Received January 12, 2019, accepted February 1, 2019, date of publication February 5, 2019, date of current version February 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2897615

Deployment Optimization of Data Centers in Vehicular Networks

BAIXIANG HUANG[1], WEI LIU[2], TIAN WANG[3], XIONG LI[4], HOUBING SONG[5], (Senior Member, IEEE), AND ANFENG LIU[1]

[1]School of Information Science and Engineering, Central South University, Changsha 410083, China
[2]School of Informatics, Hunan University of Chinese Medicine, Changsha 410208, China
[3]College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China
[4]School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China
[5]Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA

Corresponding author: Anfeng Liu (afengliu@mail.csu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61772554, Grant 61572528, and Grant 6157256.

ABSTRACT Due to the ubiquitous utilization of GPS devices, traffic cameras, and sensing devices, data are collected more readily in a smart city. If all the cabs of this city are used as data carriers, data generated by sensing devices will be collected to data centers efficiently and economically. Therefore, vehicular networks joint sensing devices provide more perspectives for a variety of vehicle-based applications. In this paper, three data centers deployment optimization schemes are proposed to optimize data centers' deployment, which can considerably enhance the performance of data collection and code dissemination tasks. Each proposed scheme uses different criteria to optimize the deployment of data centers: 1) locations of sensing devices; 2) locations that have high traffic flow; and 3) locations that have high valid flow, and these schemes are called Scheme 1, Scheme 2, and Scheme 3, respectively. In addition, another scheme that produces the deployment of data centers randomly is called Scheme 4, and it is used as a contrast in experiments. After performing extensive experiments and simulations based on two real-world datasets of cabs' GPS coordinates, the experiment results demonstrate that Scheme 3 noticeably outperformed remaining schemes under various circumstances. The results of February 3 were taken (Dataset 1) for instance. In comparison with Scheme 4, Scheme 3 enhances the total number of collected data packets by 57.71% when the number of the data center is ten, the speed of code dissemination and the coverage of cabs are ameliorated by 23.92% and 12.93%, respectively. Compared with Scheme 1, the total number of Scheme 3's collected packets is 65.00% higher than that of Scheme 1 when there are ten data centers deployed. The figures for Scheme 3's code dissemination and the coverage of sensing devices exceed that of Scheme 1 by 18.98% and 10.21%, respectively.

INDEX TERMS Vehicular networks, data collection, data centers, code dissemination, smart city.

I. INTRODUCTION

Along with the dramatic development of electronic technologies, various kinds of sensing devices have been invented, and they have become increasingly powerful and cheaper. Sensing devices are widely used in a variety of applications to build a smart city [1]–[4], from which the status of infrastructures can be collected and used to subsidize the decision making regarding infrastructures. According to [5], the number of devices (the majority of them are sensing-based devices) connected to the Internet of Things on the earth has reached 9 billion since 2011. It is estimated that by the year 2020, the number of devices connected to the Internet will reach 24 billion [6], [7]. Such a trend has brought many far-reaching impacts. In the initial place, regulators are able to obtain all kinds of data through sensing devices more easily. Both the coverage and the number of collected data have witnessed a huge leap, which enables the regulators of infrastructures to react to sensed data more swiftly. As a result, a city becomes smarter [8]–[11]. On the other hand, by analyzing the underlying patterns of these data, considerable improvements can be made associated with transportation management [3], [7], [12], urban planning, epidemic control, and mobile platform applications [13]–[16].

The associate editor coordinating the review of this manuscript and approving it for publication was Xiangjie Kong.

The task of sensing devices is to obtain ubiquitous data [3], [7], [17]–[19]. However, how to deliver the collected data from sensing devices to data centers effectively remains a challenging issue [12], [17]. Some efforts of collecting data from sensing devices using Vehicular Networks (VN) have been made by some researchers [3], [4], [20], [21]. Vehicular Networks and their applications have recently attracted much attention from both industry and academia [3], [4], [20], [21]. With the pervasive applications of intelligent equipment such as GPS devices, traffic cameras, smart cards, smartphones, and road deceleration devices, multisource big data can be collected more easily than ever before [22]–[24]. On the other hand, the number of vehicles in urban areas is also immense. Many cosmopolitans contain more than one million cars, which move around all day long restlessly. For example, Beijing has more than 544 million vehicles. Vehicles are also capable of collecting data, and they have strong communication capacity [3], [4], [20], [21]. Thus, the way of collecting data could be revolutionized by using moving vehicles to obtain data from infrastructures embedded with sensing devices [3], [4], [20], [25], [26], and a new perspective of collecting data in a swift, economical way is provided. Following that, those obtained data are integrated into an online platform where data can be further processed and analyzed [27]–[30].

A few previous research had tried to implement the scenarios mentioned above. To collect and sense a variety of useful data, the number of deployed sensing devices is ever-increasing [3], [4], [20], [21], [25], [26], [31]. For instance, sensing devices are embedded into street lamps; they are deployed inside garbage bins to obtain their filling level [31]; they are also installed in important constructions like bridges, theaters, and roads to record their physical deformations [25], [26]. In other applications such as monitoring status of trees and green belts, sensing devices can report the temperature and humidity information to the relevant regulators so that they can decide how often to irrigate vegetation [25], [26]. A common attribute of these applications is that sensing devices spread broadly across the entire city, which makes them hard to collect. Moreover, their distribution is sparse and dynamic, which means that deploying a static network connected to the Internet would be expensive and inefficient. Also, it takes a lot of time to deploy such a static network. For example, to monitor physical status of a road that is being constructed, sensing devices needed to be added to the road as the construction goes on, the locations and the number of garbage bins may also change under different circumstances [25], [26]. Therefore, deploying a static network is not feasible [25], [26].

Bonola *et al.* [31] proposed a scheme that uses cabs as data mules. Nowadays, a large number of cabs are equipped with onboard communication equipment due to the dispatching purpose. The equipment can be used to collect data from sensing devices and transmit collected data to data centers. Thus, when a cab passes an infrastructure embedded with sensing devices, the status data about the infrastructure are transmitted to this cab through short-distance communication technologies such as Bluetooth and 802.15.4. Later, when the cab passes a data center, all the data stored in its buffer are transmitted to the data center. In this way, data produced by ubiquitous sensing devices are collected efficiently. In comparison with traditional methods, such a scheme has the following advantages: (1) Oblivious. Cabs work on a 24/7 basis, and the paths that they travel cover the vast majority of roads and streets in a city. Therefore, the behavior of cabs does not have to be intervened. (2) Economical. It would be expensive to hire a fleet of vehicles to perform data collection and code dissemination tasks. However, using cabs to perform these tasks is economical since their behavior does not need to be intervened. Also, unlike wireless sensor networks, there is no need to deploy a network connected to the Internet. (3) Adaptive. The movement of cabs is highly adaptive to the changes of circumstance quickly. For instance, if a new residential area is constructed, then cabs will visit that place more often and interact with newly deployed sensing devices. If many residences left one area for holidays, the number of times cabs visit that area will decline. Vice versa, if many tourists come to a city, infrastructures around tourist spots and hotels will produce more data. Since tourists are likely to take cabs to these places, cabs will pass sensing devices located there more frequently.

The ways of data collection are further investigated in [25] and [26]. To increase the quality of collected data, these researchers tried to reduce redundancy rate by assigning priorities to sensing devices. Since sensing devices that are located in city centers are visited more frequently, the data collected from city centers tend to be redundant. Thus these devices are assigned with low priority. In contrast, data generated by sensors located in suburban areas are difficult to collect because those devices are less visited by vehicles. To collect data from suburbs more easily, sensing devices located there are assigned with high priority. When new data are received, and the buffer of a cab is full, data with low priority are replaced by high-priority data.

Meanwhile, with the emerging of a technology called software-defined networking [20], [21], code can be loaded to hardware and enable it to modify the settings or gain new functions after running the received code. This technology can reduce the cost and time it takes to replace old hardware. It can provide more flexibility for sensing devices since it is possible for them to upgrade or modify their functions such as collecting different types of data, or adding simple data-processing functions to themselves. Facilitated by the software-defined networking technology, normal sensing devices are upgraded into smart sensing devices [32]. This technology can significantly reduce the cost and time of deploying new devices, and it can also extend the life-span of existing sensing devices. Ren *et al.* [20] and Liu *et al.* [21] proposed a scheme to conduct code dissemination using vehicles. They focused on the scarcity of cabs' storage size and made optimization for it.

According to the analyses above, the colossal amount of sensing devices and vehicles can be combined into a powerful network to address many existing problems. Specifically, those applications that can benefit from vehicular networks are consisted of following procedures: (1) sensing devices sense and perceive useful data; (2) sensing devices transmit the collected data to passing cabs; (3) cabs upload the collected data to data centers through opportunistic communications; (4) the collected data are analyzed by regulators and can be later provided to public for academic research. Besides, data centers are able to publish and disseminate new code to sensing devices: (1) a piece of code is written to modify the settings or upgrade the functions of sensing devices, and the new code is published to data centers; (2) the code is transmitted to cabs when they enter the communication range of data centers; (3) cabs disseminate the code to sensing devices that they pass; (4) sensing devices upgrade their functions by running the received code. Hence, vehicular networks joint sensing devices have a promising perspective.

Previous research mainly focuses on the processes of data collection or code dissemination solely, and most of the previous work is based on the assumption that data centers are already deployed. In contrast, however, this work focus on the deployment of data centers. Moreover, our work considers both data collection and code dissemination tasks and optimize both of them. On the one hand, the optimization of data centers' deployment can noticeably enhance the speed and efficiency of data collection at a low cost. On the other hand, it can also reduce the delay of code dissemination, which makes sensing devices more intelligent. While inefficient deployment of data centers can render a prolonged process of code dissemination, which causes the incongruity among upgraded sensing devices and devices that have not received the code. The incongruities are illustrated in the preciseness, sampling frequency, and the number of data that needed to be collected. Therefore, there would be a period of time when new code and outdated code both exist. Our goal is to reduce the length of such a period of incongruity. Although previous research tried to promote the performance of code dissemination, they did not address this issue by adjusting the deployment of data centers. The deployment of data centers is not only of considerable significance for the efficiency of data collection but also for the speed of code dissemination. In this paper, three novel Data Centers Deployment Optimization (DCDO) schemes are proposed to address the challenges mentioned above. The principal innovations of this paper are:

1. Three DCDO schemes are proposed. The optimization of data centers' deployment consists of two aspects: (1) Location of data centers. The proposed schemes can optimize the location of data centers by using a clustering algorithm, and each of them use three different criteria: (a) Locations of smart devices; (b) Locations that have high traffic flow; (c) Locations with high valid flow (valid flow is explained in detail in Section IV.A). Clustering algorithms are particularly suitable to determine the location of data centers due to some attributes of the distribution of smart devices. For instance, sensing devices embedded in garbage bins are located more intensely in residential and central business area. To obtain the optimal locations of data centers, the clustering algorithm use locations obtained using various criteria as input, and these locations are clustered into different clusters, the centroids of which are the locations of data centers. Since the traffic speed is limited in a city, it is not feasible to reduce the delay of collecting data by increasing the speed of cabs. Besides, cabs move obliviously, so their ordinary behavior is not intervened. Therefore, enhance the performance of data collection by adjusting the deployment of data centers is a more feasible strategy. If a data center are deployed into better locations, data will be collected and uploaded more swiftly. Additionally, the speed of code dissemination will also see a rise. (2) Total number of data centers. It is indisputable that the sum of collected data packets is positively correlated to the quantity of data centers. Nevertheless, according to the experiment results, we found that when the number of data centers increases to 10 to 30, the growth rate of total data packets becomes noticeably slower. In other words, once the number of deployed data center reach a certain level, deploying more data centers becomes uneconomical. Thus, one of the goals of this paper is to obtain the appropriate number of data centers that can guarantee high performance at low cost.

2. The proposed DCDO schemes are able to ameliorate the performance of both data collection and code dissemination tasks by optimizing the deployment of data centers. None of previous works address these two issues jointly. For instance, Fang *et al.* [32], Tang *et al.* [25], and Xu *et al.* [26] did research to collect data more swiftly and cover more broadly. Ren *et al.* [20] and Liu *et al.* [21] only considered how to disseminate code to a network. The DCDO schemes are able to facilitate the data collection and code dissemination tasks simultaneously. Therefore it is of high practical values.

3. The evaluations and simulations of the proposed schemes are based on two real-life trajectories datasets. They have 7,871,498 and 17,478,202 GPS points within approximately 400 square kilometers of areas that are chosen to perform simulations. After conducting extensive experiments, the optimal deployment of data centers is obtained. Moreover, comparisons are made among three proposed schemes and a random scheme, which produces locations of data centers randomly. Experiment results reveal that Scheme 3 noticeably outperformed remaining schemes. Taking results on February 3 (Dataset 1) for instance, in comparison with Scheme 4, Scheme 3 enhances the total number of collected data packets by 57.71% when the number of data center is ten, the speed of code dissemination and the coverage of cabs are ameliorated by 23.92% and 12.93% respectively. Regarding Scheme 1, Scheme 3's total collected packets are 65.00% higher than that of Scheme 1 when there are ten data centers deployed. The figure for Scheme 3's code dissemination and coverage exceed that of Scheme 1 by 18.98% and 10.21% respectively.

The rest of this paper is organized as follows. In Section II, related work is reviewed. Following that, the system model and problem statements are described in Section III. The detailed design of the DCDO schemes is discussed in Section IV. The performance analyses and experiment results are presented in Section V. Finally, the conclusion is made in Section VI.

