to have an advantage in this type of interaction. Data collectors with high bid will always lose the chance to win the bid competition. However, this approach is often difficult to realize in practice. For example, for vehicular communication networks, each vehicle submits data independently, and the data is generally real-time data. It is difficult for DD to obtain global information and then optimize the choice. Therefore, in practice, most of the data collector-centric payment means that the server platform does not select the participants according to the information of all participants but pay the participant directly according to each participant's individual quotation or completion quality.

Auction and game are the most commonly used mechanisms. Reverse auctions are the most popular mechanism used by crowdsourcing. Reverse auctions refer to the auction

of one buyer and many potential sellers. In a reverse auction, the buyer presents the data they want and then awaits the contact of the sellers who hold the data they want. The potential seller continually shouts a lower price until no more seller calls out for a lower price. In the crowdsourcing system, DD is the buyer and the data collector is the seller. DD publishes sensing tasks, and the participants bid based on the expected reward of completing the tasks. Eventually, DD chooses the lowest-priced set of data collectors as winners and pays them off. Inverse Auction Incentives is a subset selection problem, in which the server platform chooses the subset of participants with the lowest payment under the premise of maximizing utility.

The Stackelberg game model is also commonly used in crowdsourcing, which is suitable for describing networks with multiple DDs [62]. This game model includes leaders and followers. Leaders act first and then followers adjust their strategies based on leader's actions to maximize their utility. In crowdsourcing, one of the DDs announces its own incentive price first, while the other DD adjusts its price according to the leader's published price to optimize its own profits. In such a competitive model, if a certain DD (a leader) takes the lead in raising his or her incentive price, the data collector will select the high-priced DD to submit the data, resulting in the other DD collecting no data or little data. So, other DD have to raise prices eventually. Similarly, there is also a leader-follower relationship between DD and data collectors. As DD adjusts the incentive price, data collectors also adjust their data collection strategies [62].

The essence of the above method is to adjust the incentive to get the specified number of data samples. The general process is as follows: assume the number of data applications expected to collect is D. If the actual received data is greater than D, the current reward for each data sample will be cut to reduce the cost of applications. On the contrary, if the actual received data is less than D, which means the current incentive is not enough, system will raise the reward for data sample to motivate the participants to collect adequate data samples. There are differences in the application of specific methods. However, this method did not fully consider the quality of the data. So, in the subsequent study, the researchers put forward some data collection methods to improve the quality of the data.

(3) Quality Assurance Strategies and Methods for Data Collection Strategies. In the incentive mechanism of crowdsourcing, the recruitment of many participants cannot guarantee that the sensing task is accomplished with high quality. While increasing the participation rate, it is important to ensure the quality of tasks. The distribution of the participants' position will affect the quality of the task. The quality of data uploaded by participants is different. Establishing an incentive mechanism based on data quality and user performance can motivate participants to upload high quality data. Incentives for quality-of-completion mainly include location-based, user-based (behaviors, contributions, and reputations), and data quality-based incentive mechanisms.

DDs generally serve service consumers by constructing data into services. Thus, DD's evaluation of data mainly lies in examining how the collected data contribute to the construction of services. Thus, the concept Quality of Contributed Service (QCS) is proposed by Tham and Luo [44] to measure the contribution of collected data to the combined services. And in the applications, such as noise mapping [13], traffic condition reporting [10], and environmental impact monitoring [5], the metric named Quality-of-Information (QoI) is proposed by Song et al. [3] to evaluate the quality of data samples. The disadvantage of this type of research is that the contribution of the collected data to the service is often difficult to evaluate which makes it hard to apply in practice.

In addition to the amount and the quality of sensing data, space-time coverage is often used as a standard by researchers to evaluate the collected data. The server platform not only recruits more participants at the lowest possible payment but also considers the user's location distribution. The broader the coverage of recruiters and the greater the range of sensing, the better the data quality. Therefore, Jaimes et al. [63] proposed the GIA algorithm based on the GBMC (greedy budgeted maximum coverage) algorithm which increases the coverage of interested area with a given budget.

In addition, the data collection not only needs to consider the quantity and quality of data but also needs to control the credibility of the participants because participants will misrepresent data or personal information to obtain more payment return. The literature [64, 65] proposes an incentive mechanism based on the user's reputation. In [66], the quality of sensing data is used as the participant's reputation value which is managed by the sever. Participants' rewards are related to their reputation ranking. This motivates participants to provide high quality data to maintain reputation value and get a higher return on payments. Literature [65] introduced existing social networks into crowdsourcing. It encourages participants to submit reliable data to establish a trusted mechanism via the reputation of participants in social networks.

3. System Models and Problem Statements

3.1. Network Model. The vehicular communication networks include n vehicle which compose mobile data collection network [67] via vehicle set $\mathscr{V}_{\mathbb{V}} = \{V_1, V_2, \ldots, V_n\}$, $\mathbb{V} \triangleq \{j = 1, 2, \ldots, n\}$. A vehicle can be regarded as an IoT node, moving in a smart city. Vehicle is equipped with detection and sensing equipment. So, vehicle can detect and perceive some interesting physical phenomena, events, and so on. $\mathscr{A}$ is on behalf of the application based on sensing data, such as VTrack [11] and Waze [12] for providing omnipresent traffic information, WeatherLah [13] for giving fine-grained situation on the ground, and NoiseTube [14] for making noise maps. Application $\mathscr{A}$ divided the entire inspection area into $\mathfrak{H}$ grids. $\mathbb{A}_A^{\mathscr{A}}$ denotes the set of grids, $A \triangleq \{i = 1, 2, \ldots, \mathfrak{H}\}$. $\mathbb{A}_i^{\mathscr{A}} \mid i \in \{1, 2, \ldots, \mathfrak{H}\}$ represents the grid i. At the same time, the whole sensing time is divided into a series of time periods, which are expressed as $\Gamma \triangleq \{t = 1, 2, \ldots, T\}$. Each time t consists of many smaller time intervals τ. In the period t, the number of data samples that application $\mathscr{A}$ needs to collect in the sensing area i can be expressed as $\mathfrak{R}_{i,t}^{\mathscr{A}} \mid \forall i \in A, \forall t \in \Gamma$.

The required exact number of data samples is dictated by the specific application needs.

Definition 1 (data reporter bid). For the vehicle, they want to get some real money or virtual reward in return. The vehicle needs to pay a certain cost such as the additional time it takes to collect data, communications, electricity, and so on. The vehicle also needs to bear the risk of privacy leak. The vehicle has an expected return on the data it submits. The reward that the vehicle claimed is the bid which varies according to the data collection conditions and the area. The bid of vehicle j could be denoted as $b_{j,i,t}^{\mathscr{A}}$, $\forall j \in \mathbb{V}$, $\forall i \in A$, $\forall t \in \Gamma$.

Definition 2 (data quality). Different data have different data quality. The quality of collected data is affected by the equipment performance, weather, and the signal strength when the data is submitted. The system can quantify the quality of data by integrating the past behavior of the vehicle and the quality of the submitted data which is denoted as $r_{i,t}^{\mathscr{A},j}$, $\forall j \in \mathbb{V}$, $\forall i \in A$, $\forall t \in \Gamma$.

The data quality and price of each data collected by the vehicles are different. So, the profile of vehicle j could be expressed as

$$V_j = \left\{\mathfrak{D}_{i,t}^{\mathscr{A},j}, b_{i,t}^{\mathscr{A},j}, r_{i,t}^{\mathscr{A},j}\right\}, \quad \forall j \in \mathbb{V}, \ i \in A, \ \forall t \in \Gamma. \quad (1)$$

$\mathfrak{D}_{i,t}^{\mathscr{A},j}$ indicates the grid data owned by vehicle j. $b_{i,t}^{\mathscr{A},j}$ denotes the bid based on the cost of vehicle data reporter. $r_{i,t}^{\mathscr{A},j}$ indicates the data quality of submitted by data reporter.

Definition 3 (data demander). Data demander affords a corresponding payment when publishing the data collection application. According to the requirement and the budget of application, each demander could provide their expected payment to the system which could be denoted as

$$\mathscr{P}^{\mathscr{A}} = \mathscr{P}_{A,\Gamma}^{\mathscr{A}}. \quad (2)$$

That is,

$$\mathscr{P}^{\mathscr{A}} = \mathscr{P}_{i,t}^{\mathscr{A}} \mid \forall i \in A, \ \forall t \in \Gamma. \quad (3)$$

Demander could announce the request of data quality in addition to the payment. The data quality designated by the data demander could be expressed as $\mathscr{L}^{\mathscr{A}} = L_{A,\Gamma}^{\mathscr{A}} = L_{i,t}^{\mathscr{A}} \mid \forall i \in A$, $\forall t \in \Gamma$. In the system, data demanders can choose to make a request for the quality and the reward or data, or not. The profile of a demander k could be expressed as

$$U_k^{\mathscr{A}} = \left\{\mathfrak{R}_{A,\Gamma}^{\mathscr{A}}, \mathscr{P}_{A,\Gamma}^{\mathscr{A}}, L_{A,\Gamma}^{\mathscr{A}}\right\} = \left\{\mathfrak{R}_{i,t}^{\mathscr{A}}, \mathscr{P}_{i,t}^{\mathscr{A}}, L_{i,t}^{\mathscr{A}}\right\} \mid$$
$$\forall i \in A, \ \forall t \in \Gamma. \quad (4)$$

After data demander submits the profile, the system needs to select the appropriate vehicle to complete the task according to the demander requirement. However, due to the lack of data, market law, and other reasons, the requirements of data demander not always can be completely satisfied.

The system optimizes the task allocation scheme for data collection based on demander requirements.

If vehicle k is selected and provides data sample $\mathfrak{D}_{i,t}^{\mathscr{A},k}$ for application $\mathscr{A}$ in time t, then

$$x_{i,t}^{\mathscr{A},k} = \begin{cases} 1, & \text{if vehicle } k \text{ was selected at time } t \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

$X^{\mathscr{A},k}$ denotes the amount of data samples reported by the data reporter k:

$$X^{\mathscr{A},k} = \sum_{\forall i \in A, \forall t \in \Gamma} x_{i,t}^{\mathscr{A},k} \mid \forall k \in \mathbb{V}. \quad (6)$$

The aggregate amount of data collected in all areas can be expressed as

$$X^{\mathscr{A}} = \sum_{\forall k \in \mathbb{V}} X^{\mathscr{A},k}. \quad (7)$$

The total payment of data demander in all the area could be expressed as

$$\mathscr{C}^{\mathscr{A}} = \sum_{\forall i \in A, \forall t \in \Gamma, k \in \mathbb{V}} \left(x_{i,t}^{\mathscr{A},k} \mathscr{P}_{i,t}^{\mathscr{A}}\right). \quad (8)$$

The quality of all areas is

$$\mathscr{Q}^{\mathscr{A}} = \sum_{\forall i \in A, \forall t \in \Gamma, \forall k \in \mathbb{V}} \left(x_{i,t}^{\mathscr{A},k} r_{i,t}^{\mathscr{A},k}\right). \quad (9)$$

To simplify the comparison, take the average quality as output.

The allocation of the system could be expressed as

$$T^{\mathscr{A}} = \left\{X^{\mathscr{A}}, \mathscr{C}^{\mathscr{A}}, \overline{\mathscr{Q}^{\mathscr{A}}}\right\}. \quad (10)$$

$X^{\mathscr{A}}$ indicates the amount of data the system can provide for application $\mathscr{A}$; $\mathscr{C}^{\mathscr{A}}$ represents the payment required for the system to collect data; $\overline{\mathscr{Q}^{\mathscr{A}}}$ represents the average quality of data collected by the system.

3.2. Problem Statements. The publisher of data collection is data demander. In common data collection models, data demander often only considers the data quantity and data collection costs but ignores the data quality requirements of the data demander.

$X^{\mathscr{A}}$ denote the amount of collected data samples for application $\mathscr{A}$ and $X_i^{\mathscr{A}}$ is the number of samples collected for task $\mathscr{A}$ on grid i, within a certain time slot t.

$$X_i^{\mathscr{A}} = \sum_{\forall t \in \Gamma, \forall k \in \mathbb{V}} x_{i,t}^{\mathscr{A},k}. \quad (11)$$

Definition 4 (quality utilization r). "Quality Utilization" (QU) is to quantify the ratio of quality of the collected sensing data to the cost of the system. According to the relationship between quality and bid, we can divide the data into four categories: data with high quality and low bid; data with high quality and high bid; data with low quality and low bid; and data with low quality and high bid.

Application is desperate for the data with high quality and low bid. And the data with low quality and high bid should be eliminated. The data collected in grid area i, $i \in A$ for application A can be denoted as

$$D_i = \{d_1, d_2, \ldots, d_n\}. \tag{12}$$

The Quality Utilization $r_{i,t}^{\mathscr{A},j}$ of data which is collected in grid area i, $i \in A$ during time slot t by data reporter j can be denoted as

$$r_{i,t}^{\mathscr{A},j} = \frac{b_{i,t}^{\mathscr{A},j}}{r_{i,t}^{\mathscr{A},j}} \quad \forall A \in \mathbb{A}_{\mathscr{A}}. \tag{13}$$

(1) $U_k^{\mathscr{A}} = \{\mathfrak{R}^{\mathscr{A}}\}$. When the data demander does not propose the collection payment $\mathscr{P}^{\mathscr{A}}$ and the data average quality requirement $\mathscr{L}^{\mathscr{A}}$ and only the data amount $\mathfrak{R}^{\mathscr{A}}$ is proposed, the goal of system is how to allocate the data collection task to obtain the higher data quality with lower payment.

