Joint Computation Offloading and Wireless Resource Allocation in Vehicular Edge Computing Networks

Jiao Zhang[1(✉)], Zhanjun Liu[1], Bowen Gu[2], Chengchao Liang[1], and Qianbin Chen[1]

[1] School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China
zjj.0502@foxmail.com

[2] Faculty of Information Technology, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, China

Abstract. In vehicular edge computing (VEC) networks, a promising strategy for in-vehicle user equipments (UEs) with limited battery capacity is to offload data-intensive or/and latency-sensitive services to VEC servers via vehicle-to-infrastructure (V2I) links to reduce their energy consumption (EC). However, limited power supply, inadequate computation capability, and dynamic task latency make it extremely challenging to implement. In this work, we focus on designing a system-centric EC scheme in a VEC network. In particular, a joint computation offloading and wireless resource allocation problem is formulated to minimize the system EC by optimizing power control, offloading decision as well as subcarrier allocation while taking into account the dynamic and fickle time delay of each UE's task. In light of the intractability of the problem, we propose an effective block coordinate descent (BCD)-based algorithm with greedy search to find a high-quality sub-optimal solution. Simulation results illustrate that the proposed algorithm has better energy savings than the benchmarks.

Keywords: Vehicular network · Resource allocation · Energy consumption · Edge computing

1 Introduction

The recent advancement of vehicular networks has spurred various new applications in the domains of autonomous driving, video-aided real-time navigation, video streaming, and augmented reality [1,2]. However, it also means higher requirements for vehicular networks to handle such latency-sensitive and data-intensive services [3]. Although the processing capacity of the vehicle's central processing unit (CPU) becomes stronger, it is still unable to handle enormous applications in a short time, and the quality of service (QoS) is hard to be

H. Gao et al. (Eds.): ChinaCom 2021, LNICST 433, pp. 377–391, 2022.
https://doi.org/10.1007/978-3-030-99200-2_29

effectively guaranteed, which has become one of the bottlenecks for the further implementation of the vehicular networks [4].

To tackle this challenge, vehicular edge computing (VEC), as a promising technology, has been proposed. Specifically, the computation tasks generated by applications can be allowed to offload to the edges of the networks, so that long-distance transmission and excessive network hops can be eliminated [5]. In addition, the computational response time is reduced to better ensure the QoS of UEs. As a result, VEC-based computation offloading has attracted a large amount of attentions and has been extensively investigated from different perspectives, e.g., task computation delay minimization [6], vehicular terminals and servers cost minimization [7] and system utility maximization [8].

On the other hand, VEC also provides an opportunistic solution for energy-limited in-vehicle UEs to reduce their energy consumption (EC) and extend their lifetime [9]. Specifically, traditional solutions are tending to process all computing tasks locally by UEs, which leads to tremendous EC and ephemeral lifetime for UE. Thanks to the flourish of VEC, users' tasks can be offloaded to VEC servers for computation with the help of vehicle-to-infrastructure (V2I) links, which not only improves computational efficiency but also compensates for the inherent weaknesses in the traditional approach.

Resource allocation, as an essential technology in VEC networks, can achieve spectrum management, power allocation, and resource scheduling, etc. Such as the authors in [10] maximized the sum computation rate of all UEs via the binary offloading mode in wireless-powered mobile-edge computing networks. However, binary offloading may no longer be suitable for data-ultra-intensive computing tasks in practical. To this end, the authors in [11] minimized the weighted sum computation delay by invoking partial offloading for D2D-enabled mobile edge computing networks. We notice that [10] and [11] mainly focus on computation rate and latency, ignoring system energy consumption, which may be contrary to the green communication initiative. To deal with it, EC has been studied in [12,13]. In [12], the authors studied the resource-management policy to minimize the EC of all tasks for a mobile-edge computation offloading system under time constraints for mobile UEs. In [13], the authors jointly optimizing the transmit beamforming, the processing unit frequencies, and the offloading decision for energy minimization in wireless-powered multi-user mobile-edge computing systems. Although the aforementioned works [10–13] investigated resource allocation for UEs based on static cellular networks, the reached conclusions are not applicable to vehicular networks, due to the highly dynamic and unstable connectivity. To solve this issue, the authors in [14] minimized the EC of in-vehicle UEs and proposed an effective workload offloading scheme in VEC networks. However, the power allocation has not been involved, which may lead to biased resource scheduling and mismatched EC.

Motivated by the mentioned facts, in this paper, we investigate a joint computation offloading and wireless resource allocation problem for VEC networks, where energy savings of UEs is evaluated. And the contributions of this paper are summarized as follows.

