

Network Slicing Enabled Resource Management for Service-Oriented Ultra-Reliable and Low-Latency Vehicular Networks

Yuanbin Chen, Ying Wang, *Member, IEEE*, Man Liu,
Jiayi Zhang, *Senior Member, IEEE*, and Lei Jiao, *Member, IEEE*

Abstract—Network slicing has been considered as a promising candidate to provide customized services for vehicular applications that have extremely high requirements of latency and reliability. However, the high mobility of vehicles poses significant challenges to resource management in such a stochastic vehicular environment with time-varying service demands. In this paper, we develop an online network slicing scheduling strategy for joint resource block (RB) allocation and power control in vehicular networks. The long-term time-averaged total system capacity is maximized while guaranteeing strict ultra-reliable and low-latency requirements of vehicle communication links, subject to stability constraints of task queues. The formulated problem is a mixed integer nonlinear stochastic optimization problem, which is decoupled into three subproblems by leveraging Lyapunov optimization. In order to tackle this problem, we propose an online algorithm, namely JRPSV, to obtain the optimal RB allocation and power control at each time slot according to the current network state. Furthermore, rigorous theoretical analysis is conducted for the proposed JRPSV algorithm, indicating that the system capacity and the system average latency obey a $[O(1/V), O(V)]$ trade-off with the control parameter V . Extensive simulation results are provided to validate the theoretical analysis and demonstrate the effectiveness of JRPSV as well as the impacts of various parameters.

Index Terms—Vehicular networks, network slicing, resource management, reliability, latency, stochastic optimization.

I. INTRODUCTION

GIVEN that vehicles have evolved into an obligato part of people’s modern life, vehicular networks, driven by novel technology such as big data analysis and artificial intelligence, have developed rapidly in recent years [1]. The time people usually spend on vehicles (i.e., car, bus, or train) daily can last several hours or more. In this context, it is expected to be able to enjoy high-quality entertainment services (e.g.,

watching movies) as well as information services (e.g., road safety and traffic news) during the travelling [2]. The emergence of vehicle-to-everything (V2X) communications aims to make people’s daily travelling safer and more convenient, thereby paving the way for cutting edge intelligent applications such as intelligent transportation systems, intelligent vehicle interconnection and autonomous driving.

The advent of the fifth-generation (5G) mobile communication network has greatly encouraged V2X communications. Vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communications aim to realize information exchange, which makes enhanced mobile broad band (eMBB) and ultra reliable and low latency communications (URLLC) widely used [3]. Cloud access, streaming video and vehicle social networks that usually involve a large amount of data exchange need to frequently access to radio access points and core network servers so as to acquire data, which desires high-speed transmission of communication links for eMBB applications. At the same time, safety-critical information, such as cooperative awareness messages (CAM), decentralized environmental notification messages (DENM) and messages for autonomous driving in vehicular networks naturally requires communication links to be reliable and low latency.

To meet the strict reliability requirements of vehicular applications, many studies have been carried out. For example, radio resource management (RRM) [4] [5] issues have been investigated in vehicular networks and the total capacity of V2I links is maximized while guaranteeing the signal-to-interference-plus-noise ratio (SINR) of V2V links [6] [7]. The RRM scheme in [8] [9] can still achieve the SINR reliability requirement of V2V links. It should be pointed out that although the above literature has considered the reliability of V2V links from the perspective of SINR outage probability, in the face of diverse and heterogeneous applications in vehicular networks, it is not sufficient to consider only the reliability of communication links. Therefore, low-latency requirements of the information exchanged through V2I and V2V links also deserve serious attention. Given that the latency caused by data packets queuing in the buffer dominates the end-to-end latency [10], it is necessary to employ the queuing theory to study the average latency of the system. In [11], the network is modeled using $M/M/m$ queues and the average queuing delay is analyzed. It is studied in [12] that the latency based on the queuing theory and the steady-state latency expression of vehicle communication links is derived. In [13], the latency of V2V

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This work is supported by National Key Research and Development Program of China (2017YFE0120600), in part by National Natural Science Foundation of China under Grant 61971027. (*Corresponding author: Ying Wang.*)

Y. Chen, Y. Wang and M. Liu are with State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China 100876 (e-mail: chen_yuanbin@163.com; wangyinj@bupt.edu.cn; ieliuman@163.com).

J. Zhang is with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China (e-mail: jiaiyizhang@bjtu.edu.cn).

L. Jiao is with Department of Information and Communication Technology, University of Agder, N-4879 Grismstad, Norway (e-mail: lei.jiao@uia.no).

links has been described by the average packet sojourn time that consists of the queuing time plus the time for transmission and retransmission. Although the average queuing latency has been studied from different perspectives [11]–[13], average latency is just the first moment of random latency experienced by packets so that the tail behavior of the random latency cannot be known [14], which does not accurately correspond to practical situations sometimes. Therefore, latency violation probability characterizing the tail behavior of packet latency, especially in vehicular communication links, is worthy of special consideration.

In view of the above mentioned requirements and considering the scarce spectrum resources, the way to reasonably allocate radio resource for each vehicular application has been examined by many researchers. Network slicing, which is one of the key technologies of 5G, is a powerful means for resource allocation due to its customized services and flexible scheduling capabilities, especially considering the challenges posed by diverse and heterogeneous vehicular applications as well as wireless channel uncertainty caused by high-speed movement in vehicular networks. In this scenario, certain effective radio access network (RAN) slicing scheduling schemes have been proposed [15]–[17]. In short, network slicing technology is an effective approach to conquer challenges brought by the heterogeneity and diversity of applications in vehicular networks. However, most of the existing studies utilizing network slicing have not considered the stochastic nature of communication traffic.

Due to the fact that vehicular networks operate in a stochastic environment and vehicular applications arrive randomly in time and space domains, the future network status can hardly be predicted. In particular, the dynamic behaviors of vehicular applications such as arrival and transmission are random processes over the time. In addition, the coupling of random arrival tasks in the time dimension cannot be ignored, which makes long-term system performance more important than that of short-term. Considering the efficiency and the flexibility of network slicing technology in resource scheduling, it can be employed for resolving the challenges of diversity and heterogeneity in vehicular applications with stochastic nature. Therefore, it is of great importance to make resource allocation decision delicately in a stochastic environment without foreseeing the future network status.

Inspired by the above-mentioned open issues, in this paper, we focus on the radio resource allocation by employing network slicing technology in stochastic vehicular networks to fulfill the demand of heterogeneous applications. The long-term time-averaged system capacity is maximized while satisfying the high-rate, ultra-reliable and low-latency QoS requirements of diverse applications at the same time. Our contributions are summarized as follows:

- According to the QoS requirements of various applications in vehicular networks, two types of network slicing, namely eMBB slicing and URLLC slicing, are constructed to provide customized services for vehicular applications with high-rate transmission, low latency and high reliability. Considering that vehicular networks operate in a stochastic environment where applications

arrive randomly, RB allocation and power control are tuned jointly in order to maximize the long-term time-averaged system capacity, subject to the constraints of latency violation probability and outage probability as well as stability of task queues.

- The studied problem is a mixed integer nonlinear stochastic optimization problem. In order to solve this problem, Markov inequalities and variable substitution are firstly used to transform the probabilistic latency constraints. Then leveraging Lyapunov optimization theory, the stochastic optimization problem is decoupled into three sub-problems, which are optimized alternately to obtain the optimal solution at each time slot. Furthermore, we propose a joint RB allocation and power control slicing scheduling algorithm called JRPSV. The JRPSV algorithm provides an online slicing resource scheduling strategy that can instantaneously determine the optimal RB allocation and power control strategy without any prior knowledge of future task arrival or network status information.
- We evaluate the JRPSV algorithm from the perspectives of stability, convergence and optimality, respectively. The theoretical analysis results show that the JRPSV algorithm has a convergence behavior when searching for optimal solutions at each time slot, which can achieve an asymptotic optimal system capacity while enforcing the stability of the networks. In addition, the simulation results validate the theoretical analysis, demonstrate the effectiveness of the proposed JRPSV algorithm as well as the impact of QoS parameters on system performance. Compared with the set baselines, when the constraints of latency and reliability are completely ignored, the JRPSV algorithm can reduce the time-averaged queue length by at least 20.74%, and achieve a trade-off between the optimal system capacity and the average latency.

The remainder of this paper is organized as follows. Section II introduces the system model and problem formulation. The problem is transformed and reformulated in Section III. In Section IV, we propose an effective iterative scheme to obtain the approximate optimal solution, while Section V theoretically analyzes the optimality and stability of the proposed JRPSV. Simulation results are provided in Section VI to evaluate the performance of JRPSV. Finally, Section VII concludes the paper.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. Vehicular Network Slicing Model

As shown in Fig. 1, we consider a vehicular network with ground vehicles and a base station (BS), where the vehicular communications can be divided into two types: V2I and V2V communications. The V2I links are the communication links between the vehicles and the BS with high-speed uplink connections. The V2V communication links are the communication links among the vehicles, which are designed as ultra-reliable and low-latency communication links. In this network, two types of network slicing are constructed: eMBB slicing and URLLC slicing. The V2I links that perform high-speed

transmission to the BS are served by the eMBB slicing. Since some V2I links have strict requirements for latency [18], not only can URLLC slicing carries V2I services, but also V2V services.

The vehicular network consists of M cellular vehicle user equipments (VUEs) served by eMBB slicing and L cellular VUEs served by URLLC slicing, which are denoted by C^e -VUEs and C^u -VUEs, respectively and all of which connect with BS in V2I communication form. Denote C^e -VUE set and C^u -VUE set by $\mathcal{M} = \{1, 2, \dots, M\}$ and $\mathcal{L} = \{1, 2, \dots, L\}$, respectively. Device to device vehicle user equipments (D-VUEs) communicate with each other directly through device-to-device (D2D) mode when they are close to each other, and denote the D2D-V2V pairs set by $\mathcal{N} = \{1, 2, \dots, N\}$. All smart VUEs can select V2I and V2V communication modes based on demand. We assume that all vehicles in this paper are equipped with a single antenna.

Denote the radio frequency resource RB set by $\mathcal{K} = \{1, 2, \dots, K\}$, where K is the total number of RBs and $K \geq M + L$ needs to be satisfied. The RBs are designed to be allocated to C^e -VUEs, C^u -VUEs and D2D-V2V pairs in vehicular communication networks, respectively. The RB resource subsets allocated to C^e -VUEs constitute eMBB slicing, and URLLC slicing consists of RB resource subsets allocated to C^u -VUEs. The network adopts orthogonal frequency division multiple access (OFDMA) technology to serve VUEs in the uplink scenario, and the RBs in each slicing are orthogonal to each other.

Let $g_{m,B}$ and $g_{l,B}$ denote the channel gains from the m th C^e -VUE and the l th C^u -VUE to the BS, respectively. Let g_n denote the channel gain of the desired transmission for the n th D2D-V2V pair. Let $g_{l,n,B}$ denote the interference channel gain from the n th D2D-V2V transmitter to the BS on the RB of the l th C^u -VUE and $g_{l,n}$ denote the interference channel gain from the l th C^u -VUE to the n th D2D-V2V receiver. Without loss of generality, it is assumed that all channel gains include path loss, shadow fading and small-scale fading [7].

The vehicular network operates over a continuous period of time, which is divided into discrete time slots denoted by $\mathcal{T} = \{0, 1, 2, \dots, T-1\}$, $t \in \mathcal{T}$, and the duration of each time slot is considered as unit time.

B. Vehicular Network Communication Model

1) *eMBB Slicing*: The RB resource in the eMBB slicing can provide services for V2I links with high-rate connections and wide band access. The rules of RB allocation we define is that the RBs in the eMBB slicing can only be used by one C^e -VUE, and cannot be reused by other VUEs. The binary variable $x_{m,k}^e(t) \in \{0, 1\}$ is adopted to indicate whether the k th RB is allocated to the m th C^e -VUE at the time slot t . If the allocation action is performed, $x_{m,k}^e(t) = 1$ and vice versa. Then, the received instantaneous uplink signal-to-noise ratio (SNR) at the BS for the m th C^e -VUE is given by

$$\gamma_{m,k}^{c,e}(t) = \frac{P_{m,k}^{c,e}(t) g_{m,B}(t)}{\sigma^2}, \quad (1)$$

where $P_{m,k}^{c,e}(t)$ indicates the instantaneous transmit power of the m th C^e -VUE when occupying the k th RB at time slot

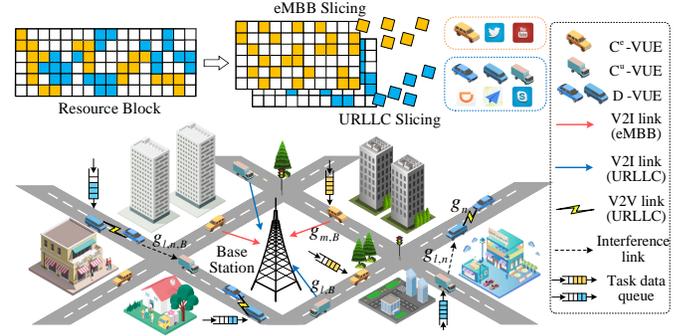


Fig. 1. Service-oriented vehicular network slicing.

t , and $0 \leq P_{m,k}^{c,e}(t) \leq P_{\max}^{c,e}$ holds. Here, the superscript “ c ” means cellular link. σ^2 is the noise power, and it is assumed that all VUEs have the same noise power. The instantaneous uplink transmission rate of the m th C^e -VUE can be given by

$$R_{m,k}^{c,e}(t) = \log_2 \left(1 + \gamma_{m,k}^{c,e}(t) \right), \quad (2)$$

and the maximum achievable uplink transmission rate of the m th C^e -VUE is $R_m^{c,e}(t) = \sum_{k \in \mathcal{K}} x_{m,k}^e(t) R_{m,k}^{c,e}(t)$.

Let $A_m^{c,e}(t) \in [0, A_{\max}^{c,e}]$ be the data arrival rate of the m th C^e -VUE at time slot t , and it is assumed that $\{A_m^{c,e}(t) | \forall t \geq 0\}$ obeys Poisson process while $A_m^{c,e}(t)$ in the different time slots is independent and identically distributed (i.i.d.). Then the data queue backlog of the m th C^e -VUE at time slot t is denoted by $Q_m^{c,e}(t)$, which evolves according to

$$Q_m^{c,e}(t+1) = \max \{Q_m^{c,e}(t) - R_m^{c,e}(t), 0\} + A_m^{c,e}(t). \quad (3)$$

By definition in [19], queue stability needs to satisfy

$$\bar{Q}_m^{c,e} = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E[Q_m^{c,e}(t)] < \infty. \quad (4)$$

It should be noted that the queue backlog has the same meaning as queue length, and in the context we will interchangeably use these two terms.

2) *URLLC Slicing*: With respect to URLLC services, there are both V2I and V2V communication links. The RBs in the URLLC slicing can be allocated to C^u -VUEs and D2D-V2V pairs on demand, and each VUE can only select one communication mode. Due to the fact that the number of V2V links is far greater than that of V2I links [6], the uplink RB resource in the URLLC slicing can be directly used by C^u -VUEs, and can be reused by two close D-VUEs. Let binary variables $x_{l,k}^u(t), x_{n,k}^u(t) \in \{0, 1\}$ indicate whether the k th RB is allocated to the l th C^u -VUE and the n th D2D-V2V pair, respectively. Then, the uplink SINR of the l th C^u -VUE at time slot t can be given by

$$\gamma_{l,k}^{c,u}(t) = \frac{P_{l,k}^{c,u}(t) g_{l,B}(t)}{\sum_{n \in \mathcal{N}} x_{n,k}^u(t) P_{n,k}^{d,u}(t) g_{l,n,B}(t) + \sigma^2}, \quad (5)$$

where $P_{l,k}^{c,u}(t)$ and $P_{n,k}^{d,u}(t)$ denote the transmit power of the m th C^u -VUE and the transmitter of the n th D2D-V2V pair on the k th RB at time slot t , and $0 \leq P_{l,k}^{c,u}(t) \leq$

$P_{\max}^{c,u}, 0 \leq P_{n,k}^{d,u}(t) \leq P_{\max}^{d,u}$ hold. Here, the superscript “ d ” in $P_{n,k}^{d,u}(t)$ means D2D-V2V link. The received uplink SINR at the receiver of the n th D2D-V2V pair when the RB of the l th C^u -VUE is reused can be given by

$$\gamma_{n,k}^{d,u}(t) = \frac{P_{n,k}^{d,u}(t) g_n(t)}{\sum_{l \in \mathcal{L}} x_{l,k}^u(t) P_{l,k}^{c,u}(t) g_{l,n}(t) + \sigma^2}. \quad (6)$$

Then the uplink transmission rate of the l th C^u -VUE and the n th D2D-V2V pair can be given by

$$R_{l,k}^{c,u}(t) = \log_2 \left(1 + \gamma_{l,k}^{c,u}(t) \right), \quad (7a)$$

$$R_{n,k}^{d,u}(t) = \log_2 \left(1 + \gamma_{n,k}^{d,u}(t) \right). \quad (7b)$$

Similarly, the maximum achievable uplink transmission rate of the l th C^u -VUE and the n th D2D-V2V pair are $R_l^{c,u}(t) = \sum_{k \in \mathcal{K}} x_{l,k}^u(t) R_{l,k}^{c,u}(t)$ and $R_n^{d,u}(t) = \sum_{k \in \mathcal{K}} x_{n,k}^u(t) R_{n,k}^{d,u}(t)$, respectively.

Let $A_l^{c,u}(t) \in [0, A_{\max}^{c,u}]$ and $A_n^{d,u}(t) \in [0, A_{\max}^{d,u}]$ be the data arrival rate of the l th C^u -VUE and the n th D2D-V2V pair respectively, while $A_l^{c,u}(t)$ and $A_n^{d,u}(t)$ are i.i.d. over different time slots. The mean arrival rate of Poisson process $\{A_l^{c,u}(t) | \forall t \geq 0\}$ and $\{A_n^{d,u}(t) | \forall t \geq 0\}$ can be denoted by $\lambda_l^{c,u}$ and $\lambda_n^{d,u}$, respectively. Let $Q_l^{c,u}(t)$ and $Q_n^{d,u}(t)$ be the data queue backlog of the l th C^u -VUE and the n th D2D-V2V pair, which evolve according to

$$Q_l^{c,u}(t+1) = \max \{Q_l^{c,u}(t) - R_l^{c,u}(t), 0\} + A_l^{c,u}(t), \quad (8a)$$

$$Q_n^{d,u}(t+1) = \max \{Q_n^{d,u}(t) - R_n^{d,u}(t), 0\} + A_n^{d,u}(t). \quad (8b)$$

Similar to (4), the queue stability conditions that should be satisfied are $\bar{Q}_l^{c,u} < \infty$ and $\bar{Q}_n^{d,u} < \infty$, respectively.

3) *Vehicle Communication Mode*: Based on the above analysis, we can divide the communication mode of vehicles into reused mode and unreused mode.

In the reused mode, the n th D2D-V2V pair usually reuses the RB of the l th C^u -VUE, and this characteristic can be described as $0 \leq \sum_{k \in \mathcal{K}} x_{l,k}^u(t) \leq 1$ and $0 \leq \sum_{k \in \mathcal{K}} x_{n,k}^u(t) \leq 1$.

The RB allocated to the C^u -VUE can only be reused by one D2D-V2V pair at most, which can be expressed as $0 \leq \sum_{l \in \mathcal{L}} x_{l,k}^u(t) \leq 1$ and $0 \leq \sum_{n \in \mathcal{N}} x_{n,k}^u(t) \leq 1$.

In the unreused mode, the C^e -VUE, the C^u -VUE or the D2D-V2V pair occupies one RB for communication and the RB is not allocated to any other VUEs. Considering that one RB is allocated to each C^e -VUE at most, we have $0 \leq \sum_{k \in \mathcal{K}} x_{m,k}^e(t) \leq 1$. In addition, eMBB and URLLC slicing cannot own the same RB, and the RBs in the eMBB slicing cannot be reused by VUEs in the URLLC slicing. Therefore, the rules can be expressed as $0 \leq \sum_{m \in \mathcal{M}} x_{m,k}^e(t) + \sum_{l \in \mathcal{L}} x_{l,k}^u(t) \leq 1$ and $0 \leq \sum_{m \in \mathcal{M}} x_{m,k}^e(t) + \sum_{n \in \mathcal{N}} x_{n,k}^u(t) \leq 1$.

C. QoS Requirements for Vehicular Networks

1) *QoS Requirements for eMBB Slicing*: eMBB vehicular applications require very high data transmission rate, so the

data transmission rate of the m th C^e -VUE needs to satisfy the transmission rate threshold $R^{e,\text{req}}$, which can be described by

$$R_{m,k}^{c,e}(t) \geq R^{e,\text{req}}. \quad (9)$$

2) *QoS Requirements for URLLC Slicing*: With respect to URLLC services, the requirements of low latency and high reliability must be satisfied. According to Little’s theorem, when the system reaches steady state and the arrival rate of the data traffic is constant, the average latency is proportional to the data queue length [19]. The average latency, i.e., the first moment of random latency experienced by packets, is not that accurate to portray the tail behavior of packets, especially considering random actions of packets. Thus the latency violation probability is introduced to constrain latency. The latency violation probability should be less than the tolerable threshold, which can be expressed by

$$\lim_{t \rightarrow \infty} \Pr(Q_l^{c,u}(t) \geq Q^{c,\text{req}}) \leq P_{\max}^{c,\text{laten}}, \quad (10a)$$

$$\lim_{t \rightarrow \infty} \Pr(Q_n^{d,u}(t) \geq Q^{u,\text{req}}) \leq P_{\max}^{u,\text{laten}}, \quad (10b)$$

where $Q^{c,\text{req}}$ and $Q^{u,\text{req}}$ are queue length thresholds with latency constraints. $P_{\max}^{c,\text{laten}}$ and $P_{\max}^{u,\text{laten}}$ indicate tolerable threshold probabilities that meet different low latency requirements of URLLC services.

Then, the reliability requirement of the D2D-V2V pair n can be satisfied by controlling the outage probability, which can be given by

$$\Pr(R_{n,k}^{d,u}(t) \leq R^{u,\text{req}}) \leq P_{\max}^{\text{outage}}, \quad (11)$$

where $R^{u,\text{req}}$ is the target threshold for the transmission rate of the V2V link, and P_{\max}^{outage} is tolerable outage probability.

D. Problem Formulation

Let $\mathbf{X} = \{x_{m,k}^e(t), x_{l,k}^u(t), x_{n,k}^u(t), \forall m, l, n, k\}$ and $\mathbf{P} = \{P_{m,k}^{c,e}(t), P_{l,k}^{c,u}(t), P_{n,k}^{d,u}(t), \forall m, l, n, k\}$ be the RB allocation variable matrix and vehicle transmit power matrix, respectively. The object is to maximize the long-term time-averaged system capacity, and the optimization variables are the transmit power of all VUEs (i.e. \mathbf{P}) and the RB allocation variable (i.e. \mathbf{X}). The optimization problem is formulated as follows:

$$\max_{\{\mathbf{X}, \mathbf{P}\}} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E \left[\sum_{m \in \mathcal{M}} R_{m,k}^{c,e}(t) + \sum_{l \in \mathcal{L}} R_{l,k}^{c,u}(t) + \sum_{n \in \mathcal{N}} R_{n,k}^{d,u}(t) \right] \quad (12a)$$

$$s.t. \quad x_{m,k}^e(t), x_{l,k}^u(t), x_{n,k}^u(t) \in \{0, 1\}, \forall m, l, n, k, \quad (12b)$$

$$0 \leq \sum_{n \in \mathcal{N}} x_{n,k}^u(t) \leq 1, 0 \leq \sum_{l \in \mathcal{L}} x_{l,k}^u(t) \leq 1, \forall k, \quad (12c)$$

$$0 \leq \sum_{k \in \mathcal{K}} x_{m,k}^e(t) \leq 1, 0 \leq \sum_{k \in \mathcal{K}} x_{l,k}^u(t) \leq 1, \forall m, \forall l, \quad (12d)$$

$$0 \leq \sum_{k \in \mathcal{K}} x_{n,k}^u(t) \leq 1, \forall n, \quad (12d)$$

$$0 \leq \sum_{m \in \mathcal{M}} x_{m,k}^e(t) + \sum_{l \in \mathcal{L}} x_{l,k}^u(t) \leq 1, \forall k, \quad (12e)$$

$$0 \leq \sum_{m \in \mathcal{M}} x_{m,k}^e(t) + \sum_{n \in \mathcal{N}} x_{n,k}^u(t) \leq 1, \forall k, \quad (12e)$$

$$\begin{aligned}
0 \leq P_{m,k}^{c,e}(t) &\leq P_{\max}^{c,e}, 0 \leq P_{l,k}^{c,u}(t) \leq P_{\max}^{c,u}, \forall m, l, k, \\
0 \leq P_{n,k}^{d,u}(t) &\leq P_{\max}^{d,u}, \forall n, k, \quad (12f) \\
R_{m,k}^{c,e}(t) &\geq R^{e,req}, \forall m, k, \quad (12g) \\
\lim_{t \rightarrow \infty} \Pr(Q_l^{c,u}(t) \geq Q^{c,req}) &\leq P_{\max}^{c,laten}, \forall l, \quad (12h) \\
\lim_{t \rightarrow \infty} \Pr(R_n^{d,u}(t) \geq Q^{u,req}) &\leq P_{\max}^{u,laten}, \forall n, \quad (12i) \\
\Pr(R_{n,k}^{d,u}(t) \leq R^{u,req}) &\leq P_{\max}^{outage}, \forall n, k, \quad (12j) \\
\bar{Q}_m^{c,e}, \bar{Q}_l^{c,u}, \bar{Q}_n^{d,u} &< \infty, \forall m, l, n. \quad (12k)
\end{aligned}$$

Problem (12) is rather challenging to solve, mainly due to the three reasons below. Firstly, the optimization variables \mathbf{X} for allocating RBs to different slices are binary and thus (12b)-(12e) involve integer constraints. Secondly, even with fixed RB allocation variables \mathbf{X} , probability constraints (12h)-(12j) are non-convex constraints with respect to the transmit power variables \mathbf{P} . Finally, the constraint (12k) guarantees the stability of queues. Due to random arrival tasks, the optimal RB allocation and power control decisions are time dependent. Therefore, problem (12) is a mixed integer nonlinear stochastic optimization problem, which is difficult to be optimally solved in general.

III. PROBLEM TRANSFORMATION

A. Delay Constraints Transformation

In order to make problem (12) more tractable, we first deal with two non-convex constraints (12h) and (12i) on latency, so that the two probabilistic constraints become deterministic constraints. According to Markov's inequality [20], for a non-negative random variable X , we have

$$\Pr\{X \geq a\} \leq \frac{E\{X\}}{a}. \quad (13)$$

Since $Q_l^{c,u}(t)$ in constraint (12h) is always non-negative, constraints (12h) can be relaxed as

$$\lim_{t \rightarrow \infty} \Pr(Q_l^{c,u}(t) \geq Q^{c,req}) \leq \frac{E[Q_l^{c,u}(t)]}{Q^{c,req}}, \quad (14)$$

which can be further written as $E[Q_l^{c,u}(t)] \leq P_{\max}^{c,laten} Q^{c,req}$.

Considering that $E[Q_l^{c,u}(t)] = t\lambda_l^{c,u} - \sum_{t=0}^{T-1} R_l^{c,u}(t)$, we have

$$t\lambda_l^{c,u} - \sum_{t=0}^{T-1} R_l^{c,u}(t) \leq P_{\max}^{c,laten} Q^{c,req}. \quad (15)$$

From (15), the lower bound of the instantaneous transmission rate of the l th C^u -VUE at time slot t can be expressed by

$$R_l^{c,u}(t) \geq t\lambda_l^{c,u} - \sum_{t=0}^{T-2} R_l^{c,u}(t) - P_{\max}^{c,laten} Q^{c,req}. \quad (16)$$

It can be seen that the probabilistic constraint on latency in (12h) is transformed into a deterministic tractable constraint in (15), which can further help us analyze the impact of latency constraints. In order to strictly guarantee the low latency characteristics of URLLC services, the queue backlogs arriving at the server in the current time slot should not be greater than the processing rate of the server. Otherwise it

will cause queuing and result in system latency. The queue backlog evolution is always a non-increasing expression, that is, the queue backlog at the next moment will always be less than or equal to the queue backlog at the current moment. In order to accurately represent this characteristic, we introduce corresponding auxiliary variables and virtual queues in the following to strictly guarantee the latency constraints.

We introduce auxiliary variable vector $\varphi = \{\varphi_l^{c,u}(t) | \forall l \in \mathcal{L}\}$, which satisfies the following constraint:

$$\bar{\varphi}_l^{c,u} = \lim_{T \rightarrow \infty} \sup \frac{1}{T} \sum_{t=0}^{T-1} E[\varphi_l^{c,u}(t)] \leq \bar{R}_l^{c,u}, \quad (17)$$

where $\bar{R}_l^{c,u} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E[R_l^{c,u}(t)]$, and we have

$$\varphi_l^{c,u}(t) \geq \max \left\{ (t+1)\lambda_l^{c,u} - \sum_{\tau=0}^{t-1} \varphi_l^{c,u}(\tau) - P_{\max}^{c,laten} Q^{c,req}, 0 \right\}. \quad (18)$$

Similarly, with regard to constraint (12i), we can use the same operation to convert it into a deterministic constraint. Introduce auxiliary variable vector $\psi = \{\psi_n^{d,u}(t) | \forall n \in \mathcal{N}\}$, which satisfies the following constraint:

$$\bar{\psi}_n^{d,u} = \lim_{T \rightarrow \infty} \sup \frac{1}{T} \sum_{t=0}^{T-1} E[\psi_n^{d,u}(t)] \leq \bar{R}_n^{d,u}, \quad (19)$$

where $\bar{R}_n^{d,u} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E[R_n^{d,u}(t)]$ and we have

$$\psi_n^{d,u}(t) \geq \max \left\{ (t+1)\lambda_n^{d,u} - \sum_{\tau=0}^{t-1} \psi_n^{d,u}(\tau) - P_{\max}^{u,laten} Q^{u,req}, 0 \right\}. \quad (20)$$

In order to strictly guarantee the low-latency of URLLC slicing, we impose constraints $\bar{\varphi}_l^{c,u} \leq \bar{R}_l^{c,u}$ and $\bar{\psi}_n^{d,u} \leq \bar{R}_n^{d,u}$. Next, the concept of virtual queues [19] was introduced to transform corresponding constraints into queue stability problems. $H_l^{c,u}(t)$ and $Z_n^{d,u}(t)$ are defined as virtual queues with evolving equations as follows:

$$H_l^{c,u}(t+1) = \max \{H_l^{c,u}(t) + \varphi_l^{c,u}(t) - R_l^{c,u}(t), 0\}, \quad (21a)$$

$$Z_n^{d,u}(t+1) = \max \{Z_n^{d,u}(t) + \psi_n^{d,u}(t) - R_n^{d,u}(t), 0\}. \quad (21b)$$

For simplicity, let $R_{total}(t) = \sum_{m \in \mathcal{M}} R_m^{c,e}(t) + \sum_{l \in \mathcal{L}} \varphi_l^{c,u}(t) + \sum_{n \in \mathcal{N}} \psi_n^{d,u}(t)$ be the equivalent total system capacity. Then, the original question (12) is transformed into the following problem (22):

$$\max_{\{\mathbf{X}, \mathbf{P}, \varphi, \psi\}} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E[R_{total}(t)] \quad (22a)$$

$$s.t. \quad (12b) - (12g), (12j), (18), (20), \quad (22b)$$

$$\bar{Q}_m^{c,e} < \infty, \bar{\varphi}_l^{c,u} \leq \bar{R}_l^{c,u}, \bar{\psi}_n^{d,u} \leq \bar{R}_n^{d,u}, \forall m, l, n. \quad (22c)$$

Although the probability constraints (12h) and (12i) are transformed into deterministic linear constraints (18) and (20), integer constraints and queue stability constraints included in this problem still make the problem difficult to solve.

B. Lyapunov Optimization

In this subsection, an online algorithm based on Lyapunov optimization theory is proposed to solve the above problems. We can effectively solve the formulated stochastic optimization problem by solving the deterministic problem of each time slot based on the current status information, and do not need any information about the future moments of task arrival, network queue status, etc.. The key of Lyapunov optimization is to use the Lyapunov function to optimally control the dynamic system while ensuring the stability of the system.

Define queue vectors $\mathbf{Q}(t) = \{Q_1^{c,e}(t), \dots, Q_M^{c,e}(t)\}$, $\mathbf{H}(t) = \{H_1^{c,u}(t), \dots, H_L^{c,u}(t)\}$ and $\mathbf{Z}(t) = \{Z_1^{d,u}(t), \dots, Z_N^{d,u}(t)\}$. Let vector $\Theta(t) = \{\mathbf{Q}(t), \mathbf{H}(t), \mathbf{Z}(t)\}$ be backlogs of task queues containing all types of applications. The Lyapunov function is defined as

$$F(\Theta(t)) = \frac{1}{2} \sum_{m \in \mathcal{M}} Q_m^{c,e}(t)^2 + \frac{1}{2} \sum_{l \in \mathcal{L}} H_l^{c,u}(t)^2 + \frac{1}{2} \sum_{n \in \mathcal{N}} Z_n^{d,u}(t)^2. \quad (23)$$

The Lyapunov drift function is given by

$$\Delta(\Theta(t)) \triangleq E[F(\Theta(t+1)) - F(\Theta(t)) | \Theta(t)], \quad (24)$$

where $\Delta(\Theta(t))$ is the expected change of the Lyapunov function in a time slot, which is, the difference between the arrival and the transmission of each VUE data queue according to the current network status.

In addition to stabilizing the task data queue for each VUE, we should also maximize network capacity at time slot t . Therefore, in conjunction with problem (22), we introduce a control parameter V and define drift-plus-penalty $G(\Theta(t))$ at time slot t , which is given by

$$G(\Theta(t)) \triangleq \Delta(\Theta(t)) - VE[R_{total}(t) | \Theta(t)], \quad (25)$$

where V is a non-negative control parameter representing the trade-off between network stability and system capacity. A large V indicates that more network stability may be sacrificed. More specifically, a larger data queue length is required to achieve higher system capacity. When V is sufficiently large, a progressively optimal system capacity can be obtained by minimizing $G(\Theta(t))$.

Nevertheless, it can be seen that $G(\Theta(t))$ is a quadratic function of the length of the data queue, which is difficult to minimize directly. Thus we theoretically deduce the upper bound of $G(\Theta(t))$ in the following theorem. According to the analysis in [19], if we can effectively minimize the upper bound of $G(\Theta(t))$ instead of directly minimizing itself, we can still achieve near-optimal system capacity.

Theorem 1: Given the network status at time slot t , i.e. $\Theta(t)$, the upper-bound of $G(\Theta(t))$ can be given by

$$G(\Theta(t)) \leq C + E[\Phi(t) | \Theta(t)], \quad (26)$$

where C is a non-negative constant as follows

$$C = \frac{1}{2} \sum_{m \in \mathcal{M}} \left\{ (R_m^{\max})^2 + (A_m^{\max})^2 \right\} + \frac{1}{2} \sum_{l \in \mathcal{L}} \left\{ (R_l^{\max})^2 + (\varphi_l^{\max})^2 \right\} + \frac{1}{2} \sum_{n \in \mathcal{N}} \left\{ (R_n^{\max})^2 + (\psi_n^{\max})^2 \right\}, \quad (27)$$

and $\Phi(t)$ is defined as

$$\begin{aligned} \Phi(t) \triangleq & - \sum_{m \in \mathcal{M}} (Q_m^{c,e}(t) + V) R_m^{c,e}(t) \\ & + \sum_{l \in \mathcal{L}} (H_l^{c,u}(t) - V) \varphi_l^{c,u}(t) + \sum_{n \in \mathcal{N}} (Z_n^{d,u}(t) - V) \psi_n^{d,u}(t) \\ & - \sum_{l \in \mathcal{L}} H_l^{c,u}(t) R_l^{c,u}(t) - \sum_{n \in \mathcal{N}} Z_n^{d,u}(t) R_n^{d,u}(t). \end{aligned} \quad (28)$$

Proof: Please refer to Appendix A. \blacksquare

In *Theorem 1*, we derive the upper bound of $G(\Theta(t))$ for a given network state $\Theta(t)$. Therefore, the problem of optimizing $G(\Theta(t))$ is converted to optimize its upper bound $\Phi(t)$. In addition, the constraints of network stability in (22c) is relaxed by leveraging the Lyapunov optimization theory, and we focus on optimizing $\Phi(t)$ in each time slot t regardless of variables that change randomly over different time slots. Based on this, the problem is further formulated as

$$\min_{\{\mathbf{X}, \mathbf{P}, \varphi, \psi\}} \Phi(t) \quad (29a)$$

$$s.t. \quad (12b) - (12g), (12j), (18), (20). \quad (29b)$$

IV. PROBLEM SOLUTION

In order to make the problem (29) tractable, we first relax binary variables in (12b) into continuous variables, which leads to the following problem

$$\min_{\{\mathbf{X}, \mathbf{P}, \varphi, \psi\}} \Phi(t) \quad (30a)$$

$$s.t. \quad x_{m,k}^e(t), x_{l,k}^u(t), x_{n,k}^u(t) \in [0, 1], \forall m, l, n, k, \quad (30b)$$

$$(12c) - (12g), (12j), (18), (20). \quad (30c)$$

Such a relaxation usually demonstrates that the objective value of problem (30) is the lower bound of the objective value of problem (29). Although relaxed, the problem is still a non-convex optimization problem due to the non-convexity of the objective function and constraint (12j), which is rather challenging to solve effectively. Hereinafter, the slack problem (30) is decoupled into three sub-problems and an effective alternating iterative algorithm is proposed.

A. Equivalent Transmission Rate in URLLC Slicing

For any given RB allocation and transmit power $\{\mathbf{X}, \mathbf{P}\}$, the equivalent transmission rate of URLLC slicing $\{\varphi, \psi\}$ in problem (30) can be optimized by solving the following problem:

$$\min_{\{\varphi, \psi\}} \sum_{l \in \mathcal{L}} (H_l^{c,u}(t) - V) \varphi_l^{c,u}(t) + \sum_{n \in \mathcal{N}} (Z_n^{d,u}(t) - V) \psi_n^{d,u}(t) \quad (31a)$$

$$s.t. \quad (18), (20). \quad (31b)$$

It can be understood that problem (31) is a standard linear programming (LP) problem, which can be effectively solved by CVX.

$$1 - \exp \left\{ - \frac{(2^{R^{u,\text{req}}} - 1) \sigma^2}{P_{n,k}^{d,u}(t) g_n(t)} \right\} \prod_{l \neq n} \frac{1}{1 + \frac{(2^{R^{u,\text{req}}} - 1) x_{l,k}^u(t) P_{l,k}^{c,u}(t) g_{l,n}(t)}{P_{n,k}^{d,u}(t) g_n(t)}}} \leq P_{\max}^{\text{outage}}. \quad (33)$$

$$\begin{aligned} & 1 - \left\{ \exp \left(\frac{(2^{R^{u,\text{req}}} - 1) \sigma^2}{P_{n,k}^{d,u}(t) g_n(t)} \right) \prod_{l \neq n} \left(1 + \frac{(2^{R^{u,\text{req}}} - 1) x_{l,k}^u(t) P_{l,k}^{c,u}(t) g_{l,n}(t)}{P_{n,k}^{d,u}(t) g_n(t)} \right) \right\}^{-1} \\ & \leq 1 - \exp \left\{ - \frac{(2^{R^{u,\text{req}}} - 1) \sigma^2}{P_{n,k}^{d,u}(t) g_n(t)} - \sum_{l \in \mathcal{L}} \frac{(2^{R^{u,\text{req}}} - 1) x_{l,k}^u(t) P_{l,k}^{c,u}(t) g_{l,n}(t)}{P_{n,k}^{d,u}(t) g_n(t)} \right\} \leq P_{\max}^{\text{outage}}, \forall l, n, k. \end{aligned} \quad (34)$$

B. Vehicle Transmit Power Control

For any given equivalent transmission rate and RB allocation $\{\varphi, \psi, \mathbf{X}\}$, the transmission power $\{\mathbf{P}\}$ can be optimized by solving the following problem

$$\begin{aligned} \min_{\{\mathbf{P}\}} & - \sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} (Q_m^{c,e}(t) + V) x_{m,k}^e(t) R_{m,k}^{c,e}(t) \\ & - \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} H_l^{c,u}(t) x_{l,k}^u(t) R_{l,k}^{c,u}(t) \\ & - \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} Z_n^{d,u}(t) x_{n,k}^u(t) R_{n,k}^{d,u}(t) \quad (32a) \\ \text{s.t.} & \quad (12f), (12g), (12j). \quad (32b) \end{aligned}$$

It can be seen that due to the non-convexity of the objective function and the non-convex constraint (12j), the problem (32) is a difficult non-convex problem. First, the non-convex probabilistic constraint (12j) is converted into a deterministic constraint according to the lemma in [21], which can be written as (33) at the top of this page. However, the constraint (33) is still non-convex, which is difficult to solve. To avoid operating on such complex inequality constraints, we further use the conclusions in [22] to bound the derived constraint of outage probability, which is given by (34) at the top of this page. Then we have a tight upper bound of the outage probability in (34), which can be further written as

$$\frac{P_{n,k}^{d,u}(t) g_n(t)}{\sigma^2 + \sum_{l \in \mathcal{L}} x_{l,k}^u(t) P_{l,k}^{c,u}(t) g_{l,n}(t)} \geq \frac{(2^{R^{u,\text{req}}} - 1)}{\ln \left(\frac{1}{1 - P_{\max}^{\text{outage}}} \right)}. \quad (35)$$

It can be observed that constraint (35) is linear.

Through analyzing the objective function of problem (32), it can be told that the convexity of $R_{l,k}^{c,u}(t)$ and $R_{n,k}^{d,u}(t)$ cannot be directly judged, but they can be written as the differences of two concave functions, respectively. To concise the process representation, slack variables $\mathbf{S} = \left\{ S_{k,l}^{c,u}(t) = \sum_{n \in \mathcal{N}} x_{n,k}^u(t) P_{n,k}^{d,u}(t) g_{l,n,B}(t) + \sigma^2, \forall k, l \right\}$ and $\mathbf{W} = \left\{ W_{k,n}^{d,u}(t) = \sum_{l \in \mathcal{L}} x_{l,k}^u(t) P_{l,k}^{c,u}(t) g_{l,n}(t) + \sigma^2, \forall k, n \right\}$ are introduced, which follows that

$$R_{l,k}^{c,u}(t) = \log_2 \left(S_{l,k}^{c,u}(t) + P_{l,k}^{c,u}(t) g_{l,B}(t) \right) - \hat{R}_{l,k}^{c,u}(t), \quad (36a)$$

$$R_{n,k}^{d,u}(t) = \log_2 \left(W_{n,k}^{d,u}(t) + P_{n,k}^{d,u}(t) g_n(t) \right) - \hat{R}_{n,k}^{d,u}(t). \quad (36b)$$

Since any concave function can find its global upper bound at any point through the first-order Taylor expansion [23], with given local points $P_{l,k}^{c,u}(t)^r$, $P_{n,k}^{d,u}(t)^r$, $S_{l,k}^{c,u}(t)^r$ and $W_{n,k}^{d,u}(t)^r$ in the r th iteration, we obtain the following upper bounds for $R_{l,k}^{c,u}(t)$ and $R_{n,k}^{d,u}(t)$, respectively, i.e.,

$$\begin{aligned} \hat{R}_{l,k}^{c,u}(t) & \leq \log_2 \left(S_{l,k}^{c,u}(t)^r \right) + \frac{S_{l,k}^{c,u}(t) - S_{l,k}^{c,u}(t)^r}{S_{l,k}^{c,u}(t)^r \ln 2} \\ & \triangleq \hat{R}_{l,k}^{c,u,ub}(t), \end{aligned} \quad (37a)$$

$$\begin{aligned} \hat{R}_{n,k}^{d,u}(t) & \leq \log_2 \left(W_{n,k}^{d,u}(t)^r \right) + \frac{W_{n,k}^{d,u}(t) - W_{n,k}^{d,u}(t)^r}{W_{n,k}^{d,u}(t)^r \ln 2} \\ & \triangleq \hat{R}_{n,k}^{d,u,ub}(t). \end{aligned} \quad (37b)$$

With given local points \mathbf{P}^r , \mathbf{S}^r , \mathbf{W}^r and the upper bounds in (37a) and (37b), problem (32) is approximated as follows

$$\begin{aligned} \max_{\{\mathbf{P}, \mathbf{S}, \mathbf{W}\}} & \sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} (Q_m^{c,e}(t) + V) x_{m,k}^e(t) R_{m,k}^{c,e}(t) \\ & + \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} H_l^{c,u}(t) x_{l,k}^u(t) \\ & \cdot \left[\log_2 \left(S_{l,k}^{c,u}(t) + P_{l,k}^{c,u}(t) g_{l,B}(t) \right) - \hat{R}_{l,k}^{c,u,ub}(t) \right] \\ & + \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} Z_n^{d,u}(t) x_{n,k}^u(t) \\ & \cdot \left[\log_2 \left(W_{n,k}^{d,u}(t) + P_{n,k}^{d,u}(t) g_n(t) \right) - \hat{R}_{n,k}^{d,u,ub}(t) \right] \end{aligned} \quad (38a)$$

$$\text{s.t.} \quad \sum_{n \in \mathcal{N}} x_{n,k}^u(t) P_{n,k}^{d,u}(t) g_{l,n,B}(t) + \sigma^2 \leq S_{l,k}^{c,u}(t), \forall l, k, \quad (38b)$$

$$\sum_{l \in \mathcal{L}} x_{l,k}^u(t) P_{l,k}^{c,u}(t) g_{l,n}(t) + \sigma^2 \leq W_{n,k}^{d,u}(t), \forall n, k, \quad (38c)$$

$$(12f), (12g), (35). \quad (38d)$$

Problem (38) is a convex optimization problem. The Lagrange dual method is used to solve it, and closed-form solutions for each variable can be obtained according to Karush-Kuhn-Tucker (KKT) conditions. Considering the limitation of the article length, we omit the detailed steps and the similar process can be found in [24]. Since problem (38) is a convex optimization problem, the dual gap is zero. It is worth noting that the upper bound adopted in problem (38) shows that any feasible solution in problem (38) is also feasible for problem (32), but the converse is usually not true. As a result, the optimal objective value obtained from the approximate problem (38) is usually used as the upper bound of that of

problem (32).

C. RB Resource Allocation

For any given equivalent transmission rate of URLLC slicing and transmit power of vehicles $\{\varphi, \psi, \mathbf{P}\}$, the RB allocation strategy $\{\mathbf{X}\}$ in problem (30) can be optimized by solving the following subproblem:

$$\begin{aligned} \min_{\{\mathbf{X}\}} \quad & - \sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} (Q_m^{c,e}(t) + V) x_{m,k}^e(t) R_{m,k}^{c,e}(t) \\ & - \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} H_l^{c,u}(t) x_{l,k}^u(t) R_{l,k}^{c,u}(t) \\ & - \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} Z_n^{d,u}(t) x_{n,k}^u(t) R_{n,k}^{d,u}(t) \quad (39a) \\ \text{s.t.} \quad & (12c) - (12e), (30b). \quad (39b) \end{aligned}$$

It can be seen that (12c)-(12e) and (30b) are linear constraints. However, due to the non-convexity of the objective function, problem (39) is still a non-convex problem. Next, we use appropriate means to convert the problem. Auxiliary slack variables $\alpha = \{\alpha_{l,k}^{c,u}(t) = R_{l,k}^{c,u}(t), \forall l, k\}$ and $\beta = \{\beta_{n,k}^{d,u}(t) = R_{n,k}^{d,u}(t), \forall n, k\}$ are introduced, which leads to the non-convexity of the objective function due to bilinear parts $x_{l,k}^u(t) \alpha_{l,k}^{c,u}(t)$ and $x_{n,k}^u(t) \beta_{n,k}^{d,u}(t)$. Meanwhile, constraints $R_{l,k}^{c,u}(t) \geq \alpha_{l,k}^{c,u}(t)$ and $R_{n,k}^{d,u}(t) \geq \beta_{n,k}^{d,u}(t)$ should be satisfied. To deal with these non-convex factors, we define $\Xi_{l,k}^{c,u}(t) = -x_{l,k}^u(t) \alpha_{l,k}^{c,u}(t)$ and $\Gamma_{n,k}^{d,u}(t) = -x_{n,k}^u(t) \beta_{n,k}^{d,u}(t)$, respectively, which can be further written as

$$\Xi_{l,k}^{c,u}(t) = \frac{1}{2} x_{l,k}^u(t)^2 + \frac{1}{2} \alpha_{l,k}^{c,u}(t)^2 - \frac{1}{2} (x_{l,k}^u(t) + \alpha_{l,k}^{c,u}(t))^2, \quad (40a)$$

$$\Gamma_{n,k}^{d,u}(t) = \frac{1}{2} x_{n,k}^u(t)^2 + \frac{1}{2} \beta_{n,k}^{d,u}(t)^2 - \frac{1}{2} (x_{n,k}^u(t) + \beta_{n,k}^{d,u}(t))^2. \quad (40b)$$

Successive convex approximation (SCA) technique is used to approximate $\Xi_{l,k}^{c,u}(t)$ and $\Gamma_{n,k}^{d,u}(t)$ in each iteration. Recall that any concave function can find its global upper bound at any point through the first-order Taylor expansion. Therefore, with given local points $x_{l,k}^u(t)^r, x_{n,k}^u(t)^r, \alpha_{l,k}^{c,u}(t)^r$ and $\beta_{n,k}^{d,u}(t)^r$ in the r th iteration, the following concave upper bounds are given by

$$\begin{aligned} \Xi_{l,k}^{c,u}(t) &\leq \frac{1}{2} x_{l,k}^u(t)^2 + \frac{1}{2} \alpha_{l,k}^{c,u}(t)^2 - \frac{1}{2} (x_{l,k}^u(t)^r + \alpha_{l,k}^{c,u}(t)^r)^2 \\ &\quad - (x_{l,k}^u(t)^r + \alpha_{l,k}^{c,u}(t)^r) (x_{l,k}^u(t) - x_{l,k}^u(t)^r) \\ &\quad - (x_{l,k}^u(t)^r + \alpha_{l,k}^{c,u}(t)^r) (\alpha_{l,k}^{c,u}(t) - \alpha_{l,k}^{c,u}(t)^r) \\ &\triangleq \Xi_{l,k}^{c,u,ub}(t), \quad (41a) \end{aligned}$$

$$\begin{aligned} \Gamma_{n,k}^{d,u}(t) &\leq \frac{1}{2} x_{n,k}^u(t)^2 + \frac{1}{2} \beta_{n,k}^{d,u}(t)^2 - \frac{1}{2} (x_{n,k}^u(t)^r + \beta_{n,k}^{d,u}(t)^r)^2 \\ &\quad - (x_{n,k}^u(t)^r + \beta_{n,k}^{d,u}(t)^r) (x_{n,k}^u(t) - x_{n,k}^u(t)^r) \\ &\quad - (x_{n,k}^u(t)^r + \beta_{n,k}^{d,u}(t)^r) (\beta_{n,k}^{d,u}(t) - \beta_{n,k}^{d,u}(t)^r) \\ &\triangleq \Gamma_{n,k}^{d,u,ub}(t). \quad (41b) \end{aligned}$$

With respect to non-convex constraints $R_{l,k}^{c,u}(t) \geq \alpha_{l,k}^{c,u}(t)$ and $R_{n,k}^{d,u}(t) \geq \beta_{n,k}^{d,u}(t)$, we can obtain the lower bound of

$R_{l,k}^{c,u}(t)$ and $R_{n,k}^{d,u}(t)$ through SCA technique respectively, which can be given by

$$\begin{aligned} R_{l,k}^{c,u}(t) &\geq \log_2 \left(1 + \frac{P_{l,k}^{c,u}(t) g_{l,B}(t)}{\sum_{n \in \mathcal{N}} x_{n,k}^u(t)^r P_{n,k}^{d,u}(t) g_{l,n,B}(t) + \sigma^2} \right) \\ &\quad - \Delta_{l,k}^{c,u}(t) \sum_{n \in \mathcal{N}} P_{n,k}^{d,u}(t) g_{l,n,B}(t) (x_{n,k}^u(t) - x_{n,k}^u(t)^r) \\ &\triangleq R_{l,k}^{c,u,lb}(t), \quad (42a) \end{aligned}$$

$$\begin{aligned} R_{n,k}^{d,u}(t) &\geq \log_2 \left(1 + \frac{P_{n,k}^{d,u}(t) g_n(t)}{\sum_{l \in \mathcal{L}} x_{l,k}^u(t)^r P_{l,k}^{c,u}(t) g_{l,n}(t) + \sigma^2} \right) \\ &\quad - \Delta_{n,k}^{d,u}(t) \sum_{l \in \mathcal{L}} P_{l,k}^{c,u}(t) g_{l,n}(t) (x_{l,k}^u(t) - x_{l,k}^u(t)^r) \\ &\triangleq R_{n,k}^{d,u,lb}(t), \quad (42b) \end{aligned}$$

where $\Delta_{l,k}^{c,u}(t)$ and $\Delta_{n,k}^{d,u}(t)$ are the coefficients related to the derivative that is not dominant and omitted here. Then problem (39) is approximated as the following problem

$$\begin{aligned} \min_{\{\mathbf{X}, \alpha, \beta\}} \quad & - \sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}} (Q_m^{c,e}(t) + V) x_{m,k}^e(t) R_{m,k}^{c,e}(t) \\ & + \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} H_l^{c,u}(t) \Xi_{l,k}^{c,u,ub}(t) \\ & + \sum_{n \in \mathcal{N}} \sum_{k \in \mathcal{K}} Z_n^{d,u}(t) \Gamma_{n,k}^{d,u,ub}(t) \quad (43a) \end{aligned}$$

$$\text{s.t.} \quad \alpha_{l,k}^{c,u}(t) - R_{l,k}^{c,u,lb}(t) \leq 0, \forall l, k, \quad (43b)$$

$$\beta_{n,k}^{d,u}(t) - R_{n,k}^{d,u,lb}(t) \leq 0, \forall n, k, \quad (43c)$$

$$(12c) - (12e), (40b). \quad (43d)$$

Problem (43) is a convex optimization problem, and the optimal solution can be effectively obtained by the convex optimization solving tools such as CVX [23]. It should also be noted that problem (43) is always a subset of problem (39). Therefore, the optimal objective value of problem (43) can be used as the upper bound of that of problem (39).

Remark 1: Variables \mathbf{X} obtained by the problem (43) are continuous with value between 0 and 1, so the binary variables of RB allocation need to be reconstructed, which indicate that the allocation of RBs should minimize the objective function of problem (40) and can be given by $m^* = \arg \min \frac{\partial \Phi(t)}{\partial x_{m,k}^e(t)}$, $l^* = \arg \min_l \frac{\partial \Phi(t)}{\partial x_{l,k}^u(t)}$ and $n^* = \arg \min_n \frac{\partial \Phi(t)}{\partial x_{n,k}^u(t)}$, where $x_{m^*,k}^e(t) = 1$, $x_{l^*,k}^u(t) = 1$ and $x_{n^*,k}^u(t) = 1$ indicate the suboptimal RB allocation variables, respectively.

V. PROPOSED ALGORITHM

Due to the random and unpredictable behaviors of vehicular applications, it is great of importance to make resource allocation decision without foreseeing the future network status. Motivated by this, in this section, the proposed slicing scheduling algorithm, namely JRPSV, is tailored to provide an online and asymptotically optimal resource allocation decision without requiring prior statistical knowledge of any random variables. More specifically, JRPSV employs jointly RB allocation and power control to effectively maximize network system capacity, as summarized in Algorithm 1. JRPSV can

Algorithm 1 Joint Optimization of RB Allocation and Power Control Slicing Scheduling Algorithm in Vehicular Networks (JRPSV)

- 1: At the beginning of the t th time slot, observe the current queues $Q_m^{c,e}(t)$, $H_l^{c,u}(t)$ and $Z_n^{d,u}(t)$ respectively, as well as channel states.
 - 2: Determine \mathbf{P} and \mathbf{X} at current time slot t according to Algorithm 2.
 - 3: Base on the above obtained optimization results, update queues $Q_m^{c,e}(t)$, $H_l^{c,u}(t)$ and $Z_n^{d,u}(t)$ according to (3), (21a) and (21b), respectively.
-

Algorithm 2 Alternating Iterative Algorithm for Problem (30)

- 1: At current time slot t :
 - 2: Initialize \mathbf{P}^0 and \mathbf{X}^0 . Let $r = 0$.
 - 3: **repeat**
 - 4: Solve problem (31) for given $\{\mathbf{P}^r, \mathbf{X}^r\}$, and denote the optimal solution as $\{\varphi^{r+1}, \psi^{r+1}\}$.
 - 5: Solve problem (38) for given $\{\varphi^{r+1}, \psi^{r+1}, \mathbf{P}^r, \mathbf{X}^r\}$, and denote the optimal solution as $\{\mathbf{P}^{r+1}\}$.
 - 6: Solve problem (43) for given $\{\varphi^{r+1}, \psi^{r+1}, \mathbf{P}^{r+1}, \mathbf{X}^r\}$, and denote the optimal solution as $\{\mathbf{X}^{r+1}\}$.
 - 7: $r = r + 1$.
 - 8: **until** The change of the objective value is below a threshold $\zeta > 0$. Return the optimal solution $\{\varphi^*, \psi^*, \mathbf{P}^*, \mathbf{X}^*\}$.
-

obtain the optimal determination of RB allocation \mathbf{X}^* and transmit power \mathbf{P}^* at each time slot according to Algorithm 2. In particular, with respect to Algorithm 2, all optimization variables in problem (30) are divided into three parts, i.e. $\{\varphi, \psi\}$, $\{\mathbf{P}\}$ and $\{\mathbf{X}\}$. When solving these variables, two parts of variables are fixed while solving the other two parts of variables, and the three sub-problems are iterated alternately. The details are summarized in Algorithm 2.

A. Performance Analysis on JRPSV

In this subsection, we make a theoretical analysis of the performance of the proposed algorithm JRPSV. *Theorem 2* and *Theorem 3* not only explain how the control parameter V affects system capacity and system queue length, but also reveal the trade-off between optimal system capacity and system latency.

Theorem 2: Let $R_{total}^{opt} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E[R_{total}^{opt}(t)]$ be the theoretical optimal system capacity of origin problem (22), and \bar{R}_{total}^* is the achievable system capacity obtained by Algorithm 1, where $\bar{R}_{total}^* = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E[R_{total}^*(t)]$. Then we have

$$\bar{R}_{total}^* \geq R_{total}^{opt} - \frac{C}{V}, \quad (44)$$

where C is a non-negative constant defined by (27).

Proof: Please refer to Appendix B. ■

Theorem 3: Suppose there exists $\delta > 0, \Omega(\delta)$ and a stationary and randomized scheme Π that satisfies (12b)-(12j), and the following Slater Conditions [19]:

$$E[R_{total}^{\Pi}(t)] = \Omega(\delta), \quad (45a)$$

$$E[A_m^{c,e}(t)] \leq E[R_m^{c,e,\Pi}(t)] - \delta, \forall m \in \mathcal{M}. \quad (45b)$$

Then we have

$$\bar{Q}_m^{c,e} \leq \frac{C + V [\Omega(\delta) + R_{total}^{opt}]}{\delta}. \quad (46)$$

Proof: Please refer to Appendix C. ■

Theorem 2 and *Theorem 3* manifest that the total capacity of the system has a controllable gap between the proposed algorithm JRPSV and the theoretical optimal value, i.e. $\mathcal{O}(1/V)$, and when V is larger, the total system capacity obtained by JRPSV is closer to the theoretical optimal value. Meanwhile, it can be observed that the upper bound of the time-averaged queue length increases with the increase of V . Therefore, there is a $[\mathcal{O}(1/V), \mathcal{O}(V)]$ trade-off between the optimal system capacity and latency. In other words, when we want to approach the optimal capacity of the system, actions such as RB allocation and power control determined by the JRPSV algorithm will result in a larger average system latency, which illustrates that the optimal performance is achieved at the expense of system latency. Therefore, we can adjust the control parameter V to achieve the trade-off between them.

B. Convergence of Algorithm 2

This section mainly discusses the convergence of Algorithm 2. Since Algorithm 2 is completed in a time slot, for simplicity, we omit the symbol representing time slot t after each variable. It is worth noting that for the power control sub-problem (32) and the RB allocation sub-problem (39), we only optimally solve their approximate problems (38) and (43). In order to show the convergence behavior of Algorithm 2, we have the following analysis. Define $\Phi(\varphi^r, \psi^r, \mathbf{P}^r, \mathbf{X}^r)$ the value of the objective function at the r th iteration, and define the objective function of problem (38) and (43) as $\Phi_{pow}^{ub,r}(\varphi, \psi, \mathbf{P}, \mathbf{X})$ and $\Phi_{alloc}^{ub,r}(\varphi, \psi, \mathbf{P}, \mathbf{X})$, respectively.

First, in Step 4 of Algorithm 2, since problem (31) is optimized with given $\{\mathbf{P}^r, \mathbf{X}^r\}$, the optimal solution $\{\varphi^{r+1}, \psi^{r+1}\}$ is obtained, and we have

$$\begin{aligned} \Phi(\varphi^r, \psi^r, \mathbf{P}^r, \mathbf{X}^r) &= \Phi(\varphi^r, \psi^r) + \Phi(\mathbf{P}^r, \mathbf{X}^r) \\ &\geq \Phi(\varphi^{r+1}, \psi^{r+1}) + \Phi(\mathbf{P}^r, \mathbf{X}^r). \end{aligned} \quad (47)$$

We define $\Phi(\varphi^{r+1}, \psi^{r+1}) + \Phi(\mathbf{P}^r, \mathbf{X}^r) \triangleq \Phi(\varphi^{r+1}, \psi^{r+1}, \mathbf{P}^r, \mathbf{X}^r)$. Next, the first-order Taylor expansions in (37a) and (37b) will provide a tight upper bound at given local points $\{\varphi^{r+1}, \psi^{r+1}\}$ and \mathbf{X}^r , which means that problem (38) at \mathbf{X}^r has the same objective value as that of problem (32), i.e.,

$$\begin{aligned} \Phi(\varphi^{r+1}, \psi^{r+1}, \mathbf{P}^r, \mathbf{X}^r) &= \Phi_{pow}^{ub,r}(\varphi^{r+1}, \psi^{r+1}, \mathbf{P}^r, \mathbf{X}^r) \\ &= \Phi_{pow}^{ub,r}(\mathbf{P}^r, \mathbf{X}^r) + \Phi_{pow}^{ub,r}(\varphi^{r+1}, \psi^{r+1}). \end{aligned} \quad (48)$$

In Step 5 of Algorithm 2, having obtained the optimal solution \mathbf{P}^{r+1} with given $\{\varphi^{r+1}, \psi^{r+1}\}$ and \mathbf{X}^r , it follows that

$$\begin{aligned} & \Phi_{pow}^{ub,r}(\mathbf{P}^r, \mathbf{X}^r) + \Phi_{pow}^{ub,r}(\varphi^{r+1}, \psi^{r+1}) \\ & \geq \Phi_{pow}^{ub,r}(\mathbf{P}^{r+1}, \mathbf{X}^r) + \Phi_{pow}^{ub,r}(\varphi^{r+1}, \psi^{r+1}). \end{aligned} \quad (49)$$

Since the objective value of problem (38) is the upper bound of that of problem (32) at \mathbf{P}^{r+1} , we have

$$\begin{aligned} & \Phi_{pow}^{ub,r}(\mathbf{P}^{r+1}, \mathbf{X}^r) + \Phi_{pow}^{ub,r}(\varphi^{r+1}, \psi^{r+1}) \\ & \geq \Phi(\varphi^{r+1}, \psi^{r+1}, \mathbf{P}^{r+1}, \mathbf{X}^r). \end{aligned} \quad (50)$$

As can be seen from the above inequality, the object value of problem (32) is non-increasing after each iteration. Similarly, in Step 6 of Algorithm 2, it follows that

$$\begin{aligned} & \Phi(\varphi^{r+1}, \psi^{r+1}, \mathbf{P}^{r+1}, \mathbf{X}^r) = \Phi_{alloc}^{ub,r}(\varphi^{r+1}, \psi^{r+1}, \mathbf{P}^{r+1}, \mathbf{X}^r) \\ & = \Phi_{alloc}^{ub,r}(\mathbf{P}^{r+1}, \mathbf{X}^r) + \Phi_{alloc}^{ub,r}(\varphi^{r+1}, \psi^{r+1}) \\ & \geq \Phi_{alloc}^{ub,r}(\mathbf{P}^{r+1}, \mathbf{X}^{r+1}) + \Phi_{alloc}^{ub,r}(\varphi^{r+1}, \psi^{r+1}) \\ & \geq \Phi(\varphi^{r+1}, \psi^{r+1}, \mathbf{P}^{r+1}, \mathbf{X}^{r+1}). \end{aligned} \quad (51)$$

Based on the above analysis, we obtain

$$\Phi(\varphi^r, \psi^r, \mathbf{P}^r, \mathbf{X}^r) \geq \Phi(\varphi^{r+1}, \psi^{r+1}, \mathbf{P}^{r+1}, \mathbf{X}^{r+1}), \quad (52)$$

which indicates that the objective function of problem (30) is non-increasing after each iteration in Algorithm 2. Due to the fact that the value of the objective function of problem (30) is lower bounded by a finite value, Algorithm 2 is guaranteed to converge.

C. Complexity Analysis

This subsection mainly discusses the computational complexity of the proposed JRPSV algorithm. The computational complexity of the JRPSV algorithm mainly stems from the determination of power control \mathbf{P} and RB allocation \mathbf{X} in Step 2, which is dominated by Algorithm 2. Thus, we mainly analyze the computational complexity of Algorithm 2.

In Step 4 of Algorithm 2, the complexity of solving the LP problem (31) by employing the interior point method is $\mathcal{O}((LN)^{3.5})$ [23]. In Step 5, Lagrange dual method is used to solve the problem (38). According to the similar process in [24], the computational complexity of the Lagrange dual method is $\mathcal{O}(i_{\max}MLNK)$, where i_{\max} is the maximum number of iterations to obtain the KKT points. In Step 6, solving convex problem (43) with the interior point method results in the complexity of $\mathcal{O}((MLNK)^{3.5})$. Denote the maximum number of iterations that allows Algorithm 2 to converge as r_{\max} . The computational complexity of Algorithm 2 can be estimated by $\mathcal{O}(r_{\max}((LN)^{3.5} + i_{\max}MLNK + (MLNK)^{3.5})) = \mathcal{O}(r_{\max}(MLNK)^{3.5})$.

VI. SIMULATION RESULTS

In this section, we will provide simulation results to verify the proposed JRPSV algorithm. We refer to the simulation setup of the highway case detailed in 3GPP TR 36.885 [25], and model a multi-lane highway. The highway passes through

TABLE I
SIMULATION PARAMETERS [25], [26].

Parameters	Value
Cell radius	500 m
Carrier frequency	2 GHz
Bandwidth	4 MHz
Number of lanes	6
Lane width	4 m
Vehicle speed	70 km/h
Vehicle drop model	spacial Poisson process
Average distance between vehicles	$2.5v$, v in m/s
Reliability for V2V (P_{\max}^{outage})	0.01
Latency for V2I ($P_{\max}^{\text{c,laten}}$), V2V ($P_{\max}^{\text{u,laten}}$)	0.01, 0.01
Number of C ^e -VUEs, C ^u -VUEs and D2D-V2Vs (M, L, N)	10, 10, 10
Number of RBs (K)	20
Noise power (σ^2)	-114 dBm
Maximum transmit power of C ^e -VUE, C ^u -VUE and D2D-V2V ($P_{\max}^{\text{c,e}}, P_{\max}^{\text{c,u}}, P_{\max}^{\text{d,u}}$)	23, 23, 23 dBm
Pathloss model of V2I links	$128 + 37.6\log_{10}d$, d in km
Shadow fading of V2I links	Log-normal, standard deviation 8 dB
Pathloss model of V2V links	LOS in WINNER + B1 [27]
Shadow fading of V2V links	Log-normal, standard deviation 3 dB

a single cell, with the base station at the center of the cell. The scattering point of the vehicle follows the spatial Poisson process, and the density of the vehicle depends on the speed. The main simulation parameters are listed in the Table I. It is noted that by default, all parameters are set to the values specified in the table, and the parameter settings in each figure take precedence in the corresponding case.

In the following subsections, we will first evaluate the stability of the proposed JRPSV and the convergence of Algorithm 2, and then demonstrate how the control parameter V affects the system capacity and queue length. Thereafter, we will study the impact of corresponding QoS requirements on system performance. Finally, compared with the set multiple baselines, we manifest the effectiveness of the proposed algorithm.

A. Stability and Convergence

In Fig. 2, we observe the total queue length in the eMBB slicing $\sum_{m \in \mathcal{M}} Q_m^{\text{c,e}}(t)$ and the total queue length in the URLLC slicing $\sum_{l \in \mathcal{L}} H_l^{\text{c,u}}(t) + \sum_{n \in \mathcal{N}} Z_n^{\text{d,u}}(t)$ over time as well as the impact of control parameter V . Let the initial queue length be $\Theta(0) = \{\mathbf{Q}(0), \mathbf{H}(0), \mathbf{Z}(0)\} = 0$. It is not difficult to see that the sum of the queue length increases at the beginning, and the total queue length of each slice will eventually fluctuate around a constant when it reaches as a stable state. In addition, it can be seen that for both approaches of slicing, an increase of V will increase the total queue length and fluctuate around a larger value, which validates the content of *Theorem 2*.

Next, we show the convergence behavior of Algorithm 2 in each slot in Fig. 3. The set convergence condition is $(\Phi^{r+1} - \Phi^r) / \Phi^{r+1} \leq \zeta$ ($\zeta = 1e - 05$). The initial variables are randomly assigned at the beginning, so that the objective

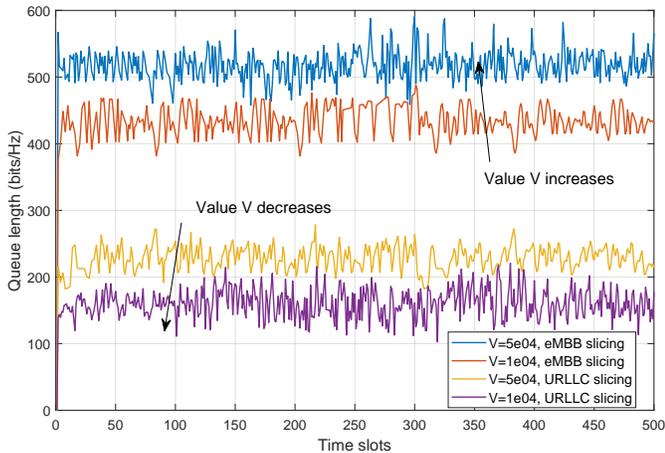


Fig. 2. Queue length of eMBB slicing and URLLC slicing versus time slots.

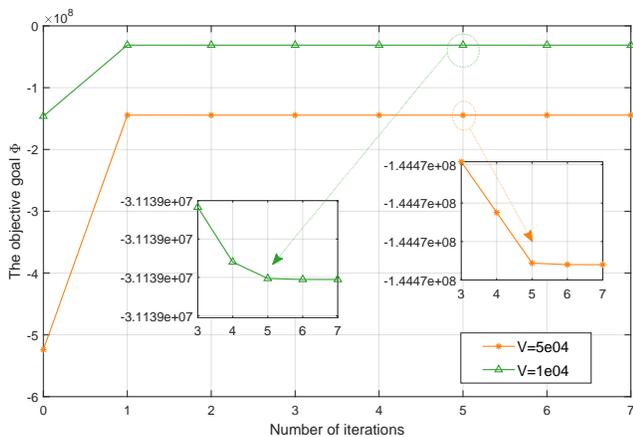


Fig. 3. Convergence behavior of Algorithm 2 parameterized by different V .

goal is a small value. Such a small value cannot represent any information of optimization, and it is just an initial value of the objective goal in the 0-th iteration. From the first iteration, we obtain the local optimal solution step by step and then approach to the optimal solution gradually. The objective goal has a non-increasing trend until the convergence condition is reached after about five iterations. In addition, the control parameter V also affects the objective goal. The larger the value of V , the larger the absolute value $|\Phi|$ of the objective goal.

B. Performance of JRPSV Versus Control Parameter V

Figure 4 depicts the CDF of the time-averaged system capacity and the time-averaged system queue length under different control parameters V , respectively. It can be seen that in (a), as V becomes larger, the mean of the time-averaged system capacity data samples also increases. In other words, a higher V will make the time average total capacity of the system larger, and when V is large enough, the optimal system capacity will be reached. The time-averaged system queue length $\sum_{m \in \mathcal{M}} \bar{Q}_m^{c,e} + \sum_{l \in \mathcal{L}} \bar{H}_l^{c,u} + \sum_{n \in \mathcal{N}} \bar{Z}_n^{d,u}$ in (b) also increases as V increases. A higher V results in a longer system queue and a higher average system latency as well. The above

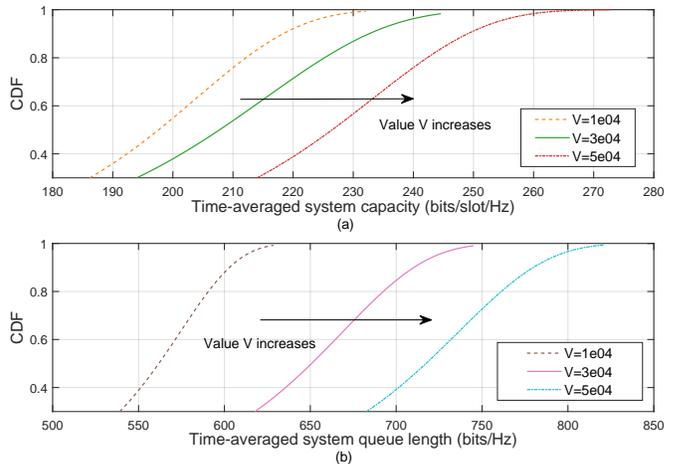


Fig. 4. Effect of control parameter V on system performance.

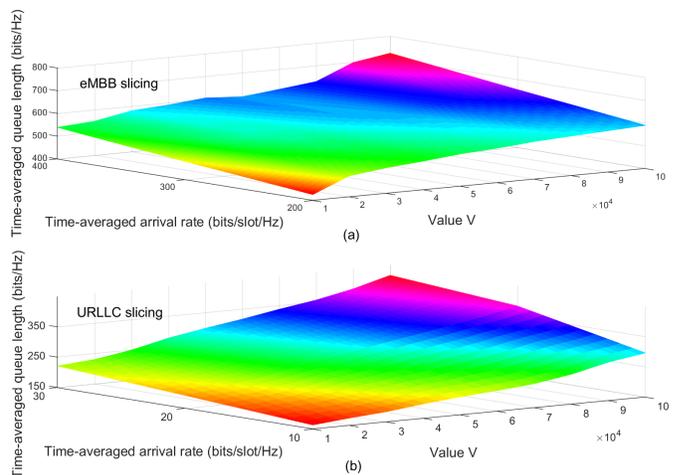


Fig. 5. Time-averaged queue length of slicing parameterized by different time-averaged arrival rate and V .

experimental results show that when V is large enough, the proposed algorithm will infinitely approach the optimal system capacity, but at the same time, it will also bring higher system latency, which indicates the trade-off relationship between optimal system capacity and system average latency, and further validates the *Theorem 2* and *Theorem 3*.

C. The Impact of QoS Requirements on Algorithm Performance

By changing $\bar{A}_m^{c,e}$, $\lambda_l^{c,u}$, $\lambda_n^{d,u}$ and the control parameter V , we show the relationship between the time-averaged queue length and the time-averaged arrival rate for each slicing and V in Fig. 5. In particular, we configure the time-averaged arrival rate of applications served by the eMBB slicing as $\bar{A}_m^{c,e} = [200, 250, 300, 350, 400]$ bits/slot/Hz, and the adopted time-averaged arrival rate of applications served by URLLC slicing are $(\lambda_l^{c,u} + \lambda_n^{d,u}) = [11, 15, 19, 23, 27]$ bits/slot/Hz. Similar to the results in Fig. 4, when the time-averaged arrival rate is constant, the time-averaged queue length of eMBB slicing ($\sum_{m \in \mathcal{M}} \bar{Q}_m^{c,e}$) and URLLC slicing ($\sum_{l \in \mathcal{L}} \bar{H}_l^{c,u} + \sum_{n \in \mathcal{N}} \bar{Z}_n^{d,u}$) in-

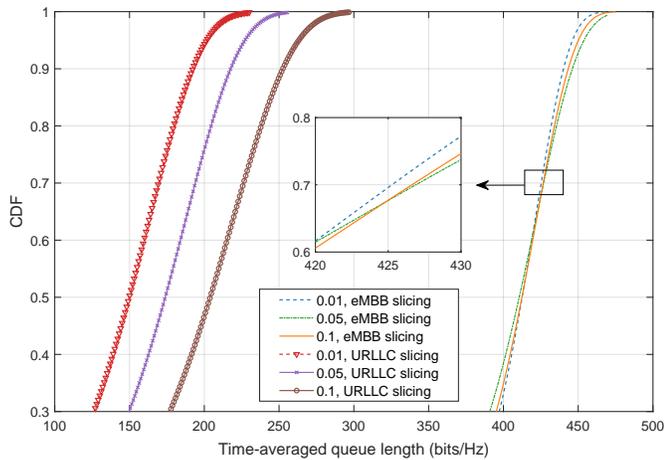


Fig. 6. CDF of time-averaged queue length with different $P_{\max}^{c,laten}$ and $P_{\max}^{u,laten}$.

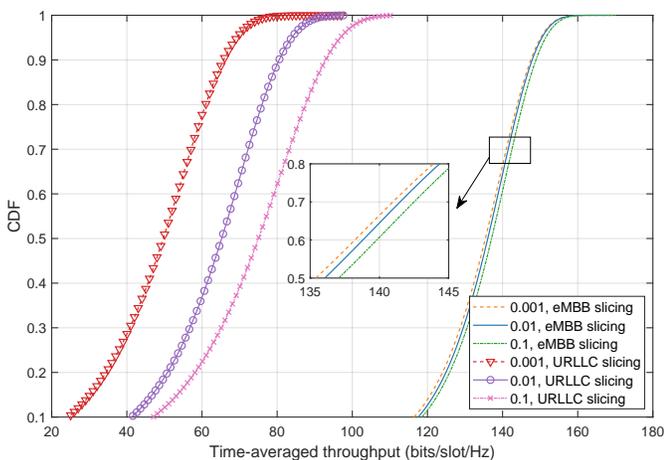


Fig. 7. CDF of time-averaged queue length with P_{\max}^{outage} .

crease almost linearly with V , which can be characterized as a $\mathcal{O}(V)$ relationship. When the control parameter V is constant, an increment of the time-averaged arrival rate also causes a growth of the time-averaged queue length. This is because when the processing rate of system is constant, a higher arrival rate will result in a larger queue backlog of system, which will cause higher system latency.

In order to demonstrate the effect of the probability $P_{\max}^{c,laten}$ and $P_{\max}^{u,laten}$ on the time-averaged queue, we compare the CDF of the time-averaged queue length of each slicing under different $P_{\max}^{c,laten}$ and $P_{\max}^{u,laten}$ in Fig. 6, where the value of $P_{\max}^{c,laten}$ is the same as $P_{\max}^{u,laten}$ in each experiment. It can be seen intuitively that as the value of probability rises, the average queue of URLLC slicing lengthens, but there is little change in eMBB slicing. This happens because the constraint in (12h) and (12i) affects virtual queues $H_l^{c,u}(t)$ and $Z_n^{d,u}(t)$. When the value of the probability tends to be larger, the restrictions on the low-latency are loosened. At the same time, the probability that the actual queues $Q_l^{c,u}(t)$ and $Q_n^{d,u}(t)$ in the network exceeds the threshold queue is greater, but it has no effect on $Q_m^{c,e}(t)$. In each experiment, no matter how

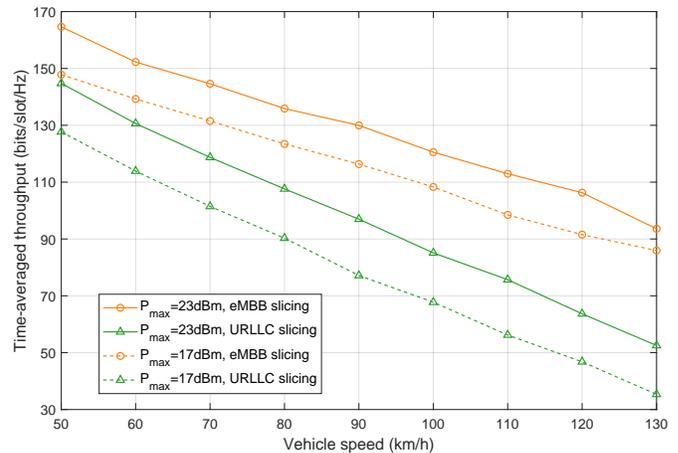


Fig. 8. Time-averaged throughput with varying vehicle speed v .

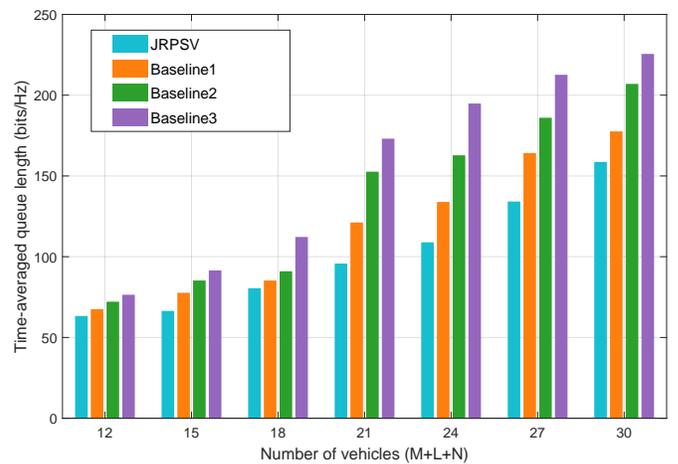


Fig. 9. Time-averaged queue length with varying number of vehicles under different schemes.

the value of $P_{\max}^{c,laten}$ and $P_{\max}^{u,laten}$ change, $Q_m^{c,e}(t)$ is still an unaffected random and stable data queue.

Figure 7 presents the impact of the D2D-V2V link outage probability on the average time throughput, and compares the time-averaged CDF of each slicing under different P_{\max}^{outage} . It is noted that when a higher outage probability of the D2D-V2V communication link is allowed, the time-averaged throughput of URLLC slicing ($\sum_{l \in \mathcal{L}} \bar{R}_l^{c,u} + \sum_{n \in \mathcal{N}} \bar{R}_n^{d,u}$) becomes larger, but the time-averaged throughput of eMBB slicing ($\sum_{m \in \mathcal{M}} \bar{R}_m^{c,e}$) hardly changes. This is due to the following facts. A higher outage probability of D2D-V2V communication link makes them more tolerant to the interference from C^u -VUEs, which encourages them to increase their transmit power. As a result, the increase of $\sum_{l \in \mathcal{L}} \bar{R}_l^{c,u}$ causes the time-averaged throughput of URLLC slicing larger. From another perspective, according to the constraint (35), a higher P_{\max}^{outage} causes $\gamma_{n,k}^{d,u}(t)$ to have a smaller lower bound, and increasing $P_{l,k}^{c,u}(t)$ appropriately can adapt to the change of the lower bound of the inequality, which leads to an increment in $\sum_{l \in \mathcal{L}} \bar{R}_l^{c,u}$.

Figure 8 illustrates that as the speed of the vehicle increases,

the time-averaged throughput of both slicing decreases. The reason is that according to the simulation settings, higher vehicle speeds result in sparse traffic, which will increase the average distance between vehicles. In order to guarantee the reliability of V2V links, the transmit power of the D2D-V2V transmitter needs to be increased to compensate for the higher path loss of the V2V link, which reduces the system capacity. In addition, it is worth mentioning that if the maximum transmit power enlarges, there will be a certain gain. For instance, when the vehicle speed is $v = 50\text{km/h}$, increasing the maximum transmit power of the vehicle by 6 dBm, the values of throughput for eMBB slicing and URLLC slicing become strengthened by 11.38% and 13.37%, respectively.

D. Scheme Comparison

In order to verify the superiority of the proposed algorithm in reducing the average system latency under the premise of maximizing the system capacity, we set up multiple schemes to evaluate the performance of the JRPSV algorithm. Fig. 9 mainly shows the change of time-averaged queue length in the URLLC slicing ($\sum_{l \in \mathcal{L}} \bar{H}_l^{c,u} + \sum_{n \in \mathcal{N}} \bar{Z}_n^{d,u}$) under distinct QoS schemes for different numbers of vehicles. Baseline 1: Disregard outage probability constraint for D2D-V2V communication links, which is equivalent to removing only the constraint (12j) in problem (12). Baseline 2: Latency constraints on URLLC slicing is not considered, which is equivalent to removing the constraints (12h) and (12i) in problem (12). Baseline 3: Any QoS constraint of URLLC slicing is not considered. It can be observed from Fig. 9 that as the number of vehicles grows, the time-averaged queue length of each scheme increases. When the number of vehicles is constant, compared with the other three baselines, the JRPSV algorithm can reduce the queue length to different degrees, which makes the average latency of the system lower. It is worth mentioning that when there are fewer vehicles in the system, the benefits brought by the JRPSV algorithm are lower than when there are more vehicles in the system. For example, when the total number vehicle is $M + L + N = 12$, compared to baseline 3, the JRPSV algorithm reduces the time-averaged queue length by 20.74%. When $M + L + N = 30$, the proposed algorithm brings 42.15% gain compared to baseline 3.

The above results can be attributed to the fact that when the number of vehicles in the system is small, in order to maximize the system capacity, the D2D-V2V pairs communicate in the unused mode as much as possible, and do not reuse the RBs of C^u -VUEs, which improves the SINR of the V2I and V2V links and meets the QoS requirements to a certain extent. When the number of vehicles in the system gradually increases, the D2D-V2V pairs start to reuse the RB of C^u -VUEs. The growing number of V2V links will cause more interference to V2I links, which further reduces the SINR of the vehicle communication link, thereby reducing the transmission rate. When the arrival rate is greater than the transmission rate, the task data queue enlarges rapidly, which triggers to undesirable latency. The JRPSV Algorithm considers the QoS requirements of multiple services comprehensively

and reduces the average system latency by decreasing queue backlogs.

VII. CONCLUSION

In this paper, we investigate the RB allocation and power control resource management in stochastic vehicular networks. Network slicing is utilized to provide customized and flexible services for each vehicular application, aiming to maximize the network system capacity while guaranteeing high reliability and low latency of vehicle communication links. The proposed online algorithm, JRPSV, can make asymptotically optimal resource allocation decisions based on the current network state information. Through comprehensive theoretical analysis, we prove that the gap between the optimal system capacity obtained by the JRPSV algorithm and the theoretical optimal system capacity is bounded. Meanwhile, the asymptotic optimal system capacity is reached at the expense of system latency, and there is a $[\mathcal{O}(1/V), \mathcal{O}(V)]$ trade-off between system capacity and system average latency. Furthermore, the impact of various parameters is revealed, which confirms the importance of comprehensive consideration of multiple QoS requirements, and provides valuable guidelines for the practical deployment of vehicular networks.

APPENDIX A

PROOF FOR THEOREM 1

First, the expression of $Q_m^{c,e}(t+1)^2$, $H_l^{c,u}(t+1)^2$ and $Z_n^{d,u}(t+1)^2$ are derived respectively according to the fact that $(\max[a-b, 0] + c)^2 \leq a^2 + b^2 + c^2 + 2a(c-b)$, $\forall a, b, c \geq 0$, which is given by

$$Q_m^{c,e}(t+1)^2 \leq Q_m^{c,e}(t)^2 + R_m^{c,e}(t)^2 + A_m^{c,e}(t)^2 + 2Q_m^{c,e}(t)[A_m^{c,e}(t) - R_m^{c,e}(t)], \quad (53a)$$

$$H_l^{c,u}(t+1)^2 \leq H_l^{c,u}(t)^2 + \varphi_l^{c,u}(t)^2 + R_l^{c,u}(t)^2 - 2H_l^{c,u}(t)R_l^{c,u}(t) + 2H_l^{c,u}(t)\varphi_l^{c,u}(t), \quad (53b)$$

$$Z_n^{d,u}(t+1)^2 \leq Z_n^{d,u}(t)^2 + \psi_n^{d,u}(t)^2 + R_n^{d,u}(t)^2 - 2Z_n^{d,u}(t)R_n^{d,u}(t) + 2Z_n^{d,u}(t)\psi_n^{d,u}(t). \quad (53c)$$

Then the expression of $F(\Theta(t+1)) - F(\Theta(t))$ is given by

$$\begin{aligned} F(\Theta(t+1)) - F(\Theta(t)) &= \frac{1}{2} \sum_{m \in \mathcal{M}} \{Q_m^{c,e}(t+1)^2 - Q_m^{c,e}(t)^2\} \\ &\quad + \frac{1}{2} \sum_{l \in \mathcal{L}} \{H_l^{c,u}(t+1)^2 - H_l^{c,u}(t)^2\} \\ &\quad + \frac{1}{2} \sum_{n \in \mathcal{N}} \{Z_n^{d,u}(t+1)^2 - Z_n^{d,u}(t)^2\}. \end{aligned} \quad (54)$$

Substituting (53a)-(53c) into (54), we have the inequality in (55), which is at the top of the next page, and non-negative constant C is defined as in (27). Therefore, the upper bound of drift-plus-penalty $G(\Theta(t))$ is given in (56) at the top of the next page.

Then we have

$$\begin{aligned} G(\Theta(t)) &= \Delta(\Theta(t)) - VE[R_{total}(t) | \Theta(t)] \\ &\leq C + E[\Phi(t) | \Theta(t)], \end{aligned} \quad (57)$$

where $\Phi(t)$ is defined the same as in (28).

$$\begin{aligned}
F(\Theta(t+1)) - F(\Theta(t)) &\leq \frac{1}{2} \sum_{m \in \mathcal{M}} \left\{ (R_m^{\max})^2 + (A_m^{\max})^2 \right\} + \frac{1}{2} \sum_{l \in \mathcal{L}} \left\{ (R_l^{\max})^2 + (\varphi_l^{\max})^2 \right\} + \frac{1}{2} \sum_{n \in \mathcal{N}} \left\{ (R_n^{\max})^2 + (\psi_n^{\max})^2 \right\} \\
&+ \sum_{m \in \mathcal{M}} Q_m^{c,e}(t) [A_m^{c,e}(t) - R_m^{c,e}(t)] + \sum_{l \in \mathcal{L}} H_l^{c,u}(t) [\varphi_l^{c,u}(t) - R_l^{c,u}(t)] + \sum_{n \in \mathcal{N}} Z_n^{d,u}(t) [\psi_n^{d,u}(t) - R_n^{d,u}(t)]. \quad (55)
\end{aligned}$$

$$\begin{aligned}
G(\Theta(t)) = \Delta(\Theta(t)) - VE[R_{total}(t) | \Theta(t)] &\leq C + E \left[\sum_{m \in \mathcal{M}} Q_m^{c,e}(t) A_m^{c,e}(t) - \sum_{m \in \mathcal{M}} Q_m^{c,e}(t) R_m^{c,e}(t) \right. \\
&\left. + \sum_{l \in \mathcal{L}} H_l^{c,u}(t) [\varphi_l^{c,u}(t) - R_l^{c,u}(t)] + \sum_{n \in \mathcal{N}} Z_n^{d,u}(t) [\psi_n^{d,u}(t) - R_n^{d,u}(t)] | \Theta(t) \right] - VE[R_{total}(t) | \Theta(t)]. \quad (56)
\end{aligned}$$

APPENDIX B PROOF FOR THEOREM 2

The parameters for optimal solution at each time slot obtained by Algorithm 2 are denoted by φ^* , ψ^* , \mathbf{P}^* and \mathbf{X}^* . Substituting these optimal variables into (56), we can obtain the following formula,

$$\begin{aligned}
&\Delta(\Theta(t)) - VE[R_{total}^*(t) | \Theta(t)] \\
&\leq C + E \left[\sum_{m \in \mathcal{M}} Q_m^{c,e}(t) (A_m^{c,e,*}(t) - R_m^{c,e,*}(t)) \right. \\
&+ \sum_{l \in \mathcal{L}} H_l^{c,u}(t) (\varphi_l^{c,u,*}(t) - R_l^{c,u,*}(t)) \\
&+ \left. \sum_{n \in \mathcal{N}} Z_n^{d,u}(t) (\psi_n^{d,u,*}(t) - R_n^{d,u,*}(t)) | \Theta(t) \right] \\
&- VE[R_{total}^*(t) | \Theta(t)]. \quad (58)
\end{aligned}$$

According to Theorem 4.5 in [19], the following properties can be obtained:

$$E[R_{total}^*(t)] \leq R_{total}^{opt} + \delta', \quad (59a)$$

$$E[A_m^{c,e,*}(t) - R_m^{c,e,*}(t)] \leq \delta', \forall m \in \mathcal{M}, \quad (59b)$$

$$E[\varphi_l^{c,u,*}(t) - R_l^{c,u,*}(t)] \leq \delta', \forall l \in \mathcal{L}, \quad (59c)$$

$$E[\psi_n^{d,u,*}(t) - R_n^{d,u,*}(t)] \leq \delta', \forall n \in \mathcal{N}. \quad (59d)$$

Substituting this set of inequalities into (58), and letting $\delta' \rightarrow 0$, it follows that

$$\Delta(\Theta(t)) - VE[R_{total}^*(t) | \Theta(t)] \leq C - VR_{total}^{opt}. \quad (60)$$

Taking expectation on both sides of (60) and summing up the time term $t = 0, 1, \dots, T-1$, we have

$$\begin{aligned}
&E[L(\Theta(T))] - E[L(\Theta(0))] - V \sum_{t=0}^{T-1} E[R_{total}^*(t)] \\
&\leq CT - VTR_{total}^{opt}, \quad (61)
\end{aligned}$$

where $\Theta(0) = 0$. Then dividing on both sides of (61) by VT , the result is

$$\frac{E[L(\Theta(T))]}{VT} - \frac{1}{T} \sum_{t=0}^{T-1} E[R_{total}^*(t)] \leq \frac{C}{V} - R_{total}^{opt}. \quad (62)$$

Let $T \rightarrow \infty$, and then we have $\bar{R}_{total}^* \geq R_{total}^{opt} - \frac{C}{V}$.

APPENDIX C PROOF FOR THEOREM 3

Substituting Slatter Conditions in Theorem 3 into (58), it follows that

$$\begin{aligned}
&\Delta(\Theta(t)) - VE[R_{total}^*(t) | \Theta(t)] \\
&\leq C + V\Omega(\delta) - \delta \sum_{m \in \mathcal{M}} Q_m^{c,e}(t). \quad (63)
\end{aligned}$$

Taking expectation on both sides of (63) and summing up the time term $t = 0, 1, \dots, T-1$, we have

$$\begin{aligned}
\bar{Q}_m^{c,e} &\leq \frac{C + V \left[\Omega(\delta) + \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E[R_{total}^*(t)] \right]}{\delta} \\
&+ \frac{E[L(\Theta(0))]}{\delta T}. \quad (64)
\end{aligned}$$

Let $T \rightarrow \infty$, and then $\bar{Q}_m^{c,e} \leq (C + V[\Omega(\delta) + R_{total}^{opt}]) / \delta$ is proven.

REFERENCES

- [1] K. Liu, X. Xu, M. Chen, B. Liu, L. Wu, and V. C. Lee, "A hierarchical architecture for the future Internet of vehicles," *IEEE Commun. Mag.*, vol. 57, no. 7, pp. 41–47, July 2019.
- [2] S. A. Kazmi, T. N. Dang, I. Yaqoob, A. Ndikumana, E. Ahmed, R. Hussain, and C. S. Hong, "Infotainment enabled smart cars: A joint communication, caching, and computation approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8408–8420, Mar. 2019.
- [3] S. Parkvall, E. Dahlman, A. Furuskar, and M. Frenne, "NR: The new 5G radio access technology," *IEEE Communications Standards Magazine*, vol. 1, no. 4, pp. 24–30, Dec. 2017.
- [4] Y. Xu, G. Li, Y. Yang, M. Liu, and G. Gui, "Robust resource allocation and power splitting in SWIPT enabled heterogeneous networks: A robust minimax approach," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10799–10811, Dec. 2019.
- [5] Y. Xu, X. Zhao, and Y. Liang, "Robust power control and beamforming in cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1834–1857, 4th Quart., 2015.
- [6] L. Liang, S. Xie, G. Y. Li, Z. Ding, and X. Yu, "Graph-based resource sharing in vehicular communication," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4579–4592, Jul. 2018.
- [7] L. Liang, G. Y. Li, and W. Xu, "Resource allocation for D2D-enabled vehicular communications," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 3186–3197, Jul. 2017.
- [8] W. Sun, E. G. Ström, F. Brännström, K. C. Sou, and Y. Sui, "Radio resource management for D2D-based V2V communication," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6636–6650, Aug. 2016.
- [9] W. Sun, D. Yuan, E. G. Ström, and F. Brännström, "Cluster-based radio resource management for D2D-supported safety-critical V2X communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2756–2769, Apr. 2016.
- [10] D. P. Bertsekas, R. G. Gallager, and P. Humblet, *Data networks*. Prentice-Hall International New Jersey, 1992, vol. 2.

- [11] M. Patra, R. Thakur, and C. S. R. Murthy, "Improving delay and energy efficiency of vehicular networks using mobile femto access points," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1496–1505, Feb. 2017.
- [12] C. Guo, L. Liang, and G. Y. Li, "Resource allocation for vehicular communications with low latency and high reliability," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 3887–3902, Aug. 2019.
- [13] C. Guo, L. Liang, and Y. Li, "Resource allocation for high-reliability low-latency vehicular communications with packet retransmission," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6219–6230, Jul. 2019.
- [14] J. Mei, K. Zheng, L. Zhao, Y. Teng, and X. Wang, "A latency and reliability guaranteed resource allocation scheme for LTE V2V communication systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3850–3860, Jun. 2018.
- [15] K. Xiong, S. Leng, J. Hu, X. Chen, and K. Yang, "Smart network slicing for vehicular Fog-RANs," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3075–3085, Apr. 2019.
- [16] S. Zhang, W. Quan, J. Li, W. Shi, P. Yang, and X. Shen, "Air-ground integrated vehicular network slicing with content pushing and caching," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 2114–2127, Sept. 2018.
- [17] X. Ge, "Ultra-reliable low-latency communications in autonomous vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 5005–5016, May 2019.
- [18] H. Yang, X. Xie, and M. Kadoch, "Intelligent resource management based on reinforcement learning for ultra-reliable and low-latency IoV communication networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4157–4169, May 2019.
- [19] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.
- [20] T. K. Vu, C.-F. Liu, M. Bennis, M. Debbah, M. Latva-Aho, and C. S. Hong, "Ultra-reliable and low latency communication in mmwave-enabled massive MIMO networks," *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 2041–2044, Sept. 2017.
- [21] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 46–55, Jan. 2002.
- [22] J. Papandriopoulos, J. Evans, and S. Dey, "Optimal power control for Rayleigh-faded multiuser systems with outage constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2705–2715, Nov. 2005.
- [23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.
- [24] Z. Li, Y. Wang, M. Liu, R. Sun, Y. Chen, J. Yuan, and J. Li, "Energy efficient resource allocation for UAV-assisted space-air-ground internet of remote things networks," *IEEE Access*, vol. 7, pp. 145 348–145 362, Oct. 2019.
- [25] *Technical Specification Group Radio Access Network: Study on LTE Based V2X Services; (Release 14)*, document TR 36.885 V2.0.0, 3GPP, Jun. 2016.
- [26] *WF on SLS Evaluation Assumptions for eV2X*, document R1-165704, 3GPP TSG RAN WG1 Meeting, May 2016.
- [27] P. Kyosti, "WINNER II channel models," *IST, Tech. Rep. IST-4-027756 WINNER II D1. 1.2 VI. 2*, Sept. 2007.



Yuanbin Chen received the B.S. degree in communications engineering from the Beijing Jiaotong University, Beijing, China, in 2019. He is currently pursuing the master degree in information and communication systems with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. His current research interests are in the area of vehicular networks, network slicing, intelligent reflecting surface and resource management in future wireless networks.



Ying Wang received the Ph.D. degree in circuits and systems from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2003. In 2004, she was invited to work as a Visiting Researcher with the Communications Research Laboratory (renamed NiCT from 2004), Yokosuka, Japan. She was a Research Associate with the University of Hong Kong, Hong Kong, in 2005. She is currently a Professor with BUPT and the Director of the Radio Resource Management Laboratory, Wireless Technology Innovation Institute, BUPT. Her research interests are in the area of the cooperative and cognitive systems, radio resource management, and mobility management in 5G systems. She is active in standardization activities of 3GPP and ITU. She took part in performance evaluation work of the Chinese Evaluation Group, as a Representative of BUPT. She was a recipient of first prizes of the Scientific and Technological Progress Award by the China Institute of Communications in 2006 and 2009, respectively, and a second prize of the National Scientific and Technological Progress Award in 2008. She was also selected in the New Star Program of Beijing Science and Technology Committee and the New Century Excellent Talents in University, Ministry of Education, in 2007 and 2009, respectively. She has authored over 100 papers in international journals and conferences proceedings.

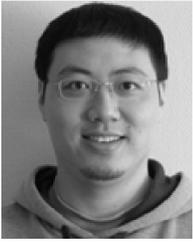


Man Liu received the B.S. degree in communications engineering from Zhengzhou University, China, in 2018. She is currently pursuing the master degree in telecommunication and information systems with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. Her research interests are in the areas of IoT, SAG Networks, and mobile cloud computing in future wireless networks.



Jiayi Zhang (S'08-M'14-SM'20) received the B.Sc. and Ph.D. degrees in communication engineering from Beijing Jiaotong University, China, in 2007 and 2014, respectively.

Since 2016, he has been a Professor with the School of Electronic and Information Engineering, Beijing Jiaotong University. From 2014 to 2016, he was a Postdoctoral Research Associate with the Department of Electronic Engineering, Tsinghua University, China. From 2014 to 2015, he was also a Humboldt Research Fellow with the Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nürnberg, Germany. From 2012 to 2013, he was a Visiting Scholar with the Wireless Group, University of Southampton, U.K. His current research interests include massive MIMO, large intelligent surface, communication theory, and applied mathematics. He received the Best Paper Awards at WCSP 2017 and IEEE APCC 2017. He was recognized as an Exemplary Reviewer of IEEE COMMUNICATIONS LETTERS in 2015 and 2016. He was also recognized as an Exemplary Reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS in 2017-2019. He was the Lead Guest Editor of the Special Issue on Multiple Antenna Technologies for Beyond 5G of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE COMMUNICATIONS LETTERS, IEEE ACCESS, and *IET Communications*.



Lei Jiao (S'08-M'12) received the B.E. degree in telecommunications engineering from Hunan University, Changsha, China, in 2005, the M.E. degree in communication and information system from Shandong University, Jinan, China, in 2008, and the Ph.D. degree in information and communication technology from the University of Agder (UiA), Agder, Norway, in 2012. He is currently with the Department of Information and Communication Technology, UiA, as an Associate Professor. His research interests include cognitive radio networks, wireless sensor networks, and artificial intelligence.