Wireless Communications and Mobile Computing

Emerging Technologies for Vehicular Communication Networks

Lead Guest Editor: Ning Zhang
Guest Editors: Ning Lu, Tao Han, Yi Zhou, and Dajiang Chen

Emerging Technologies for Vehicular Communication Networks

Wireless Communications and Mobile Computing

Emerging Technologies for Vehicular Communication Networks

Lead Guest Editor: Ning Zhang
Guest Editors: Ning Lu, Tao Han, Yi Zhou, and Dajiang Chen

Editorial Board

Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 2938121, 2 pages
https://doi.org/10.1155/2018/2938121

Editorial

Emerging Technologies for Vehicular Communication Networks

Ning Zhang,[1] **Ning Lu**,[2] **Tao Han**,[3] **Yi Zhou**,[4] **and Dajiang Chen**[5]

[1]*Texas A&M University at Corpus Christi, Texas, USA*
[2]*Thompson Rivers University, Kamloops, Canada*
[3]*University of North Carolina-Charlotte, Charlotte, USA*
[4]*Henan University, Kaifeng, China*
[5]*University of Electronic Science and Technology of China, Chengdu, China*

Correspondence should be addressed to Ning Zhang; ning.zhang@tamucc.edu

Received 2 September 2018; Accepted 2 September 2018; Published 22 October 2018

Next-generation intelligent transportation systems (ITS) are envisioned to greatly improve the transportation safety and efficiency by incorporating wireless communication and informatics technologies in the transportation system [1–3]. As the cornerstone for ITS, vehicular communication networks enable vehicles to exchange information with other vehicles and the external environments and play a significant role in supporting a variety of services such as road safety, traffic management, and entertainment

Vehicular communication networks face many technical challenges such as network scalability, highly dynamic topology, vulnerable wireless links, energy consumption of roadside units, poor network coverage, and bursty traffic. To address these challenges, various emerging technologies have been introduced in vehicular communication networks, such as software defined space-air-ground integrated vehicular network [4], fog computing in vehicular networks [5], drone-assisted vehicular networks [6], and machine learning for data delivery [7]. This special issue collection aims to present the vision, research, and dedicated efforts on the emerging technologies for vehicular communication networks. In this special issue, there are 15 submissions in total. After peer-review, 6 papers are selected for publication.

The first article, "Software-Defined Collaborative Offloading for Heterogeneous Vehicular Networks" by W. Quan et al., proposes a software-defined collaborative offloading (SDCO) solution for heterogeneous vehicular networks, to efficiently manage the offloading nodes and paths. The offloading controller is equipped with two specific functions: the hybrid awareness path collaboration (HPC) and the graph-based source collaboration (GSC). HPC is in charge of selecting the suitable paths based on the round-trip time, packet drop rate, and path capacity, while GSC optimizes the offloading nodes according to the minimum vertex cover for effective offloading.

The second article, "Performance Analysis of Space Information Networks with Backbone Satellite Relaying for Vehicular Networks" by J. Jiao et al., *studies the Space Information Network (SIN) with backbone satellites to relay information for vehicular networks, in order to support diverse vehicular services in a seamless, efficient, and cost-effective manner in rural areas and highways. The authors investigate the performance of SIN aided communications via an amplify-and-forward (AF) backbone satellite for VN communications*.

The third article, "Quality Utilization Aware Based Data Gathering for Vehicular Communication Networks" by Y. Ren et al., proposes a fine-grained data collection framework for achieving high-quality data gathering for vehicular communication networks. A metric named "Quality Utilization" (QU) is introduced to quantify the quality of the collected data. A Quality Utilization Aware Data Gathering (QUADG) scheme is proposed to collect the most appropriate data and to best satisfy the multidimensional requirements (mainly including data gathering quantity, quality, and cost) of vehicular applications.

The fourth article, "Concurrently Deniable Group Key Agreement and Its Application to Privacy-preserving VANETs" by S. Zeng and Y. Chen, presents a novel

transformation from an unauthenticated group key agreement to a deniable (authenticated) group key agreement to deal with the security issues in vehicular networks. An authenticated and privacy-preserving communication protocol is proposed for VANETs by using the proposed deniable group key agreement.

The fifth article, "Routing protocol in VANETs Equipped with Directional Antennas: Topology-Based Neighbor Discovery and Routing Analysis" by H. Li and Z. Xu, proposes a novel neighbor discovery algorithm which makes vehicles sense the topology changes around them and arrange their directional antennas accordingly. Moreover, a routing protocol is proposed for vehicular networks, which is based on the conventional epidemic routing protocol, whereby the vehicles make their routing decisions according to the information collected during the neighbor discovery process.

The sixth article, "A Fuzzy-Rule Based Data Delivery Scheme in VANETs with Intelligent Speed Prediction and Relay Selection" by Y. Zhou et al., aims to deal with the issues of data delivery in vehicular networks, such as high mobility and constant topological changes. A fuzzy-rule based wireless transmission approach is designed to optimize the relay selection, considering vehicle speed, driving direction, hop count, and connection time. In the proposed solution, each roadside unit (RSU) is equipped with a machine learning system (MLS) to make the selected relay link more reliable without GPS through predicting vehicle speed at next moment.

Conflicts of Interest

We declare that there are no conflicts of interest or private agreements with companies regarding our work for this special issue. We have no financial relationships through employment, consultancies, or either stock ownership or honoraria, with industry.

Ning Zhang
Ning Lu
Tao Han
Yi Zhou
Dajiang Chen

References

[1] A. Lei, H. Cruickshank, Y. Cao, P. Asuquo, C. P. A. Ogah, and Z. Sun, "Blockchain-Based Dynamic Key Management for Heterogeneous Intelligent Transportation Systems," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1832–1843, 2017.

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

[3] M. Azees, P. Vijayakumar, and L. J. Deboarh, "EAAP: Efficient Anonymous Authentication with Conditional Privacy-Preserving Scheme for Vehicular Ad Hoc Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 9, pp. 2467–2476, 2017.

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

[5] S. Bitam, A. Mellouk, and S. Zeadally, "VANET-cloud: A generic cloud computing model for vehicular Ad Hoc networks," *IEEE Wireless Communications Magazine*, vol. 22, no. 1, pp. 96–102, 2015.

[6] Y. Zhou, N. Cheng, N. Lu, and X. S. Shen, "Multi-UAV-Aided Networks: Aerial-Ground Cooperative Vehicular Networking Architecture," *IEEE Vehicular Technology Magazine*, vol. 10, no. 4, pp. 36–44, 2015.

[7] N. Kumar, N. Chilamkurti, and J. J. P. C. Rodrigues, "Learning automata-based opportunistic data aggregation and forwarding scheme for alert generation in vehicular ad hoc networks," *Computer Communications*, vol. 39, pp. 22–32, 2014.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 7637059, 15 pages
https://doi.org/10.1155/2018/7637059

Research Article

A Fuzzy-Rule Based Data Delivery Scheme in VANETs with Intelligent Speed Prediction and Relay Selection

Yi Zhou,[1,2] **Huanhuan Li**,[1,2] **Chenhao Shi**,[1,2] **Ning Lu**,[3] **and Nan Cheng**[4]

[1]*School of Computer and Information Engineering, Henan University, Kaifeng 475004, China*
[2]*International Joint Research Laboratory for Cooperative Vehicular Networks of Henan, Henan, China*
[3]*Department of Computing Science, Thompson Rivers University, Kamloops, BC V2C 0C8, Canada*
[4]*Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 1A1, Canada*

Correspondence should be addressed to Yi Zhou; zhouyi@henu.edu.cn

Received 15 December 2017; Revised 29 May 2018; Accepted 9 July 2018; Published 8 August 2018

Academic Editor: Enrico Natalizio

Data delivery in vehicular networks (VANETs) is a challenging task due to the high mobility and constant topological changes. In common routing protocols, multihop V2V communications suffer from higher network delay and lower packet delivery ratio (PDR), and excessive dependence on GPS may pose threat on individual privacy. In this paper, we propose a novel data delivery scheme for vehicular networks in urban environments, which can improve the routing performance without relying on GPS. A fuzzy-rule-based wireless transmission approach is designed to optimize the relay selection considering multiple factors comprehensively, including vehicle speed, driving direction, hop count, and connection time. Wireless V2V transmission and wired transmissions among RSUs are both utilized, since wired transmissions can reduce the delay and improve the reliability. Each RSU is equipped with a machine learning system (MLS) to make the selected relay link more reliably without GPS through predicting vehicle speed at next moment. Experiments show the validity and rationality of the proposed method.

1. Introduction

Vehicular Network (VANET) is a special type of the Mobile ad-hoc network (MANET) where every single node is a vehicle moving on the road. In addition to safety and privacy concerns [1], the challenges of studying VANET are mainly from the special characteristics of such networks: frequently link disconnections, rapidly topology changing, and large-scale sizes.

There are many important applications of VANETs, which are related to safety and nonsafety [2, 3]. Almost all applications are inseparable from message transmission; that is, data delivery is the cornerstone for the wider deployment of VANET applications [4]. In VANETs, the network protocols might fail due to frequent link disruptions caused by various factors, such as severe interference, interceptions, radio channel fading, and frequent topology changes. As a result, the connections are intermittent and many network services fail to function properly and their performance is seriously degraded. Therefore, developing efficient data delivery schemes with the presence of link disruptions is of great practical importance [5].

Fuzzy logic, known for its ability to deal with complex and imprecise problems, is a very promising technology in such a dynamic and complex context [6]. Meanwhile, massive traffic data in the transport system have important guiding significance to data delivery [7]. To efficiently utilize the dense data, algorithms skilled in dealing with historical data such as machine learning (ML) are needed. In order to improve the data transmission quality (DTQ), we propose a novel message delivery scheme which optimizes the V2V communication by optimizing the selection of relay nodes with the help of fuzzy rules and avoids the usage of GPS with the assistance of machine learning algorithms. For the unstable communication links in wireless transmission, we design a fuzzy-rule-based approach to select an optimal path from all the possible paths with considering multiple factors comprehensively. To reduce the usage of V2V

communication, RSUs are installed in the scene, which will decrease the network delay and packets loss probability by means of the rapid and stable transmission in a wired network. To deal with the dynamic of VANET, every RSU is embedded with a KNN-based machine learning system (MLS) to provide estimations about movements of vehicles and travelling paths.

The contributions of the paper are threefold, (1) our fuzzy-rule-based wireless transmission method can determine the optimal relay node in V2V communication, (2) a novel vehicle-based short-term speed prediction method with high practicability and flexibility is designed to enhance link stability, and (3) to defend client privacy and ensure system scalability our KNN-based machine learning system embedded in RSU supports GPS-free dynamic vehicle location prediction. The remainder of this paper is organized as follows. Section 2 describes the related work. Section 3 introduces the proposed model. Section 4 gives the detailed analysis of our system. Section 5 presents related simulation results. Section 6 concludes this paper.

2. Related Work

Xiang X et al. proposed two self-adaptive on-demand geographic routing protocols [8]. By adopting different schemes, the two protocols can obtain and maintain local topology information on data traffic demand. SOGR-HR (SOGR with Hybrid Reactive Mechanism) purely relies on one-hop topology information for forwarding as other geographic routing schemes; SOGR-GR (SOGR with Geographic Reactive Mechanism) combines both geographic and topology-based mechanisms for more efficient path building. The proposed SOGR-GR protocol achieves a better balance between control overhead and packet forwarding overhead. However, the core mechanism of SOGR-GR is purely based on traffic conditions and demands without considering the nodes' dynamicity and density. Our proposed strategy intelligently captures the dynamic changes in traffic operation through our machine learning system embedded in each RSU.

In [9], a machine learning assisted route selection (MARS) system is proposed to design routing protocols for urban environment. In order to predict the moves of vehicles and choose some suitable routing paths with better transmission capacity, the widely applied machine learning algorithm K-means is adopted. As an unsupervised cluster approach, K-means will judge the similarity of the data and decide which data can be grouped into the same cluster. However, sometimes this algorithm may be trapped into local optimum and different initial condition may produce different result because it is sensitive to initial centroids. Moreover, arithmetic mean is not robust to outliers, and very far data from the centroid may pull the centroid away from the real one. In our proposal, to avoid falling into local optimum, we select KNN algorithm instead of K-means.

In [10], a novel Fuzzy Logic based Greedy Routing (FLGR) protocol which focuses on transmitting safety messages with minimum delay is designed. The proposed FLGR is a position based greedy routing protocol that uses multiple metrics of neighbor vehicles to decide which vehicle is the best relay node by employing fuzzy logic. It selects the node with maximum distance, speed, and progress and minimum angular deviation from current forwarding node towards destination as the next hop. However, this protocol only considered the current state of vehicles and did not take the impact of future state into account when making decisions. In our fuzzy-rule-based method, we design a novel vehicle-based short-term speed prediction method to take future speed of vehicles into consideration.

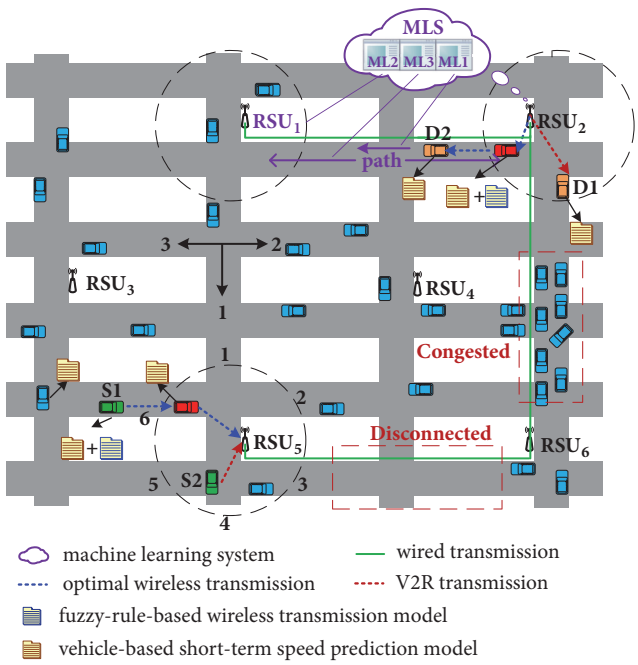


Figure 1: Data delivery model.

3. Proposed Model

In traditional routing protocols, V2V communication is the primary transmission mode. However, the vehicle density can greatly affect such protocols: (a) if the vehicle density is too high, competition for channel resources may incur collisions and result in lower packet delivery rate (PDR), and (b) if the vehicle density is too low, carry-and-forward technique will increase transmission delay in road segments with poor communication connectivity. In case the road segment is congested or disconnected, as shown in Figure 1, DTQ will be degraded. In addition, the selection of relay node plays the decisive role in the effect of wireless transmission. Not all the vehicles in transmission range are qualified for the relay node. A bad choice not only fails to deliver packet accurately and rapidly but also leads to high delay even packet loss. In general, the vehicle with the same driving direction as the source vehicle and having similar travelling speed with the source vehicle is more promising to be a good choice. Of course, other factors should also be taken into account if we intend to find the optimal choice. In order to connect to a certain vehicle, GPS is leveraged in most routing protocols to locate that node. However, excessive dependence

on GPS is unreliable: (a) GPS information is not always available especially in shielded areas, (b) GPS may induce some security issues and invade personal privacy, and (c) GPS sensor is power utilization equipment that sometimes users may turn off to decrease the power consumption, e.g., when a mobile phone is used as GPS device [11].

To solve the problems mentioned above, we propose a novel message delivery scheme which optimizes the V2V communication with the help of RSUs and fuzzy rules, and avoids the usage of GPS with the assistance of machine learning algorithms. Wired links between RSUs via backbone network can decrease the use of V2V communication obviously so as to reduce wireless transmission, so that not only higher packet delivery ratio but also lower network delay time are guaranteed. For the unavoidable wireless transmission portion, we design a fuzzy-rule-based method which takes driving direction, vehicle speed, connection time, and hop count into consideration to make the relay nodes selection more reasonable. To enhance the stability of V2V transmission link, a novel vehicle-based short-term vehicle speed prediction method is proposed to predict vehicle speed dynamically. In our proposed scheme, every RSU is embedded with a specially designed machine learning system to process dynamic traffic information and provide routing decisions. Our GPS-free dynamic vehicle location prediction can acquire the location of destination vehicle to eliminate those potential troubles caused by GPS.

We use the following notations:

(i) S: the source node

(ii) D: the destination node

(iii) RSU_s: the RSU which the source vehicle is connecting to

(iv) RSU_d: the RSU which the destination vehicle is connecting to or has just left

(v) RSU_n: the RSU which the destination will visit next

(vi) Blind zone: areas not covered by any RS

In the process of delivering packets, the source vehicle will first attempt to access a RSU. If source vehicle is covered by RSU_s, V2R communication will be selected, since On Board Unit (OBU) can communicate with RSU employing Dedicated Short Range Communication (DSRC) technology to implement information transmission [12, 13]. If source node is in the blind zone, the fuzzy-rule-based approach will be employed for establishing the communication link. First, it will find potential paths among all the routes. On the basis of fuzzy mathematics, four factors (driving direction, vehicle speed, connection time, and hop count) are considered. During this process, a novel vehicle-based short-term vehicle speed prediction method is proposed to predict vehicle speed dynamically. And then, evaluate every potential path. Each path will be evaluated according to the fuzzy comprehensive evaluation method (FCEM) and assigned an integrated assessment value and its corresponding evaluation grade based on the maximum membership principle. Finally, select wireless transmission path according to the evaluation results. Our scheme will first select the optimal path to delivery packets; in case the RSU_s fails to receive the packets after a tolerable threshold time, the suboptimum path will be selected.

After receiving the packets, RSU_s will send the Transmission Request (TREQ) to other RSUs to find RSU_d. Each RSU will maintain a Node Table, which lists the vehicles in its coverage in this moment and those vehicles before with the time when they left. Every Node Table stores the traffic information collected by RSU in real time and is updated periodically. By checking respective tables, each RSU can judge whether to respond to the TREQ or not. Only RSU_d sends the Transmission Reply (TREP) to RSU_s to inform the location state of the destination vehicle, and then RSU_s delivers packets to RSU_d. RSUs installed on the road sides are important components for the Intelligent Transportation System (ITS). Backbone network makes it possible to interconnect between RSUs which is used between RSU_s and RSU_d. Compared with wireless networks, wired networks are more stable and speedy, especially in such a highly dynamic environment.

Finally, the packets will be delivered to the D in V2R mode in case it is covered by a RSU, otherwise the transmission path will be predicted with the assist of machine learning system. First, machine learning 1 (ML1) will predict destination's turning direction at next intersection after leaving RSU_d: go straight, turn left, or turn right. And then, machine learning 2 (ML2) will predict the probability that destination node travels into each RSU to determine RSU_n. Last, machine learning 3 (ML3) will predict the travelling path from a certain exit of RSU_d to RSU_n. After the travelling path is determined, packets will be transferred in a two-way mode; that is, both RSU_d and RSU_n are dedicate to searching connected ways to D along the predicted path. In this section, fuzzy-rule-based approach is used again. If RSU_d finds the path faster than RSU_n, it will deliver packets to D in unicast mode and, meanwhile, inform RSU_n to stop searching. Otherwise, RSU_d will relay packets to RSU_n by wired way to complete the dissemination. In case the destination vehicle fails to receive packets after a tolerable threshold time, the scheme will select the RSU with the second highest possibility and perform the two-way transfer again and so on.

As shown in Figure 1, S2 can transmit packets to RSU_5 via wireless V2R communication, while S1 needs fuzzy-rule-based approach for the selection of a relay vehicle. And then after checking Node Table, RSU_5 delivers packets to RSU_2 by means of wired transmission. For destination vehicle D1, RSU_2 can deliver packets to it directly. For destination vehicle D2, a transmission path will be established with the assistance of machine learning system in RSU_2. Our proposal makes data transmission less dependent on vehicle density and avoids the negative issues resulted from the usage of GPS. Algorithm 1 presents the whole message delivery process concisely. More detailed information will be described in next part.

4. Data Delivery Scheme

4.1. Fuzzy-Rule-Based Wireless Transmission Method. Multihop broadcasting schemes are particularly preferred in

(1) initialization: determine the relationship between RSUs and vehicles periodically
(2) **if** S is covered by a RSU
(3) deliver packets to RSU_s in V2R mode
else
fuzzy-rule-based wireless transmission method
vehicle-based short-term vehicle speed prediction
(4) *(1) find potential paths*
(5) *(2) evaluate potential paths*
(6) *(3) determine wireless transmission path (optimum)*
(7) S sends packets to RSU_s
(8) **if** RSU_s fails to receive packets after threshold time
(9) select another path (sub-optimum)
(10) go to step (7)
else
(11) end
(12) RSU_s sends packets to RSU_d
(13) **if** D is covered by a RSU
(14) deliver packets to RSU_d in V2R mode
else
machine learning system
(15) *(1) ML1 predicts D's turning direction*
(16) *(2) ML2 predicts RSU_n (highest possibility)*
(17) *(3) ML3 predicts travelling path*
(18) two-way mode transfer
fuzzy-rule-based wireless transmission method
(19) **if** D fails to receive packets after threshold time
(20) select RSU_n again (second highest possibility)
(21) go to step (17)
else
(22) end

Algorithm 1

wireless transmission to transmit information to vehicles or infrastructures that cannot communicate directly in VANET. However, ordinary broadcasting may suffer from frequent contention and serious collision [14] and thus cause broadcast storms [15]. How to suppress broadcast storm is one of the important issues in the wireless transmission process [16, 17]. In order to alleviate broadcast storm and guarantee fast and efficient messages dissemination, this paper proposes a fuzzy-rule-based wireless transmission approach to provide an optimal forwarder selection scheme.

4.1.1. Structure of Wireless Transmission Method

(a) Finding Potential Paths. First, the source vehicle will initiate a Routing Request (RREQ) in its transmission range, which indicates whether or not it can reach a RSU. Then neighboring nodes, hearing this request advertisement, will rebroadcast this beacon to their neighbors. Finally RSU receiving this beacon will send back a Routing Reply (RREP) to announce the RSU_s and the route. This process is repeated until any one of the terminal conditions is met, (a) preset end time and (b) preset the number of routes that satisfy the following conditions. All the potential routes found should guarantee that the RREQ is heard by some RSU within a certain number of hops (default two hops). As shown in

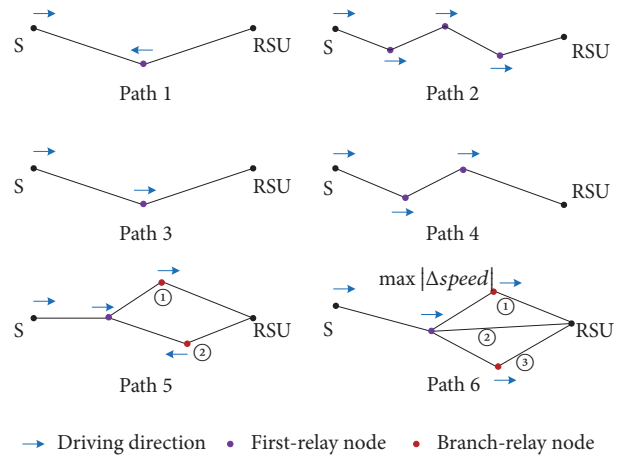


Figure 2: Judgment of potential paths.

Figure 2, all the found paths are qualified as potential routes except for path1 and path2. If there are two or more branching paths after the first-relay node, pre-evaluation is needed to select branch-relay node. The principle of pre-evaluation is defined as follows: (1) abandon branch-relay nodes driving in the opposite direction or with the maximum speed difference

with the first-relay node; (2) the subpath without branch-relay node is deemed to have the same direction and speed with the first-relay node; (3) the upper bound of the number of subpath is two. As shown in Figure 2, after pre-evaluation, path5 will select ① as the branch-relay node and path6 will select ② and ③, and path3 and path4 do not need pre-evaluation.

(b) Evaluating Potential Paths. Paths are evaluated according to the fuzzy comprehensive evaluation method, wherein the index weight vector is determined based on the analytic hierarchy process (AHP). Fuzzy logic, first mentioned by Zadeh in 1965 [18], allows uncertain information to be processed by using simple IF-THEN rules. This laid the foundation for the future development of fuzzy theory [19]. Known for its ability to deal with complexity and imprecision problems, fuzzy logic is a powerful mathematical tool to deal with multiparameter problems in such a dynamic and complex context [20]. On the basis of fuzzy mathematics, four factors are considered to derive a fuzzy relation matrix R for each potential path. And according to this matrix, our method will assign every potential route an integrated assessment value and its corresponding evaluation grade.

(c) Determining Wireless Transmission Path. The last stage of our fuzzy-ruled-based wireless transmission approach is to select wireless transmission path according to the evaluation results. Our scheme will first select the optimal path to delivery packets; in case the RSU fails to receive the packets after a tolerable threshold time, the suboptimum path will be selected. For paths belonging to different evaluation grades, higher grades have priority over lower grades; for paths belonging to the same grade, higher assessment value has priority over lower assessment value.

For example, in Figure 3, vehicle s sends out a RREQ and its neighbors will rebroadcast this beacon. As such, through vehicle a and vehicle b as the relay nodes, a RSU can be found to receive the RREQ and then reply to RREP. In this way, we find a potential path $s \longrightarrow a \longrightarrow b \longrightarrow RSU$. After route evaluation, the optimal path for the wireless communication connection to a RSU will be determined. Now, source node can unicast the packets to that RSU along the path carried by RREP.

4.1.2. Relay Node Selection. With a view to select a more reliability relay node, we employ FCEM to combine several influential factors to conduct an evaluation on each potential path. FCEM is a scientific assessment method based on fuzzy mathematics with the considerations of a plurality of influence factors [21]. It uses fuzzy logic to systematically perform evaluation of real world systems that are not clearly defined.

(a) Confirming the Evaluation Index Set. In order to evaluate whether or not a vehicle is qualified for the next relay node, index set is defined as $U = \{f_1, f_2, f_3, f_4\} = \{|\Delta V_n|, |\Delta D|, HC, CT\}$, where f_1 is the absolute value of speed difference at next moment between current vehicle and the potential vehicle, f_2 is the absolute value of driving direction difference between that two vehicles, f_3 is the hop count from the potential vehicle to a RSU, and f_4 is the corresponding connection time. The speed difference is the first index to measure a potential vehicle since similar speed is a necessary condition for stable vehicle distance. The concrete description of vehicle speed prediction will be presented in the next section. The driving direction difference is another key index, if two vehicles are driving in the opposite direction and the stability of the link must be hard to ensure. In addition, hop count and connection time are another two important influencing factors, and too many hops or too long connection time will affect the link stability seriously.

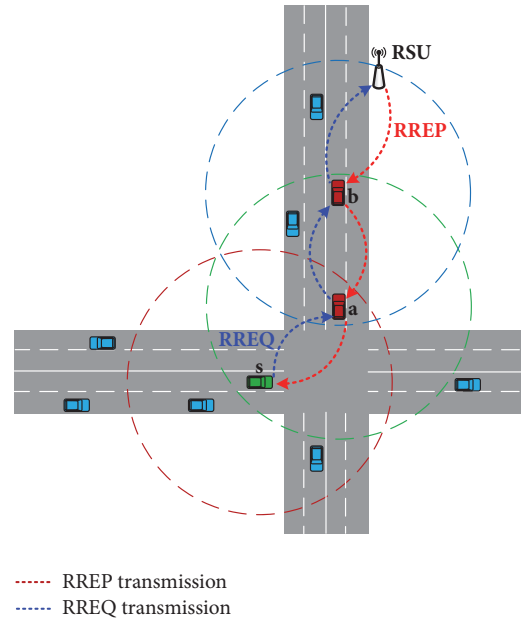


Figure 3: Finding potential wireless transmission path.

(b) Confirming the Evaluation Criteria Set. The degree of satisfaction is measured using the so-called remark set that consists of a set of linguistic variables such as "good" or "bad" [22]. In order to evaluate whether or not a vehicle is qualified for the next relay node, we define the evaluation criteria set with four ratings $v = \{very\ good, good, general, bad\}$.

(c) Determining the Index Weight Vector. According to theory of the analytic hierarchy process, the index weight vector is set to be $P = [0.32 \quad 0.4492 \quad 0.0957 \quad 0.1351]$. More detailed information about the AHP method will be introduced in the following.

(d) Constructing the Fuzzy Relation Matrix. The fuzzy relation matrix is defined as (1), where r_{ij} indicates the membership of the i_{th} index belonging to the j_{th} rate. The membership function is established according to the characteristics of the index system. For the discrete variables f_2 and f_3, membership grade is determined according to Tables 1 and 2. For continuous variables f_1 and f_4, the membership function is defined as shown in Figures 4 and 5. For variable f_4, fuzzy set function is calculated as (2)–(5), and we can get the

Table 1: Membership function of f_2.

f_2	v			
	v1	v2	v3	v4
0	1	0	0	0
1	0	0.5	0.5	0
2	0	0	0	1

Table 2: Membership function of f_3.

f_3	v			
	v1	v2	v3	v4
0	1	0	0	0
1	0	1	0	0
2	0	0	1	0

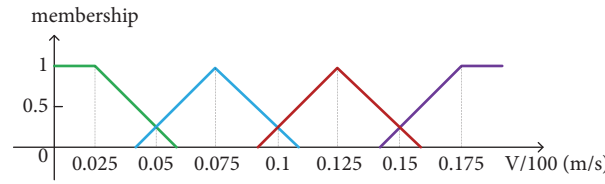


Figure 4: Membership function of f_1.

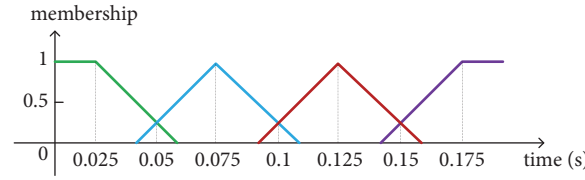


Figure 5: Membership function of f_4.

membership function of variable f_1 by substituting variate $v/100$ for t.

$$R = \begin{bmatrix} \mu_{\widetilde{v1}}^{f_1} & \mu_{\widetilde{v2}}^{f_1} & \mu_{\widetilde{v3}}^{f_1} & \mu_{\widetilde{v4}}^{f_1} \\ \mu_{\widetilde{v1}}^{f_2} & \mu_{\widetilde{v2}}^{f_2} & \mu_{\widetilde{v3}}^{f_2} & \mu_{\widetilde{v4}}^{f_2} \\ \mu_{\widetilde{v1}}^{f_3} & \mu_{\widetilde{v2}}^{f_3} & \mu_{\widetilde{v3}}^{f_3} & \mu_{\widetilde{v4}}^{f_3} \\ \mu_{\widetilde{v1}}^{f_4} & \mu_{\widetilde{v2}}^{f_4} & \mu_{\widetilde{v3}}^{f_4} & \mu_{\widetilde{v4}}^{f_4} \end{bmatrix} = \left(r_{ij}\right)_{m\times n} \tag{1}$$

$$\mu_{\widetilde{v1}}^{f_4} = \begin{cases} 1 & 0 \le t \le \dfrac{1}{40} \\ -30t + 1.75 & \dfrac{1}{40} \le t \le \dfrac{7}{120} \\ 0 & others \end{cases} \tag{2}$$

$$\mu_{\widetilde{v2}}^{f_4} = \begin{cases} 30t - 1.25 & \dfrac{1}{24} \le t \le \dfrac{3}{40} \\ -30t + 3.25 & \dfrac{3}{40} \le t \le \dfrac{13}{120} \\ 0 & others \end{cases} \tag{3}$$

$$\mu_{\widetilde{v3}}^{f_4} = \begin{cases} 30t - 2.75 & \dfrac{11}{120} \le t \le \dfrac{1}{8} \\ -30t + 4.75 & \dfrac{1}{8} \le t \le \dfrac{19}{120} \\ 0 & others \end{cases} \tag{4}$$

$$\mu_{\widetilde{v4}}^{f_4} = \begin{cases} 30t - 4.25 & \dfrac{17}{120} \le t \le \dfrac{7}{40} \\ 1 & \dfrac{7}{40} \le t \\ 0 & others \end{cases} \tag{5}$$

(e) Determining Comprehensive Evaluation Class. By performing the fuzzy composite operation between the index weight vector and the fuzzy relation matrix, a comprehensive evaluation vector model is established as shown in (6), where b_1, b_2, b_3, and b_4 represent the four ranks of the evaluation set, respectively. We determine the corresponding evaluation grade according to the maximum membership principle.

$$B = P \circ R = [b_1, b_2, b_3, b_4] \tag{6}$$

4.1.3. Index Weight Vector. The index weight vector represents the different weights of the selected four impact factors in the forwarding node selection. In order to perform the fuzzy composite operation between the index weight vector P and the fuzzy relation matrix R, we need to determine values for vector P exactly. In this paper, we employ analytic hierarchy process [23] to assign the optimal weight for every index to identify optimal forwarder. AHP decomposes the complex problem into a hierarchy of subproblems to evaluate the relative importance of each criterion [24]. The alternatives are chosen according to their weights towards each criterion and ultimately towards the goal [25].

(a) Analytic Hierarchy Structure. An important part of AHP is to structure the analytic hierarchy: (i) to state the objective; (ii) to define the criteria; (iii) to choose the alternatives [15]. Based on requirements in the scenario stated in this paper, the objective (the top level) is to select an optimal forwarder. The criteria (the medium level) include the absolute value of speed difference between current vehicle and the potential vehicle, the absolute value of direction difference between two vehicles, the hop count from the potential vehicle to a RSU, and the corresponding connection time. The alternatives (the bottom level) include all the candidate nodes within the communication range of the source node. The hierarchical tree is formed based on these three layers as shown in Figure 6, where node1, ..., nodeC represent the candidate forwarder vehicles.

(b) Judgment Matrix. According to the degree of importance to relay node selection, the importance ranking of four influence factors is defined as $f_2 > f_1 > f_4 > f_3$. The criteria are pairwise compared to find their importance towards the goal, and such a pairwise comparison is represented as a judgment matrix shown in (7), where n is the total number

Table 3: Relative importance between criteria.

A_{ij}	definition
1	i and j are equally important
3	i is moderately more important than j
5	i is strongly more important than j
1/3	i is moderately less important than j
1/5	i is strongly less important than j
2,4, 1/2, 1/4	intermediate values

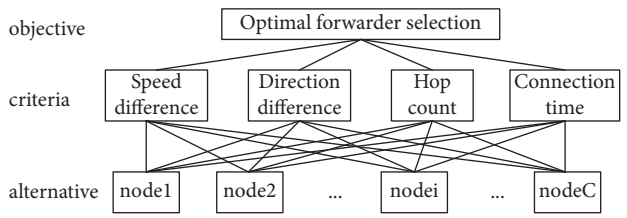


Figure 6: Analytic hierarchy structure.

of criteria. In the matrix, A_{ij} denotes the relative importance of criteria i to j. $A_{ij} = 1$ indicates that index i is as important as index j, $A_{ij} > 1$ indicates that index i is more important than index j, and $0 < A_{ij} < 1$ indicates that index i is less important than index j. The property of judgment matrix is also shown in (7) [25]. The relative importance of one criterion over another can be expressed in pairwise comparison matrix according to Table 3.

$$\begin{array}{c|cccc} A & f_1 & f_2 & f_3 & f_4 \\ \hline f_1 & A_{11} & A_{12} & A_{13} & A_{14} \\ f_2 & A_{21} & A_{22} & A_{23} & A_{24} \\ f_3 & A_{31} & A_{32} & A_{33} & A_{34} \\ f_4 & A_{41} & A_{42} & A_{43} & A_{44} \end{array} \tag{7}$$

$$(1)\ A_{ij} > 0,\ (2)\ A_{ij} = \frac{1}{A_{ji}}\ (i \neq j),\ (3)\ A_{ij} = 1\ (i = j)$$

As such, we get the judgment matrix A, and matrix B is the result after normalization.

$$A = \begin{bmatrix} 1 & \frac{1}{2} & 4 & 3 \\ 2 & 1 & 4 & 3 \\ \frac{1}{4} & \frac{1}{3} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{3} & 2 & 1 \end{bmatrix} \tag{8}$$

$$B = \begin{bmatrix} 0.2857 & 0.2308 & 0.3636 & 0.4 \\ 0.5714 & 0.4615 & 0.3636 & 0.4 \\ 0.0714 & 0.1538 & 0.0909 & 0.0667 \\ 0.0714 & 0.1538 & 0.1818 & 0.1333 \end{bmatrix}$$

(c) Consistency Examination. The last step of AHP is consistency examination to check whether the comparison matrix A is consistent or not [26]. The judgment errors are detected using the consistency ratio (CR), which is the ratio of the consistency index (CI) to the random index (RI). The CI value is calculated as (9), where n is the number of decision factors, $\lambda_{\max}$ is maximal eigenvalue of matrix A, and RI values are shown in Table 4 [27]. The errors in judgments are considered tolerable when CR ≤ 0.1; otherwise, the pairwise comparisons need to be adjusted [25]. After calculation, the consistency ratio $CR = 0.0526 < 0.1$ in the judgment matrix, so the matrix A meets the compliance requirements. After normalization, the index weight vector is set to be $P = [0.32\ \ 0.4492\ \ 0.0957\ \ 0.1351]$.

$$CI = \frac{\lambda_{\max} - n}{n - 1} = \frac{4.1421 - 4}{4 - 1} = 0.0474 \tag{9}$$

$$\begin{aligned} CR &= 0.0526 < 0.1 \\ \lambda_{\max} &= 4.1421 \\ CR &= \frac{CI}{RI} < 0.1 \end{aligned} \tag{10}$$

4.2. Vehicle-Based Short-Term Speed Prediction. Short-term vehicle speed prediction is one of the most critical components of an ITS. Real-time and accurate vehicle speed prediction is the key to traffic control and traffic guidance and provides important information for intelligent vehicles and transportation applications.

Although there has been a growing body of studies on short-term vehicle speed prediction approaches, most methods belong to segment-based methods [28, 29]. Segment-based models predict vehicle speed for a certain road by analyzing historical traffic data collected from one or more road fracture surfaces. For example, a short-term traffic speed prediction model which predicts the traffic speed on a route containing more than one road link is developed based on a support vector machine model in [28]. Based on the SVM algorithm, the temporal information of the target road link and traffic speed of upstream/downstream road links are considered. Segment-based methods are suitable for vehicle navigation if historical traffic data about related road segments are acquired, but the inherent disadvantage of these methods is the poor scalability. The predictive ability is restricted by the vehicle location that is whether or not our database contains the historical traffic information about the road vehicle is travelling on. Besides, segment-based vehicle speed prediction model cannot capture the subtle fluctuations caused by routine traffic flows and the sudden disruption caused by accidents, since those abnormal data are often eliminated in data process stage.

In order to deal with these problems, a novel vehicle-based short-term vehicle speed prediction model, based on the weighted K-Nearest Neighbor algorithm (W-KNN), is introduced in this study. The predictive ability of our model is restricted by related traffic data which are obtained by OBU employing DSRC technology instead of the vehicle location. Since vehicle speed prediction is based on the latest data, the impact caused by subtle fluctuations will soon be reflected. The travel speed at next moment is affected by speed of the

Table 4: Mean random consistency index RI.

n	1	2	3	4	5	6	7	8	9	...
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	...

moment and speed in the past, and the closer the time, the greater the impact. In addition, in an urban road network, road links do not exist in isolation. Traffic conditions on both upstream and downstream road segments can affect the vehicle speed of the current road segment [28]. To improve the prediction accuracy, we synthetically consider limited spatial and temporal influence factors in our proposal. For the temporal domain, we dynamically select real-time traffic data and historical data within certain time lags to provide valuable sample data for each prediction. For the spatial domain, the testing vehicle and vehicles in its communication range are considered since vehicle's travel speed is affected by other vehicles around.

Based on the analysis above, we determine the feature vector $[V, a, C, \Delta C]$, which is composed of velocity, acceleration, the vehicle count in the testing sample's communication range, and the vehicle count gradient. For each potential neighbor $x(i, j)$, the state vector is defined as

$$\begin{aligned} S(i,j) &= \left[V(i,j), a(i,j), C(i,j), \Delta C(i,j)\right] \\ &= \left[V_i(t - j\Delta t), a_i(t - j\Delta t), C_i(t - j\Delta t), \right. \\ &\quad \left. \Delta C_i(t - j\Delta t)\right] \quad j = 1, 2, \ldots, l,\ i = 0, 1, 2, \ldots, C, \end{aligned} \tag{11}$$

where $i = 0$ indicates the testing vehicle, $i = 1, 2, \ldots, C$ indicates other vehicles in its communication range, and C is the count of vehicles in its communication range. Δt is the time interval between two data collections. l is the number of time interval Δt to determine the time lag of the historical data. The label for potential neighbor $x(i, j)$ will be $y(i, j) = V_i(t - (j - 1)\Delta t)]$. For example, a training data collected from the third vehicle in the testing vehicle's communication range with two time interval lags can be written as $S(3, 2) = [V_3(t - 2\Delta t), a_3(t - 2\Delta t), C_3(t - 2\Delta t), \Delta C_3(t - 2\Delta t)]$, and the corresponding label is $y(3, 2) = V_3(t - \Delta t)$. In order to predict a vehicle's travel speed at next moment, we should provide the input vector at time t $X = [V_0(t), a_0(t), C_0(t), \Delta C_0(t)]$ to the prediction model, and then predicted result $\widehat{V_0}(t + \Delta t)$ will be output for this testing vehicle $X_0(t)$.

We have analyzed and compared three different prediction models to select a suitable one with higher accuracy in this section. The moving average data-based (MAD) model adopts one of the simplest techniques, a straight average of the dependent variables. In this model, the travel speed at next moment is predicted by the speed in the previous time period. The data used for prediction is the previous data closest to the testing time t [28]. The observed equation of the moving average data-based model is shown in

$$\widehat{V_0}(t + \Delta t) = \frac{1}{1 + n} \sum V_0(t - j\Delta t) \quad j = 0, 1, 2, \ldots, n, \tag{12}$$

where n is the number of previous time periods used in model. This does consider all the dependent variables evenly but does not consider the relationship between the testing data and each training data. The differences between samples are neglected, and each training sample is considered to make the same contribution. And also, this model does not take the impact from spatial domain into account. Several simulation experiments to explore the relationship between parameter n and the prediction accuracy are conducted as shown in Figure 2.

K-Nearest Neighbor learning, one of the most popular realizations of IBL (instance based learning), combines the target values of K selected neighbors to predict the target value of a given test pattern. Once the state vector is defined, the next step is the selection of a suitable determinant to measure the closeness between the testing sample and each candidate neighbor in the training data set. Rank results based on the closeness information will determine the member of the neighborhood. The similarity is usually based on the Minkowski distance metrics, wherein the Lr distance, as written in (13), is referred to as r = {1, 2, ..., max} in the {Manhattan distance, Euclidean distance, ..., Chebyshev distance} metrics [29]. Due to the dynamic nature of VANETs, the traffic condition shows successive fluctuations. In other words, the time-series traffic state is a highly dynamical system with uncertain noise, which is a meaningful signal for the future state. The Euclidean distance is sensitive to noise, so the abnormal variation can be real-timely captured when the current state is either seriously disturbed or rapidly changed. And this is why Euclidean distance has been most frequently used to measure the similarity in NPR-based (non-parametric regression) traffic variable prediction [30–32].

$$L_r = \left[\sum \left|x(i,j) - X_0(t)\right|^r\right]^{(1/r)} \tag{13}$$

In this study, the Euclidean distance ($r = 2$), $ED(i, j)$, between $x(i, j)$ and $X_0(t)$, is used, which can be written as (14). Given a testing sample $V_0(t)$, the Euclidean distance metric is used to obtain the K-nearest neighbors and their corresponding labels from the training data set. Hence, the potential output vector Y_i, associated with the *ith* nearest neighbors, is made up of two values as shown in (15). The observed equation with nonweighted KNN model is shown in (16).

$$\begin{aligned} ED(i,j) &= d\left(x(i,j), X_0(t)\right) = \left[\left|V(i,j) - V_0(t)\right|^2 \right. \\ &\quad + \left|a(i,j) - a_0(t)\right|^2 + \left|C(i,j) - C_0(t)\right|^2 \\ &\quad \left. + \left|\Delta C(i,j) - \Delta C_0(t)\right|^2\right]^{1/2} \end{aligned} \tag{14}$$

$$Y_i = \left[V_i, ED_i\right] \quad i = 1, 2, \ldots, K \tag{15}$$

$$\widehat{V_0}(t + \Delta t) = \frac{1}{K} \sum V_i \quad i = 1, 2, \ldots, K \tag{16}$$

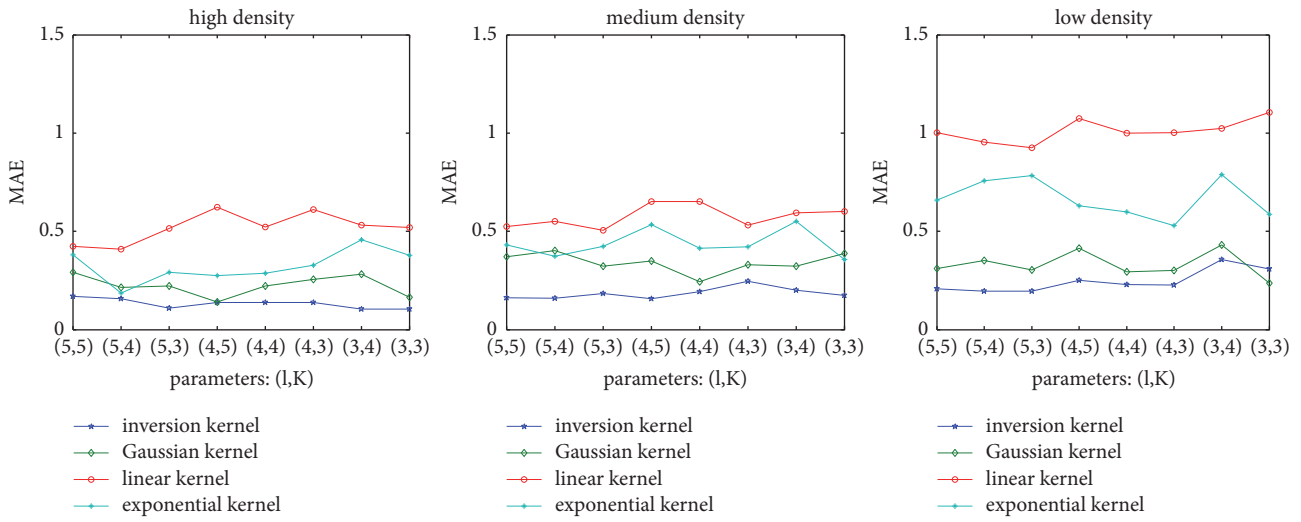


Figure 7: Results of three kernel functions.