II. RELATED WORK

A. DATA COLLECTION IN WIRELESS SENSOR NETWORKS

Wireless sensor networks are one of the areas that have been researched for the longest time by far, and they are also the oldest networks used for collecting data [32]–[36]. In such a network, sensor nodes are deployed into the area that needs to be monitored, and the network is composed of these sensor nodes. Then, data are collected and forwarded by multi-hop routing among nodes [37]–[39]. For instance, to obtain the status (temperature, humidity, etc.) of crops, sensor nodes need to be installed in farmland which needs to be monitored. A special node called sink also need to be deployed, and it is connected to the Internet and has endless energy while normal sensor nodes are energy-restricted. Other sensor nodes forward data to sink as soon as data are collected, the data are then be transmitted to a control center to facilitate the regulation and decision-making of crops [40]–[42].

Perceiving and collecting data is the primary function of a wireless sensor network, and many research work is based on this topic [43], [44]. However, sensor nodes are extremely energy-constrained because they are powered by batteries. Therefore, a significant concern in wireless sensor networks is how to save energy while ensuring the quality of the collected data. For sensor nodes, both receiving and forwarding data consume energy. Since the nodes that are closer to the sink have more data traffic, these nodes consume energy more quickly. The death of these nodes could negatively influence the lifetime of the entire network, and this phenomenon is known as the energy hole. Many research has been proposed to avoid such a phenomenon [41], [45]. A feasible method is data aggregation. In this method, collected data are integrated with data packets collected from nearby nodes because they are usually relevant and have similar content. Consequently, the total amount of data packets are reduced, and the life span of the entire network is extended [22], [44]. Data merging process can be divided into two categories. The first one is infinite data merging mode, in which the number of data packets that can be merged together is infinite [44]. This mode is usually applied to applications that sense only the maximum and minimum data. For instance, regulation centers of crops are more concern about the maximum and minimum daily temperature. Hence, energy consumption of networks could be noticeably reduced, and regulation centers receive less redundant data. In another mode of data aggregation, which is widely used. In this mode, data packets are aggregated together into packets that have sizes in proportion ($\emptyset | 0 < \emptyset \leq 1$) to their original size, $\emptyset$ is also referred to as the ratio of data merging.

In wireless sensor networks, another major concern is the quality of collected data. The quality of data can be measured in terms of delay [2], [10], [22], the amount of energy consumption per bit, and the reliability of collected data. There is plenty of research has been conducted to ameliorate the quality of data. Due to the low efficiency of wireless sensor nodes, their communication range is constrained. Also, it is found that long-range communications consume more energy than short-range communications using multi-hop routing [33]. Therefore, multi-hop routing is widely used in wireless sensor networks. However, it often causes more latency. Besides, the reliability of wireless sensor networks is significantly lower than that of wired networks due to the path loss caused by wireless transmission. Moreover, data need to be retransmitted if failures occurs, which further increase the delay. The reliability of data transmission also depends on the transmission power of sensors nodes. In other words, higher efficiency yields higher reliability of networks since the number of transmission failure is reduced. Although the amount of consumed energy is positively correlated to the transmission power, relevant research can be found in [33]. Furthermore, the quality of collected data is influenced by routing strategies. Thus, many research aims to ameliorate the quality of collected data by devising more effective routing protocols. In a simple point-to-point routing process, one sender chooses a receiver and send data packets to it. However, this method is not reliable due to the inherent instability of wireless transmission. To address this issue, some researchers proposed the opportunistic routing scheme [46], in which a sender can choose multiple receivers to perform data transmission. As a result, latency can be reduced while the reliability of transmission is guaranteed.

To reduce total energy consumption, wireless sensor networks often use the duty cycle [10], [32]. In a duty cycle, sensor nodes shift between two states of the duty cycle: awake and sleep. Using duty cycle can considerably expand the lifespan of entire networks because when sensor nodes are idle, they shift into the asleep mode, which consumes less energy than the awake state [10], [32]. Nevertheless, using duty cycle may render more delay because when a sender sends a data packet to a receiver which is under the sleep mode, it takes some time for the receiver to shifts back into the awake mode to receive the data.

B. DATA COLLECTION IN PARTICIPATORY SENSING NETWORKS

Participatory Sensing Network (PSN) [12], [17], [47], [48] is a network that is similar to the vehicular network used in this paper. In a typical PSN, sensing devices are usually smartphones or other portable sensors [12], [17]. On the one hand, there is a considerable number of mobile phones that have sensing abilities, and it is estimated that there are more than 20 billion mobile phones in the world [12], [17]. On the other hand, mobile phones have potent capacities. The CPU processing speed, storage size, and communication capability of mobile phones are even faster than personal computers that were manufactured ten years ago. As a large number of mobile phones move along with people, a variety of data are collected.

Moreover, the types of collected data cover temperature, humidity, and other forms of data such as sounds and videos [12], [17]. Therefore, data sources of PSN are comprehensive, and those data to be collected are usually real-time or semi-real-time. Therefore, it is inexpensive and convenient to obtain data in PSN. Such an attribute make it possible to observe some objects across a broad geographical scale continuously [12], [17]. For instance, it used to be a hard and costly task to trace the migration of migratory birds because the time and the area that they start to migrate varies a lot. Traditional methods such as deploying observation stations are very time-consuming and ineffective. To facilitate applications like this, industry and academia have made many efforts. It is found that the utilization of participatory sensing is an effective way to obtain data at a low cost. To conduct participatory sensing, data consumers firstly send demands and requirements of data, and reporters who obtain the requested data will upload them and get paid. By reporting captured data (in forms of pictures and videos), big tasks such as tracking the migration of animals can be achieved [12], [17].

In comparison with wireless sensor networks, participatory sensing networks usually use 4G or even 5G cellular networks. In other words, networking infrastructures such as sensor nodes and sinks do not have to be deployed [49]–[51].

One of the primary goals in the data collection perspective of participatory sensing networks is to motivate participants to collect comprehensive and high-quality sensing data. Since many research and applications (e.g. tracking the migration of birds) are based on sufficient data, the volume of collected data has to be large enough. In the bird-migration example, patterns of the migration can only be revealed when the amount of data about their migrating routes is big enough. Also, the collected data are expected to have high quality since redundant data is useless and increase the burden of networks. For instance, when an emergency event occurs, a few high-quality data can cover all the critical information about this event while data of low quality usually contains much irrelevant information, which could distract readers and cost more resources to collect them.

Many research about the data collection in participatory sensing network has been conducted. One of the most prevalent methods is the money incentives scheme. The core idea of this scheme is to pay back participants who collect useful data [12], [17], while the amount of award is determined by the quantity and quality of collected data. Generally speaking, more reward will be provided to stimulate more people to get involved if the number of reported data is low. Conversely, there will be fewer rewards if the reported data are already sufficient. Later research does not pay back participants based on the number of provided data, the quality of data is used as criteria instead. Thus the performance of applications is guaranteed due to high-quality data and low cost. For instance, when the required data is about weather conditions. If rewards for reported data are same, population-dense regions are likely to produce more redundant data and cost more money. While population-sparse areas may have insufficient data samples. Therefore, some researchers proposed schemes that can adjust the number of rewards according to regions and time to ameliorate system performance [12], [17]. Tham and Luo [52] proposed a data-collecting scheme based on ''Quality of Contributed'' to collect data that can evenly cover an entire region. Reference [53] introduced a scheme based on ''Quality of Information'' to collect data. Similar research can also be found in [12] and [17].

C. DATA COLLECTION IN VEHICULAR NETWORKS

In wireless sensor networks, the data collection can be performed only after the network required by an application is deployed in advance [54]. However, the cost of network deployment is high, and it is inconvenient to share the collected data with other applications. There is no need for participatory sensing networks to deploy any network in advance since the data collection can be facilitate by cellular networks. However, since most of participants may not process professional skills (e.g. measuring the physical deformations of bridges or detecting geographical disasters such as landslide and earthquake), participatory sensing networks are not suitable for performing data-collecting tasks that need these skills [12], [17].

The vehicular networks studied in this paper combine the benefits of wireless sensor networks and participatory sensing networks. Similar to participatory sensing networks, it is not necessary for vehicular networks to set up a specific network in advance since the task of transmitting data is performed by vehicles. Therefore, the cost of deploying and maintaining networks is saved. On the other hand, vehicular networks can collect data from sensing devices, which are able to record and monitor the physical conditions of the environment. The locations of sensing devices can be adjusted according to the requirements of applications. Unlike participatory sensing networks and wireless sensor networks, vehicular networks are connected by moving vehicles. Thus, vehicular networks are more flexible and adaptive [3], [4], [20], [21], [25], [26], [31].

Bonola *et al.* [31] proposed a scheme concerning the feasibility of utilizing cabs as data mules to collect data. In later research, some QoS-based schemes are proposed to perform data collections. Xu *et al.* [26] found that in the data-collecting scheme proposed by Bonola *et al.* [31], data collections are completed only when data are sent from vehicles to Data centers. For those vehicles that never pass a data center, the data collected by them are wasted. Even those data can finally reach Data centers, they are likely to be outdated. Such a scheme is not efficient enough for many applications. To make the data-collecting processes more efficient, Xu *et al.* [26] proposed the Latency and Coverage Optimized Data Collection (LCODC) scheme. In the LCODC scheme, data can be uploaded to Data centers not only when vehicles pass a data center, but they can also be forwarded among vehicles before they reach Data centers after a few relays. In other words, a better performance of data collection can be yielded if data can be transmitted among vehicles. Tang *et al.* [25] proposed the Simulated Annealing for Priority Assignment Algorithm (SA-PA algorithm) scheme to improve social welfare by maximizing the rate of data collection and minimizing the number of redundant data. The core idea of the SA-PA algorithm is to set priorities for sensors because the buffer capacities of data mules (vehicles) are limited. When the remaining buffer size of vehicles is sufficient, all the data sent by sensors are stored. Once a vehicle runs out of buffer capacity, data collection is conducted according to the priorities of sensors. When a vehicle passes a sensor which has high priority, new data are stored, and the data collected from sensors with lower priority are dropped. Moreover, the priorities of sensors can be adjusted dynamically. For those sensors that generate data about similar content, their priorities should be low. Conversely, those sensors that generate more crucial data usually have high priorities. For example, data collected from sensors located in city centers usually are likely to be redundant because they are collected by a large number of vehicles that pass them. However, data generated by sensors located in suburban areas are difficult to collect because those devices are less visited by vehicles. Therefore, sensors located in suburbs should have higher priorities compared with those located in downtown areas in order to collect comprehensive data, which cover most sensors of a city [55].

D. CODE DISSEMINATION IN WIRELESS SENSOR NETWORKS

Code dissemination is another area that is closely relevant to this paper. There is already a few research about code dissemination in wireless sensor networks [20], [21]. The goal of investigating code dissemination is to disseminate code effectively while keeping the number hops and delay as low as possible. Code dissemination is particularly effective using wireless sensor networks due to its broadcast ability. When a piece of code is broadcasted, all the devices within in the communication radius can receive the code and update their settings. Nevertheless, it is arduous to minimize the times of broadcast it takes to broadcast a piece of code to all the nodes in a wireless sensor network. Some researchers found that it is an NP-hard problem [56], and they proposed a connected dominating set based optimization algorithm for code dissemination [56]. The main idea of this algorithm is to find a connected dominating set, which contains nodes that can dominate all of the other nodes in the entire network. As long as a piece of code is received by nodes belong to the connected dominating set, the code is guaranteed to be received by remaining nodes.

All the aforementioned schemes that aim to address code dissemination are based on the assumption that all the nodes are working all the time. However, receivers may be unable to receive the code in a duty-cycle-based wireless sensor network during their sleep mode. Therefore, a sender may have to broadcast multiple times in order to make sure all the nodes from a duty-cycle-based wireless sensor network can receive the code. As a result, a longer delay is caused. Fang *et al.* [32] proposed an Adaption Broadcast Radius-based Code Dissemination (ABRCD) scheme to reduce delay and improve energy efficiency in duty cycle-based wireless sensor networks. After performing data collection, the ABRCD scheme utilizes the energy left in nodes which are far from the sink to enlarge the range of code broadcasting. With a larger communication radius, more nodes are able to receive the code in each broadcast. Consequently, the latency decreases considerably, and the lifespan of the network is not influenced.

E. CODE DISSEMINATION IN VEHICULAR NETWORKS

Code dissemination in vehicular networks is a newly proposed research area [20], [21]. Its goal is to disseminate code to a large number of sensing devices embedded in infrastructures to update their settings.

Liu *et al.* [21] proposed a Complete Software Update based on Trust and Priority (CSUTP) scheme to update the code of software for edge devices in a city. The CSUTP scheme mainly contains two procedures. The first procedure is to check whether edge devices need to update their code by collecting the status information of them. The second procedure is to disseminate code to edge devices. During the first procedure, each device is assigned with a priority. Code dissemination gives priority to devices with higher priority. As a result, CSUTP scheme not only reduces dissemination delay and data redundancy but also improves the coverage ratio of collecting status information of edge devices and success arrival ratio of the code.

Besides collecting data from sensing devices to vehicles, uploading data from vehicles to data centers is also a crucial procedure. Although, none of the research mentioned above addresses the deployment optimization of data centers, which is also of great significance for both data collection and code dissemination tasks. The locations and number of data centers have a considerable influence on the overall network performance. Moreover, ameliorating deployment of data centers can reduce the cost and improve the efficiency of collecting data and disseminating code. Hence, the research on their deployment is critical and useful.