- Our goal is to minimize the total EC of UEs, by jointly optimizing offloading decision, power allocation, and subcarrier selection, where transmit power threshold, the latency constraint, and computing ability of UEs are taking into account. The resulting problem is non-convex and challenging to solve.
- To tackle its non-convexity, we decompose the original problem into three subproblems based on block coordinate descent (BCD) method and solved them respectively. In specific, the interior-point method is used to address the subproblem of offloading decision. And for the subproblem of transmit power, we transform it into an equivalent convex optimization problem via Dinkelbach's method and successive convex approximation (SCA) method and then solve it by using Lagrangian dual decomposition theory. Besides, the greedy search method is employed to solve the subproblem of subcarrier selection.
- Simulation results demonstrate that the proposed algorithm not only has a close-to-optimal solution but also is better than the benchmarks in energy saving.

The rest of this paper is organized as follows. In Sect. 2, we introduce the system model. In Sect. 3, we formulate the EC minimization problem and propose a resource allocation algorithm to solve this problem. After that, simulation results are provided in Sect. 4. Finally, the paper is concluded in Sect. 5.

2 System Model

2.1 The Overall System Model

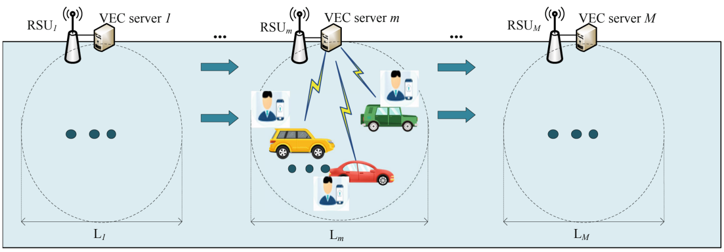

Fig. 1. System model

The task offloading framework for vehicular networks is shown in Fig. 1. There is a unidirectional road, where M RSUs are located along the road. Each RSU is equipped with a VEC server for remote edge computing. The road is divided into M segments based on the coverage areas of the M RSUs, represented as $\{L_1, L_2, ..., L_M\}$. A set of vehicles $\mathcal{K} = \{1, 2, ..., K\}$ in the m-th segment traveling towards the same direction. Define $U_{k,m}$ and $V_{k,m}$ as the UE k and the vehicle k within the RSU m coverage, respectively. We assume that there is only one task denoted by $\phi_{k,m} = \{D_{k,m}, C_{k,m}, \tau_{k,m}\}$ needs to be executed within a period of time for UE $U_{k,m}$, where $D_{k,m}$ represents the data size of the task, $C_{k,m}$ is the required computing ability for processing the task, and $\tau_{k,m}$ is the maximum tolerable latency of the task. In this paper, similar to [11], we assume that the task-input bits are bit-wise independent and can be arbitrarily divided into different groups. Thus the system support data offloading and local computing simultaneously.

2.2 The Transmission Model

The OFDMA is introduced to our system thus the co-channel interference among different vehicles can be ignored. The total bandwidth is divided into N subcarriers, denoted as $n \in N = \{1, 2, ...N\}$, where $N \geq K$. To ensure the quality of service (QoS) and reduce the EC of UE $U_{k,m}$, the task of UE $U_{k,m}$ can be offloaded to the VEC Server m through vehicle $V_{k,m}$. Specifically, the data transmission is from UE $U_{k,m}$ to RSU m in a two-hop. (i.e., the data is firstly sent from UE $U_{k,m}$ to the vehicle $V_{k,m}$, then is forwarded from vehicle $V_{k,m}$ to RSU m). For the UE $U_{k,m}$, the signal to noise ratio (SNR) of the first-hop link and the second-hop link are calculated as

$$\gamma_{k,m}^{n,1} = \frac{p_{k,m}^n h_{k,m}^n}{\sigma^2} \tag{1}$$

$$\gamma_{k,m}^{n,2} = \frac{P_{k,m}^n g_{k,m}^n}{\sigma^2} \tag{2}$$

where $p_{k,m}^n$ and $P_{k,m}^n$ are the transmit power over the subcarrier n of UE $U_{k,m}$ and vehicle $V_{k,m}$, respectively. $h_{k,m}^n$ and $g_{k,m}^n$ are the channel gain between UE $U_{k,m}$ and vehicle $V_{k,m}$ and between vehicle $V_{k,m}$ and RSU m, respectively. σ^2 denotes the noise power.

Based on the full-duplex amplification and forwarding (AF) protocol [15], the effective SNR of the two-hop link is calculated as

$$\gamma_{k,m}^{n} = \frac{\gamma_{k,m}^{n,1}\gamma_{k,m}^{n,2}}{\gamma_{k,m}^{n,1} + \gamma_{k,m}^{n,2} + 1} \tag{3}$$

Then, the corresponding data transmission rate can be obtained as

$$R_{k,m}^n = \alpha_{k,m}^n B \log_2(1 + \gamma_{k,m}^n) \tag{4}$$

where B denotes the subcarrier bandwidth, and $\alpha_{k,m}^n \in \{0, 1\}$ is the subcarrier selection indicator with $\alpha_{k,m}^n = 1$ implying the data of UE $U_{k,m}$ is transmitted over the subcarrier n and $\alpha_{k,m}^n = 0$ otherwise.