For each grid G_i specified by application, the data amount $X_A^{\mathscr{A}}$ collected by system should be bigger than the demand of the data demander: that is,

$$X_{i,t}^{\mathscr{A}} \geq R_{i,t}, \quad \forall i \in A, \ \forall t \in \Gamma. \tag{14}$$

At the same time, the system should minimize the payment and maximize the quality of data. Therefore, the goal of the system could be expressed as

$$X_A^{\mathscr{A}} \geq R_{i,t}, \quad \forall i \in A, \ \forall t \in \Gamma$$
$$\min\left(\mathscr{C}^{\mathscr{A}}\right) = \min \sum_{\forall i \in A, \forall t \in \Gamma, k \in \mathbb{V}} \left(x_{i,t}^{\mathscr{A},k} b_{i,t}^{\mathscr{A},j}\right)$$
$$\max\left(\mathcal{Q}^{\mathscr{A}}\right) = \max \sum_{\forall i \in A, \forall t \in \Gamma, \forall k \in \mathbb{V}} \left(x_{i,t}^{\mathscr{A},k} r_{i,t}^{\mathscr{A},k}\right) \tag{15}$$
$$\max\left(\mathcal{Q}U^{\mathscr{A}}\right) = \max \sum_{\forall t \in \Gamma, A \in \mathbb{A}_{\mathscr{A}}, j \in (1,\ldots,n)} \frac{b_{i,t}^{\mathscr{A},j}}{r_{i,t}^{\mathscr{A},j}}.$$

(2) $U_k^{\mathscr{A}} = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{P}^{\mathscr{A}}\}$. Because the data demander only proposed the request for data collection payment and data amount, the goal of the system is to maximize Quality Utilization under designated payment according to the requirement of data demander. This is QUADG scheme which could maximize the Quality Utilization. And the request of data demander could be expressed as

$$X_A^{\mathscr{A}} \geq R_{i,t}, \quad \forall i \in A, \ \forall t \in \Gamma$$
$$\mathscr{C}^{\mathscr{A}} = \sum_{\forall i \in A, \forall t \in \Gamma, k \in \mathbb{V}} \left(x_{i,t}^{\mathscr{A},k} b_{i,t}^{\mathscr{A},j}\right) \leq \mathscr{P}^{\mathscr{A}} \tag{16}$$
$$\max\left(\mathcal{Q}U^{\mathscr{A}}\right) = \max \sum_{\forall t \in \Gamma, A \in \mathbb{A}_{\mathscr{A}}, j \in (1,\ldots,n)} \frac{b_{i,t}^{\mathscr{A},j}}{r_{i,t}^{\mathscr{A},j}}.$$

(3) $U_k^{\mathscr{A}} = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$. When the data demander proposed the requirement for the data average quality $\mathscr{L}^{\mathscr{A}}$ and data amount

$\mathfrak{R}^{\mathscr{A}}$, the goal of system is how to allocate the data collection task to obtain the higher data quality with lower payment. With scheme QUADG, the goal of system could be expressed as

$$X_A^{\mathscr{A}} \geq R_{i,t}, \quad \forall i \in A, \ \forall t \in \Gamma$$
$$\overline{\mathcal{Q}^{\mathscr{A}}} = \frac{1}{N} \sum_{\forall i \in A, \forall t \in \Gamma, \forall k \in \mathbb{V}} \left(x_{i,t}^{\mathscr{A},k} r_{i,t}^{\mathscr{A},k}\right) \geq \mathscr{L}^{\mathscr{A}} \tag{17}$$
$$\max\left(\mathcal{Q}U^{\mathscr{A}}\right) = \max \sum_{\forall t \in \Gamma, A \in \mathbb{A}_{\mathscr{A}}, j \in (1,\ldots,n)} \frac{b_{i,t}^{\mathscr{A},j}}{r_{i,t}^{\mathscr{A},j}}.$$

(4) $U_k^{\mathscr{A}} = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{P}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$. When the data demander proposed the requirement for payment $\mathscr{P}^{\mathscr{A}}$, the data average quality $\mathscr{L}^{\mathscr{A}}$, and data amount $\mathfrak{R}^{\mathscr{A}}$, the goal of system is how to allocate the data collection task to satisfy the data demander according the request. With scheme QUADG, the goal of system could be expressed as

$$X_A^{\mathscr{A}} \geq R_{A,t}, \quad \forall A \in \mathbb{A}_{\mathscr{A}}, \ \forall t \in \Gamma$$
$$\overline{\mathcal{Q}^{\mathscr{A}}} = \frac{1}{N} \sum_{\forall i \in A, \forall t \in \Gamma, \forall k \in \mathbb{V}} \left(x_{i,t}^{\mathscr{A},k} r_{i,t}^{\mathscr{A},k}\right) \geq \mathscr{L}^{\mathscr{A}}$$
$$\mathscr{C}^{\mathscr{A}} = \sum_{\forall i \in A, \forall t \in \Gamma, k \in \mathbb{V}} \left(x_{i,t}^{\mathscr{A},k} b_{i,t}^{\mathscr{A},j}\right) \leq \mathscr{P}^{\mathscr{A}} \tag{18}$$
$$\max\left(\mathcal{Q}U^{\mathscr{A}}\right) = \max \sum_{\forall t \in \Gamma, A \in \mathbb{A}_{\mathscr{A}}, j \in (1,\ldots,n)} \frac{b_{i,t}^{\mathscr{A},j}}{r_{i,t}^{\mathscr{A},j}}.$$

4. Scheme Design

To state the parameter of this paper clearly, the main notions introduced in this paper can be found in Parameter Description.

4.1. Motivation. In time slot t, the amount of data sample in sensing area A could be expressed as $\mathscr{D}^{\mathscr{A}} = D_{A,t}$, $\forall A \in \mathbb{A}_{\mathscr{A}}$, $\forall t \in \Gamma$. Figure 2 shows the common process of participatory data collection. In the data collection of crowdsourcing, data reporter sense data with mobile sensing devices and then report the data to the data center nearby. Data center processes the data for the specific industrial sensing applications and returns the service to the customer. Applications submit the request of the data demander to the platform. The platform allocates the tasks to the data reporters.

Each vehicle can be regarded as a data reporter. And when a vehicle passes through a grid and submits the data, it shows that the vehicle owns the grid's resources.

At present, most crowdsourcing system always takes the auction as intensive mechanism. Most methods always only concentrate on the payment of data collection when using auction mechanism as intensive mechanism but the data quality is ignored.

Actually, the quality of different application is different. Besides, the quality of the data in the same area which is
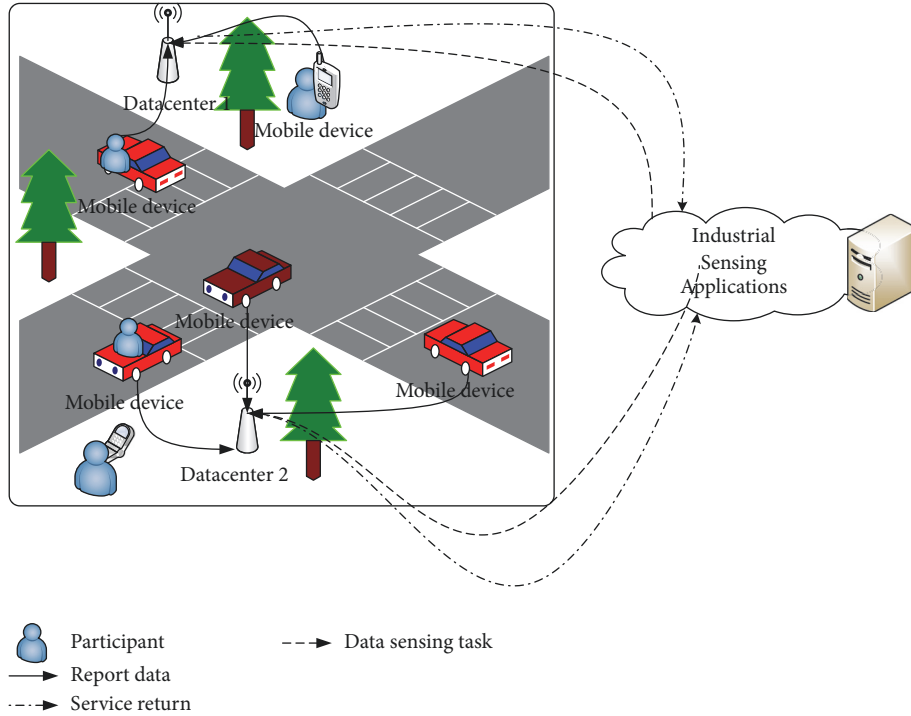
FIGURE 2: The process of participatory data collection.

provided by different data reporter is also different. If we take no account of the difference in the quality of data and the data request of different applications, the quality of data for certain application may not satisfy the demand of application. It is necessary to come up a new plan to allocate task.

Due to the neglect of data quality, most data collection schemes often use Random Selection Data Gathering (RSDG) to select data. In this scheme, the quality differences between different data in the same region are ignored.

When allocating data collection tasks, if the amount of data is small, the best allocation solution could be found by enumerating. But for areas where data amount is big, such as airports whose amount of data is up to 100,000, the needs of the data demander could not be satisfied. It is necessary to find a suitable allocation strategy to allocate the data collection task according to the request of data demander.

4.2. Quality Utilization Based Programming. Because of the difference in data gathering payment, environment, and gathering conditions, the quality of data is different. And different data reporters have different expectation for the reward which makes bid different. Quality Utilization Aware Data Gathering (QUADG) can obtain a balanced plan between payment and quality.

QUADG uses greedy strategy to allocate task for the data reporter. When making original selection, the system greedily chooses data with higher $r_{i,t}^{j,A}$ as original winner set. If the original winner set cannot satisfy the request of data demander, QUADG will replace part of data in the winning set. The main steps of algorithms are as follows.

(1) $U_i = \{\mathfrak{R}^{\mathscr{A}}\}$

Step 1. Choose grid area i, $\forall i \in A$, sort all the data collected in area i, $\forall i \in A$ according to r in descending order. The ordered Quality Utilization list can be denoted as

$$\mathbb{R}_i = \{r_{(1)}, r_{(2)}, \ldots, r_{(\mathfrak{R}^{\mathscr{A}})}, \ldots, r_{(n)}\}. \tag{19}$$

The corresponding ordered data list can be denoted as

$$\mathbb{D}_i = \{d_{(1)}, d_{(2)}, \ldots, d_{(\mathfrak{R}^{\mathscr{A}})}, \ldots, d_{(n)}\}. \tag{20}$$

Step 2. Choose first $\mathfrak{R}_i^{\mathscr{A}}$ data as winning set $\mathbb{W}$ which can be denoted as follows:

$$\mathbb{W} = \{d_{(1)}, d_{(2)}, \ldots, d_{(\mathfrak{R}^{\mathscr{A}})}\}. \tag{21}$$

Step 3. Repeat Step 1 until all the areas have been traversed.

(2) $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{P}^{\mathscr{A}}\}$, $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$. We take demander worker's profile as $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$: for example, the details of algorithm are as follows.

Step 1. Choose grid area i, $\forall i \in A$, sort all the data collected in area i, $\forall i \in A$ according to r in descending order. The ordered Quality Utilization list can be denoted as

$$\mathbb{R}_i = \{r_{(1)}, r_{(2)}, \ldots, r_{(\mathfrak{R}^{\mathscr{A}})}, \ldots, r_{(n)}\}. \tag{22}$$

The corresponding ordered data list can be denoted as

$$\mathbb{D}_i = \{d_{(1)}, d_{(2)}, \ldots, d_{(\mathfrak{R}^{\mathscr{A}})}, \ldots, d_{(n)}\}. \tag{23}$$

```
Input: data requester profile: U_i = {ℜ^𝒜, ℒ^𝒜}
Output: Winner set 𝕎, T^𝒜 = {X^𝒜, 𝒞^𝒜, \overline{Q^𝒜}}
(1)    Q^𝒜 = 0, 𝒞^𝒜 = 0, 𝕎 = ∅// to initialize the winner set and the achieved quality and payment
(2)    For each grid i | ∀i ∈ A
(3)        For n = 1 to |D_i|
(4)            r_{i,t}^{j,A} = b_{i,t}^{𝒜,j}/r_{i,t}^{𝒜,j}// to compute the Quality Utilization of each data in gird A.
(5)        End for
(6)        𝔻_i = Sort D_i according to r in ascending order // to get the ordered data set
(7)        ℝ_A = Sort {r_{i,t}^{j,A}} in ascending order
(8)        For m = 1 to ℜ_A^𝒜
(9)            x_{(m)}^{𝒜,A,t} = 1; // to mark the selected data
(10)           𝕎 = 𝕎 ∪ {d_{(m)}}// to select the winning data
(11)           Q_i^𝒜 = Q_i^𝒜 + r_{(m)}
(12)           \overline{Q_i^𝒜} = ∑ x_{(m)}r_{(m)}/X_i^𝒜
(13)           𝒞_i^𝒜 = 𝒞_i^𝒜 + b_{(m)}
(14)       End for
(15)           m = ℜ^𝒜 + 1
(16)       While \overline{Q_i^𝒜} < ℒ_i^𝒜 && m < X_i^𝒜
(17)           find d_e in 𝕎 whose r_e is smallest// to find the data which will be eliminated from the winner set
(18)           If r_e < r_{(m)}
(19)               delete d_e from 𝕎
(20)               x_{(m)}^{𝒜,A,t} = 1// to mark the data which is newly selected
(21)               𝕎 = 𝕎 ∪ {d_{(m)}}// to select the new data
(22)               Renew \overline{Q_i^𝒜}, 𝒞_i^𝒜
(23)           End if
(24)               i = i + 1
(25)       End while
(26) End for
(27)       𝒞^𝒜 = ∑ 𝒞_i^𝒜, ∀i ∈ A
(28)       \overline{Q^𝒜} = ∑ \overline{Q_i^𝒜}/ℌ
(29)       T^𝒜 = {X^𝒜, 𝒞^𝒜, \overline{Q^𝒜}}
(30) Return 𝕎, T^𝒜
```

ALGORITHM 1: The Quality Utilization Based Programming (QUADG), $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$.

