Received November 30, 2019, accepted January 20, 2020, date of publication January 31, 2020, date of current version February 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2970750

Joint Computation Offloading and URLLC Resource Allocation for Collaborative MEC Assisted Cellular-V2X Networks

LEI FENG, WENJING LI, YINGXIN LIN, LIANG ZHU, SHAOYONG GUO, AND ZERUI ZHEN

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Shaoyong Guo (syguo@bupt.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFE0205502, in part by the Fundamental Research Funds for the Central Universities under Grant 2019RC09, and in part by the State Grid Science and Technology project ''Analysis of Power Wireless Private Network Evolution and 4G/5G Technology Application'' under Grant 5700-201941235A-0-0-00.

ABSTRACT By leveraging the 5G enabled V2X networks, the vehicles connected by cellular base-stations can support a wide variety of computation-intensive services. In order to solve the arisen challenges in end-to-end low-latency transmission and backhaul resources, mobile edge computing (MEC) is now regarded as a promising paradigm for 5G-V2X communications. Considering the importance of both reliability and delay in vehicle communication, this article innovatively envisions a joint computation and URLLC resource allocation strategy for collaborative MEC assisted cellular-V2X networks and formulate a jointly power consumption optimization problem while guaranteeing the network stability. To solve this NP hard problem, we decouple it into two sub-problems: URLLC resource allocation for multi-cells to multi-vehicles and computation resource decisions among local vehicle, serving MEC server and collaborative MEC server. Secondly, non-cooperative game and bipartite graph are introduced to reduce the inter-cell interference and decide the channel allocation, which aims to maximize the throughput with a guarantee of reliability in URLLC V2X communication. Then, an online Lyapunov optimization method is proposed to solve computation resource allocation to get a trade-off between the average weighted power consumption and delay where CPU frequency are calculated using Gauss-Seidel method. Finally, the simulation results demonstrate that our proposed strategy can get better trade-off performance among power consumption, overflow probability and execution delay than the one based on centralized MEC assisted V2X.

INDEX TERMS Cellular V2X networks, URLLC radio resource management, collaborative mobile edge computing, power optimization, latency and reliability.

I. INTRODUCTION

As the increasing amount of connected autonomous vehicles, a wide variety of computation-intensive, latency sensitive and power-hungry applications are emerging, such as autonomous driving, image or video-aided real-time navigation, real-time traffic monitoring, etc. These applications need a significantly large amount of energy consumption, radio and computation resources, which brings great challenges to the operator with limited computation ability. As a result, cloud-based vehicular networks have been proposed as a solution to address this problem. Compared with local vehicular processing, remote servers have abundant storage space and resources for computation. However, the latency caused by the long distance between the remote cloud servers and the vehicles cannot be ignored and it may easily result in a considerable communication error probability. Vehicular cloud requires a scalable and reliable mobile communication network. LTE and Dedicated Short Range Communication (DSRC) have been trying to fit for such role, yet neither could be capable of meeting all requirements on account of their inherent architectural limitations.

In order to satisfy the severe latency requirements of these vehicular application scenarios, mobile edge computing (MEC) assisted V2X networks are now regarded as a promising paradigm to improve vehicular services through offloading computation-intensive tasks between edge servers and local vehicular terminals. The V2X communication in the

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

future should provide ultra-high reliability and low latency support for critical and broadband applications. The fifth generation (5G) mobile communication system will introduce ultra-reliable and low-latency communication (URLLC) to overcome these shortages, which can be fully employed as the data transmission method for computation offloading in MEC assisted V2X networks. Because the MEC servers operate at the edge of radio access networks, the rapid interactive computation offloading is available for vehicles in proximity. However, this solution suffers from the limitation of resource and radio access coverage. The computing offloading performance highly depends on the wireless transmission of data offloading from local vehicles to the MEC servers. Thus, effective task scheduling and resource allocation schemes are needed to improve system performance.

Obviously, the computing offloading performance is affected by the quality of wireless transmission. Therefore, it is important to properly allocate radio resources of wireless network to the multiple vehicles in the system. With the development of broadband services, in order to improve the utilization efficiency of radio resources, the channels in the multi-cells system are always reused, which may cause co-channel interference. The existence of the interference will result in the transmission rate attenuation of each vehicle, thereby affecting the data transmission reliability. In addition, the randomness of moving direction and variable speeds make it very complicated and challenging for dynamic resource scheduling. Therefore, the optimal allocation of radio resources including spectrum and power for each cell plays an important role in the multi-cell cellular-V2X networks.

A. RELATED WORK

Recently, there are a large number of studies on MEC-based vehicular networks and exploring the advantages of collaborative mobile edge computing networks [1]. The heterogeneous requirement on the mobility of vehicles are considered in [2] for vehicular network based on MEC framework and this work mainly focus on the MEC server selection and tasks transmission process. In [3], the author investigates the vehicular network with mobile edge servers deployed at the roadside units and using Lagrangian relaxation, in which the latency and workload requirements are well explored. In order to minimize the overall system costs of the vehicular network, [4] propose a mobility-aware mobile edge system and then solve the computational resource optimization and select the optimal offloading time. To satisfy users' experience in vehicular mobile edge computing, an adaptive computational resource allocation method is investigated in [5]. [6] is presented to jointly consider the cost at vehicle terminals and MEC servers under the system stability constraint using Lagrangian dual decomposition and relaxation.

Nevertheless, mobile edge computing can help improve the performance of vehicular, the MEC servers still have limited computational resources. Thus, some works proposed a cloud and edge collaborative system. In order to adopt the large computational resources in the central cloud server, [7] propose a game-theoretic collaboration task offloading algorithm. Reference [8] provides a partial computation offloading scheme for minimizing the delay and allocating optimal computational resources. In [9], it mainly focuses on maximizing the system utility by the optimal resource allocation and tasks offloading strategy. Reference [10] focus the multi-server scenario, but the scheme proposed only schedules one mobile edge server to the user.

Although many studies investigate MEC based vehicular network and collaboration between MEC servers and cloud. Very few articles study the collaboration between MEC servers. In 5G D-RAN, the base stations have the ability to communicate with each other. That is to say, the MEC servers installed at the base stations can transmit data with each other. This collaboration scheme can well explore the characteristic of collaborative MEC assisted cellular-V2X networks, and improve the processing ability and reliability required by the V2X network. But in previous researches, this collaborative MEC scheme has been ignored. Besides, the mobility of vehicles, the inter-interference between multi-vehicles, the low latency and ultra-high reliability of the system have not been presented simultaneously with mobile edge computing either. Therefore, in this article, we study the joint computation offloading and URLLC resource allocation strategy for collaborative MEC assisted cellular-V2X networks.

B. CONTRIBUTIONS

In this paper, we jointly optimize the URLLC radio and computational resources in collaborative MEC assisted cellular-V2X networks to minimize the overall power consumption for data offloading. The main contributions of this paper are presented as below:

1) A novel MEC task offloading model for 5G-V2X networks is proposed. To the best of our knowledge, it is the first time that collaboration between distributed MEC servers and URLLC transmission for task offloading are jointly considered in cellular V2X networks. We formulate the task offloading problem to minimize the power consumption of collaborative MEC servers and vehicles, which considers the constraint of the task buffer stability for hard delay in V2X. This model can cover the V2X performance requirement on the reliability, power consumption and latency.

2) In order to avoid co-channel interference among multi-cells and control the transmission power consumption for a reliable URLLC transmission, bipartite graph optimal matching algorithm is introduced solve the resource multiplex matching problem. We design a non-cooperative game power control algorithm to get the optimal edge weight of the bipartite graph. Both the utility and cost are considered in the pricing scheme so that the cellular vehicle communication system reaches a Nash equilibrium to pursue the maximized overall rate under the reliability guaranteeing.

3) An online computational resource allocation algorithm is proposed by using Lyapunov optimization method. We find the optimal average weighted power consumption and

execution delay trade-off under tasks buffers stability constrained by the hard delay in V2X. Under each deterministic process, the optimal solution to computational resources is decided, which can reasonably distribute the tasks into local vehicle, serving MEC server and collaborative MEC server. The simulation results demonstrate that the proposed algorithm can achieve efficient power consumption and execution delay. redFurthermore, the scheme based on distributed MEC collaboration is more reliable in overflow probability than the centralized MEC scheme in cellular V2X network.