Defining $\lambda_{k,m}$ as the offloading ratio of task $\phi_{k,m}$, the corresponding offloading time is expressed as

$$T_{k,m}^{\text{trans}} = \frac{\lambda_{k,m} D_{k,m}}{\sum_{n=1}^{N} R_{k,m}^n} \tag{5}$$

The EC of UE $U_{k,m}$ transmitting the data to the vehicle m is expressed as

$$E_{k,m}^{\text{trans}} = \sum_{n=1}^{N} \alpha_{k,m}^{n} p_{k,m}^{n} T_{k,m}^{\text{trans}} = \frac{\sum_{n=1}^{N} \alpha_{k,m}^{n} p_{k,m}^{n} \lambda_{k,m} D_{k,m}}{\sum_{n=1}^{N} R_{k,m}^{n}} \tag{6}$$

2.3 The Computation Model

In what follows, the specific process of task computation with partial offloading mode is introduced, which includes local computing and VEC.

1) Local Computing Model

The UE $U_{k,m}$ executes $(1-\lambda_{k,m})C_{k,m}$ parts of its computation task $\phi_{k,m}$ locally, and the local computing time $T_{k,m}^{\text{loc}}$ can be obtained as

$$T_{k,m}^{\text{loc}} = \frac{(1-\lambda_{k,m})\, C_{k,m}}{f_{k,m}} \tag{7}$$

where $f_{k,m}$ is the CPU cycle frequency of UE $U_{k,m}$.

Motivated by the observation on [16], the power consumption for local computing can be modeled as $P_{k,m}^{\text{loc}} = \delta (f_{k,m})^3$, where δ is the coefficient depending on the chip architecture. Based on (7), the EC of local computation is expressed as

$$E_{k,m}^{\text{loc}} = P_{k,m}^{\text{loc}} T_{k,m}^{\text{loc}} = \delta\,(1-\lambda_{k,m})\, C_{k,m} f_{k,m}^{2} \tag{8}$$

The overall EC of UE $U_{k,m}$, which contains the energy consumed by local computing and data offloading, is expressed as

$$\begin{aligned} E_{k,m} &= E_{k,m}^{\text{trans}} + E_{k,m}^{\text{loc}} \\ &= \sum_{n=1}^{N} \frac{\alpha_{k,m}^{n} p_{k,m}^{n} \lambda_{k,m} D_{k,m}}{R_{k,m}^{n}} + \delta\,(1-\lambda_{k,m})\, C_{k,m} f_{k,m}^{2} \end{aligned} \tag{9}$$

2) The VEC Model

The UE $U_{k,m}$ offloads the rest of computation task $\lambda_{k,m}C_{k,m}$ to the VEC server m, thus, the processing time of VEC server m can be calculated as

$$T_{k,m}^{\text{comp}} = \frac{\lambda_{k,m} C_{k,m}}{f_{k,m}^{\text{VEC}}} \tag{10}$$

where $f_{k,m}^{\text{VEC}}$ is the CPU-cycle frequency of VEC Server m when executing the edge computing task $\phi_{k,m}$ [17].

According to (5) and (10), the total time for computation offloading is $T_{k,m}^{\text{vec}} = T_{k,m}^{\text{trans}} + T_{k,m}^{\text{comp}}$. To ensure the QoS of UE $U_{k,m}$, we have $T_{k,m}^{\text{vec}} \le \tau_{k,m}$. However, in vehicular networks, the effective execution time for one task may be different under different vehicle velocities. Specifically, the practical execution time $T_{k,m}^{\text{pra}}$ relies on the dwell time $\tau_{k,m}^{\text{o}}$ and the maximum tolerance delay $\tau_{k,m}$, such as

$$T_{k,m}^{\text{pra}} = \min\left\{\tau_{k,m}, \tau_{k,m}^{\text{o}}\right\} \tag{11}$$

where $\tau_{k,m}^{\text{o}} = \frac{d_{k,m}}{v_{k,m}}$, $d_{k,m}$ is the distance between vehicle $V_{k,m}$ and the edge of RSU m in the vehicle heading direction, $v_{k,m}$ is the velocity of vehicle $V_{k,m}$.