III. SYSTEM MODELS AND PROBLEM STATEMENTS

A. NETWORK MODEL

The vehicular network used in this paper is similar to [3], [4], [25], [26], and [31]. As demonstrated in Figure 1, smart devices are embedded in a variety of infrastructures to collect their status data. Such as filling levels of garbage bins, the deformations of physical structures about roads and bridges, and humidity and temperature information around vegetation along either side of roads. When cabs pass by smart devices, the collected data are transmitted to cabs through wireless communications. After receiving the collected data, cabs temporarily stored the data in their buffers, and all the stored data are sent to data centers once

A. NETWORK MODEL

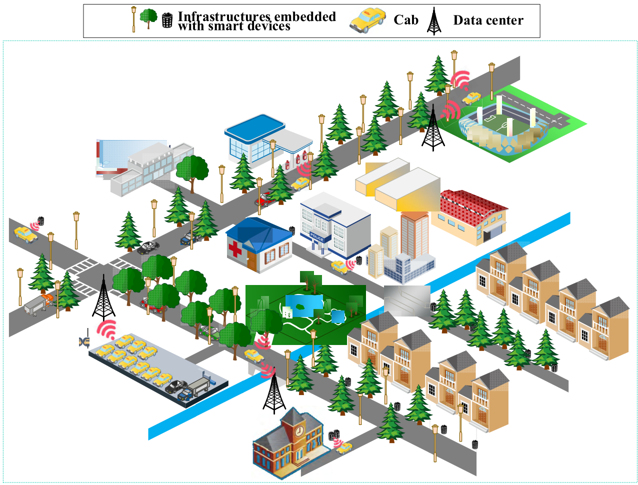

FIGURE 1. Network model.

cabs enter the communication range of data centers. For those delay-tolerant applications, using vehicular networks to collect data is efficient and economical [3], [4], [25], [26], [31]. Moreover, DCDO schemes can also guarantee the performance of applications that require less delay. Since some smart devices are software-defined devices. Code can be disseminated to these smart devices by cabs. The network model used in this paper consists of the following components:

(1) Data centers. They are storage and processing centers of collected status data. As demonstrated in right-upper part of Figure 1, a data center sends a piece of code to a cab in order to adjust the working period of street lamps. Other data centers in Figure 1 illustrate the processes of data collection. Firstly, cabs collect status data from smart devices. Following that, cabs transmit collected data to data centers when they enter the communication range.

(2) Data carriers. As discussed in Section I, due to the advantages such as long working periods and wide geographical coverage, this paper use cabs as data carriers. Cabs collect and disseminate data using wireless communication technologies. As for emergent applications that need low latency, onboard Internet connection could be used. As shown in Figure 1, they are able to collect data from smart devices installed in green belts and garbage bins. They can also disseminate code to smart devices to modify their settings or upgrade their functions.

(3) Smart devices. Sensors and actuators embedded in infrastructures of a smart city are referred to as smart devices in this paper. Facilitated by the software-defined networking technology, smart devices become more intelligent because they can be upgraded after receiving the new code. As can be seen from Figure 1, smart devices generate status data such as the humidity and temperature data around trees while they are working, and they sense the filling level of garbage bins. Then, smart devices communicate with passing cabs using low power wireless technologies such as Bluetooth Low Energy. After receiving the code sent by cabs, smart devices are capable of modifying their settings. To illustrate, the smart devices that originally collect temperature data can be reset to collect humidity data around vegetation. The working period of street lamps can be reset in foggy weathers.

As it is discussed in Section I, the majority of data collection tasks is delay-tolerant, the reason is that the status data produced by infrastructures are usually not urgent. Taking garbage bins for example, when one of them is full, it is not necessary for the regulators of garbage collection to be informed immediately. Regulators can only plan a better collection path after receiving enough status data. However, code dissemination tasks usually expect lower latency because the period of incongruity should be reduced as we discussed in Section I.

Cabs can also make use of their LTE connection and stay online while they are working. The DCDO schemes can be achieved by both online and offline approaches. For the online approach, cabs can immediately upload the collected data. This approach is particularly suitable for application that requires a low delay. However, this may be unnecessary since the majority of application are delay-tolerant and cab drivers may be unwilling to share their LTE connection due to the concerns of cost. Hence, the offline approach is used in the evaluations of DCDO schemes. In the offline approach, data are firstly stored in buffers of cabs before they are finally transmitted to data centers. As for code dissemination process, the online and offline approaches are similar to that of data collection. Under the online approach, cabs can receive the new published code immediately before sending them to smart devices. Although, for the applications which are delay-tolerant (e.g. updating the content in billboards), the offline approach is sufficient. Cabs obtain the code when they get into the communication range of data centers. Then they forward the received code to smart devices. As opposed to collecting data, only a piece of code need to be spread, so the code usually does not occupy much space of cab' buffer. Next, the standardized definitions of the network model are given.

In this smart city, it is assumed that there are m cabs involved in the data collection and code dissemination processes. The quantity of data centers is k, and m cabs are represented as $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m$. There are n smart devices, and they consist a set $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_n\}$, among which $\mathcal{S}_i$ represents the i-th smart devices. Besides the set of locations that are made up of smart devices, two other schemes proposed in this paper use other criteria. The number of locations that are used as input in all of three schemes is n. To obtain n locations for Scheme 2 and Scheme 3, thresholds are used to limit the number of locations to n (top n locations with higher traffic or valid flow is obtained). A set of locations with high traffic flow are represented as $\mathcal{F} = \{\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n\}$, and $\mathcal{F}_i$ refers to the i-th location in this set. Similarly, $\mathcal{V} = \{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_n\}$ is a set of locations that have high valid flow, among which the i-th location is referred to $\mathcal{F}_i$. The detailed explanations of valid flow are discussed in Section IV.A. The aims of smart devices embedded into infrastructures are to collect status data and send them to data centers. Data centers are represented as $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_k\}$. The role of data carriers is played by cabs, they collect and temporarily store data packets while they are moving around the city restlessly. The data stored in their buffers are sent to data centers when they get close to data centers. To obtain the locations of data centers, a clustering algorithm is used (discussed in detail in Section IV). Since the number of data centers is k, k clusters consist a set of clusters $\mathfrak{U} = \{\mathfrak{U}^1, \mathfrak{U}^2, \ldots, \mathfrak{U}^k\}$. The j-th cluster that contains r elements is $\mathfrak{U}^j = \{\mathfrak{U}^j_1, \mathfrak{U}^j_2, \ldots, \mathfrak{U}^j_r\}$.

From the moment data are produced until they are received by data centers, latency inevitably occurs. $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_w\}$ is a set of data packets sent to data centers, the total number of data packets stored in data centers is w. t_i denotes the time it takes from a data packet is sent to a cab (τ_{start}) until it is received by a data center (τ_{end}).

$$t_i = \tau_{end} - \tau_{start}$$

For the convenience of readers, the notations used in this paper are summarized in Table 1.

TABLE 1. Notations.

Symbol	Description
m	The number of cabs in a smart city
n	The number of smart devices in a smart city
k	The number of data centers in a smart city
$\mathcal{C}$	A set of cabs
$\mathcal{C}_i$	The i-th cab in $\mathcal{C}$
$\mathcal{S}$	A set of smart devices
$\mathcal{S}_i$	The i-th smart devices in $\mathcal{S}$
$\mathcal{F}$	A set of locations with high traffic flow
$\mathcal{F}_i$	The i-th element in $\mathcal{F}$
$\mathcal{V}$	A set of locations with high valid flow
$\mathcal{V}_i$	The i-th element in $\mathcal{V}$
$\mathcal{D}$	A set of data centers
$\mathcal{D}_i$	The i-th data center in $\mathcal{D}$
$\mathcal{P}$	A set of data packets stored in data centers
$\mathcal{P}_i$	The i-th data packet in $\mathcal{P}$
w_i	The number of data packets stored in $\mathcal{D}_i$
w	The sum of data packets stored in all of data centers
$\mathfrak{U}$	A set of clusters
$\mathfrak{U}^i$	The i-th cluster in $\mathfrak{U}$
$\mathfrak{U}^i_j$	The j-th element in $\mathfrak{U}^i$
r_i	The number of elements in $\mathfrak{U}^i$
DCs	Data centers

B. PROBLEM STATEMENTS

The principal goals of the DCDO schemes are to maximize the number of collected packets and reduce the delay of code dissemination. The optimization objectives in this paper can be categorized into the following aspects:

(1) Locations of data centers

While the deployment of smart devices depends on the plans and requirements of a smart city. They are embedded in infrastructures like garbage bins, street lamps, and billboards. They usually located on either side of roads. Similar to data centers, the locations of smart devices usually do not frequently change once they are installed. Therefore, the deployment of data centers is significant because it can reduce the total distance between smart devices and data centers. The total distance among each location and its closest data centers to them for three schemes are:

$$\mathrm{L} = \sum_{i=1}^{k} \sum_{j=1}^{r_i} dist(\mathcal{D}_i - \mathfrak{U}^i_j)$$

In the formulas above, $\mathcal{D}_i$ denotes the centroid of the i-th cluster that $\mathfrak{U}^i_j$ belongs to. $\mathfrak{U}^i_j$ is $\mathcal{S}_j$, $\mathcal{F}_j$, and $\mathcal{V}_j$ in Scheme 1, Scheme 2, and Scheme 3 respectively. $dist(\mathrm{D}_i - \mathfrak{U}^i_j)$ represents the Euclidean distance between $\mathfrak{U}^i_j$ and $\mathcal{D}_i$. The optimization problem is converted to:

$$min\,(\mathcal{L}) = min\left(\sum_{i=1}^{k} \sum_{j=1}^{r_i} dist(\mathcal{D}_i - \mathfrak{U}^i_j)\right) \tag{1}$$

(2) Latency

The delay of data packet $\mathcal{P}_i$ is t_i, this needed to be minimized:

$$min(t_i) = min(\tau_{end} - \tau_{start})$$

In this paper, Delay Satisfaction Degree (DSD) is used to measure the delay:

$$\mathcal{T}_{dsd} = \sum_{i=1}^{w} \frac{t_i}{w}$$

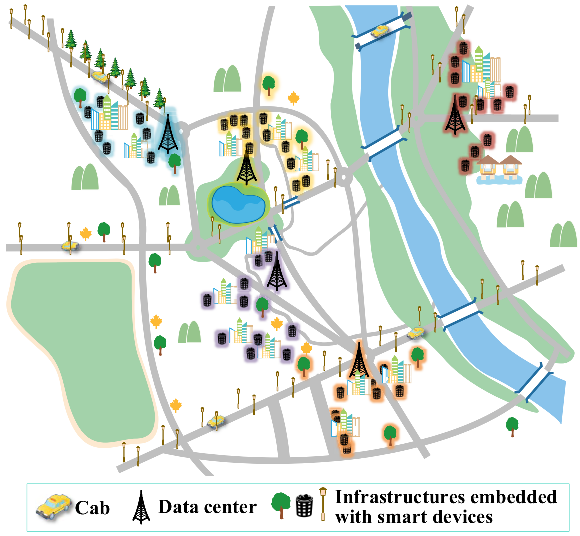

FIGURE 2. The smart scenario used in experiments.

Thus the objective is to minimize the overall average latency:

$$min\,(\mathcal{T}_{dsd}) = min\left(\sum_{i=1}^{w} \frac{t_i}{w}\right) \tag{2}$$

(3) Number of collected data packets

There are k data centers deployed in this city, and the number of data packets collected by the data center $\mathcal{D}_i$ is w_i.

$$\max\,(w) = \sum_{i=1}^{k} w_i \tag{3}$$

In summary, the optimization targets of this paper are:

$$\begin{cases} min\,(\mathcal{L}) = min\left(\sum_{i=1}^{k}\sum_{j=1}^{r_i} dist(\mathcal{D}_i - \mathfrak{U}_j^i)\right) & (1) \\ min\,(\mathcal{T}_{dsd}) = min\left(\sum_{i=1}^{w} \frac{t_i}{w}\right) & (2) \\ \max\,(w) = \sum_{i=1}^{k} w_i & (3) \end{cases}$$

IV. MAIN DESIGN OF DCDO SCHEMES

A. OVERVIEW

In this section, the overall design of the proposed schemes is discussed. In Subsection B, the clustering algorithm we used in DCDO schemes are discussed.

The distributions of smart devices in residential areas and central business district tend to be denser, and such an attribute makes them ideal to be clustered. Deploying data centers to the centroids of clusters can facilitate cabs to collect data from smart devices to a data center within a same cluster more efficiently. To optimize objective (1), the distances between each locations and its closest centroid should be minimized. In other words, if each cluster has lower total distance among its centroid and the locations that belong to the cluster, the overall distance among locations and their closest data centers will decline.

In order to comprehensively investigate the possible ways to plan the deployment of data centers, three schemes use different criteria. The first scheme directly use the locations of smart devices as input of a clustering algorithm. Besides, we combine traffic data analysis into our schemes and it is found that using locations with high traffic flow or high valid flow as input can yield better performance. Another two schemes firstly have to calculate which locations have high traffic flow and valid flow. Then, these locations are used to conclude the locations of data centers, which are centroids of each cluster. In order to measure the traffic flow and valid flow, a city is divided into 25,000 grids, each of which is 40 meters $*$ 40 meters. Considering the communication range of technologies such as Bluetooth Low Energy and 802.15.4, which is 10-30 meters in free space, such a design is feasible. To simplify the discussion, GPS points within each grids is considered as a same location. Using the grid system, the traffic flow and valid flow of each grid can be measured. Three schemes are described as follows:

1. Scheme 1. The locations of smart devices are used as input of a clustering algorithm in Scheme 1. The locations of smart devices are determined by external factors such as the layout of a city, it is assumed that smart devices are all installed in either sides of roads which are reachable for cabs in experiments. The detailed algorithm of Scheme 1 is illustrated in Subsection B.