One major challenging issue of the nonweighted KNN model is that its performance strongly depends on one key model parameter: the number of nearest neighbors K. In fact, no well-established method exists for selecting an optimal K when using the KNN algorithm. The number of nearest neighbors is often chosen empirically by cross-validation or domain experts in practice [33]. Therefore, a simple empirical or experimental test [3–5] is sufficient to find a suitable K-value. KNN considers the correlation between the testing vehicle and other vehicles by selecting the K-nearest neighbors according to the Euclidean distance metrics, but the K selected neighbors are treated equally without considering their differences according to (16).

For weighted-KNN model, in addition to the number of neighbors, the weights about those neighbors are another key parameter. Regarding the weights to the neighbors, the rule of thumb so far is "A father neighbor gets a smaller weight" [33]. That is a farther neighbor receives a smaller weight, which reduces its effect on the prediction results compared to other closer neighbors. There are a number of well-known kernel functions, which decrease monotonically as distance increases, such as the linear kernel $w_i = 1 - d(x_i, X_0(t))$ [34], the inversion kernel $w_i = (d(x_i, X_0(t)))^{-1}$ [35], the exponential kernel $w_i = \exp(-d(x_i, X_0(t)))$ [36], and the Gaussian kernel $w_i = \exp(-d(x_i, X_0(t))^2)$ [37]. Atkerson et al. [35] claimed that there is no clear evidence that any kernel function is always superior to the others, but some outperformed others on some data sets [38]. In order to select an appropriate kernel function to acquire better prediction precision for our weighted-KNN model, we performed simulation experiments using these four kernel functions. MAE (Mean Absolute Error) [24], as shown in (19), is introduced to evaluate each kernel. The larger the value of MAE, the greater the prediction error, and in contrary, the lower prediction accuracy. As shown in Figure 7, in terms of MAE, the inversion kernel which endows different weights for every element in the K select samples by the inverse of the corresponding Euclidean distances (18) is the best choice in our model.

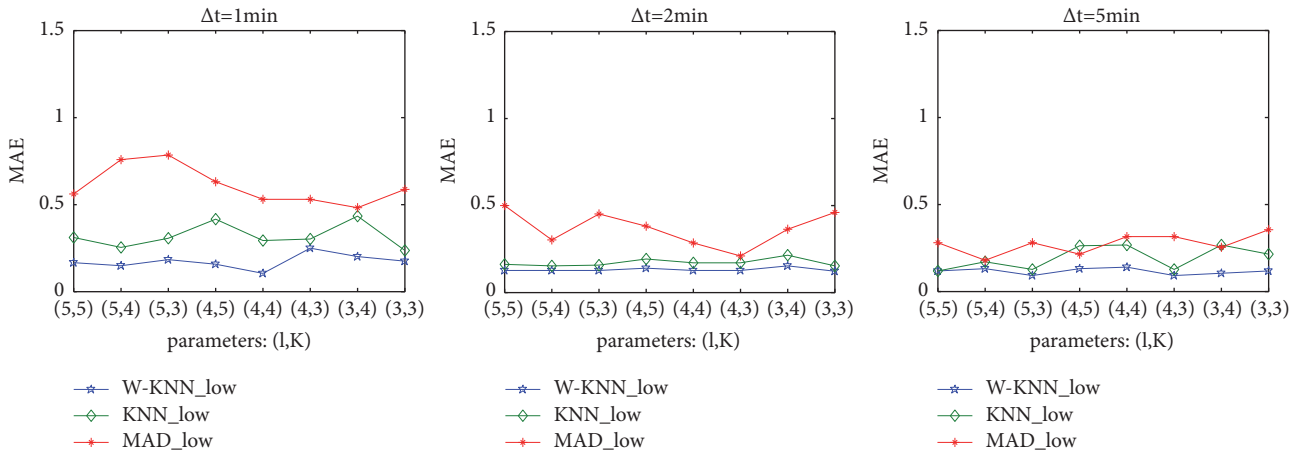
$$\widehat{V_0}(t + \Delta t) = \sum w_i V_i \quad i = 1, 2, \ldots, K \tag{17}$$

$$w_i = \frac{d_i^{-1}}{\sum d_i^{-1}} \quad \text{s.t.} \sum_i w_i = 1 \tag{18}$$

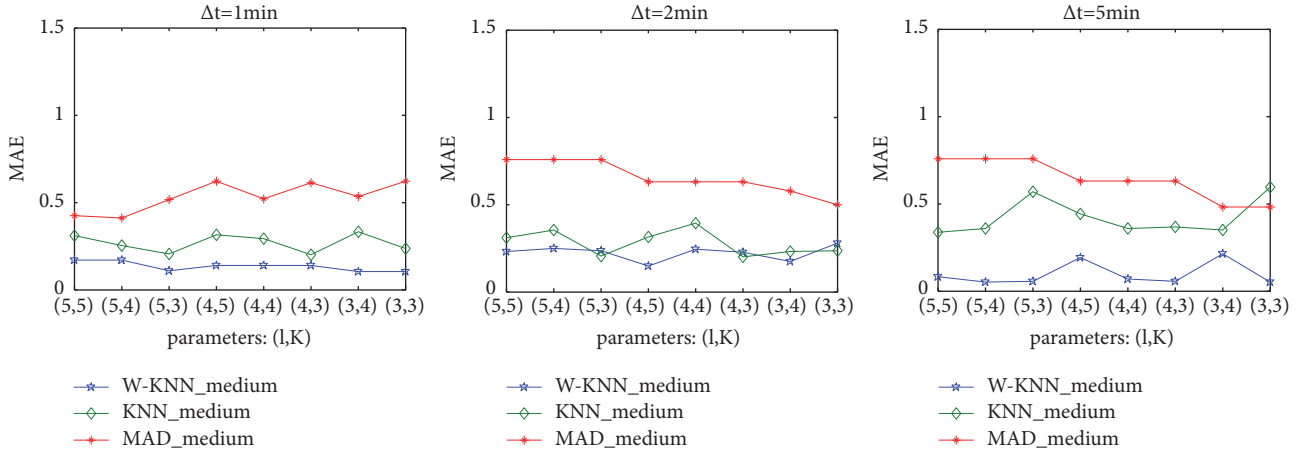
$$MAE = \frac{1}{n} \sum \left| \widehat{V_i} - V_i \right| \quad i = 1, 2, \ldots, n \tag{19}$$

To evaluate the prediction performances of the short-term traffic speed prediction models, we compare the moving average data-based model, the pure KNN model, and the weighted-KNN model utilizing the traffic data from SUMO microscopic traffic simulator. We simulate different traffic conditions by changing the number of vehicles: 150 vehicles for sparse status in Figure 8(a), 450 vehicles for normal status in Figure 8(b), and 750 vehicles for congested status in Figure 8(c). To optimize the prediction performance for each model, a series of contrastive experiments with different combinations of influential parameters are built in different traffic conditions. Results in Figure 8 show that our weighted-KNN model with spatial-temporal parameters exhibits better performance compared with the pure KNN model and the MAD model in terms of prediction accuracy. And for different time interval Δt, the MAD model acquires the worst MAE no matter the vehicle density is low, medium, or high. The results of the W-KNN model are better than or equal to that of the KNN model under all the various experimental conditions listed. Based on the experimental results, we adopted the W-KNN model to predict the short-term vehicle speed in this paper.

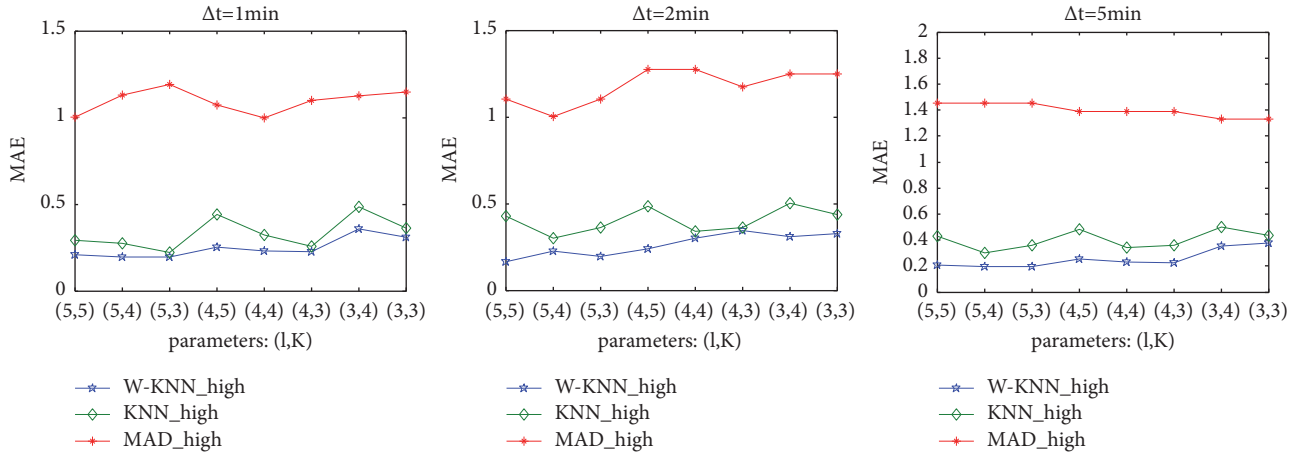
4.3. Machine Learning System. In our proposed mechanism, machine learning system is indispensable when the destination is in the blind zone. The main duty of our specially designed machine learning system embedded in every RSU is



(a) Sparse status



(b) Normal status



(c) Congested status

Figure 8: Results of three prediction models.

to process real-time traffic information and provide routing decisions dynamically.

In order to connect to a certain vehicle, GPS is leveraged in most routing protocols to locate that node. However, as described in the above, excessive dependence on GPS is unreliable. Due to the unavoidable defects caused by GPS, researchers are exploring new localization methods avoiding or decreasing the usage of GPS. In [39], authors designed a novel grid-based on-road localization system (GOT), where vehicles with and without accurate GPS signals self-organize

into a VANET. Vehicles in this small size VANET exchange location and distance information and help each other to calculate an accurate position for all the vehicles inside the network. This paper develops the fuzzy geometric relationship among vehicles, and utilizes a novel grid-based mechanism to evaluate the geometric relationships and calculate vehicle locations. Although a light-load grid-based calculation mechanism, which incurs only linear error propagation, is proposed in this paper, there is still a part of vehicles acquiring location information relying on GPS. A GPS-free localization framework that uses two-way time of arrival with partial use of dead reckoning to locate the vehicles based on communication with a single RSU is proposed in [20]. This proposed localization framework consists of two phases, determining the driving direction and computing the vehicle location in the Y-dimension. Compared to existing localization schemes which use multiple RSUs for vehicle localization, this paper decreased the required number of RSUs getting a higher accuracy compared to existing single RSU techniques in the same time. The weakness of this framework is that RSU must be installed at entry or exit. To eliminate those potential troubles caused by GPS and ensuring system scalability, our KNN-based machine learning system serves as a GPS-free dynamic vehicle location prediction system.

Machine learning techniques can learn from training data set automatically to identify rules, and we can use these rules to predict results for testing data. Among all the applications of machine learning methods, classification is one of the most major branches. Classification algorithms are assigned the responsibility of learning an objective function f (20) that maps each set of attributes $\{A_1, A_2, \ldots, A_m\}$ to one of the predefined categories or classes $\{C_1, C_2, \ldots, C_n\}$ [40].

$$\{A_1, A_2, \ldots, A_m\} \xrightarrow{f} C_i \quad 1 \le i \le n \tag{20}$$

K-Nearest Neighbor, a simple yet effective classification algorithm, which has been adopted in numerous regression and classification problems, is applied in our location method. As an instance based learning method, KNN classifies each testing data based on a certain amount of instances, and it is based on the principle that the instance within a data set will generally exist in close proximity to other instances that have similar properties. For each testing data, KNN identifies the K-nearest training data from the training data set and stores them in the N_k set. The class of the testing data is same with the majority vote of the N_k set. The only parameter in this algorithm is the number of K neighbors, which can be customized according to concrete applications [41].

In general, machine learning system needs a training process in advance such that it can generalize new instance better. During the training process, interested information will be collected to make up the training data set, and then the training set will serve as input to train the machine learning system after necessary data pre-processing. KNN's high efficiency benefits from its lazy learning characteristic; i.e., we do not need to fix any generic model in advance, so the training phase will be shorter compared with other

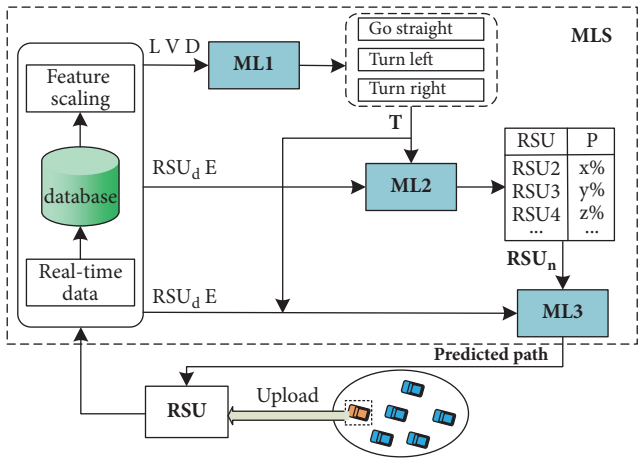


Figure 9: Machine learning system.

machine learning algorithms. During the training phase, data collection is done by RSUs and vehicles travelling between these RSUs in coordination with one another. If a vehicle that just left the coverage of RSU_a is entering the coverage of RSU_b, it will transmit its traffic information to RSU_b. According to the received information, RSU_b can inform the previous RSU that vehicle has just travelled through and transmit that information back to RSU_a. The traffic data collected by RSUs compose our training data set and will be retrieved by our machine learning system in the testing phase. Driving features monitored to train the machine learning system include the lane number (L), the vehicle velocity (V), the driving direction (D), the exit of RSU_a (E), the turning direction at the first intersection after leaving RSU_a (T), and the travelling path from the exit E of RSU_a to RSU_b.

After the training process is completed, when a new sample data arrives, machine learning system will make predictions and meanwhile store it as a training data to update the training data set. The designed system can process dynamic traffic information so as to locate the destination vehicle roughly without GPS. Although the system cannot provide accurate position of a vehicle, it can determine which road the predicted vehicle is travelling on precisely, which can lend enough support to making routing decisions for data delivery and eliminate those potential troubles caused by GPS at the same time. For description convenience, in the predicting phase we illustrate an example where the target vehicle has just left RSU_1 from exit 3 and is travelling in the blind zone now. As mentioned, when the destination vehicle was about to leave RSU_1, it has uploaded its real-time traffic information to RSU_1. After receiving TREQ from another RSU, RSU_1 will send the latest data of the destination node to its embedded machine learning system database to make predictions. As shown in Figure 9, our machine learning system is composed of three parts.

4.3.1. Predicting Destination's Turning Direction. The first problem that needs to be solved is predicting vehicle's turning direction at intersection, and this will determine the general orientation for our location. In machine learning 1, we select

KNN algorithm to predict which direction the destination vehicle will turn to at next intersection after leaving RSU_d through one of the six exits. In the dynamic prediction process, the target vehicle will first upload its real-time traffic information to RSU_1. After feature selection, target vehicle's turning direction related training data stored in the database and real-time data just uploaded will be sent into machine learning 1 in RSU_1 together. When approaching an intersection, many elements will influence a vehicle's turning direction more or less. Among all the influencing factors, we select three variables that play leading roles, the lane number (L), the vehicle velocity (V), and the driving direction (D), to predict which category the output should be assigned to. The output data will be one of the three classes: go straight, turn left, and turn right. The nearest neighbors K are set to be 10 which bring the best forecast accuracy according to our simulation results.

4.3.2. Predicting the Probability into Each RSU. Next, our system will determine every potential RSU that destination vehicle might travel to and predict the probability of each potential RSU based on the training data set. In machine learning 2, the predicting outcome from machine learning 1 will be merged with other necessary data uploaded to RSU_1 to predict which RSU coverage area the destination vehicle will move into. The input data will be the previous RSU (RSU_d), the exit of RSU_1 (E), and the predicted turning direction (T). The output will be a table consisting all the possible RSUs as well as their relevant possibilities. Our proposed mechanism will first select the one with the highest possibility to relay packets. If the destination node does not receive packets after a threshold time, RSU with the second highest possibility will be selected and so on. Suppose the predicted turning direction from machine learning 1 is going straight; the most likely RSU that destination vehicle will visit next is RSU_2 as we can learn from Figure 1. Not all vehicles travelling on roads observe the traffic regulations especially in an emergency, so there may be special little results which seem counterintuitive. Vehicles turning around halfway can explain the existence of other RSUs in the table.

4.3.3. Predicting the Travelling Path. Finally, the machine learning system in RSU_1 will predict the travelling path of the destination by locating which road the target vehicle is travelling on. On the basis of the two results obtained above, machine learning 3 will perform this step. As mentioned above, vehicles' running traces are collected into training data set in the training process; therefore, when supplied related information, the database will provide matching paths. Machine learning 3 will take input the exit of RSU_d ($E = 3$), the previous RSU ($RSU_d = RSU_1$), and the predicted turning direction ($T = go\ straight$) and the predicted next RSU ($RSU_n = RSU_2$) to generate output traces from the exit 3 of RSU_1 to RSU_2. According to the predicted trace, our system can analyze the destination vehicle's location roughly without GPS. After outputting the final outcome to RSU_d, our KNN-based machine learning system has completed a systematic prediction task. Following, the two-way mode transfer will be employed to deliver packets to the destination. Our machine learning system can work as a GPS-free dynamic vehicle location prediction method to acquire the location of vehicle so as to eliminate those potential troubles caused by GPS.

Table 5: Simulation parameters.

Parameters	Value
Simulated area	3 km ∗ 2.5 km
Number of vehicles	[150, 300, 450, 600, 750]
Maximum speed of vehicles	80km/h
RSU communication radius	500 m
Vehicle communication radius	500 m
No. lanes of each direction	2
RSU coverage ratio	41%
Packet size	512 bytes
Vehicle beacon interval	1s

5. Simulation and Evaluation

In this section, we present and discuss the performance of the proposed system through network simulations. In order to evaluate the proposed scheme, we compare it with STAR (Shortest-Path-Based Traffic-Light-Aware Routing) [27] and modified STAR. During the data delivery process, STAR adopts the most common V2V technology; therefore, in terms of wireless transmission, this scheme is representative. When reaching an intersection, STAR attempts to forward packets to a connected red light road segment instead of forwarding packets to the green light road segment. To validate the effectiveness of wired transmission between RSUs, RSUs are involved to deliver packets in the modified STAR.

5.1. Simulation Environment. To simulate the mobility of vehicles and vehicular network, Simulation of Urban Mobility (SUMO) is used for simulating vehicles' mobility traces and road topology, and Network Simulator (NS, version 3.0) is used for simulating vehicular networks. There are six RSUs in our layout as shown in Figure 1, and each RSU is equipped with a dedicated machine learning system. In each simulation experiment, we determine 10 source nodes and 10 destination nodes randomly. Each scheme is tested in the different vehicle densities to analyze their performances when running in various road conditions. Each result in all scenarios is the average of 10 runs. The detailed simulation parameters are shown in Table 5.

5.2. Results and Analysis. To evaluate the performance of these data delivery strategies, three performance metrics are employed: (1) packet delivery ratio, the ratio of the number of the packets successfully received by destination nodes to the total number of the packets sent by source vehicles, (2) network delay, the average latency of the data packets that travel from their source vehicles to the destination vehicles, and (3) control overhead, the number of extra packets generated in the delivery process per minute. In addition, a

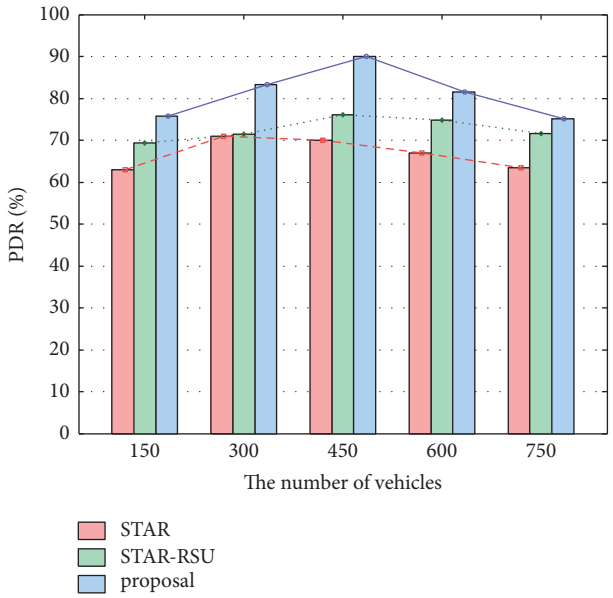


Figure 10: Comparison of packet delivery ratio.

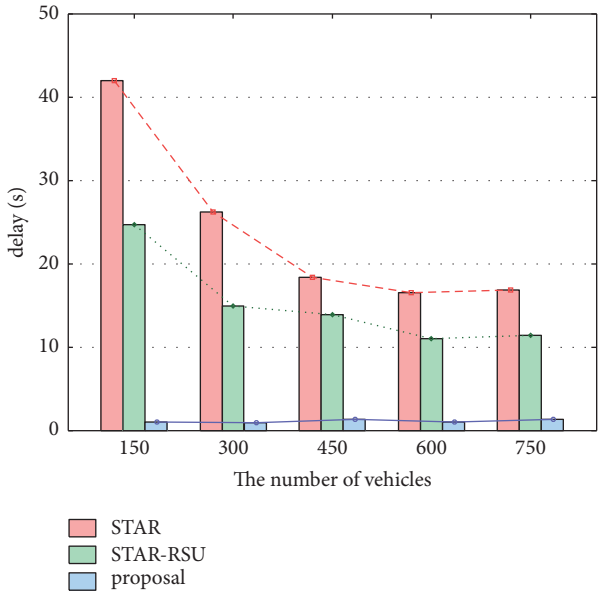


Figure 11: Comparison of network delay.

set of simulations were conducted to evaluate the impact of vehicle density.

The variation of packet delivery ratio with different vehicle densities is illustrated in Figure 10. These results illustrate that, with the increase of vehicle density, the packet delivery ratios rise firstly and then decrease in all the three methods. This is because relay vehicles may not be available to establish communication links when the vehicle density is too low, and channel collisions or traffic jams may occur when the vehicle density is too high. The pure STAR suffers the lowest delivery ratio since V2V is the main communication mode. The modified STAR reduces the packet loss delivery by replacing a part of wireless transmission with RSUs, which proves the advantage of utilizing backbone network for vehicular network. Our scheme outperforms the other two strategies throughout the whole running process because our fuzzy-rule-based wireless transmission method optimize V2V communication and the specially designed machine learning system can process dynamic traffic information effectively in different vehicle densities. In our scheme, the packet delivery ratio can be as high as 90% when traffic condition is good (450 vehicles) and can still reach more than 75% even in the worst case (150 vehicles and 750 vehicles).

Figure 11 shows the results of packet delivery delay with three compared schemes in different vehicle densities. The figure shows that, by advancing the number of vehicles from 150 to 750, the average delay for STAR and modified STAR decreases sharply firstly and then levels off. This is because the lower the vehicle density is, the more possible carry-and-forward is adopted, which leads to higher delay. As we compare horizontally, the network delay of our proposal outperforms the other two methods observably and holds steady without obvious fluctuations. The delay in our scheme remains fluctuating around 1 second in the whole process instead of soaring to dozens of seconds as in other two

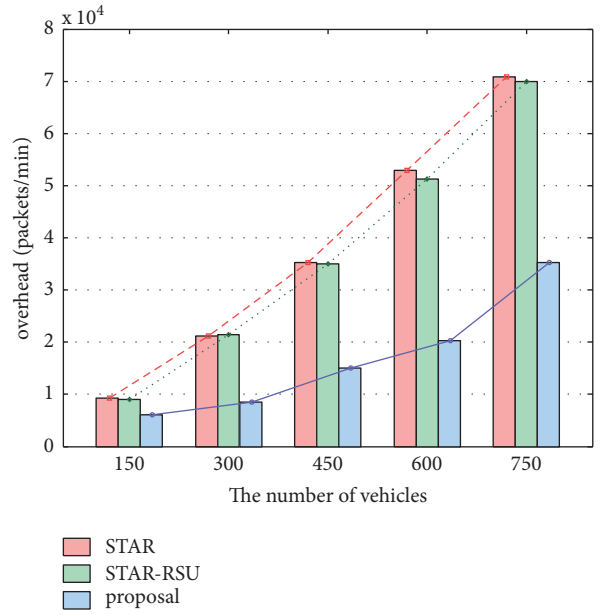


Figure 12: Comparison of control overheads.

schemes. This is because the optimized wireless communications and the backbone networks support the fast and efficient transmission of packets. With the participation of RSU, the result of modified STAR is much better than the pure STAR. When running in different vehicle densities, the average delay of STAR and modified STAR varied intensely, instead of keeping stable in our scheme, which means that our proposed method can adapt to different traffic conditions.

Figure 12 compares the control overhead among different schemes to evaluate the costs. As the figure shows, with the increase of the vehicle density, the overheads rise for all three strategies. However, the results of the proposed

strategy are much better than the other two methods in the whole simulation experiment. When the number of vehicles is greater than or equal to 300, overheads in STAR and modified STAR are twice or more that in our scenario. In STAR, both delivering packets in V2V communication mode and sending messages between intersections to check the connectivity will increase the cost significantly. Compared with STAR, the modified STAR decreases overhead slightly with the assistance of RSU in the delivery process. In contrast, our proposal causes the least overhead, because the teamwork of machine learning system and RSU will analyze traffic data timely to determine the delivery path instead of sending a great deal of communication signals between vehicles. And the vehicle-based short-term speed prediction method makes the relay node selection more reliable so as to reduce overhead caused by building communication links continually.

In summary, our proposed data delivery scheme can improve the packet delivery ratio, guarantee the timeliness of messages, and reduce the control overhead significantly. Consequently, our proposed scheme is suitable for data delivery in urban scenario.

6. Conclusion

In this paper, we propose a novel data delivery scheme for vehicular networks in urban environments, and we focus on the analysis that both the source node and the destination node are in the blind zone.

In order to set up delivery paths for vehicles in the blind zone, we designed a fuzzy-rule-based wireless transmission method. This key technology will select an optimal option from all the possible paths with considering multiple factors comprehensively. By optimizing V2V communications, the DTQ can be improved. One of the key technologies in this fuzzy-rule-based approach is the vehicle speed prediction approach. Different from common segment-based prediction method, we designed a vehicle-based short-term vehicle speed prediction method. Taking full consideration of velocity by comparing predicted speeds at next moment will make the relay node selection more reliable and increase the stability of selected transmission link. Another key technology in our data delivery scheme is the specially designed machine learning system embedded in each RSU, which provides routing decisions by processing dynamic traffic information delivered to it. The combination of machine learning system and RSU empowers our system to abandon GPS without degrading the network performance. The wired communication between RSUs can reduce the delay resulted from the unreliable carry-and-forward manner in the pure V2V communication network.

The performance of our proposal has been verified through simulations in NS-3. For future work, we intend to apply more machine learning methods to the study of VANETs.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Henan International Science & Technology Cooperation Program (182102410050), Henan Young Scholar Promotion Program (2016GGJS-018), the Program for Science & Technology Development of Henan Province (162102210022), Key Project of Science and Technology Research of the Education Department of Henan Province (17A413001), and CERNET Innovation Project (NGII20151005).

References

[1] W. Quan, Y. Liu, H. Zhang, and S. Yu, "Enhancing crowd collaborations for software defined vehicular networks," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 80–86, 2017.

[2] N. Cheng, N. Zhang, N. Lu, X. Shen, J. W. Mark, and F. Liu, "Opportunistic spectrum access for CR-VANETs: A game-theoretic approach," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, pp. 237–251, 2014.

[3] N. Cheng, F. Lyu, and J. Chen, *Big Data Driven Vehicular Networks*, IEEE Network, 2018.

[4] C. Silva, M. Nogueira, D. Kim, E. Cerqueira, and A. Santos, "Cognitive radio based connectivity management for resilient end-to-end communications in VANETs," *Computer Communications*, vol. 79, pp. 1–8, 2016.

[5] J.-S. Li, I.-H. Liu, C.-K. Kao, and C.-M. Tseng, "Intelligent Adjustment Forwarding: A compromise between end-to-end and hop-by-hop transmissions in VANET environments," *Journal of Systems Architecture*, vol. 59, no. 10, pp. 1319–1333, 2013.

[6] I. Tal and G. Muntean, "Towards Reasoning Vehicles," *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–37, 2018.

[7] N. Cheng, N. Lu, N. Zhang, X. S. Shen, and J. W. Mark, "Vehicular WiFi offloading: Challenges and solutions," *Vehicular Communications*, vol. 1, no. 1, pp. 13–21, 2014.

[8] X. Xiang, X. Wang, and Z. Zhou, "Self-Adaptive On-Demand Geographic Routing for Mobile Ad Hoc Networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 9, pp. 1572–1586, 2012.

[9] W. K. Lai, M. Lin, and Y. Yang, "A Machine Learning System for Routing Decision-Making in Urban Vehicular Ad Hoc Networks," *International Journal of Distributed Sensor Networks*, vol. 11, no. 3, article 374391, pp. 1–13, 2015.

[10] S. Agrawal, R. S. Raw, N. Tyagi, and A. K. Misra, "Fuzzy logic based greedy routing (flgr) in multi-hop vehicular adhoc networks," *Indian Journal of Science and Technology*, vol. 8, no. 30, pp. 1–14, 2015.

[11] A. Jahangiri and H. A. Rakha, "Applying Machine Learning Techniques to Transportation Mode Recognition Using Mobile Phone Sensor Data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2406–2417, 2015.

[12] H. Zhou, N. Cheng, Q. Yu, X. S. Shen, D. Shan, and F. Bai, "Toward multi-radio vehicular data piping for dynamic DSRC/TVWS spectrum sharing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 10, pp. 2575–2588, 2016.

[13] Z. Su, Y. Hui, and Q. Yang, "The next generation vehicular networks: a content-centric framework," *IEEE Wireless Communications Magazine*, vol. 24, no. 1, pp. 60–66, 2017.

[14] O. K. Tonguz, N. Wisitpongphan, J. S. Parikh, F. Bai, P. Mudalige, and V. K. Sadekar, "On the Broadcast Storm Problem in Ad hoc Wireless Networks," in *Proceedings of the 2006 3rd International*

Conference on Broadband Communications, Networks and Systems, pp. 1–11, San Jose, CA, USA, October 2006.

[15] D. Ou, Y. Yang, L. Xue, T. Shen, and L. Zhang, "The analytic hierarchy process-based optimal forwarder selection in multi-hop broadcasting scheme for vehicular safety," in *Proceedings of the 2014 IEEE Intelligent Vehicles Symposium (IV)*, pp. 999–1004, MI, USA, June 2014.

[16] A. T. Reza, T. A. Kumar, and T. Sivakumar, "Position Prediction based Multicast Routing (PPMR) using Kalman filter over VANET," in *Proceedings of the 2nd IEEE International Conference on Engineering and Technology, ICETECH 2016*, pp. 198–206, India, March 2016.

[17] A. M. Vegni and E. Natalizio, "Forwarder smart selection protocol for limitation of broadcast storm problem," *Journal of Network and Computer Applications*, vol. 47, pp. 61–71, 2015.

[18] L. A. Zadeh, "Fuzzy sets," *Information and Computation*, vol. 8, pp. 338–353, 1965.

[19] P. Lopez-Garcia, E. Onieva, E. Osaba, A. D. Masegosa, and A. Perallos, "A hybrid method for short-term traffic congestion forecasting using genetic algorithms and cross entropy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 557–569, 2016.

[20] A. A. Wahab, A. Khattab, and Y. A. Fahmy, "Two-way TOA with limited dead reckoning for GPS-free vehicle localization using single RSU," in *Proceedings of the 13th International Conference on ITS Telecommunications (ITST '13)*, vol. 12(1), pp. 244–249, Tampere, Finland, November 2013.

[21] B. Cheng, G. Liao, C. Chen, and H. Chen, "A Multi-level Fuzzy Comprehensive Evaluation Approach for Message Verification in VANETs," in *Proceedings of the 2012 Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing (MUSIC)*, pp. 176–181, Vancouver, Canada, June 2012.

[22] A. Afful-Dadzie, E. Afful-Dadzie, S. Nabareseh, and Z. K. Oplatková, "Tracking progress of African Peer Review Mechanism (APRM) using fuzzy comprehensive evaluation method," *Kybernetes*, vol. 43, no. 8, pp. 1193–1208, 2014.

[23] T. L. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, New York, NY, USA, 1980.

[24] K. Katsaros, M. Dianati, Z. Sun, and R. Tafazolli, "An evaluation of routing in vehicular networks using analytic hierarchy process," *Wireless Communications and Mobile Computing*, vol. 16, no. 8, pp. 895–911, 2016.

[25] M. Patra, S. Mishra, and C. S. Murthy, "An Analytic Hierarchy Process Based Approach for Optimal Road Side Unit Placement in Vehicular Ad Hoc Networks," in *Proceedings of the 2014 IEEE Vehicular Technology Conference (VTC 2014-Spring)*, pp. 1–5, Seoul, South Korea, May 2014.

[26] N. V. Dharani Kumari and B. S. Shylaja, "AMGRP: AHP-based Multimetric Geographical Routing Protocol for Urban environment of VANETs," *Journal of King Saud University - Computer and Information Sciences*, 2017.

[27] J.-J. Chang, Y.-H. Li, W. Liao, and I.-C. Chang, "Intersection-based routing for urban vehicular communications with traffic-light considerations," *IEEE Wireless Communications Magazine*, vol. 19, no. 1, pp. 82–88, 2012.

[28] B. Yao, C. Chen, Q. Cao et al., "Short-term traffic speed prediction for an urban corridor," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 2, pp. 154–169, 2016.

[29] H. Chang, Y. Lee, B. Yoon, and S. Baek, "Dynamic near-term traffic flow prediction: System-oriented approach based on past experiences," *IET Intelligent Transport Systems*, vol. 6, no. 3, pp. 292–305, 2012.

[30] S. Clark, "Traffic prediction using multivariate nonparametric regression," *Journal of Transportation Engineering*, vol. 129, no. 2, pp. 161–168, 2003.

[31] B. L. Smith and M. J. Demetsky, "Multiple-interval freeway traffic flow forecasting," *Transportation Research Record*, no. 1554, pp. 136–141, 1996.

[32] H. Chang, D. Park, S. Lee, H. Lee, and S. Baek, "Dynamic multi-interval bus travel time prediction using bus transit data," *Transportmetrica*, vol. 6, no. 1, pp. 19–38, 2010.

[33] P. Kang and S. Cho, "Locally linear reconstruction for instance-based learning," *Pattern Recognition*, vol. 41, no. 11, pp. 3507–3518, 2008.

[34] D. P. Solomatine, M. Maskey, and D. L. Shrestha, "Instance-based learning compared to other data-driven methods in hydrological forecasting," *Hydrological Processes*, vol. 22, no. 2, pp. 275–287, 2008.

[35] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning," *Artificial Intelligence Review*, vol. 11, no. 1–5, pp. 11–73, 1997.

[36] D. W. Aha and R. L. Goldstone, "Concept learning and flexible weighting," in *Proceedings of 14th Annual Conference of the Cognitive Science Society*, pp. 534–539, Mahwah, USA, 1992.

[37] M. P. Wand and W. R. Schucany, "Gaussian-based kernels," *Canadian Journal of Statistics*, vol. 18, no. 3, pp. 197–204, 1990.

[38] V. V. Fedorov, P. Hackl, and W. G. Müller, "Moving local regression: the weight function," *Journal of Nonparametric Statistics*, vol. 2, no. 4, pp. 355–368, 1993.

[39] T. Yan, W. Zhang, and G. Wang, "A grid-based on-road localization system in VANET with linear error propagation," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 861–875, 2014.

[40] L. P. Portugal-Poma, C. A. Marcondes, H. Senger, and L. Arantes, "Applying Machine Learning to Reduce Overhead in DTN Vehicular Networks," in *Proceedings of the 2014 Brazilian Symposium on Computer Networks and Distributed Systems (SBRC)*, pp. 94–102, Florianopolis, Brazil, May 2014.

[41] S. B. Kotsiantis, "Supervised machine learning: a review of classification techniques," in *Proceedings of the Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word Ai Systems with Applications in Ehealth, Hci, Information Retrieval and Pervasive Technologies IOS Press*, pp. 3–24, 2007.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 3810350, 9 pages
https://doi.org/10.1155/2018/3810350

Research Article

Software-Defined Collaborative Offloading for Heterogeneous Vehicular Networks

Wei Quan,[1,2] Kai Wang,[3] Yana Liu,[1] Nan Cheng,[2] Hongke Zhang,[1] and Xuemin (Sherman) Shen[2]

[1]*School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China*
[2]*Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada*
[3]*School of Computer and Control Engineering, Yantai University, Shandong, China*

Correspondence should be addressed to Kai Wang; wangkai.phd@outlook.com

Received 28 December 2017; Accepted 20 March 2018; Published 24 April 2018

Academic Editor: Tao Han

Vehicle-assisted data offloading is envisioned to significantly alleviate the problem of explosive growth of mobile data traffic. However, due to the high mobility of vehicles and the frequent disruption of communication links, it is very challenging to efficiently optimize collaborative offloading from a group of vehicles. In this paper, we leverage the concept of Software-Defined Networking (SDN) and propose a software-defined collaborative offloading (SDCO) solution for heterogeneous vehicular networks. In particular, SDCO can efficiently manage the offloading nodes and paths based on a centralized offloading controller. The offloading controller is equipped with two specific functions: the hybrid awareness path collaboration (HPC) and the graph-based source collaboration (GSC). HPC is in charge of selecting the suitable paths based on the round-trip time, packet loss rate, and path bandwidth, while GSC optimizes the offloading nodes according to the minimum vertex cover for effective offloading. Simulation results are provided to demonstrate that SDCO can achieve better offloading efficiency compared to the state-of-the-art solutions.

1. Introduction

With the significant advance of automobile technologies, the Internet of vehicles (IoV) is becoming a standard feature in the near future, where vehicles are equipped with advanced communication devices, and exchange data and information through vehicle-to-everything (V2X) communications [1, 2]. Enabled by IoV technologies, a myriad of emerging vehicular applications occur to enrich the transport ecosystem, such as real-time high-definition navigation, multimedia services on wheels, and self-driving system [3, 4]. However, these applications and services require an exponential escalation of mobile data, which should be efficiently supported by the vehicular networks.

Many wireless communications technologies are developed to support the potential vehicular scenarios. For instance, the Dedicated Short-Range Communication (DSRC) enables efficient real-time information exchange among vehicles [5]. Wireless Access for Vehicular Environment (WAVE) has been applied to the Vehicle-to-Infrastructure (V2I) communications [6]. The Long-Term Evolution (LTE) has the ability to support vehicular communications in a wide range [7, 8]. Not surprisingly, cellular network technologies play a vital role in providing ubiquitous and reliable Internet access for mobile vehicles. However, the LTE network nowadays is straining to meet the current mobile data demand. With the explosive growth of vehicular data, LTE networks are easily overloaded and degrade the experience of both nonvehicular and vehicular users.

Therefore, simply using the single type of access technologies is far from enough to satisfy all the requirements of vehicular services. Researchers have considered solutions to offload traffic from the core network to other wireless networks [9, 10]. This solution can effectively reduce the congestion of transmission paths in the core network. Figure 1 describes three typical offloading scenarios in a heterogeneous vehicular network, where the traffic is offloaded from macro base station via small cell base stations, ad hoc networks

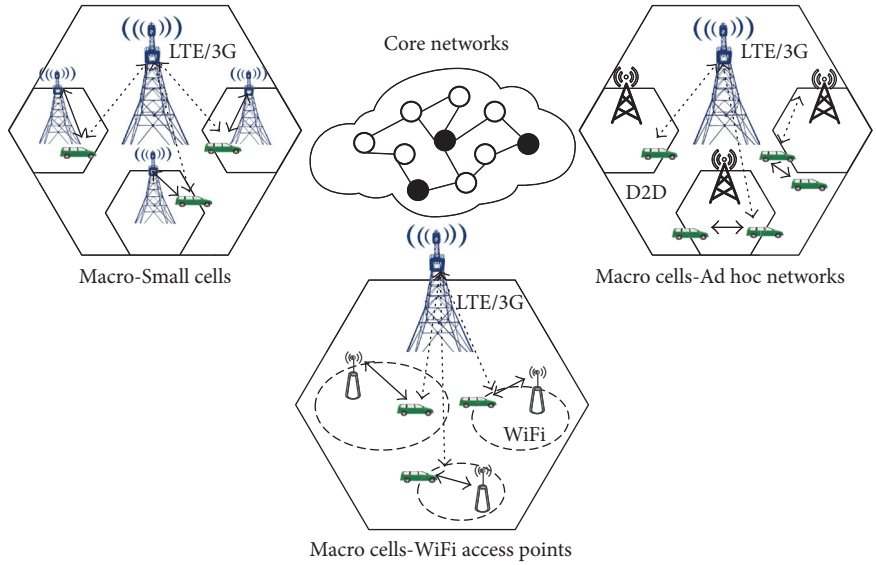


Figure 1: Offloading scenarios in heterogeneous vehicular networks.

(no base station), and WiFi access points, respectively. For example, WiFi offloading can alleviate the congestion in cellular networks by delivering data through WiFi access instead of the cellular base station [11]. Statistics show that more than 60 percent of the cellular traffic can be offloaded by WiFi networks.

Recently, the fifth generation (5G) mobile communications system is also designed to integrate networking, computing, and storage resources into one programmable and unified infrastructure. It is not only an evolution of mobile broadband networks but also a collaborative ecosystem of heterogeneous communications technologies. Therefore, the data offloading via various communications technologies has been considered in the 5G deployment. With this feature, 5G has a great potential of providing massive access, high reliability, global coverage, and very low latency for challenging situations like vehicular networks.

However, efficiently offloading the vehicular data is a challenging issue. In the vehicular environment, data offloading is opportunistic due to the limited coverage of heterogeneous wireless technologies and high mobility of vehicles [12]. The connection time between a vehicle and an offloading point (e.g., WiFi access points and small cell base stations) might be very short; therefore the content sources and transmission paths will be highly dynamic. Therefore, with such dynamics, the quality of service (QoS) requirements are difficult to guarantee.

Software-defined networking (SDN) provides a promising and referable approach due to the features of centralized controlling, global decision, and high execution power. Our previous work has investigated a new networking paradigm named Smart Identifier NETworking (SINET) and proposed a SINET customized solution for software-defined vehicular networks (SINET-V) [13, 14]. Based on such a SINET-V architecture, in this paper, we propose a software-defined collaborative offloading (SDCO) solution, which employs a centralized software-defined controller to globally distribute offloading policies to vehicular nodes in heterogeneous vehicular networks. Specifically, the hybrid awareness path collaboration (HPC) and the graph-based source collaboration (GSC) are considered in the offloading controller. HPC is in charge of selecting the suitable paths based on the round-trip time, packet loss rate, and path bandwidth, while GSC optimizes the offloading nodes according to the minimum vertex cover for effective offloading. These specific functions are detailed in the following sections.

The remainder of this paper is organized as follows. In Section 2, we analyze the related works on offloading management. In Section 3, we describe the proposed SDCO solution and the optimization model. In Sections 4 and 5, we detail the specific functions for collaborative offloading: hybrid-awareness path collaboration and graph-based source collaboration. Simulation results validate the performance of SDCO in Section 6. Finally, we conclude the paper and discuss the future works in Section 7.

2. Related Works

Currently, many research efforts have begun towards realizing efficient heterogeneous vehicular communications in a cooperative manner. In our previous work, a collaborative vehicular edge computing framework named as CVEC is introduced to support more scalable vehicular services and applications [15]. Chen et al. presented a novel architecture of software-defined Internet of vehicles (SD-IoV) and developed a centralized vehicular connection management approach to guarantee the QoS [16]. Zheng et al. presented a multilayer and soft-defined HetVNET with a far more flexible configuration capability [17]. Dong et al. analyzed the issues of energy-efficient cluster management in heterogeneous vehicular networks [18]. Zhang et al. proposed a software-defined space-air-ground integrated network architecture

for supporting diverse vehicular services [19]. However, the above newest research mainly concentrated on the general architecture designs for heterogeneous vehicular networks yet failed to analyze the specific offloading mechanisms.

In the aspect of offloading management, Liu et al. proposed an incentive mechanism for computation offloading by using edge computing to deliver computation to the edge of pervasive networks nearby mobile users [20]. Cheng et al. discussed data offloading technologies for big data driven vehicular network [21]. Aijaz et al. made a survey on mobile data offloading from both technical and business views [22]. Zhuo et al. proposed an incentive framework to motivate users to offload cellular traffic with minimal cost [23]. To minimize the incentive cost, sources with large offloading potential and delay tolerance are preferred in their solution. Lee et al. focused on a quantitative study on the performance of mobile data offloading through WiFi access [24]. Singh et al. built a general and tractable model to analyze data offloading in heterogeneous networks [25]. Besides, Cheng et al. used a queueing model to analyze the opportunistic WiFi offloading in vehicular networks [26]. Rebecchi et al. made a complete survey on the data offloading techniques in cellular networks, including classification of mobile data offloading, nondelayed offloading, delayed offloading, assessing mobile data offloading, and open challenges [9]. However, there are few works on how to collaborate among different offloading schemes in heterogeneous network scenarios.

Thanks to these efforts, some progress has been made on understanding the collaborative offloading in vehicular networks. However, there are still several challenges waiting to be overcome due to the intractable characteristics in vehicular environment. This paper leverages the advantages of SDN and proposes a SDN-based collaborative offloading for heterogeneous vehicular networks, which can globally and efficiently distribute offloading policies to vehicular nodes.

3. System Modelling and Problem Formulation

3.1. System Modelling. The typical offloading is to use local communication to offload the opportunistic traffic of the base station. Neighboring peers can communicate with each other without using a network infrastructure and can retrieve the expected content from the peer. It can support to dynamically adjust the access mode for the user equipment and then offloads the data traffic of the macro base station to the small base station.

In SDCO, we employ a centralized software-defined controller to globally distribute offloading policies to vehicular nodes. This design aims to guarantee the transmission reliability by avoiding a large delay in distributed negotiation. It can also improve the overall network throughput and alleviate the problem of mobile data traffic congestion. Figure 2 shows the SDCO system architecture. Based on this architecture, we will focus on how the controller makes optimization for efficient offloading policies.

3.2. Problem Formulation. Typical offloading allows users to relay data through other users within the network coverage, indirectly accessing the base station for data resources. In addition, the offloading scheme allows UEs with certain QoS requirements, in particular, to guarantee UEs with high-speed mobility to access the macro base station so as to avoid frequent handovers. We transferred this problem as a 0-1 linear programming problem, by which to determine the best source and preferred access mode for each user, for example, to select the base station or relay nodes. The objective is to maximize the amount of access users by offloading some users to the small base station so that the system can serve more UEs. Based on the above considerations, the optimization model is defined as follows:

$$\max_{x,y,z} \quad N = \sum_{k=1}^{K} \left(\sum_{j=1}^{J} y_{jk} + \sum_{m=1}^{M} z_{mk} \right), \tag{1}$$

$$\sum_{j=1}^{J} x_{mj} + \sum_{k=1}^{K} z_{mk} \le C_m, \tag{2}$$

$$\sum_{m=1}^{M} z_{mk} + \sum_{j=1}^{J} y_{jk} \le 1, \tag{3}$$

$$\sum_{k=1}^{K} y_{jk} \le \sum_{m=1}^{M} x_{mj} \le 1, \tag{4}$$

$$x_{mj} \le \left(1 - \sum_{k=1}^{K} y_{jk} g_{mk} \right) e_{mj}, \quad \forall j \forall m > 0. \tag{5}$$

x_{mj} denotes the binary variable characterizing the connection status between base station (b_m) and relay node (r_j), 0 for without connection and 1 with connection. y_{jk} denotes the binary variable characterizing the connection status of user d_k and r_j, 0 for without connection and 1 for with connection. z_{mk} denotes the binary variable characterizing the connection state between d_k and b_m, 0 for without connection and 1 for with connection. In formula (1), $\sum_{k=1}^{K}(\sum_{m=0}^{M} z_{mk})$ indicates the number of users connected to the network through the relay node. Formulas (2)–(5) are constraints. Wherein, formula (2) indicates that the maximum number of users and relay nodes which the base station can connect with is C_m, that is, the capacity of the base station. Formulas (3) and (4) indicate that each user node can only be connected with one base station or relay node at most at any time.

If a user node needs to connect to a base station through a relay node, the user must be out of coverage of the base station, so it is limited by formula (5). g_{mk} indicates whether d_k is within the coverage of b_m, 1 means it is within its range, and 0 means it is not. e_{mj} indicates whether r_j is within the coverage of b_m, 1 means it is within its range, and 0 means it is not.

To resolve this optimization problem, a greedy search can be used to achieve the optimized solution for the offloading problem. However, it might be a time-consuming task to achieve a globally optimized solution, especially for a large-scale scenario. In practical operations, the controller can aware more information, like the dynamic topology, mobility trace, connection relationship, and so on. Therefore, in this

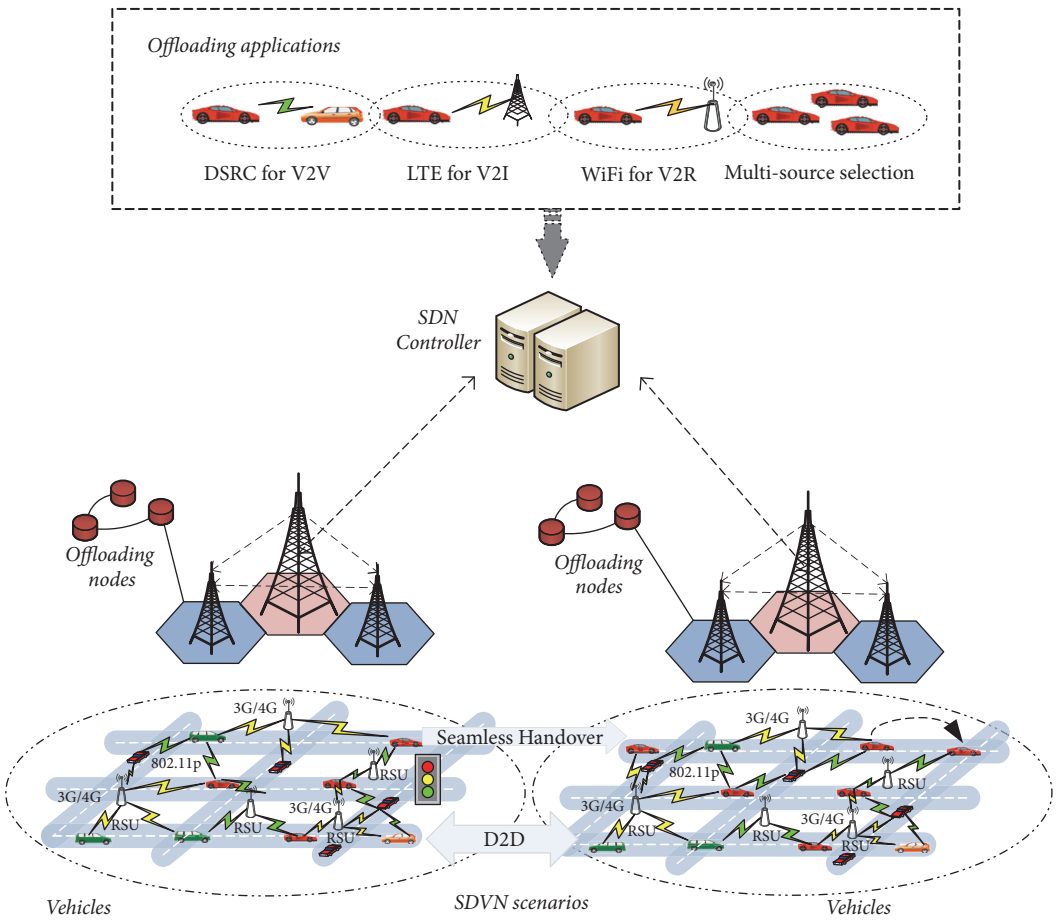


Figure 2: The SDCO system architecture.

work, we proposed two mechanisms to assist the controller to make offloading decisions. One is the hybrid awareness path collaboration (HPC) function, and the other one is the graph-based source collaboration (GSC) function. Both of them will be detailed in the following sections.

4. Hybrid Awareness Path Collaboration Function

The controller with global network information will promote the path collaboration more efficiently. SDCO adopts the HPC mechanism to address the path collaboration issue. Different from other existing solutions, the HPC mechanism is based on a new hybrid path metric, termed as Q. This path metric is able to synthetically depict the path status by considering three different parameters: Round-Trip Time (RTT), the packet loss rate (LOSS), and the path bandwidth (BWT). In the initial phase, the controller can obtain the real-time path status Q_i for the path i through sensing and notification from the nodes. Then, the controller will check whether the path status Q_i is bigger than a threshold τ to determine whether to add the path into the path scheduling list (PSL). Besides, the controller will monitor the variation of the status of each path in the PSL and then trigger an update to maintain the list. If the variation exceeds a threshold ω, this path will be updated in the PSL. Figure 3 illustrates the process of the HPC mechanism. In the following, we introduce the hybrid awareness path collaboration in detail.

4.1. Path Metric Calculation. In this section, we define a new path metric to synthetically depict the path quality. The metric of each path relies on three link parameters: RTT, LOSS, and BWT.

To be general, both RTT and LOSS need to be considered to evaluate link quality. Based on this, we make a further revision. We believe that there is a close relation between the path bandwidth and the path quality. The path bandwidth reflects the maximum throughput that the link can achieve within a unit time. Therefore, it is necessary to correctly reflect the actual situation of the paths by considering the above factors comprehensively.

$$\text{PATH} = \{P_1, P_2, \ldots, P_i, \ldots, P_n,\ n \geq 2\}. \tag{6}$$

In the following, we take PATH as a group of available paths between the vehicular users and the services providers, as shown in formula (6). We define Q_i as the path metric of P_i, as shown in formulas (7), (8), and (9). It is easily to see that the better the path quality is, the bigger the metric value Q_i is.

$$Q_i = \alpha \cdot \frac{1}{\text{RTT}_i} + \beta \cdot \text{BWT}_i + \theta \cdot \frac{1}{\text{LOSS}_i}, \tag{7}$$

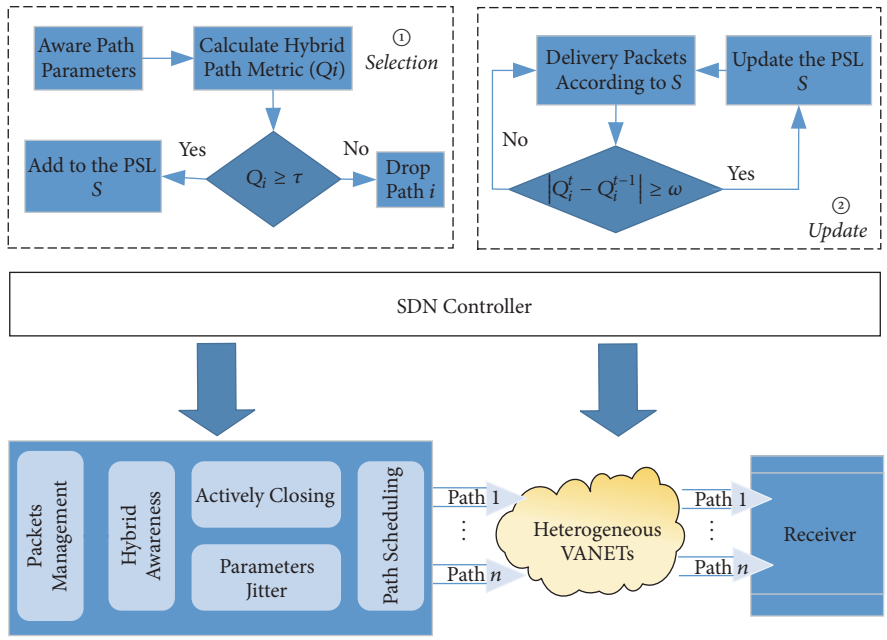


Figure 3: The process of the HPC mechanism.

$$\alpha + \beta + \theta = 1, \tag{8}$$

$$0 \leq \alpha, \beta, \theta \leq 1. \tag{9}$$

In the formula (7), RTT_i represents the RTT value of P_i, BWT_i is the path bandwidth of P_i, and LOSS_i is the packet loss rate of P_i. In addition, α, β, and θ are all positive constants less than 1 determined by the weight of each parameter. For example, if the RTT of each path affects the path quality more than the other two parameters, α will be accounted for the largest proportion compared with β and θ. The following gives a method to calculate the value of α, β, and θ in this paper.

$$\widehat{X}_t = \frac{X_t - X_{\min}}{X_{\max} - X_{\min}}, \quad X \in \{\mathrm{RTT}, \mathrm{BWT}, \mathrm{LOSS}\}, \tag{10}$$

$$\varepsilon_X = \frac{\sigma_{\widehat{X}}}{\mu_{\widehat{X}}}. \tag{11}$$

We introduce the definition of fluctuation intensity, which reflects the influence coefficient of a path parameter. If a path parameter has a big fluctuation intensity, it means this path parameter affects the path greatly. In this paper, we calculate the value of α, β, and θ by using the jitter intensity of each path parameter. Firstly, three parameters are normalized by min-max normalization as shown in formula (10). Then, the standard deviation factor ε is used to indicate the fluctuation intensity as shown in formula (11); σ represents the standard deviation of the parameter and μ is defined as the mean value. Finally, formula (12) shows the calculation method of α, β, and θ.

$$\alpha = \frac{\varepsilon_{\mathrm{RTT}}}{(\varepsilon_{\mathrm{RTT}} + \varepsilon_{\mathrm{BWT}} + \varepsilon_{\mathrm{LOSS}})},$$

$$\beta = \frac{\varepsilon_{\mathrm{BWT}}}{(\varepsilon_{\mathrm{RTT}} + \varepsilon_{\mathrm{BWT}} + \varepsilon_{\mathrm{LOSS}})},$$

$$\theta = \frac{\varepsilon_{\mathrm{LOSS}}}{(\varepsilon_{\mathrm{RTT}} + \varepsilon_{\mathrm{BWT}} + \varepsilon_{\mathrm{LOSS}})}. \tag{12}$$

With the above knowledge, the path metric for each path can be obtained by formula (7). Besides, the offloading controller will constantly monitor and update the path metric for each path. If there are packets needed to be sent, available paths will be scheduled according to the path metric. In the following, the detailed scheduling policy will be presented.

4.2. Path Collaboration Policy. The offloading controller will guide each vehicle user to select suitable paths according to the path management. The information includes the sequence of available paths and the path metric of these paths. The quality of path P_i is calculated based on formula (7). After that, available paths list S can be obtained. We take $S = \{S_1, S_2, \ldots, S_j, \ldots, S_m\}$ as a group of available paths and sort them by the path metric. The Q_i value of S_i is bigger than S_{i+1}. The better the path quality is, the bigger the path metric value is.

In the HPC mechanism, the parameter τ is proposed as the path selection threshold. On the one hand, paths with $Q_i \geq \tau$ can be treated as available paths. In other words, paths with $Q_i \geq \tau$ will be abandoned temporarily to ensure that packets are able to be sent first via the better paths. On the other hand, Q_i^t is dynamic with the time t. When $|Q_i^t - Q_i^{t-1}| \geq \omega$, the path scheduling list S updates, where the parameter ω works as the switching threshold. If the link quality varies slightly ($|Q_i^t - Q_i^{t-1}| < \omega$), path scheduling sequence will not update. This design is able to minimize the number of link switching and ensure the stability of path scheduling.

Initialization: Path array $S = \phi$;
getSD() is the function of calculating standard deviation factor;
getC() is the function of calculating weight coefficient α, β and θ;
getPW() is the function of calculating path metric value Q_i;
Sort(Ω, Q_i) is to sort the members in Ω based on Q_i in ascending.
void HPCFunction(PATH = $\{P_1, P_2, \ldots, P_i, \ldots, P_n,\ n \geq 2\}$){
  for each $P_i \in$ PATH **do**
    ε_{RTT} = getSD(P_i, RTT);
    ε_{BWT} = getSD(P_i, BWT);
    $\varepsilon_{\text{Loss}}$ = getSD(P_i, LOSS);
    α = getC(ε_{RTT});
    β = getC(ε_{BWT});
    θ = getC($\varepsilon_{\text{LOSS}}$);
    Q_i = getPW(α, β, θ);
    if $Q_i \geq \tau$;
      S.append(P_i);
    end
  end
  S = Sort(S, Q_i);
  for each $P_i \in S$ **do**
    if $|Q_i^t - Q_i^{t-1}| \geq \omega$;
      S = Sort(S, Q_i);
    end
  end
  deliver packets based on the S sequence.
end
}

Algorithm 1: Hybrid awareness path collaboration.

It should be noted that τ and ω can be obtained by the experience of a number of experiments. Before the data sending, the path scheduling policy will select the paths according to the order of the array S. The first one in the array will be the best path. Here, an efficient lookup method can be used to match the delivery path [27]. The controller can select a best path based on its integral link quality and can avoid the blind roll polling in Round Robin mechanism.

The specific path scheduling is shown in Algorithm 1.

The above HPC function guides the controller to select the optimal paths to balance the traffic load for vehicle users. Afterwards, we will further consider how to determine the number of offloading nodes and select them in an optimal manner.

5. Graph-Based Source Collaboration Function

Efficient offloading relies on a set of suitable offloading sources that can quickly distribute data content to subscribing nodes via an ad hoc network. An intuitive idea is to choose the nodes with high probability of connecting other users as the offloading sources.

We consider the source collaboration to minimalize the number of offloading nodes to serve all vehicular nodes in a targeted region. Therefore, source selection strategy can utilize the idea of the minimum vertex cover set in graph theory. It means that we choose the minimum vertex cover set as the offloading sources to connect all the nodes of the network. In order to solve the offloading source selection problem by using the minimum vertex cover set, we first need to convert the network topology into a bipartite graph. The process of bipartite graph transformation divides the vertices in a vertex set into two sets of disjoint sets. The division is based on the fact that two vertices with side relations do not exist in one set. Therefore, the network topology can be transformed into a bipartite graph.

To formulate this problem, we let $G = (V, E)$ be an undirected graph. The vertex V can be divided into two disjoint subsets (A, B) and each edge (i, j) in the graph is associated with two vertexes i and j belonging to these two different vertex sets (i in A, j in B). Then, the graph G is transferred as a bipartite graph.

Theorem 1. *It is necessary and sufficient to judge that the undirected graph G is a bipartite graph, that G has at least two vertices, and that all loops have an even length.*

Let G be the bipartite graph $\langle X, E, Y\rangle$. X and Y represent two subsets in G. Both X and Y are not empty, so G has at least two vertices. If C is any one of the loops in G, let $C = (v_0, v_1, v_2, \ldots, v_i,\ v_0 = v_i)$, where v_i must appear in either X or Y. For example:

$$\{v_0, v_2, v_4, \ldots, v_l = v_0\} \in X, \tag{13}$$

$$\{v_1, v_3, v_5, \ldots, v_{l-1} = v_1\} \in Y. \tag{14}$$

Let all loops of G have an even length, and let G be a connected graph. Without loss of generality, if G is not connected, we can discuss the following conditions for each

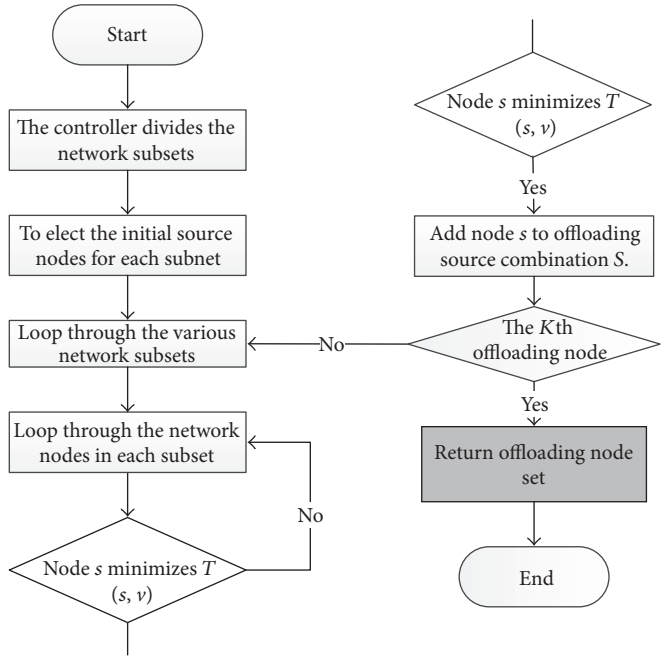


Figure 4: Flowchart of source selection.

connected branch of G. Let the vertices of G be V and the set of edges be E, so that X, Y will be constructed and $\langle X, E, Y\rangle = G$. *Take* $v_0 \in V$ *and set* $X = \{v \mid v = v_0$ *or the path from v to* v_0 *have an even length*$\}$. *Since* $Y = V - X$, *X is obviously not empty. Now we should prove Y is not empty, and neither end of the two edges is in X or in Y. Since* $|V| \geq 2$ *and G is a connected graph,* v_0 *must have adjacent vertices; take* v_1 *as an example; then* $v_1 \in Y$; *therefore Y is not empty.*

On the other hand, suppose that there are edges (u, v), $u \in X$, $v \in Y$. *Then, the path from* v_0 *to u has an even length, or* $u = v_0$; *otherwise, the path from* v_0 *to v has an even length, or* $v = v_0$. *In either case, there always is an odd length closed path from* v_0 *to* v_0 *which contradicts with the supposition. Therefore, it is impossible to have edges* (u, v) *to make both u and v be in X.*

Based on the above analysis, once we get the knowledge of a certain network topology, we can consider converting the network topology into a bipartite graph.

If there are multiple offloading sources to assist the transmission of data content, the vertex cover problem is to find a vertex cover of minimum size in a given undirected graph. This problem is an NP-complete decision problem. It may be very difficult to find an optimal vertex cover in a graph G. Therefore, we can use a heuristic algorithm to find a near-optimal vertex cover, which is the set of collaborative sources. This process is conducted in a software-defined controller.

Figure 4 introduces the flow chart of source selection, which mainly includes four steps in a controller:

(1) At the starting time $t = 0$, the controller determines the connectivity of each network node according to the collected network node information and divides the entire network into multiple subsets.

(2) According to the subdivision result, the controller selects the initial source randomly in each subset. The initial source nodes are characterized by the fact that each node in the subset can communicate with the source.

(3) There may be many different initial source nodes according to the former steps. In this case, an optimal offloading source needs to be selected in each subset according to the restricted condition. The optimal offloading source should guarantee the delivery of data from s to v in a short time.

(4) Repeat step (3) until the selected offloading source nodes keep the same as the last iteration, and exit the program. In this case, the total offloading source number should be a fixed value K.

6. Simulation Results

We conduct a group of simulations to verify the feasibility and efficiency of SDCO under the NS-3 open source network simulator. In our experiments, we used the Routes Mobility Model to simulate the vehicular networks. The Routes Mobility Model can generate realistic mobility traces by querying the Google maps directions API. The information, obtained from the Google maps services, allows the simulator to generate realistic mobility traces based on real-world locations and road networks.

We assess the performance of the proposed SDCO solution by comparing with the state-of-the-art solutions, the typical (TYP) offloading mechanism, and the RSRP/RSRQ offloading mechanism [28]. The offloading sources selected are all user nodes, which are endpoint nodes in this experiment. We compare the offloading performance of the three solutions in terms of the offloading source number and the offloading efficiency. The experimental results are shown in Figures 5, 6, and 7.

As shown in Figure 5, the number of offloading sources is obviously reduced compared with the TYP mechanism. The

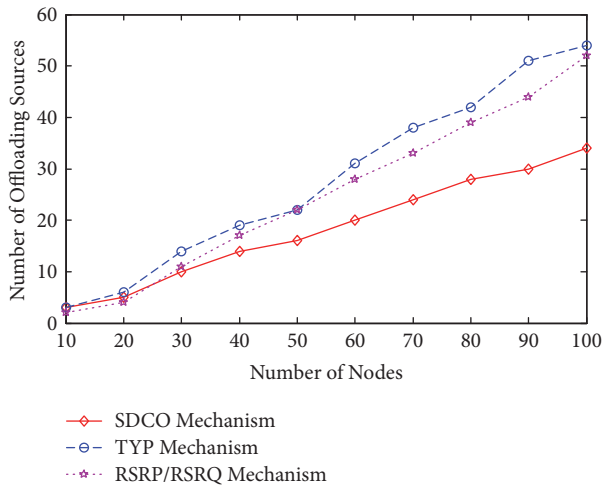


Figure 5: Number of offloading sources analysis.

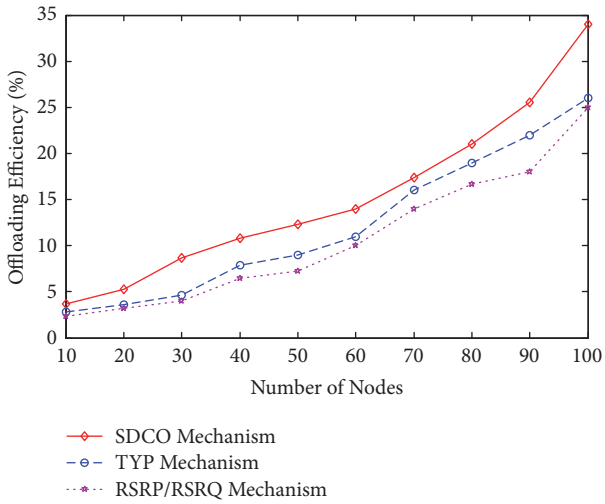


Figure 6: Offloading efficiency analysis (10–100).

number of offloading sources for SDCO is always the lowest, and TYP is always the highest. The RSRP/RSRQ mechanism has a moderate performance and it is a bit better than the one of TYP. In this figure, when the number of vehicle nodes is less than or equal to 10, the offloading sources of TYP and SDCO are almost the same. When the number of vehicle nodes increases more than 10, the number of offloading sources in SDCO becomes significantly different from the other two mechanisms. That is because SDCO adopts the GSC method to guarantee the delivery quality with the least offloading sources.

Figures 6 and 7 show the offloading efficiency of the three offloading mechanisms with a different number of vehicle nodes. Offloading efficiency is the ratio of offloaded traffic to the total traffic. In Figure 6, we can see the overall trend that the offloading efficiency rises as the number of vehicle nodes increases in all three offloading mechanisms. When the number of nodes is less than 100, the difference between TYP and RSRP/RSRQ is not obvious, but SDCO is

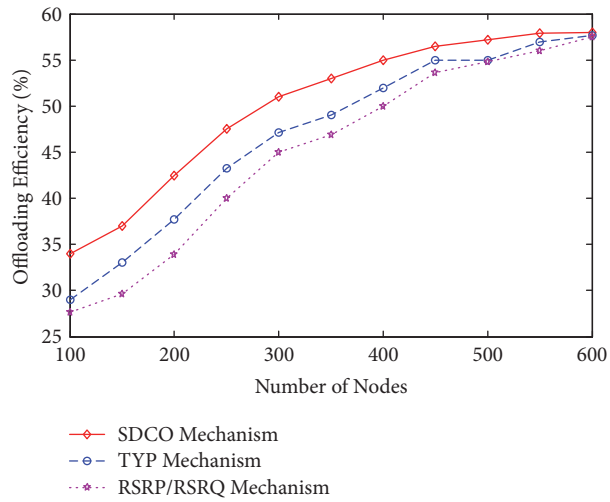


Figure 7: Offloading efficiency analysis (100–600).

much better than both of them. When the number of nodes is more than 100, the advantages of SDCO become more prominent. Compared with the RSRP/RSRQ mechanism, SDCO achieves an average improvement by 18% on the offloading efficiency. The reason is that the TYP mechanism transfers more vehicle node requests to the relay nodes (wireless access points) than the RSRP/RSRQ mechanism. The SDCO mechanism can promote the communications between nodes and the communications between the vehicle nodes and the offloading sources.

When the number of nodes adds to 600, the offloading efficiency keeps at 57% in all three mechanisms. Figure 7 shows there is a saturation status for the offloading efficiency, where the three offloading mechanisms have little difference. That is because there are too many nodes, which result in a high density for the network. Therefore, any one of the three solutions can work well for offloading, and the collaborative operations play a small effect on the performance. It is noted that when some traffic is offloaded from the base station to other relay nodes, the base station will have more resource to process the high-priority requests to further improve the experience of users.

7. Conclusion and Future Work

This paper has proposed a software-defined collaborative offloading solution for heterogeneous vehicular networks, named SDCO. Specifically, SDCO adopts a centralized software-defined controller to globally manage offloading policies to vehicular nodes. In addition, the hybrid awareness path collaboration and the graph-based source collaboration are further introduced to optimize the transmission path and offloading sources selection. Simulation results have been provided to demonstrate that SDCO can achieve better improvement in terms of number of offloading sources and offloading efficiency. In the future works, more efficient offloading control functions should be designed based on the collaboration of vehicular nodes, such as the social-based collaboration.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC) (nos. 61602030, 61702439, and 91638204), the National Key R&D Program (no. 2016YFE0122900), the Fundamental Research Funds for the Central Universities of China (no. 2016RC036), Shandong Provincial Natural Science Foundation, China (no. ZR2017BF018), the China Scholarship Council, and Natural Sciences and Engineering Research Council (NSERC), Canada.

References

[1] K. Zheng, Q. Zheng, P. Chatzimisios, W. Xiang, and Y. Zhou, "Heterogeneous vehicular networking: a survey on architecture, challenges, and solutions," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2377–2396, 2015.

[2] N. Cheng, N. Lu, N. Zhang, X. Zhang, X. S. Shen, and J. W. Mark, "Opportunistic WiFi offloading in vehicular environment: a game-theory approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 7, pp. 1944–1955, 2016.

[3] C. Xu, S. Jia, M. Wang, L. Zhong, H. Zhang, and G.-M. Muntean, "Performance-aware mobile community-based VoD streaming over vehicular Ad Hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 3, pp. 1201–1217, 2015.

[4] C. Xu, S. Jia, L. Zhong, H. Zhang, and G.-M. Muntean, "Ant-inspired mini-community-based solution for video-on-demand services in wireless mobile networks," *IEEE Transactions on Broadcasting*, vol. 60, no. 2, pp. 322–335, 2014.

[5] H. Zhou, N. Cheng, Q. Yu, X. S. Shen, D. Shan, and F. Bai, "Toward multi-radio vehicular data piping for dynamic DSRC/TVWS spectrum sharing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 10, pp. 2575–2588, 2016.

[6] N. Cheng, H. Zhou, L. Lei et al., "Performance analysis of vehicular device-to-device underlay communication," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5409–5421, 2017.

[7] H. Zhang, P. Dong, W. Quan, and B. Hu, "Promoting efficient communications for high speed railway using smart collaborative networking," *IEEE Wireless Communications Magazine*, vol. 22, no. 6, pp. 92–97, 2015.

[8] H. Zhou, N. Cheng, N. Lu et al., "WhiteFi infostation: Engineering vehicular media streaming with geolocation database," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 8, pp. 2260–2274, 2016.

[9] F. Rebecchi, M. Dias de Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti, "Data offloading techniques in cellular networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 580–603, 2015.

[10] G. Mao, Z. Zhang, and B. Anderson, "Cooperative content dissemination and offloading in heterogeneous mobile networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6573–6587, 2016.

[11] N. Lu, N. Zhang, N. Cheng, X. Shen, J. W. Mark, and F. Bai, "Vehicles meet infrastructure: toward capacity-cost tradeoffs for vehicular access networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1266–1277, 2013.

[12] Y. Liu, Z. Yang, T. Ning, and H. Wu, "Efficient quality-of-service (QoS) support in mobile opportunistic networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 9, pp. 4574–4584, 2014.

[13] H. Zhang, W. Quan, H.-C. Chao, and C. Qiao, "Smart identifier network: A collaborative architecture for the future internet," *IEEE Network*, vol. 30, no. 3, pp. 46–51, 2016.

[14] W. Quan, Y. Liu, H. Zhang, and S. Yu, "Enhancing crowd collaborations for software defined vehicular networks," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 80–86, 2017.

[15] K. Wang, H. Yin, W. Quan, and G. Min, "Enabling collaborative edge computing for software defined vehicular networks," *IEEE Network*, vol. PP, no. 99, pp. 1–6, 2018.

[16] J. Chen, H. Zhou, N. Zhang et al., "Service-oriented dynamic connection management for software-defined internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 10, pp. 2826–2837, 2017.

[17] K. Zheng, L. Hou, H. Meng, Q. Zheng, N. Lu, and L. Lei, "Soft-defined heterogeneous vehicular network: architecture and challenges," *IEEE Network*, vol. 30, no. 4, pp. 72–80, 2016.

[18] P. Dong, X. Du, J. Sun, and H. Zhang, "Energy-efficient cluster management in heterogeneous vehicular networks," in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 644–649, San Francisco, Calif, USA, April 2016.

[19] N. Zhang, S. Zhang, P. Yang, O. Alhussein, W. Zhuang, and X. S. Shen, "Software defined space-air-ground integrated vehicular networks: challenges and solutions," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 101–109, 2017.

[20] Y. Liu, C. Xu, Y. Zhan, Z. Liu, J. Guan, and H. Zhang, "Incentive mechanism for computation offloading using edge computing: a Stackelberg game approach," *Computer Networks*, vol. 129, pp. 399–409, 2017.

[21] N. Cheng, F. Lyu, J. Chen et al., "Big data driven vehicular networks," *IEEE Network*, vol. PP, no. 99, pp. 1–8, 2018.

[22] A. Aijaz, H. Aghvami, and M. Amani, "A survey on mobile data offloading: technical and business perspectives," *IEEE Wireless Communications Magazine*, vol. 20, no. 2, pp. 104–112, 2013.

[23] X. Zhuo, W. Gao, G. Cao, and S. Hua, "An incentive framework for cellular traffic offloading," *IEEE Transactions on Mobile Computing*, vol. 13, no. 3, pp. 541–555, 2014.

[24] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: how much can wifi deliver?" *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 536–550, 2013.

[25] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: modeling, analysis, and design insights," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2484–2497, 2013.

[26] N. Cheng, N. Lu, N. Zhang, X. S. Shen, and J. W. Mark, "Opportunistic WiFi offloading in vehicular environment: A queueing analysis," in *Proceedings of the 2014 IEEE Global Communications Conference, GLOBECOM 2014*, pp. 211–216, Austin, Tex, USA, December 2014.

[27] W. Quan, C. Xu, J. Guan, H. Zhang, and L. A. Grieco, "Scalable name lookup with adaptive prefix bloom filter for named data networking," *IEEE Communications Letters*, vol. 18, no. 1, pp. 102–105, 2014.

[28] L. Al-Kanj, H. V. Poor, and Z. Dawy, "Optimal cellular offloading via device-to-device communication networks with fairness constraints," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4628–4643, 2014.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 6870742, 9 pages
https://doi.org/10.1155/2018/6870742

Research Article

Concurrently Deniable Group Key Agreement and Its Application to Privacy-Preserving VANETs

Shengke Zeng[1,2] **and Yong Chen**[1,3]

[1]*School of Computer and Software Engineering, Xihua University, Chengdu, China*
[2]*Center for Cyber Security, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China*
[3]*School of Electrical Engineering, Southwest Jiaotong University, Chengdu, China*

Correspondence should be addressed to Shengke Zeng; zengshengke@gmail.com

Received 5 December 2017; Accepted 4 March 2018; Published 15 April 2018

Academic Editor: Ning Zhang

VANETs need secure communication. Authentication in VANETs resists the attack on the receipt of false information. Authenticated group key agreement (GKA) is used to establish a confidential and authenticated communication channel for the multiple vehicles. However, authentication incurs privacy leakage, that is, by using digital signature. Therefore, the deniability is deserved for GKA (which is termed as DGKA) due to the privacy protection. In the DGKA protocol, each participant interacts with intended partners to establish a common group session key. After this agreement session, each participant can not only be regarded as the intended sender but also deny that it has ever participated in this session. Therefore, under this established key, vehicles send confidential messages with authentication property and the deniability protects the vehicles privacy. We present a novel transformation from an unauthenticated group key agreement to a deniable (authenticated) group key agreement without increasing communication round. Our full deniability is achieved even in the concurrent setting which suits the Internet environment. In addition, we design an authenticated and privacy-preserving communication protocol for VANETs by using the proposed deniable group key agreement.

1. Introduction

Vehicular ad hoc networks (VANETs) [1] refer to the peer-to-peer networks formed by roadside units and adjacent vehicles for sharing information, including traffic information (the speed and flow of vehicles, etc.) and warning information. VANETs provide a safe and comfortable driving environment for the drivers, which avoids the congestion and traffic accidents. VANETs should provide secure communication in case false information is inserted into the network. Besides that, VANETs should also have privacy issues as vehicles are reluctant to expose the sensitive information while sharing their own traffic information.

The group key agreement (GKA) provides a secure channel for the vehicles communication. GKA protocol [2] allows a group of participants to establish a common session key for a secure communication channel over an insecure network by agreement. However, key agreement without authentication incurs man-in-middle attack. In order to handle this problem, the authentication is necessary. However, authentication binds the identity, which causes the privacy leakage. In many cases, the participants do not want the third party to know their involvements in some key agreements. In other words, they want to have the capacity to deny that they have ever participated in some sessions after the key agreement execution. Hence, the deniable GKA (DGKA) protocol was presented by introducing deniability into GKA protocol. Bohli and Steinwandt first formalized the DGKA protocol in [3]. In the DGKA protocol, it is not feasible to convince a third party that these participants in a group key agreement session have been involved in the conversation from the communication transcript. In other words, each participant can deny its involvement to the third party.

1.1. Related Work. The deniability is formalized by introducing a simulator that can simulate the conversation transcript

without secrets. Therefore, the participants can deny this as someone else would produce this indistinguishable conversation transcript. If this simulator can be run by anyone, it is denoted by the full deniability. Deniable authentication was first introduced by Dolev et al. in [4] and formally studied by Dwork et al. in [5]. The general technique to realize the deniable authentication is that the sender uses its secret (i.e., the private key) to generate a value v_s. If the receiver produces a value v_r which equals v_s by using a related witness, the receiver is convinced of the sender's authentication. In order to simulate the transcript (for the deniability), this witness has to be revoked. Thus, the early works such as [5, 6] require more rounds to revoke the witness upon the receipt of the committed v_s (i.e., COM(v_s)). In this way, the simulator run by anyone can extract this witness to simulate the transcript by rewinding steps. However, the deniability does not hold in the concurrent scenario due to the rewinding. Therefore, the timing assumption is necessary to be considered to realize concurrent deniability, such as [5, 6].

However, the timing assumption is farfetched for the Internet which is a fully concurrent environment. Some related works have to handle this problem by avoiding rewinding. Di Raimondo et al. [7] showed that the plaintext-awareness [8] of the underlying encryption can extract the witness for the simulation without rewinding steps. Jiang's work [9] depends on the public random oracle to extract the witness and therefore the rewinding steps are not necessary. Yao and Zhao [10, 11] proposed the deniable Internet key exchange based on the knowledge of exponent assumption (KEA). By the KEA assumption, the witness can be extracted and the transcripts are perfectly simulated. Tian et al. [12] made use of the selectively unforgeable but existentially forgeable signature to simulate the transcript. Zeng et al. [13] presented a multireceiver encryption under KEA assumption and used it as a building block to propose a concurrently deniable ring authentication. Jiang [14] made use of a moderate encryption to avoid rewinding to construct the concurrently deniable key exchange. These approaches achieve the deniability without rewinding steps; thus the simulation even in the concurrent scenario is normal. However, as we see above, these works suffer the limitations such as the strong assumptions, inefficiency, or random oracles.

Deniable authentication has been applied to many occasions nowadays [13, 15, 16] and was first introduced into the two-party key exchange protocol [17] in [18]. Mao and Paterson [18] informally defined the deniable key exchange (DKE) protocol and obtained its deniability by using identity-based techniques. Later, how to use the approach in [19] to design a DKE protocol was discussed and a concrete approach was proposed in [20], where the technique based on the public information was used to derive a symmetric key for authentication. Following this work, a series of DKE protocols were proposed in [10, 21, 22].

When a two-party DKE protocol is extended to a group setting, there may be some troubles [23, 24]. If there exists malicious insiders, the use of a common symmetric key for deniable authentication is infeasible as the malicious participants may impersonate other participants [3]. A solution provided in [3] is to make use of Schnorr's zero-knowledge identification scheme [25]. This approach needs 4 rounds to complete the establishment of session key and its efficiency was improved by Zhang et al. [26], which reduced the communication round to 3. Some approaches [27–29] transform the passively secure group key establishment to an actively secure one by adding one more round and the deniability was achieved as well.

DGKA protocols can be applied to VANETs to provide the security and privacy protection for vehicles. In recent years, wireless networks (WN) have achieved rapid development [30], and their security issues have been extensively studied [31–33]. As a kind of WN, the security of VANETs should be taken seriously due to the high risk. Some related schemes for secure communication in VANETs have been presented [34–36]. Huang et al. [35] proposed a communication scheme based on GKA protocol that the roadside unit generated session key for adjacent vehicles in batches. This scheme could effectively reduce the cost of computation and communication. In [36], a representative selected from the adjacent vehicles was arranged to communicate with the roadside unit, thus making that the security of other vehicles guaranteed. Nevertheless, the public verification in these works leaks the privacy of vehicles. Hence, the deniable group key agreement is necessary to apply to privacy-preserving communication for VANETs.

1.2. Contribution. We focus on the full deniability of the authenticated group key agreement. We provide a generic transformation from unauthenticated GKA to DGKA without increasing any additional communication rounds. Moreover, our deniability does not require rewinding steps; thus it holds even in the concurrent environment such as Internet. We also do not depend on any strong assumptions to reach the full deniability. The contribution of this work is as follows.

(1) We present a generic transformation from an unauthenticated GKA (named as DB protocol [37]) to a deniable (authenticated) GKA. The existing works achieve the full deniability by the rewinding steps, KEA assumption (which is strong), or the public random oracles. It results in inefficiency or insecurity (strong assumption). Our approach does not resort to these ways. We do not require that the underlying primitive is PA secure and the random oracles are not necessary.

(2) Our work achieves the concurrent deniability without timing constraint. In concurrent setting, the adversary can open and schedule sessions arbitrarily. Indeed, our simulation does not require extracting the witness by rewinding steps. Thus, the concurrent session attack (i.e., adversaries schedule the executions or delay messages in arbitrary ways) does not work in our scheme.

(3) We realize the optimal communication complexity. Our transformation does not increase the round of the unauthenticated one (original DB) although it realizes the property of privacy-preserving authentication in GKA, while the related works such as [3, 26]

Round 1: Participant U_i performs the following steps:
(1) Choose $x_i \in Z_q^*$ and compute $X_i = g^{x_i}$.
(2) Broadcast message (U_i, X_i).
Round 2: Upon receiving messages (U_{i-1}, X_{i-1}) and (U_{i+1}, X_{i+1}), each U_i does as following:
(1) Compute $Y_i^L = X_{i-1}^{x_i}$, $Y_i^R = X_{i+1}^{x_i}$, $Y_i = Y_i^R / Y_i^L$.
(2) Broadcast message (U_i, Y_i).
Session Key Generation: Upon receiving all messages $(U_j, Y_j)_{j\in\{1,\ldots,n\}, j\neq i}$, each U_i carries out the following steps:
(1) Compute orderly $\widehat{Y}_{i+1}^R = Y_{i+1} \cdot Y_i^R$, $\widehat{Y}_{i+2}^R = Y_{i+2} \cdot \widehat{Y}_{i+1}^R, \ldots, \widehat{Y}_{i+(n-1)}^R = Y_{i+(n-1)} \cdot \widehat{Y}_{i+(n-2)}^R$.
(2) Check $Y_i^L \stackrel{?}{=} \widehat{Y}_{i+(n-1)}^R$. If it is true, continue; Otherwise, abort.
(3) Generate the session key $\mathrm{sk} = \widehat{Y}_1^R \cdot \widehat{Y}_2^R \cdot \ldots \cdot \widehat{Y}_n^R = g^{x_1x_2 + x_2x_3 + \cdots + x_nx_1}$.

Algorithm 1: DB-GKA protocol without authentication.

have to increase the additional rounds to obtain the deniability.

(4) We also design a privacy-preserving communication protocol for VANETs using the proposed DGKA protocol. In this communication protocol, the vehicles share their information without leaking any identity privacy and without leaving any evidence in the transcript of authentication.

Organization. This paper is organized as follows. Section 2 introduces some preliminaries which are the building blocks in our protocol. Section 3 describes the adversarial model and related security definitions of the DGKA. We propose an efficient DGKA protocol with 2 rounds in Section 4. The security of DGKA protocol is proven and the performance is analyzed in Section 5. We design a privacy-preserving protocol for VANETs in Section 6. Section 7 concludes this work.

2. Preliminaries

We show the notations and introduce the building blocks in this section.

2.1. Notations. The notations used in this paper are listed in Notations in DGKA Protocol.

2.2. DB-GKA Protocol. Our deniable group key agreement (DGKA) protocol is developed on the basis of Dutta-Barua (DB) GKA protocol [37], which is a 2-round unauthenticated GKA protocol. It is a variant of [38]. We now review the original DB-GKA protocol [37]. Each participant U_i chooses x_i as its short-term private key, computes $X_i = g^{x_i}$, and broadcasts X_i in the first round. In Round 2, upon the receipt of messages (X_{i-1}, X_{i+1}), U_i computes $Y_i = f(X_{i-1}, X_{i+1})$ and broadcasts it. Finally, each U_i generates the common session key sk with the received Y_i and its secret x_i. The concrete DB-GKA protocol is presented in Algorithm 1. The security of DB-GKA protocol has been proven in [37].

2.3. Ring Signature with 2 Members. Our deniable group key agreement protocol provides the deniability based on the ring signature with 2 members. Now we introduce the syntax and the security properties of the ring signature with 2 members.

The ring signature scheme was used to sign a message privately. Given a valid ring signature σ with respect to a message M and a set of public keys $\mathcal{PK} = \{\mathrm{PK}_1, \ldots, \mathrm{PK}_n\}$, any verifier cannot decide which member in set $\mathcal{PK}$ is the actual signer.

We consider the ring signature with n members where $n = 2$. The syntax of the ring signature is as follows.

(1) A probabilistic key generation algorithm KGen: given the security parameter κ, output the key pair $(\mathrm{PK}_i, \mathrm{SK}_i)$ for U_i $(i = 1, 2)$. That is, $(\mathrm{PK}_i, \mathrm{SK}_i) \leftarrow \mathrm{KGen}(1^\kappa)$.

(2) A probabilistic ring signing algorithm RSig: given a message M, two public keys $(\mathrm{PK}_1, \mathrm{PK}_2)$, and a private (signing) key of U_k $(k \in \{1, 2\})$, output the ring signature σ. That is, $\sigma \leftarrow \mathrm{RSig}(M, (\mathrm{PK}_1, \mathrm{PK}_2); \mathrm{SK}_k)$.

(3) A deterministic verification algorithm RVer: given the ring signature σ, the message M, and the two public keys $(\mathrm{PK}_1, \mathrm{PK}_2)$, determine whether σ is valid with respect to $(M, \mathrm{PK}_1, \mathrm{PK}_2)$. That is, check $1 \stackrel{?}{=} \mathrm{RVer}(\sigma, M, \mathrm{PK}_1, \mathrm{PK}_2)$.

The properties of a secure ring signature with 2 members contain the *unconditional anonymity* and *unforgeability* as follows.

(i) *Unconditional Anonymity*. The distributions of the two ring signatures $\sigma_1 \leftarrow \mathrm{RSig}(M, (\mathrm{PK}_1, \mathrm{PK}_2); \mathrm{SK}_1)$ and $\sigma_2 \leftarrow \mathrm{RSig}(M(\mathrm{PK}_1, \mathrm{PK}_2); \mathrm{SK}_2)$ are `statistic`, identical. It implies that, given a ring signature σ with respect to $(M, \mathrm{PK}_1, \mathrm{PK}_2)$, no one can decide the signer although the private keys $(\mathrm{SK}_1, \mathrm{SK}_2)$ are revealed.

(ii) *Unforgeability*. A forger without the signing key SK_1 or SK_2 forges a ring signature $\widehat{\sigma}$ with respect to $(M, \mathrm{PK}_1, \mathrm{PK}_2)$. The probability that $1 \leftarrow \mathrm{RVer}(\widehat{\sigma}, M, \mathrm{PK}_1, \mathrm{PK}_2)$ is negligible.

3. Model of Deniable Group Key Agreement Protocol

3.1. Syntax. The syntax of the deniable group key agreement (DGKA) protocol is as follows. Let $\mathcal{U} = \{U_1, \ldots, U_n\}$

denote the set of n potential participants who would like to build a common session key to communicate securely. Each participant $U_i \in \mathcal{U}$ has a private/public key pair $(\mathrm{SK}_i, \mathrm{PK}_i)$ and the public keys are authenticated and can be accessed by any member. The DGKA protocol may be executed among any subsets of $\mathcal{U}$ at any time. At the end of this execution, the common session key is built. Each participant is convinced of the identity of his partners. In addition, all of them can also deny the involvement in this conversation of this session.

3.2. Security Model. We formalize the underlying adversarial behaviors in this subsection.

(i) Execute($\mathrm{pid}_i^{l_i}$): this query models the passive attacks in which the adversary can only eavesdrop the execution of protocol among the participants in $\mathrm{pid}_i^{l_i}$ and outputs the transcript of the session $\mathrm{sid}_i^{l_i}$. The transcript consists of the messages that are exchanged during the honest execution of the protocol.

(ii) Send(d, i, l_i, M): this query models the active attacks which the adversary can arbitrarily eavesdrop, delay, modify, and insert on any message M. The output of this query is the reply generated by instance $\pi_i^{l_i}$. When $d = 0$, the query initializes the execution of the instance $\pi_i^{l_i}$.

(iii) Reveal(i, l_i): if instance $\pi_i^{l_i}$ has successfully accepted the session key $\mathrm{sk}_i^{l_i}$, then $\mathrm{sk}_i^{l_i}$ is returned. Otherwise, NULL is returned.

(iv) Corrupt(i): the long-term private key of participant U_i is returned, and the future action will be fully taken by adversary. This query implies that there exists the malicious insiders.

(v) Test(i, l_i): the query is allowed only once. The queried instance $\pi_i^{l_i}$ must be fresh and $\mathrm{sk}_i^{l_i}$ is not NULL. Furthermore, this session as well as its partnered session should not be issued when a Corrupt query or Reveal query occurs. When the Test query occurs, a bit $b \in \{0, 1\}$ is randomly chosen. The session key $\mathrm{sk}_i^{l_i}$ is returned if $b = 1$; otherwise a random value of the same length is returned if $b = 0$.

(vi) Response: the adversary outputs a guess bit b'. We say that the adversary wins the game if $b' = b$. Let $\mathrm{Succ}_{\mathscr{A}}$ denote the event that the adversary wins the game and $\mathrm{Adv}_{\mathscr{A}}$ denote the advantage of the adversary by $\mathrm{Adv}_{\mathscr{A}} = |\Pr[\mathrm{Succ}_{\mathscr{A}}] - 1/2|$.

Freshness. An instance $\pi_i^{l_i}$ is fresh if none of the following happens: (1) A Reveal(i, l_i) query or a Reveal(j, l_j) query happens, where $\pi_i^{l_i}$ is partnered with $\pi_j^{l_j}$. (2) A Corrupt(j) query happens, where $U_j \in \mathrm{pid}_i^{l_i}$.

Partnering. The instances $\pi_i^{l_i}$ and $\pi_j^{l_j}$ are said to be partnered if $\mathrm{sid}_i^{l_i} = \mathrm{sid}_j^{l_j}$ and $\mathrm{pid}_i^{l_i} = \mathrm{pid}_j^{l_j}$.

Communicational Networks. We assume that our protocol is executed in the broadcasting channel; thus the adversaries can arbitrarily eavesdrop, delay, modify, and insert any message.

A secure DGKA protocol should satisfy the *correctness*, *deniability*, *authentication*, and *secrecy*.

Correctness. This property states that the protocol will establish a session key without adversarial interference. The DGKA protocol is said to be correct if for any pair of instances $\pi_i^{l_i}$ and $\pi_j^{l_j}$ ($i, j = 1, \ldots, n$ and $i \neq j$), which have been accepted with $\mathrm{sid}_i^{l_i} = \mathrm{sid}_j^{l_j}$ and $\mathrm{pid}_i^{l_i} = \mathrm{pid}_j^{l_j}$, the condition $\mathrm{sk}_i^{l_i} = \mathrm{sk}_j^{l_j} \neq \mathsf{NULL}$ holds.

Deniability. This deniability states that the adversary cannot convince anyone that the honest participants have indeed joined in some sessions. Let $\mathscr{A}_d$ be the adversary that violates the deniability. We use the *simulation* paradigm to formally define the deniability. We construct a simulator $\mathcal{S}$ that is a probabilistic polynomial time (PPT) Turing machine. The simulator $\mathcal{S}$ can answer all queries from the adversary $\mathscr{A}_d$, and its inputs only involve the public information and the long-term private keys of the corrupted participants. Let $\mathtt{View}_{\mathcal{S}}$ denote the outputs of the adversary $\mathscr{A}_d$ after interacting with the simulator $\mathcal{S}$. Let $\mathtt{View}_{\mathrm{R}}$ denote the outputs of the adversary $\mathscr{A}_d$ in the real world. The protocol is said to be deniable if, for any PPT adversary $\mathscr{A}_d$ and the distinguisher $\mathscr{D}$ with unbounded computation, there exists a simulator $\mathcal{S}$, such that $|\Pr[\mathscr{D}(\mathtt{View}_{\mathcal{S}}) = 1] - \Pr[\mathscr{D}(\mathtt{View}_{\mathrm{R}}) = 1]| = \mathrm{negl}(\kappa)$.

Authentication. The authentication of the protocol guarantees that the received messages of the participants come from the intended participants. If an adversary $\mathscr{A}_a$ that may even be a malicious insider can impersonate an uncorrupted participant U_i and succeed to accomplish the protocol, then we say the adversary violates the authentication of DGKA protocol. We use Forge to denote the event that the adversary succeeds in cheating the honest participants. The protocol is said to be authenticated if $\Pr[\mathrm{Forge}] \approx \mathrm{negl}(\kappa)$ for any PPT adversary.

Secrecy. The secrecy of the protocol states that the session key is known only to participants but is random to outsiders. Formally, let $\mathscr{A}$ be the adversary that violates the secrecy and $\mathrm{Succ}_{\mathscr{A}}$ denote the success of $\mathscr{A}$ in the Test query, who decides the session key from a random value successfully. We say the protocol meets the secrecy if $\Pr[\mathrm{Succ}_{\mathscr{A}}] \approx 1/2 + \mathrm{negl}(\kappa)$.

4. Our Deniable Group Key Agreement Protocol

We construct the deniable GKA protocol based on Dutta-Barua (DB) GKA protocol [37], which is elaborated in Section 2. Our DGKA protocol achieves the deniable authentication by employing a ring signature with 2 members. We first give the high level description of our DGKA protocol.

Let (SK_i, PK_i) denote the private/public key pair for the participant U_i and n is the number of the participants of this session.

Round 1: Participant U_i performs the following steps:

(1) Choose $x_i, t_i \in Z_q^*$ and compute $X_i = g^{x_i}$, $T_i = g^{t_i}$.

(2) Broadcast message $M_i^1 = (U_i, X_i, T_i)$.

Round 2: Upon the receipt of all messages $\{M_j^1\}_{j\in\{1,\ldots,n\},j\neq i}$, U_i parses X_{i-1}, X_{i+1} and $\{T_j\}_{j\in\{1,\ldots,n\},j\neq i}$. Next, U_i executes the following operations:

(1) Compute $Y_i^L = X_{i-1}^{x_i}$, $Y_i^R = X_{i+1}^{x_i}$, $Y_i = Y_i^R/Y_i^L$, $T = \prod_{j=1}^n T_j$.

(2) Generate a two-member ring signature on the message $M = (X_1, \ldots, X_n, Y_i)$: $\sigma_i = \mathrm{RSig}(M, (PK_i, T); SK_i)$.

(3) Broadcast message $M_i^2 = (U_i, Y_i, \sigma_i)$.

Session Key Generation: Upon the receipt of all messages $\{M_j^2\}_{j\in\{1,\ldots,n\},j\neq i}$, each U_i carries out the following steps:

(1) Compute orderly $\widehat{Y}_{i+1}^R = Y_{i+1} \cdot Y_i^R$, $\widehat{Y}_{i+2}^R = Y_{i+2} \cdot \widehat{Y}_{i+1}^R, \ldots, \widehat{Y}_{i+(n-1)}^R = Y_{i+(n-1)} \cdot \widehat{Y}_{i+(n-2)}^R$. Check $Y_i^L \stackrel{?}{=} \widehat{Y}_{i+(n-1)}^R$. If it is true, continue; Otherwise, abort.

(2) Check $1 = \mathrm{RVer}(\sigma_i, M, PK_j, T)$ hold or not for $j = 1, \ldots, n$ and $j \neq i$. If it fails to any participant, abort; Otherwise, continue.

(3) Generate the session key $sk = \widehat{Y}_1^R \cdot \widehat{Y}_2^R \cdot \ldots \cdot \widehat{Y}_n^R = g^{x_1x_2+x_2x_3+\cdots+x_nx_1}$.

Algorithm 2: Our deniable group key agreement protocol.

Considering a ring signature scheme with two members: a real participant and a logic entity. In the first round, each participant follows DB-GKA protocol to generate X_i. Besides that, each one also produces another group element T_i. The product of each T_i is viewed as the public key of the logic entity. Therefore, in the second round, each participant gathers all T_i to result T. Then each one uses its own public key PK_i and the logic public key T to form a ring to generate a ring signature on the message $(X_1, \ldots, X_n, Y_i)$ with its signing key SK_i. The corresponding private key of the logic public key T is unknown to any participant and the third party. Thus, a valid ring signature implies that the authentication to $(X_1, \ldots, X_n, Y_i)$ can be completed only by the participant U_i. The authentication is achieved. On the other hand, the simulator can simulate the value T by its random choice of the exponent t to get $T = g^t$. Then, the simulator produces a ring signature $\sigma' = \mathrm{RSig}(M, (PK_i, T); t)$ with the "private key" of T. By the unconditional anonymity property of the ring signature, the two distributions of $\sigma = \mathrm{RSig}(M, (PK_i, T); SK_i)$ and $\sigma' = \mathrm{RSig}(M, (PK_i, T); t)$ are statistic, identical, where the former one is the real transcription. Therefore, the simulation is perfect and the deniability is achieved. Since the rewinding steps are not necessary in the simulation, the deniability can also hold in the concurrent setting. We give a detailed description of our protocol in Algorithm 2.

Remark 1. The ring signature is with 2 members. One is the participant U_i, and the other one is a logic entity whose public key is $T = \prod_{j=1}^n T_j$. Obviously, the private key of T is $t = \sum_{j=1}^n t_j$ and it is unknown to anybody. In the real conversation, U_i uses its private key SK_i to generate the ring signature σ. Since σ is only bounded to 2 public keys and one of the public key is logic with unknown secret, the partner can be convinced of U_i's signing. The authentication is completed. Meanwhile, in the simulation, the simulator simulates t (as no secret value is required) to produce the ring signature. Obviously, this simulation is perfect without any rewinding steps; concurrent deniability is realized.

5. Security and Performance

In this section, we analyze the security and performance of our protocol. Since the verification of *correctness* of our protocol is straightforward, in what follows we will prove that our protocol meets the other three properties: *deniability*, *authentication*, and *secrecy*, which have been presented in the security model. Then we give the performance comparisons of the related deniable key agreements regarding the communication round and the deniability.

5.1. Security

5.1.1. Deniability. This property states that all the participants can deny the fact that they have joined in the generation of the session key. We use the *simulation* fashion to prove that our protocol satisfies the deniability. That is, if a simulator without any participant's secret can simulate the transcript and the simulated transcript is indistinguishable from the real one, then we say the deniability is proven. Formal proof is presented as follows.

Theorem 2. *The DGKA protocol is concurrently deniable if the underlying ring signature is secure.*

Proof. In order to prove our protocol satisfying the deniability, we have to show the real view and the simulated view are indistinguishable. Formally, we construct a simulator $\mathcal{S}$, whose inputs involve the public information and the long-term private keys of the corrupted participants. $\mathcal{A}_d$ is an adversary that violates the deniability of the protocol. Use $\mathtt{View}_R$ to denote the view of $\mathcal{A}_d$ in the real conversation and $\mathtt{View}_{\mathcal{S}}$ to denote the view of $\mathcal{A}_d$ in the simulated setting performed by $\mathcal{S}$. We show that any distinguisher $\mathcal{D}$

with unbounded computation cannot distinguish $\mathtt{View}_{\mathrm{R}}$ and $\mathtt{View}_{\mathcal{S}}$.

With the inputs of $\{\mathrm{PK}_i\}$ and the long-term private keys of the corrupted participants, $\mathcal{S}$ simulates the Send, Corrupt, and Reveal queries for $\mathcal{A}_d$ as follows.

(i) Send$(0, i, l_i, M)$: $\mathcal{S}$ normally performs protocol and answers the query as it does not require any secrets. $\mathcal{S}$ randomly chooses $x_i, t_i \in Z_i^*$ to compute X_i, T_i, respectively. Then, $\mathcal{S}$ broadcasts message $M_i^1 = (U_i, X_i, T_i)$ and records $\mathrm{stat}_i^{l_i} = (x_i, X_i, t_i, T_i)$.

(ii) Send$(1, i, l_i, M)$: $\mathcal{S}$ checks if U_i has been corrupted.

(a) If U_i has been corrupted, $\mathcal{S}$ with the known private key SK_i simulates M_i^2 normally.

(b) If U_i is uncorrupted, $\mathcal{S}$ retrieves x_j, t_j from $\mathrm{stat}_j^{l_j}$ to compute Y_j (where $j = 1, \ldots, n$), $t = \sum_{j=1}^{n} t_j$, and $T = g^t$. Then $\mathcal{S}$ produces a ring signature $\sigma_i = \mathrm{RSig}(M, (\mathrm{PK}_i, T); t)$. $\mathcal{S}$ updates $\mathrm{stat}_i^{l_i} = (x_i, t_i, X_i, T_i, Y_i, \sigma_i)$.

(iii) Send$(2, i, l_i, M)$: $\mathcal{S}$ normally answers the query no matter whether U_i has been corrupted or not as there is no secret required.

(iv) Reveal(i, l_i): $\mathcal{S}$ computes the session key $\mathrm{sk}_i^{l_i}$ according to the protocol and returns it to $\mathcal{A}_d$.

(v) Corrupt(i): $\mathcal{S}$ returns the private key SK_i of participant U_i and the fact that U_i is corrupted is marked.

Now, we argue that $\mathtt{View}_{\mathrm{R}}$ and $\mathtt{View}_{\mathcal{S}}$ are perfectly identical. It is obvious that $\mathcal{S}$ does not introduce any difference from the view of real one when Send$(0, i, l_i, M)$, Send$(2, i, l_i, M)$, Reveal(i, l_i), and Corrupt(i) are asked. Let us consider Send$(1, i, l_i, M)$. In the real transcript, Send$(1, i, l_i, M)$ is performed using U_i's private key SK_i. In the simulation, this oracle is answered using t, which is the private key of the logic party (whose public key is $T = g^t$). This is a ring signing with U_i and the logic party. Since the underlying ring signature scheme with two members is secure, it implies that the unconditional anonymity property holds. If Send$(1, i, l_i, M)$ introduces any difference, which means the ring signature under SK_i and the ring signature under t can be distinguishable, obviously, it breaks the unconditional anonymity of this ring signature scheme. It is a contradiction. □

5.1.2. Authentication. Authentication states that each U_i can ensure that the message it received is authenticated by the intended partner. This property can prevent the man-in-middle attack which exists in the unauthenticated key agreement protocol. In our protocol, we apply the ring signature with two members to preserve the authentication. Indeed, the generated ring signature σ_i is bounded to two public keys PK_i and T. Due to the unforgeability of the ring signature, anyone who knows SK_i or t can generate a valid signature. Given a valid σ_i, the partner is convinced that $M = (X_1, \ldots, X_n, Y_i)$ which is used to generate the common session key is signed by U_i as t is unknown to anyone. Obviously, our protocol is authenticated due to the unforgeability of the underlying ring signature scheme.

5.1.3. Secrecy. This property ensures the security of the session key. That is, any member without participating in the session cannot obtain the session key. Obviously, our DGKA protocol satisfies the secrecy if DB-GKA protocol produces the session key securely. It is easy to see that our DGKA protocol equals the original DB-GKA protocol only except that we provide a ring signing on $(X_1, \ldots, X_n, Y_i)$ in DGKA. We denote the game $\mathtt{G}_0$ as the environment of DB-GKA protocol and the game $\mathtt{G}_1$ as the environment of our DGKA. Let Forge be the event that $\mathcal{A}$ succeeds in forging a valid message after Round 2. The difference between the games $\mathtt{G}_0$ and $\mathtt{G}_1$ is that the challenger in $\mathtt{G}_1$ would stop the simulation when the event Forge occurs. However, Pr[Forge] is negligible as the authentication property states. Therefore, the secrecy of our DGKA protocol can be reduced to the secrecy of DB-GKA, which is proven in [37].

5.2. Performance. The obvious advantage of our construction is the optimal communication round. We transfer the unauthenticated DB-GKA protocol to the deniable GKA without increasing round. While other related DGKA protocols are more than 2 rounds.

One DGKA protocol [3] is based on Schnorr's zero-knowledge identification scheme; the participants make the commitments in Round 1. Next, an unauthenticated GKA protocol is executed in Rounds 2 and 3. The deniable authentication is achieved in Round 4. It needs 4 rounds to complete the protocol. Similarly, in [26], the participants also make commitments in Round 1. At the same time, the participants begin to execute the unauthenticated GKA protocol in this round. Finally, the deniable authentication is executed in Round 3. It is easy to see that the deniable authentication depends on the generated session key in [3, 26]. This is the reason that these two protocols require more rounds than the unauthenticated GKA to realize the deniable authentication.

Our protocol makes use of the unconditional anonymity of the ring signature to achieve the concurrent deniability. This ring signature is bounded to 2 members. The one is the actual participant and the other one is a logic party. This logic public key is accumulated by all participants with its own secret in Round 1. Then each participant uses the logic public key and its own public key to form a ring and signs the elements which are used to generate the common session key in Round 2. Obviously, the deniability is no longer dependent on the session key. Therefore, our work does not increase the communication round of the unauthenticated GKA.

We also focus on the concurrent deniability. However, both [3, 26] depend on the rewinding steps to simulate the transcript. Therefore, the deniability cannot hold in the concurrent setting. Some other deniable authentication protocols or deniable key exchange protocols which realize the concurrent deniability depend on the strong assumptions/primitives, such as KEA assumption, public random

Table 1: Comparisons of deniable key agreement protocols.

Scheme	Scale	Round	Concurrency	RO	Deniability realization
[3]	group	4	×	√	Rewinding
[26]	group	3	×	-	Rewinding
[10]	2-party	2	√	√	KEA assumption
[22]	2-party	2	√	√	Public RO
Proposed	group	2	√	-	Ring signature

oracle, or timed commitment/encryption to extract the witness for the simulations. Compared with them, our DGKA protocol is not restricted to these limitations.

The comparisons of the related protocols with deniability are listed in the Table 1.

6. A Privacy-Preserving Communication Protocol for VANETs

In this section, we design a privacy-preserving communication protocol for VANETs by using the proposed deniable group key agreement protocol. Our protocol guarantees the secure communication between vehicles and vehicles and vehicles and roadside unit. VANETs are composed of Trusted Authority (TA), roadside unit (RSU, which is the infrastructure), and On-board Units (OBUs, with which vehicles are equipped). Our security model for privacy-preserving VANETs is as follows.

(i) Authentication: in the VANETs environment, RSU and OBUs should ensure that only legitimate (certificated by TA) vehicles can join this networks. Similarly, RSU should be also authenticated by vehicles in order to prevent pseudo base stations.

(ii) Anonymity: OBUs receive the information without knowing the sender identity, but only to confirm that this message is from an authenticated group.

(iii) Privacy: the conversations among the OBUs do not leave any paper trail. This "off-the-record" property prevents the shared information from being maliciously used.

(iv) Secrecy: during the process of communication, the sent messages are only known to receivers but are random to any third parties.

Our privacy-preserving communication protocol for VANETs is mainly divided into three steps. The first step is to initialize a group of VANETs. Then, OBUs and RSU in this group authenticate mutually to generate a session key. Finally, they communicate with each other with this session key under an authenticated and privacy-preserving environment.

Let U_i be one of vehicles and U_R be RSU. $(\mathrm{SK}_i, \mathrm{PK}_i) = (s_i, g^{s_i})$ denote the private/public key pair for vehicle U_i; $(\mathrm{SK}_0, \mathrm{PK}_0) = (s_0, g^{s_0})$ denote the private/public key pair for U_R. $H : \{0,1\}^* \to \{0,1\}^l$, where l is the length of a message. A detailed protocol is given as follows.

Initialization Step. The members of a group of VANETs are decided.

(i) U_R randomly chooses id as the session ID and forms a group $\mathscr{R}$ by using its public key PK_0 and the public keys of adjacent vehicles $\{\mathrm{PK}_i\}$. Finally, broadcast message $V_{\mathrm{init}} = \mathrm{id} \parallel \mathscr{R}$.

Authentication Step. The identities of members are authenticated.

(i) *Round 1 (OBUs and RSU)*. Choose x_i, t_i, and compute $X_i = g^{x_i}$, $T_i = g^{t_i}$. Broadcast message $V^1_{\mathrm{auth}} = \mathrm{id} \parallel \mathrm{PK}_i \parallel X_i \parallel T_i$.

(ii) *Round 2 (OBUs and RSU)*. Compute Y_i and σ_i as the proposed DGKA protocol (described in Algorithm 2). Broadcast message $V^2_{\mathrm{auth}} = \mathrm{id} \parallel Y_i \parallel \sigma_i$.

(iii) *Key Generation (OBUs and RSU)*. Authenticate the identities of other members and get the session key sk as in Algorithm 2.

Communication Step. With this session key sk, all the members in this group $\mathscr{R}$ can communicate securely. There are two cases in this step, including broadcast from U_R or U_i to all members and communication from U_R to U_i, U_i to U_R, or U_i to U_j.

(i) Broadcast (one-to-many):

(a) RSU or OBUs send message m_b: compute $\nu = H(\mathrm{id}, \mathrm{sk})$ and $e = m_b \oplus \nu$. Broadcast message $V_{\mathrm{bro}} = \mathrm{id} \parallel e$.

(b) RSU and OBUs recover m_b: compute $\nu = H(\mathrm{id}, \mathrm{sk})$ and $m_b = e \oplus \nu$.

(ii) Communication (one-to-one):

(a) RSU or OBUs send m_c to U_i: choose $r \leftarrow Z_q^*$ and compute $R = g^r$, $\nu = H(\mathrm{id}, R, \mathrm{PK}_i{}^r, \mathrm{sk})$, and $e = m_c \oplus \nu$. Broadcast message $V_{\mathrm{com}} = \mathrm{id} \parallel e \parallel \mathrm{PK}_i \parallel R$.

(b) U_i recovers m_c: compute $\nu = H(\mathrm{id}, R, R^{s_i}, \mathrm{sk})$ and $m_c = e \oplus \nu$.

By employing the proposed DGKA protocol, each receiver can identify the source of the received information without knowing the actual sender by using the session key sk. Moreover, this session key sk can be simulated by anyone; the vehicles involved in the above communication can deny this. There is no paper trail; thus the vehicle privacy is protected.

7. Conclusions

This paper presents a 2-round fully deniable group key agreement protocol. We provide a novel approach to transfer an unauthenticated GKA to a deniable GKA without increasing round. The transcript simulation does not require the rewinding steps; thus our deniability also holds even in the concurrent setting. We also design a privacy-preserving communication protocol for VANETs using the proposed DGKA protocol.

Notations in DGKA Protocol

κ: The security parameter
G: A multiplicative group of prime order q
g: The generator of group G
U_i: The ith participant
PK_i: U_i's public key
SK_i: U_i's private key
$\pi_i^{l_i}$: A session of U_i called an instance—a participant may have many instances and denotes the instance l_i of U_i as $\pi_i^{l_i}$
$\mathrm{sid}_i^{l_i}$: The session ID of instance $\pi_i^{l_i}$
$\mathrm{pid}_i^{l_i}$: A set containing the identities of the participants in the group with whom $\pi_i^{l_i}$ intends to establish a session key, including U_i
$\mathrm{stat}_i^{l_i}$: The current state of instance $\pi_i^{l_i}$
$\mathrm{sk}_i^{l_i}$: The common key generated by instance $\pi_i^{l_i}$ after the protocol finished
$\mathrm{negl}(\kappa)$: A negligible function for the security parameter κ.

Conflicts of Interest

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

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61402376, U1433130), the Ministry of Education "Chunhui Plan" (Z2016150), and the National Key R&D Program of China (2017YFB0802300, 2017YFB0802000).

References

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

[2] I. Ingemarsson, D. T. Tang, and C. K. Wong, "A conference key distribution system," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 28, no. 5, pp. 714–720, 1982.

[3] J. Bohli and R. Steinwandt, "Deniable Group Key Agreement," in *Progress in Cryptology - VIETCRYPT 2006*, vol. 4341 of *Lecture Notes in Computer Science*, pp. 298–311, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[4] D. Dolev, C. Dwork, and M. Naor, "Nonmalleable cryptography," *SIAM Journal on Computing*, vol. 30, no. 2, pp. 391–437, 2000.

[5] C. Dwork, M. Naor, and A. Sahai, "Concurrent zero-knowledge," *Journal of the ACM*, vol. 51, no. 6, pp. 851–898, 2004.

[6] M. Di Raimondo and R. Gennaro, "New approaches for deniable authentication," *Journal of Cryptology*, vol. 22, no. 4, pp. 572–615, 2009.

[7] M. Di Raimondo, R. Gennaro, and H. Krawczyk, "Deniable authentication and key exchange," in *Proceedings of the CCS 2006: 13th ACM Conference on Computer and Communications Security*, pp. 400–409, Alexandria, Va, USA, November 2006.

[8] M. Bellare and A. Palacio, "Towards plaintext-aware public-key encryption without random oracles," in *Advances in cryptology—ASIACRYPT 2004*, vol. 3329 of *Lecture Notes in Comput. Sci.*, pp. 48–62, Springer, Berlin, 2004.

[9] S. Jiang, "Deniable authentication on the internet (extended abstract)," in *Information security and cryptology*, vol. 4990 of *Lecture Notes in Comput. Sci.*, pp. 298–312, Springer, Berlin, 2008.

[10] A. C.-C. Yao and Y. Zhao, "Privacy-preserving authenticated key-exchange over internet," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 1, pp. 125–140, 2014.

[11] A. C.-C. Yao and Y. Zhao, "OAKE: A new family of implicitly authenticated Diffie-Hellman protocols," in *Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security, CCS 2013*, pp. 1113–1128, Germany, November 2013.

[12] H. Tian, X. Chen, and W. Susilo, "Deniability and forward secrecy of one-round authenticated key exchange," *The Journal of Supercomputing*, vol. 67, no. 3, pp. 671–690, 2014.

[13] S. Zeng, Y. Chen, S. Tan, and M. He, "Concurrently deniable ring authentication and its application to LBS in VANETs," *Peer-to-Peer Networking and Applications*, vol. 10, no. 4, pp. 844–856, 2017.

[14] S. Jiang, "Timed encryption with application to deniable key exchange," *Theoretical Computer Science*, vol. 560, no. part 2, pp. 172–189, 2014.

[15] L. Harn and J. Ren, "Design of fully deniable authentication service for E-mail applications," *IEEE Communications Letters*, vol. 12, no. 3, pp. 219–221, 2008.

[16] W. Susilo and Y. Mu, "Non-interactive deniable ring authentication," in *Information security and cryptology—ICISC 2003*, vol. 2971 of *Lecture Notes in Comput. Sci.*, pp. 386–401, Springer, Berlin, 2004.

[17] W. Diffie, W. Diffie, and M. E. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory*, vol. 22, no. 6, pp. 644–654, 1976.

[18] W. Mao and K. G. Paterson, "On the plausible deniability feature of internet protocols," Tech. Rep., 2002.

[19] R. Canetti and H. Krawczyk, "Analysis of key-exchange protocols and their use for building secure channels," in *Advances in cryptology—EUROCRYPT 2001 (Innsbruck)*, vol. 2045 of *Lecture Notes in Comput. Sci.*, pp. 453–474, Springer, Berlin, 2001.

[20] C. Boyd, W. Mao, and K. G. Paterson, "Deniable authenticated key establishment for Internet Protocols," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 3364, pp. 255–271, 2005.

[21] M.-H. Lim, S. Lee, Y. Park, and S. Moon, "Secure deniable authenticated key establishment for Internet protocols," in *Proceedings of the 2nd International Conference on Information Security and Assurance, ISA 2008*, pp. 3–6, Republic of Korea, April 2008.

[22] S. Jiang and R. Safavi-Naini, "An efficient deniable key exchange protocol, in 12th International Conference on Financial Cryptography and Data Security," in *Proceedings of the 12th International Conference on Financial Cryptography and Data Security*, vol. 2008, pp. 47–52.

[23] D. R. Brown, "Deniable authentication with RSA and multicasting," *IACR Cryptology*.

[24] J. Katz and J. S. Shin, "Modeling insider attacks on group key-exchange protocols," in *Proceedings of the CCS 2005 - 12th ACM Conference on Computer and Communications Security*, pp. 180–189, USA, November 2005.

[25] C.-P. Schnorr, "Efficient identification and signatures for smart cards," in *Advances in cryptology*, vol. 435 of *Lecture Notes in Comput. Sci.*, pp. 239–252, Springer, New York, 1990.

[26] Y. Zhang, K. Wang, and B. Li, "A deniable group key establishment protocol in the standard model," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 6047, pp. 308–323, 2010.

[27] Y.-Z. Zhang, H.-X. Xu, and B. Li, "Generic construction of deniable group key establishment from group key establishment," *Tongxin Xuebao/Journal on Communication*, vol. 32, no. 3, pp. 143–149, 2011.

[28] K. Neupane, R. Steinwandt, and A. Suárez Corona, "Scalable deniable group key establishment," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 7743, pp. 365–373, 2013.

[29] Y. Chen, L. Lu, S. Zeng, and M. He, "Deniable authentication for multiparty protocol," *Chinese Journal of Network and Information Security*, vol. 2, no. 6, pp. 71–78, 2016.

[30] N. Zhang, P. Yang, S. Zhang et al., "Software defined networking enabled wireless network virtualization: Challenges and solutions," *IEEE Network*, vol. 31, no. 5, pp. 42–49, 2017.

[31] D. Chen, N. Zhang, and Z. Qin, "S2M: a lightweight acoustic fingerprints based wireless device authentication protocol," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 88–100, 2017.

[32] Q. Jiang, S. Zeadally, J. Ma, and D. He, "Lightweight three-factor authentication and key agreement protocol for internet-integrated wireless sensor networks," *IEEE Access*, vol. 5, pp. 3376–3392, 2017.

[33] Q. Wang, D. Chen, and N. Zhang, "LACS: A Lightweight Label-Based Access Control Scheme in IoT-Based 5G Caching Context," *IEEE Access*, 4027 pages, 2017.

[34] S. Zeng, M. He, M. Xia, X. Li, Y. Chen, and S. Tan, "Privacy-preserving location-based service based on deniable authentication," in *Proceedings of the 9th IEEE/ACM International Conference on Utility and Cloud Computing, UCC 2016*, pp. 276–281, chn, December 2016.

[35] J.-L. Huang, L.-Y. Yeh, and H.-Y. Chien, "ABAKA: An anonymous batch authenticated and key agreement scheme for value-added services in vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 248–262, 2011.

[36] C.-Y. Chow, M. F. Mokbel, and X. Liu, "A peer-to-peer spatial cloaking algorithm for anonymous location-based service," in *Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS '06)*, pp. 171–178, ACM, November 2006.

[37] R. Dutta and R. Barua, "Provably secure constant round contributory group key agreement in dynamic setting," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 54, no. 5, pp. 2007–2025, 2008.

[38] M. Burmester and Y. Desmedt, "A secure and efficient conference key distribution system," in *Advances in Cryptology — EUROCRYPT'94*, vol. 950 of *Lecture Notes in Computer Science*, pp. 275–286, Springer Berlin Heidelberg, Berlin, Heidelberg, 1995.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 7635143, 13 pages
https://doi.org/10.1155/2018/7635143

Research Article

Routing Protocol in VANETs Equipped with Directional Antennas: Topology-Based Neighbor Discovery and Routing Analysis

Haipeng Li and Zhen Xu

School of Electronic and Information Engineering, Beihang University, Beijing, China

Correspondence should be addressed to Zhen Xu; xuzhen@buaa.edu.cn

Received 14 December 2017; Accepted 11 March 2018; Published 10 April 2018

Academic Editor: Leyre Azpilicueta

In Vehicular Ad Hoc Networks (VANETs), directional antenna is a good solution if a longer transmission distance is needed. When vehicles are equipped with directional antennas, however, complete paths from the sources to the destinations there may not exist. Epidemic routing protocol is considered as one of the well-performed routing protocols when the networks are intermittently connected but it can cause a heavy load to the network and great energy consumption to the nodes. First in this paper, we propose a novel neighbor discovery algorithm which makes nodes be able to sense the topology changes around them and arrange their directional antennas accordingly. Secondly, we propose a routing protocol which is based on the conventional epidemic routing protocol, and nodes make their routing decisions according to the information collected during the neighbor discovery process. Experimental results show that the proposed neighbor discovery algorithm has better performance especially in the scenario where the node density is low. Moreover, the matching routing protocol can effectively reduce the load of the network and successfully deliver the packets to its destination in a reasonable short delay.

1. Introduction

Directional antennas have been widely used in wireless ad hoc networks recently. It can produce higher gain, provide greater transmission range, and improve the network spatial reuse and throughput as well. Furthermore, the directional selectivity also reduces the co-channel interference of neighboring nodes, and directional antennas bring potential performance improvement to the mobile ad hoc networks. Despite the improvements brought by the directional antennas, their deployments also bring severe challenges to the nodes such as the neighbor discovery, and the network suffers from frequent disruptions, making the topology unstable, that is, no consistent paths between each pair of nodes.

Considering that, in a considerable proportion of real scenarios, the mobility of vehicles is not purely random. Vehicles show strong location and group preference. For example, for the vehicles in the earthquake relief sites, they work at the areas which are under serious destruction and these areas are not uniformly distributed. Moreover, vehicles can change their locations from one area to another. Based on this kind of scenario, we formulate the neighbor discovery algorithm and the matching routing protocol in this paper. The main contribution of this paper is about two aspects. Firstly, we design a novel neighbor discovery algorithm for intermittent networks which are equipped with directional antennas. In the proposed neighbor discovery algorithm, nodes monitor the HELLO message receiving frequency for each sector to estimate the topology variations and adjust the steering time of each sector according to the history records maintained in their neighbor node lists. Secondly, we propose a routing protocol which is derived from the proposed neighbor discovery algorithm. In the routing protocol, source node makes its routing decisions according to the status of the nodes in its two hops. The stability of neighbor relationships, the communicating probability, and the time schedule for the

directional antennas of each participating node are all taken into consideration when the source node makes its routing decisions.

The rest of the paper is organized as follows. In Section 2, we introduce the related researches about neighbor discovery algorithms and routing protocols for opportunistic networks. In Section 3, we present the system models. In Section 4, we introduce and analyze our proposed neighbor discovery algorithm. Our proposed routing protocol is demonstrated in Section 5. Section 6 provides simulation results. Finally, we conclude the paper in Section 7.

2. Related Work

2.1. Neighbor Discovery. As we discussed in Section 1, neighbor discovery is the key initial step for establishing the connections among nodes in intermittent networks. In recent research papers about neighbor discovery protocols with pure directional antennas, we categorize them into two classes, *deterministic* and *random* protocols. In deterministic protocols [1–3], nodes steer their directional antennas in preset sequences. Deterministic protocols can guarantee the bounded neighbor discovery delay. However, most of them need time synchronized which is not practical in some applications. In random protocols [4–10], nodes randomly select a direction to transmit or receive HELLO messages. Compared to deterministic protocols, the most obvious advantage of probabilistic protocols is that they do not need to be synchronized, in which they are more robust and adaptive to complex scenarios. However, in most probabilistic or random protocols, the topology change of the network is not concerned. Reference [1] focuses on the oblivious neighbor discovery problems and has proposed an oblivious discovery protocol which achieves guaranteed oblivious discovery with order-minimal worst-case discovery delay. However, the topology change is not its concern. References [2–10] focus on the MAC protocols to achieve a better neighbor discovery performance. References [2–8] all need time synchronized to ensure their performance and their directional antennas just simply choose a direction when they run the neighbor discovery process. References [11, 12] take the topology into consideration. Reference [11] works on topology control problems. In this work, nodes track only a subset of their discovered neighbor nodes and it proposes a scheme called Di-ATC which tries to minimize the variance of the angular separation between the tracked neighbors of a node. In [12], nodes discover their neighbor nodes in one sector with increasing power so as to identify the closest nodes in the sector. However, [11] assumes that the network is connected at all times and the path stretch after topology control is low and [12] is only suitable for a static ad hoc network which is not realistic in VANETs with directional antennas equipped.

2.2. Routing Protocols. Since there are no complete paths in intermittent vehicle ad hoc networks with directional antennas, traditional end-to-end routing protocols for mobile ad hoc network (e.g., DSR, AODV) are not suitable anymore. The existing routing protocols that are concerned with intermittent networks mostly focus on the load easement while the successful transmission is guaranteed. In [13, 14], nodes only generate one copy of the data packets to reduce the load of network. In methods [15–18], nodes selectively send the copies of packets to their neighbor nodes. In these *replication-based* methods, nodes carefully make the routing decisions according to several rules such as *utility-based* routing [15] or *probability-based* routing [17]. They assume that movement patterns of node are not purely random and that future contacts depend on the past information. Reference [16] proposes a method called *Spray and Focus* which refers to the fact that nodes firstly spray a fixed number of their data packets to the selected relay nodes and then nodes with only one copy of the packet can only forward this message further using a *single-copy utility-based scheme*. In [18], they investigate the TTL on the copy of packets in order to reduce the load of the networks. Due to the dynamic topology of VANETs, the size of HELLO packets and the mobility should be taken into consideration for the routing decisions. Reference [19] investigates the MAC layer contention issue that it proposes BBNC, a backbone based routing protocol with interflow Network Coding. The position information of each backbone node is critical when nodes in [19] make their routing decisions. Reference [20] focuses on the effect of packet size on the performance of routing protocols; it uses a fuzzy logic-based algorithm to select relay nodes and uses a Q-learning based approach to tune the fuzzy membership functions. For the VANETs where nodes are equipped with directional antennas, the dynamic topology is mostly caused by the scan pattern of the directional antennas, and this is the emphasis of this paper. For the routing protocols which are focusing on the directional antenna, such as in [21], nodes make their routing decisions based on the information about the geographical position of the participating nodes, which needs the use of GPS. Reference [22] focuses on analyzing the relationship between the throughput and the beam width of the directional antenna. In general, they rarely consider the switch pattern of the directional antenna.

3. System Model and Problem Formulation

3.1. System Model

3.1.1. Directional Antenna Model. We assume that the set of nodes in the network is denoted by $\check{N}$. Each node $i \in \check{N}$ is equipped with only one set of directional antennas. For node i, we approximate the directional antenna as a circular sector of angle θ_i ($0 < \theta_i \leq 2\pi$), and there are $N_i \triangleq 2\pi/\theta_i$ sectors around it, indexed clockwise from 0 to $N_i - 1$. When $\theta_i = 2\pi$, the directional antenna degenerates to an omnidirectional antenna. The set of sectors of node i is denoted by $\widehat{S}_i$; $S_i \in \widehat{S}_i$ maintains a timer C_{S_i}. When C_{S_i} expires, the rotation mode of the directional antenna is decided by the status of the node:

(i) Pure neighbor discovery mode (PND-M): if there is no transmission task for the node, then it switches its directional antenna randomly to the next sector.

(ii) Data transmission mode (DT-M): if there are data packets conserved in the node and waiting to be transmitted, then the node begins to switch its directional

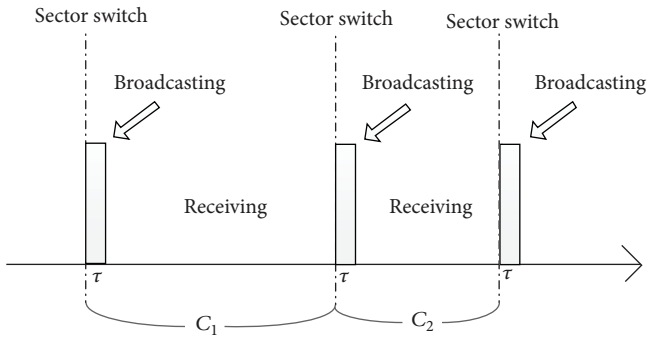


Figure 1: Communication model.

antenna clockwise. In this mode, the scan pattern of each node is ordered and predictable so that the routing decisions are more precise.

The initial C_{S_i} is equal for every sector and the rotation period is a constant number, T, where $T = \sum_{S_i \in \widehat{S}_i} C_{S_i}$. The cover range of each node is R.

3.1.2. Communication Model. We assume that the time unit is τ, which refers to the maximal time that a HELLO message needs to be broadcasted to its neighbor successfully. As shown in Figure 1, when node i points its directional antenna to sector S_i, it broadcasts a HELLO message and then it enters the receiving state until it switches its direction antenna to the next sector. When node i receives a HELLO message from node j, then node j is a discovered neighbor node for node i. Only when two nodes point their directional antennas to the opposite direction and they are located in the transmission range of each other, the HELLO message can be received successfully. Moreover, when a node receives a HELLO message from another node, it will send back a response message conditionally. This mechanism is detailed in Section 4.

3.1.3. Mobility Model. In this paper, groups of nodes are scattered randomly in a square area at first, and a community is formed around the central position of each group. Then, nodes begin to walk inside their own community randomly. We assume that a node can walk beyond its current community to another one with a very small probability, P. The scenario is shown in Figure 2.

3.1.4. Problem Formulation. As shown in Figure 2, the relative positions among the vehicle nodes are stable since they show strong location and group preference. For example, for node 0, a great number of nodes walk on its north side and its south side, in which case it is unnecessary for node 0 to point its directional antenna to the east side and west side for a long time if it wants to discover more neighbor nodes in a short time. Moreover, as we mentioned, nodes can change their communities so that the scan pattern of node 0 can be out of date sometimes, in which case node 0 needs to adjust the scan pattern of its directional antenna to the updated topology around itself.

In order to successfully deliver the packets in an intermittent network, the source node has to send plenty of data packet copies to the neighbors that it meets. However, it can cause heavy load of the network. Otherwise, if the source node simply chooses one of its neighbor nodes as the relay node, it cannot guarantee the successful delivery to the destination and leads to long delay. Therefore, the routing decision determines the transmission delay and the load and the energy consumption of the network.

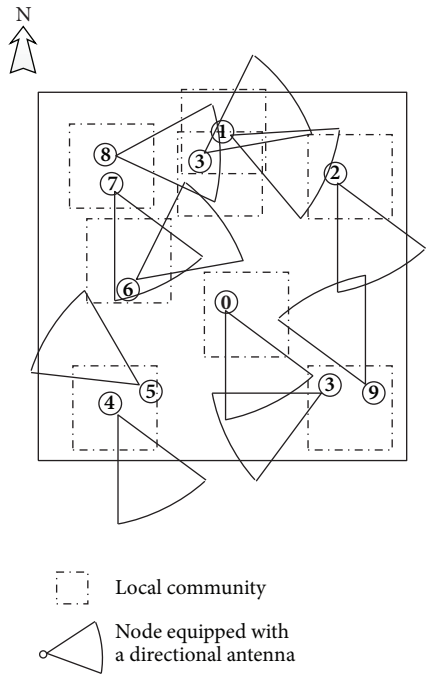


Figure 2: Mobility model.

For example, for node 8 in Figure 2, nodes 1, 3, 6, and 7 are the neighbor nodes maintained in its neighbor node list. When node 8 wants to send data packets to node 2, it needs to determine which neighbor nodes should be chosen to relay these packets. The answers of three questions about the neighbor nodes are critical when node 8 makes its routing decisions.

For a specific possible relay node:

(1) What is the probability that node 8 can communicate with it?

(2) If node 8 sends its data packets to this relay node, what is the probability that it can relay these data packets to node 2 in one hop?

(3) What is the expected delay that it can successfully transmit these data packets to node 2?

Here, we can formulate the research objective of this paper:

> *To make the vehicle be able to adjust the scanning pattern of its directional antenna according to the topology around it and find out the optimal routing decisions to minimize the delay and the number of packet copies while successful packets delivery and short delay are guaranteed.*

4. Neighbor Discovery

In this section, we devise a novel neighbor discovery algorithm with directional antennas which can make the nodes respond quickly to the topology changes. A priori knowledge is unnecessary when the nodes begin to detect their neighbor nodes.

To clearly demonstrate the protocol we devised, we assume that node j is within the cover range of node i and node k is within the cover range of node j while out of the cover range of node i.

4.1. Neighbor Node List. Neighbor node list is the fundamental part of our approach; sector distribution and routing decision are both based upon the information from the neighbor node list.

In our approach, nodes arrange their sectors according to the information in the neighbor node list. The neighbor node list of each node is built up and updated based on the interactive messages that it receives each time. The structure of the interactive message is shown in Figure 3:

(1) ID: the ID of node i;

(2) Type: to indicate that the message is either a HELLO message or a response message;

(3) Sector_S: to record the sector that the antenna of node i points to when node i broadcasts this message; it is denoted by S_i ($S_i \in \widehat{S}_i$);

(4) TNS (time to next sector): to set the timer C_{S_i} when it broadcasts this message.

The *Neighbor List* records all the neighbor nodes that node i has ever detected:

(1) Neighbor ID: to indicate the ID of the neighbor node, such as the ID of node j;

(2) DES_N (duration for every sector of the neighbor node): to record all the sectors where node i has ever detected node j; the total number of messages received from node j is denoted by $A_{ij}^{\widehat{S}_i}$;

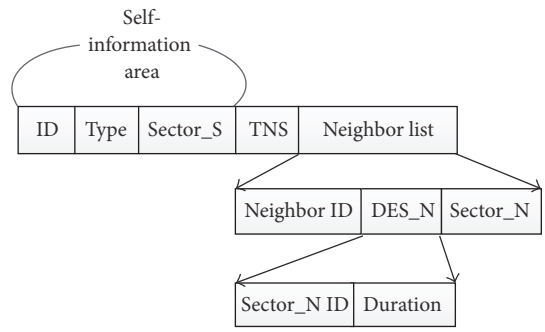


Figure 3: Structure of interactive message.

(3) Sector_N (sector of the neighbor node for the last communication): to record the sector that the directional antenna of node i points to when node i detected node j last time, which is denoted by S_{ij}.

Every node maintains a neighbor node list. The format of each item in the neighbor node list is shown in Figure 4. The *Neighbor-Information Area* of node i records the information about node j. The *Two-Hop Node List* records all the information of the neighbor nodes that node j has ever detected, such as the information about node k. Their formats are the same as the HELLO message and their descriptions are similar as well, except that the *Time_Flag* indicates the update time of this item, which is updated based on the timer kept in node i since nodes are not time synchronized. It is denoted by T_{Flag}^{i}.

In order to improve the efficiency of neighbor discovery, nodes will send back a response message when they receive a HELLO message so that they can take every chance to make themselves discovered by others. However, collision issue is not concerned in this paper since nodes are not time synchronized like in [2–8]. We propose a mechanism to reduce the collision probability, which is that nodes are not forced to send back a response message every time when it receives a HELLO message. When a node, say j, receives a HELLO message from another node, say i, there are two steps to be taken:

(1) Node j updates its neighbor node list according to the HELLO message from node i. The *Neighbor-Information Area* is updated according to the *Self-Information Area* of the HELLO message and the *Two-Hop Node List* is updated according to the *Neighbor List* of the HELLO message.

(2) If node i is listed in the *neighbor node list* of node j and the information about itself is the latest, node j does not respond to node i. If not, node j sends back a response packet to node i after a random time period ε whether it is in the transmitting state or not. ε is supposed to be smaller than the *TNS* in the HELLO message from node i in case that node i turns its directional antenna to the next sector and it cannot receive the response message from node j.

As HELLO messages and response messages being broadcasted in the network, nodes are able to collect plenty of information to arrange their directional antennas.

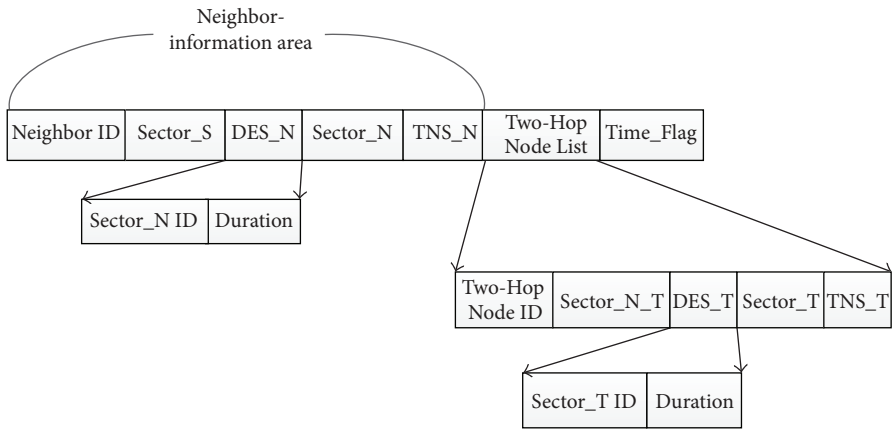


Figure 4: Neighbor node list.

4.2. Sector Steering Schedule Arrangement. In this part, we present how nodes adjust the steering time of their sectors based upon the neighbor node list.

As we mentioned in Section 3, when the mobility model of the nodes in the network shows strong location and group preference, nodes can waste time on the sectors where few neighbor nodes exist if they switch their directional antennas randomly. To make the nodes sensitive to the topology around them and arrange their directional antennas intelligently, we propose an algorithm called *History-Based Sector Distribution (HSD)*. At the beginning of *HSD*, node i switches its directional antenna to sectors randomly and the timer C_{S_i} ($S_i \in \widehat{S}_i$) is equal to $T\theta_i/2\pi$. As the neighbor discovery proceeds, node i will gather enough information to arrange its directional antennas and C_{S_i} ($S_i \in \widehat{S}_i$) is going to be adjusted. According to the *DES_N* in the neighbor node list maintained in node i, we can know which sectors communicate with other nodes frequently. They are the areas where groups of nodes may exist and are more likely for node i to discover new neighbor nodes.

Then, we can formulate the *HSD* as follows. When node i points its directional antenna to S_i, we can get that

$$R_{S_i} = \frac{\sum_{k\in L_i} A_{ik}^{S_i}}{\sum_{S\in\widehat{S}_i}\sum_{k\in L_i} A_{ik}^{S}}, \tag{1}$$

where R_{S_i} represents the proportion of interactive messages received through S_i compared to the total messages received through all sectors of node i and L_i represents the set of neighbor nodes recorded in the neighbor node list of node i. We define $Q_{S_i} \triangleq \sum_{k\in L_i} A_{ik}^{S_i}$ and $E_i \triangleq \sum_{S\in\widehat{S}_i}\sum_{k\in L_i} A_{ik}^{S}$. The relationship between R_{S_i} and C_{S_i} is expressed as follows:

$$C_{S_i} = R_{S_i}T = \frac{Q_{S_i}}{E_i}T. \tag{2}$$

In the preliminary stage, Q_{S_i} and E_i in (2) are both very small and the sector which is the first to discover neighbor nodes tends to have a bigger R_{S_i} since Q_{S_i} and E_i can be all generated by this sector, in which case the node possibly points its antenna to this sector for an inappropriate C_{S_i} which is close to T. With the process of neighbor discovery, according to (1) and (2), C_{S_i} of each sector would be too stable to make a response to the topology changes. For example, C_{S_i} of the sectors where few neighbor nodes are discovered tend to be small or even close to zero. However, nodes can change their communities with a small probability as we mentioned in Section 3. Moreover, the relative positions among nodes can change dramatically sometime even if they walk in the same community, such as when nodes walk to an opposite direction. Therefore, neighbor nodes can appear in the sectors which used to be empty and the previous C_{S_i} of these sectors would be obsolete.

To make (2) sensitive to the topology of the network, in *HSD*, we propose a mechanism to estimate the topology changes by monitoring the frequency of message receiving.

Firstly, in a specific monitoring time period, M_i, then, we can get

$$V_{S_i} = \frac{Q_{S_i}}{M}, \tag{3}$$

where V_{S_i} indicates the message receiving frequency of S_i in the recent time period of M_i. Obviously, V_{S_i} is a stable value when the relative positions among the nodes are stable. In that case, the instant receiving frequency, V'_{S_i}, is stable as well and it holds that

$$V'_{S_i} = \frac{\Delta Q_{S_i}}{\Delta t} \cong V_{S_i}. \tag{4}$$

According to (1) and (3), we notice that

$$\frac{V_{S_i}}{R_{S_i}} = \frac{E_i}{M_i}. \tag{5}$$

Therefore, V_{S_i}/R_{S_i} is only determined by the status of node i but not the status of S_i. Due to this, we define $\varphi_i \triangleq E_i/M_i$

which is the message receiving frequency of node i regardless of the status of its sectors.

According to (4) and (5), when the topology in the network is stable, the following is established between every two sectors of node i:

$$\frac{V'_{S_i^m}}{R_{S_i^m}} \cong \frac{V'_{S_i^n}}{R_{S_i^n}} \cong \frac{E_i}{M} \quad \left(S_i^m, S_i^n \in \widehat{S}_i;\ m \neq n\right). \tag{6}$$

Therefore, we can compare V'_{S_i} among $\widehat{S}_i$ to estimate the topology variation around node i based on (6). The functioning process of (6) is analyzed as follows.

In the preliminary stage of the neighbor discovery process, according to (6), the adjustment of R_{S_i} would be started after the first monitoring time period. And node i is supposed to scan more than two sectors during the first monitoring time period to apply (6). Hence, the initial C_{S_i} would be restricted within a reasonable range due to the comparison of C_{S_i} among sectors. After that, when the topology around the node i changes, there are two situations:

(i) The relative positions among nodes are changed while the neighbor nodes are still in the transmission range of node i; that is, the number of neighbor nodes within the transmission range of node i remains the same;

(ii) Some neighbor nodes are no longer in the transmission range of node i or new neighbor nodes come into the transmission range; that is, the number of neighbor nodes within the transmission range of node i is changed.

The simulation result of [1] shows that in some cases nodes cannot discover each other even though they are within the transmission range of each other due to the equipment of directional antenna, while, in general, more neighbor nodes within the transmission range of node i can lead to more frequent communications. Thus, we assume that φ_i is only determined by the number of nodes within the transmission range of node i.

For the first situation, for example, some nodes move from sector S_i^m to sector S_i^n. Obviously, $V'_{S_i^m}$ will decrease while $V'_{S_i^n}$ will increase. Then we can readjust $R_{S_i^m}$ and $R_{S_i^n}$ according to (6) since φ_i is stable.

For the second situation, φ_i is going to change as well, and we need to rewrite (6) to the following form:

$$R_{S_i} \cong \frac{V'_{S_i}}{\varphi_i}. \tag{7}$$

When the second situation happens, to clearly figure out the variation of R_{S_i}, we separately analyze the variation of V'_{S_i} and φ_i as follows:

$$\begin{aligned} \Delta V'_{S_i} &= \frac{\Delta Q_{S_i}}{\Delta t} - \left(V'_{S_i}\right)_{\text{before}}, \\ \Delta\varphi_i &= \frac{E_i + \Delta E_i}{M + \Delta t} - (\varphi_i)_{\text{before}}; \end{aligned} \tag{8}$$

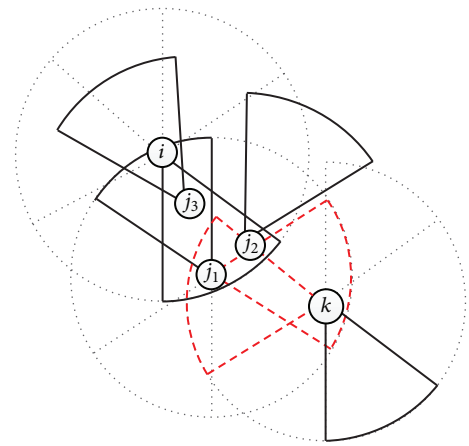


Figure 5: Relay node selection.

then,

$$\begin{aligned} \Delta V'_{S_i} - \Delta\varphi_i = \frac{\Delta Q_{S_i}}{\Delta t} - \frac{E_i + \Delta E_i}{M + \Delta t} + (\varphi_i)_{\text{before}} \\ - \left(V'_{S_i}\right)_{\text{before}}. \end{aligned} \tag{9}$$

Simplifying (9) and letting

$$\begin{aligned} \frac{\Delta Q_{S_i} - \Delta E_i}{M} &\longrightarrow 0, \\ \frac{\Delta t}{M} &\longrightarrow 0 \end{aligned} \tag{10}$$

yield

$$\Delta V'_{S_i} - \Delta\varphi_i \cong V'_{S_i} - \left(V'_{S_i}\right)_{\text{before}}. \tag{11}$$

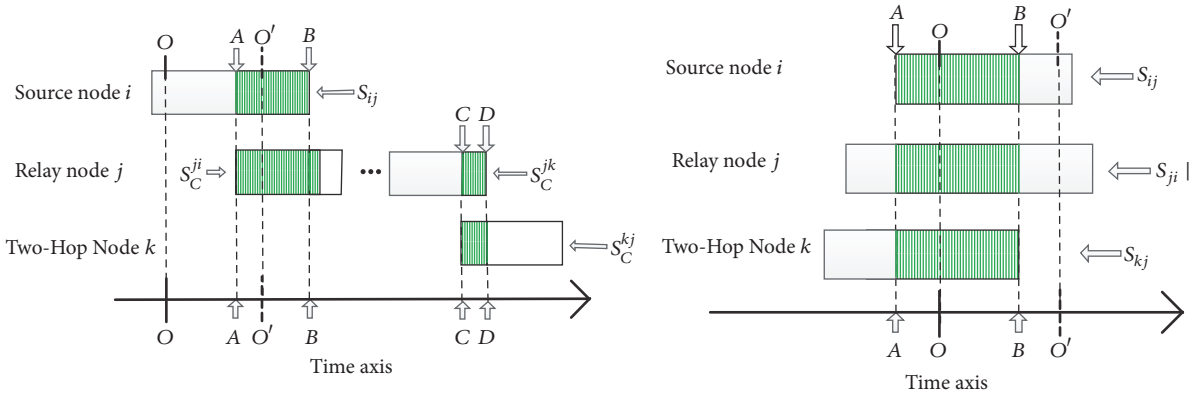
According to (11), when new neighbor nodes come into the transmission range of node i such as appearing in sector S_i, it is obvious that V'_{S_i} would be bigger than $(V'_{S_i})_{\text{before}}$ so that $\Delta V'_{S_i} > \Delta\varphi_i$. According to (7), R_{S_i} would increase. Contrarily, when nodes flee from S_i, V'_{S_i} would be smaller than $(V'_{S_i})_{\text{before}}$ so that R_{S_i} would decrease. By far we can conclude that (6) works well for dynamic topology.

5. Routing Protocol

In the network where nodes are equipped with directional antennas, there is no consistent path from the source node to the destination node. In this section, we devise a special routing protocol called *HSD-R* which is formulated based upon HSD. The routing decisions are made based on the knowledge maintained in the neighbor node list.

5.1. Neighbor Node Table Processing. We take Figure 5 as an example to demonstrate how we analyze the neighbor node table when there are data packets waiting to be transmitted.

As illustrated in Figure 5, when node i plans to transmit its data packets to node k, it can send its packets to nodes j_1 and



Duration of the communicating sector S_c

Able-transmitting time period

O, O' Zero point (beginning point of data transmission)

(a) Node j should wait for a period time until it can communicate with node k

(b) Node j relays packets to node k immediately

Figure 6

j_3 immediately. But node j_3 is not qualified to be a relay node since node k is beyond its transmission distance. However, for node j_1, we need to take a further consideration because the connecting time between it and node i could be too short to finish a complete data transmission when they point their directional antennas to each other. Moreover, when data packets arrive at node j_1, the time consumption for node j_1 and node k to point their directional antennas to a proper direction also needs to be considered since the data packets are conserved in node j_1 during that time and cannot be transmitted.

In order to make an appropriate routing decision, we need to find out which nodes can relay the data packets to the destination successfully with a high probability. When node i wants to communicate with node j through sector S_i, the communicating probability can be expressed as follows:

$$P_{ij}^{S_i} = \frac{A_{ij}^{S_i}}{(M/(n+1)\tau) R_{S_i}}, \tag{12}$$

where $P_{ij}^{S_i}$ refers to the communication probability between node i and node j through S_i. Furthermore, we need to make sure that S_i is the best option for node i to communicate with node j. For example, in the most recent monitoring time period, node i and node j have a frequent communication in S_i but this kind of situation could be a result of that node i and node j just happen to detect each in the recent M while they seldom encountered with each other before. In other words, the neighbor relationship between node i and node j existing in S_i is only during the most recent monitoring time period. For this kind of neighbor nodes, they are not good options for relaying data packets due to the weak neighbor relationship between them and the source node. Therefore, we introduce a coefficient, σ_{ij}, which refers to the stability of neighborhood relationship between node i and node j. It can be calculated as follows:

$$\sigma_{ij} = \frac{\sum_{S \in \widehat{S}_i} A_{ij}^S}{E_i}. \tag{13}$$

Before analyzing the time metric of every possible path, we need to detail how the directional antennas operate during the data transmission. As we mentioned in Section 3, nodes are supposed to switch their directional antennas in a predicable sequence in order to make the routing decisions more precise, while the source node can begin its data transmission at any time point and nodes are not time synchronized in this paper. Therefore, we propose a mechanism to make nodes to change their rotation pattern spontaneously. When there are data packets conserved in the nodes and waiting to be transmitted, data packets could be generated by the nodes themselves or received from other neighbors, these nodes switch their rotation mode to *DT-M*. Otherwise, they still keep their scan pattern as *PTN-M*. The broadcast of interactive messages is still going on whether the rotation mode is *DT-M* or *PTN-M* since the topology changes can also happen during the data transmissions.

Considering that nodes make their routing decisions based on the status of nodes in two-hop distance, there are at least three sets of directional antennas participating in data transmission. According to the neighbor node list maintained in the source node i, we can construct the relationships about sectors among the source node, the neighbor nodes, and the two-hop nodes. As illuminated in Figure 6, we can draw the position and length of the frame for *Source Node i* according to the *Sector Distribution* in neighbor node list; the position and length of the frames for *Relay Node j* and *Two-Hop Node k* are separately decided by the *Neighbor-Information Area* and the *Two-Hop Node List*. In Figure 6, the zero point of the

time axis, T_0, refers to the time point when the source node prepares to send its data packets. The green zone between point A and point B or between point C and point D refers to the time period when the directional antennas of two nodes cover each other; that is, the two nodes can communicate with each other during the green zone.

When the relay node receives data packets successfully, there are two possibilities: (1) the relay node waits for a period of time until it can communicate with node k, such as shown in Figure 6(a); (2) the relay node sends them to node k immediately, which means $S_{jk} = S_{kj}$, as shown in Figure 6(b).

Source node, like node i, can send data packets at any time. In other words, the position of point O can be anywhere within the duration of S_{ij}, while, in general, it can be categorized into two cases: (1) it is ahead of the able-transmitting period, like point O in Figure 6(a); (2) it is during the able-transmitting period, like point O' in Figure 6(a) and point O in Figure 6(b). If the node misses the two kinds of time positions, like point O' in Figure 6(b), it has to wait for the next communicating sector and it can be considered as point O again.

From Figure 6 we can see that, in order to figure out the relationship among the three sets of directional antennas, it is necessary to figure out the beginning time points, T^{mn}_{begin} $(m, n \in \{i, j, k\};\ m \neq n)$, and the ending time points, T^{mn}_{end} $(m, n \in \{i, j, k\}; m \neq n)$, for these sectors:

(1) S_{ji}:

$$\begin{aligned} T^{ji}_{\text{end}} &= T^{j}_{\text{Flag}} + C_{S_{ji}} \\ T^{ji}_{\text{begin}} &= T^{ji}_{\text{end}} - R_{S_{ji}} T. \end{aligned} \tag{14}$$

(2) S_{jk}:

$$T^{jk}_{\text{end}} = \begin{cases} T^{ji}_{\text{end}}, & S_{ji} = S_{jk} \\ T^{ji}_{\text{end}} + \sum_{m=1+S_{ji}}^{S_{jk}} R_m T, & S_{ji} \neq S_{jk} \end{cases} \tag{15}$$

$$T^{jk}_{\text{begin}} = T^{jk}_{\text{end}} - R_{S_{jk}} T.$$

(3) S_{kj}:

$$\begin{aligned} T^{kj}_{\text{end}} &= T^{k}_{\text{Flang}} + C_{S_{jk}} \\ T^{kj}_{\text{begin}} &= T^{kj}_{\text{end}} - R_{S_{kj}} T. \end{aligned} \tag{16}$$

Then we present an algorithm to predict the waiting and transmitting time for every possible path.

After the waiting time and able-transmitting time of each path are calculated according to Algorithm 1, the source node is able to make its routing decisions.

5.2. Routing Decisions. When the source node, say node i, wants to send its data packets to the destination node, say node k, we assume that the set of the neighbor nodes discovered by node i is denoted by L_i and there are β_m $(m \in L_i)$ two-hop nodes in the *Two-Hop Node List* of each neighbor node. Therefore, the number of potential two-hop paths for the node i is $\sum_{m \in L_i} \beta_m$. And there are four possibilities when node i goes through its neighbor node table:

(a) Node k is only recorded in the *Neighbor-Information Area*.

(b) Node k is only recorded in the *Two-Hop Node List* of some neighbor nodes.

(c) Node k is recorded not only in the *Neighbor-Information Area* but also in the *Two-Hop Node Lists*. In other words, node i can send its data packets to node k directly or via the relay nodes.

(d) There is no record for node k in the neighbor node list.

For the first three possibilities, every possible path should be taken into a further consideration because direct paths may have less connecting time or lower connecting ratio. The set of paths that can reach the destination node is denoted by $\widetilde{U}$.

From Algorithm 1 we can see that the time consumption for a complete transmission only consists of the able-transmitting time and the waiting time. For a specific path from the source node to the destination node, obliviously, the decisive part for a successful delivery is the weakest link among the paths, which refers to the most unstable neighbor relationship between two nodes. The calculation results of Algorithm 1 are theoretical time based upon the most recently updated history records. To make it more suitable for a general situation, we propose an inequation to filter every path in $\widetilde{U}$:

$$\min \left\{ P^{S_{ij}}_{ij} T^{h}_{\text{trans1}}, P^{S_{jk}}_{jk} T^{h}_{\text{trans2}} \right\} > H, \tag{17}$$

where $P^{S_{ij}}_{ij} T^{h}_{\text{trans1}}$ and $P_{jk} T^{h}_{\text{trans2}}$ indicate the predicted time period for delivery of each hop and H indicates the sum of transmission delay and propagation delay for one hop. In this paper, the propagation delay between the nodes is assumed to be small and the size of data packets is the same; therefore, we hypothesize that H is the same for every node. Equation (17) indicates that the minimum of the able-transmitting time period of an ideal path should be enough to complete one packet transmission.

In order to reduce the load of the network, we set a TTL, $\mathbb{T}$, for every data packet. Based upon the previous discussion, we can conclude that the TTL for the data packets in a specific path should be the sum of H and the waiting time in the path. To be more tolerant of the complex environment condition, we set $\mathbb{T}$ as follows:

$$\mathbb{T} = \max_{m \in \widetilde{U}} T^{m}_{\text{waiting}} + H. \tag{18}$$

Algorithm 2 details the procedure for the path selection. The set of qualified paths is denoted by $\widetilde{V}$.

After the path selection process, the source node generates v copies of its data packets and transmits them through the first v paths in $\widetilde{V}$. v is a preset constant number which is detailed in Section 6.

Input: T_O, T^{ij}_{begin}, T^{ij}_{end}, T^{j}_{Flag}, T^{k}_{Flag}
Output: T^{h}_{waiting}, T^{h}_{transl}, T^{h}_{trans2}
for *every row of the neighbor node table* **do**
Calculate the beginning and ending time points for S_{ji},S_{jk} (Eq. (14)-(15)).
for *every row of the Two-hop Node List* **do**
Calculate the beginning and ending time points for S_{kj} (Eq. (16)).
if $S_{ji} = S_{jk}$, $T^{ij}_{\text{end}} > T^{kj}_{\text{begin}}$ **then**
if $T_O < \max\{T^{ij}_{\text{begin}}, T^{ji}_{\text{begin}}, T^{kj}_{\text{begin}}\}$ **then**
Set $T^{h}_{\text{waiting}} = \max\{T^{ij}_{\text{begin}}, T^{ji}_{\text{begin}}, T^{kj}_{\text{begin}}\} - T_O$
Set $T^{h}_{\text{transl}} = T^{h}_{\text{trans2}} = \min\{T^{ij}_{\text{end}}, T^{ji}_{\text{end}}, T^{kj}_{\text{end}}\} - \max\{T^{ij}_{\text{begin}}, T^{ji}_{\text{begin}}, T^{kj}_{\text{begin}}\}$
else then
Set $T^{h}_{\text{waiting}} = 0$
Set $T^{h}_{\text{transl}} = T^{h}_{\text{trans2}} = \min\{T^{ij}_{\text{end}}, T^{ji}_{\text{end}}, T^{kj}_{\text{end}}\} - T_O$
end if
else then
Identify T'^{h}_{waiting} as the waiting time for the first hop of path h
Identify T''^{h}_{waiting} as the waiting time for the second hop of path h
if $T_O < \max\{T^{ij}_{\text{begin}}, T^{ji}_{\text{begin}}\}$ **then**
$T'^{h}_{\text{waiting}} = \max\{T^{ij}_{\text{begin}}, T^{ji}_{\text{begin}}\} - T_O$
Set $T^{h}_{\text{transl}} = \min\{T^{ij}_{\text{end}}, T^{ji}_{\text{end}}\} - \max\{T^{ij}_{\text{begin}}, T^{ji}_{\text{begin}}\}$
else then
$T'^{h}_{\text{waiting}} = 0$
Set $T^{h}_{\text{transl}} = \min\{T^{ij}_{\text{end}}, T^{ji}_{\text{end}}\} - T_O$
end if
Set $T''^{h}_{\text{waiting}} = \max\{T^{jk}_{\text{begin}}, T^{kj}_{\text{begin}}\} - \min\{T^{ij}_{\text{end}}, T^{ji}_{\text{end}}\}$
Set $T^{h}_{\text{waiting}} = T'^{h}_{\text{waiting}} + T''^{h}_{\text{waiting}}$
Set $T^{h}_{\text{trans2}} = \min\{T^{jk}_{\text{end}}, T^{kj}_{\text{end}}\} - \max\{T^{jk}_{\text{begin}}, T^{kj}_{\text{begin}}\}$
end if
end for
end for

Algorithm 1: Calculation for the waiting and transmitting time.

Input: $\widetilde{U}$, S_{ij}, k
Output: $\widetilde{V}$
for every path $\widetilde{u} \in \widetilde{U}$ **do**
Identify *Sector_S* in $\widetilde{u}$ as S_s
Identify $\omega^h = \min\{P^{S_{ij}}_{ij} T^{h}_{\text{transl}}, P^{S_{jk}}_{jk} T^{h}_{\text{trans2}}\}$
if $S_i = S_s$ **and** $\omega^{\widetilde{u}} > H$ **then**
$\widetilde{V} \leftarrow \widetilde{V} + \{\widetilde{u}\}$
end if
end for
if $\widetilde{V}$ is null **then**
The source node holds its data packets and continue the neighbor discovery process until it finds qualified paths.
end if
Sort $\widetilde{V}$ in descending order according to $\sigma_{ij}\sigma_{jk}\omega^h$

Algorithm 2: Path selection.

```
Input: x
Output: Ñ
  for every neighbor node j ∈ L_i do
    if β_j > x then
      Ñ ← Ñ + {j}
    end if
  end for
  if Ñ is null then
    for every neighbor node j ∈ L_i do
      if j has different neighbor nodes compared to L_i then
        Ñ ← Ñ + {j}
      end if
    end for
  end if
```

Algorithm 3: Extreme situation treatment.

Then, we discuss the fourth situation that there is no record for node k in the neighbor node list. Apparently, a node that has encountered more neighbor nodes may get more topology information in its two-hop distance. Hence, when this kind of extreme situation occurs, node i sends its data packets to all the neighbor nodes who have discovered more neighbor nodes than node i dose; that is, node i sends its data packets to node j if β_j is bigger than the size of L_i. If there is still no qualified neighbor node after that, then we choose those nodes who have encountered different neighbor nodes compared to L_i. Algorithm 3 describes this procedure. The size of L_i is denoted by x, and the set of qualified neighbor nodes is denoted by $\widetilde{N}$.

6. Simulation Results

In this section, we evaluate the performance of *HSD* and *HSD-R* separately. We use a simulation setup that consists of a uniform distribution of nodes over a 2-dimensional plane of area 20 km × 20 km. The fixed transmission range is $R = 10$ km and the angle of each directional antenna is $\pi/3$, that is, 6 sectors for each vehicle node. The rotation period of the directional antennas is $T = 180\tau$ and the duration of each receiving state is $n = 9$ which indicates that each node can broadcast HELLO messages 18 times in a circle totally. The monitoring time $M = 360\tau$. The moving speed of the vehicle nodes is 50 km/h which is nearly 15 meters every 180τ with a random direction.

We vary the total number of nodes inside this area with 5, 10, and 50, and they represent three levels of node density which are extremely low, low, and high, respectively. We assume that the area of the community is 4 km × 4 km, and the number of nodes in one community is random which is from 2 to 5 and $P = 0.1$. In the extremely low density scenario, we artificially divide the 5 nodes into two groups with 2 and 3 separately. We run the simulation 20 times and the simulation duration is $10^4\tau$. Nodes are randomly scattered inside the area every time. There is a chance that nodes cannot discover all other nodes because their directional antennas cannot cover the whole area.

6.1. Evaluation of HSD. Firstly, we compare and analyze the performance of *HSD* and the random algorithm. When the nodes switch their sectors clockwise with an equal and fixed C_{S_i}, it turns out that their average time consumption is much larger than *HSD* and the random algorithm. Therefore, we only concentrate on the simulation comparison of *HSD* and the random algorithm. Figure 7 shows the performances of the two algorithms on every node to successfully discover a certain percentage of neighbor nodes inside the area.

In Figure 7(c), *HSD* and the random algorithm show no obvious difference on time consumption when the discovered nodes are under 70% since, in the high density scenario, there are plenty of neighbor nodes in any sectors so that the scan pattern of *HSD* is similar to the random algorithm. When the discovered nodes reach 70%, the performance of *HSD* is worse. Because higher density of nodes leads to more frequent communication among nodes, the nodes especially which are located at the central part of the scenario can go through turbulence on the schedule of sectors. However, as shown in Figures 7(a) and 7(b), the *HSD* achieves better performance since the probability of communication is smaller and the change of topology is easier to be detected when the density of nodes is low. Especially in Figure 7(a), when nodes in one community have discovered a node in another community, they can point their directional antennas to the certain sector where another community locates for a long time since there are few interferences from other sectors. Hence, the average time consumption and the standard deviation of time consumption of *HSD* are both smaller than those of the random algorithm. According to [20], the packet size has obvious influence on the performance of protocols. However, the interactive action among nodes is rare since the node density in this paper is quite low; for example, only one neighbor node exists within a sector for the low density scenario.

In summary, simulation results demonstrate that *HSD* achieves better performance than the random algorithm on neighbor discovery when nodes are sparsely scattered.

6.2. Evaluation of HSD-R. From the simulation results of *HSD*, we conclude that *HSD* achieves better performance

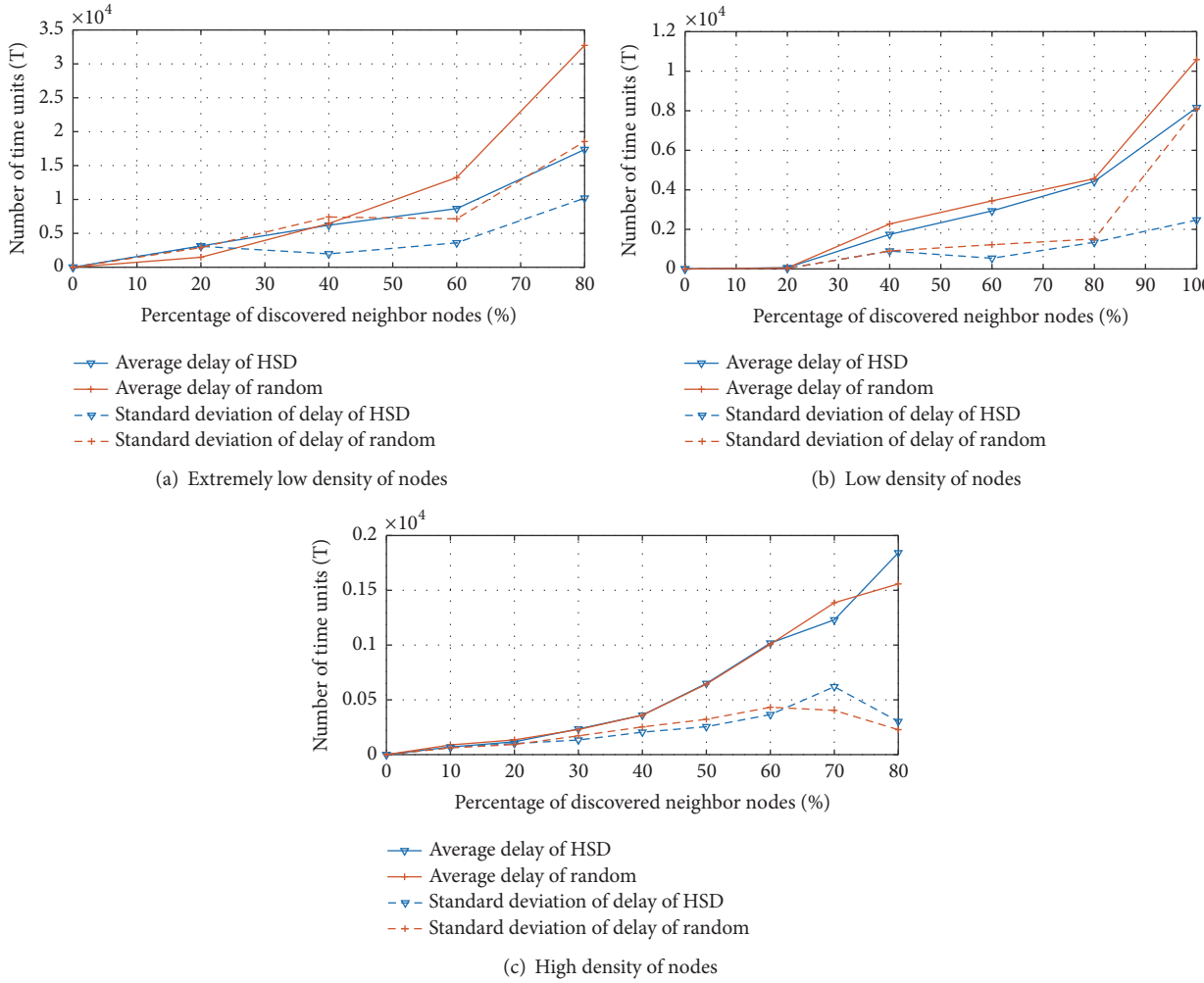


Figure 7: Simulation result comparison on different node densities.

when nodes are sparsely scattered. Reference [16] proposes a routing protocol called *Spray and Focus*. Though *Spray and Focus* is not designed for the networks where directional antennas are equipped, it is a classic routing protocols for opportunistic networks and the most important is that nodes in [16] show location and group preference as well. Therefore, we compare and analyze the performance of *HSD-R* and *Spray and Focus* in low density scenario and high density scenario separately. *Spray and Focus* consists of two phases: in the first phase it distributes a fixed number of copies to the first few relays encountered, and in the second phase each relay can forward its copy to a potentially more appropriate relay, using a carefully designed utility-based scheme.

According to [16], the number of packet copies equal to about 5–10% of the total nodes serves as a useful rule of thumb for good performance. Because nodes in [16] are not equipped with the directional antennas, the connecting probability among nodes in our scenario is much lower than that in [16]. Under this consideration, we increase the number of packets copies, that is, the number of selected paths, v, gradually from approximately 5% to 20% of the total nodes to compare their performance. The beginning time of data transmission for every node is random and the destination for every source node is selected randomly as well.

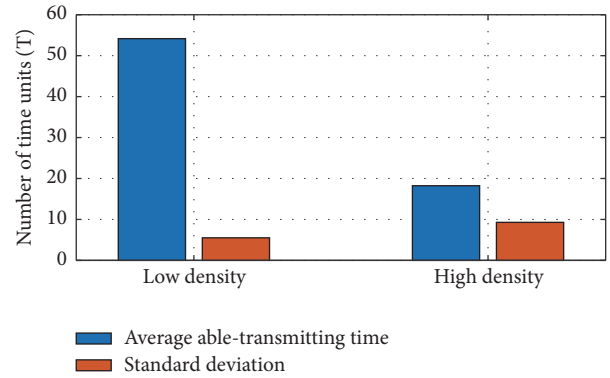


Figure 8: Average able-transmitting time in low density scenario and high density scenario with v equal to 25% of the total nodes.

When the node density is low, only two paths can be selected from $\widetilde{V}$ and around ten paths for high density scenario. As illuminated in Figure 8, the average able-transmitting

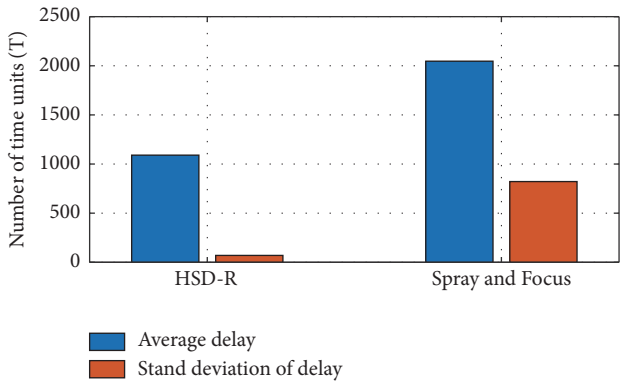


Figure 9: Average delay in low density scenario.

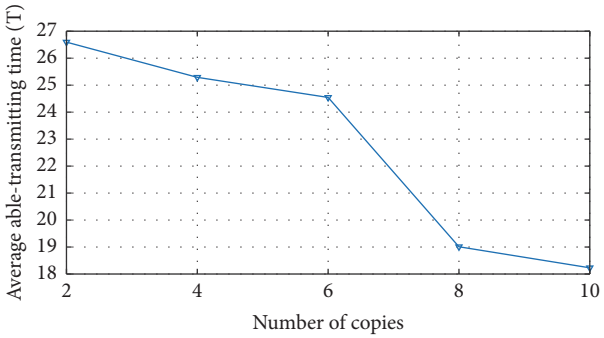


Figure 10: Average able-transmitting time in high density scenario with different v.

time, T^h_{trans}, among nodes in low density scenario is bigger since nodes mostly point their directional antennas to certain sectors where another community locates for a long time and there are few interferences from other sectors.

Figure 9 depicts the performance of the two protocols in low density scenario. Compared to *Spray and Focus*, the transmission delay of *HSD-R* is much lower and it shows no obvious differences for every node. Figure 9 indicates that the selected paths in *HSD-R* do reduce the delay. The successful transmission rates of the two protocols in low density scenario are all above 90%.

When the node density gets higher, in Figure 10, we can see an obvious decrease on T^h_{trans} when the number of copies is bigger than 6, that is, around 10% of the total nodes. It indicates that the average able-transmitting time of the first 6 paths is much bigger than the left paths. Figure 11 proves that the minimum number of copies for *HSD-R* to achieve a reasonable success rate and delay is around 10% as well. Moreover, we can see that the success rate of *Spray and Focus* decreases dramatically when the number of copies is small while the transmission delay of *HSD-R* is slightly changed. And the minimum of number copies for *HSD-R* to achieve a reasonable performance is smaller than *Spray and Focus*. What needs to be explained is that the obvious success rate difference between *HSD-R* and *Spray and Focus* is caused by small number of total nodes in the scenario, and few data packets failing to be delivered to the destination can lead to a great decrease in success rate. We do not have the simulation

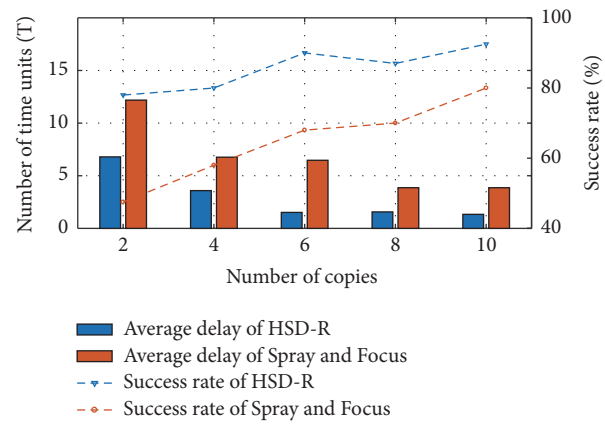


Figure 11: Success rate and average delay in high density scenario with different v.

on extremely high density scenarios because the directional antenna is not designed for that kind of scenarios.

In summary, compared to *Spray and Focus*, *HSD-R* can guarantee a successful delivery with a smaller number of packet copies while directional antennas are equipped.

7. Conclusion

In this paper, we concentrate on the networking of vehicles. Firstly, we have proposed a novel neighbor discovery algorithm called *History-Based Sector Distribution (HSD)* for vehicle ad hoc networks where directional antenna is equipped. The proposed *HSD* can make the vehicles be able to arrange their directional antennas according to the topology of the network. We have proved that our algorithm can achieve better performances than the traditional random algorithm especially in the scenario where vehicle nodes are sparsely scattered. Secondly, we design a routing protocol called *HSD-R* which is derived from *HSD*. In *HSD-R*, vehicle nodes make their routing decisions by analyzing the link quality of each possible path between the source node and the destination node based on the information collected during the neighbor discovery process. The evaluation results of *HSD-R* prove that high successful transmission rate and low transmission delay are guaranteed with a small number of packet copies.

Conflicts of Interest

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

Acknowledgments

This work was supported in part by the National Key R&D Plan of China (Grants 2016YFB1200100).

References

[1] L. Chen, Y. Li, and A. V. Vasilakos, "Oblivious neighbor discovery for wireless devices with directional antennas," in

Proceedings of the 35th Annual IEEE International Conference on Computer Communications, IEEE INFOCOM 2016, pp. 1–9, April 2016.

[2] S. Huang, M. Li, and L. Zhao, "An intelligent neighbor discovery algorithm for ad hoc networks with directional antennas," in *Proceedings of the 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer, MEC 2013*, pp. 302–305, chn, December 2013.

[3] W. Xiong, B. Liu, and L. Gui, "Neighbor discovery with directional antennas in mobile ad-hoc networks," in *Proceedings of the 54th Annual IEEE Global Telecommunications Conference: "Energizing Global Communications", GLOBECOM 2011*, USA, December 2011.

[4] S. Zhao, Y. Liu, T. Yang, Z. Feng, Q. Zhang, and C. Gao, "3-Way multi-carrier asynchronous neighbor discovery algorithm using directional antennas," in *Proceedings of the 2016 IEEE Wireless Communications and Networking Conference, WCNC 2016*, Qatar, April 2016.

[5] H. Cai and T. Wolf, "On 2-way neighbor discovery in wireless networks with directional antennas," in *Proceedings of the 34th IEEE Annual Conference on Computer Communications and Networks, IEEE INFOCOM 2015*, pp. 702–710, Hong Kong, May 2015.

[6] H. Cai, B. Liu, L. Gui, and M.-Y. Wu, "Neighbor discovery algorithms in wireless networks using directional antennas," in *Proceedings of the 2012 IEEE International Conference on Communications, ICC 2012*, pp. 767–772, Canada, June 2012.

[7] Z. Zhang and B. Li, "Neighbor discovery in mobile ad hoc self-configuring networks with directional antennas: Algorithms and comparisons," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1540–1549, 2008.

[8] Z. H. Mir, W.-S. Jung, and Y.-B. Ko, "Continuous neighbor discovery protocol in wireless ad hoc networks with sectored-antennas," in *Proceedings of the 29th IEEE International Conference on Advanced Information Networking and Applications, AINA 2015*, pp. 54–61, Republic of Korea, March 2015.

[9] F. Tian, R. Q. Hu, Y. Qian, B. Rong, B. Liu, and L. Gui, "Pure asynchronous neighbor discovery algorithms in ad hoc networks using directional antennas," in *Proceedings of the 2013 IEEE Global Communications Conference, GLOBECOM 2013*, pp. 498–503, USA, December 2013.

[10] F. Tian, B. Liu, H. Cai, H. Zhou, and L. Gui, "Practical Asynchronous Neighbor Discovery in Ad Hoc Networks with Directional Antennas," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3614–3627, 2016.

[11] E. Gelai, G. Jakllari, S. V. Krishnamurthy, and N. E. Young, "An integrated scheme for fully-directional neighbor discovery and topology management in mobile ad hoc networks," in *Proceedings of the 2006 IEEE International Conference on Mobile Ad Hoc and Sensor Sysetems, MASS*, pp. 139–149, Canada, October 2006.

[12] Z. Huang, C.-C. Shen, C. Srisathapornphat, and C. Jaikaeo, "Topology control for ad hoc networks with directional antennas," in *Proceedings of the 11th International Conference on Computer Communications and Networks, ICCCN 2002*, pp. 16–21, USA, October 2002.

[13] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: the single-copy case," *IEEE/ACM Transactions on Networking*, vol. 16, no. 1, pp. 63–76, 2008.

[14] V. Conan, J. Leguay, and T. Friedman, "Fixed point opportunistic routing in delay tolerant networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 5, pp. 773–782, 2008.

[15] S. Shaghaghian and M. Coates, "Optimal forwarding in opportunistic delay tolerant networks with meeting rate estimations," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 2, pp. 104–116, 2015.

[16] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and focus: efficient mobility-assisted routing for heterogeneous and correlated mobility," in *Proceedings of the 5th Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops '07)*, pp. 79–85, March 2007.

[17] Z. Chen and C. Chen, "Self-stopping epidemic routing in cooperative wireless mobile sensor networks," in *Proceedings of the 12th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob 2016*, USA, October 2016.

[18] C. Y. Aung, I. W.-H. Ho, and P. H. J. Chong, "Store-Carry-Cooperative Forward Routing with Information Epidemics Control for Data Delivery in Opportunistic Networks," *IEEE Access*, vol. 5, pp. 6608–6625, 2017.

[19] C. Wu, S. Ohzahata, Y. Ji, and T. Kato, "How to Utilize Interflow Network Coding in VANETs: A Backbone-Based Approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 8, pp. 2223–2237, 2016.

[20] C. Wu, X. Chen, Y. Ji et al., "Packet size-aware broadcasting in VANETs with fuzzy logic and RL-based parameter adaptation," *IEEE Access*, vol. 3, pp. 2481–2491, 2015.

[21] T. Noguchi and T. Kobayashi, "Adaptive Location-Aware Routing with Directional Antennas in mobile adhoc networks," in *Proceedings of the 2017 International Conference on Computing, Networking and Communications, ICNC 2017*, pp. 1006–1011, USA, January 2017.

[22] J. Yoon, W. Shin, and S. Jeon, "Elastic Routing in Ad Hoc Networks with Directional Antennas," *IEEE Transactions on Mobile Computing*, vol. 16, no. 12, pp. 3334–3346, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 6353714, 25 pages
https://doi.org/10.1155/2018/6353714

Research Article

Quality Utilization Aware Based Data Gathering for Vehicular Communication Networks

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

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

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

Received 13 December 2017; Accepted 14 February 2018; Published 27 March 2018

Academic Editor: Tao Han

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

1. Introduction

Vehicles which equipped various types of sensing devices can sense all kinds of information in the surrounding environment [1–3]. At the same time, due to its strong mobility, the use of participatory sensing can obtain a large amount of sensing data in a very economical way [4–9]. These data could build various applications based on bid data. The examples of the classical use of such applications are as follows: VTrack [10] and Waze [11] is a system which can provide omnipresent traffic information by collecting vehicle operating status information; WeatherLah [12] is for giving fine-grained situation on the ground; and NoiseTube [13] is for making noise maps. In such applications, the time and space for data gathering are very large which leads to the general use of participatory sensing to get the data. This type of network is called crowdsourcing network or participatory network [14–19]. Apparently, vehicular communication networks (VCNs) have a greater advantage in this regard [20–22]. First of all, it is a good participatory network for the reason that it has big number of vehicles which equipped various sensing devices which can sense different data efficiently; second, vehicles generally have strong communication skills [22, 23], which enable vehicles to exchange information with other vehicles, data center, and the external environments. They also play a significant role in supporting various applications such as road safety, traffic management, and a wide range of applications based on perceptual data [10–13]. On the other way, vehicular communication networks are also regarded as vital key technologies for next-generation

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

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

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

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

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

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

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

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

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

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

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

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

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

2. Related Work

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

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

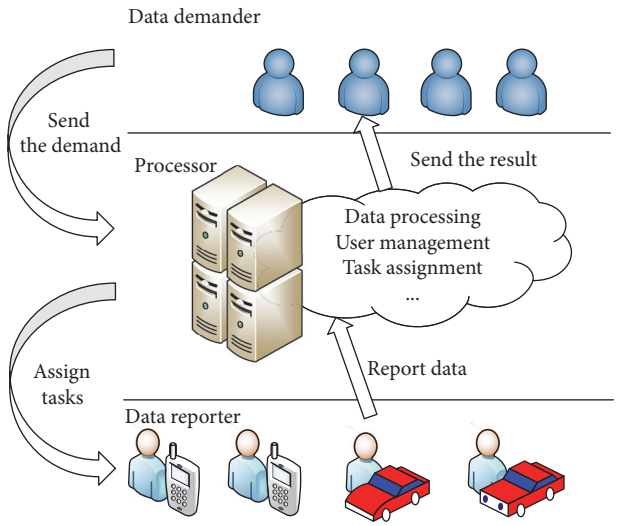


Figure 1: The classical architecture of crowdsourcing.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3. System Models and Problem Statements

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

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

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

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

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

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

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

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

$$\mathscr{P}^{\mathscr{A}} = \mathscr{P}_{A,\Gamma}^{\mathscr{A}}. \tag{2}$$

That is,

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

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

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

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

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

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

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

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

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

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

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

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

The quality of all areas is

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

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

The allocation of the system could be expressed as

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

4. Scheme Design

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

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

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

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

Actually, the quality of different application is different. Besides, the quality of the data in the same area which is

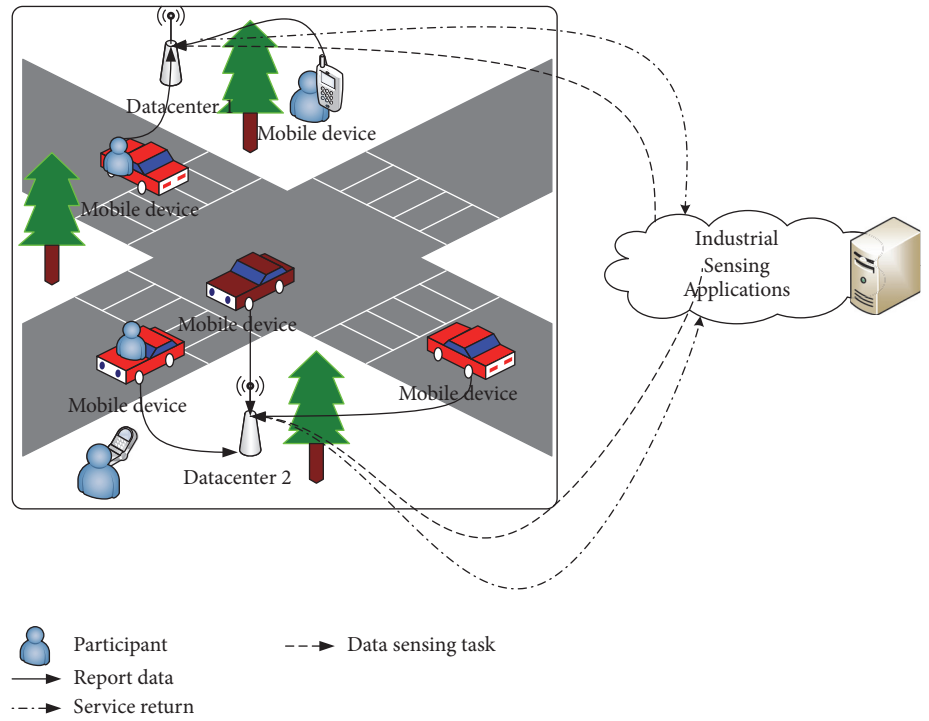


FIGURE 2: The process of participatory data collection.

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

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

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

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

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

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

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

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

The corresponding ordered data list can be denoted as

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

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

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

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

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

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

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

The corresponding ordered data list can be denoted as

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

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

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

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

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

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

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

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

(3) $U_i = \{\mathfrak{R}^{\mathscr{A}}, \mathscr{P}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$. In this situation, the demand of user is tight. Data demander proposed the request for payment and quality at the same time. But the request of data demander may not be satisfied at the same time. To quantify the satisfying degree of data demander, the index satisfaction $\mathscr{é}$ is used to measure the demander's satisfaction:

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

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

In order to facilitate the comparison of data, besides the overall satisfaction $\mathscr{é}$, for a single data, $\mathscr{d}_m \in D_i$ system also computes the data demander satisfaction τ_m^i.

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

Then the algorithm steps are as follows.

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

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

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

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

The corresponding ordered data list can be denoted as

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

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

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

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

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

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

5. Experimental Evaluation

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

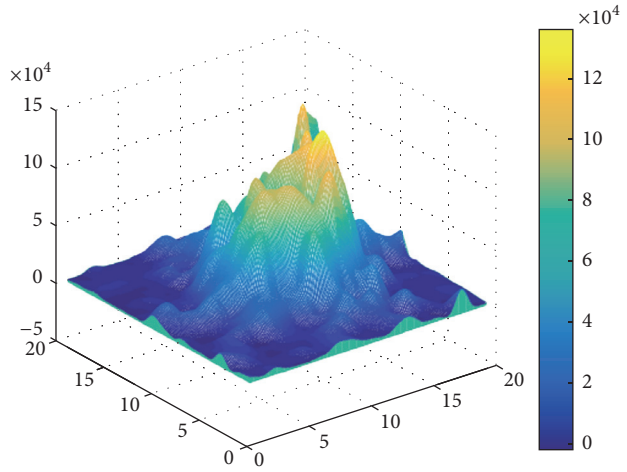


Figure 3: The frequency of vehicles.

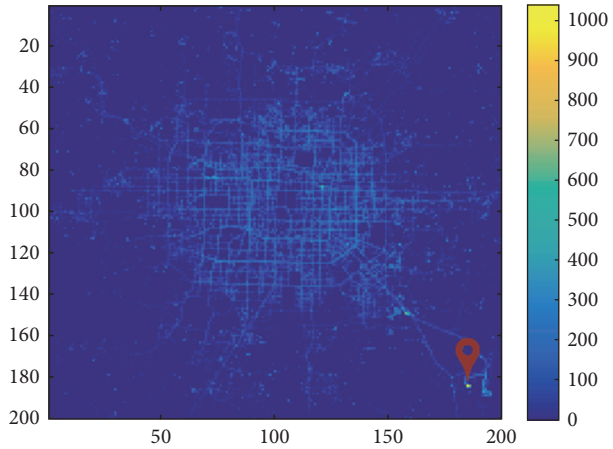


Figure 4: The distribution of vehicles in the city.

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

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

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

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

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

5.2. The Estimation of a Certain Grid i. The bid and the data quality of each data reporter can be generated by the Normal Distribution. We assume that the amount of submitted data of gird area i, $\forall i \in A$ is 100 for convenience. The generated quality of data reporter is shown in Figure 5. The generated bid of each data reporter is shown in Figure 6. And the

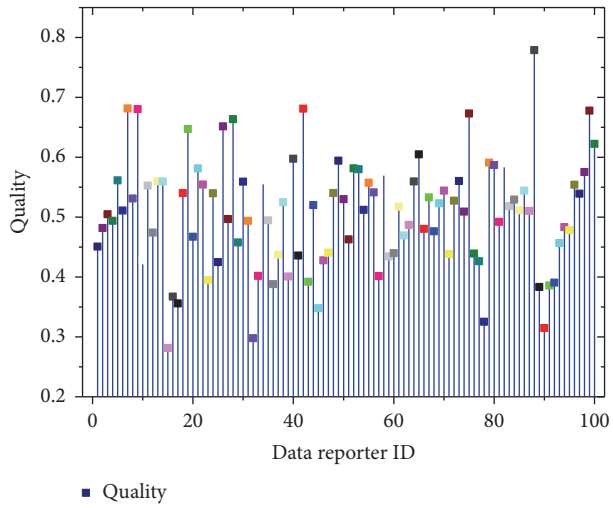


Figure 5: The quality of 100 data reporters. The colored boxes, other than "the black box," refer to the different value of quality and bid.

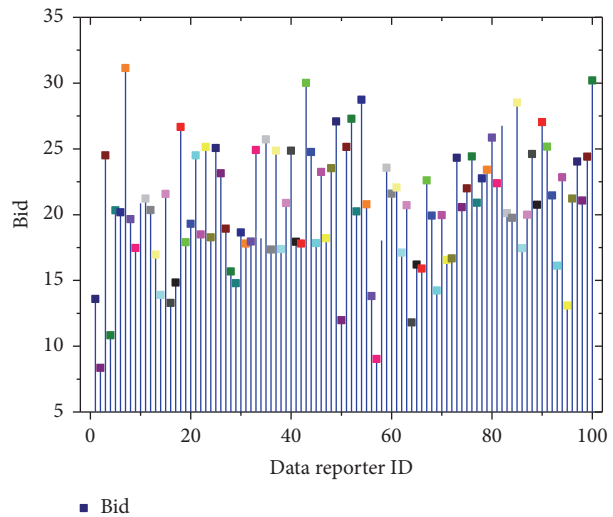


Figure 6: The bids of 100 data reporters. The colored boxes, other than "the blue box," refer to the different value of bid.

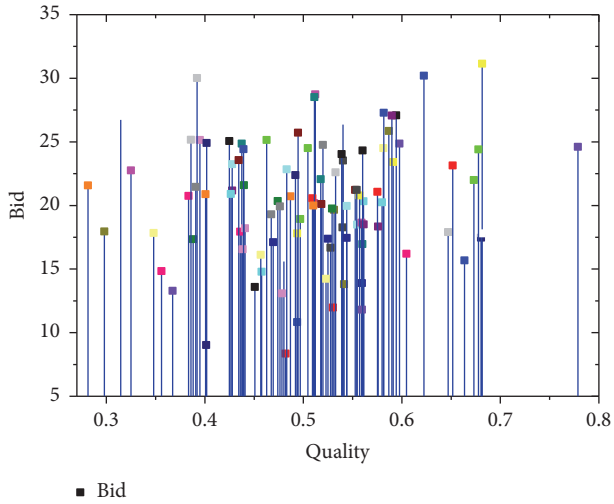


FIGURE 7: The relationship between quality and bid. The colored boxes, other than "the black box," refer to the different value of quality and bid.

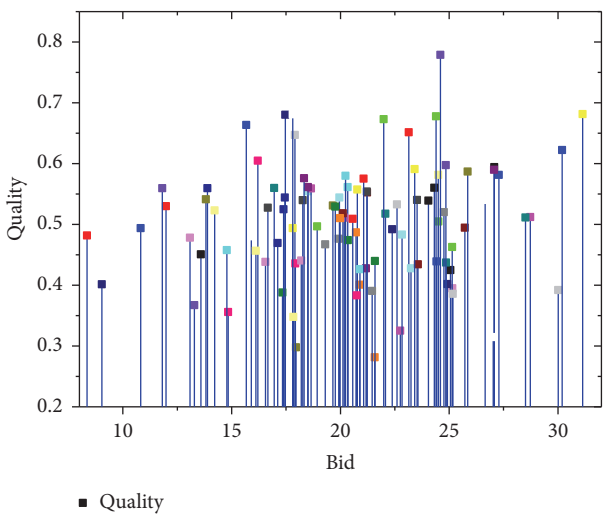


FIGURE 8: The relationship between bid and quality. The colored boxes, other than "the black box," refer to the different value of quality and bid.

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

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

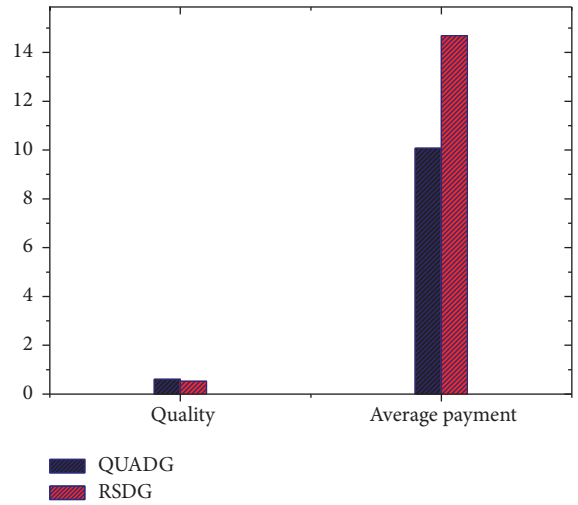


FIGURE 9: The comparison of quality and average payment.

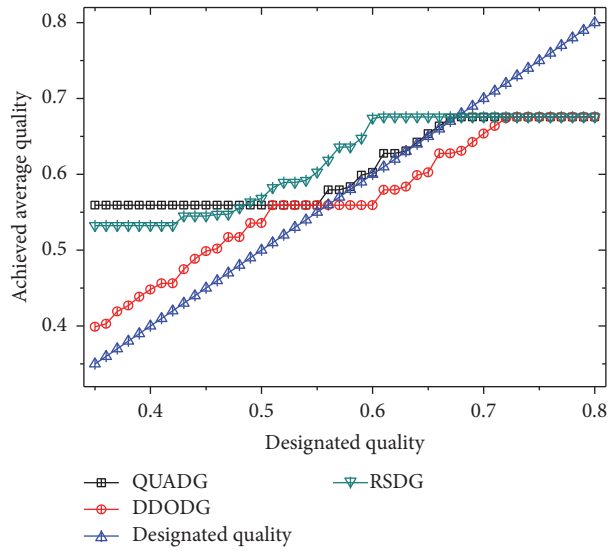


FIGURE 10: The comparison of achieved data quality.

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

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

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

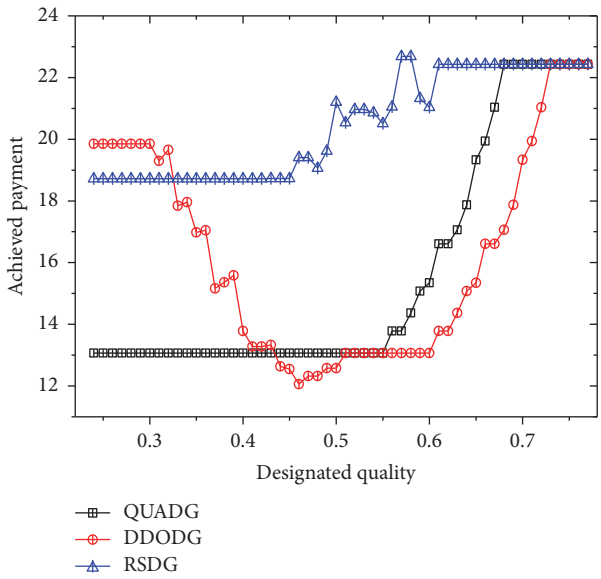


Figure 11: The comparison of achieved payment.

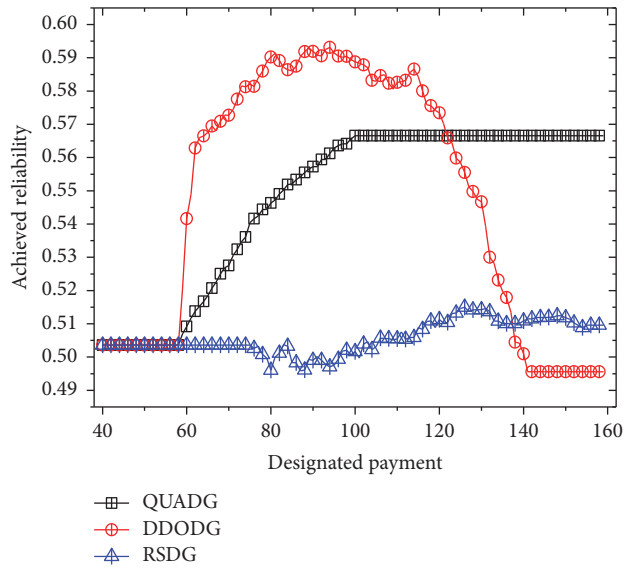


Figure 12: The comparison of the achieved quality.

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

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

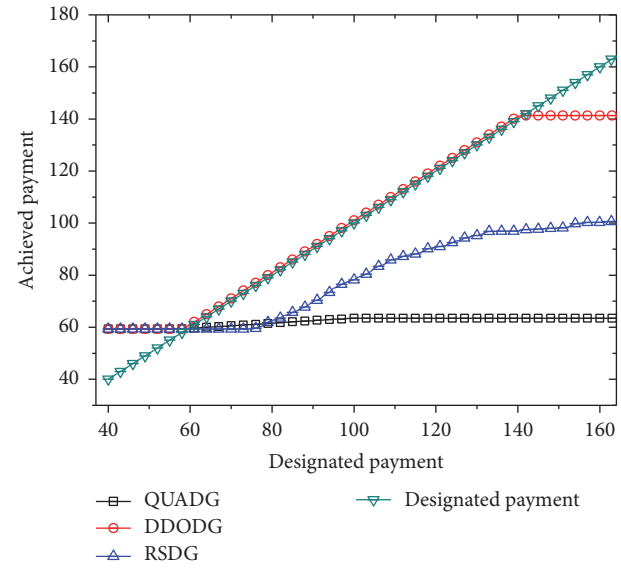


Figure 13: The comparison of the achieved payment.

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

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

When the satisfaction $\mathscr{E}$ is designated as 0.1 and $\mathscr{L}_A^{\mathscr{A}} = 0.5$ the achieved quality and the achieved payment by designating different payment are shown in Figures 15 and 16, respectively.

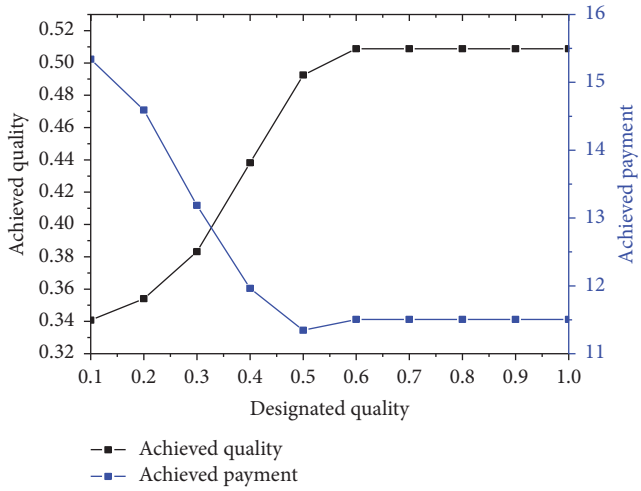


FIGURE 14: The achieved quality and payment under different deviate.

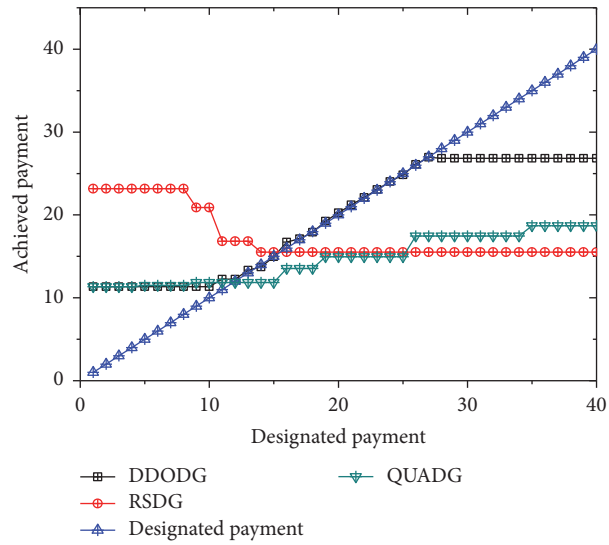


FIGURE 16: The achieved payment by designating payment.

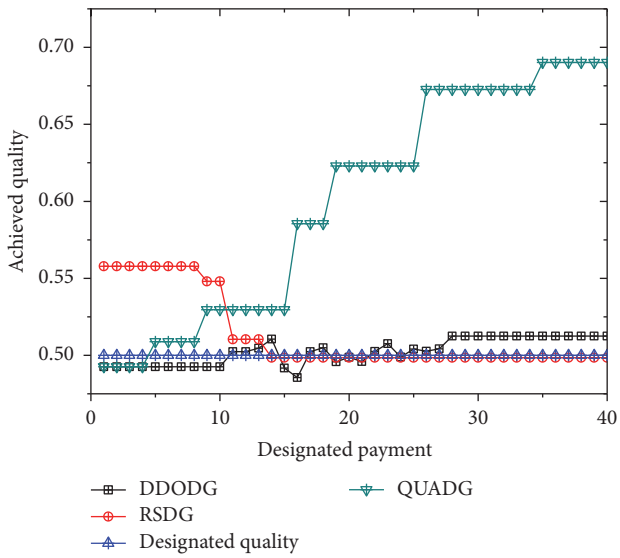


FIGURE 15: The achieved quality by designating payment.

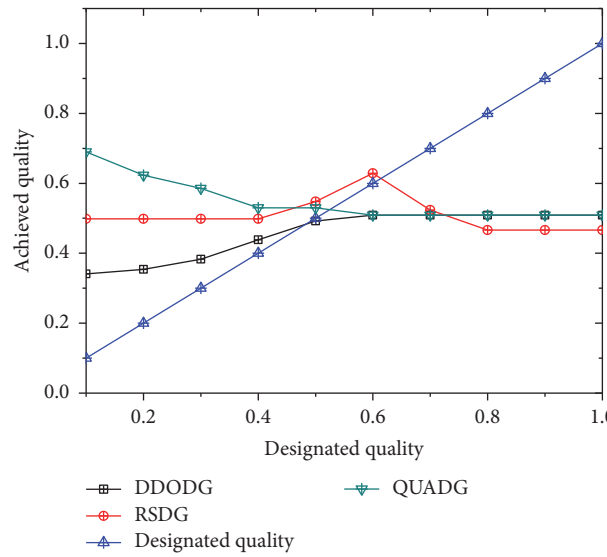


FIGURE 17: The achieved quality by designating quality.

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

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

5.3. The Estimation of All Grids

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

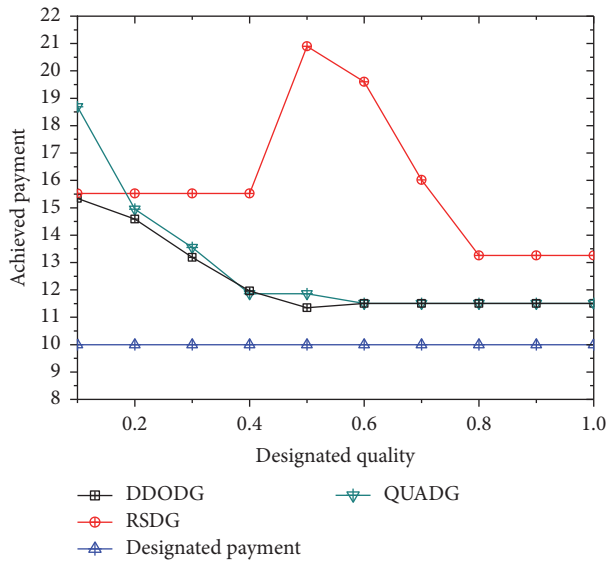


Figure 18: The achieved payment by designating quality.

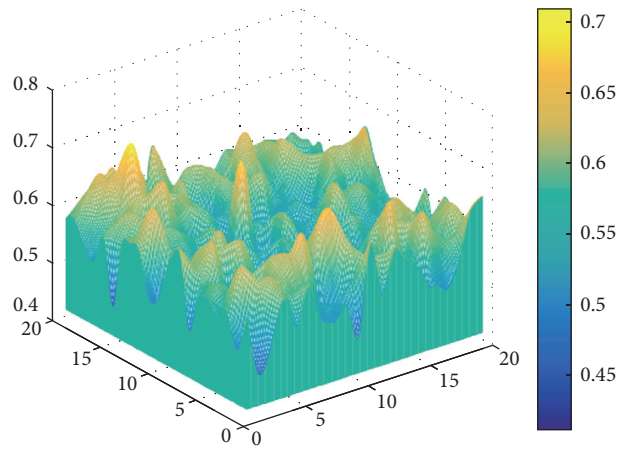


Figure 19: The overall quality of QUADG.

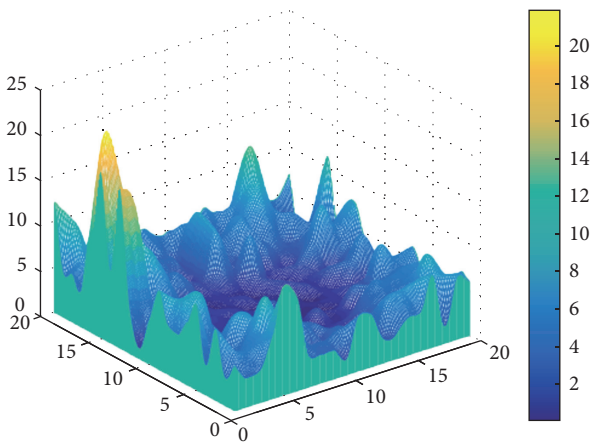


Figure 20: The overall payment of QUADG.

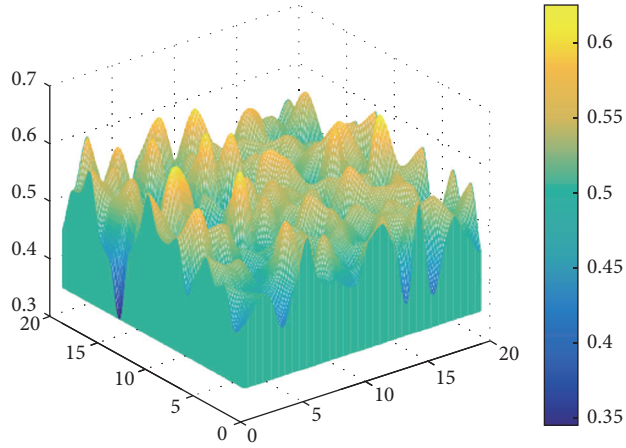


Figure 21: The overall quality of RSDG.

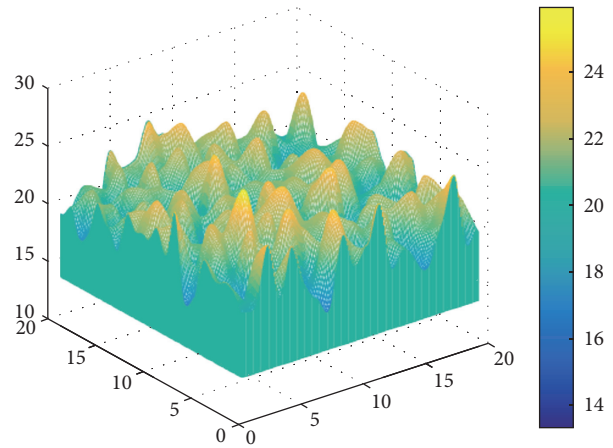


Figure 22: The overall payment of RSDG.

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

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

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

From Figures 25 and 29, we could see that the payment by QUADG in the center area is less when the requested

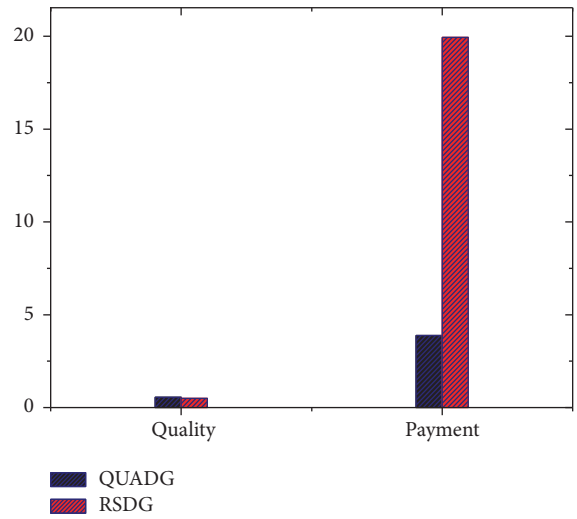


Figure 23: The comparison of quality and payment.

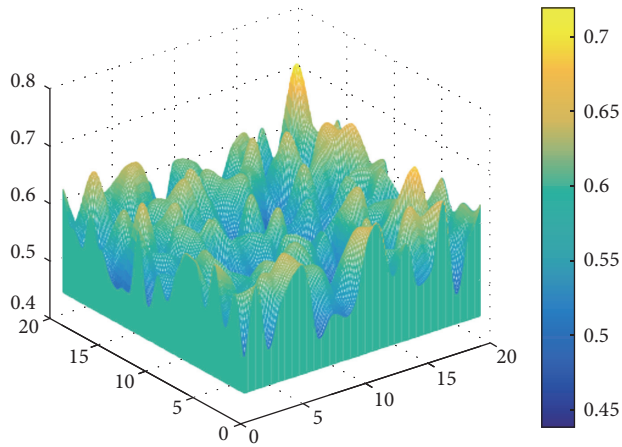


Figure 24: The quality of QUADG.

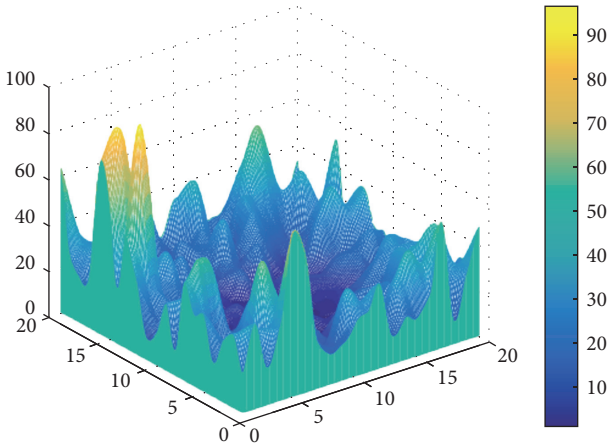


Figure 25: The payment of QUADG.

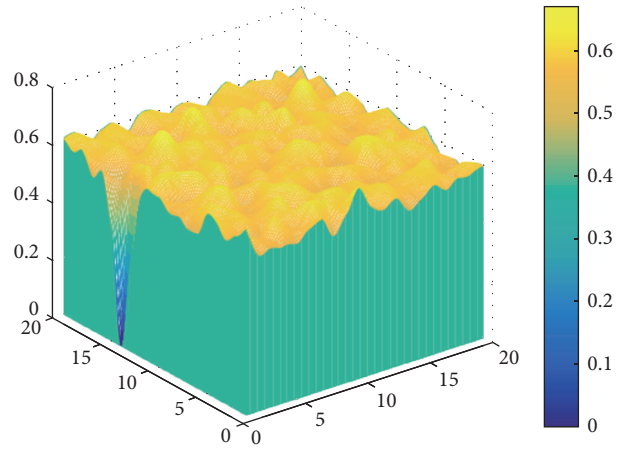


Figure 26: The quality of RSDG.

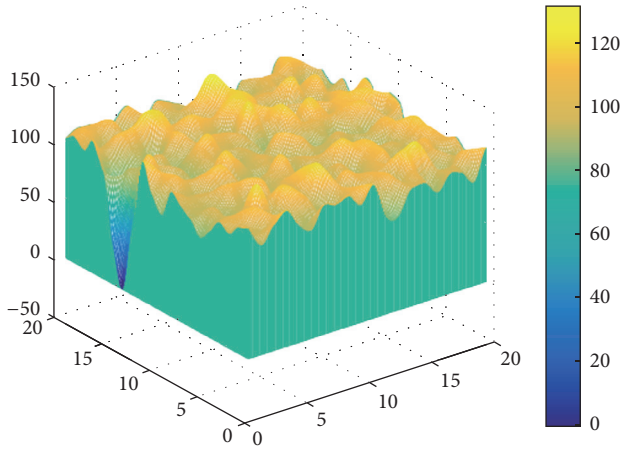


Figure 27: The payment of RSDG.

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

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

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

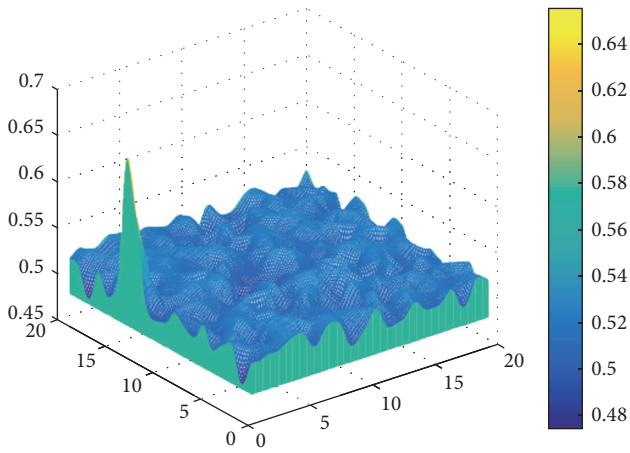


Figure 28: The quality of DDODG.

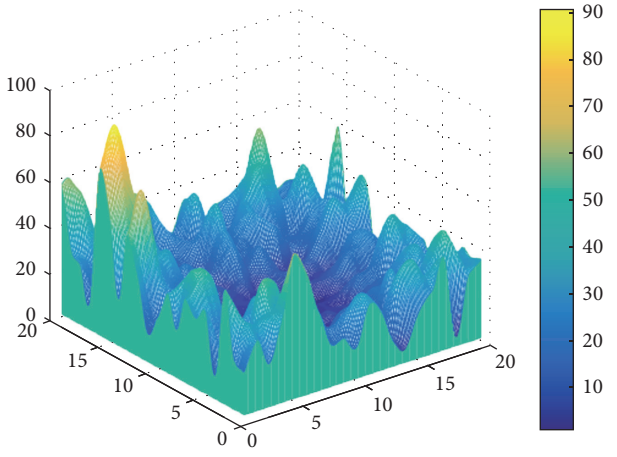


Figure 29: The payment of DDODG.

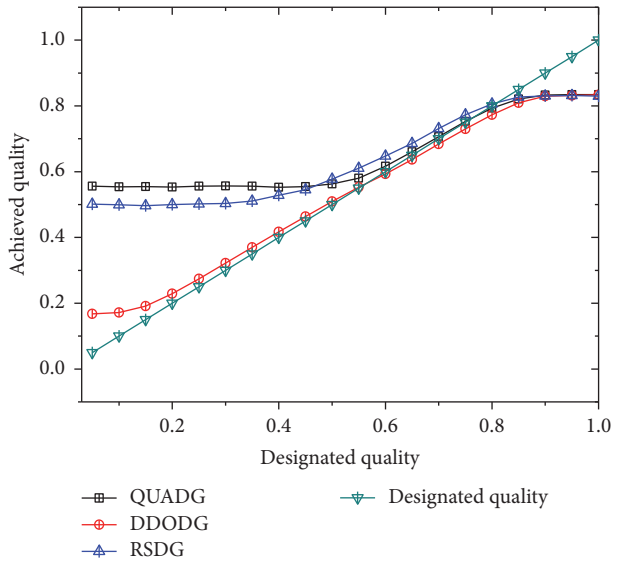


Figure 30: The comparison of achieved quality.

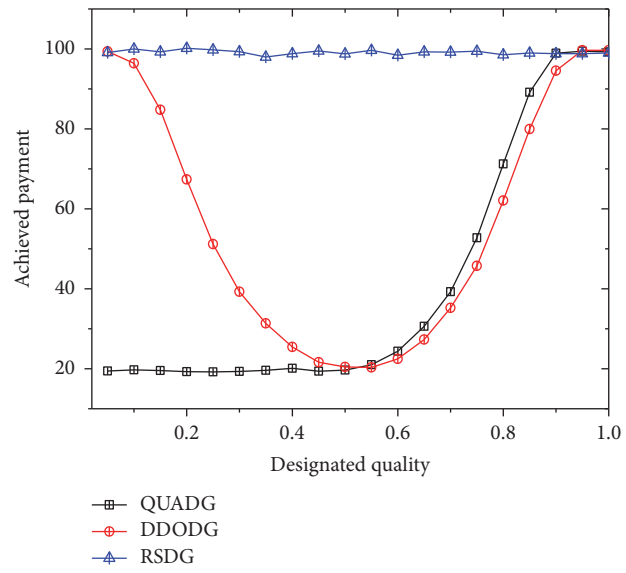


Figure 31: The comparison of achieved payment.

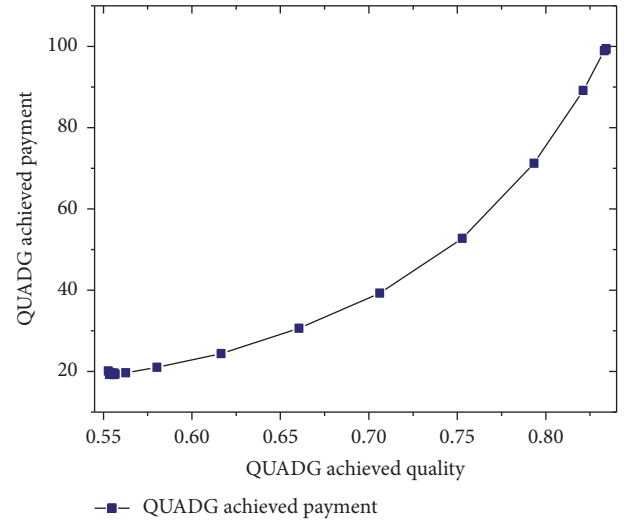


Figure 32: The relationship between achieved quality and payment with QUADG.

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

5.3.3. $U_i = \{\mathfrak{R}_i^{\mathscr{A}}, \mathscr{P}_i^{\mathscr{A}}\}$. If both $\mathfrak{R}_i^{\mathscr{A}}$ and $\mathscr{P}_i^{\mathscr{A}}$ are designated, to compare more conveniently we assume the requested payment of each data is 5 and the amount of requested data is 5.

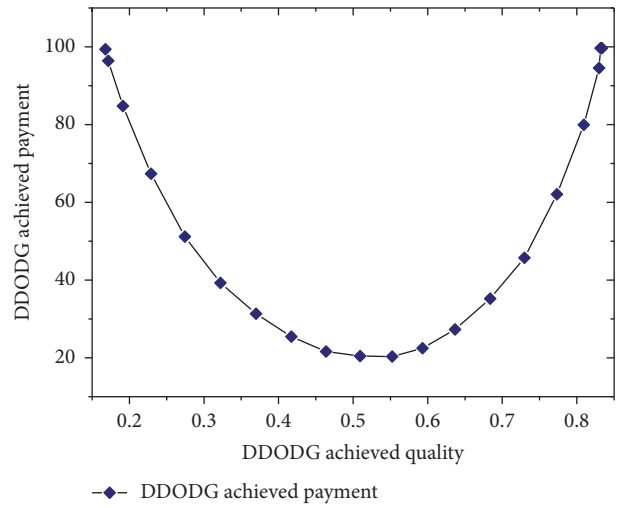


Figure 33: The relationship between achieved quality and payment with DDODG.

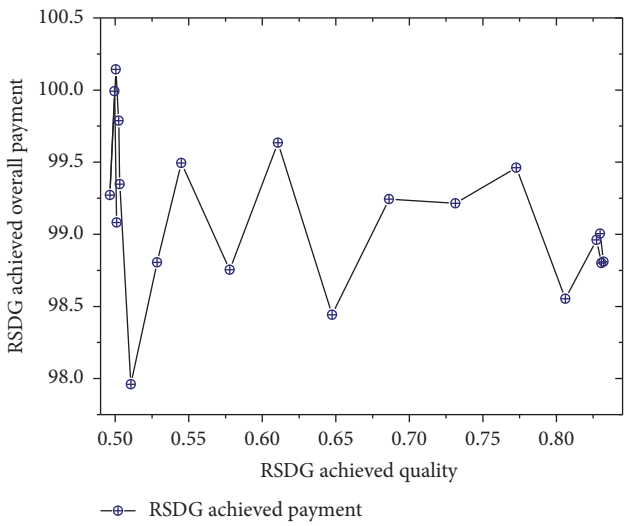


Figure 34: The relationship between achieved quality and payment with RSDG.

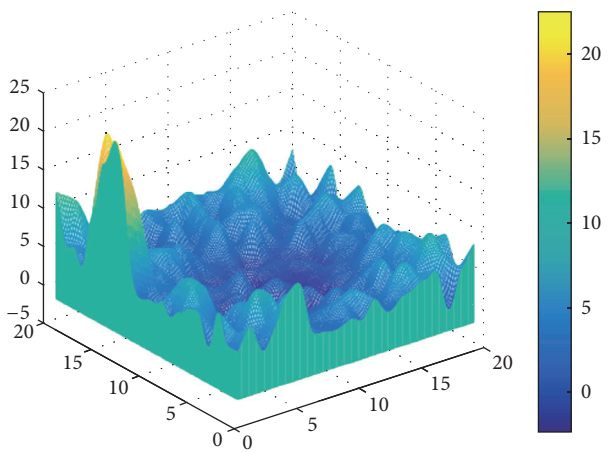


Figure 35: The payment of QUADG.

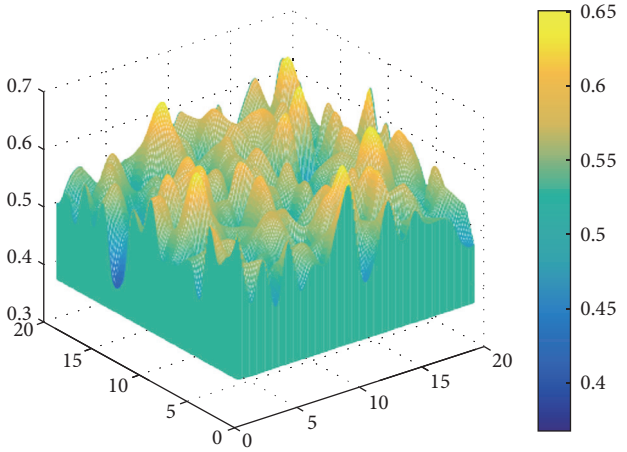


Figure 36: The quality of QUADG.

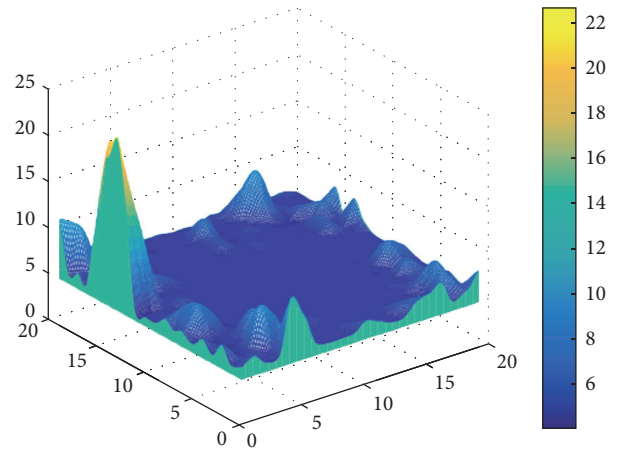


Figure 37: The payment of DDODG.

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

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

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

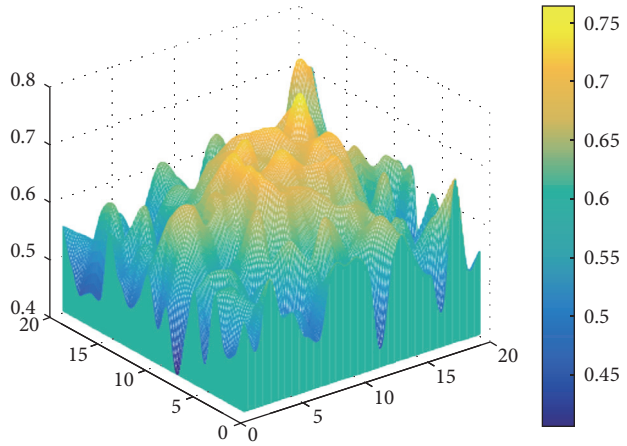


Figure 38: The quality of DDODG.

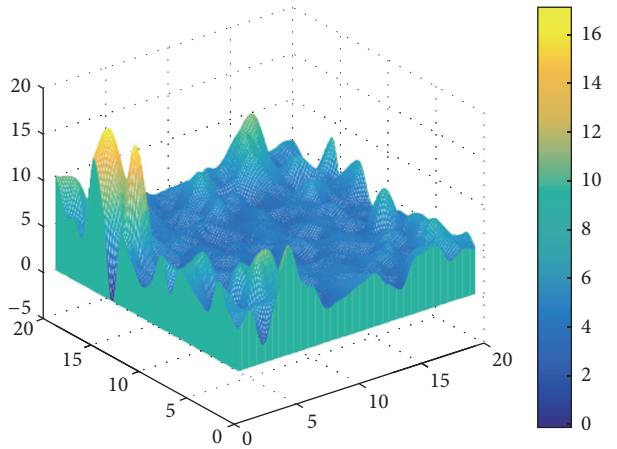


Figure 39: The payment of RSDG.

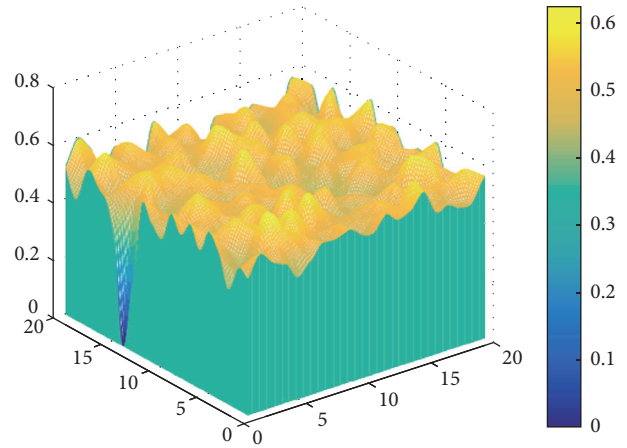


Figure 40: The quality of RSDG.

of data demander, DDODG strategy will choose the data with the higher bid despite data quality. The QUADG strategy can be used to obtain the best solution of the combination of payment and quality when the data demander only makes a request for the payment.

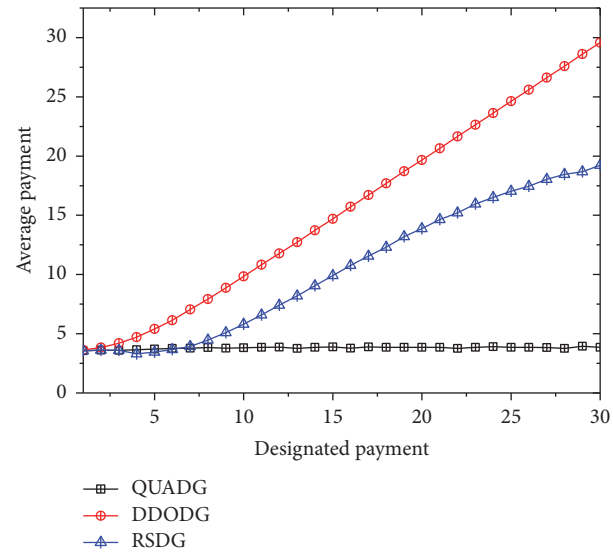


Figure 41: The achieved average payment by designating payment.

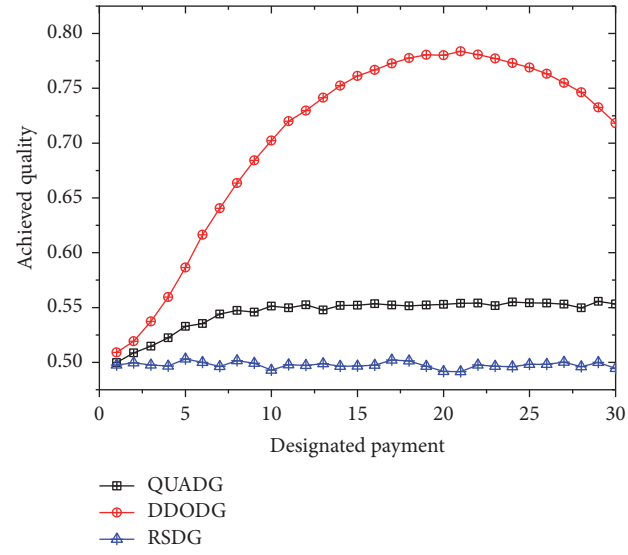


Figure 42: The achieved quality by designating payment.

5.3.4. $U_i = \{\Re^{\mathscr{A}}, \mathscr{P}^{\mathscr{A}}, \mathscr{L}^{\mathscr{A}}\}$. To simplify the process of comparison, we first assume the average payment of a single data is 10. And the requested quality is 0.45.

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

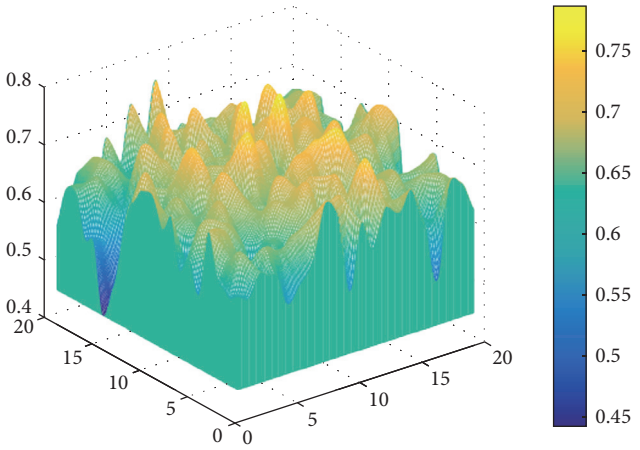


Figure 43: The achieved quality of QUADG.

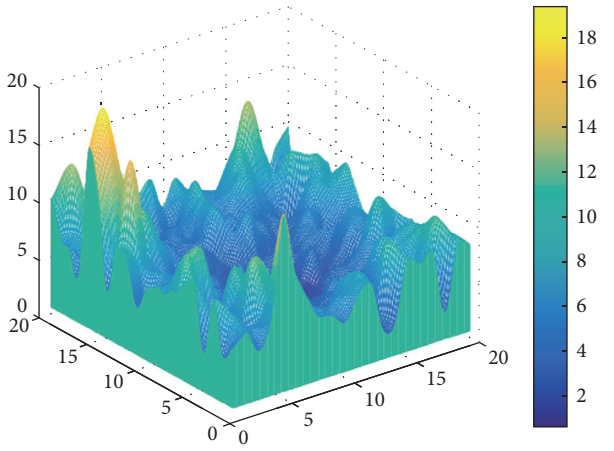


Figure 44: The achieved payment of QUADG.

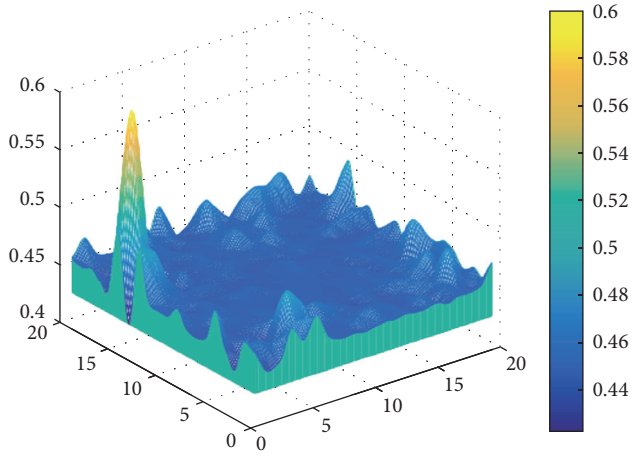


Figure 45: The achieved quality of DDODG.

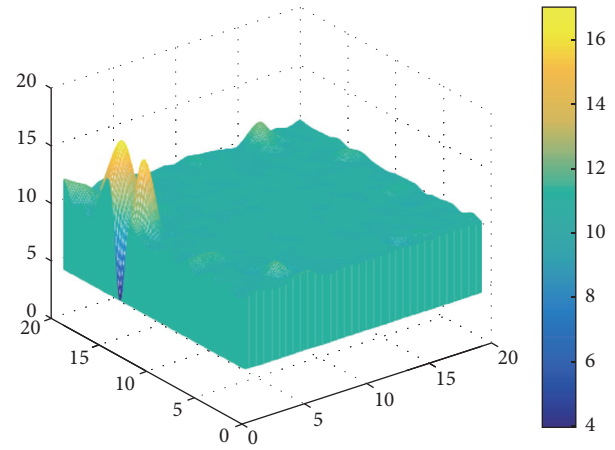


Figure 46: The achieved payment of DDODG.

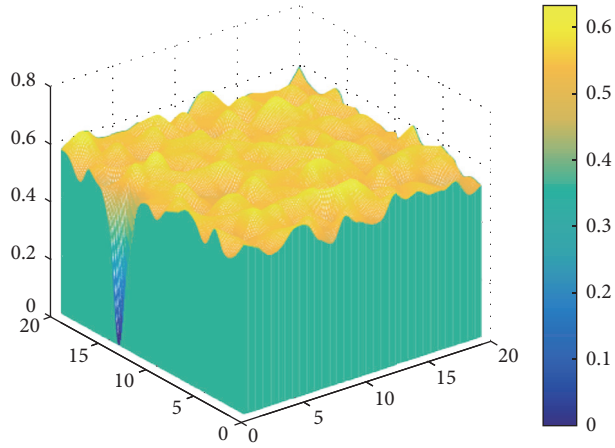


Figure 47: The achieved quality of RSDG.

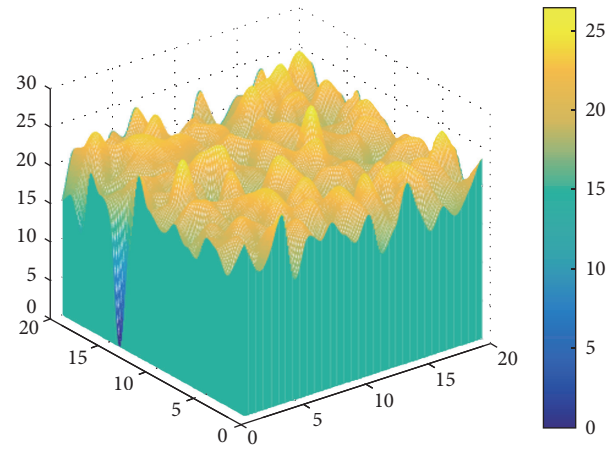


Figure 48: The achieved payment of RSDG.

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

And next, the control variate method is used to compare the differences between QUADG strategy, DDODG strategy, and RSDG strategy. Figures 49 and 50 are the results obtained under the conditions of fixed payment and changing quality. In Figures 49 and 50, the data quality

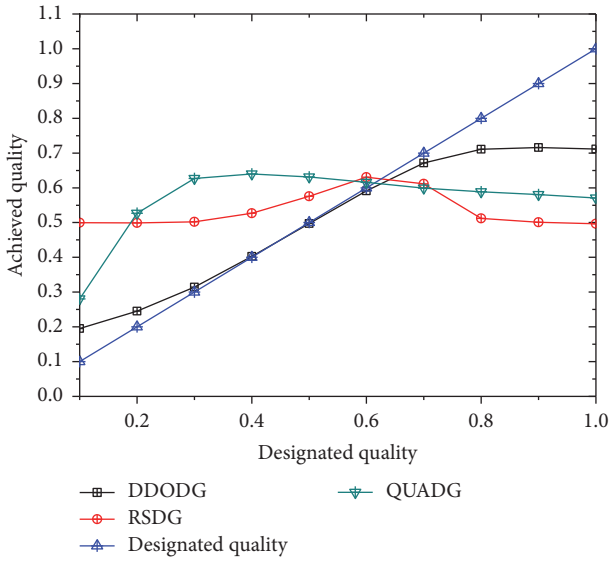


Figure 49: The achieved quality by designating quality.

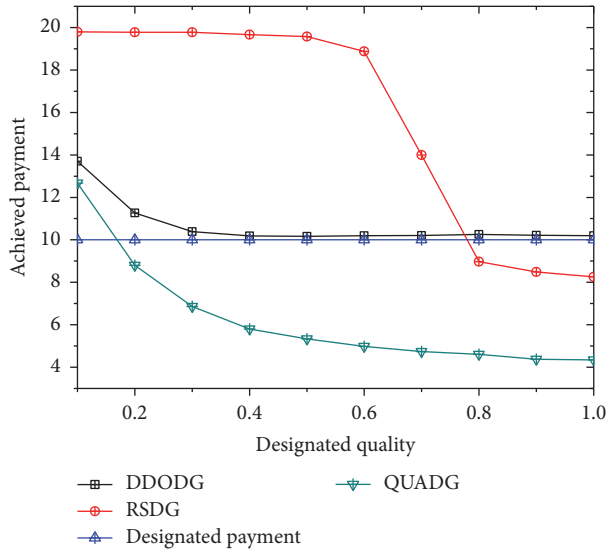


Figure 50: The achieved payment by designating quality.

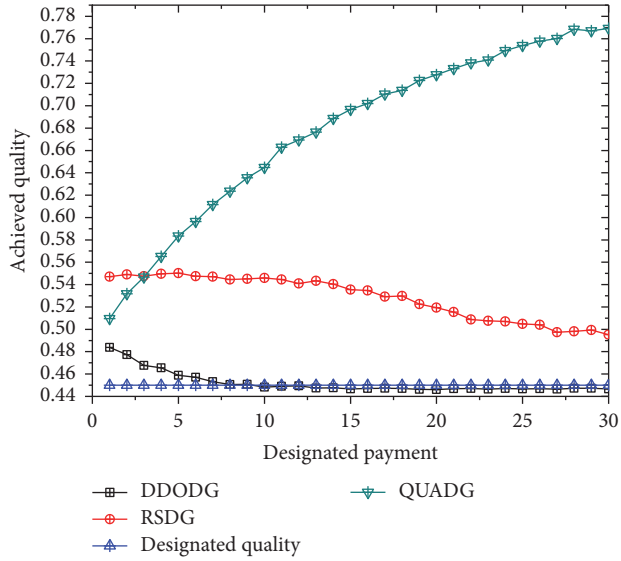


Figure 51: The achieved quality by designating payment.

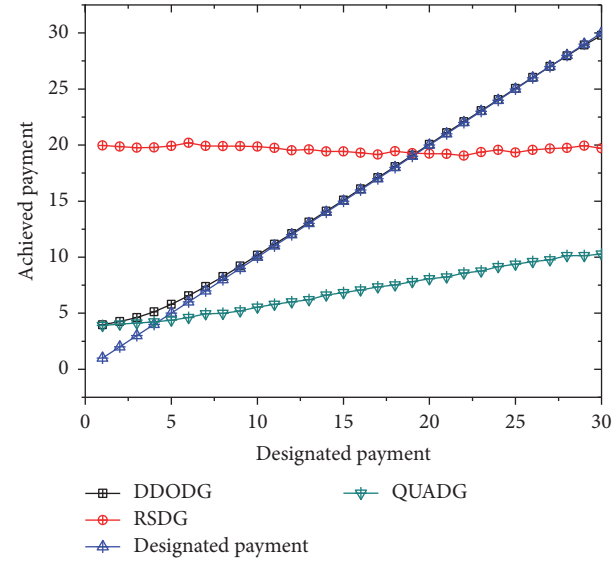


Figure 52: The achieved payment by designating payment.

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

6. Conclusion

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

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

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

Parameter Description

D_i: The data set collected in grid area i, $\forall i \in A$
$r_{i,t}^{j,A}$: The Quality Utilization of data collected by data reporter j in grid area i, $\forall i \in A$ at time t
$b_{i,t}^{\mathscr{A},j}$: The bid of data collected by data reporter j in grid area i, $\forall i \in A$ at time t
$r_{i,t}^{\mathscr{A},j}$: The quality of data collected by data reporter j in grid area i, $\forall i \in A$ at time t
$\mathfrak{R}^{\mathscr{A}}$: The designated data amount of application $\mathscr{A}$ by data demander
$\mathscr{P}^{\mathscr{A}}$: The designated payment of application $\mathscr{A}$ by data demander
$\mathscr{L}^{\mathscr{A}}$: The designated quality of application $\mathscr{A}$ by data demander
$\mathbb{R}_i$: The ordered list of Quality Utilization in grid i, $\forall i \in A$
$\mathbb{D}_i$: The list of data in grid i, $\forall i \in A$ ordered by the Quality Utilization
$\mathscr{O}$: The satisfaction of data demander towards the QUADG strategy
τ_m^i: The satisfaction of data demander towards the single data
$\mathbb{W}$: The winning data set.

Conflicts of Interest

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

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (61772554, 61370229, and 61370178), the National Basic Research Program of China (973 Program) (2014CB046305), the Science and Technology Projects of Guangdong Province, China (2016B010109008 and 2016B030305004), and the Science and Technology Projects of Guangzhou Municipality, China (201604010054 and 201604016019).

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[62] X. Fang, J. Tang, D. Yang, and G. Xue, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking (MobiCom '12)*, pp. 173–184, August 2012.

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

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

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

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

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

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

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 4859835, 13 pages
https://doi.org/10.1155/2017/4859835

Research Article

Performance Analysis of Space Information Networks with Backbone Satellite Relaying for Vehicular Networks

Jian Jiao, Houlian Gao, Shaohua Wu, and Qinyu Zhang

Communication Engineering Research Centre, Harbin Institute of Technology, Shenzhen, Guangdong, China

Correspondence should be addressed to Shaohua Wu; hitwush@hit.edu.cn

Received 9 September 2017; Accepted 15 November 2017; Published 10 December 2017

Academic Editor: Tao Han

Space Information Network (SIN) with backbone satellites relaying for vehicular network (VN) communications is regarded as an effective strategy to provide diverse vehicular services in a seamless, efficient, and cost-effective manner in rural areas and highways. In this paper, we investigate the performance of SIN return channel cooperative communications via an amplify-and-forward (AF) backbone satellite relaying for VN communications, where we assume that both of the source-destination and relay-destination links undergo Shadowed-Rician fading and the source-relay link follows Rician fading, respectively. In this SIN-assisted VN communication scenario, we first obtain the approximate statistical distributions of the equivalent end-to-end signal-to-noise ratio (SNR) of the system. Then, we derive the closed-form expressions to efficiently evaluate the average symbol error rate (ASER) of the system. Furthermore, the ASER expressions are taking into account the effect of satellite perturbation of the backbone relaying satellite, which reveal the accumulated error of the antenna pointing error. Finally, simulation results are provided to verify the accuracy of our theoretical analysis and show the impact of various parameters on the system performance.

1. Introduction

Nowadays, the connected vehicles paradigm is to form a vehicular network (VN) to communicate with the surrounding environment and the VN plays a vital role in the next generation intelligent transportation system (ITS) [1]. Generally, the long-term evolution (LTE) can provide reliable access to the Internet for VN communications in the urban areas. However, LTE network has poor coverage in rural areas and highways due to the costly network infrastructure [2, 3]. Moreover, the high mobility of vehicles can suffer from frequent handovers as the networks become even denser.

Space Information Network (SIN) is regarded as an effective strategy to provide diverse vehicular services in a seamless, efficient, and cost-effective manner in rural areas and highways. For instance, satellites and high altitude platforms (HAPs) in SIN can help achieve ubiquitous coverage in rural areas. Further, they can provide road information and transport information to assist ITS, entertainment services dissemination as relays, and relieve the demands on terrestrial networks through data offloading [4].

The return channels of the low/medium Earth orbit (L/MEO) satellites are unstable and discontinuous intrinsically to the ground-based stations and vehicles, which limit the throughput as well as the delay sensitive services of SIN-assisted VN communications. Recently, high throughput backbone satellites (such as the Ka/Q/V-band geostationary Earth orbit (GEO) satellites) relaying for SIN communications are regarded as an effective strategy to improve the continuity of return channels as well as the throughput performance.

Theoretically, three GEO satellites which are 120° apart in the SIN backbone networks can provide coverage of the space between Earth ground and GEO orbit and achieve high-speed data relay through the intersatellite and satellite-terrestrial millimeter/terahertz/laser links.

With the development of high throughput satellites (HTS), several GEO HTS can establish the backbone network of SIN, where the backbone HTS relaying for SIN-assisted VN is able to provide a global seamless broadband transmission by developing the intersatellite links. People believe that the SIN will enable a "terabit data rate capacity" broadband

access, which was previously possible only with fiber-optic links, and offer the access availability of "anywhere and anytime" inherent to the satellites [5]. Furthermore, the SIN will be a significant enabling factor as well as an important component of the upcoming 5th-generation (5G) networks [6].

Therefore, considering the backbone HTS relaying communication undergoes the large-scale and complex SIN dual-hop channel properties, such as rain attenuation [7], solar scintillation [8], perturbation factors [9], and interference [10–13], this paper investigates the performance of SIN return channel cooperative communications via an amplify-and-forward (AF) backbone satellite relaying for VN communications.

1.1. Background and Motivation. In our SIN communication scenario, space-based nodes (i.e., source nodes, like space mission explorers, orbiters and landers, space stations, spacecraft, manned and unmanned aircraft, etc.) can establish cooperative communications via an AF backbone HTS relaying.

Recently, SINs have attracted considerable research interest, and substantial effort has been devoted to investigating the performance of the research works of the hybrid satellite-terrestrial cooperative/relay networks (HSTC/RNs) by analyzing the complex multihop channel models. For that, by applying maximal ratio combining (MRC) at the destination, [14, 15] studied the outage probability (OP) performance of HSTCNs with an AF relaying protocol. In [16], the decode-and-forward (DF) relaying protocols for HSTCNs was investigated. Further, with the help of the moment generating function (MGF), [14, 15, 17] have presented the analytical expression of average symbol error rate (ASER) for HSTCNs with an AF relaying protocol. Besides, the performance of optimal selection algorithm of multiple relays for HSTRNs was presented in [18, 19].

Moreover, to achieve higher system capacity and energy efficiency, multiantenna technique was investigated in [20–22], and HTS with Ka/Q/V-band frequency with multiple antennas have attracted significant attention [23]. Authors in [24–31] investigate the performance of relay-based multiple antenna HSTC/RNs, since relay transmissions can effectively improve the throughput and the coverage of satellite communications. Further, the cognitive radio (CR) needs to be investigated since the HTS already suffer from spectrum scarcity in Ka band [32].

Besides, the SINs backbone GEO satellites are subjected to various satellite perturbation forces (e.g., Earth oblateness perturbation, third-body gravitational perturbation, atmospheric perturbation, and solar perturbation), which leads to position drift and result in the beam center of the ground station antenna unfocused [33]. The accumulated error of the antenna pointing error will cause the satellite elevation error, which may deteriorate the signal-to-noise ratio (SNR), decrease link margin [34] and bit error rate (BER) [35], and so forth. To the best of our knowledge, this is the first work on GEO satellite perturbation that reveals the effect of satellite elevation error for SIN backbone satellite relaying.

1.2. Contributions and Novelty. In this paper, we investigate the performance of SIN return channel cooperative communications via an AF backbone satellite relaying for VN, where both of the source-destination and relay-destination links undergo Shadowed-Rician fading, and the source-relay link follows Rician fading, respectively. By applying MRC at the destination, the equivalent end-to-end SNR of the system is first obtained, and then analytical expressions as well as the satellite perturbation effect are derived to evaluate the system performance. The detailed contributions of this paper are outlined as follows:

(i) The system model of SIN return channel cooperative communications via an AF backbone satellite relaying for VN is first built, and we present a new analytical expression for the approximate statistical distributions of the equivalent end-to-end SNR of system (7).

(ii) To gain further insight, the effect of the satellite perturbation of the relaying GEO satellite is considered for the first time, which reveals the accumulated error of the antenna pointing error leads to the satellite elevation error. And the accumulated satellite elevation error is taking into account the derivation of the ASER expression.

(iii) The closed-form expression for the end-to-end ASER (31) is derived, which can efficiently evaluate the system performance. Moreover, simulation results prove the rationality of our theoretical analysis.

Notations. X-Y describes the link from node X to node Y. X-Y-Z represents the dual-hop link from node X to node Z through relay node Y. $\mathrm{E}[\cdot]$ denotes the expectation operator. $N(\mu, \sigma^2)$ denotes a complex Gaussian distribution with mean μ and variance σ^2. $\exp(\cdot)$ represents the exponential function. $M_\lambda(\cdot)$ denotes the moment generating function (MGF) of γ. $f_x(\cdot)$ and $F_x(\cdot)$ denote the probability distribution function (PDF) and cumulative distribution function (CDF) of x, respectively. ${}_1F_1(a; b; c)$ represents the confluent hypergeometric function of first kind [36, Eq. (9.210.1)]. $F(a, b; c; d)$ is Gauss hypergeometric function [36, Eq. (9.100)], and $K_n(\cdot)$ represents the modified Bessel function of the second kind with order n [36, Eq. (8.446)]. $M_{a,b}(\cdot)$ is the Whittaker function defined as [36, Eq. (9.220.2)].

2. System Model

Our system model of the SIN return channel cooperative communications via an AF GEO HTS relaying for VN is considered as shown in Figure 1, where a source node (S), that is, space node, communicates with a terrestrial destination (D) via a GEO HTS relay (R) and h_0, h_1, and h_2 are the channel gains of the S-R, S-D, and R-D links, respectively.

The space node S is generally on the stratosphere layer and above, and R is a GEO HTS in our SIN communication scenario. In the S-R link, since the line of sight (LOS) signal is much stronger than the others, which is different from terrestrial networks, the channel gain h_0 of the S-R link is considered as a Rician fading with additive white Gaussian noise (AWGN) [37, 38]. On the other hand, the channel gains

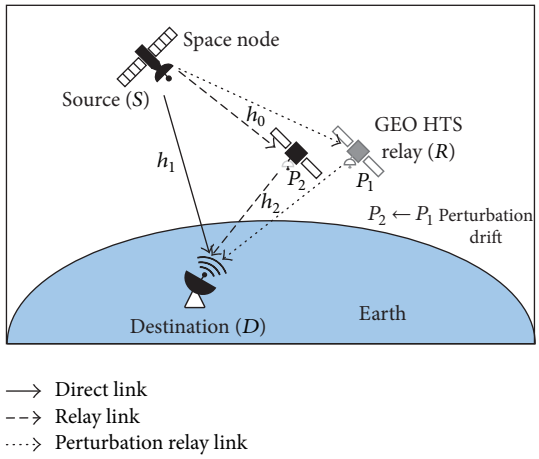


Figure 1: The proposed system model of SIN return channel cooperative communications via an AF GEO HTS relaying for VN, where each node is equipped with a single antenna, the relay point P_1 with an arrow point to P_2 shows the satellite drift effect caused by the satellite perturbation.

h_1 and h_2 of satellite-terrestrial links S-D and R-D are usually modeled by Shadowed-Rician fading distribution [14–16, 24–27, 35]. It approaches the LOS communication using the Rician fading, whereas the amplitude is Nakagami-m distributed [39], and it sufficiently agrees with experimental data and is computationally less complex than other land mobile satellite channel models.

As illustrated in Figure 1, in such a backbone GEO HTS relaying SIN-assisted VN system, the communication occurs during two time phases. In the first time phase, the space node S broadcasts its signal to the relay R and the destination D, where h_0 and h_1 are the channel gains of the S-R and S-D links, respectively. The received signals at the relay y_0 and the destination y_1 from S are given by

$$y_0 = \sqrt{E_1} h_0 x + n_0, \tag{1}$$

$$y_1 = \sqrt{E_1} h_1 x + n_1, \tag{2}$$

where x is the transmitted signal with unit power, E_1 is the transmitted power at S, and n_0 and n_1 are the AWGN of S-R and S-D links with zero mean and variance σ_0^2 and σ_1^2, respectively.

During the second time phase, R first amplifies the received signal y_0 by an amplifying factor G and then forwards it to D through R-D link of which the channel gain is h_2, and the received signal at the destination y_2 is given by

$$\begin{aligned} y_2 &= \sqrt{E_2} G h_2 y_0 + n_2 \\ &= G\sqrt{E_1 E_2} h_2 h_0 x + G\sqrt{E_2} h_2 n_0 + n_2, \end{aligned} \tag{3}$$

where E_2 is the transmit power at R and n_2 is in AWGN at D obeying $n_2 \sim N(0, \sigma_2^2)$.

Assuming that perfect channel state information (CSI) is available at D and R and MRC is applied at the destination, thus, the end-to-end SNR at D can be expressed as

$$\gamma_{e2e} = \gamma_1 + \gamma_{02}, \tag{4}$$

where γ_1 is the SNR of S-D link and γ_{02} is the SNR of S-R-D link. From (2), we have

$$\gamma_1 = \frac{E_1 |h_1|^2}{\sigma_1^2} = \rho_1 |h_1|^2. \tag{5}$$

From (3), γ_{02} can be expressed as

$$\gamma_{02} = \frac{E_1 |h_0|^2 / \sigma_0^2 \cdot E_2 |h_2|^2 / \sigma_2^2}{E_2 |h_2|^2 / \sigma_2^2 + 1/G^2 \sigma_0^2} = \frac{\gamma_0 \gamma_2}{\gamma_2 + C}, \tag{6}$$

where $C = 1/(G\sigma_0)^2$ and $\gamma_0 = E_1|h_0|^2/\sigma_0^2 = \rho_0|h_0|^2$ and $\gamma_2 = E_2|h_2|^2/\sigma_2^2 = \rho_2|h_2|^2$ denote the SNR of S-R and R-D link, respectively. Thus, (4) can be rewritten as

$$\gamma_{e2e} = \frac{\gamma_0 \gamma_2}{\gamma_2 + C} + \rho_1 |h_1|^2. \tag{7}$$

2.1. Satellite Perturbation. In this paper, the GEO HTS satellite is considered as backbone relaying node for SIN-assisted VN communications to enhance the continuity as well as the throughput of the return channel. This is the first work to analyze the performance of cooperative communication for SIN dual-hop channel properties. To gain further insight, the effect of the satellite perturbation of the GEO HTS is analyzed, which reveals that the accumulated error of the antenna pointing error leads to the satellite elevation error.

2.1.1. Principle and Law of Satellite Perturbation Drift. The satellite is always subjected to a variety of perturbation forces, especially to the GEO satellites, which will lead to perturbation drift and accumulate the antenna pointing error. The satellite perturbation forces include the Earth oblateness perturbation [40, 41], the third-body attraction perturbation [42] such as lunisolar gravitational perturbation [43], the solar radiation pressure perturbation [44, 45], and the atmospheric drag perturbation.

The Earth oblateness perturbation is caused by the facts that the Earth is not an ideal sphere and it has uneven internal density distribution. It affects the long-term change of the right ascension of ascending node (RAAN) r and argument of perigee ω of satellite orbit. The lunisolar gravitational perturbation can reduce the satellite orbit radius and may increase the orbital inclination, while the semimajor axis changes with half-day cycle. On the contrary, the solar radiation pressure perturbation mainly affects the orbital eccentricity, which directly determines the satellite center distance and the satellite height.

2.1.2. Satellite Elevation Error. The diagram of antenna pointing error is shown in Figure 2, and let $\theta = \theta_e + \theta_c$ denote the elevation of a GEO HTS, where θ_e represents the elevation

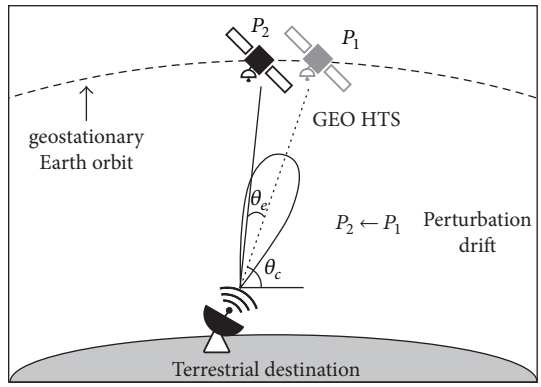


Figure 2: The diagram of antenna pointing error caused by satellite perturbation, P_1 means the satellite position located in antenna beam center, and P_2 represents the satellite position offsets from antenna beam center.

error, which is the angle between centerline of antenna beam pointed to P_1 and the real LOS channel between the actual position P_2 of the satellite to the destination. θ_c denotes the satellite elevation angle if the GEO satellite is not affected by the satellite perturbation. Considering the drift caused by satellite perturbation in the eastwest and northsouth directions, the elevation of satellite θ can be calculated by [46]

$$\theta = \arctan\left[\frac{\sin\Delta\varphi\sin\phi + \cos\Delta\varphi\cos\phi\cos(\lambda+\Delta\lambda) - 0.151}{\cos\left\{\arcsin\left[\sin\Delta\varphi\sin\phi + \cos\Delta\varphi\cos\phi\cos(\lambda+\Delta\lambda)\right]\right\}}\right],$$
$$\theta_c = \arctan\left[\frac{\cos\phi\cos\lambda - 0.151}{\cos\left[\arcsin\left(\cos\phi\cos\lambda\right)\right]}\right]_{\Delta\varphi=0}, \tag{8}$$

where ϕ is the latitude of the destination, $\lambda = \lambda_{\text{sat}} - \lambda_{\text{des}}$ is the longitude difference between the subsatellite point and the ground station, λ_{sat} represents the longitude of subsatellite point, λ_{des} is the longitude of destination, $\Delta\varphi$ is the drift of satellite in the northsouth direction, and $\Delta\lambda$ is the drift of satellite in the eastwest direction. Therefore, the elevation error θ_e can be calculated as $\theta_e = \theta - \theta_c$.

In general, λ_{des} can be calculated by the six elements of the satellite orbit [47]. When the satellite orbit is elliptical, the longitude (L_{lon}) and the latitude (L_{lat}) of the subsatellite point can be expressed as

$$L_{\text{lon}} = r + \arctan\left\{\cos i \cdot \tan\left[\omega + 2\arctan\left(\sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2}\right)\right]\right\} - \omega_e\left(t - t_N\right)$$
$$L_{\text{lat}} = \arcsin\left\{\sin i \cdot \sin\left[\omega + 2\arctan\left(\sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2}\right)\right]\right\}, \tag{9}$$

where ω_e is the average angular velocity of Earth, t is the time, t_N is the time when the satellite passes the ascending node, E is the eccentric anomaly, i is the inclination of satellite orbit, and e is eccentricity. When the satellite orbit is a circle, (9) can be simplified as

$$L_{\text{lon}} = r + \arctan\left\{\cos i\tan\left[\omega + \frac{2\pi}{T}(t-\tau)\right]\right\} - \omega_e\left(t - \tau + \frac{T\omega}{2\pi}\right)$$
$$L_{\text{lat}} = \arcsin\left\{\sin i\sin\left[\omega + \frac{2\pi}{T}(t-\tau)\right]\right\}, \tag{10}$$

where τ is the time when the satellite perigee passes.

Therefore, from (9) and (10), it is worth noting that the subsatellite point is related to the parameters r, i, ω, and τ. If these variables have been affected by the satellite perturbation, the subsatellite point will have drift. Let L_p^{lon} and L_p^{lat} represent the longitude and latitude of subsatellite point considering the effect of satellite perturbation, respectively, and L_{np}^{lon} and L_{np}^{lat} denote the longitude and latitude of unperturbed subsatellite point, respectively. It is clear that the drift of subsatellite point affected by satellite perturbation causes the elevation error of satellite, and we have

$$\Delta\varphi = L_p^{\text{lon}} - L_{np}^{\text{lon}}$$
$$\Delta\lambda = L_p^{\text{lat}} - L_{np}^{\text{lat}}. \tag{11}$$

2.2. Channel Model

2.2.1. S-R Link. As the *S-R* link is modeled by using the Rician fading distribution, the probability distribution function (PDF) of γ_0 is given by [48]

$$f_{\gamma_0}(x) = \frac{(K+1)}{\rho_0\Omega}\exp\left(-\frac{x(K+1)}{\rho_0\Omega} - K\right) \cdot I_0\left(2\sqrt{\frac{K(K+1)x}{\rho_0\Omega}}\right), \tag{12}$$

where Ω is the average power of received signal at R and $\Omega = (A^2 + 2\sigma^2)/2$ and $K = A^2/(2\sigma^2)$ are known as the Rician factor, A is the amplitude of LOS signal, σ^2 is the average power of multipath component, and $I_0(\cdot)$ is the modified Bessel function of the first kind with order zero.

Then, the cumulative distribution function (CDF) of γ_0 can be expressed as

$$F_{\gamma_0}(x) = 1 - Q_1\left(\sqrt{2K}, \sqrt{\frac{2(K+1)x}{\rho_0\Omega}}\right), \tag{13}$$

where $Q_M(\cdot,\cdot)$ is the Marcum Q-function of the Mth order, which is defined as

$$Q_M(a,b) = \int_b^\infty x\left(\frac{x}{a}\right)^{M-1}\exp\left(-\frac{x^2+a^2}{2}\right)I_{M-1}(ax)\,dx. \tag{14}$$

Moreover, the relationship between the Rician factor K of the Rician fading *S-R* link (12) and the satellite elevation (θ_c) was simulated in [49] through a large number of experiments. Then the relationship between the Rician factor and the satellite elevation ($\theta_c \in [10^\circ, 90^\circ]$) was fitted as an empirical formula

$$K(\theta_c) = K_0 + K_1(\theta_c) + K_2(\theta_c), \tag{15}$$

where K_0, K_1, and K_2 are empirical constant and $K_0 = 2.731$, $K_1 = -1.074 \times 10^{-1}$, and $K_2 = 2.771 \times 10^{-3}$, respectively. Considering the effect of satellite perturbation, (15) can be rewritten as

$$K(\theta_c, \theta_e, t) = K_0 + K_1(\theta_c + \theta_e \sin t) + K_2(\theta_c + \theta_e \sin t), \tag{16}$$

where $\theta_e \sin t$ indicates the elevation error θ_e affected by the satellite perturbation at time t.

2.2.2. S-D and R-D Links. The *S-D* link is usually modeled as a composite fading distribution to describe the amplitude fluctuation of the signal envelope. Considering the tradeoffs between accuracy and computational complexity, the satellite-destination *S-D* link and the relay-destination *R-D* link are modeled by using the Shadowed-Rician fading distribution [11–13] in our SIN-assisted VN system.

Let γ_1 and γ_2 denote the SNR of the *S-D* and *R-D* links, respectively. The PDF of γ_i ($i = 1, 2$) is given by [39]

$$f_{\gamma_i(x)} = \frac{\alpha_i}{\rho_i}\exp\left(-\frac{\beta_i x}{\rho_i}\right) \cdot {}_1F_1\left(m_i; 1; \frac{\delta_i}{\rho_i}x\right), \tag{17}$$

where

$$\alpha_i = \frac{\left(2b_i m_i / \left(2b_i m_i + \Omega_i\right)\right)^{m_i}}{2b_i}, \quad \beta_i = \frac{1}{2b_i}, \quad \delta_i = \frac{0.5\Omega_i}{\left(2b_i^2 m_i + b_i\Omega_i\right)}, \tag{18}$$

where ${}_1F_1(\cdot;\cdot;\cdot)$ is the confluent hypergeometric function of first kind [36, Eq. (9.210.1)]. Moreover, Ω_i and $2b_i$ are the average power of the LOS and the multipath components, respectively, and m_i ($m_i \in [0, \infty)$) is the fading severity parameter.

Recall the definition of ${}_1F_1(\cdot;\cdot;\cdot)$, and we have

$${}_1F_1\left(m_i; 1; \frac{\delta_i}{\rho_i}x\right) = \sum_{n=0}^{\infty}\frac{(m_i)_n}{(1)_n}\frac{1}{n!}\left(\frac{\delta_i}{\rho_i}x\right)^n. \tag{19}$$

For the analytical tractability, we retain our focus in the case when the channel severity parameters take integer values in the rest of this paper; that is, $m_i \in \mathrm{N}$. Hence, with the aid of [50, Eq. (07.20.03.0009.01), Eq. (07.02.03.0014.01)], (19) becomes

$${}_1F_1\left(m_i; 1; \frac{\delta_i}{\rho_i}x\right) = \exp\left(\frac{\delta_i}{\rho_i}x\right)\sum_{n=0}^{m_i-1}\frac{(-1)^n\left(1-m_i\right)_n}{(n!)^2}\left(\frac{\delta_i}{\rho_i}x\right)^n, \tag{20}$$

where $(z)_n = \Gamma(z+n)/\Gamma(z)$ denotes the Pochhammer symbol with $n \in N$ [51, Eq. (6.1.22)].

To solve the three parameters (Ω_i, b_i, and m_i) in (18), we assume θ_{ci} is the elevation at GEO HTS R, when the center line of the receiving antenna beam in different link aims at R. $\theta_{ei}\sin t$ is the elevation error which affects the satellite perturbation. When $\theta_{ci} \in (20^\circ, 30^\circ)$, Ω_i, b_i, and m_i can be calculated by the empirical formulas [39, Eq. (19)]. Therefore,

considering the effect of satellite perturbation, Ω_i, b_i, and m_i can be calculated as follows:

$$\begin{aligned} b_i(\theta_{ci},\theta_{ei},t) &= -4.7943\times10^{-8}(\theta_{ci}+\theta_{ei}\sin t)^3 \\ &\quad + 5.5784\times10^{-6}(\theta_{ci}+\theta_{ei}\sin t)^2 \\ &\quad - 2.1344\times10^{-4}(\theta_{ci}+\theta_{ei}\sin t) \\ &\quad + 3.2710\times10^{-2}, \\ m_i(\theta_{ci},\theta_{ei},t) &= 6.3739\times10^{-5}(\theta_{ci}+\theta_{ei}\sin t)^3 \\ &\quad + 5.8533\times10^{-4}(\theta_{ci}+\theta_{ei}\sin t)^2 \\ &\quad - 1.5973\times10^{-4}(\theta_{ci}+\theta_{ei}\sin t) \\ &\quad + 3.5156, \\ \Omega_i(\theta_{ci},\theta_{ei},t) &= 1.4428\times10^{-5}(\theta_{ci}+\theta_{ei}\sin t)^3 \\ &\quad - 2.3798\times10^{-3}(\theta_{ci}+\theta_{ei}\sin t)^2 \\ &\quad + 1.2702\times10^{-1}(\theta_{ci}+\theta_{ei}\sin t) \\ &\quad - 1.4864. \end{aligned} \tag{21}$$

3. Performance Analysis

In order to exactly measure the effect of satellite perturbation on the SIN return channel cooperative communications via an AF GEO HTS relaying, the important quality-of-service (QoS) metric, that is, average symbol error probability (ASER), is analytically studied and evaluated in our proposed SIN-assisted VN systems.

Since MRC is applied at the destination, we derive the close-form expression by using MGF according to [52], where the ASER of an M-ary phase-shift keying (MPSK) modulated system is given by

$$P_{\mathrm{MPSK}} = \frac{1}{\pi}\int_0^{\theta_M} M_{\gamma_{e2e}}\left(\frac{g_{\mathrm{MPSK}}}{\sin^2\theta}\right), \tag{22}$$

where $\theta_M = \pi(M-1)/M$ and $g_{\mathrm{MPSK}} = \sin^2(\pi/M)$. The MGF of instantaneous SNR (γ) is defined as

$$M_\gamma = E_\gamma\{e^{-s\gamma}\} = \int_0^\infty e^{-s\gamma} f_\gamma(\gamma)\, d\gamma. \tag{23}$$

Considering h_0, h_1, and h_2 are independent and the relationship between γ_1, γ_{02}, and γ_{e2e} is presented in (4), we can express $M_{e2e}(s)$ as

$$M_{\gamma_{e2e}}(s) = M_{\gamma_1}(s)\, M_{\gamma_{02}}(s), \tag{24}$$

where $M_{\gamma_1}(s)$ and $M_{\gamma_{02}}(s)$ are the MGF of γ_1 and γ_{02}.

In the following, we derive the expressions for $M_{\gamma_1}(s)$ and $M_{\gamma_{02}}(s)$. Then, we use (22) and (24) to obtain the ASER of our SIN backbone satellite relaying VN system.

3.1. MGF of the SNR for the S-D Link. By using the definition of MGF and substituting (17) into (23), we can evaluate the MGF of the *S-D* link as follows

$$\begin{aligned} M_{\gamma_1}(s) &= \mathrm{E}\{e^{-sx}\} \\ &= \int_0^\infty e^{-sx}\frac{\alpha_1}{\rho_1}e^{-(\beta_1/\rho_1)x}\cdot {}_1F_1\left(m_1;1;\frac{\delta_1}{\rho_1}x\right)dx \\ &= \frac{\alpha_1}{\rho_1}\int_0^\infty e^{-(\beta_1/\rho_1+s)x}\cdot {}_1F_1\left(m_1;1;\frac{\delta_1}{\rho_1}x\right)dx. \end{aligned} \tag{25}$$

By using [36, Eq. (7.621.4)], the result of (25) can be easily obtained as follows:

$$M_{\gamma_1}(s) = \frac{\alpha_1}{\rho_1 s+\beta_1}F\left(m_1,1;1;\frac{\delta_1}{\rho_1 s+\beta_1}\right). \tag{26}$$

3.2. MGF of the SNR for the S-R-D Link. By substituting γ_{02} in (6) into (23), we can evaluate the MGF of cooperative link as presented in (27). The proof is provided in the Appendix.

$$\begin{aligned} M_{\gamma_{02}}(s) &= \frac{u_2(s)\,\alpha_2 e^{-K+u_2(s)+((\beta_2-\delta_2)/\rho_2)u_1(s)}}{K\rho_2} \\ &\cdot \sum_{n=0}^{m_2-1}\frac{(-1)^n(1-m_2)_n}{(n!)^2}\left(\frac{\delta_2}{\rho_2}\right)^n\sum_{k=0}^{n}\left[-u_1(s)\right]^{n-k} \\ &\cdot \left[2\left(\frac{\varepsilon\rho_2 u_2(s)}{\beta_2-\delta_2}\right)^{(k+1)/2}\right. \\ &\cdot K_{k+1}\left(2\sqrt{\frac{\varepsilon u_2(s)(\beta_2-\delta_2)}{\rho_2}}\right) \\ &\left.+\,2\varepsilon\left(\frac{\varepsilon\rho_2 u_2(s)}{\beta_2-\delta_2}\right)^{k/2}K_k\left(2\sqrt{\frac{\varepsilon u_2(s)(\beta_2-\delta_2)}{\rho_2}}\right)\right], \end{aligned} \tag{27}$$

where the definitions $u_1(s)$, $u_2(s)$, and ε are shown in (A.7).

3.3. Derivation of ASER. Based on the definition of ASER of an MPSK modulated system (i.e., (27) and (24)), (22) can be written as

$$\begin{aligned} P_{\mathrm{MPSK}} &= \frac{1}{\pi}\int_0^{\theta_M} M_{\gamma_{e2e}}\left(\frac{g_{\mathrm{MPSK}}}{\sin^2\theta}\right)d\theta \\ &= E_\gamma\left[\frac{1}{\pi}\int_0^{\pi/2}\exp\left(\frac{g_{\mathrm{MPSK}}}{\sin^2\theta}\right)d\theta\right. \\ &\quad \left.+\,\frac{1}{\pi}\int_{\pi/2}^{\theta_M}\exp\left(\frac{g_{\mathrm{MPSK}}}{\sin^2\theta}\right)d\theta\right] \\ &= \frac{1}{\pi}\int_0^{\theta_M} M_{\gamma_1}\left(\frac{g_{\mathrm{MPSK}}}{\sin^2\theta}\right)M_{\gamma_{02}}\left(\frac{g_{\mathrm{MPSK}}}{\sin^2\theta}\right)d\theta. \end{aligned} \tag{28}$$

Alternatively, the following approximation of (28) can be used [37]:

$$P_{\mathrm{MPSK}} = \frac{1}{\pi}\int_0^{\theta_M} M_{\gamma_1}\left(\frac{g_{\mathrm{MPSK}}}{\sin^2\theta}\right) M_{\gamma_{02}}\left(\frac{g_{\mathrm{MPSK}}}{\sin^2\theta}\right) d\theta \approx \sum_{j=1}^{3} k_j M_{\gamma_1}\left(w_j\right) M_{\gamma_{02}}\left(w_j\right), \tag{29}$$

where

$$\begin{aligned} k_1 &= \frac{\theta_M}{2\pi} - \frac{1}{6}, \\ w_1 &= -g_{\mathrm{MPSK}}, \\ k_2 &= \frac{1}{4}, \\ w_2 &= -\frac{4}{3} g_{\mathrm{MPSK}}, \\ k_3 &= \frac{\theta_M}{2\pi} - \frac{1}{4}, \\ w_3 &= -\frac{g_{\mathrm{MPSK}}}{\sin^2\theta}. \end{aligned} \tag{30}$$

As shown in (26) and (27), the close forms of $M_{\gamma_1}(s)$ and $M_{\gamma_{02}}(s)$ have already been derived. By substituting these close forms into (29), we finally obtain the accurate closed-form expression of the ASER of a space downlink cooperative transmission system with relay GEO satellite as shown in

$$\begin{aligned} P_{\mathrm{MPSK}} = {} & \sum_{j=1}^{3} k_j \frac{\alpha_1}{\rho_1 w_j + \beta_1} F\left(m_1, 1; 1; \frac{\delta_1}{\rho_1 w_j + \beta_1}\right) \\ & \cdot \frac{u_2\left(w_j\right)\alpha_2 e^{-K+u_2(w_j)+((\beta_2-\delta_2)/\rho_2)u_1(w_j)}}{K\rho_2} \\ & \times \sum_{n=0}^{m_2-1} \frac{(-1)^n \left(1-m_2\right)_n}{(n!)^2} \left(\frac{\delta_2}{\rho_2}\right)^n \sum_{k=0}^{n} \left[-u_1\left(w_j\right)\right]^{n-k} \\ & \times \left[2\left(\frac{\varepsilon\rho_2 u_2\left(w_j\right)}{\beta_2-\delta_2}\right)^{(k+1)/2} \right. \\ & \cdot K_{k+1}\left(2\sqrt{\frac{\varepsilon u_2\left(w_j\right)\left(\beta_2-\delta_2\right)}{\rho_2}}\right) \\ & \left. + 2\varepsilon\left(\frac{\varepsilon\rho_2 u_2\left(w_j\right)}{\beta_2-\delta_2}\right)^{k/2} \right. \\ & \left. \cdot K_k\left(2\sqrt{\frac{\varepsilon u_2\left(w_j\right)\left(\beta_2-\delta_2\right)}{\rho_2}}\right) \right]. \end{aligned} \tag{31}$$

Table 1: Initial position and velocity vector of GEO HTS and other system parameters.

Parameters	Description
Earth gravity model	WGS84_EGM96.grv
Satellite mass	1000 kg
Mass-area ratio of satellite	0.1 m^2/kg
Reflection coefficient of spacecraft	1.2
Solar radiation pressure model	Spherical
Third-body gravity	Sun, Moon

4. Numerical Results

This section gives the numerical results to demonstrate the validity of the theoretical analysis and the effect of satellite perturbation on the SIN return channel cooperative communications via a GEO HTS relaying.

We assume the node S is a space node and its position is $L_{\mathrm{lat}} = 10^{\circ}$N and $L_{\mathrm{lon}} = 0^{\circ}$W. The node R is a GEO HTS and, for the initial position (X, Y, and Z) and velocity vector ($\mathrm{V}x$, $\mathrm{V}y$, and $\mathrm{V}z$) in the Cartesian coordinate system, their initial values are (X, Y, and Z) = (−32299.6, −27102.6, 0) km and ($\mathrm{V}x$, $\mathrm{V}y$, and $\mathrm{V}z$) = (1.97635, −2.35533, 0) km/sec. We adopt high precision orbit propagator (HPOP) model and the parameters of the various perturbations are shown in Table 1. The GEO HTS R is mainly affected by the Earth gravity, the third-body gravity, and the solar radiation pressure perturbation [53]. In Table 1, the Earth gravity model [54], the solar radiation pressure perturbation, and the third-body gravity perturbation adopt general settings.

The elevation error affected by the satellite perturbation is shown in Figure 3, which is mainly considering the Earth nonspherical perturbation, the lunisolar gravitational perturbation, and the solar radiation pressure perturbation. The simulation duration is one lunar month and each step is 60 seconds.

As shown in Figure 3, the elevation error accumulates on the R-D link which is about 1 degree after one lunar month. The elevation error fluctuates on the S-R link, where the range of fluctuation increases gradually and the simulation time accumulates, and the maximum of fluctuation is about 0.1 degrees at the end of the simulation.

We assume $E_1 = E_2$ and $\rho_0 = \rho_1 = \rho_2 = \rho$. In S-R and R-D links, the elevations of T_X ignoring satellite perturbation at nodes R and D are equal; that is, $\theta_{c0} = \theta_{c2} = \theta_0$. In order to simulate the influence of the elevation error caused by perturbation on the system ASER performance, the elevation error data is sampled at intervals of 12 hours due to the large amount of elevation error in Figure 3. The amplifying factor is $G = 30$ and $\theta_{c1} = 60^{\circ}$ and QPSK modulation is implemented, and the rest of the simulation parameters are the same as above. The end-to-end ASER with satellite perturbation in the SIN return channel cooperative communications via an AF GEO HTS are shown in Figures 4, 5, and 6, respectively.

Figures 4 and 5 indicate the ASER performance is improving with the increasing of receiving SNR ρ. As one can expect, with the increasing of T_X and average power of received signal at node R in the S-R and R-D links, the ASER performance is

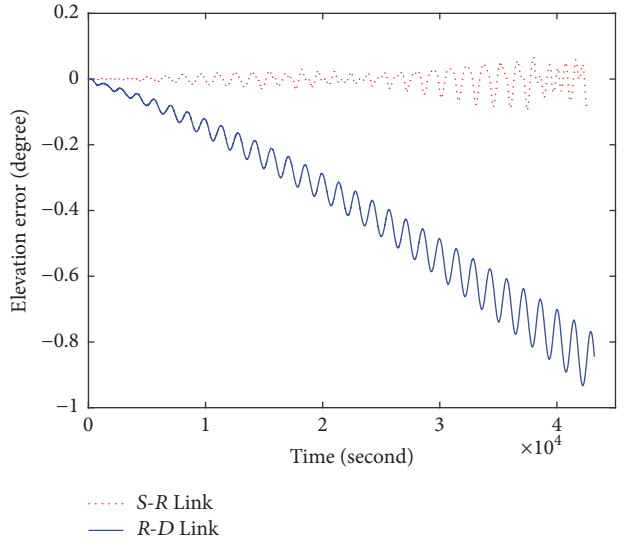


Figure 3: Elevation error caused by GEO satellite perturbation.

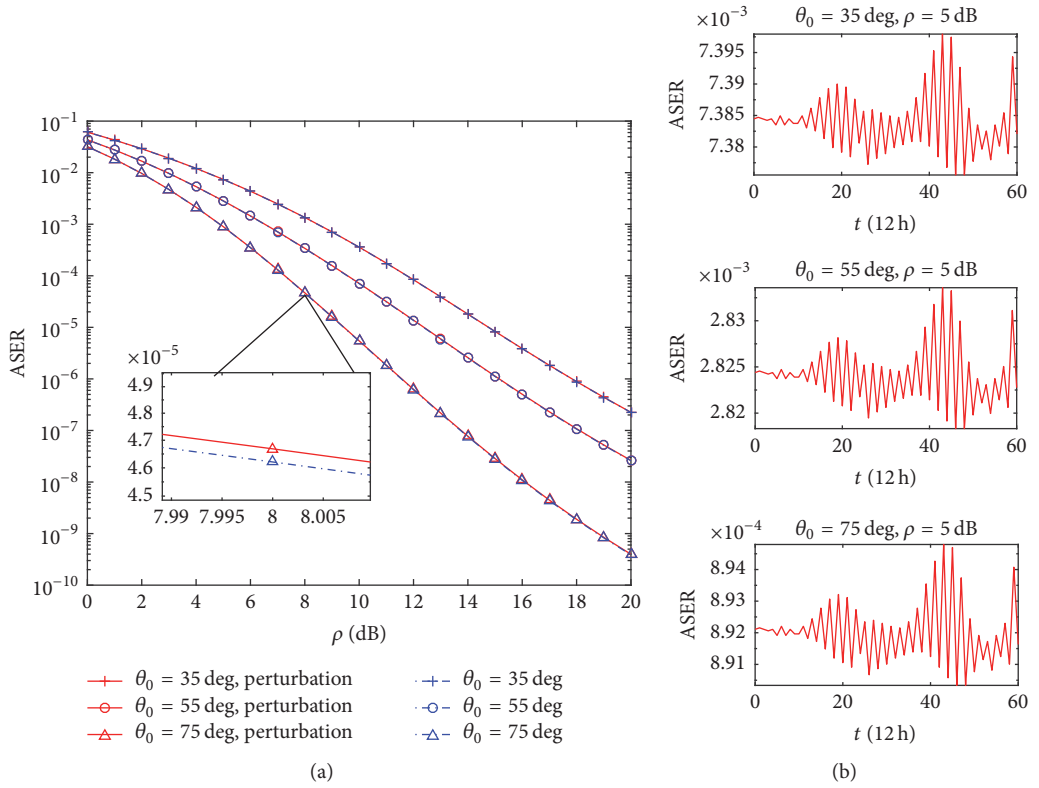


Figure 4: ASER with satellite perturbation versus various ρ; (b) shows the elevation error fluctuation and accumulation process on the ASER under different θ_0, where $\rho = 5$ dB and $\Omega = 5$ W.

improving. Comparing with the cases of ignoring the effect of the satellite perturbation, the ASER performance deteriorates a little due to the satellite perturbation.

Moreover, three subfigures in Figures 4 and 5 show the fluctuation and error accumulation process of the ASER with satellite perturbation with $\rho = 5$ dB, under different θ_0 and Ω, respectively. In Figure 4, when $\theta_0 = 35^\circ, 55^\circ$, and 75°, the fluctuation of ASER is the same, and the fluctuation ranges are $2.2379e{-}5$, $1.5260e-5$, and $4.4567e-6$, respectively, and the fluctuation range becomes tighter and θ_0 increases. Similar to Figure 5, when $\Omega = 2$ W, 5 W, and 10 W, the fluctuation of ASER is the same, and the fluctuation ranges are $2.9818e-5$, $1.5260e-5$, and $6.8832e-6$, respectively, and the fluctuation range declines when Ω increases. To be clearer, we show the

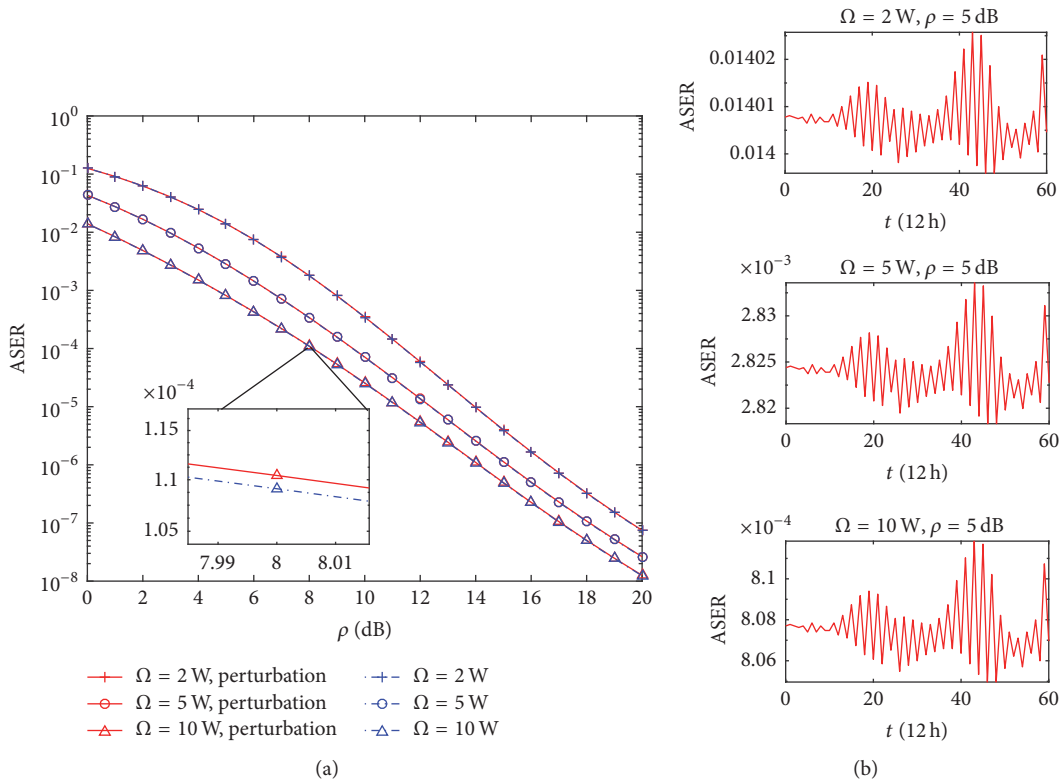


FIGURE 5: ASER with satellite perturbation versus various ρ; (b) shows the elevation error fluctuation and accumulation process on the ASER under different Ω, where $\rho = 5$ dB and $\theta_0 = 55^\circ$.

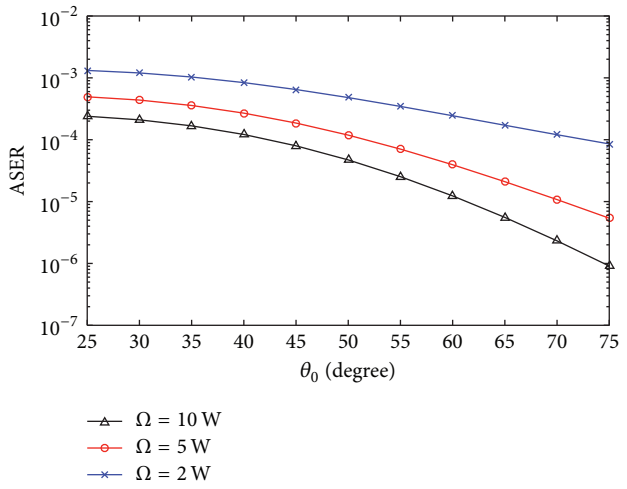


FIGURE 6: ASER performance versus various θ_0 under $\rho = 5$ dB.

ASER performance with the increasing of Ω and θ_0 as in Figure 6.

5. Conclusion

In this paper, we investigate the ASER performance of SIN return channel cooperative communications via an AF GEO HTS relaying for VN, where both of the *S-D* and *R-D* links undergo the Shadowed-Rician fading, and the *S-R* link follows Rician fading, respectively. By applying MRC at *D*, the equivalent end-to-end SNR of the system is first obtained, then the analytical expressions of ASER and the satellite perturbation effect are derived. The effect of the satellite perturbation of the relaying GEO satellite is considered for the first time, which reveals that the accumulated error of the antenna pointing error leads to the satellite elevation error. And the accumulated satellite elevation error is taking into account the derivation of the ASER expression. The closed-form expression for the end-to-end ASER can efficiently evaluate the system performance, and simulation results prove the rationality of our theoretical analysis.

Appendix

Derivation the MGF of the Cooperative Link

By the definition of MGF, from (6) and (23), the MGF of the cooperative link can be evaluated:

$$\begin{aligned} M_{\gamma_{02}} &= E\left(e^{-s\gamma_{02}}\right) \\ &= \int_0^\infty e^{-s(\gamma_0\gamma_2/(\gamma_2+C))} f_{\gamma_{02}}(\gamma)\, d\gamma \\ &= \int_0^\infty e^{-s(xy/(y+C))} f_{\gamma_0}(x)\, f_{\gamma_2}(y)\, dx\, dy. \end{aligned} \tag{A.1}$$

Considering h_0 and h_2 are independent, we first calculate the following integral of variable x such that

$$\begin{aligned} L &= \int_0^\infty e^{-s(xy/(y+C))} f_{\gamma_0}(x)\, dx \\ &= \frac{K+1}{\rho_0\Omega} e^{-K} \int_0^\infty e^{-x(sy/(y+C)+(K+1)/\rho_0\Omega)} \\ &\quad \cdot I_0\left(2\sqrt{\frac{K(K+1)x}{\rho_0\Omega}}\right) dx. \end{aligned} \tag{A.2}$$

By using [36, Eq. (6.614.3)], we get

$$\begin{aligned} L &= \frac{K+1}{\rho_0\Omega} e^{-K} \\ &\quad \cdot \frac{e^{(1/2)((K(K+1)/\rho_0\Omega)/((K+1)/\rho_0\Omega+sy/(y+C)))}}{\sqrt{((K+1)/\rho_0\Omega + sy/(y+C))(K(K+1)/\rho_0\Omega)}} \\ &\quad \cdot M_{-1/2,0}\left(\frac{K(K+1)/\rho_0\Omega}{(K+1)/\rho_0\Omega + sy/(y+C)}\right), \end{aligned} \tag{A.3}$$

where $M_{\lambda,\mu}(z)$ is the Whittaker functions defined as [36, Eq. (9.220.2)]

$$M_{\lambda,\mu}(z) = z^{\mu+1/2} e^{-z/2} \Phi\left(\mu - \lambda + \frac{1}{2}, 2\mu + 1; z\right), \tag{A.4}$$

where $\Phi(\alpha, \gamma; z)$ is a second notation of confluent hypergeometric function, and when $\alpha = \gamma$, $\Phi(\alpha, \gamma; z)$ has the relationship [36, Eq. (9.215)] as follows:

$$\Phi(\alpha, \alpha; z) = e^z. \tag{A.5}$$

Thus, (A.3) can be rewritten as

$$\begin{aligned} L &= \frac{(K+1)e^{-K}}{K+1+\rho_0\Omega s} \frac{y+C}{y+C(K+1)/(K+1+\rho_0\Omega s)} \\ &\quad \cdot e^{(K(K+1)/(K+1+\rho_0\Omega s))((y+C)/(y+C(K+1)/(K+1+\rho_0\Omega s)))}. \end{aligned} \tag{A.6}$$

In order to make the derivation more clear, we define

$$\begin{aligned} \frac{C(K+1)}{K+1+\rho_0\Omega s} &= u_1(s), \\ \frac{K(K+1)}{K+1+\rho_0\Omega s} &= u_2(s), \\ C - u_1 &= \varepsilon. \end{aligned} \tag{A.7}$$

After some algebra manipulations, we can rewrite (A.6) as

$$L = \frac{u_2(s) e^{-K+u_2(s)}}{K} \left(1 + \frac{\varepsilon}{y + u_1(s)}\right) e^{\varepsilon u_2(s)/(y+u_1(s))}. \tag{A.8}$$

Now, we can rewrite $M_{\gamma_{02}}(s)$ as the integral of variable y

$$\begin{aligned} M_{\gamma_{02}}(s) &= \int_0^\infty \frac{u_2(s) e^{-K+u_2(s)}}{K} \left(1 + \frac{\varepsilon}{y+u_1(s)}\right) \\ &\quad \cdot e^{\varepsilon u_2(s)/(y+u_1(s))} \frac{\alpha_2}{\rho_2} e^{-(\beta_2/\rho_2)y} \\ &\quad \cdot {}_1F_1\left(m_2; 1; \frac{\delta_2}{\rho_2} y\right) dy \\ &= \frac{u_2(s)\alpha_2 e^{-K+u_2(s)}}{K\rho_2} \int_0^\infty \left(1 + \frac{\varepsilon}{y+u_1(s)}\right) \\ &\quad \cdot e^{\varepsilon u_2(s)/(y+u_1(s)) - (\beta_2/\rho_2)y} \cdot {}_1F_1\left(m_2; 1; \frac{\delta_2}{\rho_2} y\right) dy. \end{aligned} \tag{A.9}$$

Let $w(s) = y + u_1(s)$, and then we have $y = w(s) - u_1(s)$ and $dy = dw$. After some algebra manipulations, (A.9) can be written as the integral of variable w

$$\begin{aligned} M_{\gamma_{02}}(s) &= \frac{u_2(s)\alpha_2 e^{-K+u_2(s)+(\alpha_2/\rho_2)u_1(s)}}{K\rho_2} \\ &\quad \cdot \int_0^\infty \left(1 + \frac{s}{w(s)}\right) e^{\varepsilon u_2(s)/w(s) - (\beta_2/\rho_2)w(s)} \\ &\quad \cdot {}_1F_1\left(m_2; 1; \frac{\delta_2}{\rho_2}(w(s) - u_1(s))\right) dw. \end{aligned} \tag{A.10}$$

By substituting (20) into (A.10), we have

$$\begin{aligned} M_{\gamma_{02}}(s) &= \frac{u_2(s)\alpha_2 e^{-K+u_2(s)+(\alpha_2/\rho_2)u_1(s)}}{K\rho_2} \\ &\quad \cdot \sum_{n=0}^{m_2-1} \frac{(-1)^n (1-m_2)_n}{(n!)^2} \left(\frac{\delta_2}{\rho_2}\right)^n \cdot \int_0^\infty \left(1 + \frac{\varepsilon}{w(s)}\right) \\ &\quad \cdot e^{\varepsilon u_2(s)/w(s) - ((\beta_2-\delta_2)/\rho_2)w(s)} (w(s) \\ &\quad - u_1(s))^n dw. \end{aligned} \tag{A.11}$$

Then, by using the Binomial expansion for $(w(s) - u_1(s))^n$, we can rewrite (A.11) as

$$\begin{aligned} M_{\gamma_{02}}(s) &= \frac{u_2(s)\alpha_2 e^{-K+u_2(s)+(\alpha_2/\rho_2)u_1(s)}}{K\rho_2} \\ &\quad \cdot \sum_{n=0}^{m_2-1} \frac{(-1)^n (1-m_2)_n}{(n!)^2} \left(\frac{\delta_2}{\rho_2}\right)^n \cdot \sum_{k=0}^{n} [-u_1(s)]^{n-k} \\ &\quad \cdot \int_0^\infty \left(1 + \frac{\varepsilon}{w(s)}\right) \\ &\quad \cdot e^{\varepsilon u_2(s)/w(s) - ((\beta_2-\delta_2)/\rho_2)w(s)} (w(s))^k dw. \end{aligned} \tag{A.12}$$

The integral part of (A.12) can be solved by using [36, Eq. (3.471.9)] as follows:

$$\begin{aligned}
&\int_0^\infty \left(1+\frac{\varepsilon}{w(s)}\right) \\
&\quad \cdot e^{\varepsilon u_2(s)/w(s)-((\beta_2-\delta_2)/\rho_2)w(s)} (w(s))^k \, dw \\
&\quad = \int_0^\infty e^{\varepsilon u_2(s)/w(s)-((\beta_2-\delta_2)/\rho_2)w(s)} (w(s))^k \, dw \\
&\quad + \varepsilon \int_0^\infty e^{\varepsilon u_2(s)/w(s)-((\beta_2-\delta_2)/\rho_2)w(s)} (w(s))^{k-1} \, dw \\
&= 2\left(\frac{\varepsilon\rho_2 u_2(s)}{\beta_2-\delta_2}\right)^{(k+1)/2} \\
&\quad \cdot K_{k+1}\left(2\sqrt{\frac{\varepsilon u_2(s)(\beta_2-\delta_2)}{\rho_2}}\right) \\
&\quad + 2\varepsilon\left(\frac{\varepsilon\rho_2 u_2(s)}{\beta_2-\delta_2}\right)^{k/2} K_k\left(2\sqrt{\frac{\varepsilon u_2(s)(\beta_2-\delta_2)}{\rho_2}}\right).
\end{aligned} \tag{A.13}$$

Therefore, by plugging (A.13) into (A.12), the result of $M_{\gamma_{02}}(s)$ can be obtained as presented in (27).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Jian Jiao, Houlian Gao, and Qinyu Zhang contributed equally to this work.

Acknowledgments

This work was supported in part by the National Natural Sciences Foundation of China (NSFC) under Grants 61771158, 61701136, 61525103, and 61371102, the National High Technology Research & Development Program no. 2014AA01A704, the Natural Scientific Research Innovation Foundation in Harbin Institute of Technology under Grant HIT.NSRIF.2017051, and the Shenzhen Fundamental Research Project under Grants JCYJ20160328163327348 and JCYJ20150930150304185.

References

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

[2] S. Zhang, N. Zhang, X. Fang, P. Yang, and X. S. Shen, "Self-Sustaining Caching Stations: Toward Cost-Effective 5G-Enabled Vehicular Networks," *IEEE Communications Magazine*, vol. 55, no. 11, pp. 202–208, 2017.

[3] S. Zhang, N. Zhang, P. Yang, and X. S. Shen, "Cost-Effective Cache Deployment in Mobile Heterogeneous Networks," *IEEE Transactions on Vehicular Technology*, In press.

[4] N. Cheng, H. Zhou, L. Lei et al., "Performance Analysis of Vehicular Device-to-Device Underlay Communication," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5409–5421, 2017.

[5] M. Á. Vázquez, A. Pérez-Neira, D. Christopoulos et al., "Precoding in multibeam satellite communications: Present and future challenges," *IEEE Wireless Communications Magazine*, vol. 23, no. 6, pp. 88–95, 2016.

[6] S. Maleki, S. Chatzinotas, B. Evans et al., "Cognitive spectrum utilization in Ka band multibeam satellite communications," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 24–29, 2015.

[7] J. Jiao, X. Sui, S. Gu, S. Wu, and Q. Zhang, "Partially Observable Markov Decision Process-Based Transmission Policy over Ka-Band Channels for Space Information Networks," *Entropy*, vol. 19, no. 10, p. 510, 2017.

[8] Y. Feria, M. Belongie, T. Mcpheeters, and H. Tan, "Solar scintillation effects on yelecommunication links at Ka-Band and X-Band," *Telecommunications and Data Acquisition Progress Report*, vol. 129, pp. 1–11, 1997.

[9] R. F. Arenstorf, "A new method of perturbation theory and its application to the satellite problem of celestial mechanics," *Journal für die reine und Angewandte Mathematik*, vol. 1966, no. 221, pp. 113–145, 1966.

[10] C. Zhong, S. Jin, and K.-K. Wong, "Dual-hop systems with noisy relay and interference-limited destination," *IEEE Transactions on Communications*, vol. 58, no. 3, pp. 764–768, 2010.

[11] L. Yang and M. O. Hasna, "Performance Analysis of Amplify-and-Forward Hybrid Satellite-Terrestrial Networks with Cochannel Interference," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 5052–5061, 2015.

[12] K. An, M. Lin, T. Liang et al., "Performance Analysis of Multi-Antenna Hybrid Satellite-Terrestrial Relay Networks in the Presence of Interference," *IEEE Transactions on Communications*, vol. 63, no. 11, pp. 4390–4404, 2015.

[13] Y. Wang, J. Jiao, X. D. Sui, S. H. Wu, Y. H. Li, and Q. Y. Zhang, "Rateless coding scheme for time-varying dying channels," in *Proceedings of the 8th Wireless Communications Signal Processing (WCSP)*, Yangzhou, China, October 2016.

[14] Y. Ruan, Y. Li, R. Zhang, and H. Zhang, "Performance analysis of hybrid satellite-terrestrial cooperative networks with distributed alamouti code," in *Proceedings of the 83rd IEEE Vehicular Technology Conference (VTC Spring)*, Nanjing, China, May 2016.

[15] A. Iqbal and K. M. Ahmed, "A hybrid satellite-terrestrial cooperative network over non identically distributed fading channels," *Journal of Communications*, vol. 6, no. 7, pp. 581–589, 2011.

[16] S. Sreng, B. Escrig, and M.-L. Boucheret, "Exact outage probability of a hybrid satellite terrestrial cooperative system with best relay selection," in *Proceedings of the 2013 IEEE International Conference on Communications (ICC)*, pp. 4520–4524, Hungary, June 2013.

[17] M. R. Bhatnagar and M. K. Arti, "Performance analysis of AF based hybrid satellite-terrestrial cooperative network over generalized fading channels," *IEEE Communications Letters*, vol. 17, no. 10, pp. 1912–1915, 2013.

[18] V. K. Sakarellos and A. D. Panagopoulos, "Outage performance of cooperative Land Mobile Satellite broadcasting systems," in

Proceedings of the IEEE European Conference on Antennas and Propagation, pp. 473–476, April 2013.

[19] V. K. Sakarellos, C. Kourogiorgas, and A. D. Panagopoulos, "Cooperative Hybrid Land Mobile Satellite–Terrestrial Broadcasting Systems: Outage Probability Evaluation and Accurate Simulation," *Wireless Personal Communications*, vol. 79, no. 2, pp. 1471–1481, 2014.

[20] M. Li, M. Lin, Q. Yu, W.-P. Zhu, and L. Dong, "Optimal beamformer design for dual-hop MIMO AF relay networks over rayleigh fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 8, pp. 1402–1414, 2012.

[21] J. Wang, H. Zhu, and N. J. Gomes, "Distributed antenna systems for mobile communications in high speed trains," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 675–683, 2012.

[22] H. Zhu, "Performance comparison between distributed antenna and microcellular systems," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 6, pp. 1151–1163, 2011.

[23] P.-D. Arapoglou, K. Liolis, M. Bertinelli, A. Panagopoulos, P. Cottis, and R. De Gaudenzi, "MIMO over satellite: a review," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 1, pp. 27–51, 2011.

[24] A. M.k. and M. R. Bhatnagar, "Beamforming and combining in hybrid satellite-terrestrial cooperative systems," *IEEE Communications Letters*, vol. 18, no. 3, pp. 483–486, 2014.

[25] N. I. Miridakis, D. D. Vergados, and A. Michalas, "Dual-Hop Communication over a Satellite Relay and Shadowed Rician Channels," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4031–4040, 2015.

[26] P. K. Upadhyay and P. K. Sharma, "Max-Max user-relay selection scheme in multiuser and multirelay hybrid satellite-Terrestrial relay systems," *IEEE Communications Letters*, vol. 20, no. 2, pp. 268–271, 2016.

[27] Y. Dhungana, N. Rajatheva, and C. Tellambura, "Performance analysis of antenna correlation on LMS-Based dual-hop AF MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 8, pp. 3590–3602, 2012.

[28] Y. Dhungana and N. Rajatheva, "Analysis of LMS based dual hop MIMO systems with beamforming," in *Proceedings of the 2011 IEEE International Conference on Communications*, Kyoto, Japan, June 2011.

[29] M. R. Bhatnagar, "Performance Evaluation of Decode-and-Forward Satellite Relaying," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4827–4833, 2015.

[30] M. K. Arti and M. R. Bhatnagar, "Two-way mobile satellite relaying: A beamforming and combining based approach," *IEEE Communications Letters*, vol. 18, no. 7, pp. 1187–1190, 2014.

[31] K. An, J. Ouyang, M. Lin, and T. Liang, "Outage Analysis of Multi-Antenna Cognitive Hybrid Satellite-Terrestrial Relay Networks with Beamforming," *IEEE Communications Letters*, vol. 19, no. 7, pp. 1157–1160, 2015.

[32] M. Jia, X. Gu, Q. Guo, W. Xiang, and N. Zhang, "Broadband hybrid satellite-terrestrial communication systems based on cognitive radio toward 5G," *IEEE Wireless Communications Magazine*, vol. 23, no. 6, pp. 96–106, 2016.

[33] F. Lu, X. S. Zhang, and Z. X. Tian, "Analysis of influence on satellite-to-earth data transmission link by directing precision of Ka-band spot beam antenna," *Spacecraft Engineering*, vol. 25, no. 6, pp. 61–68, 2016.

[34] F. Wang, X. Z. Li, and J. Meng, "Analysis of the impact of elevation on satellite data quality," in *Proceedings of the in Nineteenth Academic conference of Chinese Society of Space Research Detection Professional Committee*, 2006.

[35] J. Jiao, Y. Yang, B. Feng, S. Wu, Y. Li, and Q. Zhang, "Distributed Rateless Codes with Unequal Error Protection Property for Space Information Networks," *Entropy*, vol. 19, no. 1, p. 38, 2017.

[36] A. Jeffrey and D. Zwillinger, *Table of Integrals, Series, and Products*, Academic Press, 2007.

[37] M. R. Mckay, A. Zanella, I. B. Collings, and M. Chiani, "Error probability and SINR analysis of optimum combining in Rician fading," *IEEE Transactions on Communications*, vol. 57, no. 3, pp. 676–687, 2009.

[38] S. Gu, J. Jiao, and Q. Zhang, "Intermediate performance of rateless codes over dying erasure channel," in *Proceeding of the 6th International Conference on Communications, Signal Processing, and Systems (CSPS2017)*, Harbin, China, July 2017.

[39] A. Abdi, W. C. Lau, M. Alouini, and M. Kaveh, "A new simple model for land mobile satellite channels: First-and second-order statistics," *IEEE Transactions on Wireless Communications*, vol. 2, no. 3, pp. 519–528, 2003.

[40] K. T. Alfriend, H. Schaub, and D.-W. Gim, "Gravitational perturbations, nonlinearity and circular orbit assumption effects on formation flying control strategies," *Guidance and Control*, pp. 139–155, 2000.

[41] H. Gao, J. Jiao, R. Zhang, S. Gu, S. Wu, and Q. Zhang, "Outage performance of cooperative deep-space downlink with backbone relaying in space-based information networks," in *Proceeding of the 6th International Conference on Communications, Signal Processing, and Systems (CSPS2017)*, Harbin, China, July 2017.

[42] A. F. B. de Almeida Prado, "Third-body perturbation in orbits around natural satellites," *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 1, pp. 33–40, 2003.

[43] Y. Kozai, "A new method to compute Lunisolar perturbations in satellite motions," Sao Special Report, 1973.

[44] D. Vokrouhlicky, P. Farinella, and F. Mignard, "Solar radiation pressure perturbations for Earth satellites I: A complete theory including penumbra transitions," *Astronomy & Astrophysics* , vol. 280, no. 1, pp. 295–312, 1993.

[45] A. Martyusheva, K. Oskina, N. Petrov, and E. Polyakhova, "Solar radiation pressure influence in motion of asteroids, including near-Earth objects," in *Proceedings of the 2015 International Conference on Mechanics - Seventh Polyakhov's Reading*, Saint Petersburg, Russia, February 2015.

[46] L. M. Buchsbaum, "Pointing losses in single-axis and fixed-mount earth-station antennas due to satellite movement," *International Journal of Satellite Communications and Networking*, vol. 4, no. 2, pp. 89–96, 1986.

[47] S. Q. Min, *Design and Application of Satellite Communication System*, Publishing House of Electronics Industry, 2015.

[48] M. Pätzold, *Mobile Fading Channels*, John Wiley & Sons, Ltd, Chichester, UK, 2002.

[49] Iskandar and S. Shimamoto, "Channel characterization and performance evaluation of mobile communication employing stratospheric platforms," *IEICE Transactions on Communications*, vol. E89-B, no. 3, pp. 937–944, 2006.

[50] I. Wolfram, *Mathematica Edition: Version 8.0*, Wolfram Research Inc, Champaign, Illinois, USA, 2010.

[51] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Courier Corporation, 1964.

[52] M. K. Simon and M. Alouini, *Digital Communication Over Fading Channels*, John Wiley & Sons, Inc., New York, NY, USA, 2005.

[53] O. Montenbruck and E. Gill, *Satellite Orbits: Models, Methods and Applications*, Springer Science and Business Media, 2012.

[54] N. Zhengtao, "Progress and present status of research on Earths Gravitational Field," *Journal of Geomatics*, vol. 1, 2013.