3 Problem Formulation and Solution

Our goal is to minimize the total EC of UEs by jointly optimizing the transmit power, subcarrier selection, offloading ratio, and computing ability. Mathematically, the optimization problem is formulated as

$$\begin{aligned}
&\min_{p_{k,m}^n, \alpha_{k,m}^n, \lambda_{k,m}, f_{k,m}} \sum_{k=1}^{K} E_{k,m} \\
&s.t.\ C_1 : \sum_{k=1}^{K} \alpha_{k,m}^n \le 1,\ \alpha_{k,m}^n \in \{0,1\},\ \forall n \in \mathcal{N} \\
&\quad C_2 : \sum_{n=1}^{N} \alpha_{k,m}^n p_{k,m}^n \le P_{\max},\ \forall k \in \mathcal{K} \\
&\quad C_3 : f_{k,m} \le f_{k,m}^{\max}\ , \forall k, m \\
&\quad C_4 :\ T_{k,m}^{\text{loc}} \le \tau_{k,m}, \forall k, m \\
&\quad C_5 :\ T_{k,m}^{\text{vec}} \le T_{k,m}^{\text{pra}}, \forall k, m \\
&\quad C_6 :\ \lambda_{k,m} \in [0,1],\ \forall k \in \mathcal{K}
\end{aligned} \tag{12}$$

where C_1 denotes that each subcarrier is only allocated to one UE. C_2 represents the maximum transmit power constraint. C_3 is computing ability constraint. C_4 and C_5 ensure the latency requirements of local computing and remote executions. And C_6 enforces the offloaded task ratio of UE cannot exceed 1.

3.1 The Local Rescource Allocation

According to (7) and C3, we have $f_{k,m} \in \left[\frac{(1-\lambda_{k,m})C_{k,m}}{\tau_{k,m}}, f_{k,m}^{\max}\right]$. From (8), it is evident that the EC for local computing is monotonically increasing with $f_{k,m}$. To reduce the EC of local computing as much as possible, $f_{k,m}$ should take the smallest value within the feasible range. As a result, the optimal computing ability of UE $U_{k,m}$ is

$$f_{k,m}^{*} = \frac{(1-\lambda_{k,m})\, C_{k,m}}{\tau_{k,m}} \tag{13}$$

Correspondingly, the optimal local computing time is expressed as

$$T_{k,m}^{\text{loc},*} = \tau_{k,m} \tag{14}$$

Substituting $T_{k,m}^{\mathrm{loc},*}$ and $f_{k,m}^{*}$ into (12), as a result, problem (12) can be rewritten as

$$\min_{p_{k,m}^{n},\alpha_{k,m}^{n},\lambda_{k,m}} \overline{E}_{k,m}=\sum_{k=1}^{K}\left(\frac{\sum\limits_{n=1}^{N}\alpha_{k,m}^{n}p_{k,m}^{n}\lambda_{k,m}D_{k,m}}{\sum\limits_{n=1}^{N}R_{k,m}^{n}}+\frac{\delta(C_{k,m})^{3}}{{\tau_{k,m}}^{2}}(1-\lambda_{k,m})^{3}\right)$$

$$s.t.\ C1, C2, C5, C6 \tag{15}$$

However, problem (15) is non-convex due to the coupled variables and the binary variable. There are no standard methods to address such problems optimally. Therefore, Motivated by the idea of the BCD method [18], we decompose problem (15) into three subproblems, namely, offloading ratio, transmit power, and subcarrier selection.

3.2 The Subproblem of Offloading Ratio

With the given $\alpha_{k,m}^{n}, p_{k,m}^{n}$, the subproblem of offloading ratio is expressed as

$$\begin{aligned}&\min_{\lambda_{k,m}}\sum_{k=1}^{K}\left(\frac{\sum\limits_{n=1}^{N}\alpha_{k,m}^{n}p_{k,m}^{n}\lambda_{k,m}D_{k,m}}{\sum\limits_{n=1}^{N}R_{k,m}^{n}}+\frac{\delta(C_{k,m})^{3}}{{\tau_{k,m}}^{2}}(1-\lambda_{k,m})^{3}\right)\\&s.t.\ C6, C5:\lambda_{k,m}\left(\frac{D_{k,m}}{\sum\limits_{n=1}^{N}R_{k,m}^{n}}+\frac{C_{k,m}}{f_{k,m}^{\mathrm{VEC}}}\right)\leq T_{k,m}^{\mathrm{pra}}\end{aligned}\tag{16}$$

Observing problem (16), we find that the objective function and constraints are linear about $\lambda_{k,m}$, it is a convex problem. Thus, the interior-point method can be employed to solve it.