2. Scheme 2. It uses locations with high traffic flow as input. The traffic flow is measured by counting the frequency that cabs have passed each grid during an entire day. If some of grids have higher traffic flow, the probability that a cab passes that grid again is high. The grids with high traffic flow are usually located in hectic areas like city centers and intersections. Since cabs appear in these grids more frequently, the data exchange among cabs and data centers is also more frequent. After calculating the traffic flow of every grid, a threshold is set to limit the number of high-traffic-flow grids into a pre-defined amount so that three schemes can use same amount of locations as input. After that, the clustering algorithm is run to obtain the locations of data centers of Scheme 2.

3. Scheme 3. This scheme uses locations with high valid flow as input. The valid flow of a grid is the cumulative total number data packets carried by cabs when they reach the grid over a day. For instance, when a cab carrying two data packets pass a grid at 13:01, the valid flow of this grid adds two to its original value. It is introduced because it is found that although some grids have high traffic flow, cabs do not carry many data packets when they pass these grids. Therefore, valid flow can not only reflect the traffic flow of each locations, but also indicate which locations have high demand of data exchanges. Similar to Scheme 2, after obtaining a set of locations with high valid flow using a threshold, they are clustered into various clusters and data centers' locations are obtained.

A scheme called Scheme 4 is designed to be used as a contrast. In this scheme, locations of data centers are generated randomly in the given area.

B. CLUSTERING ALGORITHM

This sub-section describes the detailed procedures of obtaining locations of data centers by applying a clustering algorithm.

Clustering is sometimes referred to as unsupervised classification. Popular clustering algorithms include the K-means algorithm and the hierarchical clustering algorithm. In addition, the bisecting K-means algorithm is an advanced version of the K-means algorithm. One of the advantages of the bisecting K-means algorithm is that it is less susceptible to initialization problems (randomly selected initial centroids may be poor). There are three major reasons why the bisecting K-means algorithm is used in this paper:

(1) The distribution of smart devices in areas like residential blocks and central business districts usually has higher density since infrastructures such as garbage bins are usually put together closely in these areas. (2) This algorithm has high speed and extendibility.

Each location is clustered into its closest cluster $\mathfrak{U}^i$, of which the centroid (data center) is $\mathcal{D}_i$. The sum of the squared error (SSE) is used to measure the quality of a clustering. SSE represents the total Euclidean distance among $\mathcal{D}_j$ and all the smart devices that belong to this cluster:

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{r_i} dist(\mathrm{D}_i - \mathfrak{U}_j^i)^2$$

After calculating the traffic flow and valid flow as discussed in Subsection A, the goal of clustering is to minimize SSE. The following algorithms depict the detailed algorithm for each of three schemes.

Algorithm 1 Clustering Algorithm for Scheme 1

```
Input the number of data centers: K
Input a set of smart devices: 𝒮
Initialize a cluster 𝔘^i that is made up of smart devices,
  {𝔘^i_1, 𝔘^i_2, ..., 𝔘^i_{r_i}} = {𝒮_1, 𝒮_2, ..., 𝒮_r}
While the number of clusters is less than K do
  Remove the cluster 𝔘^j that has the largest SSE from 𝔘
  Select two elements from 𝔘^j as initial centroids
  While centroids change their locations do
    Form two clusters by assigning each locations to its
    closest centroid
    Compute the mean of each cluster separately as their
    centroid
  End While
  Add these two clusters to 𝔘
End While
```

Algorithm 1 is the pseudo-code of the clustering process of Scheme 1.

Algorithm 2 and Algorithm 3 are the pseudo-code of the clustering process of Scheme 2 and Scheme 3 respectively. Their difference is that their input is based on three different criteria.

V. EXPERIMENT RESULTS AND PERFORMANCE ANALYSIS

To evaluate the performance of the DCDO schemes, extensive experiments were conducted using two large datasets obtained from real life [57]–[59]. Comparisons were made among three proposed schemes and a contrast scheme. After analyzing the experiment results, it is proved that for the optimization target (2) and (3), the DCDO schemes can improve the efficiency of data collection and code dissemination tasks considerably by optimizing the deployment of data centers. As a result, the management of infrastructures in smart cities is considerably facilitated. A scenario that applies Scheme 1 to a smart city is demonstrated in Figure 2. It is

Algorithm 2 Clustering Algorithm for Scheme 2

```
Input the number of data centers: K
Input a set of locations that have high traffic flow: ℱ
Initialize a cluster 𝔘^i that is made up of locations with
  high traffic flow, {𝔘^i_1, 𝔘^i_2, ..., 𝔘^i_{r_i}} =
  {ℱ_1, ℱ_2, ..., ℱ_n}
While the number of clusters is less than K do
  Remove the cluster 𝔘^j that has the largest SSE from 𝔘
  Select two elements from 𝔘^j as initial centroids
  While centroids change their locations do
    Form two clusters by assigning each locations to its
    closest centroid
    Compute the mean of each cluster separately as their
    centroid
  End While
  Add these two clusters to 𝔘
End While
```

Algorithm 3 Clustering Algorithm for Scheme 3

```
Input the number of data centers: K
Input a set of locations with high valid flow: 𝒱
Initialize a cluster 𝔘^i that is made up of locations with
  high valid flow, {𝔘^i_1, 𝔘^i_2, ..., 𝔘^i_{r_i}} = 𝒱_1, 𝒱_2, ..., 𝒱_n
While the number of clusters is less than K do
  Remove the cluster 𝔘^j that has the largest SSE from 𝔘
  Select two elements from 𝔘^j as initial centroids
  While centroids change their locations do
    Form two clusters by assigning each locations to its
    closest centroid
    Compute the mean of each cluster separately as their
    centroid
  End While
  Add these two clusters to 𝔘
End While
```

noticeable that smart devices in the city are mostly located in residential areas, and they are installed into infrastructures located along roads. After these smart devices are clustered into clusters (clusters are represented in different colors), the locations of data centers are centroids of these clusters. It is evident that a cab can collect data from smart devices and communicate with the data center (centroid) which is located near the cluster of smart devices swiftly.

In Subsection A, the processes of clustering 4,900 locations into 25 and 150 data centers using three schemes are visualized. The number of the total collected data packets and its correlation with the figure for data centers are illustrated in Subsection B. Finally, the visualization of trajectories and the coverage of smart devices are presented Subsection C. Moreover, through the trajectories of cabs, it is evident that the DCDO schemes are adaptive under various circumstances including weekdays, weekends, and holidays.

To comprehensively evaluate the proposed schemes in real world settings, we found two trajectory datasets that

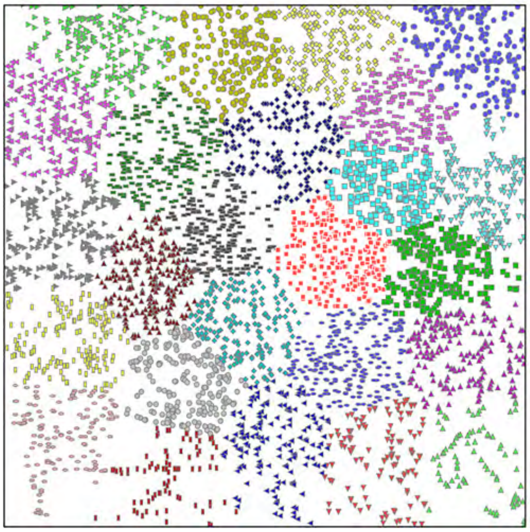

FIGURE 3. Smart devices are clustered into 25 clusters using Scheme 1.

FIGURE 4. Locations of 25 data centers generated by Scheme 1.

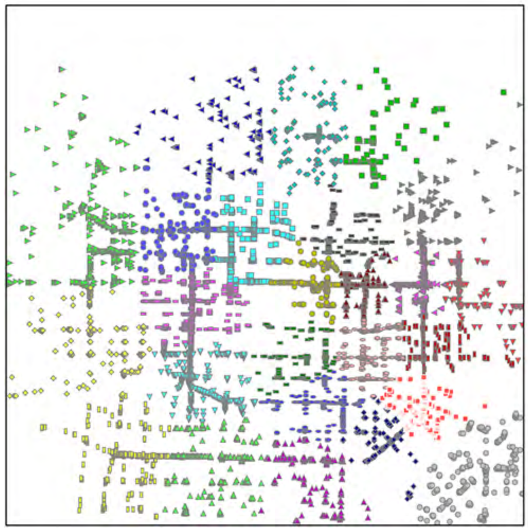

FIGURE 5. Locations with high traffic flow are clustered into 25 clusters using Scheme 2.

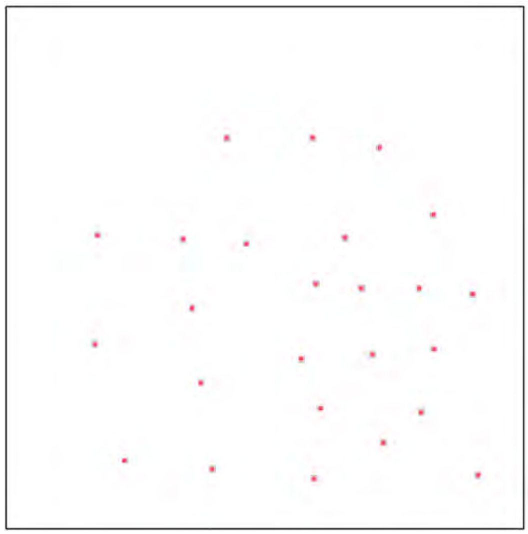

FIGURE 6. Locations of 25 data centers generated by Scheme 2.

were created by Microsoft Research Asia [57], [58] and Wireless and Sensor networks Lab, Shanghai Jiao Tong University [59], and they are referred as Dataset 1 and Dataset 2 respectively. Dataset 1 contains the GPS trajectories of Beijing. Approximately 400 square kilometers of area that located in the city center of Beijing is chosen, and its longitude and latitude lies within 116.266-116.500 and 39.810-39.990 respectively. Similarly, an area of some 400 square kilometers is chosen for Dataset 2, and the range of longitude and latitude of the chosen area is 121.390-121.600 and 31.110-31.290 respectively. Within the chosen areas, Dataset 1 and Dataset 2 have 7,871,498 and 17,478,202 GPS points respectively. In order to facilitate the experiments and visualization processes, two chosen areas are divided into 250,000 girds. Considering the communication range between smart devices and cabs, such a design is reasonable since each grid represent an area of approximately 40 meters * 40 meters. Therefore, as soon as a cab enter a gird, data exchanges between smart devices and cabs can occur. In addition, the effective communication range between data centers and cabs is approximately 250 meters according to 802.11p and 802.11a/b. Thus, cabs can transmit data to data centers when they enter the communication range of data centers.

The dates we used in experiments are also very representative. February 3 to February 7 are chosen from Dataset 1 while February 13 to February 17 are chosen from Dataset 2. Because these periods of time contain workdays, weekends, and the most significant festival in China—Spring Festival. Taking Dataset 1 for example, February 3 is Sunday, February 4 and February 5 are weekdays, February 6 and

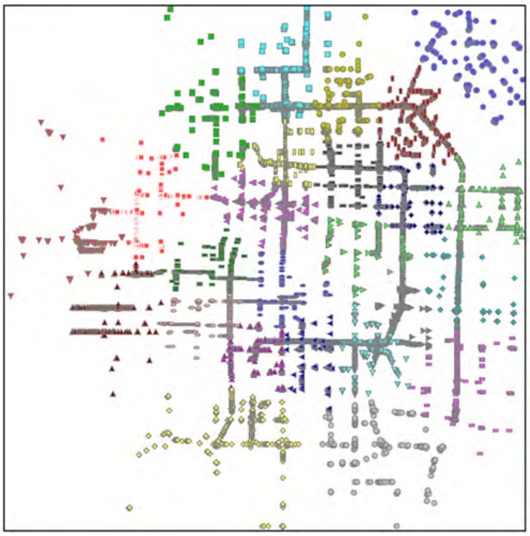

FIGURE 7. Locations with high valid flow are clustered into 25 clusters using Scheme 3.

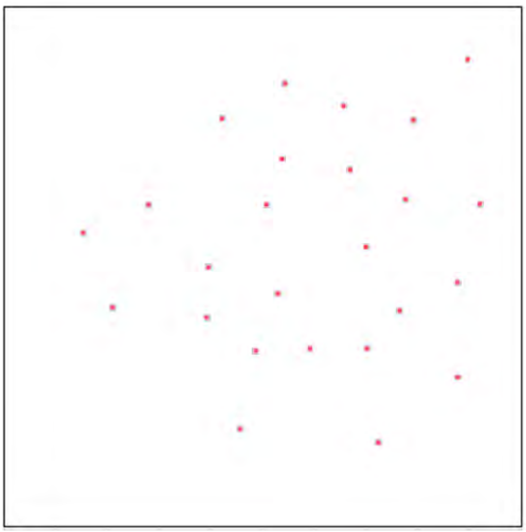

FIGURE 8. Locations of 25 data centers generated by Scheme 3.

February 7 are Spring Festival holidays. Regarding Dataset 2, February 13 and February 16 are weekdays while February 16 is the beginning of the Spring Festival holidays in that year.