Step 2. Choose first $\mathfrak{R}_i^{\mathscr{A}}$ data as winning set:

$$\mathbb{W} = \left\{ d_{(1)}, d_{(2)}, \ldots, d_{(\mathfrak{R}^{\mathscr{A}})} \right\}. \tag{24}$$

Step 3. If original winning set $\mathbb{W}$ can satisfy the demand of the application, the algorithm comes to an end. If the original winning set $\mathbb{W}$ cannot satisfy the demand of the application, QUADG will delete the data with the lowest quality. And then choose the data with the highest Quality Utilization in the remaining set. Repeat this displacement process until the quality meets the demand of the application.

Step 4. Repeat Step 1 until all the grid areas have been traversed.

The algorithm of $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$ is similar. The algorithm of $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$ is symmetric with $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{P}^{\mathscr{A}}\}$ so Algorithm 1 only shows the detail of $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$.

(3) $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{P}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$. In this situation, the demand of user is tight. Data demander proposed the request for payment and quality at the same time. But the request of data

demander may not be satisfied at the same time. To quantify the satisfying degree of data demander, the index satisfaction θ is used to measure the demander's satisfaction:

$$\theta = \lambda \frac{\left(\mathscr{P}^{\mathscr{A}} - \mathscr{C}^{\mathscr{A}} \right)}{\mathscr{P}^{\mathscr{A}} + \mathscr{C}^{\mathscr{A}}} + (1 - \lambda) \frac{\left(\overline{Q^{\mathscr{A}}} - \mathscr{L}^{\mathscr{A}} \right)}{\mathscr{L}^{\mathscr{A}} + \overline{Q^{\mathscr{A}}}}. \tag{25}$$

θ is the data demander's satisfaction with the system allocation. When $\theta < 0$, it indicates that the data collection scheme recommended by the system is completely consistent with the expectation of data demander. $\theta < 0$ indicates that the system's allocation scheme does not meet data demander's expectation. $\theta > 0$ indicates that the system's allocation plan exceeded the user's expectations.

In order to facilitate the comparison of data, besides the overall satisfaction θ, for a single data, $d_m \in D_i$ system also computes the data demander satisfaction τ_m^i.

$$\tau_m^i = \lambda \frac{\left| \left(b_m^i - \mathscr{P}_i^{\mathscr{A}} \right) \right|}{\mathscr{P}_i^{\mathscr{A}}} + (1 - \lambda) \frac{\left| \left(r_m^i - \mathscr{L}_i^{\mathscr{A}} \right) \right|}{\mathscr{L}_i^{\mathscr{A}}}. \tag{26}$$

Then the algorithm steps are as follows.

Input: data requester profile: $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{P}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$
Output: Winner set $\mathbb{W}$, $T^{\mathscr{A}} = \{X^{\mathscr{A}}, \mathscr{C}^{\mathscr{A}}, \overline{Q^{\mathscr{A}}}\}$
(1) $Q^{\mathscr{A}} = 0, \mathscr{C}^{\mathscr{A}} = 0, \mathbb{W} = \emptyset$ // to initialize the winner set and the achieved quality and payment
(2) **For** each grid $i \mid \forall i \in A$
(3)   **For** n = 1 to $|D_i|$
(4) $r_{i,t}^{j,A} = b_{i,t}^{\mathscr{A},j}/r_{i,t}^{\mathscr{A},j}$ // to compute the Quality Utilization of each data in gird A.
(5)   **End for**
(6) $\mathbb{D}_i$ = Sort D_i according to r in ascending order
     // to get the ordered data set
(7) $\mathbb{R}_A$ = Sort $\{r_{i,t}^{j,A}\}$ in ascending order
(8)   **For** m = 1 to $\mathfrak{R}_A^{\mathscr{A}}$
(9) $x_{(m)}^{\mathscr{A},A,t} = 1$ // to mark the selected data
(10) $\mathbb{W} = \mathbb{W} \cup \{d_{(m)}\}$// to select the winning data
(11) $\underline{Q_i^{\mathscr{A}}} = Q_i^{\mathscr{A}} + r_{(m)}$
(12) $\overline{Q_i^{\mathscr{A}}} = \sum x_{(m)} r_{(m)}/X_i^{\mathscr{A}}$
(13) $\mathscr{C}_i^{\mathscr{A}} = \mathscr{C}_i^{\mathscr{A}} + b_{(m)}$
(14)   **End for**
(15) $\mathscr{b}_i = \lambda((\mathscr{P}_i^{\mathscr{A}} - \mathscr{C}_i^{\mathscr{A}})/\mathscr{P}_i^{\mathscr{A}}) + (1 - \lambda)((\overline{Q_i^{\mathscr{A}}} - \mathscr{L}_i^{\mathscr{A}})/\mathscr{L}_i^{\mathscr{A}})$
(16) $\tau_m^i = \lambda(|(b_m^i - \mathscr{P}_i^{\mathscr{A}})|/\mathscr{P}_i^{\mathscr{A}}) + (1 - \lambda)(|(r_m^i - \mathscr{L}_i^{\mathscr{A}})|/\mathscr{L}_i^{\mathscr{A}})$
(17) $m = \mathfrak{R}^{\mathscr{A}} + 1$;
(18)   **While** $\mathscr{b}_i > \beta$ && $m < X_i^{\mathscr{A}}$.
(19)   find d_e in $\mathbb{W}$ whose τ_e^A is biggest// to find the data which will be eliminated from the winner set
(20)   **If** $\tau_e^i > \tau_{(m)}^i$
(21)   delete d_e from $\mathbb{W}$
(22) $x_{(m)}^{\mathscr{A},A,t} = 1$// to mark the data which is newly selected
(23) $\mathbb{W} = \mathbb{W} \cup \{d_{(m)}\}$// to select the new data
(24)   Renew $\overline{Q_i^{\mathscr{A}}}, \mathscr{C}_i^{\mathscr{A}}$
(25) $m = m + 1$
(26)   **End if**
(27) **End while**
(28) **End for**
(29) $\mathscr{C}^{\mathscr{A}} = \sum \mathscr{C}_i^{\mathscr{A}}, \ \forall i \in A$
(30) $\overline{Q^{\mathscr{A}}} = \sum \overline{Q_i^{\mathscr{A}}}/\mathfrak{H}$
(31) $T^{\mathscr{A}} = \{X^{\mathscr{A}}, \mathscr{C}^{\mathscr{A}}, \overline{Q^{\mathscr{A}}}\}$
(32) **Return** $\mathbb{W}$, $T^{\mathscr{A}}$

ALGORITHM 2: The Quality Utilization Based Programming, $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{P}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$.

Step 1. Choose grid area i, $\forall i \in A$ and sort all the data collected in area i, $\forall i \in A$ according to r in descending order. The ordered Quality Utilization list can be denoted as

$$\mathbb{R}_i = \{r_{(1)}, r_{(2)}, \ldots, r_{(\mathfrak{R}^{\mathscr{A}})}, \ldots, r_{(n)}\}. \tag{27}$$

The corresponding ordered data list can be denoted as

$$\mathbb{D}_i = \{d_{(1)}, d_{(2)}, \ldots, d_{(\mathfrak{R}^{\mathscr{A}})}, \ldots, d_{(n)}\}. \tag{28}$$

Step 2. Choose first $\mathfrak{R}_i^{\mathscr{A}}$ data as winning set:

$$\mathbb{W} = \{d_{(1)}, d_{(2)}, \ldots, d_{(\mathfrak{R}^{\mathscr{A}})}\}. \tag{29}$$

Step 3. If original winning set $\mathbb{W}$ can satisfy the demand of the application, the algorithm comes to an end. If the original winning set $\mathbb{W}$ cannot satisfy the demand of the application, QUADG will delete the data with the lowest satisfaction $\mathscr{b}$. And then choose the data with highest Quality Utilization in the remaining set. Repeat this displacement process until the satisfaction degree meets the demand of the application.

Step 4. Repeat Step 1 until all the grid areas have been traversed.

Algorithm 2 shows the details of $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{P}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$.

5. Experimental Evaluation

In this section, we compare our algorithm QUADG with RSDG (Random Selection Data Gathering) and DDODG (Data Demand Only Data Gathering). RSDG randomly selects the data from the random list. The goal of DDODG is to make the allocation as close as possible to the demand of the application, not to maximize the Quality Utilization. When comparing the differences between QUADG, DDODG, and RSDG, we first analyze the differences from a single grid sensing area under the different requirements of data demander and then compare the differences overall all grid sending area of the city overall.
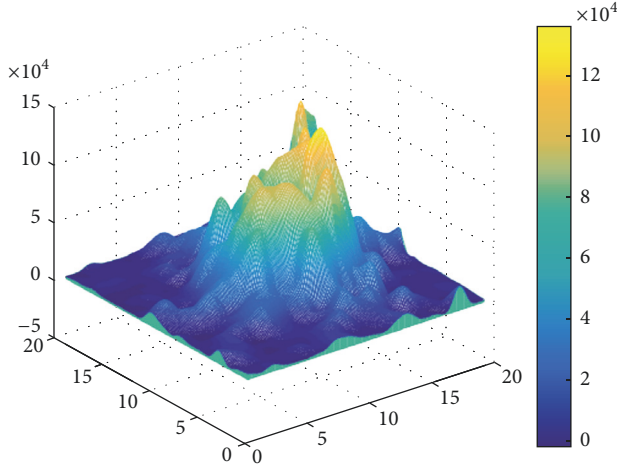
FIGURE 3: The frequency of vehicles.

FIGURE 4: The distribution of vehicles in the city.

The data set used in our experiment is T-Drive trajectories, which is provided by MSRA [68]. The original data used in the experiment is Beijing taxi GPS track data. The data packet includes approximately 10357 taxis' files. Each file is a GPS track data of a taxi within a week during the period between February 2 and February 8, 2008, in Beijing. We divide the city into grids by latitude and longitude. And according to the coordinates of vehicles, we count the frequency of taxi in each grid in time slot of a week.

5.1. The Result of Data Preprocessing. We first divide the city into grid and add up the appearance times of taxi tab in each grid; we noted it as frequency of the tab.

Before the optimization, we must filter the invalid data in the files. If the distance between the current coordinate and the last coordinate is too far, the current coordinate may be the wrong data. The speed in most places of Beijing is limited to 80 km/h. If the average speed between current coordinate and the last coordinate is bigger than the limited speed, we can regard the current coordinate as an invalid data

The three-dimensional figure of the frequency of the tab is shown as Figure 3. From Figure 3 we can observe that the frequencies of vehicles in the center of the city are far more than the frequency in the grids near the boundary. Most of vehicles appear in the longitude rage of 116.3, 116.55 and latitude rage of 39.8, 39.9. If there are not any vehicles in the area, the area might be developing.

Figure 4 is the planar graph of the frequency of the tab. From Figure 4, we could learn the trajectory of taxi tab. The icon in Figure 4 is the placed with the highest frequency which is Beijing Capital Airport. The data amount is far higher than the other area.

5.2. The Estimation of a Certain Grid i. The bid and the data quality of each data reporter can be generated by the Normal Distribution. We assume that the amount of submitted data of gird area i, $\forall i \in A$ is 100 for convenience. The generated quality of data reporter is shown in Figure 5. The generated bid of each data reporter is shown in Figure 6. And the
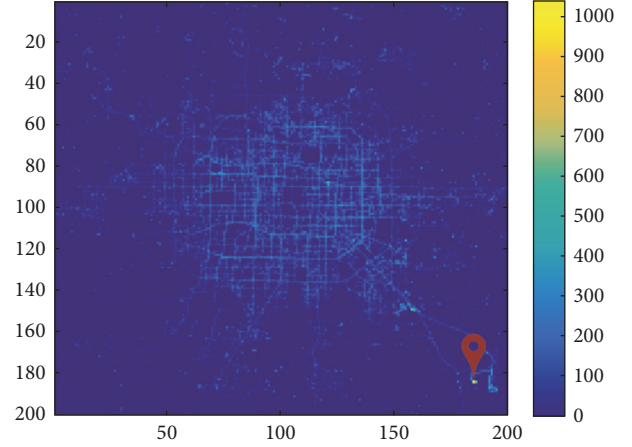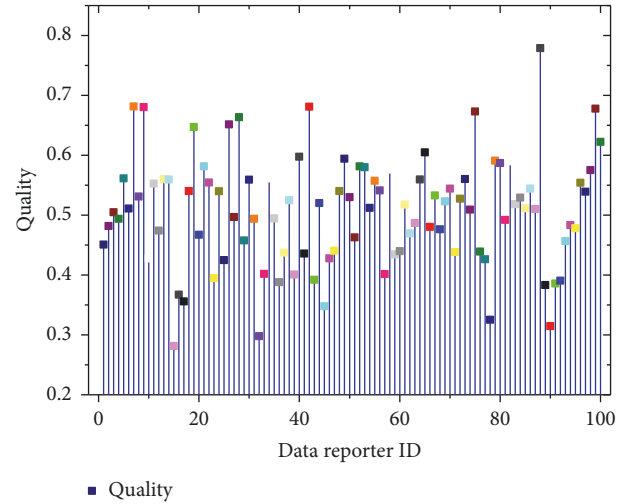
FIGURE 5: The quality of 100 data reporters. The colored boxes, other than "the black box," refer to the different value of quality and bid.
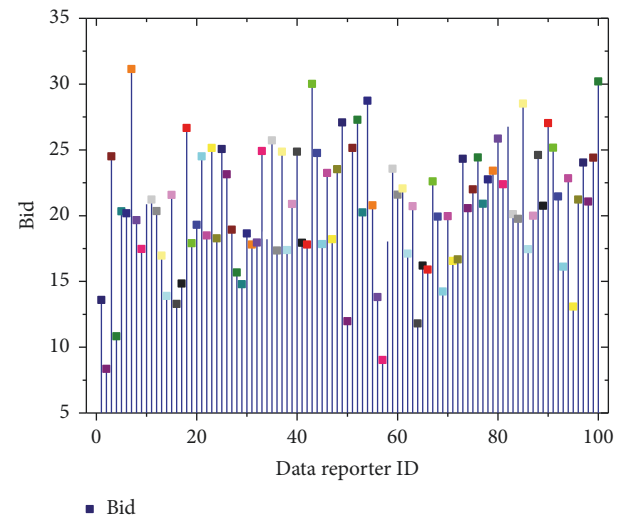
FIGURE 6: The bids of 100 data reporters. The colored boxes, other than "the blue box," refer to the different value of bid.
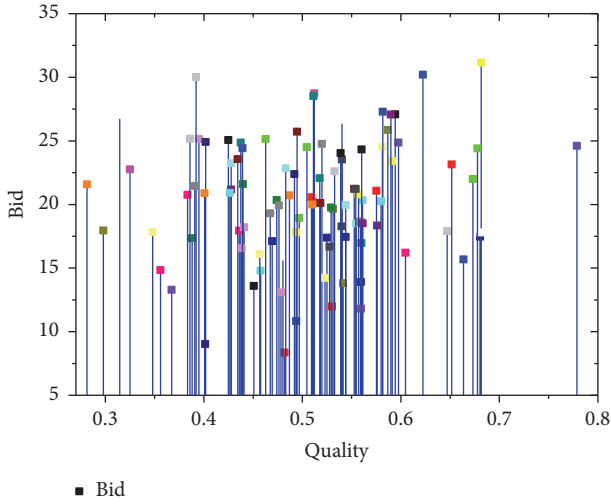
FIGURE 7: The relationship between quality and bid. The colored boxes, other than "the black box," refer to the different value of quality and bid.
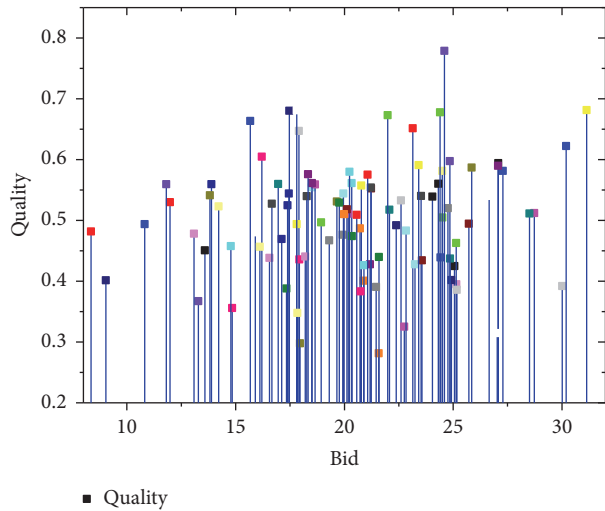
FIGURE 8: The relationship between bid and quality. The colored boxes, other than "the black box," refer to the different value of quality and bid.

FIGURE 9: The comparison of quality and average payment.
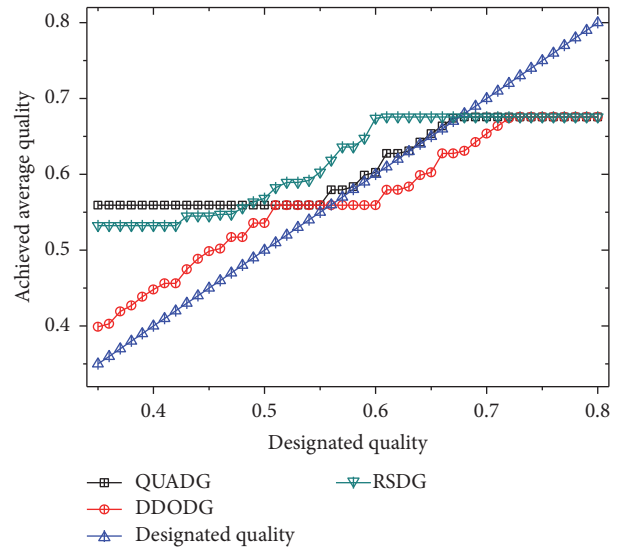
FIGURE 10: The comparison of achieved data quality.

relationship between quality and bid is shown in Figures 7 and 8. Figure 7 shows that most data quality is distributed in the interval $[0.35, 0.62]$. Figure 8 shows that most bid is distributed in the interval $[11, 22]$. And next we will evaluate the validity of the algorithm in a single grid area i, $\forall i \in A$.

5.2.1. $U_i = \{\mathfrak{R}_i^{\mathscr{A}}\}$. Firstly, we evaluate the situation when the data demander only makes request for the data amount. Different applications have different requirements for the amount of data collected in different sensing areas. Some applications need same amount of data in each sensing grid area such as environment monitoring application and haze detection applications. And some other applications need different data amount in different sensing grid area. For example, the real-time traffic applications need more data in

larger traffic area than less traffic area. QUADG is suitable for the two situations. But to simplify the comparison of different scheme, we assume the amount of needed data in each sensing grid area i, $\forall i \in A$ is 5. The sensing areas whose data amount submitted by the data reporter is less than 5 are ignored. It is impossible to raise data quality by selection in the sensing area where data amount is too small. The only way to raise data quality is to encourage more people to participate in the data collection.

The achieved quality and payment by QUADG and RSDG are shown in Figure 9. With QUADG quality has been raised by 34.28% and the payment has been reduced by 45.76%.

5.2.2. $U_i = \{\mathfrak{R}_i^{\mathscr{A}}, \mathscr{L}_i^{\mathscr{A}}\}$. We compare the task allocation result of QUADG, DDODG, and RSDG. The result of the three strategies for designating different quality is shown in Figures 10 and 11. Figure 10 shows the achieved quality of
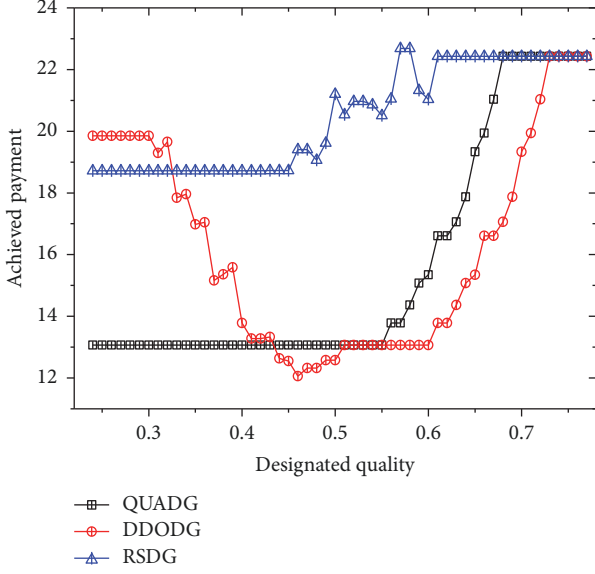
FIGURE 11: The comparison of achieved payment.

FIGURE 12: The comparison of the achieved quality.

three strategies by designating different quality. And Figure 11 shows the needed payment of three strategies by designating quality. Figures 10 and 11 illustrate that DDODG adds a lot of overhead in order to be closer to the application's needs when the specified data quality falls within a lower range. QUADG's data quality at this time is always beyond the user's expectation, but the corresponding payment is far less than RSDG. From Figure 11, we could see that when the designated quality is low, the payment of DDOGDG is high. To meet the request, DDODG will choose the data with lower Quality Utilization when the quality designated by data demander is low. So only getting closer to the request of data demander does not make sense.

5.2.3. $U_i = \{\mathfrak{R}_i^{\mathscr{A}}, \mathscr{P}_i^{\mathscr{A}}\}$. We compare the task allocation result of QUADG, DDODG, and RSDG. The achieved quality of three algorithms by designating different payment is shown in Figure 12. And the achieved payment of three algorithms by designating different payment is shown in Figure 13. Figures 12 and 13 illustrate that quality of DDODG could be close to the requirement of the data demander when the designated payment is low. But it must be noted that the quality of DDODG decreases when the payment is high. The reason why the payment of DDODG will decrease is that the algorithm chooses the data with high bid and low quality to fit the request of the data demander. Although the payment of the data demander increases, the achieved quality of DDODG will decrease. QUADG has the lowest payment in three algorithms. Although the quality of QUADG is lower than the DDODG in part of the interval, the achieved quality is stable as the payment increases. QUADG could reduce the payment redundancy at most because when the payment comes to a certain extent, the overall quality will not increase which brings the redundancy. The payment of RSDG strategy is the highest, but the quality is the lowest.
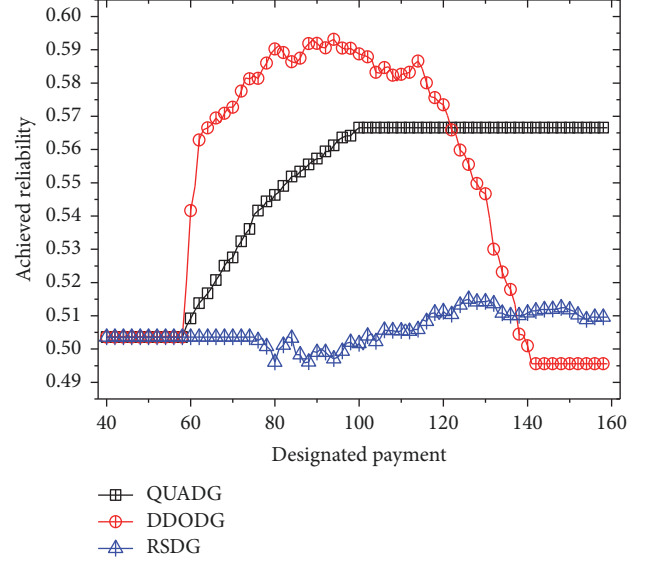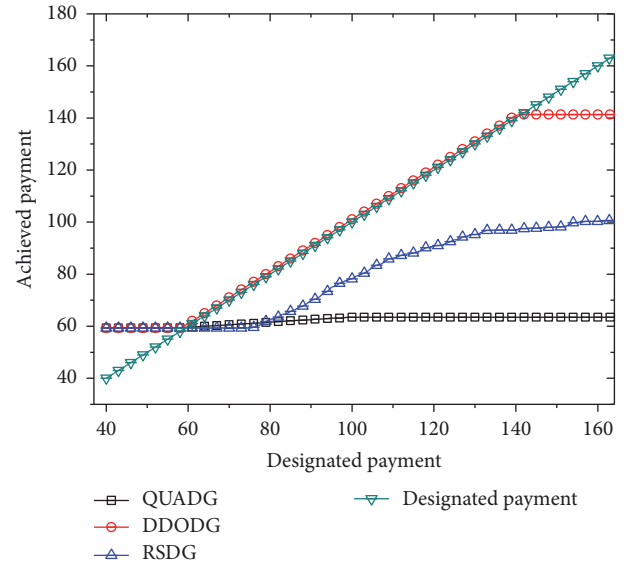
FIGURE 13: The comparison of the achieved payment.

5.2.4. $U_i = \{\mathfrak{R}_i^{\mathscr{A}}, \mathscr{P}_i^{\mathscr{A}}, \mathscr{L}_i^{\mathscr{A}}\}$. System will allocate the task of data collection with satisfaction $\mathscr{b}$ which is set by data demander or the system when data requester assigned the amount of collected data, payment, and the amount of data.

We first evaluate the validity of satisfaction $\mathscr{b}$. The achieved payment and achieved quality by designating different satisfaction $\mathscr{b}$ are shown in Figure 14 when the $U_i = \{5, 10, 0.5\}$ by QUADG. From Figure 14, we could learn that the lower the value of satisfaction $\mathscr{b}$, the higher quality and the lower the payment of data demander, which means the satisfaction $\mathscr{b}$ is valid.

When the satisfaction $\mathscr{b}$ is designated as 0.1 and $\mathscr{L}_A^{\mathscr{A}} = 0.5$ the achieved quality and the achieved payment by designating different payment are shown in Figures 15 and 16, respectively.
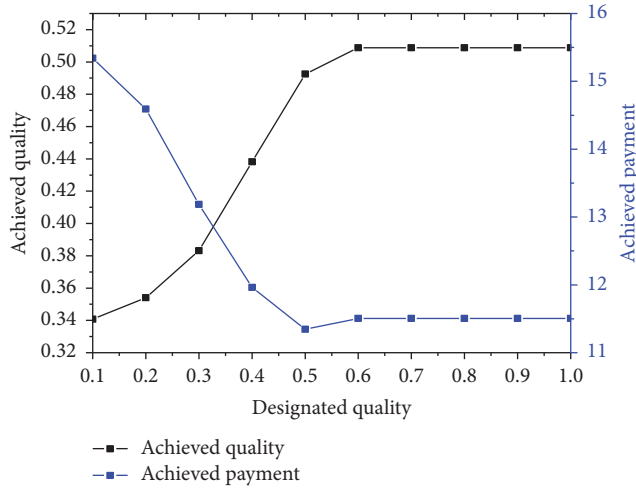
FIGURE 14: The achieved quality and payment under different deviate.
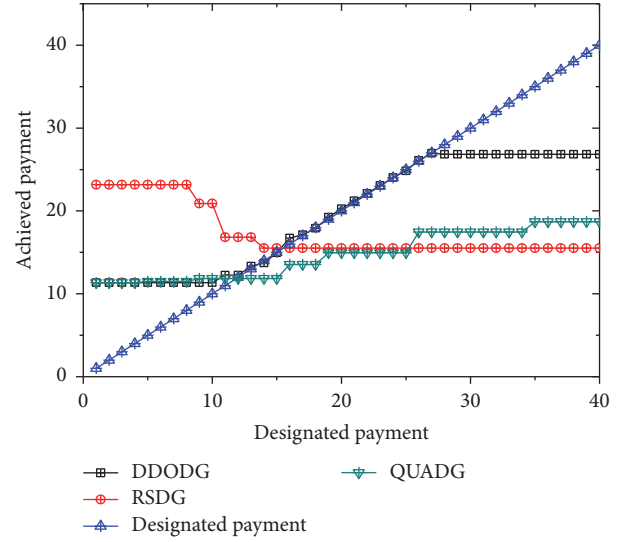
FIGURE 16: The achieved payment by designating payment.
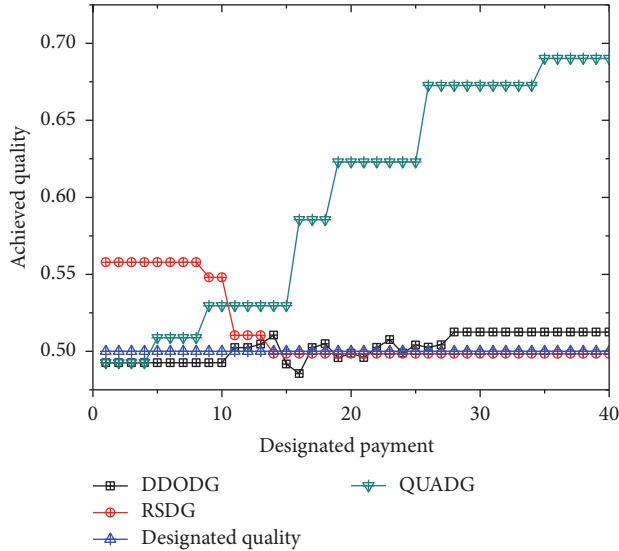
FIGURE 15: The achieved quality by designating payment.
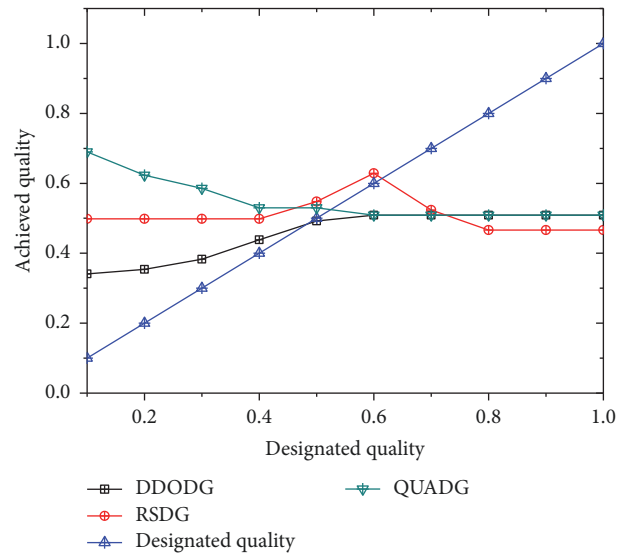
FIGURE 17: The achieved quality by designating quality.

When $\mathscr{P}_i^{\mathscr{A}} < 10$, the quality of QUADG is lower than RSDG. When the $\mathscr{P}_i^{\mathscr{A}}$ designated by the data demander lies in the interval $(10, 27)$, the quality achieved by QUADG is far higher than RSDG and DDODG because the QUADG strategy is an optimal allocation that combines quality and payment. If the DDODG strategy is used which does not consider maximizing Quality Utilization, simply taking the requirements of data demander as goal will increase unnecessary overhead.

When the satisfaction $\mathscr{E}$ is designated as 0.1 and $\mathscr{P}_i^{\mathscr{A}}$ is designated as 10, the achieved quality and the achieved payment by designating different quality are shown in Figures 17 and 18. Due to the nature of the random selection, the payment of RSDG is far more than DDODG and QUADG. The QUADG strategy has the similar effect as DDODG when the requirements of data demander are beyond the system boundary. However, the QUADG strategy can achieve high

quality when data demander designated low quality. The overall effect of QUADG strategy is better than DDODG strategy and RSDG strategy.

5.3. The Estimation of All Grids

5.3.1. $U_i = \{\mathfrak{R}_i^{\mathscr{A}}\}$. When $\mathfrak{R}_i^{\mathscr{A}}$ is designated, the achieved quality and payment of each grid by QUADG are shown in Figures 19 and 20. The achieved quality and payment of each grid by RSDG are shown in Figures 21 and 22. Figures 19 and 20 show the quality of each grid is similar when the amount of data $\mathfrak{R}_i^{\mathscr{A}}$ is designated. But the payment of the grid in the center is lower than the grid at the edge of the city. That is because the amount of data in the center grid is bigger and the choice space is larger. So, the payment in the center grid
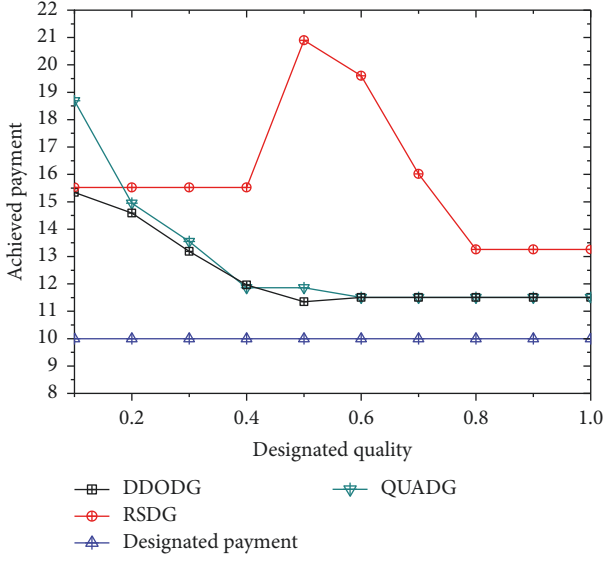
Figure 18: The achieved payment by designating quality.

Figure 21: The overall quality of RSDG.

Figure 19: The overall quality of QUADG.
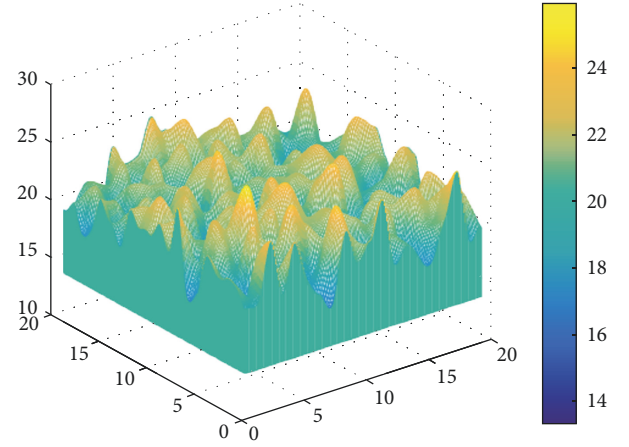
Figure 22: The overall payment of RSDG.

Figure 20: The overall payment of QUADG.

is lower. In the grid area where the amount of data is not much enough, the system can only select the limited data reporter to assign tasks. So, the payment is higher than the grid where data amount is bigger.

Figures 20 and 21 show that there is no obvious law between the achieved payment and the achieved quality. RSDG cannot achieve lower payment in the area where the data amount is big because RSDG makes the allocation randomly. The comparison of average quality and the average payment of QUADG and RSDG is shown in Figure 23. QUADG quality has been raised by 10.52% and the payment has been reduced by 80.50%.

5.3.2. $U_i = \{\mathfrak{R}_i^{\mathcal{A}}, \mathscr{L}_i^{\mathcal{A}}\}$. If both $\mathfrak{R}_i^{\mathcal{A}}$ and $\mathscr{L}_i^{\mathcal{A}}$ are designated, to compare more conveniently we first assume the requested quality of each grid is 0.5 and the amount of requested data is 5. The quality of each grid by QUADG is shown in Figure 24. There are two grid areas which have less data than the demand of demander. So, the achieved quality cannot satisfy the request of demander.

From Figures 25 and 29, we could see that the payment by QUADG in the center area is less when the requested
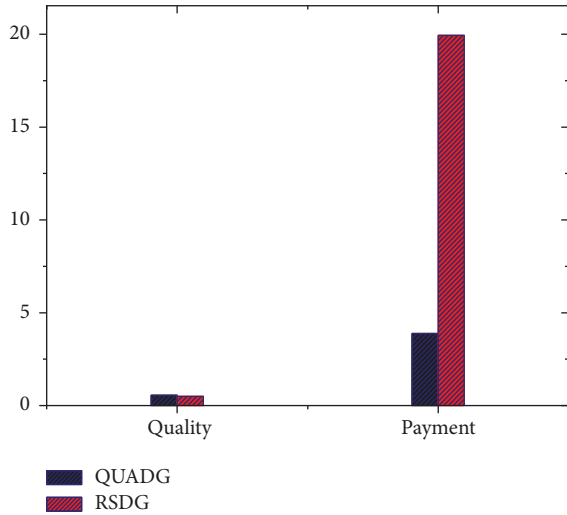
FIGURE 23: The comparison of quality and payment.
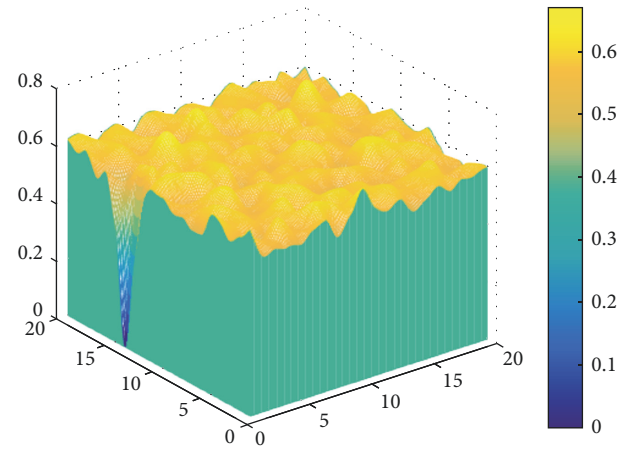
FIGURE 26: The quality of RSDG.
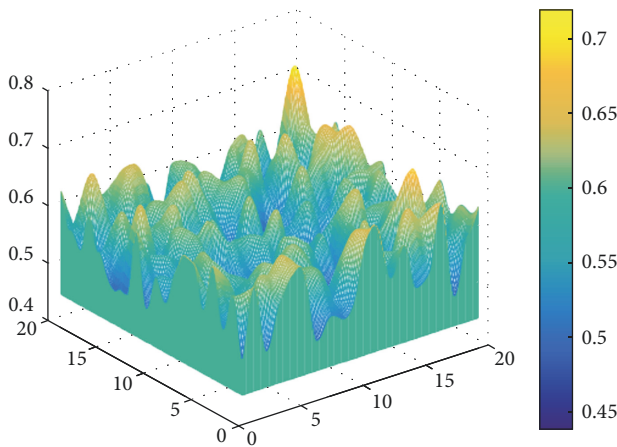
FIGURE 24: The quality of QUADG.
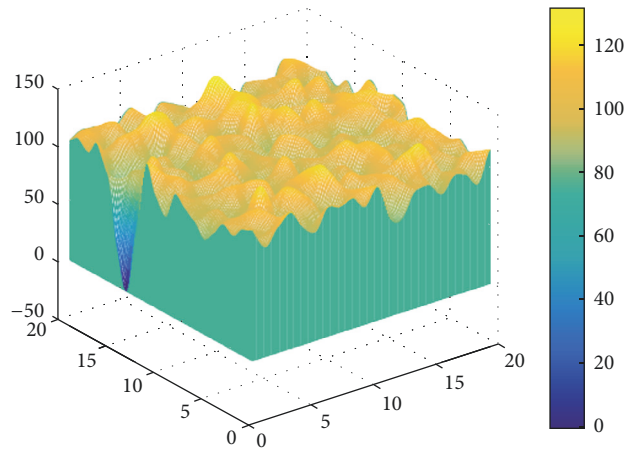
FIGURE 27: The payment of RSDG.
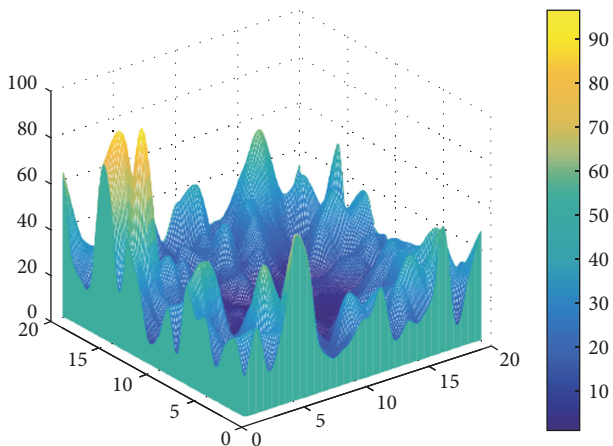
FIGURE 25: The payment of QUADG.

data amount in each grid is the same. That is because when the number of candidate data is big, there will be more room for the system to choose. When the number of workers participated in the data collection is lower, the payment for obtaining high quality is higher.

Figures 25, 27, and 29 illustrate that RSDG strategy has the additional overhead. Even in areas with a large amount of data, the cost of collected data will not be reduced. Figures 24, 26, and 28 illustrate that the DDODG strategy achieves much lower data quality than the RDBP strategy among the three strategies.

The comparison of the achieved quality and the achieved payment of three algorithms by designating different quality is shown in Figures 30 and 31. We noted that when the requested quality lies in the interval $[0, 0.47]$ where the corresponding data reporter amount is little, DDODG can satisfy the demander's quality request. But from Figure 31 we could see that to meet the request, DDODG will choose the data with lower Quality Utilization when the quality designated by data demander is low. The quality low bound of QUADG is 0.55. If the requested quality is lower than 0.55, the quality will still be 0.55. But the payment will not
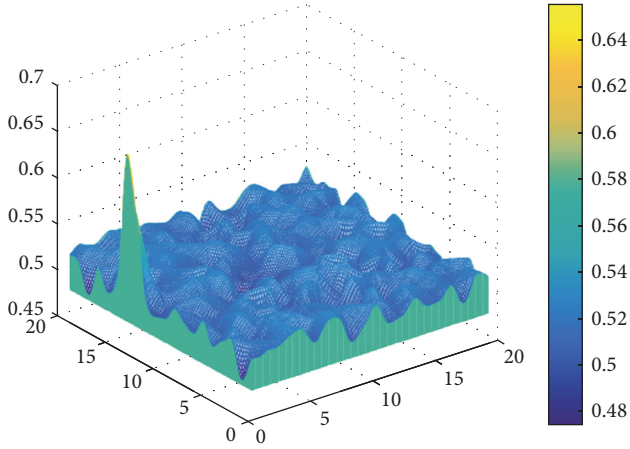
FIGURE 28: The quality of DDODG.
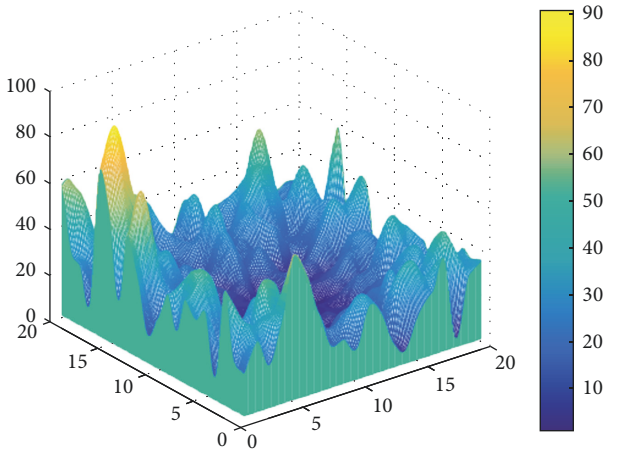
FIGURE 29: The payment of DDODG.
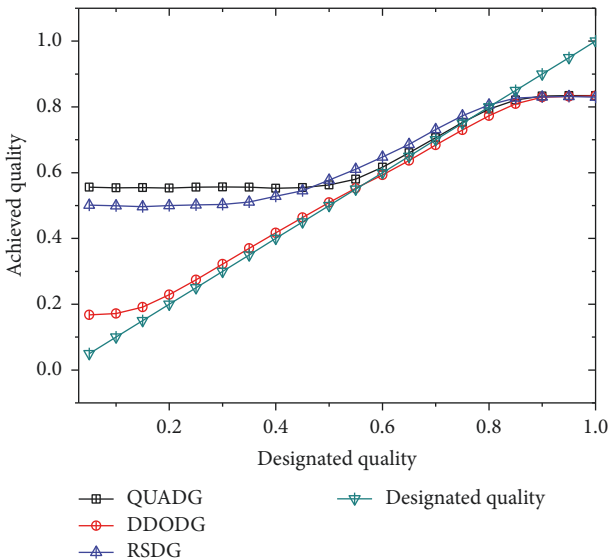
FIGURE 30: The comparison of achieved quality.
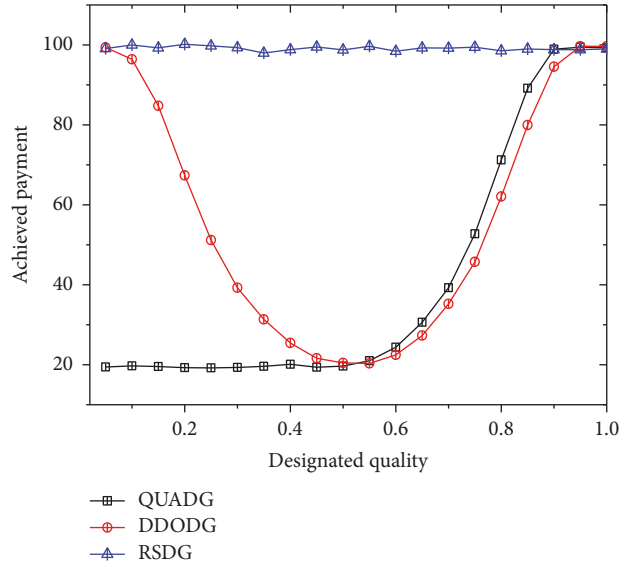
FIGURE 31: The comparison of achieved payment.
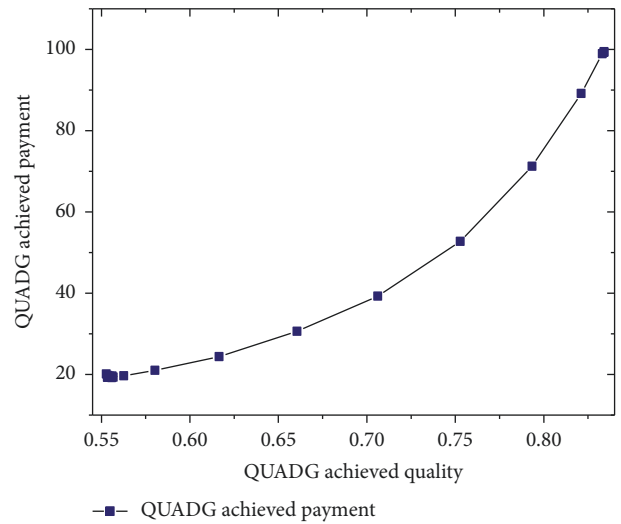
FIGURE 32: The relationship between achieved quality and payment with QUADG.

increase. Figures 34, 35, and 36 show the relationship between the achieved quality and the achieved payment of QUADG, DDODG, and RSDG, respectively. Figures 32, 33, and 34 show that there is a positive correlation between the achieved quality and the achieved payment when QUADG quality is greater than 0.55. But, DDODG have high payment when the achieved quality is low. There is no obvious correlation between the achieved payment and the achieved quality of the RSDG. The efficiency of RSDG is lower than QUADG.

5.3.3. $U_i = \{\mathfrak{R}_i^{\mathscr{A}}, \mathscr{P}_i^{\mathscr{A}}\}$. If both $\mathfrak{R}_i^{\mathscr{A}}$ and $\mathscr{P}_i^{\mathscr{A}}$ are designated, to compare more conveniently we assume the requested payment of each data is 5 and the amount of requested data is 5.
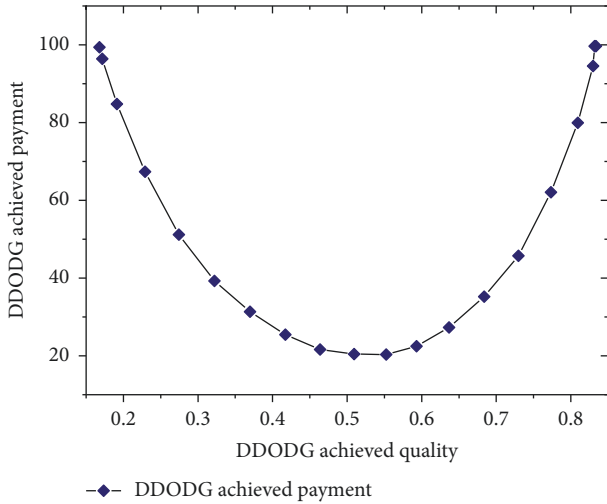
FIGURE 33: The relationship between achieved quality and payment with DDODG.

FIGURE 36: The quality of QUADG.

FIGURE 34: The relationship between achieved quality and payment with RSDG.
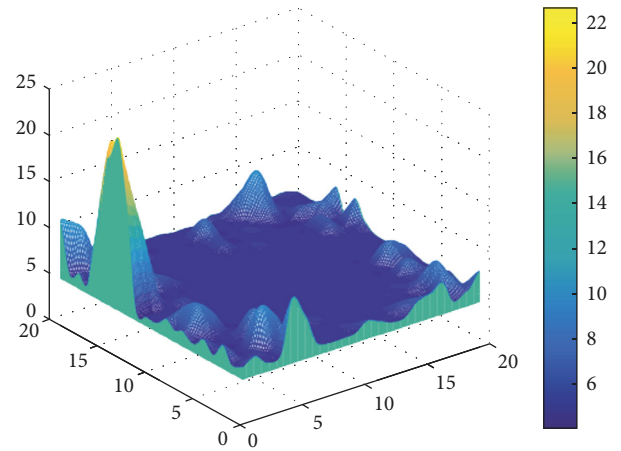
FIGURE 37: The payment of DDODG.

FIGURE 35: The payment of QUADG.

The achieved payment and the achieved quality of QUADG are shown in Figures 35 and 36, respectively. The achieved payment and the achieved quality of DDODG are shown in Figures 37 and 38, respectively. The achieved payment and the achieved quality of RSDG are shown in Figures 39 and 40, respectively.

As can be seen from Figures 35, 37, and 39, the payment of the RSDG strategy is the highest. The goal of the DDODG strategy is to as close as possible to the data demander's requirement, so the three dimensions of DDODG are smoother. From Figures 36, 38, and 40 we could see that the overall level of the quality of RSDG is lower than that of QUADG and DDODG.

The comparison of the achieved quality and the achieved payment of three strategies by designating different quality is shown in Figures 41 and 42. It can be seen from Figure 41 that the payment of the QUADG strategy is much lower than the RSDG and DDODG strategies. Figure 42 shows the quality results under the three strategies with designated payment. Although the quality of DDODG is high, the achieved quality gradually decreases when the payment of data demander gradually increases. To get close to the request
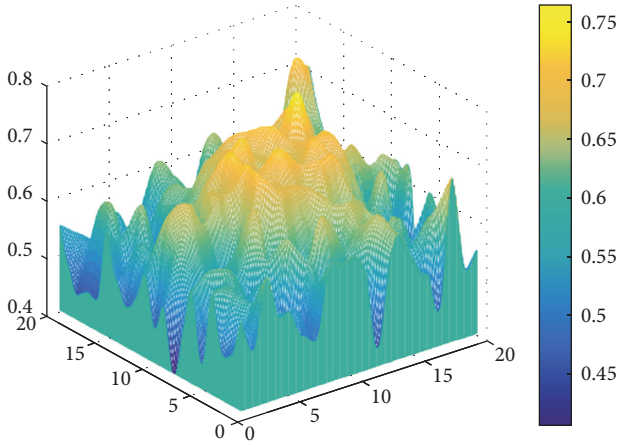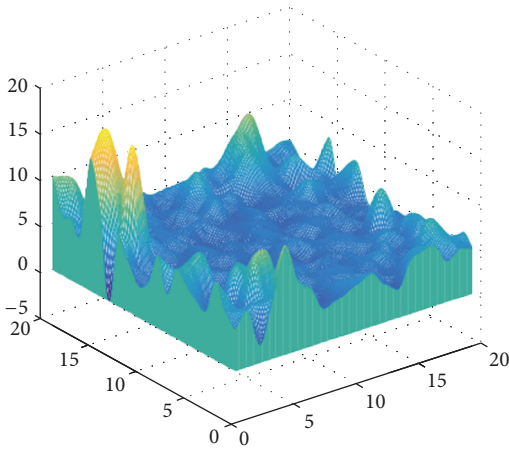
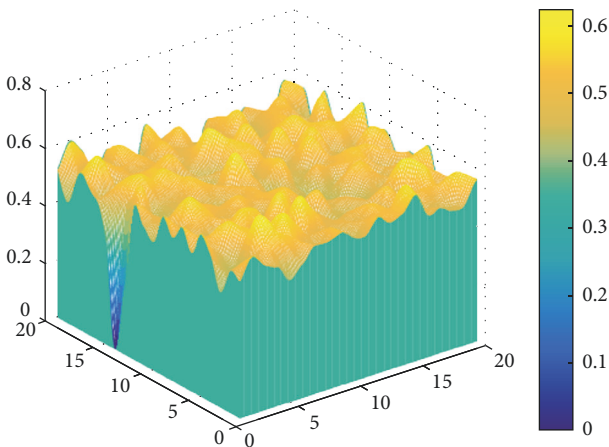FIGURE 38: The quality of DDODG.

FIGURE 39: The payment of RSDG.

FIGURE 40: The quality of RSDG.

FIGURE 41: The achieved average payment by designating payment.
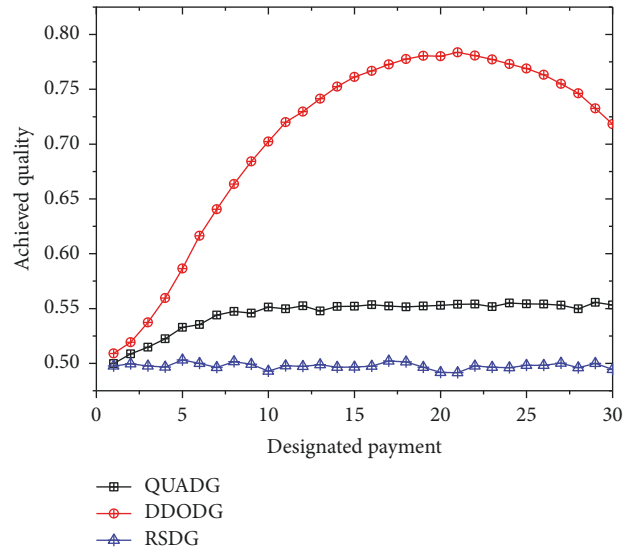
FIGURE 42: The achieved quality by designating payment.

of data demander, DDODG strategy will choose the data with the higher bid despite data quality. The QUADG strategy can be used to obtain the best solution of the combination of payment and quality when the data demander only makes a request for the payment.
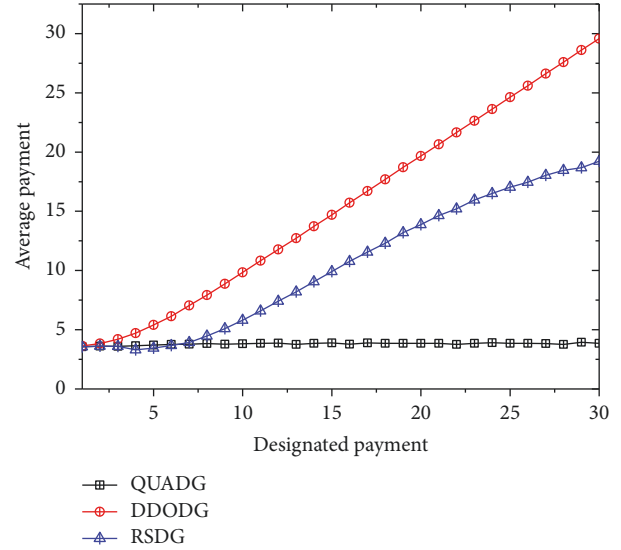
5.3.4. $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{P}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$. To simplify the process of comparison, we first assume the average payment of a single data is 10. And the requested quality is 0.45.

The achieved quality and the achieved payment of each grid by QUADG are shown in Figures 43 and 44, respectively. The achieved quality and the achieved payment of each grid by DDODG are shown in Figures 45 and 46, respectively. The achieved quality and the achieved payment of each grid by QUADG are shown in Figures 47 and 48, respectively. From Figures 43, 45, and 47, it can be seen that the QUADG strategy achieves the highest data quality. The goal of the DDODG strategy is to get as close as possible to the data demander's requirement, so the quality in each grid is similar. As can be seen from Figures 44, 46, and 48, with QUADG, the payment in the center of city is lower where data amount is bigger. It
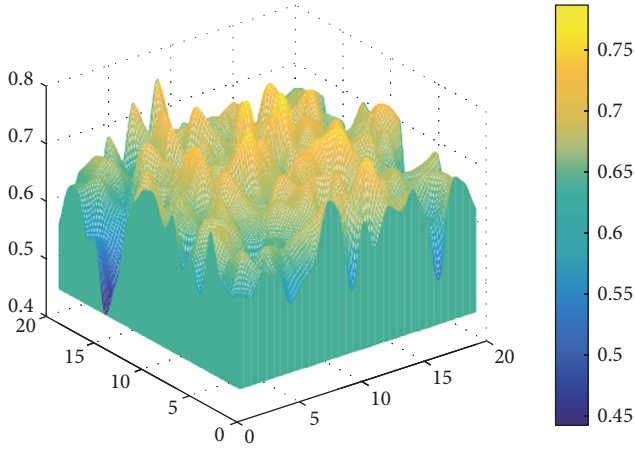
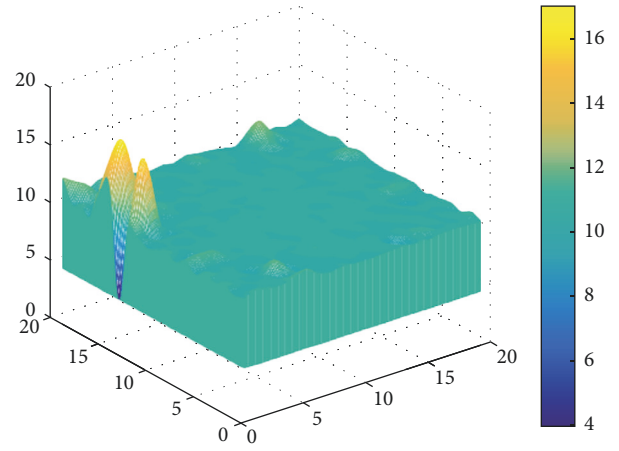FIGURE 43: The achieved quality of QUADG.

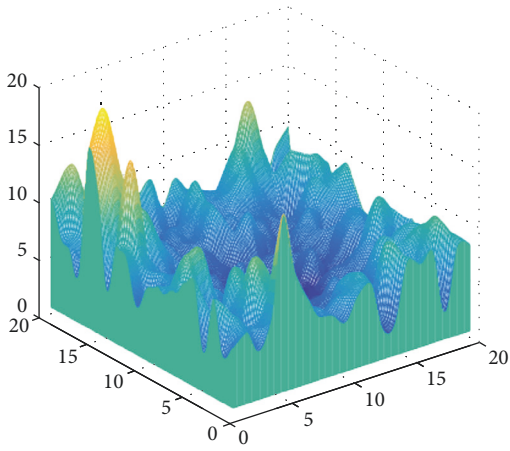FIGURE 46: The achieved payment of DDODG.

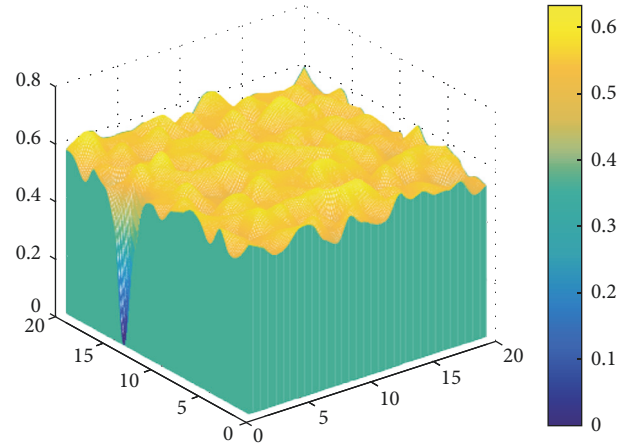FIGURE 44: The achieved payment of QUADG.

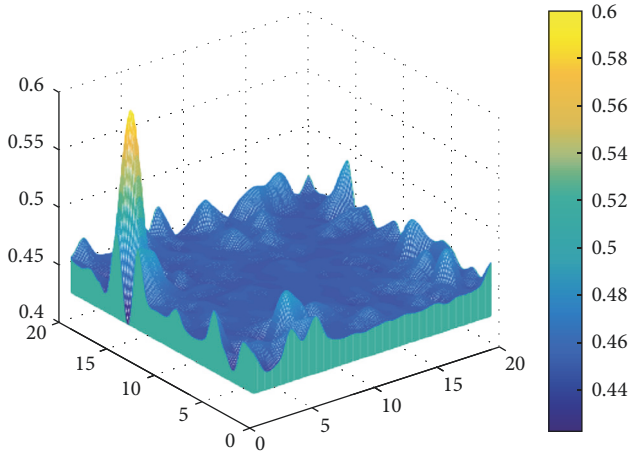FIGURE 47: The achieved quality of RSDG.
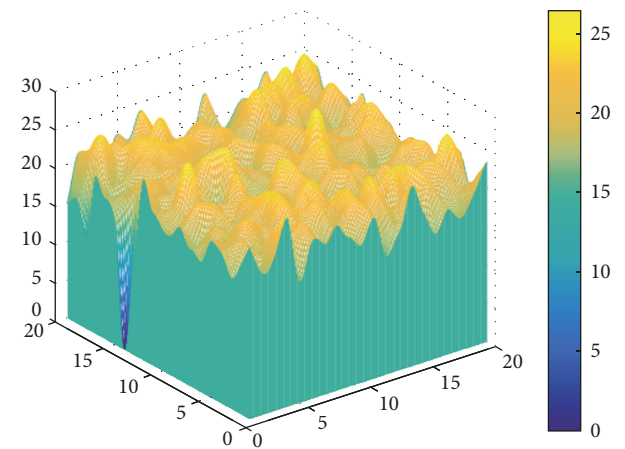

FIGURE 45: The achieved quality of DDODG.

FIGURE 48: The achieved payment of RSDG.

shows that QUADG strategy can effectively reduce the data redundancy when data amount is sufficient.

And next, the control variate method is used to compare the differences between QUADG strategy, DDODG

strategy, and RSDG strategy. Figures 49 and 50 are the results obtained under the conditions of fixed payment and changing quality. In Figures 49 and 50, the data quality
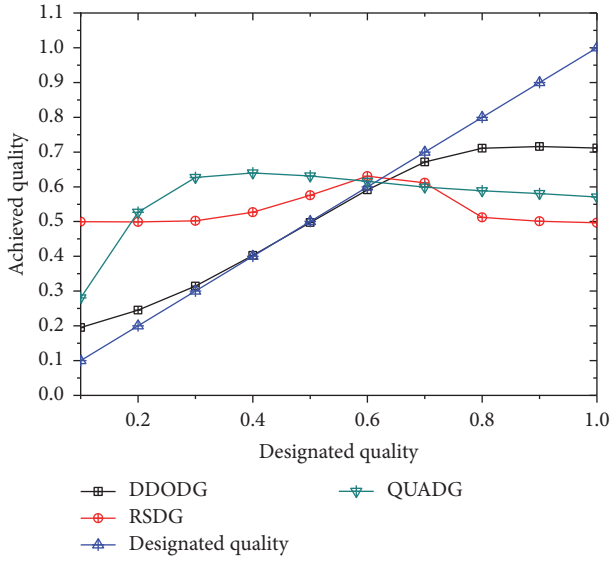
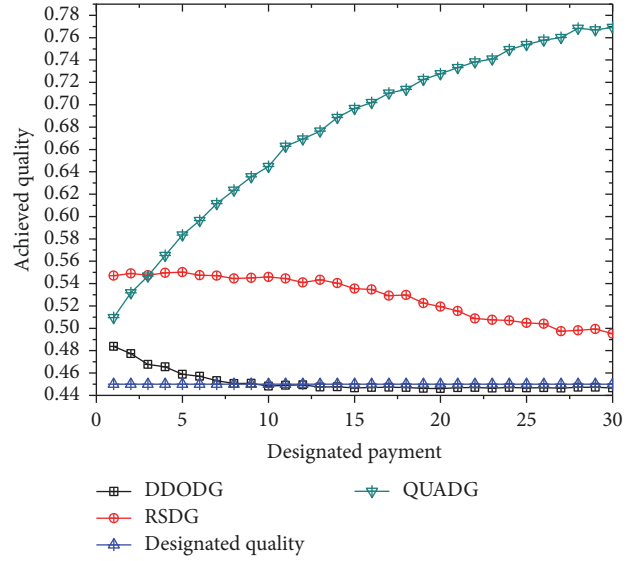FIGURE 49: The achieved quality by designating quality.

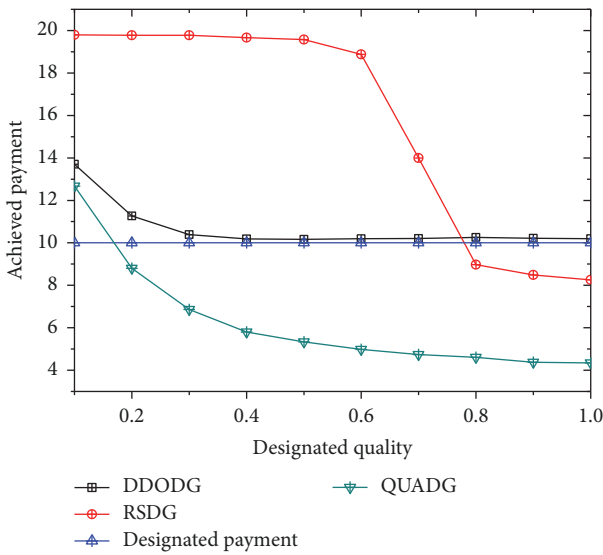FIGURE 51: The achieved quality by designating payment.

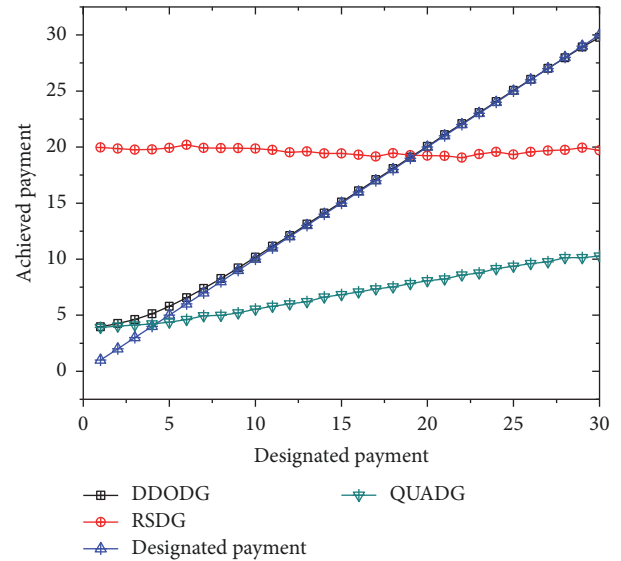FIGURE 50: The achieved payment by designating quality.

FIGURE 52: The achieved payment by designating payment.

of QUADG strategy is the highest when the data quality specified by the data demander is less than 0.6. When the data-demander-specified data quality is greater than 0.6, although quality of the DDODG strategy is higher than QUADG, the corresponding payment of QUADG is far less than the DDODG strategy. The RSDG strategy is the worst in both data payment and data quality. Figures 51 and 52 are the results obtained under the conditions of fixed quality and changing payment. From Figures 51 and 52, comparted with DDODG and RSDG, it can be seen that data demander could get the highest data quality with the lowest payment. QUADG is better than RSDG and DDODG in both data quality and payment.

6. Conclusion

In this paper, QUADG has been proposed to optimize the quality and payment in the process of data collection. In this scheme, the information will be collected when vehicles pass by the area near sensor nodes. The data will be submitted to data processing center when vehicles pass through the center. To encourage more people to participate in the process of collecting data, vehicle data reporter can propose the bid of his own data.

Most intensive mechanisms always concentrate on the payment of data gathering when using auction mechanism as intensive mechanism. But the data quality is ignored when system selects data randomly in a traditional way. In this paper, quality-density has been raised to select the data with

high quality and low bid. Data demander could designate the quality and payment according to the requirement of the application. System will allocate the data collection tasks according to the request of data demander. We compare QUADG with the common scheme RSDG and DDODG under four situations. And it has been proved that QUADG is valid.

Parameter Description

D_i: The data set collected in grid area i, $\forall i \in A$

$r_{i,t}^{j,A}$:   The Quality Utilization of data collected by data reporter j in grid area i, $\forall i \in A$ at time t

$b_{i,t}^{A,j}$:   The bid of data collected by data reporter j in grid area i, $\forall i \in A$ at time t

$r_{i,t}^{A,j}$:   The quality of data collected by data reporter j in grid area i, $\forall i \in A$ at time t

$\mathfrak{R}^{A}$:   The designated data amount of application A by data demander

$\mathscr{P}^{A}$:   The designated payment of application A by data demander

$\mathscr{L}^{A}$:   The designated quality of application A by data demander

$\mathbb{R}_i$:   The ordered list of Quality Utilization in grid i, $\forall i \in A$

$\mathbb{D}_i$:   The list of data in grid i, $\forall i \in A$ ordered by the Quality Utilization

$\mathscr{C}$:   The satisfaction of data demander towards the QUADG strategy

τ_m^i: The satisfaction of data demander towards the single data

$\mathbb{W}$:   The winning data set.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

References

[1] N. Zhang, S. Zhang, P. Yang, O. Alhussein, W. Zhuang, and X. S. Shen, "Software Defined Space-Air-Ground Integrated Vehicular Networks: Challenges and Solutions," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 101–109, 2017.

[2] N. Lu, N. Cheng, N. Zhang, X. S. Shen, J. W. Mark, and F. Bai, "Wi-Fi Hotspot at Signalized Intersection: Cost-Effectiveness for Vehicular Internet Access," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3506–3518, 2016.

[3] Z. Song, C. H. Liu, J. Wu, J. Ma, and W. Wang, "QoI-aware multitask-oriented dynamic participant selection with budget constraints," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 9, pp. 4618–4632, 2014.

[4] M. Wu, Y. Wu, X. Liu, M. Ma, A. Liu, and M. Zhao, "Learning-based synchronous approach from forwarding nodes to reduce the delay for Industrial Internet of Things," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, 2018.

[5] D. Zeng, L. Gu, L. Lian, S. Guo, H. Yao, and J. Hu, "On cost-efficient sensor placement for contaminant detection in water distribution systems," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2177–2185, 2016.

[6] J. Gui and K. Zhou, "Flexible adjustments between energy and capacity for topology control in heterogeneous wireless multi-hop networks," *Journal of Network and Systems Management*, vol. 24, no. 4, pp. 789–812, 2016.

[7] T. Wang, J. Zeng, Y. Lai et al., "Data collection from WSNs to the cloud based on mobile Fog elements," *Future Generation Computer Systems*, 2017.

[8] X. Liu, "A novel transmission range adjustment strategy for energy hole avoiding in wireless sensor networks," *Journal of Network and Computer Applications*, vol. 67, pp. 43–52, 2016.

[9] X. Liu, N. Xiong, N. Zhang, A. Liu, H. Shen, and C. Huang, "A trust with abstract information verified routing scheme for cyber-physical network," *IEEE Access*, 2018.

[10] A. Thiagarajan, L. Ravindranath, K. LaCurts et al., "VTrack: accurate, energy-aware road traffic delay estimation using mobile phones," in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys '09)*, pp. 85–98, November 2009.

[11] *Waze - outsmarting traffic, together*, 2013, http://www.waze.com/.

[12] Buuuk, *WeatherLah iPhone application*, 2012, http://itunes.apple.com/us/app/weatherlah/id411646329?mt=8.

[13] N. Maisonneuve, M. Stevens, M. E. Niessen, and L. Steels, "NoiseTube: Measuring and mapping noise pollution with mobile phones," in *Information Technologies in Environmental Engineering*, Environmental Science and Engineering, pp. 215–228, Springer Berlin Heidelberg, Heidelberg, Gremany, 2009.

[14] S. He, D. Shin, J. Zhang, J. Chen, and P. Lin, "An exchange market approach to mobile crowdsensing: pricing, task allocation, and walrasian equilibrium," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 4, pp. 921–934, 2017.

[15] G. Yang, S. He, and Z. Shi, "Leveraging crowdsourcing for efficient malicious users detection in large-scale social networks," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 330–339, 2017.

[16] X. Duan, C. Zhao, S. He, P. Cheng, and J. Zhang, "Distributed algorithms to compute walrasian equilibrium in mobile crowdsensing," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 4048–4057, 2017.

[17] M. Marjanović, L. Skorin-Kapov, K. Pripužić, A. Antonić, and I. P. Žarko, "Energy-aware and quality-driven sensor management for green mobile crowd sensing," *Journal of Network and Computer Applications*, vol. 59, pp. 95–108, 2016.

[18] A. Antonić, M. Marjanović, K. Pripužić, and I. P. Žarko, "A mobile crowd sensing ecosystem enabled by CUPUS: cloud-based publish/subscribe middleware for the Internet of Things," *Future Generation Computer Systems*, vol. 56, pp. 607–622, 2016.

[19] Y. Gao, Y. Chen, and K. J. R. Liu, "On cost-effective incentive mechanisms in microtask crowdsourcing," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 1, pp. 3–15, 2015.

[20] T. Han and N. Ansari, "Provisioning green energy for base stations in heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5439–5448, 2016.

[21] N. Zhang, S. Zhang, J. Zheng, X. Fang, J. W. Mark, and X. Shen, "QoE Driven Decentralized Spectrum Sharing in 5G Networks: Potential Game Approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 7797–7808, 2017.

[22] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: solutions and challenges," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 289–299, 2014.

[23] C. H. Liu, J. Fan, P. Hui, J. Wu, and K. K. Leung, "Toward QoI and energy efficiency in participatory crowdsourcing," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4684–4700, 2014.

[24] T. Wang, J. Zeng, M. Z. Bhuiyan et al., "Trajectory privacy preservation based on a fog structure for cloud location services," *IEEE Access*, vol. 5, pp. 7692–7701, 2017.

[25] F. Ma, X. Liu, A. Liu, M. Zhao, C. Huang, and T. Wang, "A Time and Location Correlation Incentive Scheme for Deep Data Gathering in Crowdsourcing Networks," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 8052620, 22 pages, 2018.

[26] N. Zhang, N. Cheng, A. T. Gamage, K. Zhang, J. W. Mark, and X. Shen, "Cloud assisted HetNets toward 5G wireless networks," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 59–65, 2015.

[27] H. Li, D. Liu, Y. Dai, T. H. Luan, and S. Yu, "Personalized search over encrypted data with efficient and secure updates in mobile clouds," *IEEE Transactions on Emerging Topics in Computing*, vol. PP, no. 99, 2016.

[28] Z. Su, Q. Xu, M. Fei, and M. Dong, "Game theoretic resource allocation in media cloud with mobile social users," *IEEE Transactions on Multimedia*, vol. 18, no. 8, pp. 1650–1660, 2016.

[29] T. Wang, Y. Li, G. Wang, J. Cao, M. Z. Bhuiyan, and W. Jia, "Sustainable and efficient data collection from WSNs to cloud," *IEEE Transactions on Sustainable Computing*, 2017.

[30] J. Tang, A. Liu, M. Zhao, and T. Wang, "An Aggregate Signature Based Trust Routing for Data Gathering in Sensor Networks," *Security and Communication Networks*, vol. 2018, Article ID 6328504, 30 pages, 2018.

[31] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment framework for participatory sensing data collections," in *Pervasive Computing*, vol. 6030 of *Lecture Notes in Computer Science*, pp. 138–155, Springer, Berlin, Germany, 2010.

[32] M. Huang, A. Liu, T. Wang, and C. Huang, "Green data gathering under delay differentiated services constraint for internet of things," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 9715428, 2018.

[33] L. Zheng and L. Chen, "Maximizing acceptance in rejection-aware spatial crowdsourcing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1943–1956, 2017.

[34] A. Bazzi and A. Zanella, "Position based routing in crowd sensing vehicular networks," *Ad Hoc Networks*, vol. 36, pp. 409–424, 2016.

[35] G. Merlino, S. Arkoulis, S. Distefano, C. Papagianni, A. Puliafito, and S. Papavassiliou, "Mobile crowdsensing as a service: a platform for applications on top of sensing clouds," *Future Generation Computer Systems*, vol. 56, pp. 623–639, 2016.

[36] S. Maharjan, Y. Zhang, and S. Gjessing, "Optimal Incentive Design for Cloud-Enabled Multimedia Crowdsourcing," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2470–2481, 2016.

[37] H. Gao, C. H. Liu, W. Wang et al., "A survey of incentive mechanisms for participatory sensing," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 918–943, 2015.

[38] X. Hu, T. H. S. Chu, V. C. M. Leung, E. C.-H. Ngai, P. Kruchten, and H. C. B. Chan, "A survey on mobile social networks: applications, platforms, system architectures, and future research directions," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1557–1581, 2014.

[39] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 140–150, 2010.

[40] W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang, and G. Wang, "Security and Privacy in the Medical Internet of Things," *Security and Communication Networks*, vol. 2018, Article ID 5978636, 2018.

[41] H. Zhang, Z. Cai, Q. Liu, Q. Xiao, Y. Li, and C. Cheang, "A Survey on Security-aware Measurement in SDN," *Security and Communication Networks*, vol. 2018, Article ID 2459154, 2018.

[42] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, "Vita: a crowdsensing-oriented mobile cyber-physical system," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 148–165, 2013.

[43] C. H. Liu, B. Zhang, X. Su, J. Ma, W. Wang, and K. K. Leung, "Energy-aware participant selection for smartphone-enabled mobile crowd sensing," *IEEE Systems Journal*, vol. PP, no. 99, 2015.

[44] C.-K. Tham and T. Luo, "Quality of contributed service and market equilibrium for participatory sensing," *IEEE Transactions on Mobile Computing*, vol. 14, no. 4, pp. 829–842, 2015.

[45] A. Liu, Z. Chen, and N. N. Xiong, "An adaptive virtual relaying set scheme for loss-and-delay sensitive WSNs," *Information Sciences*, vol. 424, pp. 118–136, 2018.

[46] C. Hu, M. Li, D. Zeng, and S. Guo, "A survey on sensor placement for contamination detection in water distribution systems," *Wireless Networks*, 2016.

[47] H. Xin and X. Liu, "Energy-balanced transmission with accurate distances for strip-based wireless sensor networks," *IEEE Access*, vol. 5, pp. 16193–16204, 2017.

[48] K. Ota, M. Dong, J. Gui, and A. Liu, "QUOIN: Incentive Mechanisms for Crowd Sensing Networks," *IEEE Network Magazine*, 2017.

[49] D. Zeng, P. Li, S. Guo, T. Miyazaki, J. Hu, and Y. Xiang, "Energy Minimization in Multi-Task Software-Defined Sensor Networks," *IEEE Transactions on Computers*, vol. 64, no. 11, pp. 3128–3139, 2015.

[50] X. Liu, "Node Deployment Based on Extra Path Creation for Wireless Sensor Networks on Mountain Roads," *IEEE Communications Letters*, vol. 21, no. 11, pp. 2376–2379, 2017.

[51] *ICT Facts and Figures 2016*, 2017, http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf.

[52] *Economic operation of Telecom industry November 2016*, 2016, http://www.miit.gov.cn/n1146285/n1146352/n3054355/n3057511/n3057518/c5426194/content.html.

[53] J. Gui and J. Deng, "Multi-hop Relay-Aided Underlay D2D Communications for Improving Cellular Coverage Quality," *IEEE Access*, 2018.

[54] P. Dutta, P. M. Aoki, N. Kumar et al., "Demo abstract: Common sense - Participatory urban sensing using a network of handheld air quality monitors," in *Proceedings of the 7th ACM Conference*

on Embedded Networked Sensor Systems, SenSys 2009, pp. 349-350, USA, November 2009.

[55] S. Kim, C. Robson, T. Zimmerman, J. Pierce, and E. M. Haber, "Creek watch: pairing usefulness and usability for successful citizen science," in *Proceedings of the 29th Annual CHI Conference on Human Factors in Computing Systems (CHI '11)*, pp. 2125–2134, ACM, Vancouver, Canada, May 2011.

[56] M. Mun, S. Reddy, K. Shilton et al., "PEIR, the personal environmental impact report, as a platform for participatory sensing systems research," in *Proceedings of the 7th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys '09)*, pp. 55–68, June 2009.

[57] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Ear-phone: an end-to-end participatory urban noise mapping system," in *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN '10)*, pp. 105–116, Stockholm, Sweden, April 2010.

[58] S. B. Eisenman, E. Miluzzo, N. D. Lane, R. A. Peterson, G.-S. Ahn, and A. T. Campbell, "BikeNet: a mobile sensing system for cyclist experience mapping," *ACM Transactions on Sensor Networks*, vol. 6, pp. 87–101, 2009.

[59] E. Miluzzo, N. D. Lane, K. Fodor et al., "Sensing meets mobile social netwo rks: the design, implementation and evaluati on of the cenceme application," in *Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems (SenSys '08)*, pp. 337–350, ACM, New York, NY, USA, November 2008.

[60] J.-S. Lee and B. Hoh, "Sell your experiences: A market mechanism based incentive for participatory sensing," in *Proceedings of the 8th IEEE International Conference on Pervasive Computing and Communications, PerCom 2010*, pp. 60–68, April 2010.

[61] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava, "Examining micro-payments for participatory sensing data collections," in *Proceedings of the 12th International Conference on Ubiquitous Computing, UbiComp 2010*, pp. 33–36, Denmark, September 2010.

[62] X. Fang, J. Tang, D. Yang, and G. Xue, "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, August 2012.

[63] L. G. Jaimes, I. Vergara-Laurens, and M. A. Labrador, "A location-based incentive mechanism for participatory sensing systems with budget constraints," in *Proceedings of the 10th IEEE International Conference on Pervasive Computing and Communications, (PerCom '12)*, pp. 103–108, March 2012.

[64] D. Yang, X. Fang, and G. Xue, "Truthful incentive mechanisms for k-anonymity location privacy," in *Proceedings of the IEEE INFOCOM*, pp. 2994–3002, Turin, Italy, April 2013.

[65] G. Bigwood and T. Henderson, "IRONMAN: using social networks to add incentives and reputation to opportunistic networks," in *Proceedings of the IEEE 3rd International Conference on Social Computing (SocialCom '11)*, pp. 65–72, Boston, Mass, USA, October 2011.

[66] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work?—a literature review of empirical studies on gamification," in *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS '14)*, pp. 3025–3034, Waikoloa, Hawaii, USA, January 2014.

[67] N. Lu, T. H. Luan, M. Wang, X. Shen, and F. Bai, "Bounds of asymptotic performance limits of social-proximity vehicular networks," *IEEE/ACM Transactions on Networking*, vol. 22, no. 3, pp. 812–825, 2014.

[68] S. He, J. Chen, F. Jiang, D. K. Y. Yau, G. Xing, and Y. Sun, "Energy provisioning in wireless rechargeable sensor networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 10, pp. 1931–1942, 2013.