The rest of the paper is organized as follows. The system model is presented in Section II. Then the problem is formulated as a problem for minimization of weighted energy consumption sum in Section III. The optimal algorithm is developed in Section IV. Next, we discuss and analyze the performance of the algorithm based on the simulation results in Section V. We finally draw a conclusion in Section VI.

II. SYSTEM MODEL

Fig.1 depicts a collaborative MEC assisted cellular-V2X networks, which consists of a distributed radio access network and mobile edge servers and multiple vehicles. From this graph, we can obtain the architecture of the coexistence MEC servers. There are N adjacent cells with a base station (BS) located at the center, respectively, and an MEC server operates in each BS. An edge node can be regarded as the combination of base stations and MEC servers and each node has the ability to transmit data with other nodes. In each cell, K vehicles are assumed to move independently inside. Different from the unidirectional road and RSU distribution model in above papers, the speed and direction of the vehicles in this paper are both random in their corresponding cell.

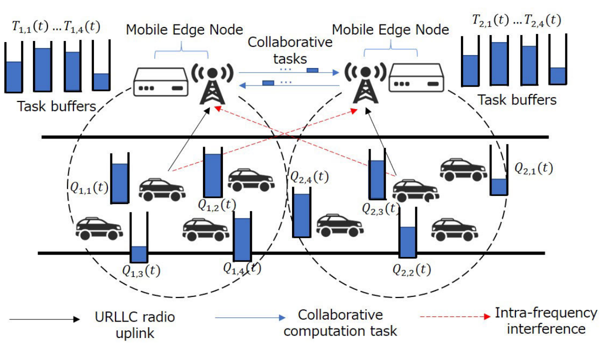


FIGURE 1. Collaborative MEC assisted cellular-V2X networks (2 MEC servers).

The overall computing tasks are supposed to be processed at both the mobile edge node and local vehicles. It is assumed that the wireless communication for data offloading from the vehicle terminals to the base station is through 5G URLLC links. And the tasks of vehicle moving in the n th cell can only be offloaded to the corresponding MEC server which is located in the base station B_n.

Due to the mini-slot scheme of URLLC, the moving distance of vehicles changes little in one mini-slot. Thus, we define a larger scheduling granularity as "time slot", which contains T mini-slots. At the beginning of each time slot, the vehicles' location in the cell changes, consequently the radio resource management scheme has to change. Respectively, at the beginning of each mini-slot, the offloading scheme is decided. The SISO-OFDMA scheme is adopted to avoid serious intra-cell interference. Thereby, the radio resources in each cell are distributed into several sub-channels or resource blocks (RBs) which are orthogonal in the time domain. But the vehicle terminals scheduled by the same sub-channel in different cells will result in inter-cell interference (i.e. co-channel interference) to each other, which will cause the transmission rate attenuation and affects the offloading efficiency.

A. VEHICLE AND MEC SERVER EXECUTION MODEL

During the task execution process, the vehicles can process the task in its local CPU or offloaded to be processed in MEC server. The task processing speed all depends on the speed of CPU frequency, which is strictly correlated with the type of vehicles and CPU. The computation ability can be calculated in off-line measurement. The k th vehicles in n th cell's CPU frequency is denoted as $f_{n,k}(t)$ and it is not larger than $f_{n,k,max}$. The CPU cycles needed in processing one bit of tasks is denoted as $L_{n,k}$. The computing ability in each time slot is defined as $D_{l,n,k}(t)$, which is the bits of tasks that can be executed by local CPU and it is positively related to frequency. Thereby, the equation of local computation ability can be expressed as,

$$D_{l,n,k}(t) = \tau f_{n,k}(t) L_{n,k}^{-1} \tag{1}$$

Similarly, the MEC computation ability is defined as $D_{s,n,k}(t) = \tau f_{C,n,k}(t) L_{C,n,k}^{-1}$. As is known to us, CPU needs the energy to maintain basic operation. CPU processing energy consumption will fluctuate when the logic gates flip and the consumption is positively related to the square of circuit voltage and frequency of CPU. In fact, the power dissipation at the output pins of a core is directly proportional to its frequency and is governed by the equation

$$P = \frac{1}{2} C V^2 f \tag{2}$$

where C is capacitance, V is voltage, and f is the effective bus frequency [11]. Under low voltage constraint, the energy consumption equation of local vehicles can be eliminated to $p_{l,n,k}(t) = k_{mob,n,k} f_{n,k}^3(t)$, and $p_{\mathrm{ser},n,k}(t) = k_{\mathrm{ser},n,k} f_{C,n,k}^3(t)$ because in this circumstance, the CPU frequency has linear correlativity with circuit voltage. In the equation, the effective switched capacitance is denoted as $k_{mob,n,k}$ and $k_{\mathrm{ser},n,k}$, and it is related to the architecture of chip.

B. URLLC TRANSMISSION FOR TASK OFFLOADING

For the n th cell, the uplink received SINR (i.e. signal-to-interference and noise-ratio) of the vehicle which is scheduled by the m th sub-channel at the time slot t can be denoted as:

$$\gamma_{n,k}^{m}(t) = \frac{G_{n,k}^{m}(t) \cdot P_{n,k}^{m}(t)}{\sigma^2 + \sum_{l=1,l\neq n}^{N} G_{l,k}^{m} \cdot P_{l,k}^{m}(t)} \tag{3}$$

where the $G_{n,k}^{m}(t)$ is the channel gain, $P_{n,k}^{m}(t)$ is the uplink transmit power of this vehicle, $\sigma^2 = N_0 \times B$ is the noise power and the N_0 denotes the noise unilateral power spectral density. $\sum_{l=1,l\neq n}^{N} G_{l,k'}^{m} \cdot P_{l,k'}^{m}(t)$ denotes the interference from the vehicle in the adjacent cell.

According to the accurate estimation of the achievable rate with the finite packet length [12]–[15], at the t time slot, the maximum uplink transmission rate of the vehicle scheduled by the m th sub-channel can be expressed as follows (units:bits/s),

$$R_{n,k}^{m}(t) = B \cdot \left\{ \log_2 \left[1 + \gamma_{n,k}^{m}(t)\right] - \sqrt{\frac{V_k}{n_0}} f_Q^{-1}\left(\varepsilon_k^d\right) \right\} \tag{4}$$

where B represents the bandwidth of a sub-channel and $V_k = 1 - \frac{1}{\left[1+\gamma_{n,k}^{m}(t)\right]^2}$ denotes the channel dispersion [12], which measures the stochastic variation of channels relative to deterministic channels with the same capacity. The upper bound of V_k is 1, especially in the high SINR-scenario of URLLC. n_0 is the length of data packet. $f_Q^{-1}(x)$ is the inverse function of Q function and $Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}t^2\right)} dt \cdot \varepsilon_k^d$ denotes the decoding error rate and it is assumed to have a small threshold for simplicity. In this paper, we decide to improve the URLLC transmission performance under the fixed requirement of decoding error rate (i.e. reliability).

In time slot t, the sum of rate at the vehicle k and the cell n in the sub-channel m can be presented as below:

$$\text{sum\_} R_{n,k}^{m}(t) = \sum_{m=1}^{M} a_{n,k}^{m}(t) \cdot R_{n,k}^{m}(t) \tag{5}$$

where $a_{n,k}^{m} = 1$ means the sub-channel m in the cell n is allocated to the vehicle k at this time slot, otherwise it is allocated to other vehicle. Thus, the total system capacity is the uplink transmission rate sum of all vehicle terminals:

$$T(t) = \sum_{n=1}^{N}\sum_{k=1}^{K} sum\_R_{n,k}^{m}(t) = \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{m=1}^{M} a_{n,k}^{m}(t) R_{n,k}^{m}(t) \tag{6}$$

C. TASK QUEUING MODEL

1) VEHICLES

Suppose that the tasks received by vehicles are running in fine-grained parallelism. At the initial moment of the t th time slot, $A_{n,k}(t)$ bits of tasks arrive at the k th vehicles in the n th cell, which will be processed since next time slot. $A_{n,k}(t)$ at distinct time slots are distributed within the range of $A_{n,k}(t)$ with $E\left[A_{n,k}(t)\right] = \lambda_{n,k}$ [16]. In every time slot, the tasks to be processed by local CPU is defined as $D_{l,n,k}(t)$, the tasks to be offloaded to MEC servers is denoted as $R_{n,k}^{m}(t)$ and the remaining tasks that have arrived but not yet been processed and offloaded will wait in the queue backlogs of vehicles with finite capacity. The queue length of task buffer at the t th time slot can be expressed as N rows K columns array $\boldsymbol{Q(t)} = [[Q_{1,1}(t), \ldots, Q_{1,K}(t)], \ldots, [Q_{N,1}(t), \ldots, Q_{N,K}(t)]]$. $Q_{n,k}(t+1)$ can be derived from $Q_{n,k}(t)$ as below.

$$Q_{n,k}(t+1) = max(Q_{n,k}(t) - D_{\sum,n,k}(t), 0) + A_{n,k}(t) \tag{7}$$

in (7), $D_{\sum,n,k}(t)$ equals to the sum of $D_{l,n,k}(t)$ and $R_{n,k}^{m}(t)$. red It stands for the number of tasks executed by local vehicles and offloaded to mobile edge servers at k th vehicle and the n th cell in the t th time slot.

2) MEC SERVER

Similar to local vehicles, the tasks offloaded to the server but have not been executed will be stored in the task buffer of edge nodes and the task buffer is also assumed to have a finite capacity. Suppose each cell has one MEC server and the queue backlogs of task buffer can be expressed as vector $\boldsymbol{C(t)} = [C_{1,1}(t), \ldots, C_{1,K}(t), \ldots, C_{N,1}(t), \ldots, C_{N,K}(t)]$. Accordingly, the expression of $C_{n,k}(t+1)$ for a non-collaborative MEC system is established as below,

$$\begin{aligned} C_{n,k}^{incop}(t+1) &= max(C_{n,k}(t) - D_{s,n,k}(t), 0) \\ &\quad + min(max(C_{n,k}(t) - D_{l,n,k}(t), 0), R_{n,k}^{m}(t)) \end{aligned} \tag{8}$$

(8) indicates that tasks that have not been handled by edge server from previous time slot will be stored in the task buffer of edge node Besides, tasks that are not processed at local CPU should be offloaded to the edge server, and the amount should be less than the maximum transmit capacity. If the number of tasks exceeds transmit capacity constraint, the excess amount will be ignored apparently, as the amount is determined by channel capacity and queue length simultaneously.

When the MEC-based 5G URLLC vehicular network is collaborative, the queue length can be expressed as,

$$\begin{aligned} &C_{n,k}^{cop}(t+1) \\ &= max(C_{n,k}(t) - D_{s,n,k}(t), 0) \\ &\quad + min(max(C_{n,k}(t) - D_{l,n,k}(t), 0), R_{n,k}^{m}(t)) + G_{n,k}(t) \end{aligned} \tag{9}$$

where $G_j(t)$ is data transmitted from collaborative MEC servers.

III. PROBLEM FORMULATION AND DECOMPOSITION

In this section, we will jointly discuss the weighted mean of energy consumption and the average queue backlogs of the whole system. Furthermore, the average energy consumption optimization problem will be established under all constraints that are already provided. Finally, we formulate this NP hard problem to minimize the power consumption of collaborative MEC servers and vehicles, which considers the constraint of the task buffer stability for hard delay in V2X introduced as the NP hard problem to be solved in this paper.

A. PERFORMANCE METRIC

When transmitting stochastic traffic flows over wireless networks, there exists an inherent tradeoff between average transmit power and corresponding queuing-delay bound. In [17], authors investigate such a tradeoff and show how average power increases as delay-bound requirement for wireless network traffics becomes stringent. Under certain QoS conditions, what we need to do is reducing the corresponding power consumption. Thus, the weighted mean of energy consumption is one of the most concerned issues in this paper, which mainly consists of the energy consumption of CPU processing and task transmission power. In addition, the energy required to maintain the basic operation of vehicles and MEC servers is irrelevant to the offloading process and it is ignored for simplicity. Therefore, the weighted mean of energy consumption in the offloading and processing process is given by,

$$\bar{P}_{\sum} \triangleq \lim_{T\to+\infty}\frac{1}{T}\sum_{t=0}^{T-1}E\Big[\sum_{n\in N}\sum_{k\in K}(P_{loc,n,k}+P_{MEC,n,k} + \omega_{N+K+1}\mu p_{n,k}^{tx}(t))\Big] \tag{10}$$

where $P_{MEC,n,k} = \omega_{K+N+1}p_{ser,n,k}(t)$ and $P_{loc,n,k} = \omega_{n,k}(p_{tx,n,k}(t)+p_{l,n,k}(t))$. It is also the performance metric of the system model. $\omega_{n,k}$ and ω_{N+K+1} are parameters used to balance the energy consumption between different nodes and μ indicates the weight of distribution. All these three notations are assumed to be constant. For example, assuming that the energy consumption of MEC has a greater impact on the whole offloading process. Thereby we can set the weight of MEC energy consumption larger during the stimulation stage, so as to achieve the goal of balancing energy consumption. The average of T is to adjust the sum of energy consumption to a function related to t and to build the foundation for the subsequent calculation process.

Based on Little's Law [18], the weighted mean processing delay of tasks is positively related to the length of tasks waiting in the task buffers both at the server and local vehicles. In [16], as a measurement of processing delay, the total queue length of task buffers at both server and local vehicles is expressed as below,

$$\bar{q}_{\sum,n} \triangleq \lim_{T\to+\infty}\frac{1}{T}\sum_{t=0}^{T-1}E\Big[\sum_{k\in K}(Q_{n,k}(t)+\mu C_{n,k}^{cop}(t) + (1-\mu)C_{n,k}^{incop}(t))\Big] \tag{11}$$

However, relying merely on the average queue length fails to take the extreme value of queue length into account. Therefore, for the better assessment of the system performance, a hard delay is proposed and it is regarded as proportional to the maximum value of queue length [19], [20].

$$delay_{n,k}(t) = \frac{q_{max}}{R_{n,k}^{m}(t)} \tag{12}$$

where q_{max} is the maximum value of queue length that appears during the whole transmission process. $delay_{n,k}(t)$ is the hard latency at each time slot. In the simulation process, there is a threshold assigned to the data traffic and thus the worst delay can be acquired now.

B. AVERAGE ENERGY CONSUMPTION OPTIMIZATION

At each time slot, a series of variable denoted as $\boldsymbol{X(t)} \triangleq [\boldsymbol{f(t)}, \boldsymbol{p_{tx}(t)}, \boldsymbol{f_c(t)}]$ will be calculated during every iteration. Then the average energy consumption optimization problem will be established as below,

$$P_1: \min_{X(t)} \bar{P}_{\sum}$$

$$0 \le \mu \le 1 \tag{13}$$

$$0 \le f_{n,k}(t) \le f_{n,k,max}, n\in N, k\in K \tag{14}$$

$$0 \le f_{C,n}(t) \le f_{C_n,max}, n\in N \tag{15}$$

$$0 \le p_{n,k}^{tx}(t) \le p_{n,k}^{tx,max}(t) \tag{16}$$

$$\lim_{T\to+\infty}\frac{E[|Q_{n,k}(t)|]}{T}=0,\quad \lim_{T\to+\infty}\frac{E[|U_n(t)|]}{T}=0, \tag{17}$$

(13) stands for the distribution weight constraint. (14) is the constraint for local CPU frequency. (15) is the constraint for MEC CPU frequency. (16) is the constraint to transmit power. (17) limits the task buffers to be mean rate stable so that all arrived offloading tasks will be handled during finite time slots.

In a word, the energy consumption and processing delay for collaborative MEC assisted cellular-V2X networks should be jointly considered and they can be solved by a decomposed algorithm. All the constraints on CPU frequency, transmit power and the stability of task buffer will be satisfied when minimizing the average energy consumption.

IV. JOINT OPTIMIZATION ALGORITHM DESIGN

Since the energy consumption optimization problem above is NP-hard in general. The target problem is supposed to be decomposed into subproblems and then iteratively and separately solve them until convergence. For example, in [21], a decomposition technique can be used to solve this problem efficiently. We decompose the power consumption minimization problem into two subproblems: URLLC resource allocation for multi-cells to multi-vehicles and task offloading decisions among local vehicle, serving MEC server and collaborative MEC server. Then, the two subproblems are solved separately and the power consumption optimization problem are solved iteratively. A common framework for the original problem is summarized as Fig. 2.

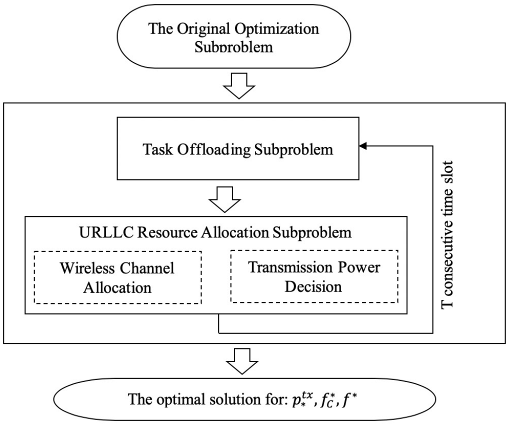


FIGURE 2. Proposed framework for solving the power consumption problem.

A. RESOURCE MANAGEMENT FOR THE MAXIMIZATION OF OVERALL URLLC TRANSMISSION RATE FOR COMPUTING OFFLOADING

Firstly, the transmission energy consumption, $p^{tx}_{n,k}(t)$ needs to be determined, which is directly related to the resource scheduling and power control scheme. Then the computation energy consumption at the local and the MEC server will be optimized. For improvement of the efficiency in data offloading to the MEC server, the throughput with a guarantee of reliability in URLLC V2X communication need to be maximized. Therefore, the sub-problem is to maximize the overall uplink throughput, which is presented as below:

$$\begin{aligned}
&\underset{P^m_{n,k}(t)\in(0,P_{\max}]}{\arg\max} \quad T(t) \\
&= \underset{P^m_{n,k}(t)\in(0,P_{\max}]}{\arg\max} \sum_{n=1}^{N}\sum_{k=1}^{K} a^m_{n,k}(t)\cdot R^m_{n,k}(t), \\
&\qquad \forall m \in \{1,2,\cdots,M\} \\
&\qquad C1: 0 < P^m_{n,k}(t) \le P_{max} \\
&\qquad C2: R^m_{n,k}(t) \ge 0 \\
&\qquad C3: \sum_{k=1}^{K} a^m_{n,k}(t) = 1 \\
&\qquad C4: 10^{-7} \le \epsilon^d_k \le 10^{-5} \qquad (18)
\end{aligned}$$

where C1 describes the range of uplink transmit power. C2 is the rate constraint, which means that the achievable transmission rate is non-negative. C3 guarantees that one sub-channel in each cell can only be scheduled to a vehicle terminal inside the cell during one time slot. C4 is the constraint of the decoding error rate. The sub-channel in the OFDM system is orthogonal in the frequency domain. Thus, the resource allocation of each sub-channel can be considered as an independent process. The problem of overall system throughput maximization can be summarized to a multi-resource-block throughput maximization problem, which is shown in (18). The vehicles assigned to the same sub-channel m in each cell cause co-channel interference to each other and they are all selfish. Higher uplink transmits power is expected to improve the achievable transmission rate as much as possible. However, a higher transmit power will make other terminals suffer a higher interference. Nevertheless, the transmission rate of other vehicles will be reduced, which results in a contradiction in power control. Thus, all the variables need to be jointly considered.

Due to the mobility of vehicles, the resource management including channel allocation and uplink power control becomes much more complicated. In order to address the problem above, all possible combinations of sub-channel assignments in different cells need to be traversed to determine the optimal one, which is computationally complex and difficult. For simplicity, in this paper, we proposed a joint uplink resource management scheme in order to lower the inter-cell interference as much as possible and make an optimal strategic decision about the channel multiplexing matching at the same time. At each scheduling time slot, the position of the vehicle is assumed to be fixed. The channel multiplexing matching set and power set that maximize the total system throughput (i.e. the sum of transmission rate) will be selected. In this paper, the two-cell channel multiplex matching problem is considered as a bipartite graph optimal matching problem. Each vehicle in the system is considered as a vertex in the bipartite graph, as shown in Fig 3. In this scenario, the two vertices connected by an edge represent the two vehicles which multiplex the same sub-channel resource (or the resource block). The weight of this edge is the overall throughput of this shared sub-channel, which is maximized through the non-cooperative power control game algorithm below. Therefore, the system overall throughput maximization problem is transformed into an optimal matching problem that maximizes the edge weight sum of the bipartite graph.

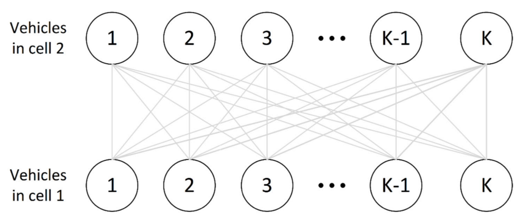


FIGURE 3. Resource allocation based on bipartite graph with K vehicles and K MEC cells.

At each time slot, we obtain the weight of each edge. For each pair of vehicles in two separate cells, the real-time position and channel gain are clear to each other and the non-cooperative power control game can be carried out to lower the inter-cell interference. When the game converges to the final equilibrium point, the overall throughput of this sub-channel can be calculated as the weight of this edge.

1) NON-COOPERATIVE POWER CONTROL GAME ALGORITHM BASED ON PRICING SCHEME

For the problem (18), the system throughput is maximized through the maximization of each subchannel's throughput.

However, for each sub-channel, the throughput maximization is a high-order derivative problem involving N power variables, which is computationally complex and does not necessarily have a solution set.

For simplicity, the non-cooperative game is introduced and the power control problem of each sub-channel is considered as an independent non-cooperative game process. All the vehicle terminals assigned to the same sub-channel are the participants of the game. Since each participant in this game is selfish and their strategy is always choosing the maximum of power to achieve their own interest maximization, which will result in non-negligible interference and the rate decay. To avoid vicious selfish competition and control the power assumption, the pricing scheme is introduced [22]. In the choice of power strategy, the vehicles need to consider both their own utility and the corresponding cost due to interference caused by themselves.

The non-cooperative game of the mth sub-channel in this paper is denotes $G_m(t) = [N, P_m(t), U_m(t)]$, where $N = \{u_m^1, u_m^2\}$ represents the sub-channel matching set. $P_m(t) = \{P_{l,k}^m(t), P_{2,k}^m(t)\}$ is the power strategy set, and $U_m(t)$ denotes the net utility function set of vehicles. The power strategy value space can be described as $P_{n,k}^m(t) = [(P_{n,k}^m(t))_{min}, P_{max}]$, where $(P_{n,k}^m(t))_{min}$ is the minimal power to guarantee that the transmit rate is non-negative and P_{max} is the budget of uplink transmit power. The strategy space is a closed bounded convex set.

The net utility of vehicle terminals is the achievable rate minus the pricing function (19). The pricing factor $c_{n,k}^m(t)$ denotes the cost to pay per unit power. Therefore, all the game participants n, k aim at maximizing their own net utility. And the optimal power response of each game participant can be denoted as $\arg \max_{P_{n,k}^m(t)} U_{n,k}^m(t)$, where Γ denotes the channel constant.

$$U_{n,k}^m(t) = B \cdot \left\{ \log_2 \left[1 + \frac{\gamma_{n,k}^m(t)}{\Gamma} \right] - \sqrt{\frac{V_k}{n_0}} f_Q^{-1}\left(\varepsilon_k^d\right) \right\} - c_{n,k}^m(t) \cdot P_{n,k}^m(t) \tag{19}$$

Theorem 1: A unique Nash equilibrium exists in the $G_m(t)$ for the sub-channel m at time slot t.

Proof: The Theorem 1 is proved if the following two assumptions are guaranteed. (i) The power strategy space $[(P_{n,k}^m(t))_{min}, P_{max}]$ is a non-empty, bounded and closed convex set. (ii) The net utility function $U_{n,k}^m(t)$ is continuous and quasi-concave in $P_{n,k}^m(t)$.

According to (19), it is obvious that the second derivative is negative so that the (i) and (ii) are proved. When the first derivative is assigned to zero, we can obtain the optimal power response (20).

$$P_{n,k}^m(t) = \frac{B}{c_{n,k}^m(t)\ln 2} - \frac{\Gamma\left(\sigma^2 + \sum_{l=1,l\neq n}^N G_{l,k}^m \cdot P_{l,k}^m(t)\right)}{G_{n,k}^m(t)} \tag{20}$$

It is easy to prove that the monotonicity of the net utility function. Thus, the maximal is obtained at the point where the first derivative is zero. Based on the limitations of power strategy, the lower bound of $c_{n,k}(t)$ is denoted as (21) and the upper bound is denoted as (22).

$$c_{n,k}^{min}(t) = \frac{B}{\left(P_{\max} + \frac{\Gamma\left(\sigma^2 + \sum_{l=1,l\neq n}^N G_{l,k'}^m \cdot P_{l,k'}^m(t)\right)}{G_{n,k}^m(t)}\right)\ln 2} \tag{21}$$

$$c_{n,k}^{max}(t) = \frac{B G_{n,k}^m(t)}{2^{c_0} \cdot \Gamma\left(\sigma^2 + \sum_{l=1,l\neq n}^N G_{l,k'}^m \cdot P_{l,k'}^m(t)\right)\ln 2} \tag{22}$$

Therefore, at each time slot t, the range of pricing factor for each vehicle's matching between two cells can be calculated by the intersection of its upper limit and the union of its lower limit is calculated. And the optimal pricing factor needs to be decided for each matching in order to obtain the maximal throughput sum, which will be set as the weight later. The best power response of the k th vehicle in cell n is showed as (23). Then each game participant will carry out the optimal power response according to other participants' power strategy till the overall power set converges to an equilibrium point.

$$P_{n,k}^m(t)^* = \begin{cases} (P_{n,k}^m(t))^{min} \dfrac{\Gamma\cdot(2^{c_0}-1)\cdot\left(\sigma^2 + \sum_{l=1,l\neq n}^N G_{l,k'}^m \cdot P_{l,k'}^m(t)\right)}{G_{n,k}^m(t)}, \\ \qquad P_{n,k}^m(t) \le \left(P_{n,k}^m(t)\right)^{\min}) \\ \dfrac{B}{c_{n,k}^m(t)\ln 2} - \dfrac{\Gamma\left(\sigma^2 + \sum_{l=1,l\neq n}^N G_{l,k'}^m \cdot P_{l,k'}^m(t)\right)}{G_{n,k}^m(t)}, \\ \qquad \left(P_{n,k}^m(t)\right)^{\min} < P_{n,k}^m(t) < \left(P_{n,k}^m(t)\right)^{\max} \\ P_{max}, \quad p_{max}^n \ge P_{max} \end{cases} \tag{23}$$

To prove the uniqueness, we have guaranteed that the best power response is a standard function which has the following three properties. (i) Positivity. (ii) Monotonicity. (iii) Scalability. According to [22], the Nash Equilibrium point is unique for the $G_m(t)$ ■

2) BIPARTITE GRAPH OPTIMAL MATCHING ALGORITHM FOR MULTI-CELL CHANNEL RESOURCE MULTIPLEXING

For the two-cell scenario, the multiplex matching scheme of the wireless channel is critical to the wireless system's transmission efficiency. In this paper, we consider adopting the Kuhn-Munkres (KM) algorithm to obtain the optimal channel reusing scheme efficiently and rapidly. It is assumed that each sub-channel can only be scheduled to one vehicle

Algorithm 1 The Non-Cooperative Game Power Control Algorithm

Input : $G^{m}_{n,k}(t), K, T_{end}, B, P_{max}, N_0, c_0, T, S, c_{min}, c_{max}$;
Output: $RS_{max}, P^1_t, P^2_t, R^1_t, R^2_t$;
for $t = 1; t \leq T_{end}; t++$ **do**
  for $k1 = 1; k1 \leq K; k1++$ **do**
    for $k2 = 1; k2 \leq K; k2++$ **do**
      for $s = 1; s \leq S; s++$ **do**
        (1) Calculate the price factor $price = c_{min}(k1, k2, t) + s * (c_{max}(k1, k2, t))/S$;
        (2) Update the power strategies iteratively according to the optimal power response (23) for k1 and k2 until it converges to a fixed point;
        (3) Calculate the transmit power and rate according to (4) and (23);
      end
      Determine the optimal pricing factor and set the corresponding throughput RS_{max} as the weight of this edge;
    end
  end
end

at each time slot to avoid intra-cell interference. And each vehicle terminal cannot occupy more than one sub-channel at the same time. In this paper, the sub-channels are considered to be identical so that it is not necessary to consider all the combinations of all sub-channels and vehicles. It is the vehicle matching set between two cells that really matters. The quantity of the sub-channels is assumed to be equal to the number of vehicles so that the problem is a perfect matching problem for the bipartite graph.

The bipartite graph is built and the vehicles in two cells are regarded as two groups of vertexes of the bipartite graph, respectively. The number of vehicles in each cell is equal to K. The steps of the bipartite graph optimal matching algorithm can be concluded as Algorithm 2.

Algorithm 2 Bipartite Graph Optimal Matching Algorithm

Input : $V, E, A, B, W_{i,j}(t), T_{end}$;
Output: $P_{final}, RS_{final}, match(t)$
for $t = 1; t \leq T_{end}; t++$ **do**
  (1) Set weight matrix r according to $W_{i,j}(t)$;
  (2) Assign the initial vertex labels: $L_{1,i}(t) = max(W_{i,j}(t))$ and $L_{2,j}(t) = 0$;
  (3) Find a complete match with the Hungarian algorithm, if it fails go to (4);
  (4) Improve the labeling by slack array: $slack(y) = min\{l(x) + l(y) - w(x, y)|x \in S\}$. Calculate the labeling changing value: $\Delta = min\{slack(y)|y \in B\backslash T\}$. Then improve the labeling of vertex r:
  if $r \in S$ **then**
    $l'(r) = l(r) - \Delta$;
  else if $r \in T$ **then**
    $l'(r) = l(r) + \Delta$;
  else if $r \notin S$ *and* $r \notin T$ **then**
    $l'(r) = l(r)$
  end
  (5) Repeat the step 3 and 4 till the complete match of equality subgraph is found.
end

The time complexity is $O(n^3)$. Now the radio resource management scheme is finished so the optimal sub-channels matching set and uplink transmit power set are obtained. The actual transmission rate for uplink data offloading can be calculated. The steps can be concluded in the pseudo-code as below.

B. ONLINE COMPUTATIONAL RESOURCE MANAGEMENT ALGORITHM

In this section, an online Lyapunove optimization method is proposed to solve computation task offloading to get a trade-off between the average weighted power consumption and delay. With this method and drift-plus-penalty, the time average of a stochastic problem can be minimized under a series of constraints. At every time slot, the optimal solution will be decided by the Lyapunov optimization framework. In a word, the algorithm and calculation process are able to achieve asymptotic optimality.

From definition of Lyapunov, we defined the quadratic Lyapunov functions as formula (24)

$$L(\Theta(t)) = \frac{1}{2}\sum_{n\in N}\sum_{k\in K}\Big((1-\mu)(C^{incop}_{n,k}(t))^2 + \mu(C^{cop}_{n,k}(t))^2 + Q^2_{n,k}(t)\Big) \quad (24)$$

where $\Theta(t) \triangleq [Q(t), C(t)]$. This function is defined to measure the total queue length in the system. By defining the Lyapunov drift as the change in the quadratic Lyapunov function from one slot to the next, a drift function can be written as,

$$\Delta L(\Theta(t)) = L(\Theta(t+1)) - L(\Theta(t)) \quad (25)$$

For stabilizing a queuing network while also minimizing the delay in network penalty function, the drift-plus-penalty method is used. Using this methodology, the current queue state is observed and the optimization actions are taken to minimize the upper limit of drift plus penalty function in the formula.

$$\Delta_V(\Theta(t)) = \Delta L(\Theta(t)) + V \times P_{\sum}(t) \quad (26)$$

where $P_{\sum}(t)$ is the penalty function and V is a non-negative weight which is chosen to adjust to optimal point, with a trade-off in the queue length of task buffer. Based on the equation, we can figure out the upper bound of (27).

$$\begin{aligned}&\Delta_V(\Theta(t))\\&\triangleq C - E\Big[\sum_{n\in N}\sum_{k\in K} Q_{n,k}(t)(D_{\sum,n,k}(t) - A_{n,k}(t))|\Theta(t)\Big]\\&\quad - E\Big[\sum_{n\in N}\sum_{k\in K}(1-\mu)C^{incop}_{n,k}(t)(D_{s,n,k}(t) - R^m_{n,k}(t))|\Theta(t)\Big]\\&\quad - E\Big[\sum_{n\in N}\sum_{k\in K}C^{cop}_{n,k}(t)(D_{s,n,k}(t) - R^m_{n,k}(t) - G_{n,k}(t))|\Theta(t)\Big]\\&\quad + V\times E[P_{\sum}(t)|\Theta(t)] \qquad (27)\end{aligned}$$

Algorithm 3 A Lyapunov Optimization-Based Online Resource Allocation Algorithm

Input : $\Theta(t), A_{n,k}(t), n = 1, 2, \cdots, N, k = 1, 2, \cdots, K$;
Output: $f(t), f_{ser}(t)$
while $t \le T$ **do**
  for $n = 1; n \le N; n++$ **do**
    for $k = 1; m \le K; k++$ **do**
      Determine $\boldsymbol{f(t), f_{ser}(t)}$ by solving,
      P2: $\min_{X(t)} \sum_{n\in N}\sum_{k\in K}\Big(Q_{n,k}(t)D_{\sum,n,k}(t) - (1-\mu)C^{incop}_{n,k}(t)(D_{s,n,k}(t) - R^m_{n,k}(t)) - \mu C^{cop}_{n,k}(t)(D_{s,n,k}(t) - R^m_{n,k}(t) - G_{n,k}(t))\Big) + V\cdot P_{\sum}(t)$
      where $P_{\sum}(t) \triangleq \sum_{n\in N}\sum_{k\in K}\Big(\omega_{n,k}(p_{tx,n,k}(t) + p_{l,n,k}(t)) + \omega_{K+N+1}(p_{ser,n,k}(t) + \mu p^{tx}_{n,k}(t))\Big)$
      $Q_{n,k}(t)$, $C^{incop}_{n,k}(t)$ and $C^{cop}_{n,k}(t)$ are updated according to (6) and (7)
    end
  end
  t = t + 1;
end

Algorithm 3 is designed to make decisions greedily at every time slot so that the upper bound will also be minimized on each time slot t. It is obvious that when the tasks waiting in the task buffer are minimized then the total energy consumption will be minimized at the same time. The object function of the algorithm is at the right-hand side of the Lyapunov function. The optimal solution to the deterministic problem will be discussed in the next section.

With vehicles' CPU frequency and the mobile edge servers' CPU frequency to be determined, the original problem is decomposed into two sub problems. The irrelevant variables in (27) can be regarded as constant.

1) OPTIMAL CPU FREQUENCY OF VEHICLES

As variables irrelevant to these problems is regarded as constant, the original equation is eliminated to (28), and the optimal solution to the original problem is transferred to the optimal solution to the following equation.

$$\begin{aligned}&\min_{f(t)} \sum_{i\in N}\big(-Q_{n,k}(t)\tau f_{n,k}(t)L^{-1}_{n,k} + V\times\omega_{n,k}k_{mob,n,k}f^3_{n,k}(t)\big),\\&s.t.\ 0 \le f_{n,k}(t) \le f_{max},\ i\in N \qquad (28)\end{aligned}$$

The first-order derivative is increasing and the second-order derivative is greater than zero, thus the problem is convex under linear constraints. Furthermore, this variable $f_{n,k}(t)$ is decoupled from other variables. According to the theory of extremum, the extreme value is achieved at the stationary point or the boundary points, which is calculated as (29).

$$f_{n,k}(t) = \begin{cases} \min\Big(f_{n,k,max}, \sqrt{\frac{Q_{n,k}(t)\tau}{3k_{mob,n,k}\omega_{n,k}VL_{n,k}}}\Big), & \omega_{n,k} > 0\\ 0, & \omega_{n,k} = 0\end{cases} \qquad (29)$$

According to the equation above, it is obviously that $f^*_{n,k}(t)$ is negatively correlated with V, $L_{n,k}$, $\omega_{n,k}$. At this point, with the larger number of $\omega_{n,k}$ and V, the weight of total energy consumption becomes larger, thus the CPU frequency need to decrease. These variables are not required to be decided by the system operation. While for the task buffers of local CPU, it is a variable that correlated with other deterministic variables and it is influenced according to the relationship in the formula.

2) OPTIMAL CPU FREQUENCY AT THE MEC SERVER

After decoupling irrelevant variables, the optimal solution to CPU frequency at MEC server can be solved by the following equation.

$$\begin{aligned}&\min_{f_c(t),D_s(t)} \sum_{n\in N} -C_{n,k}(t)D_{s,n}(t) + V\times\omega_{N+1}k_{ser,n}f^3_{C,n}(t)\\&s.t.\ 0 \le f_{C,n}(t) \le f_{C,n,max},\quad n\in N \qquad (30)\end{aligned}$$

Similar to the optimal solution of local CPU frequency, optimal frequency at the MEC server is also a convex problem. Follow the same steps, we can derive the optimal solution (31) to the sub question. We define $((1-\mu)C^{incop}_{n,k}(t) + \mu C^{cop}_{n,k}(t))\tau$ as $\mathbb{C}_{n,k}$.

$$f_{C,n,k}(t) = \begin{cases} \min\Big(f^{max}_{C,n,k}, \sqrt{\frac{\mathbb{C}_{n,k}}{3k_{ser,n,k}\omega_{N+K+1}VL_{C,n,k}}}\Big), & \omega_{N+K+1} > 0\\ 0, & \omega_{N+K+1} = 0\end{cases} \qquad (31)$$

In summary, the optimal solution demonstrates that the deterministic result is irrelevant with task arrived rate, thus it is appropriate for the complicated environment. According to the previous discussion, the optimal local and MEC server frequency can be obtained under a low complexity algorithm.

V. SIMULATION RESULT

In simulation, to evaluate the performance of our proposed resource management strategy based on collaborative

TABLE 1. Parameter settings.

Notation	Value
N	2
K	30
v	72km/h
τ	1 ms
ω	10MHz
N_0	-174dBm/Hz
$k_{mob,n,k}, k_{ser,n,k}$	10^{-27}
$f_{n,k,max}$	1GHz
$f_{c,n,k,max}$	2.5GHz
$L_{n,k}, L_{C,n,k}$	737.5cycles/bit
P_{max}	23dBm
Sub-carrier spacing	30KHz

MEC assisted cellular-V2X (CBSOA), we regard centralized MEC vehicular network without collaboration between MEC servers (NCNOA) as the compared method. For convenience, parameter settings are summarized in Table 1. We assume that N MEC servers and M vehicles are randomly located in the coverage of each base station installed with MEC servers. Besides, the speed of each cars equals to 72km/h. The length of time slot τ = 1ms, the system bandwidth $\omega = 10MHz$, $N_0 = -174dBm/Hz$. $f_{n,k,max} = 1GHz$and $f_{C,n,k,max} = 2.5GHz$ are the maximum CPU frequency for local vehicles and MEC servers. $k_{mob,n,k}$ and $k_{ser,n,k}$ are the effective switched capacitance and they equals to 10^{-27}. $L_{n,k}$ and $L_{C,n,k}$ are the CPU cycles needed to process 1 bit of tasks and they equals to 737.5 cycles/bit. The maximum transmission power is 23dBm and the sub-carrier spacing is 30KHz.

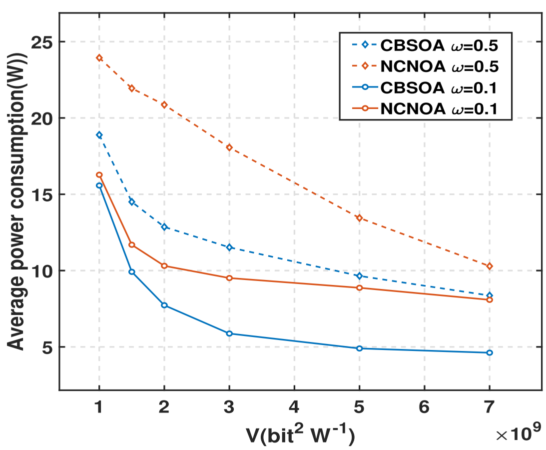


FIGURE 4. **The effect of the control parameter on the weighted mean power consumption in different scheme.**

Then in the following discussion, we will talk about the performance of these two strategies and key parameters. We first validate the theoretical result for the impact of the control parameter. The relationship between weighted mean of energy consumption and average queue length with control parameter are revealed in Fig. 4 and Fig. 5. We can easily see that the weighted mean of energy consumption is negatively correlated with the control parameter and the weighted mean of sum queue length is positively correlated with the control parameter. It verifies that there is a trade-off between weighted mean of energy consumption and average processing delay. For the balance of the system, a larger control parameter can be used in an energy consumption sensitive situation and a smaller control parameter can be used in a delay-sensitive situation. Regarding a system that uses the maximum transmit power to offload the task, our proposed optimized system has a lower weighted mean energy consumption and average queue length. In a word, some computational resources provided by MEC servers is wasted when using the maximum transmit power and can be re-allocated to other vehicles.

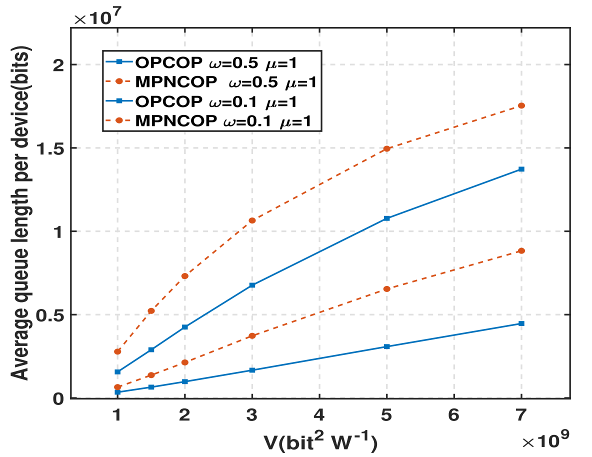


FIGURE 5. **The effect of the control parameter on the average queue length in different scheme.**

Fig. 4 and Fig. 5 also tells that when ω increases, the energy consumption and the queue length increases either. Under this circumstance, the frequencies of the CPU cores are redundant than needed and more computational resources are provided. These redundant resources can be re-allocated to other vehicles.

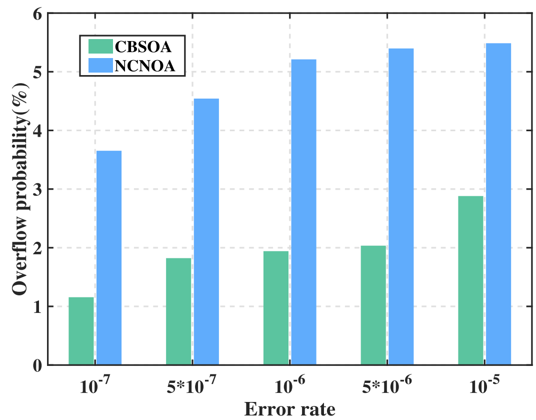


FIGURE 6. **The overflow probability of NCNOA and CBSOA with different error rate.**

Fig. 6 shows the overflow probability for two schemes. By varying the decoding error rate of the transmission, we can see that the reliability will increase with the error rate. From equation (4) we can observe that a larger decoding error rate will have a corresponding smaller transmission rate. A smaller transmission rate means that less task will be offloaded to mobile edge servers and therefore it will result in a smaller stable queue length. And then the task buffers will have less probability of overflow. In a word, a smaller overflow probability means higher reliability. Besides, the overflow probability between different schemes is evaluated. It demonstrates that the overflow probability of the collaborative MEC system is no larger than 3%, while the non-collaborative system's overflow probability is no larger than 6%. Thus, Fig. 6 shows that a collaborative MEC vehicular network with the proposed algorithm is more reliable than the non-collaborative MEC vehicular network without the proposed algorithm. As a collaborative MEC system has better processing capability, thus the task buffers have less overflow probability.

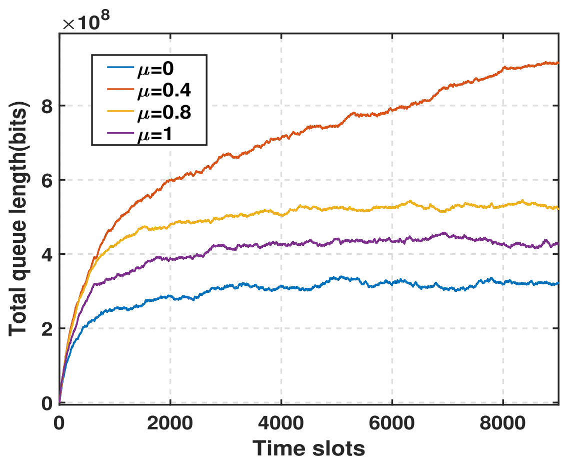


FIGURE 7. Total queue length v.s time for collaborative MEC networks with different collaborative impact factor.

Fig. 7 illustrates the amount of queue length as time goes and the impact of collaboration impact factor on queue length. As is shown in Fig 7, if it enters the collaborative mode, a collaborative impact factor will be always larger than 0 and the average queue length will always be greater than the non-collaborative system. This is because in the collaborative MEC system will receive data transmission from other base stations. The higher the collaboration impact factor becomes, the higher the processing capability is provided. In other words, the queue length will decrease with the increase in the collaborative impact factor. When μ equals to 0.4, the system's processing capability cannot satisfy the requirement. Therefore, the average queue length continues to increase as time goes, which demonstrates that the system performance cannot converge to a stable state under the system's time-constraint.

From Fig. 8 we can see the relationship between weighted mean of energy consumption and the average processing delay. The average processing delay increases while the

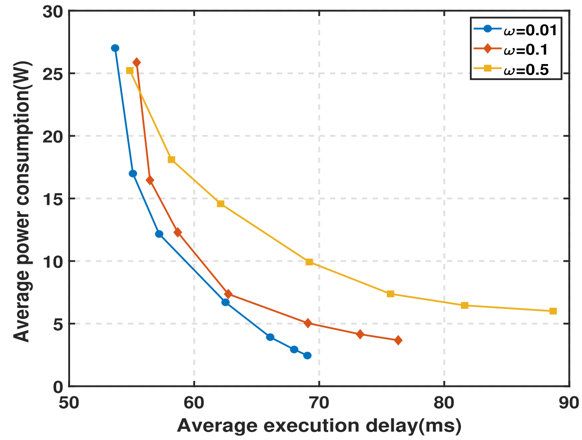


FIGURE 8. Weighted mean of energy consumption vs. Average processing delay.

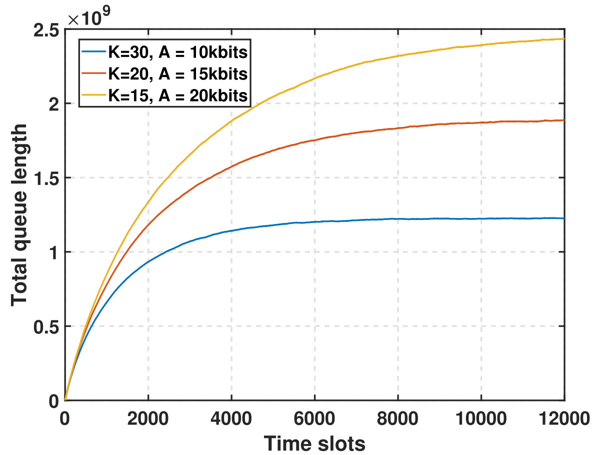


FIGURE 9. Relationship network size and average weighted total queue length.

weighted mean of energy consumption decreases. Furthermore, it shows that a proper control parameter V can be chosen properly to balance the energy consumption and processing delay. Besides, given specific processing delay, the weighted mean of energy consumption increases with the weight. It also confirms that the energy consumption can be adjusted by weight factors. Furthermore, there is a power delay trade-off in this system, which is consistent with our assumption.

In real circumstance, there will be a different number of vehicles in each cell. Thus, we discuss the system stability for different network size. Fig. 9 discuss the impact of network size on the convergence time using the proposed scheme. By tracking the average total queue length of vehicles with different numbers of vehicles, the task arrival rate is kept unchanged simultaneously. The task buffer's convergence time and stability are presented in Fig. 9. It presents that the queue length will increase at the beginning, but it will be stabilized finally. A larger A(tasks arrived rate) will result in larger convergence time, while the queue length is still able to be stabilized after several time slots for larger network size.

Therefore, the system is applicable to a reasonable number of vehicles for mobile edge computing.

VI. CONCLUSION

This paper innovatively envisions a joint computation resource allocation and URLLC resource allocation strategy for collaborative MEC assisted cellular-V2X networks, where the tasks can be exchanged feasibly between MEC base stations. A non-cooperative game and bipartite graph are introduced to reduce the inter-cell interference and decide the channel allocation, which maximize the throughput in URLLC V2X communication. Then an online Lyapunov optimization method is proposed to solve computation resource allocation to get a trade-off between the average weighted power consumption and delay. The results demonstrated that our scheme greatly reduces the energy consumption and processing delay of the system and there exists a power-delay tradeoff in the system. Nevertheless, our proposed can get better overflow probability, which means that it is more reliable than centralized MEC assisted V2X. The road environment in this paper is based on the random distribution of vehicles. However, in the real world, it is obviously impossible for vehicles to be completely distributed randomly on the road. The actual road scene is relatively complex. For example, cars on a one-way street drive in the same direction, while cars on a two-way lane drive in opposite direction. There are also intersections and more complex vehicle environments. Therefore, it is necessary to further study the complex road scenes in reality.

REFERENCES

[1] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, ''A survey on mobile edge computing: The communication perspective,'' *IEEE Commun. Surveys Tutr.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.

[2] K. Zhang, Y. Mao, S. Leng, Y. He, and Y. Zhang, ''Mobile-edge computing for vehicular networks: A promising network paradigm with predictive offloading,'' *IEEE Veh. Technol. Mag.*, vol. 12, no. 2, pp. 36–44, Jun. 2017.

[3] I. Sorkhoh, D. Ebrahimi, R. Atallah, and C. Assi, ''Workload scheduling in vehicular networks with edge cloud capabilities,'' *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8472–8486, Sep. 2019.

[4] C. Yang, Y. Liu, X. Chen, W. Zhong, and S. Xie, ''Efficient mobility-aware task offloading for vehicular edge computing networks,'' *IEEE Access*, vol. 7, pp. 26652–26664, 2019.

[5] X. Sun, J. Zhao, X. Ma, and Q. Li, ''Enhancing the user experience in vehicular edge computing networks: An adaptive resource allocation approach,'' *IEEE Access*, vol. 7, pp. 161074–161087, 2019.

[6] J. Du, F. R. Yu, X. Chu, J. Feng, and G. Lu, ''Computation offloading and resource allocation in vehicular networks based on dual-side cost minimization,'' *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1079–1092, Feb. 2019.

[7] H. Guo and J. Liu, ''Collaborative computation offloading for multiaccess edge computing over fiber–wireless networks,'' *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4514–4526, May 2018.

[8] J. Ren, G. Yu, Y. He, and G. Y. Li, ''Collaborative cloud and edge computing for latency minimization,'' *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 5031–5044, May 2019.

[9] J. Zhao, Q. Li, Y. Gong, and K. Zhang, ''Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks,'' *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7944–7956, Aug. 2019.

[10] K. Cheng, Y. Teng, W. Sun, A. Liu, and X. Wang, ''Energy-efficient joint offloading and wireless resource allocation strategy in multi-MEC server systems,'' in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[11] A. K. Datta and R. Patel, ''CPU scheduling for power/energy management on multicore processors using cache miss and context switch data,'' *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 5, pp. 1190–1199, May 2014.

[12] Y. Polyanskiy, H. V. Poor, and S. Verdu, ''Channel coding rate in the finite blocklength regime,'' *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[13] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, ''Quasi-static multiple-antenna fading channels at finite blocklength,'' *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.

[14] G. Durisi, T. Koch, and P. Popovski, ''Toward massive, ultrareliable, and low-latency wireless communication with short packets,'' *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[15] C. She and C. Yang, ''Available range of different transmission modes for ultra-reliable and low-latency communications,'' in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–5.

[16] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, ''Stochastic joint radio and computational resource management for multi-user mobile-edge computing systems,'' *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5994–6009, Sep. 2017.

[17] J. Tang and X. Zhang, ''Power-delay tradeoff over wireless networks,'' in *Proc. Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2008, pp. 1–12.

[18] M. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. San Mateo, CA, USA: Morgan & Claypool, 2010. [Online]. Available: https://ieeexplore.ieee.org/document/6813406

[19] C.-F. Liu, M. Bennis, and H. V. Poor, ''Latency and reliability-aware task offloading and resource allocation for mobile edge computing,'' in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2017, pp. 1–7.

[20] M. Bennis, M. Debbah, and H. V. Poor, ''Ultrareliable and low-latency wireless communication: Tail, risk, and scale,'' *Proc. IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct. 2018.

[21] Q.-V. Pham, L. B. Le, S.-H. Chung, and W.-J. Hwang, ''Mobile edge computing with wireless backhaul: Joint task offloading and resource allocation,'' *IEEE Access*, vol. 7, pp. 16444–16459, 2019.

[22] H. Kwon and B. Lee, ''Distributed resource allocation through noncooperative game approach in multi-cell OFDMA systems,'' in *Proc. IEEE Int. Conf. Commun.*, vol. 9, Jun. 2006, pp. 4345–4350.

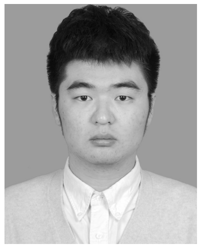
LEI FENG received the B.Eng. and Ph.D. degrees in communication and information systems from the Beijing University of Posts and Telecommunications (BUPT), in 2009 and 2015, respectively. He is currently an Associate Professor with the State Key Laboratory of Networking and Switching Technology, BUPT. His research interests are resources management in wireless networks and smart grid.

WENJING LI is currently a Professor with BUPT and serves as the Director of the Key Laboratory of Network Management Research Center. Meanwhile, she is the Leader of TC7/WG1 with the China Communications Standards Association (CCSA). Her research interests are wireless network management and automatic healing in SONs.

YINGXIN LIN received the B.Eng. degree from the Queen Mary University of London and the B.Admin. degree from the Beijing University of Posts and Telecommunications (BUPT), in 2018, both in E-communication and management, where she is currently pursuing the M.Eng. degree with Computer Science and Technology. Her research interests are resources management in mobile edge computing systems.

LIANG ZHU received the B.Eng. degree from the Beijing University of Posts and Telecommunications (BUPT), in 2018, where he is currently pursuing the M.Eng. degree with Electronics and Communications Engineering. His research interests include 5G URLLC, the Internet of Things, and vehicular networks.

SHAOYONG GUO received the B.S. degree in information and computing science from Hebei University, in 2008, and the Ph.D. degree in computer science and technology from the Beijing University of Posts and Telecommunications, in 2013. He drafts an International standard as the first accomplisher and participates in three other International/industry standards. His research interests include blockchain application technology, mobile edge computing, and the Internet of Things (IoT) in energy Internet.



ZERUI ZHEN received the B.Eng. and B.C. degrees in electromagnetic field and radio technology from Xidian University, in 2018. He is currently pursuing the M.Eng. degree in information and communication engineering with the Beijing University of Posts and Telecommunications (BUPT). His research interests are resources management in wireless networks and smart grid.