As for the deployment of smart devices, they usually depend on city planning and budget. In this paper, it is assumed that there are 4,900 smart devices installed in this smart city. If an area of 20 Km^2 is equally divided into grids with an interval of 289.6 meters, and smart devices are installed in intersections of grids, then 4,900 smart devices are needed. In the other two schemes, 4,900 is also used to obtain locations with high traffic flow and valid flow. The parameters are summarized as tables below:

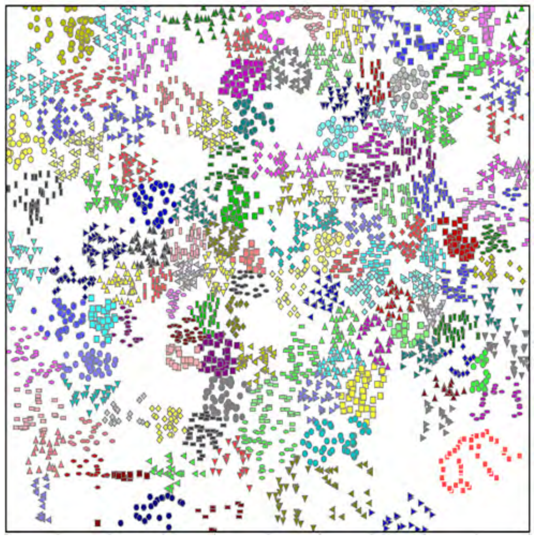

FIGURE 9. Locations of smart devices are clustered into 25 clusters using Scheme 1.

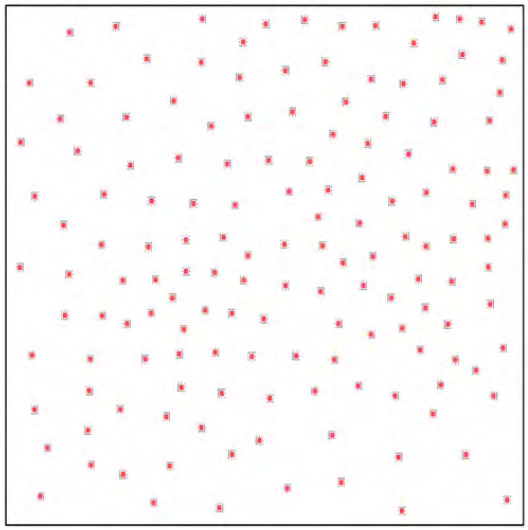

FIGURE 10. Locations of 150 data centers generated by Scheme 1.

TABLE 2. Parameters for dataset 1.

Parameter	Value
Area	20 Km^2
Longitude	116.266-116.500
Latitude	39.810-39.990
Number of GPS points	7,871,498
Number of input	4,900
Number of data centers	1-150

A. CLUSTERING PROCESSES ANALYSIS

Three schemes are implemented following the processes below:

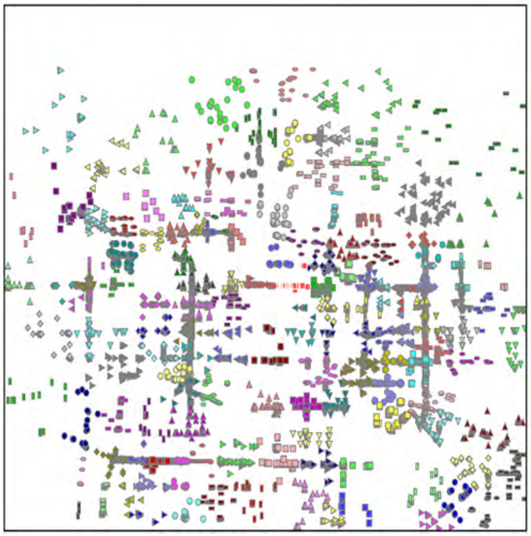

FIGURE 11. Locations with high traffic flow are clustered into 150 clusters using Scheme 2.

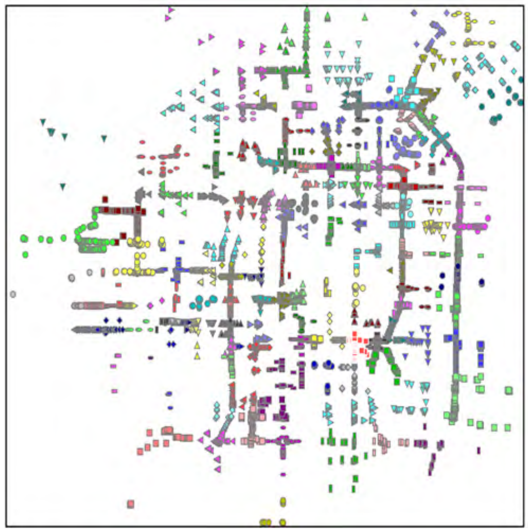

FIGURE 13. Locations with high valid flow are clustered into 150 clusters using Scheme 3.

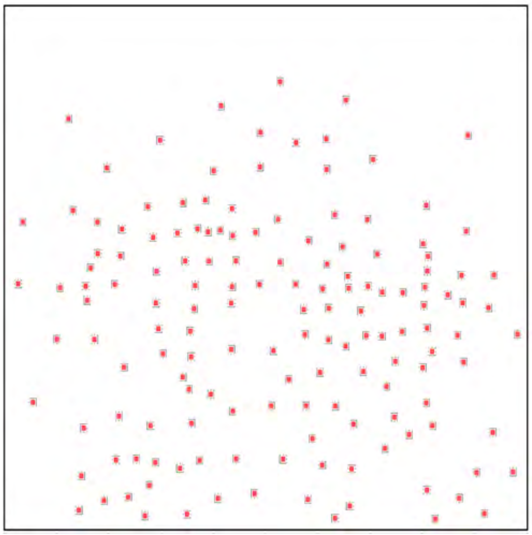

FIGURE 12. Locations of 150 data centers generated by Scheme 2.

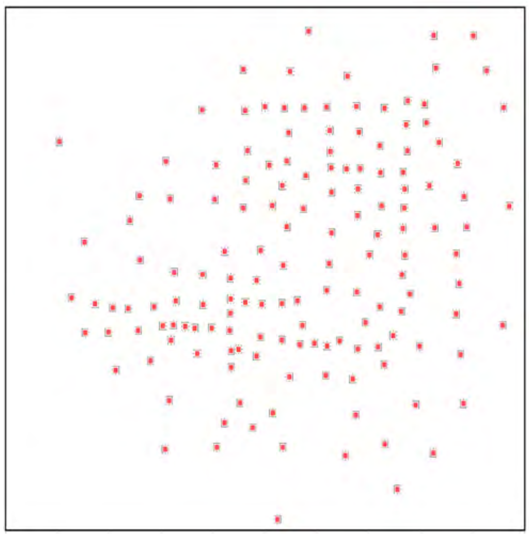

FIGURE 14. Locations of 150 data centers generated by Scheme 3.

TABLE 3. Parameters for dataset 2.

Parameter	Value
Area	20 Km^2
Longitude	121.390-121.600
Latitude	31.120-31.300
Number of GPS points	17,478,202
Number of input	4,900
Number of data centers	1-150

As discussed in Section I, the distribution of smart devices in a smart city has some patterns. Thus, they are mainly installed into infrastructures along the either sides of roads in simulations. Then, the locations of smart devices are directly used as input in Scheme 1. Regarding Scheme 2, locations with high traffic flow are generated by calculating the number of times that cabs pass each grid during an entire day as described in Section IV. Following that, a threshold is set to choose top 4,900 locations with high traffic flow, and the locations of data centers are generated by clustering the obtained locations. Similar to Scheme 2, Scheme 3 considers the frequency that cabs visit each grid. It also considers the number of data packets carried by cabs. Subsequently, a threshold is used to get top 4,900 locations that have high valid flow.

Four assumptions are made throughout the experiments in order to focus on optimizing the deployment of data centers:

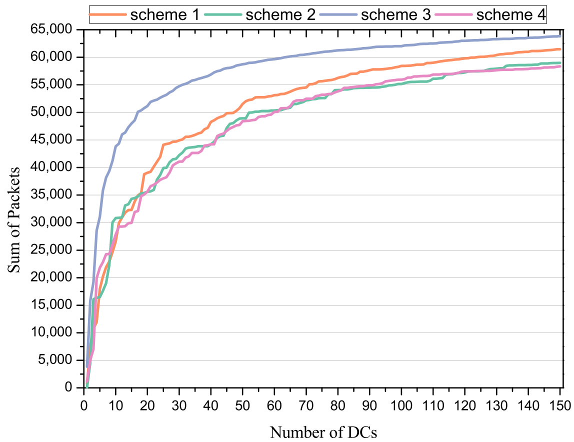

FIGURE 15. Sum of collected packets using Dataset 1 on February 3.

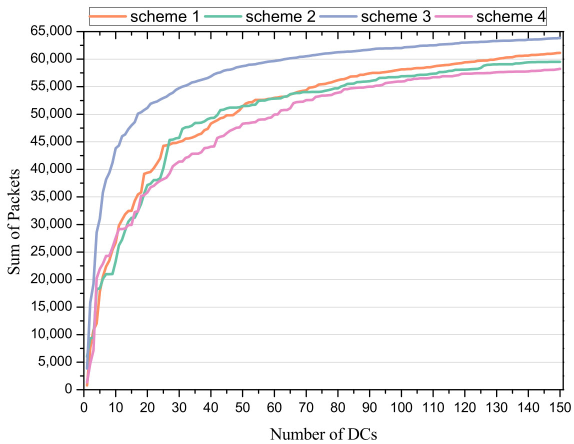

FIGURE 16. Sum of collected packets using Dataset 1 on February 4.

1. Smart devices have sufficient energy, which could be achieved by energy harvesting. Technologies such as Bluetooth Low Energy can be used to reduce energy consumption during communications.

2. Cabs only collect one data packet each time they pass a smart device.

3. There is no pairing and transferring delay between cabs and smart devices, cabs and data centers.

4. It is assumed that the capacity of cabs' buffer is infinite. However, no redundant data is stored.

In order to find the appropriate quantity of data centers that needed to be deployed, the performance was tested under a range of parameters and the number of data centers varies from 1 to 150.

Figure 3 to Figure 8 visualize the distributions of smart devices, locations with high traffic flow, and locations with high valid flow. Using these locations as input, 25 data centers are generated (illustrated in Figure 4, Figure 6, and Figure 8) under three schemes. It is evident that for Scheme 2 and Scheme 3, locations with high traffic flow and valid flow are mainly distributed in main roads and intersections.

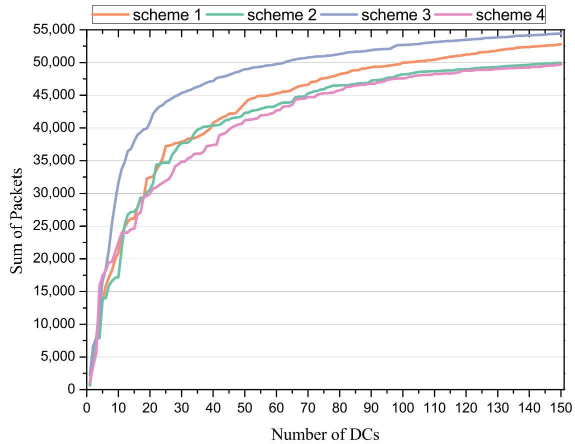

FIGURE 17. Sum of collected packets using Dataset 1 on February 5.

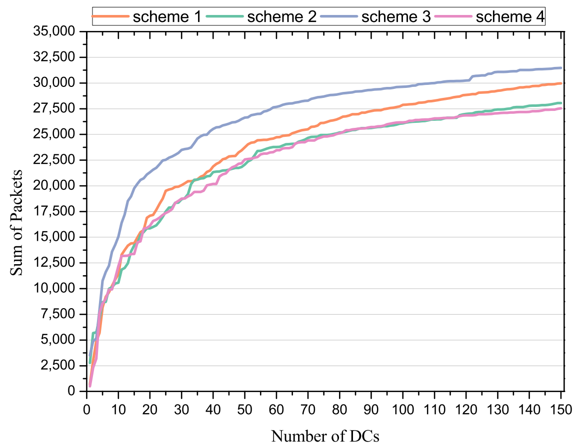

FIGURE 18. Sum of collected packets using Dataset 1 on February 6.

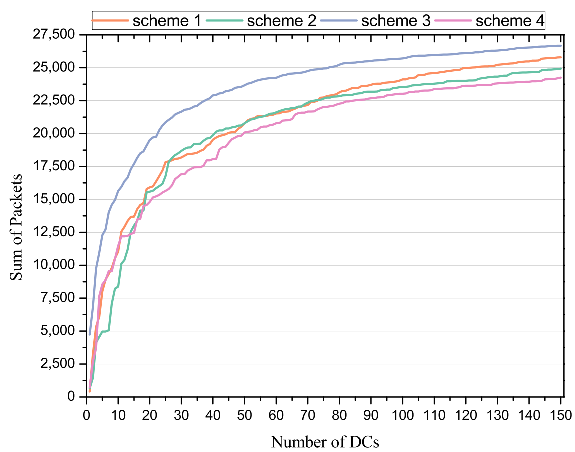

FIGURE 19. Sum of collected packets using Dataset 1 on February 7.

The experiment results which are discussed in Subsection B proved that the deployment generated by Scheme 3 outperforms other schemes.