3.3 The Subproblem of Transmit Power

With the given $\lambda_{k,m}$ and $\alpha_{k,m}^{n}$, the subproblem of transmit power is

$$\begin{aligned}&\min_{p_{k,m}^{n}}\sum_{k=1}^{K}\sum_{n=1}^{N}\left(\frac{\alpha_{k,m}^{n}p_{k,m}^{n}\lambda_{k,m}D_{k,m}}{R_{k,m}^{n}}\right)\\&s.t.\ C2, \bar{C}5:\sum_{n=1}^{N}R_{k,m}^{n}\geq\frac{D_{k,m}\lambda_{k,m}f_{k,m}^{\mathrm{VEC}}}{T_{k,m}^{\mathrm{pra}}f_{k,m}^{\mathrm{VEC}}-C_{k,m}\lambda_{k,m}}\end{aligned}\tag{17}$$

However, problem (17) is a non-convex problem since the objective function is non-smooth, which is challenging to solve. To this end, Dinkelbach's method

[19] is used to transform the objective function with fractional form into one with non-fractional form, e.g.,

$$\begin{aligned} &\min_{p_{k,m}^n} \sum_{k=1}^{K} \sum_{n=1}^{N} \left(\alpha_{k,m}^n p_{k,m}^n \lambda_{k,m} D_{k,m} - A R_{k,m}^n \right) \\ &s.t.\ C2, \bar{C}5 \end{aligned} \tag{18}$$

where $A \geq 0$. However, there are still some coupled variables in $R_{k,m}^n$, which makes problem (18) difficult to solve. Then, the SCA method [20] is employed. We have the following inequality

$$\log_2(1+\gamma_{k,m}^n) \geq a_{k,m}^n \log_2(\gamma_{k,m}^n) + b_{k,m}^n \tag{19}$$

where $a_{k,m}^n$ and $b_{k,m}^n$ are defined as

$$\begin{cases} a_{k,m}^n = \frac{\bar{\gamma}_{k,m}^n}{1+\bar{\gamma}_{k,m}^n} \\ b_{k,m}^n = \log_2(1+\bar{\gamma}_{k,m}^n) - a_{k,m}^n \log_2(\bar{\gamma}_{k,m}^n) \end{cases} \tag{20}$$

The lower bound is tight when $\gamma_{k,m}^n = \bar{\gamma}_{k,m}^n$, we define $\bar{\gamma}_{k,m}^n$ as the SINR for UE $U_{k,m}$ from the last iteration. We use the initialized value to calculate $\bar{\gamma}_{k,m}^n$ for the first iteration. As a result, the data transmission rate with the lower bound is

$$\bar{R}_{k,m}^n = B\alpha_{k,m}^n \left(a_{k,m}^n \log_2(\bar{\gamma}_{k,m}^n) + b_{k,m}^n \right) \tag{21}$$

To make the problem tractable, we define $\mathbf{p} = 2^{\mathbf{q}}$, where $\mathbf{q} = \left\{ q_{k,m}^n \right\}$. Substituting (3) into (21), the lower bound of $\bar{R}_{k,m}^n$ can be expressed as

$$\bar{R}_{k,m}^n = B\alpha_{k,m}^n \left[a_{k,m}^n \left(q_{k,m}^n + C_{k,m}^n - \log_2(2^{q_{k,m}^n} \frac{h_{k,m}^n}{\sigma^2} + \frac{P_{k,m}^n g_{k,m}^n}{\sigma^2} + 1) \right) + b_{k,m}^n \right] \tag{22}$$

where $C_{k,m}^n = \log_2(\frac{h_{k,m}^n}{\sigma^2}) + \log_2(\frac{P_{k,m}^n g_{k,m}^n}{\sigma^2})$.

Based on (21) and (22), problem (18) can be transformed as

$$\begin{aligned} &\min_{\mathrm{q}_{k,m}^n} \sum_{k=1}^{K} \sum_{n=1}^{N} \left(\alpha_{k,m}^n 2^{q_{k,m}^n} \lambda_{k,m} D_{k,m} - \mathrm{A} \bar{R}_{k,m}^n \right) \\ &s.t.\ C2, \bar{C}6 : \sum_{n=1}^{N} \bar{R}_{k,m}^n \geq \frac{D_{k,m} \lambda_{k,m} f_{k,m}^{\mathrm{VEC}}}{T_{k,m}^{\mathrm{pra}} f_{k,m}^{\mathrm{VEC}} - C_{k,m} \lambda_{k,m}} \end{aligned} \tag{23}$$

Problem (23) is a convex optimization problem, which can be solved by using the Lagrangian dual decomposition theory. The Lagrangian function of problem (23) is

$$L_m(\mathrm{q}_{k,m}^n,\mu_{k,m}^n,\varphi_{k,m}^n)=\sum_{k=1}^{K}\sum_{n=1}^{N}(\alpha_{k,m}^n 2^{q_{k,m}^n}\lambda_{k,m}D_{k,m}-\mathrm{A}\bar{R}_{k,m}^n)$$
$$+\sum_{k=1}^{K}\varphi_{k,m}^n(\frac{D_{k,m}\lambda_{k,m}f_{k,m}^{\mathrm{VEC}}}{T_{k,m}^{\mathrm{pra}}f_{k,m}^{\mathrm{VEC}}-C_{k,m}\lambda_{k,m}}-\sum_{n=1}^{N}\bar{R}_{k,m}^n) \quad (24)$$
$$+\sum_{k=1}^{K}\mu_{k,m}^n(\sum_{n=1}^{N}\alpha_{k,m}^n 2^{q_{k,m}^n}-P_{\max})$$

where $\varphi_{k,m}^n$ and $\mu_{k,m}^n$ are the non-negative Lagrange multipliers. For ease of handling, (24) can be rewritten as

$$L_m\left(q_{k,m}^n,\mu_{k,m}^n,\varphi_{k,m}^n\right)=L_{k,m}^n\left(q_{k,m}^n,\mu_{k,m}^n,\varphi_{k,m}^n\right)$$
$$-\sum_{k=1}^{K}\mu_{k,m}^n P_{\max}+\sum_{k=1}^{K}\varphi_{k,m}^n\frac{D_{k,m}\lambda_{k,m}f_{k,m}^{\mathrm{VEC}}}{T_{k,m}^{\mathrm{pra}}f_{k,m}^{\mathrm{VEC}}-C_{k,m}\lambda_{k,m}} \quad (25)$$

where

$$L_{k,m}^n\left(q_{k,m}^n,\mu_{k,m}^n,\varphi_{k,m}^n\right)=\alpha_{k,m}^n 2^{q_{k,m}^n}\lambda_{k,m}D_{k,m}-q\bar{R}_{k,m}^n$$
$$+\alpha_{k,m}^n 2^{q_{k,m}^n}\mu_{k,m}^n-\bar{R}_{k,m}^n\varphi_{k,m}^n \quad (26)$$

The dual problem of (23) is

$$\max_{\mu_{k,m}^n,\varphi_{k,m}^n} D_m\left(\mu_{k,m}^n,\varphi_{k,m}^n\right)$$
$$s.t.\ \mu_{k,m}^n\geq 0,\varphi_{k,m}^n\geq 0 \quad (27)$$

where

$$D_m\left(\mu_{k,m}^n,\varphi_{k,m}^n\right)=\min_{q_{k,m}^n,\mu_{k,m}^n,\varphi_{k,m}^n} L_m\left(q_{k,m}^n,\mu_{k,m}^n,\varphi_{k,m}^n\right) \quad (28)$$

According to KKT condition [21], we can derive the optimal transmit power of UE $U_{k,m}^n$, i.e.,

$$p_{k,m}^{n,*}=\left[\frac{-\psi_{k,m}^n+\sqrt{\left(\psi_{k,m}^n\right)^2-4h_{k,m}^n\xi_{k,m}^n}}{2h_{k,m}^n}\right]^+ \quad (29)$$

where $\psi_{k,m}^n = P_{k,m}^n g_{k,m}^n + \sigma^2$, $[\cdot]^+ = \max(0,\cdot)$, and $\xi_{k,m}^n = \frac{a_{k,m}^n\left(A+\varphi_{k,m}^n\right)\left(P_{k,m}^n g_{k,m}^n+\sigma^2\right)}{\ln 2\left(\alpha_{k,m}^n\lambda_{k,m}D_{k,m}+\alpha_{k,m}^n\mu_{k,m}^n\right)}$.

By applying the subgradient method [22], the Lagrange multipliers can be updated as

$$\mu_{k,m}^n(l+1)=\left[\mu_{k,m}^n(l)-\Delta_{\mu_{k,m}^n}\left(P_{\max}-\sum_{n=1}^{N}\alpha_{k,m}^n 2^{q_{k,m}^n}\right)\right]^+ \quad (30)$$

$$\varphi_{k,m}^{n}\left(l+1\right)=\left[\varphi_{k,m}^{n}\left(l\right)-\varDelta_{\varphi_{k,m}^{n}}\left(\sum_{n=1}^{N}\bar{R}_{k,m}^{n}-\frac{D_{k,m}\lambda_{k,m}f_{k,m}^{\mathrm{VEC}}}{T_{k,m}^{\mathrm{pra}}f_{k,m}^{\mathrm{VEC}}-C_{k,m}\lambda_{k,m}}\right)\right]^{+} \tag{31}$$

where l is the iteration number, $\varDelta_{\mu_{k,m}^{n}}$ and $\varDelta_{\varphi_{k,m}^{n}}$ are the positive gradient steps.

Algorithm 1. A BCD-based resource allocation algorithm

Initialize: K, N, $D_{k,m}$, $C_{k,m}$, $\tau_{k,m}$, $P_{k,m}$, N_0, B, $h_{k,m}^{n}$, $g_{k,m}^{n}$, δ, $v_{k,m}$, $d_{k,m}$, $f_{k,m}^{\mathrm{VEC}}$;
Output:
The task offloading ratio $\lambda_{k,m}^{*}$, the transmission power $p_{k,m}^{n,*}$
and the subcarrier selection decision $\alpha_{k,m}^{n,*}$
1. **for** $k=1$ to $k=K$ **do**
2. Initialize $E_G=0$.
3. **for** $n=1$ to $n=N$ **do**
4. Set $\alpha_{k,m}^{n}=1$
5. **for** $l=1$ to $l=L$ **do**
6. Set $E_c(l)=0$
7. With the fixed $p_{k,m}^{n}(l)$, solve the problem (16) via the interior-point method and obtain the solution $\lambda_{k,m}(l+1)$.
8. With the fixed $\lambda_{k,m}(l+1)$, solve optimization problem (23) via CVX tools to obtain $p_{k,m}^{n}(l+1)$.
9. Calculate the total EC of UEs $E_c(l+1)$.
10. **if** $|E_c(l+1)-E_c(l)|>\zeta_{k,m}$, **then**
11. update $l=l+1$ and return step 7;
12. **else**
13. update $p_{k,m}^{n,*}=p_{k,m}^{n}(l)$, $\lambda_{k,m}^{*}=\lambda_{k,m}(l)$, $\mathrm{E}_c^{*}=\mathrm{E}_c(l)$;
14. **end if**
15. **end for**
16. **if** $\mathrm{E}_c^{*}\leq E_G$ **then**
17. set $\alpha_{k,m}^{n}=0$
18. **else**
19. update $E_G=\mathrm{E}_c^{*}$, $p_{\mathrm{k},m}^{n}=p_{k,m}^{n,*}$, $\lambda_{k,m}=\lambda_{k,m}^{*}$;
20. **end if**
21. **end for**
22. **end for**

Similarly, based on the given $\lambda_{k,m}$ and $p_{k,m}$, the subcarrier selection decision can be obtained. Specifically, owing to the $\alpha_{k,m}^{n}$ is a binary variable, the optimization problem is difficult to solve. Motivated by [23], the greedy algorithm is employed to tackle this problem. The detail of the proposed algorithm is shown in Algorithm 1.

4 Simulation Results

In this section, simulation results are provided to verify the effectiveness of the proposed algorithm. We assume that there is a unidirectional road, where four RSUs are randomly located along a 2000-m road. There are 20 vehicles driving on the road at a speed of 50 km/h–90 km/h within the coverage of the same RSU. The path-loss model is $PL = 128.1 + 37.6\log_{10} d$, where $d \in (0.05, 0.1)$ km is the distance between one receiver and one transmitter. The other parameters used in our simulations are summarized in Table 1. To better show the effectiveness of the proposed algorithm, we compare two other algorithms, namely, the local computing algorithm and the full offloading algorithm.

Table 1. Parameters

Parameter	Value
Number of UEs	15
Task data size/Mb	6–10
Task tolerance latency/s	0.8–1.3
Velocity of vehicles/km/h	50–90
Transmission power of vehicles /w	0.02–0.27
Noise power/dbm	−70
$f_{k,m}^{\mathrm{VEC}}$ for VEC server/GHz	6
Maximum transmit power of UEs/w	0.9
The bandwidth of each UE/MHz	0.9
Diameter of RSU coverage/m	500
The bandwidth of each UE/MHz	0.9

Figure 2 shows the convergence performance of the proposed algorithm. It can be seen that the proposed algorithm can quickly converge to the stationary point within 3 iterations. Besides, the obtained solution of the proposed algorithm is closed to that of an exhaustive search algorithm, which shows the effectiveness of the proposed algorithm.

Figure 3 displays the EC of UEs versus the amount of input data bits. As expected, with the increase of $D_{k,m}$, the EC of UEs under three resource allocation algorithms increases. The reason is that as the input data bits increase, the required computing resources also increase, which leads to a higher EC for each UE via (9). Besides, it is also observed that the proposed algorithm consumes the least energy because the proposed algorithm adopts partial offloading mode, in which a balance between local computing and data offloading can be achieved.

Figure 4 shows the relationship between the EC of UEs and the delay of tasks. In this figure, the EC decreases with the increasing task tolerance latency

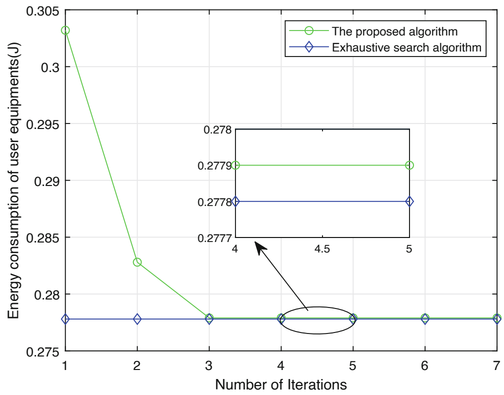

Fig. 2. The convergence performance of the proposed algorithm

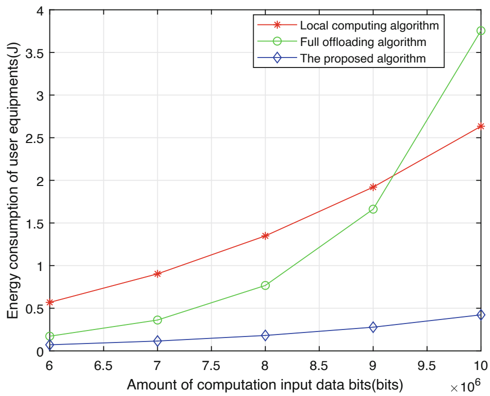

Fig. 3. EC vs input data bits

under three resource allocation algorithms. The reason is that a larger latency means that the computation burden per unit time decreases. Furthermore, the proposed algorithm has the lowest EC. This is because both the local computing algorithm and the full offloading algorithm only consider a single computation strategy, while leads to a non-flexible resource scheduling. However, this case can be overcome by the proposed algorithm with partial offloading.

Figure 5 shows vehicle velocity versus the EC of UEs. It can be seen from the figure that with the increase of vehicle velocity, the EC under all resource allocation algorithms keeps unchanged and begins to increase when vehicle velocity is bigger than 70 km/h. The reason is that the latency decreases with the increasing vehicle velocity, which leads to a higher EC for each UE accordingly. Moreover, the EC under the proposed algorithm is lower than other resource allocation algorithms. This is because the proposed algorithm can reduce the EC of UEs by dynamically adjusting the offloading portion.

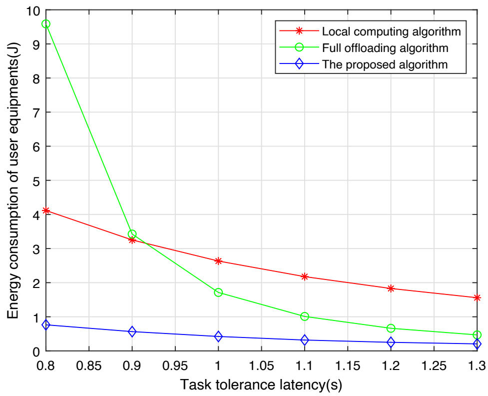

Fig. 4. EC vs delay

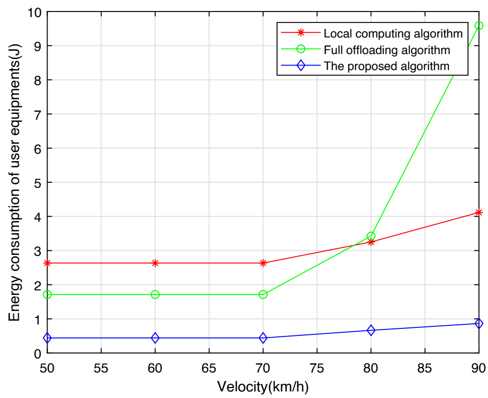

Fig. 5. EC vs velocity

5 Conclusion

In this paper, we have studied the joint computation offloading and wireless resource allocation problem for minimizing the total EC of UEs in vehicular networks with the constraints of the impact of vehicle mobility, the limited power supplies, computational capabilities of UEs, and the maximum latency of tasks. To solve this problem, a BCD-based iterative algorithm with greedy search was developed by jointly optimizing the computing ability of UEs, offloading decisions, transmit power, and subcarrier selection. Simulation results demonstrated that the proposed algorithm not only obtained a high-quality sub-optimal solution but also outperformed the benchmarks in terms of EC.

References

1. Lin, C., Han, G., Qi, X., Guizani, M., Shu, L.: A distributed mobile fog computing scheme for mobile delay-sensitive applications in SDN-enabled vehicular networks. IEEE Trans. Veh. Technol. **69**(5), 5481–5493 (2020)
2. Chen, Y., Wang, Y., Liu, M., Zhang, J., Jiao, L.: Network slicing enabled resource management for service-oriented ultra-reliable and low-latency vehicular networks. IEEE Trans. Veh. Technol. **69**(7), 7847–7862 (2020)
3. Wang, H., Lin, Z., Lv, T.: Energy and delay minimization of partial computing offloading for D2D-assisted MEC systems. In: 2021 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6 (2021)
4. Zhang, J., Guo, H., Liu, J., Zhang, Y.: Task offloading in vehicular edge computing networks: a load-balancing solution. IEEE Trans. Veh. Technol. **69**(2), 2092–2104 (2020)
5. Peng, E., Li, Z.: Optimal control-based computing task scheduling in software-defined vehicular edge networks. In: 2020 International Conference on Internet of Things and Intelligent Applications (ITIA), pp. 1–5 (2020). https://doi.org/10.1109/ITIA50152.2020.9312356
6. Misra, S., Bera, S.: Soft-VAN: mobility-aware task offloading in software-defined vehicular network. IEEE Trans. