This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2921444, IEEE Transactions on Vehicular Technology

Toward Dynamic Link Utilization for Efficient Vehicular Edge Content Distribution

Haibo Zhou, *Senior Member, IEEE,* Nan Cheng, *Member, IEEE,* Jian Wang, *Member, IEEE,* Jiacheng Chen, Quan Yu, *Senior Member, IEEE,* and Xuemin (Sherman) Shen, *Fellow, IEEE,*

***Abstract*—With the significant advance of connected vehicles, future demand for vehicular infotainment services will be greatly increased. Traditional content distribution approaches based on typical cellular architecture suffer from long latency and unstable connections in high-dynamic vehicular environments and even may cause congestion on the backhaul network due to a large amount of requested data. In the paper, we propose a novel content delivery framework by leveraging the 5G edge networks, in which the content caching and data prefetching techniques are exploited accordingly. We investigate the comprehensive dynamic link utilization problem in 5G edge networks from the perspectives of vehicle users and network operator, respectively. For the vehicle users' perspective, our aim is to maximize the vehicular content distribution throughput through optimal scheduling of vehicular access link slots, and also the utilization problem of backhaul link slots in edge networks is studied to reduce the data access delay of vehicles. For the network operator's perspective, the objective is to maximize the total profit of the network operator, and therefore, the auction model is utilized for the vehicular access link slot scheduling and the backhaul link utilization is analyzed in terms of compensations and costs. Finally, extensive simulations are conducted to demonstrate the efficiency of the proposed solutions for edge content delivery of connected vehicles.**

***Index Terms*—Connected Vehicles; Edge Networks; Caching; Auction; Vehicular Content Distribution**

I. Introduction

Information technology (IT) industry leaders have envisioned the bright future of connected vehicles [1]. According to a recent report, it is predicted that 98% of cars will be connected by 2020, and 100% in 2025 [2]. With the vehicles linked to the cyberspace, people are able to stay connected on the road, fed by the information from a variety of sources such as intelligent transportation system (ITS), news and multimedia entertainment etc. To achieve this goal, researchers have focused on the development of vehicular communication network (VCN) in recent years. Among all the approaches, the mobile telecommunications technologies have been regarded as the fastest way to implement VCN [3]. As an example, AT&T has collaborated with several major car manufacturers to make several specific car models communication-enabled and has accordingly presented its data plans. However, the current cellular networks can only be a temporal solution due to the high data cost [4]. In addition, the current mobile users have already consumed a huge amount of its capacity and the demands on vehicular infotainment services are still growing rapidly. It is imperative to design cost-effective and alternative vehicular content delivery approaches on the road.

In vehicular networks, the roadside units (RSUs) are deployed along the road to provide vehicles with location-based services, such as local traffic information, road condition notification, advertisement, etc. These RSUs work in a similar way as the WiFi networks, providing relatively small wireless coverages (several hundreds of meters) to vehicles. Inspired by the 5G edge networks [5]-[6], in this paper, we adopt the concept of "edge network" in vehicular network, where the RSUs provide high throughput and edge content cache to the vehicles, and thus are termed as vehicular edge infostations. Vehicular infotainment station is basically a vehicular content delivery system based on the edge network and vehicle-to-roadside (V2R) communication technology, which means vehicles are connected to the edge roadside infostation for data acquirement [7]-[8]. However, providing information services to vehicles via infostations are challenging. Considering the high vehicle mobility, the vehicle's sojourn within the coverage of an infostation is short. Moreover, the contention-based channel access used in vehicular communication render inefficiency when the density of vehicles is high, leading to low throughput and prolonged for a single vehicle. In addition, the limited backhaul bandwidth also constrains the network performance especially when the vehicular users are requesting data craving services such as high-quality video streaming and virtual reality. Therefore, to improve the network performance and user quality of service (QoS), we propose a novel content delivery framework which leverages the edge networks caching and schedules the wireless link slots. Vehicles only need to retrieve information from the content server instead of the Internet. In order to cater to the infotainment demands of most of the vehicles, the contents stored in the server also should be carefully selected and periodically updated, e.g., local ITS information, popular video clips, and news headlines etc. In addition, we use content caching and data prefetching techniques to further reduce the data access delay. In specific, caching works by storing certain contents in the cache spaces of infostations for a long time, while prefetching means that a piece of the content requested by a vehicle is sent to the infostations before it is transmitted to the vehicle, and the data

H. Zhou, J. Wang (corresponding author), and Q. Yu are with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China, 210023, (e-mail: {haibozhou, wangjnju}@nju.edu.cn, yuquan61@qq.com)

J. Cheng is with Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, (e-mail: jiacheng1989@gmail.com)

N. Cheng and X. Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada, N2L 3G1, (e-mail: {n5cheng, sshen}@uwaterloo.ca).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2921444, IEEE Transactions on Vehicular Technology

is deleted from the cached infostation after the transmission. Both of the techniques can save the delay between the content server and cached infostations. We also investigate the problem of wireless resource scheduling in cached infostation system, which dynamically allocates the time slots of wireless link utilization to vehicles for the sake of throughput maximization.

In this paper, we engineer connected vehicular content distribution by leveraging the edge infostations, and investigate the comprehensive dynamic link utilization problems from the perspectives of vehicle user and network operator, respectively. Generally speaking, in cached infostation system, we study the resource management of both vehicle access network and local content distribution network in each allocation period. The former consists of wireless links between vehicles and infostations, and the latter comprises the backhaul links between the local content server and infostations. From the vehicle users' perspective, we seek to maximize the throughput for V2R communications, or equivalently maximizing the total transmitted data between vehicles and infostations. The challenges are threefold: i) how to allocate the wireless link slots among the vehicles; ii) how to utilize the backhaul link for content caching and data prefetching; and iii) how to select appropriate contents to be cached at the infostations. For the first challenge, we formulate the wireless link slot scheduling problem aiming at throughput maximization, while the user fairness is also considered. For the second challenge, we identify the objective of the backhaul link utilization. For the third challenge, we aim to maximize the total utility of all the cached contents. Corresponding solutions are provided for the problems mentioned above. Also, we study cached infostation system from the network operator's perspective. Then, the challenges are how to schedule the wireless link slot and design the usage pattern of backhaul link to maximize the total profit. We adopt the auction model to describe the resource scheduling and a strategy-proof auction mechanism is presented. In addition, Backhaul link utilization process is formulated as an operational expenditure (OPEX) minimization problem, and the solution is based on the tradeoff property between compensations and costs.

We highlight the contributions of the paper in three-fold in the light of previous literature works:

- We design a system architecture for 5G edge vehicular content delivery, i.e., edge infostation, which is featured by the deployment of local content server and the use of content caching and data prefetching techniques, so as to reduce the data access delay and efficiently utilize the precious connection time of vehicles.
- We study the cached infostation from the vehicle user's perspective. To maximize the total transmitted data between vehicles and Infostations. We consider the wireless link slot scheduling, backhaul link utilization and content caching policy, and corresponding solutions are given.
- We investigate the cached infostation from the network operator's perspective. In order to maximize the total network operator's profit, wireless link slot scheduling is modeled as an auction process and a strategy-proof auction mechanism is developed. Also, the objective of backhaul link utilization is to reduce the OPEX, and an solution is provided to achieve this goal.

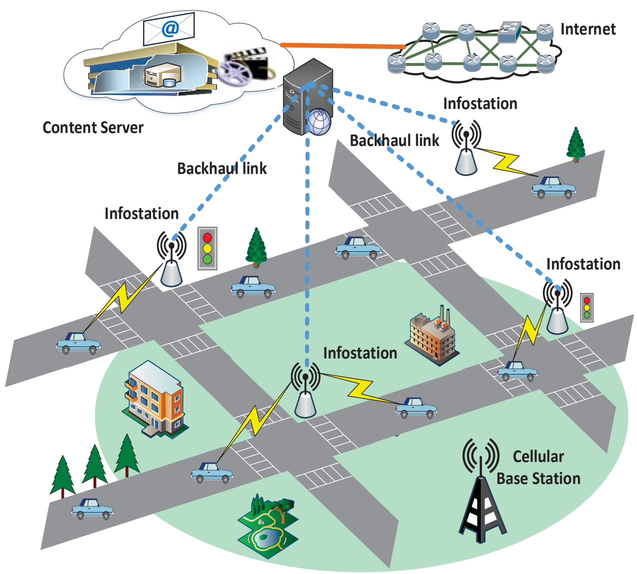


Fig. 1. Architecture of connected vehicular infotainment services on the road by leveraging the cached Infostations.

The remainder of this paper is organized as follows. In Section II, the related works are reviewed. System model is introduced in Section III. The studies on cached infostation from two different perspectives are given in Section IV and Section V, respectively. Simulation results are presented in Section VI. Finally, Section VII concludes the paper.