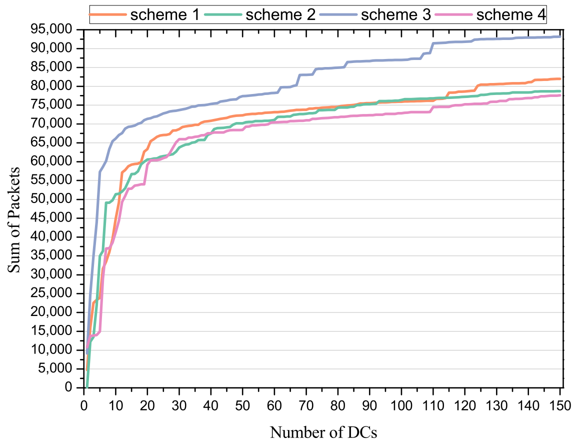

FIGURE 20. Sum of collected packets using Dataset 2 on February 13.

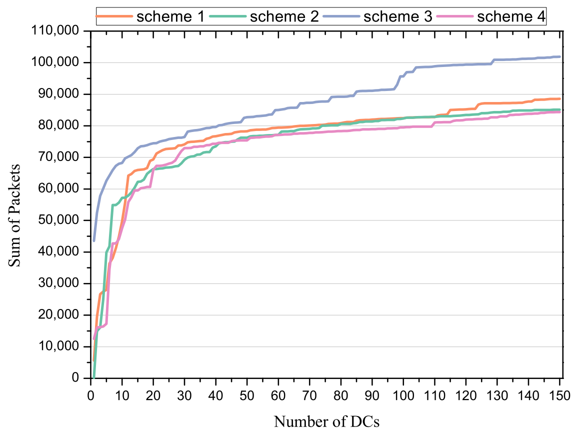

FIGURE 21. Sum of collected packets using Dataset 2 on February 14.

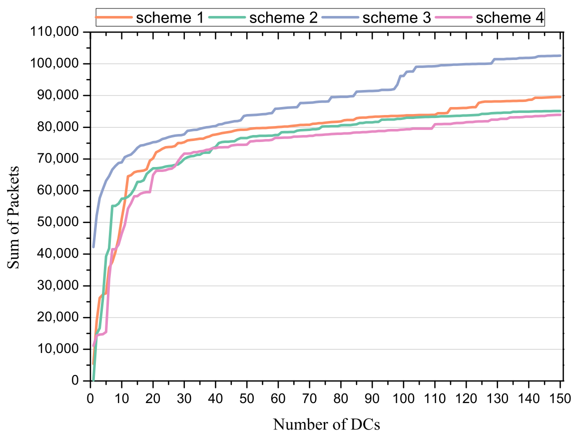

FIGURE 22. Sum of collected packets using Dataset 2 on February 15.

In comparison with obtaining 25 data centers, Figure 9 to Figure 14 visualize the clustering results when the number of data centers is 150. The input generated by three schemes is clustered into 150 clusters as shown in Figure 9, Figure 11, and Figure 13. The locations of 150 data centers obtained from three scheme are demonstrated in Figure 10, Figure 12, and Figure 14 respectively.

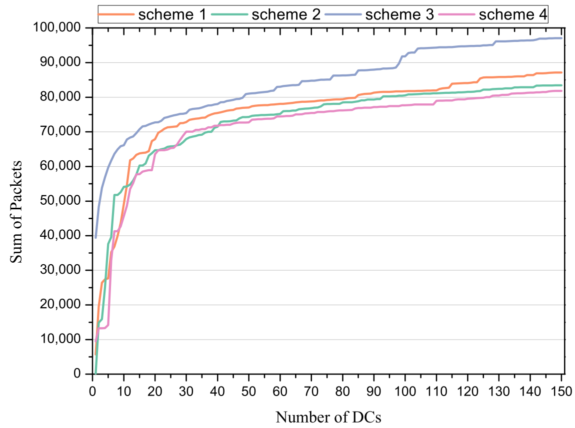

FIGURE 23. Sum of collected packets using Dataset 2 on February 16.

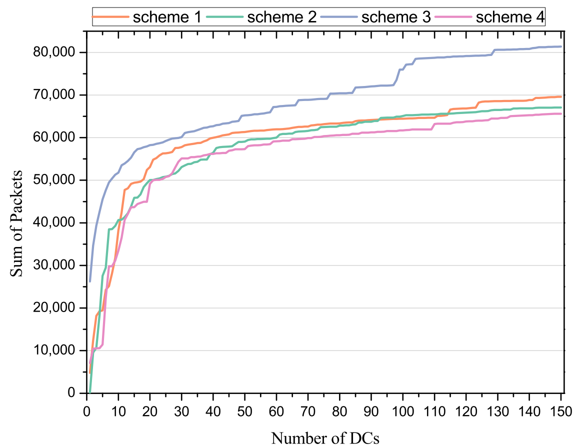

FIGURE 24. Sum of collected packets using Dataset 2 on February 17.

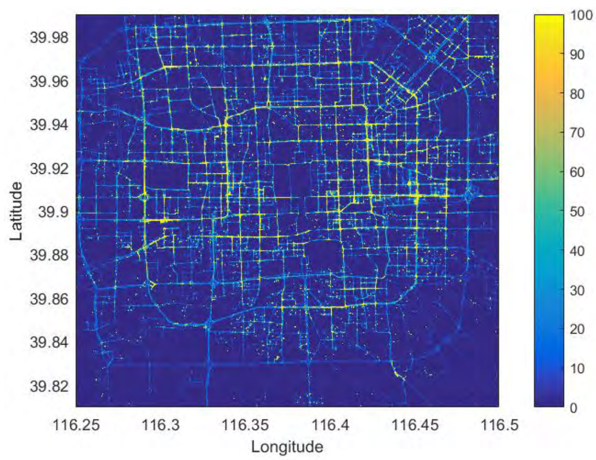

FIGURE 25. The trajectories of cabs from Dataset 1 on February 3.

B. DATA COLLECTION ANALYSIS

The number of working cabs is different due to various factors, and the status of infrastructures does not remain unchanged. Thus, the number of collected data packets also varies a lot on different dates. Figure 15 to Figure 19

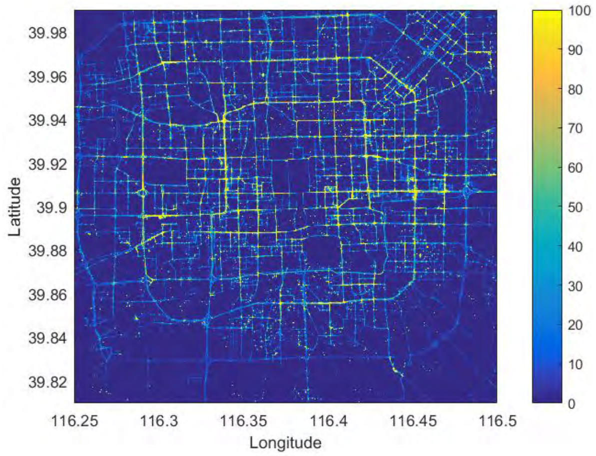

FIGURE 26. The trajectories of cabs from Dataset 1 on February 4.

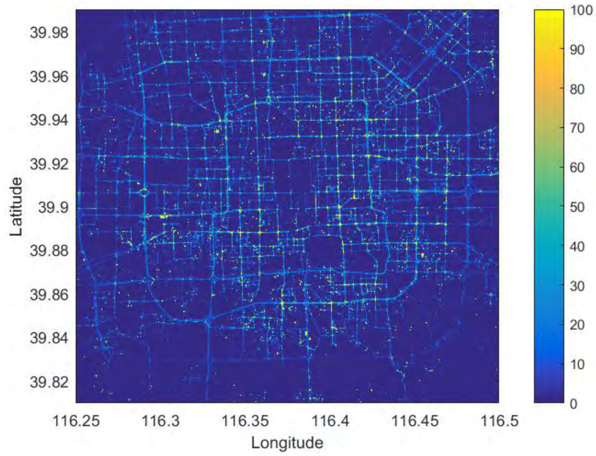

FIGURE 28. The trajectories of cabs from Dataset 1 on February 6.

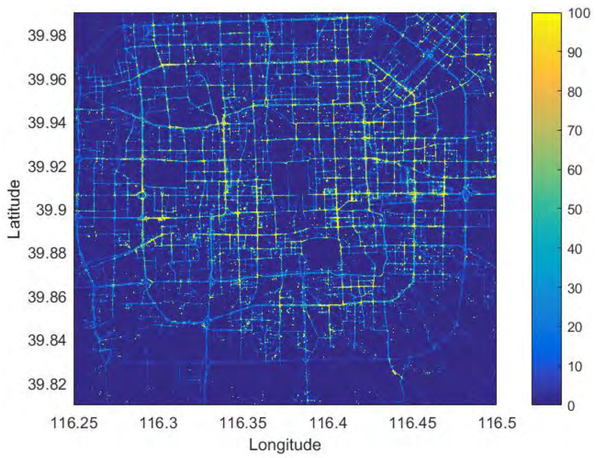

FIGURE 27. The trajectories of cabs from Dataset 1 on February 5.

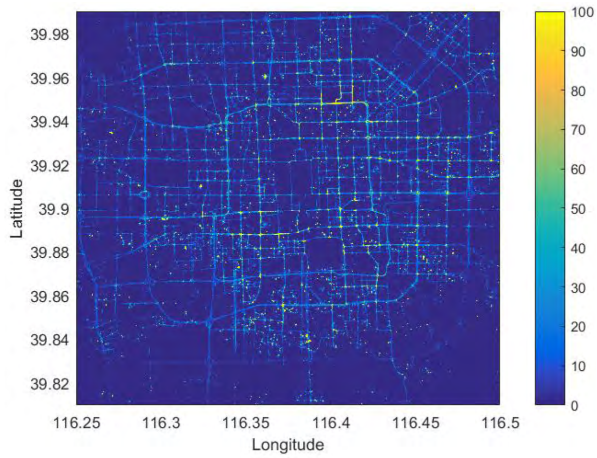

FIGURE 29. The trajectories of cabs from Dataset 1 on February 7.

illustrate the sum of collected data packets using Dataset 1, while Figure 20 to Figure 24 demonstrate that of Dataset 2. Scheme 2 and Scheme 3 are based on traffic data analysis. However, the deployment of data centers is not able to be changed frequently once it is confirmed. Therefore, we aim to find a deployment plan based on the traffic situation of a representative date so that it can be applied to other dates. The following experiments use data centers obtained from February 3 for Dataset 1 and data centers obtained from February 13 for Dataset 2. The experiment results suggest that the performance in other dates are guaranteed based on the deployment obtained from these dates. Thus the optimization target (3) is ameliorated. According to the given line charts, it is noticeable that for all of the proposed schemes, the trend of correlation between the number of data centers and the sum of collected data packets is similar: the sum of collected packets in each scheme climbs exponentially when the number of data center is between 0 to approximately 30, then it only increases slightly in the remaining range.

Among four schemes, is can be seen that Scheme 3 outperformed the remaining schemes under various dates in both datasets. Also, the sum of collected data packets in Scheme 1 is generally the second highest. In addition, the total of collected packets in all of the proposed schemes generally outnumbers Scheme 4 because locations of data centers in Scheme 4 are randomly generated.

C. CODE DISSEMINATION AND ADAPTATION ANALYSIS

In order to evaluate the speed of data collection and code dissemination. Figure 25 to Figure 29 display trajectories of Dataset 1, and Figure 30 to Figure 34 demonstrate that of Dataset 1 (the density of GPS points can be referred to the color bar on the right side of each diagram). February 7 in Dataset 1 is chosen to visualize the spreading of cabs every 4 hours as shown in Figure 35. Moreover, the correlation between time and the coverage of smart devices is evaluated.

According to Figure 28 and Figure 29, it is noticeable that trajectories on February 6 and February 7 is darker than that on February 3, February 4, and February 5. This is because February 6 in 2008 was the beginning of Spring Festival holidays, which last for seven days. During this period of holidays, a vast amount people who temporarily worked in

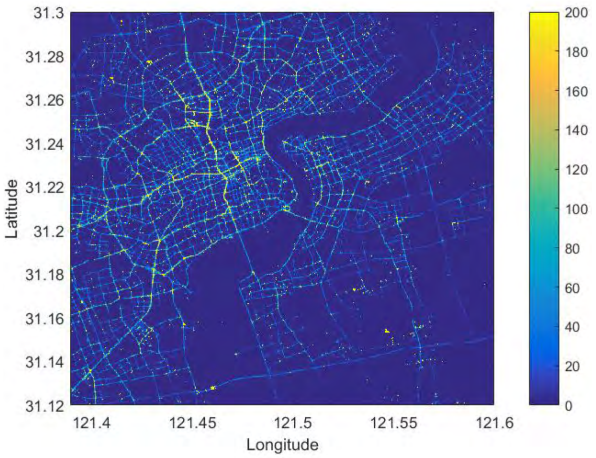

FIGURE 30. The trajectories of cabs from Dataset 2 on February 13.

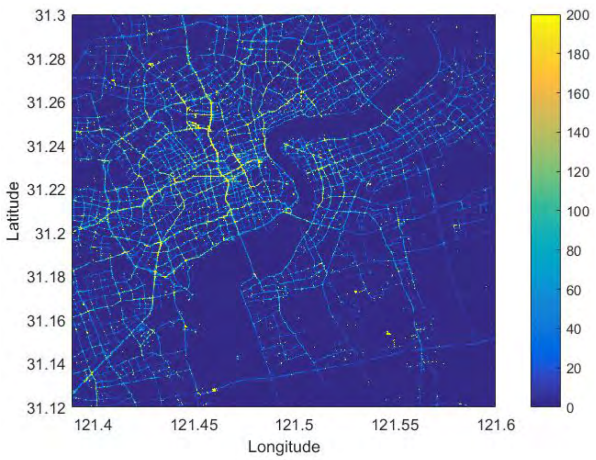

FIGURE 31. The trajectories of cabs from Dataset 2 on February 14.

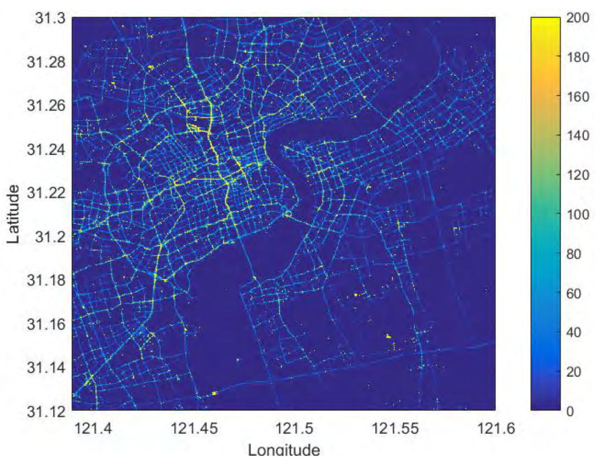

FIGURE 32. The trajectories of cabs from Dataset 2 on February 15.

Beijing went back to their hometowns for family reunions. The behavior of cabs can adjust to such changes because the number of data generated by smart devices declined as some residences left their homes in Beijing. While cabs also visited

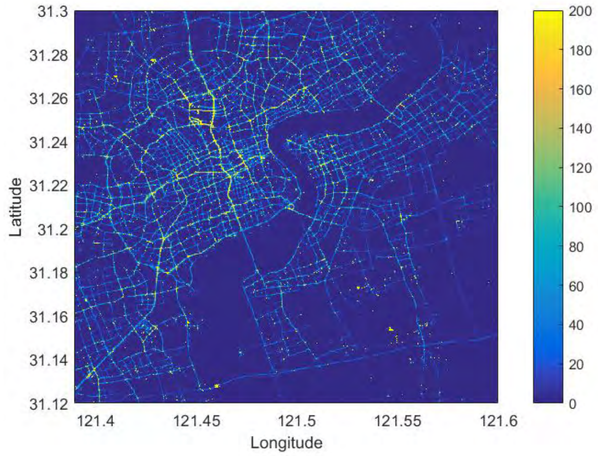

FIGURE 33. The trajectories of cabs from Dataset 2 on February 16.

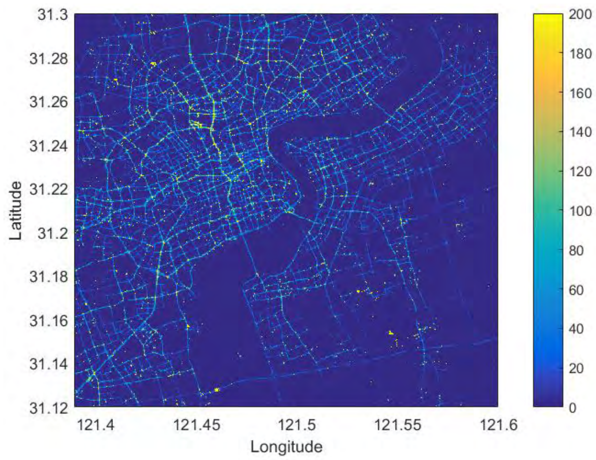

FIGURE 34. The trajectories of cabs from Dataset 2 on February 17.

residential areas less frequently since they visited the residential areas that have fewer people living there during holidays less frequently. Also, the proposed schemes are adaptive under workdays and weekends. Similar patterns can be found in Dataset 2 as shown in Figure 30 to Figure 34. Cabs moved less actively in the city center of Shanghai on February 17 (Figure 34) compared with February 13 (Figure 31) because February 17 is also the beginning of the Spring Festival holidays in the year this dataset was obtained.

Using the data of February 17 from Dataset 2 as example, Figure 35 manifests the spreading of cabs every 4 hours. It is evident that the GPS points gradually spread around the entire area every 4 hours. This figure also indicates that cabs that are moving obliviously in a city is able to cover the majority of roads. Consequently, they can perform the data collection and code dissemination tasks efficiently.

Following line charts (Figure 36 and Figure 37) illustrates the speed of code dissemination during a specific day. February 3 and February 16 are chosen to illustrate the correlation between the coverage of smart devices and time for

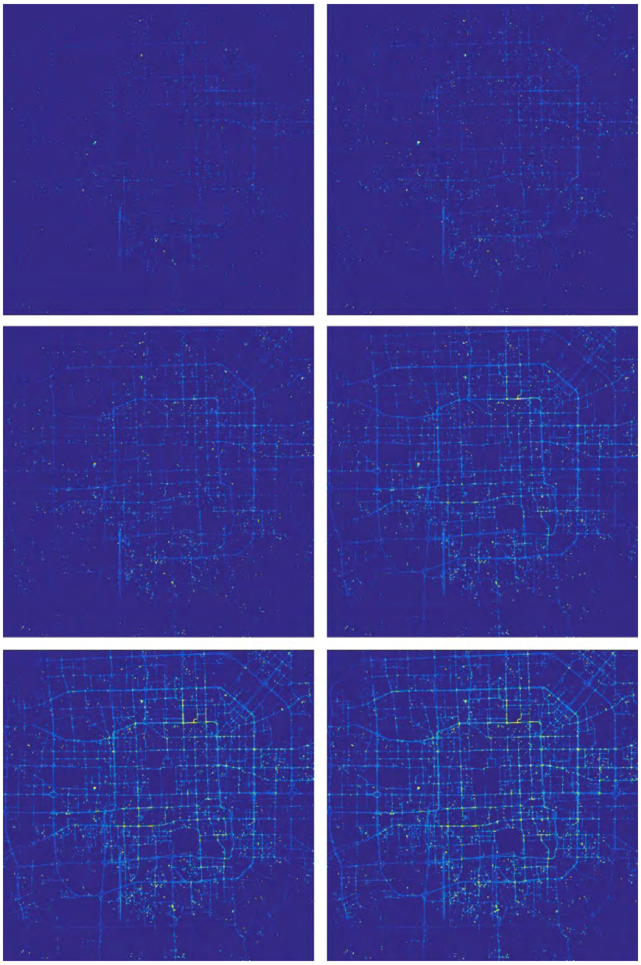

FIGURE 35. The density of GPS points every 4 hours on February 7.

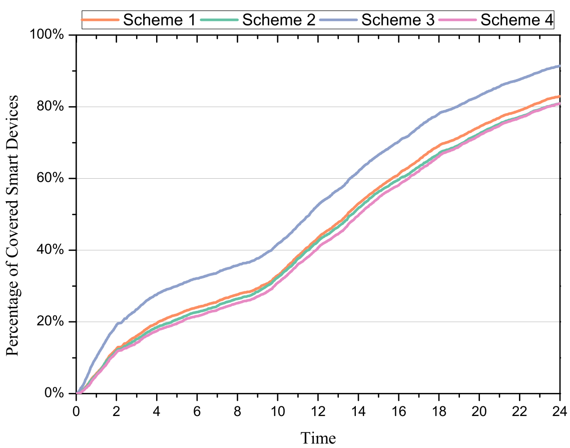

FIGURE 36. Coverage of smart devices on February 3 (Dataset 1).

Dataset 1 and Dataset 2 respectively. The coverage of smart devices denotes how many of them receive the new code when a piece of new code is published to update the configuration of smart devices. Since it is assumed that cabs do not share their LTE connection, they can only get the published code when they enter the communication range of data centers.

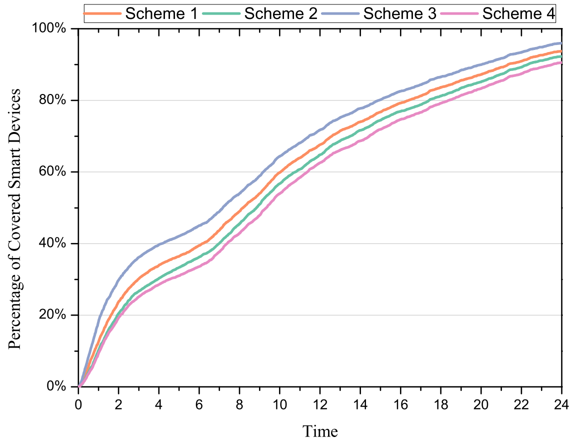

FIGURE 37. Coverage of smart devices on February 16 (Dataset 2).

As shown in Figure 36 and Figure 37, it is noticeable that for both datasets, all of the proposed schemes are able to cover more than 80% of smart devices within 24 hours. Scheme 3 outperforms remaining schemes, which denotes that the deployment of data center not only enhances the performance of data collection but also ameliorate that of code dissemination. As a result, for those applications that require less delay such as alter the working cycle of street lamps in foggy weather, the code can be disseminated quicker using Scheme 3.

VI. CONCLUSION

As sensing devices are becoming increasingly powerful and cheaper, Vehicular Networks joint sensing devices provide more perspectives for a variety of applications. Vehicular Networks can greatly facilitate the regulation and maintenance of infrastructures when they are used to collect status data of infrastructures. In a smart city, the regulation and maintenance of its infrastructures are mainly achieved by data collection and code dissemination tasks. From being generated by sensing devices to being used to subsidize decision making about the regulation of infrastructures, data collection mainly contains two procedures. Firstly, data are produced by sensing devices embedded in infrastructures, and they are collected by cabs. Following that, cabs uploaded collected to data centers. In contrast, code dissemination is achieved by firstly publishing data to data centers, which is achieved by transmitting the code to cabs within their communication range. Then cabs disseminate the received code to smart devices that need to be updated. Aiming to ameliorate the efficiency of data collection and code dissemination tasks by optimizing the deployment of data centers, three DCDO schemes are proposed in this paper. Each scheme is evaluated using two real-world datasets of cabs' GPS coordinates. Compared with Scheme 4 (February 3 of Dataset 1), Scheme 3 can enhance the total number of collected data packets by 57.71% when the number of data center is ten, and the speed of code dissemination and the coverage of cabs are ameliorated by 23.92% and 12.93% respectively. Compared with another proposed scheme—Scheme 1, the total number of Scheme 3's collected packets is 65.00% higher than that of Scheme 1 when there are ten data centers deployed. And the figures for Scheme 3's code dissemination and coverage of sensing devices exceed that of Scheme 1 by 18.98% and 10.21% respectively.

REFERENCES

[1] K. Sood and I. Mahajan, "Fog-cloud based cyber-physical system for distinguishing, detecting and preventing mosquito borne diseases," *Future Gener. Comput. Syst.*, vol. 88, pp. 764–775, Nov. 2018.

[2] T. Li, S. Tian, A. Liu, H. Liu, and T. Pei, "DDSV: Optimizing delay and delivery ratio for multimedia big data collection in mobile sensing vehicles," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3474–3486, Oct. 2018.

[3] J. Ren, H. Guo, C. Xu, and Y. Zhang, "Serving at the edge: A scalable IoT architecture based on transparent computing," *IEEE Netw.*, vol. 31, no. 5, pp. 96–105, Aug. 2017.

[4] X. Liu *et al.*, "Construction of large-scale low cost deliver infrastructure using vehicular networks," *IEEE Access*, vol. 6, pp. 21482–21497, 2018.

[5] S. Sarkar, S. Chatterjee, and S. Misra, "Assessment of the suitability of fog computing in the context of Internet of Things," *IEEE Trans. Cloud Comput.*, vol. 6, no. 1, pp. 46–59, Jan./Mar. 2018. doi: 10.1109/TCC.2015.2485206.

[6] *Internet of Things Market Forecast: Cisco*. Accessed: Sep. 21, 2018. [Online]. Available: http://postscapes.com/internet-of-things-market-size

[7] F. Xia, A. Rahim, X. Kong, M. Wang, Y. Cai, and J. Wang, "Modeling and analysis of large-scale urban mobility for green transportation," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1469–1481, Apr. 2018.

[8] X. Wang *et al.*, "A city-wide real-time traffic management system: Enabling crowdsensing in social Internet of vehicles," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 19–25, Sep. 2018.

[9] F. Xia, J. Wang, X. Kong, Z. Wang, J. Li, and C. Liu, "Exploring human mobility patterns in urban scenarios: A trajectory data perspective," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 142–149, Mar. 2018.

[10] H. Teng, W. Liu, T. Wang, A. Liu, X. Liu, and S. Zhang, "A cost-efficient greedy code dissemination scheme through vehicle to sensing devices (V2SD) communication in smart city," *IEEE Access*, to be published. doi: 10.1109/ACCESS.2019.2895899.

[11] X. Xiang, W. Liu, N. N. Xiong, H. Song, A. Liu, and T. Wang, "Duty cycle adaptive adjustment based device to device (D2D) communication scheme for WSNs," *IEEE Access*, vol. 6, pp. 76339–76373, 2018.

[12] H. Teng *et al.*, "A novel code data dissemination scheme for Internet of Things through mobile vehicle of smart cities," *Future Gener. Comput. Syst.*, vol. 94, pp. 351–367, May 2019.

[13] X. Kong, M. Li, T. Tang, K. Tian, L. Moreira-Matias, and F. Xia, "Shared subway shuttle bus route planning based on transport data analytics," *IEEE Trans. Automat. Sci. Eng.*, vol. 15, no. 4, pp. 1507–1520, Oct. 2018.

[14] J. Tan *et al.*, "A low redundancy data collection scheme to maximize lifetime using matrix completion technique," *EURASIP J. Wireless Commun. Netw.*, to be published. doi: 10.1186/s13638-018-1313-0.

[15] B. Huang, A. Liu, C. Zhang, N. Xiong, Z. Zeng, and Z. Cai, "Caching joint shortcut routing to improve quality of service for information-centric networking," *Sensors*, vol. 18, no. 6, p. 1750, May 2018. doi: 10.3390/s18061750.

[16] X. Peng, J. Ren, L. She, D. Zhang, J. Li, and Y. Zhang, "BOAT: A block-streaming app execution scheme for lightweight IoT devices," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1816–1829, Jun. 2018.

[17] T. Wang, G. Zhang, M. Bhuiyan, A. Liu, W. Jia, and M. Xie, "A novel trust mechanism based on fog computing in sensor-cloud system," *Future Gener. Comput. Syst.*, to be published. doi: 10.1016/j.future.2018.05.049.

[18] J. Ren, Y. Zhang, N. Zhang, D. Zhang, and X. Shen, "Dynamic channel access to improve energy efficiency in cognitive radio sensor networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3143–3156, May 2016.

[19] Z. Ning, X. Wang, J. Rodrigues, and F. Xia, "Joint computation offloading, power allocation, and channel assignment for 5G-enabled traffic management systems," *IEEE Trans. Ind. Informat.*, to be published. doi: 10.1109/TII.2019.2892767.

[20] Y. Ren, W. Liu, T. Wang, X. Li, N. Xiong, and A. Liu, "A collaboration platform for effective task and data reporter selection in crowdsourcing network," *IEEE Access*, to be published. doi: 1109/ACCESS.2019.2897062.

[21] Y. Liu, A. Liu, N. Zhang, X. Liu, M. Ma, and Y. Hu, "DDC: Dynamic duty cycle for improving delay and energy efficiency in wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 131, pp. 16–27, Jan. 2019.

[22] X. Li *et al.*, "Differentiated data aggregation routing scheme for energy conserving and delay sensitive wireless sensor networks," *Sensors*, vol. 18, no. 7, p. 2349, 2018. doi: 10.3390/s18072349.

[23] Z. M. Li, J. S. Gui, N. X. Xiong, and Z. W. Zeng, "Energy-efficient resource sharing scheme with out-band D2D relay-aided communications in C-RAN-based underlay cellular networks," *IEEE Access*, to be published. doi: 10.1109/ACCESS.2019.2891670.

[24] M. Huang *et al.*, "A services routing based caching scheme for cloud assisted CRNs," *IEEE Access*, vol. 6, pp. 15787–15805, 2018.

[25] Z. Tang, A. Liu, and C. Huang, "Social-aware data collection scheme through opportunistic communication in vehicular mobile networks," *IEEE Access*, vol. 4, pp. 6480–6502, 2016.

[26] Y. Xu, X. Chen, A. Liu, and C. Hu, "A latency and coverage optimized data collection scheme for smart cities based on vehicular ad-hoc networks," *Sensors*, vol. 17, no. 4, p. 888, 2017. doi: 10.3390/s17040888.

[27] Z. Ning, X. Wang, and J. Huang, "Non-orthogonal multiple access for mobile edge computing enabled vehicular networks," *IEEE Veh. Technol. Mag.*, to be published. doi: 10.1109/MVT.2018.2882873.

[28] X. Kong *et al.*, "Big trajectory data: A survey of applications and services," *IEEE Access*, vol. 6, pp. 58295–58306, 2018.

[29] Z. Liu, T. Tsuda, H. Watanabe, S. Ryuo, and N. Iwasawa, "Data driven cyber-physical system for landslide detection," *Mobile Netw. Appl.*, vol. 192, pp. 1–12, Mar. 2018. doi: 10.1007/s11036-018-1031-1.

[30] W. Zhao, "Performance optimization for state machine replication based on application semantics: A review," *J. Syst. Softw.*, vol. 112, pp. 96–109, Feb. 2016.

[31] M. Bonola, L. Bracciale, P. Loreti, P. Amici, A. Rabuffi, and G. Bianchi, "Opportunistic communication in smart city: Experimental insight with small-scale taxi fleets as data carriers," *Ad Hoc Netw.*, vol. 43, pp. 43–55, Jun. 2016.

[32] S. Fang *et al.*, "Feature selection method based on class discriminative degree for intelligent medical diagnosis," *CMC, Comput., Mater. Continua*, vol. 55, no. 3, pp. 419–433, 2018.

[33] J. Tan *et al.*, "An adaptive collection scheme-based matrix completion for data gathering in energy-harvesting wireless sensor networks," *IEEE Access*, vol. 7, pp. 6703–6723, 2019. doi: 10.1109/ACCESS.2019.2890862.

[34] F. Xiao, X. Xie, Z. Li, Q. Deng, A. Liu, and L. Sun, "Wireless network optimization via physical layer information for smart cities," *IEEE Netw.*, vol. 32, no. 4, pp. 88-93, Jul./Aug. 2018.

[35] W. Qi *et al.*, "Minimizing delay and transmission times with long lifetime in code dissemination scheme for high loss ratio and low duty cycle wireless sensor networks," *Sensors*, vol. 18, no. 10, p. 3516, 2018. doi: 10.3390/s18103516.

[36] Y. Liu, M. Ma, X. Liu, N. Xiong, A. Liu, and Y. Zhu, "Design and analysis of probing route to defense sink-hole attacks for internet of things security," *IEEE Trans. Netw. Sci. Eng.*, to be published. doi: 10.1109/TNSE.2018.2881152.

[37] A. Liu and S. Zhao, "High-performance target tracking scheme with low prediction precision requirement in WSNs," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 29, no. 4, pp. 270–289, 2018.

[38] Q. Deng *et al.*, "Compressed sensing for image reconstruction via back-off and rectification of greedy algorithm," *Signal Process.*, vol. 157, pp. 280–287, Apr. 2019.

[39] Y. Liu, A. Liu, X. Liu, and X. Huang, "A statistical approach to participant selection in location-based social networks for offline event marketing," *Inf. Sci.*, vol. 480, pp. 90–108, Apr. 2019.

[40] X. Ju *et al.*, "An energy conserving and transmission radius adaptive scheme to optimize performance of energy harvesting sensor networks," *Sensors*, vol. 18, no. 9, p. 2885, 2018. doi: 10.3390/s18092885.

[41] M. Huang, A. Liu, N. Xiong, T. Wang, and A. V. Vasilakos, "A low-latency communication scheme for mobile wireless sensor control systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 2, pp. 317–332, Feb. 2019.

[42] M. Huang, A. Liu, M. Zhao, and T. Wang, "Multi working sets alternate covering scheme for continuous partial coverage in WSNs," *Peer-Peer Netw. Appl.*, vol. 99, pp. 1–15, Apr. 2018. doi: 10.1007/s12083-018-0647-z.

[43] T. Li, K. Ota, T. Wang, X. Li, Z. Cai, A. Liu. "Optimizing the coverage via the UAVs with lower costs for information-centric Internet of Things," *IEEE Access*, to be published. doi: 10.1109/ACCESS.2019.2894172.

[44] Z. Li, Y. Liu, A. Liu, S. Wang, and H. Liu, "Minimizing convergecast time and energy consumption in green internet of things," *IEEE Trans. Emerg. Topics Comput.*, to be published. doi: 10.1109/TETC.2018.2844282.

[45] J. Ren, Y. Zhang, K. Zhang, A. Liu, J. Chen, and X. Shen, "Lifetime and energy hole evolution analysis in data-gathering wireless sensor networks," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 788–800, Apr. 2016.

[46] T. Li, N. Xiong, J. Gao, H. Song, A. Liu, and Z. Zeng, "Reliable code disseminations through opportunistic communication in vehicular wireless networks," *IEEE Access*, vol. 6, pp. 55509–55527, 2018.

[47] D. Tassos, "Privacy-respecting reward generation and accumulation for participatory sensing applications," *Pervasive Mobile Comput.*, vol. 49, pp. 139–152, Sep. 2018.

[48] T. Wang, G. Zhang, A. Liu, M. Z. A. Bhuiyan, and Q. Jin, "A secure IoT service architecture with an efficient balance dynamics based on cloud and edge computing," *IEEE Internet Things J.*, to be published. doi: 10.1109/JIOT.2018.2870288.

[49] C. Yang, Z. Shi, K. Han, J. Zhang, Y. Gu, and Z. Qin, "Optimization of particle CBMeMBer filters for hardware implementation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 9027–9037, Sep. 2018. doi: 10.1109/TVT.2018.2853120.

[50] C. Zhou, Y. Gu, S. He, and Z. Shi, "A robust and efficient algorithm for coprime array adaptive beamforming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1099–1112, Feb. 2017.

[51] P. L. Nguyen, Y. Ji, Z. Liu, H. Vu, and K. V. Nguyen, "Distributed hole-bypassing protocol in WSNs with constant stretch and load balancing," *Comput. Netw.*, vol. 129, pp. 232–250, Dec. 2017.

[52] C.-K. Tham and T. Luo, "Fairness and social welfare in service allocation schemes for participatory sensing," *Comput. Netw.*, vol. 73, pp. 58–71, Nov. 2014.

[53] Z. Song, C. H. Liu, J. Wu, J. Ma, and W. Wang, "QoI-aware multitask-oriented dynamic participant selection with budget constraints," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4618–4632, Nov. 2014.

[54] X. Deng *et al.*, "Cooperative channel allocation and scheduling in multi-interface wireless mesh networks," *Peer-Peer Netw. Appl.*, vol. 12, no. 1, pp. 1–12, Jan. 2019.

[55] X. Luo, Y. Lv, M. Zhou, W. Wang, and W. Zhao, "A laguerre neural network-based ADP learning scheme with its application to tracking control in the Internet of Things," *Pers. Ubiquitous Comput.*, vol. 20, no. 3, pp. 361–372, Jun. 2016.

[56] L. Cheng *et al.*, "Towards minimum-delay and energy-efficient flooding in low-duty-cycle wireless sensor networks," *Comput. Netw.*, vol. 134, pp. 66–77, Apr. 2018.

[57] Y. Jing *et al.*, "T-drive: driving directions based on taxi trajectories," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Jun. 2010, pp. 99–108.

[58] Y. Jing, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jun. 2011, pp. 316–324.

[59] [Online]. Available: http://wirelesslab.sjtu.edu.cn/taxi_trace_data.html

BAIXIANG HUANG is currently pursuing the degree with the School of Information Science and Engineering, Central South University, China. His research interests include information-centric networking and vehicular ad hoc networks.

WEI LIU received the Ph.D. degree in computer application technology from Central South University, China, in 2014. He is currently an Associate Professor and a Senior Engineer with the School of Informatics, Hunan University of Chinese Medicine, China. He has published over 20 papers in the related fields. His research interests include software engineering, data mining, and medical informatics.

TIAN WANG received the B.Sc. and M.Sc. degrees in computer science from Central South University, in 2004 and 2007, respectively, and the Ph.D. degree from the City University of Hong Kong, in 2011. He is currently an Associate Professor with Huaqiao University, China. His research interests include wireless sensor networks, social networks, and mobile computing.

XIONG LI received the master's degree in mathematics and cryptography from Shanxi Normal University, China, in 2009, and the Ph.D. degree in computer science and technology from the Beijing University of Posts and Telecommunications, China, in 2012. He is currently an Associate Professor with the School of Computer Science and Engineering, Hunan University of Science and Technology, China. He has published more than 70 referred journal papers in his research interests, which include cryptography, information security, and cloud computing security. He has served as a TPC Member for several international conferences on information security and as a Reviewer for more than 30 ISI indexed journals. He was a recipient of the *Journal of Network and Computer Applications* 2015 Best Research Paper Award. He is currently an Editor of *Telecommunication Systems* and the *KSII Transactions on Internet and Information Systems*.

HOUBING SONG (M'12–SM'14) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, USA, in 2012. In 2007, he was an Engineering Research Associate with the Texas A&M Transportation Institute. He served on the Faculty of West Virginia University, from 2012 to 2017. In 2017, he joined the Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL, USA, where he is currently an Assistant Professor and the Director of the Security and Optimization for Networked Globe Laboratory. He is the Editor of four books, including *Smart Cities: Foundations, Principles and Applications* (Hoboken, NJ, USA: Wiley, 2017), *Security and Privacy in Cyber-Physical Systems: Foundations, Principles and Applications* (Chichester, U.K.: Wiley-IEEE Press, 2017), *Cyber-Physical Systems: Foundations, Principles and Applications* (Boston, MA, USA: Academic Press, 2016), and *Industrial Internet of Things: Cybermanufacturing Systems* (Cham, Switzerland: Springer, 2016). He has authored more than 100 articles. His research interests include cyber-physical systems, cybersecurity and privacy, the Internet of Things, edge computing, big data analytics, unmanned aircraft systems, connected vehicle, smart and connected health, and wireless communications and networking. He serves as an Associate Technical Editor for the *IEEE Communications Magazine*.

Dr. Song is a Senior Member of the ACM. He was the very first recipient of the Golden Bear Scholar Award, the highest campus-wide recognition for research excellence at the West Virginia University Institute of Technology, in 2016.

ANFENG LIU received the M.Sc. and Ph.D. degrees in computer science from Central South University, China, in 2002 and 2005, respectively, where he is currently a Professor with the School of Information Science and Engineering. His major research interests are cloud computing, fog computing, and wireless sensor networks. He is a member of the China Computer Federation.