II. Related Work

Vehicular content delivery system plays an important role in providing high-quality infotainment services to vehicles [9]. Surveys related to this topic can be found in [10]. In addition, Malandrino *et al.* in [11] investigated the vehicular content downloading problem with a max-flow formulation and provided the optimal infostation deployment and data transfer paradigm by using vehicle-to-vehicle (V2V) communications and vehicle-to-infrastructure (V2I) communications. An extended work [12] developed a fog-of-war model to handle the mobility prediction noise and construct a time-expanded graph correspondingly. The data prefetching and transmission scheduling for both V2V and V2I communications are jointly decided through solving a linear programming problem. In addition, Wang *et al.* in [13] studied the vehicular popular contents distribution through infostations and the peer-to-peer (P2P) technology is utilized to realize the cooperative content sharing among vehicles. Zhang *et al.* in [14] introduced the cooperative vehicular content dissemination approach by allocating representative infostations to prefetch and share data with other vehicles. Similarly, the cooperation among infostations was investigated in [15], in which infostations can coordinate the data transmission among vehicles. Li *et al.* in [16] studied the vehicular content dissemination from a perspective of opportunistic network, and the vehicular content dissemination process has been formulated to maximize

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2921444, IEEE Transactions on Vehicular Technology

the submodular function. Sun *et al.* developed a quality-of-experience (QoE) centric distributed caching approach in [17], which exploits the cellular network for providing video streaming services to vehicular users. In [18], Mezghani *et al.* presented an overview on content dissemination in the context of vehicular social networking. They also outlined a framework to tackle the new challenge of supporting user satisfaction. In [19], an efficient road-oriented dissemination (ROD) protocol was proposed, which relies on cooperative V2V communications including both optimized distance defer transfer strategy and store-and-forward strategy. For the sake of cost-effective content distribution, Luan *et al.* [20] proposed to use low-cost distributed roadside buffers (RSBs) to distribute vehicular contents wirelessly. They also developed a distributed content replication strategy to optimize the caching at each RSB. We also conducted a prior work [21] on engineering the link utilization in VANET for content delivery purpose. Different from the above mentioned works, in this paper, we jointly consider the V2R wireless resource allocation and backhaul link utilization within each individual infostation. Data prefetching and caching are integrated into the above resource management process. What is more, the problems are solved for both network/throughput and operator/revenue perspectives.

III. System Model

We present the architecture of edge connected vehicular infotainment services provision by leveraging the Infostations in Fig. 1. The content server is connected with the Internet such that the cached contents can be updated. The infostations are nodes with communication as well as storage capabilities. They are located near the vehicles (i.e., at the edge) such that the transmission delay over air can be reduced. For each infostation, two wired backhaul links that connect it with the content server are considered, namely basic backhaul link and quick access backhaul link. Specifically, the basic backhaul link is used for common data transmission. The quick access backhaul link is used for content caching and data prefetching, which is our focus in the paper. In the following, backhaul link refers to the latter if not specified. Each infostation also connects vehicles within its coverage area with wireless links. We consider the wireless resources used for vehicular communications between vehicles and infostations can be DSRC spectrum bands, unlicensed ISM spectrum bands, or TV White Space spectrum bands [22]. In addition, cellular link is utilized for exchanging the control messages between vehicles and the infostation system. Although the cached infostation is primarily designed to be a vehicular content delivery system, it can also support data uploading for the vehicles. Actually, a vehicle connected with the cached infostation works as a moving Wi-Fi hotspot, providing wireless access to the mobile devices in it.

In this paper, the time slot is defined as the basic wireless and backhaul link resource utilization unit. We denote a single slot length and a scheduling period as δ and $\mathcal{T}$ respectively, and $\mathcal{T}$ contains T slots. We also suppose that the scheduling of wireless link slot is done before the backhaul link, and

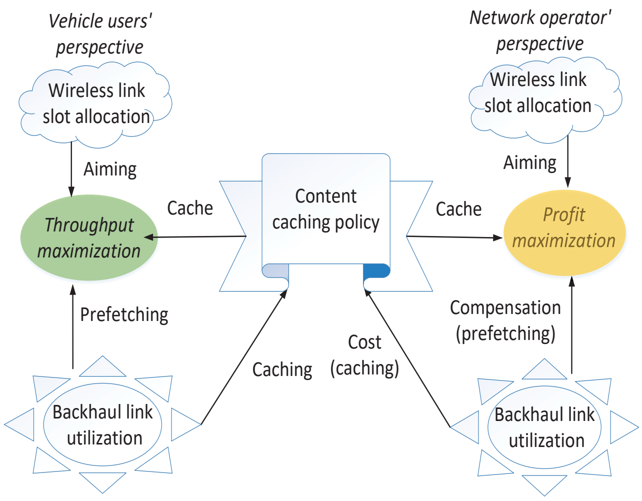


Fig. 2. An intuitive illustration of the studied problems in the paper.

both of them happen just before the beginning of each period. If the content server is updated, then the contents cached at the infostations should also be updated, and such decision is supposed to be made prior to the wireless link slot scheduling. We denote the infostations and vehicles in the studied area as the set of $\mathcal{R}$ and $\mathcal{V}$ respectively. Since the resource scheduling is done period by period, $\mathcal{V}$ actually represents the vehicles just before the beginning of the next period. Likewise, the contents in the server are represented by the set $\mathcal{C}$. The cardinality and an element of $\mathcal{R}$, $\mathcal{V}$, $\mathcal{C}$ are denoted by R, V, C and r, v, c, respectively. For convenience, we list the important notations used in this paper in Tab. I, and an intuitive illustration of the big picture of studied problems in the paper is given in Fig. 2.

IV. Study from Vehicle Users' Perspective

In this section, from the vehicle users' perspective, we first investigate the cached infostation system for connected vehicular infotainment services. We aim to maximize the throughput of vehicular access network, and specifically the total transmitted data between vehicles and infostations in each period. To this end, we investigate the problem of wireless resource scheduling in cached infostation system, which dynamically allocates the time slots of wireless link utilization to vehicles for the sake of throughput maximization. Furthermore, in order to reduce the data access delay that leads to the waste of wireless slots [23], we study the utilization problem of backhaul link slots in cached infostation system, which can be used for content caching or data prefetching, both for eliminating the delay between infostations and the content server. Moreover, due to the cache space limit of the infostations, a content caching policy is required so as to select appropriate contents that can minimize the data access delay.

A. Wireless Link Slot Scheduling in Cached Infostation System

For each $v \in \mathcal{V}$, we use RT dimensional vector $\rho^v = (\rho_1^{v,1}, \rho_1^{v,2}, ..., \rho_1^{v,R}, ...\rho_t^{v,r}, ..., \rho_T^{v,1}, \rho_T^{v,2}, ..., \rho_T^{v,R})$ as its link

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2921444, IEEE Transactions on Vehicular Technology

TABLE I
IMPORTANT NOTATIONS USED IN THE PAPER.

Vehicle Users' Perspective	
Notation	**Meaning**
$\mathcal{T}, T, t, \delta$	A scheduling period $\mathcal{T}$ containing T time slots, each slot t with length δ
$\mathcal{R}, \mathcal{V}, \mathcal{C}; R, V, C; r, v, c$	The set of Infostations, vehicles and contents; set size; set element
$\rho^v = (\rho_1^{v,1}, ...\rho_t^{v,r}, ..., \rho_T^{v,R})$	The link quality profile for v in a period
$\xi^r = (\xi_1^{1,r}, ..., \xi_t^{v,r}, ..., \xi_T^{V,r})$	The scheduling result for r in a period
$\eta^r = (\eta_1^r, ..., \eta_c^r, ..., \eta_C^r)$	The cache status of r in a period
$f_{v,r,t}^d, f_{v,r,t}^u$	The download and upload flow of v at r in t
$B^v, B_{\mathcal{T}}^v$	The total buffer size of v, the used buffer size of v at $\mathcal{T}$
$\beta^r = (\beta_1^r, ..., \beta_t^r, ..., \beta_T^r)$	Backhaul link utilization for r in a period
$h_{\mathcal{T}}^c, g^c$	The expected number of slots for transmitting c in $\mathcal{T}$, the number of slots for caching c
$U^c(m), U^{c,c'}(m)$	The utility of c, the utility of c replacing c'
Network Operator's Perspective	
Notation	**Meaning**
$\theta_{\mathcal{T}} = \{\theta_t^r : t \in \mathcal{T}, r \in \mathcal{R}\}$	The set of all the items (wireless link time slots)
$\mathcal{V}(\theta_t^r)$	The buyer group of θ_t^r
$b_v(b_t^r), q_v(q_t^r), w_v(w_t^r), p_v(p_t^r)$	The bid, valuation, bidding result and payment of v
u_v, S	The utility of v and the operator
$y_t^{v,r}, z_t^{v,r}, y^c, z^c$	The compensation reduction and cost for prefetching and content caching
$\varepsilon, \varepsilon'$	The price of compensation and backhaul link in the unit of $\$/bit$

quality profile in the period, where a positive $\rho_t^{v,r}$ denotes the transmission rate from v to r and $\rho_t^{v,r} = 0$ means the distance at time t is out of the effective transmission range from v to r . We suppose the infostations' coverage areas do not overlap with each other, hence there is at most one $\rho_t^{v,r} > 0$ in $\{\rho_t^{v,r} : r \in \mathcal{R}\}$ for $\forall t \in \mathcal{T}, \forall v \in \mathcal{V}$. The link quality profiles of all the vehicles in $\mathcal{T}$ is denoted by $\rho_{\mathcal{T}}$. For each $r \in \mathcal{R}$, we use a VT dimensional vector $\xi^r = (\xi_1^{1,r}, \xi_1^{2,r}, ..., \xi_1^{V,r}, ..., \xi_t^{v,r}, ..., \xi_T^{1,r}, \xi_T^{2,r}, ..., \xi_T^{V,r})$ to represent its time slot scheduling result in the period, where $\xi_t^{v,r} = 1$ indicates that the slot of r at t is allocated to v and $\xi_t^{v,r} = 0$ indicates otherwise. Likewise, there is at most one $\xi_t^{v,r} = 1$ in $\{\xi_t^{v,r} : v \in \mathcal{V}\}$ for $\forall t \in \mathcal{T}, \forall r \in \mathcal{R}$. The scheduling results of all the infostations in $\mathcal{T}$ is denoted by $\xi_{\mathcal{T}}$. For $\forall \xi_t^{v,r} = 1$, $\rho_t^{v,r} > 0$ should be satisfied so as to make the allocation feasibly. Also for each $r \in \mathcal{R}$, we use a C-dimension vector $\eta^r = (\eta_1^r, \eta_2^r, ..., \eta_c^r, ..., \eta_C^r)$ to represent the cache status of r, i.e., $\eta_c^r = 1$ means c is cached at r and $\eta_c^r = 0$ means otherwise.

We define a flow $f_{v,r,t}$ as the total transmitted data between v and r in t, and $f_{\mathcal{T}}$ represents all the possible flows. In order to calculate it, we consider the upload and download cases separately. $\mathcal{V}^d$ and $\mathcal{V}^u$ are used to represent the set of download and upload vehicles, respectively. In the download case, the content requested by v is denoted by $c^v \in \mathcal{C}$. Its total size is l^{c^v} and the unfinished size at the beginning of the period is $l_{\mathcal{T}}^{c^v}$. Then, the download flow is given below:

$$f_{v,r,t}^d = \rho_t^{v,r}(\delta - 1^{\{\eta_{c^v}^r = 0\}}\tau) \tag{1}$$

where τ denotes the delay between infostation and content server, and $1^{\{statement\}}$ denotes the indicator function which gives 1 if the $statement$ is true and 0 otherwise. Similarly, the upload flow is given as:

$$f_{v,r,t}^u = \rho_t^{v,r}\delta \tag{2}$$

We also consider the buffer size of each vehicle. B^v denotes the total buffer of v and $B_{\mathcal{T}}^v$ denotes the residual buffer at the beginning of $\mathcal{T}$. When a vehicle uploads data, the total upload flows in a period are at most the buffered data $B_{\mathcal{T}}^v$. On the other hand, the total download flows in a period cannot exceed the remaining size of the buffer, i.e., $B^v - B_{\mathcal{T}}^v$. Furthermore, we consider the fairness among all the vehicles. A window-based fairness constraint is adopted to achieve this [24][25][26]. Specifically, we define a window $\mathcal{W}$ containing W periods counting backward from the current period. For each vehicle, its total flow in $\mathcal{W}$ should be larger than a minimum value $f_{\min}$. Finally, we formulate the wireless link slot scheduling process as a constrained optimization problem, which is given below:

$$\max_{\xi_{\mathcal{T}}} \sum_{t\in\mathcal{T}, r\in\mathcal{R}, v\in\mathcal{V}} \xi_t^{v,r} f_{v,r,t}$$

$$s.t. \sum_{r\in\mathcal{R}} 1^{\{\rho_t^{v,r}>0\}} \leq 1, \ \forall t \in \mathcal{T}, \ \forall v \in \mathcal{V} \tag{3}$$

$$\sum_{v\in\mathcal{V}} 1^{\{\xi_t^{v,r}=1\}} \leq 1, \forall t \in \mathcal{T}, \ \forall r \in \mathcal{R} \tag{4}$$

$$\xi_t^{v,r}\rho_t^{v,r} > 0, \ \forall \xi_t^{v,r} > 0 \tag{5}$$

$$\sum_{t\in\mathcal{T}, r\in\mathcal{R}} \xi_t^{v,r} f_{v,r,t}^d \leq \min\{B^v - B_{\mathcal{T}}^v, l_{\mathcal{T}}^{c^v}\}, \ \forall v \in \mathcal{V}^d \tag{6}$$

$$\sum_{t\in\mathcal{T}, r\in\mathcal{R}} \xi_t^{v,r} f_{v,r,t}^u \leq B_{\mathcal{T}}^v, \ \forall v \in \mathcal{V}^u \tag{7}$$

$$\sum_{t\in\mathcal{W}, r\in\mathcal{R}} \xi_t^{v,r} f_{v,r,t} \geq f_{\min}, \ \forall v \in \mathcal{V} \tag{8}$$

The formulated wireless link slot scheduling process can be regarded as a "0-1" integer linear programming (ILP) problem, which is NP-hard. Therefore, we adopt a heuristic greedy method to solve the problem, which is given in Alg. 1. At the beginning, all the elements of $\xi_{\mathcal{T}}$ are initialized as 0, and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2921444, IEEE Transactions on Vehicular Technology

each flow $f_{v,r,t}$ can be calculated according to Eq. (1) or Eq. (2) such that we can obtain the value $f_{\mathcal{T}}$ (Line 1). Then, the non-zero elements of $f_{\mathcal{T}}$ can be sorted in a non-increasing order, which gives $f_{\mathcal{T}}^{\downarrow}$ (Line 2). The vehicles that do not satisfy the fairness constraint are considered first during the resource allocation. Specifically, the vehicles' flows whose total flows during the window $\mathcal{W}$ are less than f_{min} are selected from $f_{\mathcal{T}}^{\downarrow}$, which are denoted by $f_{\mathcal{T}}^{\downarrow *}$, also in non-increasing order (Line 2). Time slots are now allocated to the flows in $f_{\mathcal{T}}^{\downarrow *}$ one by one (Line 3-10). After a slot is allocated for $f_{v,r,t}$ (Line 4), we set the indicator $\xi_t^{v,r}$ as 1, and the flows conflicted with $f_{v,r,t}$ are denoted as $confl(f_{v,r,t}) = \{f_{v',r',t'} : r' = r, t' = t, v \in \mathcal{V}\}$, which are removed from both $f_{\mathcal{T}}^{\downarrow}$ and $f_{\mathcal{T}}^{\downarrow *}$ (Line 6). If either the constraint (6) or (7) is violated, i.e., $examine(f_{v,r,t}) = False$, then the flows of the same vehicle v denoted as $f_v = \{f_{v',r',t'} : v' = v, r' \in \mathcal{R}, t' \in \mathcal{T}\}$ are removed from both $f_{\mathcal{T}}^{\downarrow}$ and $f_{\mathcal{T}}^{\downarrow *}$ (Line 7-8). When the vehicle satisfies the fairness constraint after a slot is allocated to $f_{v,r,t}$, f_v is removed from $f_{\mathcal{T}}^{\downarrow *}$ (Line 9-10). When $f_{\mathcal{T}}^{\downarrow *}$ becomes empty, the above allocation process is repeated for $f_{\mathcal{T}}^{\downarrow}$ (Line 11). The whole wireless link slot scheduling process ends when $f_{\mathcal{T}}^{\downarrow}$ becomes empty.

Algorithm 1: Wireless Link Slot Scheduling—Vehicle Users' Perspective

Input: $\rho_{\mathcal{T}}$
Output: $\xi_{\mathcal{T}}$
1 $\xi_t^{v,r} \leftarrow 0$, for all $t \in \mathcal{T}, v \in \mathcal{V}, r \in \mathcal{R}$, calculate $f_{\mathcal{T}}$
2 get $f_{\mathcal{T}}^{\downarrow}$ and $f_{\mathcal{T}}^{\downarrow *}$
3 **while** $f_{\mathcal{T}}^{\downarrow *} \neq \emptyset$ **do**
4     $f_{v,r,t} \leftarrow f_{\mathcal{T}}^{\downarrow *}.pop()$
5     $\xi_t^{v,r} \leftarrow 1$
6   remove $confl(f_{v,r,t})$ from $f_{\mathcal{T}}^{\downarrow}$ and $f_{\mathcal{T}}^{\downarrow *}$
7   **if** $examine(f_{v,r,t}) = False$ **then**
8   remove f_v from $f_{\mathcal{T}}^{\downarrow}$ and $f_{\mathcal{T}}^{\downarrow *}$
9   **if** $\sum_{t \in \mathcal{W}, r \in \mathcal{R}} f_{v,r,t} \geq f_{\min}$ **then**
10         $f_{\mathcal{T}}^{\downarrow *}.remove(f_v)$
11 repeat Line 3-8 for $f_{\mathcal{T}}^{\downarrow}$

B. Backhaul Link Utilization in Cached Infostation System

On the one hand, at the beginning of the period, if some contents have not been completely cached at the infostations, then the backhaul link is responsible for doing this. On the other hand, the data requested by the vehicles that have been allocated time slots in the period can be prefetched (if it has not been cached) so as to reduce the delay. The tradeoff is that, a cached content is beneficial in the long term, which is guaranteed by the caching policy, while data prefetching can directly increase the total flow in the current period. An illustration of backhaul link utilization is shown in Fig. 3. In the first two periods, some slots of the backhaul link is used for caching content A and others are used for data prefetching. Then, from the third period, vehicles requesting A can use the full slot for transmission. However, the popularity of A is decreasing with time and new content B is updated at the server, so in the fifth period, B is cached to replace A.

The backhaul link utilization for r is represented by a T dimensional vector $\beta^r = (\beta_1^r, ..., \beta_t^r, ..., \beta_T^r)$, and that for all the infostations in the period is represented by $\beta_{\mathcal{T}}$. A slot t of backhaul link can be used in three ways, which are given below:

$$\beta_t^r = \begin{cases} \beta_t^{r,0} \ (not\ used) \\ \beta_t^{r,c}, c \in \mathcal{C} \ (caching\ c) \\ \beta_t^{r,v,t'}, v \in \mathcal{V}, t < t', t' \neq 1 \ (prefetching\ for\ t') \end{cases}$$

where the last way indicates that prefetching should be prior to the wireless link data transmission, and thus the data transmitted in the first slot cannot be prefetched. Our objective of backhaul link utilization is to cache the contents as possible so as to gain benefit early, and the total flow in the current period should also be considered for the fairness issue. Therefore, we formulate the problem of backhaul link utilization as follows:

Find $\beta_{\mathcal{T}}$ to cache the contents shortly while keeping the total flow in the period is more than a fraction $\epsilon \in (0, 1)$ of the optimal total flow f_{opt}, which can be calculated by considering all the requests are either cached or prefetched.

To solve the problem of backhaul link utilization in cached infostation system, we denote the flows that are allocated time slots at r as $f_{\mathcal{T}}^r$, and denote the transmission rates of these flows as $\rho_{\mathcal{T}}^r$. The problem of backhaul link utilization requires that the total flow should be larger than ϵf_{opt}. For simplicity, we consider the total flow of each infostation to be larger than ϵf_{opt}^r, where f_{opt}^r can be calculated by assuming the download requests are either cached or prefetched:

$$f_{opt}^r = \sum_{\rho_t^{v,r} \in \rho_{\mathcal{T}}^r} \rho_t^{v,r} \delta \tag{9}$$

Also, let $\mathcal{C}^r \subseteq \mathcal{C}$ be the set of contents that have not been completely cached in r at the beginning of the period. For $c^r \in \mathcal{C}^r$, it still needs k_{c^r} backhaul link slots to finish the caching. In addition, we can sort $\mathcal{C}^r$ in a non-decreasing order which is with respect to k_{c^r}, and we can obtain $\mathcal{C}^{r\uparrow}$.

The backhaul link utilization of r is described in Alg. 2. It firstly checks if the total flow at r is larger than ϵf_{opt}^r. If not, some of the backhaul link slots should be allocated for prefetching. Define $f_{v,r,t}^{d*}$ as the set of flows that can be prefetched, i.e., all the download flows whose requesting content is not cached, except for the flow which is allocated with the first time slot of wireless link. Sort $f_{v,r,t}^{d*}$ in a non-increasing order and get $f_{v,r,t}^{d\downarrow *}$ (Line 2). Then, at each time pick up a flow $f_{v,r,t}$ from $f_{v,r,t}^{d\downarrow *}$ and use the slot just before t to prefetch data for the flow (Line 4-7). When the total flow becomes larger than ϵf_{opt}^r, allocate the remaining backhaul link slots for caching the contents in $\mathcal{C}^{r\uparrow}$ if it is not empty (Line 11-14), or still for prefetching otherwise (Line 8-10).

C. Content Caching Policy for Cached Infostation System

Due to the cache space limit of the infostations [27], we propose a content caching policy to select appropriate contents that can make the most contributions for reducing the data access delay. First, we suppose that the cache space of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2921444, IEEE Transactions on Vehicular Technology

Algorithm 2: Backhaul Link Utilization—Vehicle Users' Perspective

Input: ξ^r,$\mathcal{C}^{r\uparrow}$
Output: β^r
$\beta_t^r \leftarrow \beta_t^{r,0}$ for all $t \in \mathcal{T}$
get $f_{\mathcal{T}}^r$ and $f_{v,r,t}^{d\downarrow *}$
$s \leftarrow sum(f_{\mathcal{T}}^r)$
while $s < \epsilon f_{opt}^r$ **and** $f_{v,r,t}^{d\downarrow *} \neq \emptyset$ **do**
  $f_{v,r,t}^d \leftarrow f_{v,r,t}^{d\downarrow *}.pop()$
  $s \leftarrow s + \rho_t^{v,r} \tau$
  $\beta_{t-1}^r \leftarrow \beta_{t-1}^{r,v,t}$
while $\mathcal{C}^{r\uparrow} = \emptyset$ **and** $f_{v,r,t}^{d\downarrow *} \neq \emptyset$ **do**
  $f_{v,r,t}^d \leftarrow f_{v,r,t}^{d\downarrow *}.pop()$
  $\beta_{t-1}^r \leftarrow \beta_{t-1}^{r,v,t}$
for $c^r \in \mathcal{C}^{r\uparrow}$ **do**
  while $\exists \beta_t^{r,0} \in \beta^r$ **and** $k_{c^r} > 0$ **do**
    change a $\beta_t^{r,0}$ into β_t^{r,c^r}
    $k_{c^r} \leftarrow k_{c^r} - 1$

infostation is not full. In order to evaluate the content, we define the utility of a content c during m periods as:

$$U^c(m) = \sum_{i=1}^{m} h_{\mathcal{T}_i}^c - g^c \tag{10}$$

where m is the content update period of the server, and we suppose that it is a multiple of the scheduling period, $\mathcal{T}_1$ is the current period, and $\mathcal{T}_m$ is the m-th period from $\mathcal{T}_1$. $h_{\mathcal{T}_i}^c$ is the benefit of c in period $\mathcal{T}_i$, namely the expected number of wireless link slots used for transmitting c in $\mathcal{T}_i$, and g^c is the number of backhaul link slots used for caching c. In the above definition, time slots are used as the metric for the utility. For data prefetching, the utility is 0, which can be used as the threshold to decide whether a content should be cached. Second, we consider the situation when there is no place for a new content. This means that an old content cached in the infostation needs to be replaced. In such case, the utility of the new content c should exclude the benefit of the old content c' that is to be replaced, and the expression is given below:

$$U^{c,c'}(m) = U^c(m) - \sum_{i=1}^{m} h_{\mathcal{T}_i}^{c'} \tag{11}$$

Finally, we can present the problem as follows:

Find the appropriate contents for caching such that the total utility is maximized, under the constraint of cache space.

In order to calculate the utility of content c with Eq. (10), we need to calculate the total backhaul link slots used for caching it, i.e.,

$$g^c = \frac{l^c}{\omega} \tag{12}$$

where l^c is the size of the content, and ω is the rate of the backhaul link. Also, we need to calculate $h_{\mathcal{T}}^c$, namely the expected number of wireless link slots used for transmitting c in a future period $\mathcal{T}$. We denote $\mathbb{P}_r(\mathcal{C}|\mathcal{T})$ as the probability

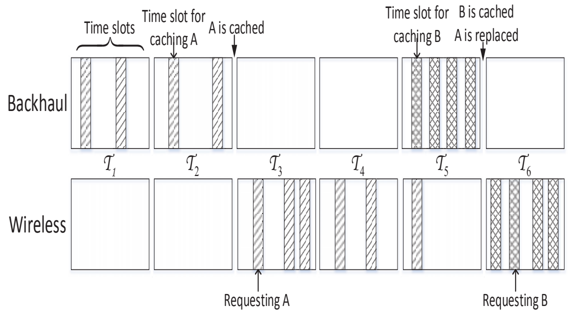


Fig. 3. Illustration of backhaul link utilization (the time slots used for prefetching is not shown).

distribution of all the contents in $\mathcal{C}$ at r in period $\mathcal{T}$, and denote $\mathbb{P}_r(c|\mathcal{T})$ as the probability that c is requested. Note that such a probability distribution can be different at different infostations, considering the location-based contents. Then, we have:

$$h_{\mathcal{T}}^c = T\mathbb{P}_r(c|\mathcal{T}), c \in \mathcal{C} \tag{13}$$

Obtaining $\mathbb{P}_r(\mathcal{C}|\mathcal{T})$ is a problem known as content popularity prediction [28]-[29] and is beyond the scope of this paper.

Let $D^r = \{(c, c') : c \in \mathcal{C}_{new}, c' \in \mathcal{C}_{old}^r\}$ be the set of content caching decisions at r, where $\mathcal{C}_{new}$ is the set of updated contents, $\mathcal{C}_{old}^r$ is the set of contents that have been cached in r, and a pair (c, c') means c' should be replaced by c. (c, c_0) is used to represent the situation when the remaining cache space is enough for c. L is the total cache space of an infostation and $L_{\mathcal{T}}^r$ is the used space of r in the current period. The procedures for obtaining D^r are shown in Alg. 3. First, calculate the utility $U^c(m)$ for $c \in \mathcal{C}_{new}$, delete the contents with non-positive utilities, sort the remaining contents in non-increasing order with respect to the utility and get $\mathcal{C}_{new}^{\downarrow}$ (Line 2). Then, sort all $c' \in \mathcal{C}_{old}^r$ in the non-decreasing order in terms of the sum of benefit in the next m periods and get $\mathcal{C}_{old}^{r\uparrow}$ (Line 2). In each iteration, the first content c in $\mathcal{C}_{new}^{\downarrow}$ is selected (Line 4). If the remaining space of the infostation is enough, then the decision (c, c_0) is made (Line 5-7). If not, the first content c' in $c' \in \mathcal{C}_{old}^{r\uparrow}$ is selected (Line 9) and the replacement decision (c, c') is made if the utility $U^{c,c'}(m)$ is above the threshold and the cache space is enough (Line 10-13).

V. Study from Network Operator's Perspective

In this section, for practical applications, we study the wireless link resource scheduling and backhaul link utilization problem from the network operator's perspective. We consider the auction mechanism for the wireless resource allocation, which has been widely adopted in secondary spectrum resource allocation [30]-[31]. We consider the design objective of network operator is to maximize the total profit, i.e., the difference between the total payments from the vehicles and OPEX. Payments come from the vehicles that use the wireless resources, so we adopt the auction model [30]-[31] in the wireless link resource scheduling for the sake of payment maximization, and a strategy-proof auction mechanism is

Algorithm 3: Content Caching Policy

Input: $\mathbb{P}_r(\mathcal{C}_{new}|\mathcal{T}_i), \mathbb{P}_r(\mathcal{C}^r_{old}|\mathcal{T}_i),\ i = 1, ..., m$
Output: D^r
$D^r \leftarrow \emptyset$
get $\mathcal{C}^{r\uparrow}_{old}$ and $\mathcal{C}^{\downarrow}_{new}$
while $\mathcal{C}^{\downarrow}_{new} \neq \emptyset$ **do**
  $c \leftarrow \mathcal{C}^{\downarrow}_{new}.pop()$
  if $l^c \leq L - L^r_{\mathcal{T}}$ **then**
    add (c, c_0) to D^r
    $L^r_{\mathcal{T}} \leftarrow L^r_{\mathcal{T}} + l^c$
  else if $\mathcal{C}^{r\uparrow}_{old} \neq \emptyset$ **then**
    $c' \leftarrow \mathcal{C}^{r\uparrow}_{old}[1]$
    if $l^c - l^{c'} \leq L - L^r_{\mathcal{T}}$ **and** $U^{c,c'}(m) > 0$ **then**
      add (c, c') to D^r
      $L^r_{\mathcal{T}} \leftarrow L^r_{\mathcal{T}} - l^{c'} + l^c$
      $\mathcal{C}^{r\uparrow}_{old}.pop()$

designed. On the other hand, OPEX mainly comes from two aspects: the backhaul link costs for either content caching or data prefetching; and the compensations for the vehicles that cannot use full wireless link time slots. Therefore, the backhaul link utilization aims to minimize the OPEX so as to increase the profit.

A. Wireless Link Slot Scheduling in Cached Infostation System

In the following, a buyer refers to a vehicle and an item refers to a wireless link slot. The auction is conducted at the beginning of each scheduling period $\mathcal{T}$ and is a direct revelation multi-item auction. The set of all the items in the period is represented by $\theta_{\mathcal{T}} = \{\theta^r_t : t \in \mathcal{T}, r \in \mathcal{R}\}$, where θ^r_t is the slot of r at t. For each item, its buyer group is different, which is given by $\mathcal{V}(\theta^r_t) = \{v : v \in \mathcal{V}, \rho^{v,r}_t > 0\}$, i.e., the set of vehicles within the coverage of r at t. Only the buyers in $\mathcal{V}(\theta^r_t)$ can bid for the item θ^r_t. The set of all the bids of a buyer v is denoted by $b_v = \{b^r_t : t \in \mathcal{T}, r \in \mathcal{R}, v \in \mathcal{V}(\theta^r_t)\}$, where b^r_t represents the bid for θ^r_t. The whole bids in $\mathcal{T}$ are denoted by $b_{\mathcal{T}} = \{b_v : v \in \mathcal{V}\}$. Likewise, the set of all the valuations of a buyer v is given by $q_v = \{q^r_t : t \in \mathcal{T}, r \in \mathcal{R}, v \in \mathcal{V}(\theta^r_t)\}$. We suppose that the valuation is mainly based on the flow size, i.e., $q^r_t \propto f_{v,r,t}$ for $q^r_t \in q_v$. In the auction, each buyer exclusively owns the item if it wins and the bidding result of each item is independent. Also, the buyers do not collude with each other.

We consider the goal of each buyer is the utility maximization, which is defined as below:

$$u_v = \sum_{w^r_t \in w_v, q^r_t \in q_v, p^r_t \in p_v} w^r_t (q^r_t - p^r_t) \tag{14}$$

where $w_v = \{w^r_t : t \in \mathcal{T}, r \in \mathcal{R}, v \in \mathcal{V}(\theta^r_t)\}$ denotes the bidding results of v and $p_v = \{p^r_t : t \in \mathcal{T}, r \in \mathcal{R}, v \in \mathcal{V}(\theta^r_t)\}$ denotes the actual payment for the items. $w^r_t = 1$ means v wins θ^r_t and the payment is p^r_t, and $w^r_t = 0$ means otherwise and the payment is 0. The utility of the operator, which is also called social welfare, is defined as below:

$$S = \sum_{v \in \mathcal{V}} \sum_{w^r_t \in w_v, q^r_t \in q_v} w^r_t q^r_t \tag{15}$$

The objective of the auction is to maximize the social welfare S, so the problem formulation is given as:

$$\max_{w_v, v \in \mathcal{V}} S$$

$$s.t. \sum_{v \in \mathcal{V}} w^r_t \leq 1, \forall \theta^r_t \in \theta_{\mathcal{T}} \tag{16}$$

$$\sum_{w^r_t \in w_v, q^r_t \in q_v} w^r_t q^r_t \leq \mathcal{B}_v, \forall v \in \mathcal{V} \tag{17}$$

where $\mathcal{B}_v$ is a constant related with the valuation of v and the constraints (6) or (7).

For an auction to be practical, the *strategy-proofness* property should be achieved. A strategy-proof auction mechanism should both satisfy *truthfulness* (or *incentive-compatibility*) and *individual rationality*, and the definitions are given below:

Truthfulness: An auction is truthful if and only if the dominant (best) strategy for each buyer is to truthfully bid its valuation, i.e., $u_v(q_v, b_{-v}) \geq u_v(b_v, b_{-v})$*, where* b_v *can be any bid and* b_{-v} *are the bids of other buyers.*

Individual rationality: An auction is individual rational if and only if for each buyer, its utility is not less than 0 *when it bids truthfully.*

Truthfulness guarantees that no buyer can manipulate its bid for the sake of higher utility and individual rationality ensures that each buyer is willing to attend the auction. Then, the wireless resource scheduling problem can be given as follows:

Find a strategy-proof auction mechanism to allocate the wireless link slots so as to maximize the social welfare.

The auction mechanism in this paper contains an allocation method and a pricing scheme. The allocation method aims to maximize the social welfare while the pricing scheme guarantees the strategy-proofness of the auction. A possible strategy-proof method is the VCG mechanism, but it requires to solve the social welfare maximization problem, which is still an ILP problem in our scenario and is NP-hard. Therefore, we design a heuristic algorithm by selecting items based on bids in a greedy manner. The procedures for the allocation are given in Alg. 4. First, all the bids from the buyers are sorted in a non-increasing order, denoted by $b^{\downarrow}_{\mathcal{T}}$. Then, the highest bid is selected (Line 3) and the corresponding buyer is found (Line 4). If the item has not been won by other buyers and the constraint (14) is not violated, i.e., $examine(\theta^r_t, v) = True$, then the buyer wins the item (Line 5-7).

Next, we discuss the pricing scheme for the auction. For the item θ^r_t, its buyer group is $\mathcal{V}(\theta^r_t)$, in which all the buyers' bids are larger than 0. Suppose that v wins the bid with $b^r_t \in b_v$, and the largest bid among the bids smaller than b^r_t is denoted as b^*. Then the payment of v is b^* if it exists, or 0 if is no bid smaller than b^r_t. By applying this pricing scheme together with the allocation method in Alg. 4, the auction is guaranteed to be strategy-proof, which is proved below.

Lemma 1: The auction in this work satisfies truthfulness.

Algorithm 4: Wireless Link Slot Scheduling—Network Operator's Perspective

```
Input: b_T^↓, θ_T
Output: ξ_T
ξ_t^{v,r} ← 0 for all t ∈ T, v ∈ V, r ∈ R
while b_T^↓ ≠ ∅ do
    b_t^r ← b_T^↓.pop()
    v ← {v : b_t^r ∈ b_v}
    if examine(θ_t^r, v) = True and θ_t^r ∈ θ_T then
        ξ_t^{v,r} ← 1 (allocate θ_t^r to v)
        remove θ_t^r from θ_T
```

Proof. First, we should prove that for a single item θ_t^r, $u_v(q_t^r, b_{-v}) \geq u_v(b_t^{r'}, b_{-v})$, where $q_t^r \in q_v$ and $b_t^{r'}$ is any bid for item θ_t^r. There exist totally four cases. In the first case, both q_t^r and $b_t^{r'}$ lose the bid. Then $u_v(q_t^r, b_{-v}) = u_v(b_t^{r'}, b_{-v}) = 0$. In the second case, both of them win the bid. Then $u_v(q_t^r, b_{-v}) = u_v(b_t^{r'}, b_{-v}) = q_t^r - b^*$. In the third case, q_t^r wins the bid while $b_t^{r'}$ loses. Then $u_v(q_t^r, b_{-v}) > 0 = u_v(b_t^{r'}, b_{-v})$. In the last case, q_t^r loses the bid while $b_t^{r'}$ wins. Then there must exist a $b_t^{r''} \in b_{v'} > q_t^r$. Also, we have $b_t^{r'} > b_t^{r''}$ for it to win the bid and the payment is $b^* = b_t^{r''}$. So $u_v(q_t^r, b_{-v}) = 0 > u_v(b_t^{r'}, b_{-v}) = q_t^r - b_t^{r''}$. Now, we have $u_v(q_t^r, b_{-v}) \geq u_v(b_t^{r'}, b_{-v})$ and therefore $u_v(q_v, b_{-v}) \geq u_v(b_v, b_{-v})$. Hence, the lemma is proved. ■

Lemma 2: The auction in this work satisfies individual rationality.

Proof. If v wins θ_t^r, then its utility for θ_t^r is $u_v(\theta_t^r) = q_t^r - p_t^r = b_t^r - b^*$, which is greater than 0 according to the pricing scheme. If it loses the bid, the utility is 0. Therefore, from Eq. (14), we have its total utility $u_v \geq 0$. Hence, the lemma is proved. ■

Theorem 1: The auction in this work is strategy-proof.

Proof. Combining Lemma 1 and Lemma 2, the theorem is proved. ■

B. Backhaul Link Utilization in Cached Infostation System

In order to maximize the total profit, the OPEX should be minimized. As mentioned earlier, the OPEX consists of compensations for vehicles and costs for using the backhaul link. In the auction, each vehicle evaluates and bids for a full time slot. If a winner vehicle cannot utilize the whole slot for data transmission, then the operator should compensate for the loss, based on the valuation of the winner. On the other hand, the use of backhaul link for content caching and data prefetching will bring costs. Therefore, we have the tradeoff property between compensations and costs. Specifically, with the increase of backhaul link usage, the costs become higher while the compensations become lower. Then, we can give the description of the backhaul link utilization problem:

Find a way to utilize the backhaul link in order to minimize the OPEX (sum of costs and compensations).

In this subsection, the vehicles only refer to those that win the bids and have download requests. For those vehicles whose requesting contents are already cached, no compensation needs to be paid. For the other vehicles, the potential compensation is denoted as $y_t^{v,r} \propto q_t^r$ ($q_t^r \in q_v$), which means higher compensation should be paid if the valuation is higher. In the case that the valuation is solely dependent on the link quality, the compensation is $y_t^{v,r} = \varepsilon \rho_t^{v,r} \tau$, where ε is the compensation price in the unit of $\$/bit$. Similarly, we suppose that the cost of backhaul link is determined by the transmitted data. Then, the cost of prefetching data is $z_t^{v,r} = \varepsilon' \rho_t^{v,r} \delta$ and the cost of caching a content is $z^c = \varepsilon' l^c$, where ε' is the price of backhaul link.

In order to evaluate the utilization of backhaul link, we define a metric as the ratio between the compensation reduction and cost (CR-C ratio), which gives the amount of compensations that can be reduced with a unit cost. The CR-C ratio should be greater than 1 so as to minimize OPEX. When the backhaul link is used for prefetching, the CR-C ratio is $y_t^{v,r}/z_t^{v,r}$. If it is used for content caching, then we need to consider the expected compensation reduction of the content in the future periods. For c, the value should be $y^c \propto (\sum_{i=1}^{m} h_{\mathcal{T}_i}^c)\overline{q_t^r}$ or $y^c \propto (\sum_{i=1}^{m} h_{\mathcal{T}_i}^c - \sum_{i=1}^{m} h_{\mathcal{T}_i}^{c'})\overline{q_t^r}$, where $\overline{q_t^r}$ is the average valuation of the vehicles that use the time slots. Then the CR-C ratio of content caching is y^c/z^c.

The procedures of backhaul link utilization are shown in Alg. 5. First, for each $r \in \mathcal{R}$, find the vehicles that can be prefetched and calculate their CR-C ratios assuming that their requests are prefetched. Also, calculate the CR-C ratios of the contents that have not been completely cached yet. Then, find the CR-C ratios larger than 1, sort them in a non-increasing order and obtain $\alpha^{r\downarrow}$ (Line 3). In each iteration, pop the current largest CR-R ratio α_1^r and reserve backhaul link slots for it (Line 4-6). Specifically, if α_1^r corresponds to a content caching, then reserve the needed slots for it as many as possible. If α_1^r corresponds to a prefetching, then reserve one slot. The reservation ends when there is no slot or $\alpha^{r\downarrow}$ is empty. Finally, the allocation of backhaul link slots is straightforward (Line 7). The slots used for prefetching are allocated first since they should be prior to the slots of corresponding vehicles. The remaining slots are allocated for content caching.

Algorithm 5: Backhaul Link Utilization—Network Operator's Perspective

```
  Input: ξ_T, C^r, r ∈ R
  Output: β_T
1 β_t^r ← β_t^{r,0} for all t ∈ T, r ∈ R
2 for r ∈ R do
3     generate α^{r↓}
4     while α^{r↓} ≠ ∅ and ∃β_t^{r,0} ∈ β^r do
5         α_1^r ← α^{r↓}.pop()
6         reserve backhaul link slots for α_1^r
7     allocate the slots
```

C. Tradeoff for Network Throughput

In this subsection, we will show that the operator can easily make the tradeoff between its total profit and the network throughput. On the one hand, we know that the wireless resource scheduling implicitly takes into account the network throughput, since the valuation and bid of the wireless link is mainly based on the flow size, and the auction process greedily selects the vehicles with larger bids. Therefore, if the operator wants to guarantee the network throughput, a straightforward tradeoff method is to add constraints on the other factors that can affect the valuation of the wireless resource. For example, by defining a range on the value of vehicles' access priority, the network throughput can be guaranteed to some extent, at the cost of more potential profits. On the other hand, for the backhaul link utilization, the CR-C threshold can be tuned (a value between 0 and 1) such that more slots are used for either caching or data prefetching, which can increase the network throughput at the cost of total OPEX increment.

VI. Simulation Results

We show the simulation results from the perspective of both vehicle users and network operator in this section. Matlab is used as the simulation tool to build the whole environment and test our algorithms. We present the values of important simulation parameters as shown in Tab. II. The environment is shown in Fig. 4, in which there are four crossing roads. The infostations are located at the intersections, and each infostation covers a 1km-radius circle. Vehicles are randomly distributed on the topology at the beginning, and then the Manhattan mobility model is used to simulate the microscopic mobility of each vehicle. The wireless resource includes four 10MHz channels and the transmission rate is referred from the IEEE 802.11p standard under speed limit of 60km/h [32], i.e. $\{27, 24, 18, 12, 9\}$Mbps for different signal-to-noise ratios. Zipf distribution is adopted to characterize the content popularity [33].

TABLE II
Simulation setups.

Parameter	Value	Parameter	Value
δ	0.1s	T	10
Vehicle density	20km^{-1}	Speed	[30,60]km/h
C	100	Upload ratio	20%
Content size	[40,60]MB	Upload size	[10,30]MB
OBU cache	200MB	Infostation cache	1000MB
τ	0.3δ	Backhaul rate	20MB/s
ϵ	0.9	f_{min}	{0.315, 0}MB
W	10	m	50

A. Vehicle Users' Perspective

We consider four different cases, which are described below:

- Case 1: This is the ideal case, which means the maximal transmission rate is applied each time slot.
- Case 2: There is no backhaul link used for content caching or data prefetching.
- Case 3: Suppose that all requests are either cached or prefetched, i.e., f_{opt}.
- Case 4: There is backhaul link used for content caching or data prefetching.

First, we suppose that the infostations do not cache any content. Fig. 5 shows the accumulated total flow at the end of simulation from each case is compared to case 1. From the figure, we can see that the total flow in case 4 is much larger than that in case 2 and is close to that in case 3. Hence, the advantage for utilizing backhaul link for data prefetching through Alg. 2 is demonstrated. Also, we can find that the total flow in case 3 is close to case 1, which validates the efficiency of the wireless resource scheduling given in Alg. 1.

When caching is considered, the results are given in Fig. 6. Compared with Fig. 5, it can be seen that the total flow in case 2 increases significantly, due to the existence of cached contents selected via Alg. 3 and delivered to the infostations through the backhaul link with Alg. 2. Fig. 7 illustrates the utilization of backhaul link, showing that content caching is finished soon after the update of contents, which is consistent with Alg. 2. Moreover, owing to the cached contents, the number of backhaul link slots used for data prefetching decreases dramatically, and so does the number of total used slots. Specifically, the backhaul link utilization ratio is 87.34% without caching, and this number drops down to 14.71% when caching is adopted.

Then, we evaluate the effect of vehicle speed in our system. The results are given in Fig. 8. The average speed is set to $\{$35,45,55$\}$km/h, respectively. From the figure, we can find that as the simulation goes on, the average total flow in each period becomes steady, and the results show that the total flow is smaller when the speed is higher. Intuitively, when a vehicle with higher speed passes an infostation, its sojourn time is shorter, thus the opportunity that it is allocated resources is also smaller. However, when there exist multiple infostations, the time that it spends beyond the coverage areas of the infostations is also smaller, which means it takes less time to get into another infostation. From Fig. 4, we know that the ratio between the road length that is covered by infostations and not is 16:12. Therefore, in a certain period of time, if the vehicle speed is higher, the ratio between the time that it spends within the coverage of an infostation and not is more likely to converge to the above ratio. In other words, if the above ratio is larger than 1, then vehicles with higher speed will have more chance to get more resources, which is shown in Fig. 8.

Next, we compare the total flow of case 4 with the case when the methods in [14] are adopted, as shown in Fig. 9. In [14], the cooperative content distribution system for vehicles (CCDSV) was proposed, which includes three major steps, namely representative access point selection, storage management, and estimate prefetching volume. Since the scenarios of our work and [14] are not exactly the same, we do some simplifications so as to adopt the CCDSV in our scenario, which can make it perform better while keep its core ideas. Specifically, given our relatively simple topology, accurate lookahead-AP prediction can be realized. We also aim to maximize the vehicular download volume (VDV). We suppose all the backhaul link slots are used for implementing CCDSV and assume infinite bandwidth such that the CCDSV decisions can be updated in the infostations without any delay. From Fig. 9, we can see that the methods in our work perform

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2921444, IEEE Transactions on Vehicular Technology

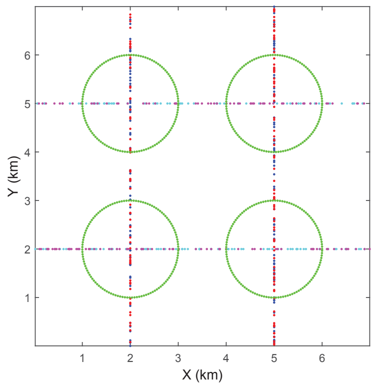


Fig. 4. A topology with 4 crossing roads and 4 Infostations, each covering a 1km-radius circle. Dots with different colors represent vehicles moving in different directions.

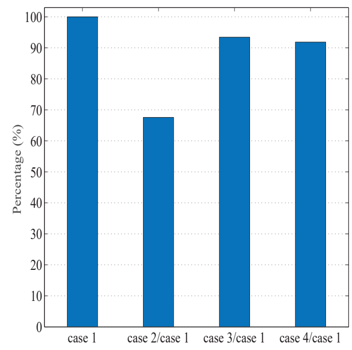


Fig. 5. Accumulated total flow compared to case 1 (without caching).

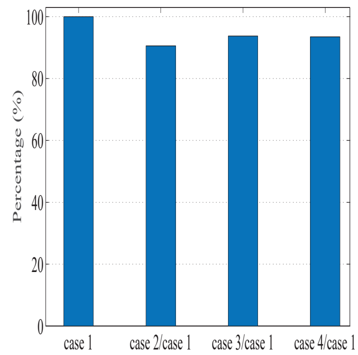


Fig. 6. Accumulated total flow compared to case 1 (with caching).

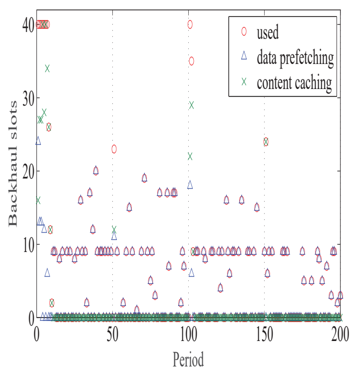


Fig. 7. Backhaul link utilization in each period.

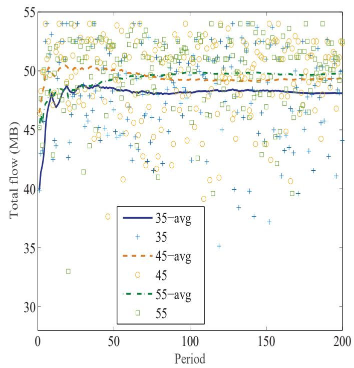


Fig. 8. Total flow each period with different vehicle speeds.

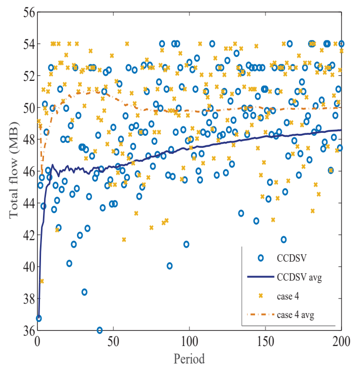


Fig. 9. Total flow each period compared with methods in [14].

better than CCDSV. There are several reasons for this. CCDSV prefetches data as cache for the vehicles that are coming into the infostation, which means the cache space is taken even before the vehicles enter the infostation. Also, the contents prefetched for the vehicles are unlikely to be used by others, if they are not the popular contents.

Finally, we evaluate the fairness among the vehicles through the Jain's fairness index. f_{min} is respectively set to 0 and 0.315 (the minimum data that can be transmitted in one slot), so as to validate the window-based fairness constraint. In Fig. 10 and Fig. 11, the Jain's index and accumulated total flow of each period are compared, respectively. The tradeoff between total flow and fairness is shown in the figures. Specifically, with less constraint of flow, the vehicles' fairness becomes better at the cost of total flow, since vehicles with worse link qualities get more chance to have the allocated resources.

B. Network Operator's Perspective

In the following simulation, we let each vehicle have a different access priority, which influences its valuation on the items. In Fig. 12, the accumulated social welfare at the end of simulation from three different cases is compared to the ideal case, which assumes that each item is sold at the

TABLE III
SCENARIOS WITH DIFFERENT CONTENT CACHING POLICIES.

Scenario No.	Description
1	$\overline{q_t^r} = 18$ (mean rate) CR-C threshold = 1 (following CR-C threshold)
2	$\overline{q_t^r} = 27$ (max rate) CR-C threshold = 1 (following CR-C threshold)
3	$\overline{q_t^r} = 18$ (mean rate) CR-C threshold = 0 (following Alg. 3)
4	$\overline{q_t^r} = 27$ (max rate) CR-C threshold = 0 (following Alg. 3)
5	No cache

highest valuation. It can be seen that the result from Alg. 4 (i.e., the strategy-proof case) is close to the result from the ideal case and is much better than the result from the random bidding process, thereby demonstrating the efficiency of Alg. 4. Furthermore, the CDF of vehicle utility is shown in Fig. 13 in order to validate the strategy-proofness of the auction. For the "maximal" case in the figure, we let a randomly selected buyer bid the item with the largest valuation among all the bids. It can be seen that when all the buyers bid truthfully, the utilities are above 0, satisfying the individual rationality

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2921444, IEEE Transactions on Vehicular Technology

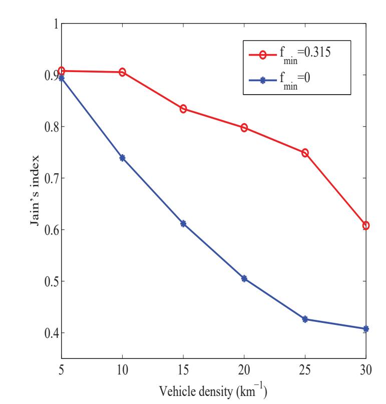


Fig. 10. Jain's fairness index with different f_{min}.

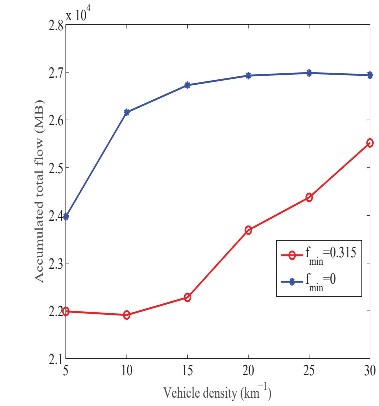


Fig. 11. Accumulated total flow with different f_{min}.

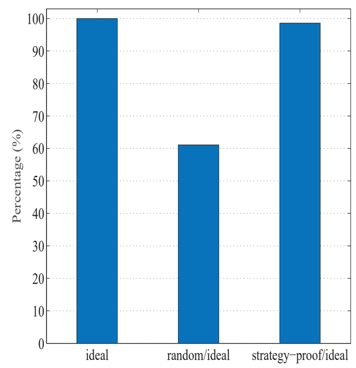


Fig. 12. Accumulated social welfare compared to the ideal case.

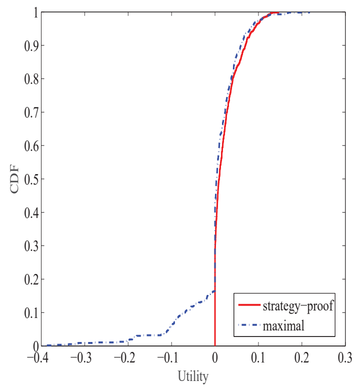


Fig. 13. CDF of vehicle utility.

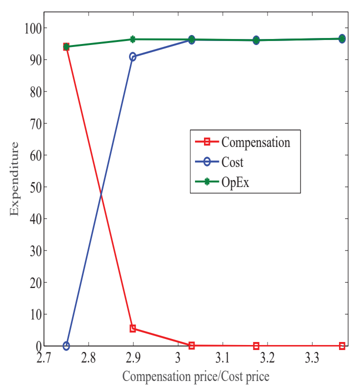


Fig. 14. Total compensations, costs and OPEX with different ratios between compensation price and cost price.

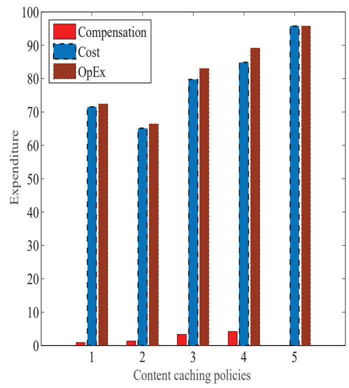


Fig. 15. Effect of content caching policy on compensations, costs and OPEX.

property. Also, these utilities are larger than the "maximal" case, demonstrating the truthfulness property.

We also evaluate the efficiency of the backhaul link utilization given in Alg. 5. Fig. 14 and Fig. 15 show the total OPEX and its components within a certain duration under different conditions. From Fig. 14, when the ratio between compensation price and cost price increases, the total compensations drop while total costs increase. Also, it can be seen that the transition is sensitive to the price ratio since buyers with higher priorities (and thereby higher bids and valuations) are more likely to win the bids, which means their CR-C ratios are higher. In Fig. 15, we fix the price ratio and evaluates the effect of content caching policy. Totally five different scenarios are compared and they are described in Tab. III. The smallest OPEX appears in the second scenario, which uses the $\overline{q_t^r}$ calculated from maximal rate and only caches the contents with CR-C ratios lower than the threshold. The reasons are that the benefits of cached contents can be truly evaluated by considering the link qualities of winners instead of all vehicles, and the unworthy contents with respect to OPEX minimization can be filtered.

VII. Conclusion

In this paper, we have investigated the dynamic link utilization problem within the 5G edge framework of cached infostation for cost-effective and efficient edge vehicular infotainment services. We have studied the framework from two different perspectives, i.e., vehicle users' perspective and network operator's perspective, respectively. For the vehicle users perspective, we aim to maximize the vehicular access network throughout. First, we have formulated the wireless resource scheduling problem to deal with the allocation of wireless link time slots, where a heuristic optimization method has been developed. Secondly, we have proposed the backhaul link utilization, which can be responsible for determining the usage of backhaul link slots for either content caching or data prefetching, and an optimal allocation method has been presented. In addition, we have proposed the content caching policy to decide which contents should be cached at the infostations, and an utility-based selection method has been given. For the network operator's perspective, we have modeled the wireless resource scheduling problem via the auction process, and a strategy proof auction mechanism has been developed. Also, an OPEX minimization goal has been achieved through a new backhaul link utilization method, which considers the tradeoff between compensations and costs. For our future work, we will investigate the potential of edge infostation cooperation to further improve the content delivery. We will also study the handover process if the coverage areas of infostations are overlapped.

ACKNOWLEDGMENT

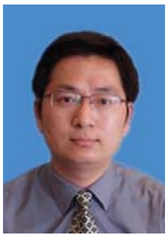
This work is supported in part by the National Natural Science Foundation of China (No. 61871211, No. 91638204), Natural Science Foundation of Jiangsu Province Youth Project (BK20180329) and Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] 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.
[2] T. S. Portal. (2019) Connected cars as a share of total vehicles 2015-2025. [Online]. Available: www.statista.com/statistics/275849/number-of-vehicles-connected-to-the-internet
[3] K. Abboud, H. A. Omar, and W. Zhuang, "Interworking of DSRC and cellular network technologies for V2X communications: A survey," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9457–9470, 2016.
[4] N. Cheng, N. Lu, N. Zhang, X. Shen, and J. W. Mark, "Vehicular wifi offloading: Challenges and solutions," *Vehicular Communications*, vol. 1, no. 1, pp. 13–21, 2014.
[5] K. Zhang, S. Leng, Y. He, S. Maharjan, and Y. Zhang, "Cooperative content caching in 5G networks with mobile edge computing," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 80–87, 2018.
[6] X. Huang, R. Yu, J. Kang, Y. He, and Y. Zhang, "Exploring mobile edge computing for 5G-enabled software defined vehicular networks," *IEEE Wireless Communications*, vol. 24, no. 6, pp. 55–63, 2017.
[7] H. Zhou, N. Cheng, N. Lu, L. Gui, D. Zhang, Q. Yu, F. Bai, and X. Shen, "Whitefi infostation: Engineering vehicular media streaming with Geolocation database," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 8, pp. 2260–2274, 2016.
[8] H. Liang and W. Zhuang, "Efficient on-demand data service delivery to high-speed trains in cellular/infostation integrated networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 4, pp. 780–791, 2012.
[9] Z. Ning, X. Hu, Z. Chen, M. Zhou, B. Hu, J. Cheng, and M. S. Obaidat, "A cooperative quality-aware service access system for social internet of vehicles," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2506–2517, 2018.
[10] S. Panichpapiboon and W. Pattara-Atikom, "A review of information dissemination protocols for vehicular ad hoc networks," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 3, pp. 784–798, 2012.
[11] F. Malandrino, C. Casetti, C. Chiasserini, and M. Fiore, "Optimal content downloading in vehicular networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 7, pp. 1377–1391, 2013.
[12] ——, "Content download in vehicular networks in presence of noisy mobility prediction," *IEEE Transactions on Mobile Computing*, vol. 13, no. 5, pp. 1007–1021, 2014.
[13] T. Wang, L. Song, and Z. Han, "Coalitional graph games for popular content distribution in cognitive radio VANETs," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 8, pp. 4010–4019, 2013.
[14] D. Zhang and C. K. Yeo, "Enabling efficient wifi-based vehicular content distribution," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 3, pp. 479–492, 2013.
[15] W. Saad, Z. Han, A. Hjorungnes, D. Niyato, and E. Hossain, "Coalition formation games for distributed cooperation among roadside units in vehicular networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 1, pp. 48–60, 2011.
[16] Y. Li, X. Zhu, D. Jin, and D. Wu, "Multiple content dissemination in roadside unit aided vehicular opportunistic networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 8, pp. 3947–3956, 2014.
[17] F. Sun, B. Liu, F. Hou, H. Zhou, J. Chen, Y. Rui, and L. Gui, "A QoE centric distributed caching approach for vehicular video streaming in cellular networks," *Wireless Communications and Mobile Computing*, vol. 16, no. 12, pp. 1612–1624, 2015.
[18] F. Mezghani, R. Dhaou, M. Nogueira, and A.-L. Beylot, "Content dissemination in vehicular social networks: Taxonomy and user satisfaction," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 34–40, 2014.
[19] M. O. Cherif, S.-M. Senouci, and B. Ducourthial, "Efficient data dissemination in cooperative vehicular networks," *Wireless Communications and Mobile Computing*, vol. 13, no. 12, pp. 1150–1160, 2013.
[20] T. H. Luan, L. X. Cai, J. Chen, X. Shen, and F. Bai, "Engineering a distributed infrastructure for large-scale cost-effective content dissemination over urban vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1419–1435, 2014.
[21] J. Chen, B. Liu, H. Zhou, L. Gui, and F. Sun, "Engineering link utilization in cellular offloading oriented VANETs," *In Proceeding of the IEEE GLOBECOM*, pp. 1–6, 2015.
[22] H. Zhou, N. Zhang, Y. Bi, Q. Yu, X. Shen, D. Shan, and F. Bai, "TV white space enabled connected vehicle networks: Challenges and solutions," *IEEE Network*, vol. 31, no. 3, pp. 6–13, 2017.
[23] T. Liu, J. Li, F. Shu, M. Tao, W. Chen, and Z. Han, "Design of contract-based trading mechanism for a small-cell caching system," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6602–6617, 2017.
[24] 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.
[25] E. Lim and S. Kim, "Transmission rate scheduling with fairness constraints in downlink of CDMA data networks," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 1, pp. 328–337, 2005.
[26] P. Wang and W. Zhuang, "A token-based scheduling scheme for WLANs supporting voice/data traffic and its performance analysis," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, 2008.
[27] M. Tao, W. Yu, W. Tan, and S. Roy, "Communications, caching, and computing for content-centric mobile networks: Part 1 [guest editorial]," *IEEE Communications Magazine*, vol. 54, no. 8, pp. 14–15, 2016.
[28] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Communications of the ACM*, vol. 53, no. 8, pp. 80–88, 2010.
[29] Y. Borghol, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti, "The untold story of the clones: Content-agnostic factors that impact YouTube video popularity," in *In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1186–1194.
[30] Y. Chen, J. Zhang, K. Wu, and Q. Zhang, "Tames: A truthful double auction for multi-demand heterogeneous spectrums," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 11, pp. 3012–3024, 2014.
[31] Q. Wang, B. Ye, S. Lu, and S. Guo, "A truthful QoS-aware spectrum auction with spatial reuse for large-scale networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 10, pp. 2499–2508, 2014.
[32] Y. L. Morgan, "Notes on DSRC & WAVE standards suite: Its architecture, design, and characteristics," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 4, pp. 504–518, 2010.
[33] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "Analyzing the video popularity characteristics of large-scale user generated content systems," *IEEE/ACM Transactions on Networking*, vol. 17, no. 5, pp. 1357–1370, 2009.

Haibo Zhou (M'14-SM'18) received the Ph.D. degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. From 2014 to 2017, he worked as a Postdoctoral Fellow with the Broadband Communications Research Group, Electrical and Computer Engineering Department, University of Waterloo. He is currently an Associate Professor with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests include resource management and protocol design in cognitive radio networks and vehicular networks.

Nan Cheng (M'16) received the B.E. and the M.S. degrees from Tongji University, Shanghai, China, in 2009 and 2012, respectively, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, in 2015. He is currently working as a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, under the supervision of Prof. Ben Liang. His current research focuses on big data in vehicular networks and self-driving systems. His research interests also include performance analysis, MAC, opportunistic communication for vehicular networks, cognitive radio, WiFi, smart grid, and cellular traffic offloading.



Jian Wang (M'11) is currently an Associate Professor in the School of Electronic Science and Engineering and Institute of Space-Terrestrial Intelligent Networks. He received a Ph.D. degree in 2006 from Nanjing University, China. His research interests include Mobile Ad-Hoc Networks, Spatial Information Network, Satellite Communications and Networks, Video Coding and Transmission.

Jiacheng Chen received his Ph.D. degree in information and communications engineering from Shanghai Jiao Tong University, Shanghai, China, in 2018. From Dec. 2015 to Dec. 2016, he was a visiting scholar at BBCR group, University of Waterloo, Canada. Currently, he is an assistant researcher in Peng Cheng Laboratory, Shenzhen, China. His research interests include future network design, 5G/6G network, and resource management.

Quan Yu received the B.S. degree in information physics from Nanjing University in 1986, the M.S. degree in radio wave propagation from Xidian University in 1988, and the Ph.D. degree in fiber optics from the University of Limoges in 1992. Since 1992, he joined the faculty of the Institute of China Electronic System Engineering Corporation, where he has been a Senior Engineer, and currently a Research Fellow. His main areas of research interest are the architecture of wireless networks, optimization of protocols and cognitive radios. Dr. Yu is an academician of Chinese Academy of Engineering (CAE).

Xuemin (Sherman) Shen (M'97-SM'02-F'09) received the B.Sc. degree from Dalian Maritime University, Dalian, China, in 1982, and the M.Sc. and Ph.D. degrees from Rutgers University, New Brunswick, NJ, USA, in 1987 and 1990, respectively, all in electrical engineering. He is the University Professor and an Associate Chair of Graduate Studies, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research focuses on resource management, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. He served as the Technical Program Committee Chair/Co-Chair for the IEEE Global Communications Conference (Globecom) 2016, International Conference on Computer Communication 2014, the IEEE Vehicular Technology Conference (VTC) 2010 Fall, and Globecom 2007, the Symposia Chair for the IEEE International Conference on Communications (ICC) 2010, the Tutorial Chair for the IEEE VTC 2011 Spring, and the IEEE ICC 2008, the General Co-Chair for ACM Mobihoc 2015, Chinacom 2007 and the Chair for the IEEE Communications Society Technical Committee on Wireless Communications. He also serves/served as an Editor-in-Chief for the IEEE Internet of Things Journal, the IEEE Network, Peer-to-Peer Networking and Application, and IET Communications; a Founding Area Editor for the IEEE Transactions on Wireless Communications; an Associate Editor for the IEEE Transactions on Vehicular Technology, Computer Networks, and ACM/Wireless Networks, etc.; and the Guest Editor for the IEEE JSAC, IEEE Wireless Communications, and the IEEE Communications Magazine, etc. He was the recipient of the Excellent Graduate Supervision Award in 2006, and the Premiers Research Excellence Award in 2003 from the Province of Ontario, Canada. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.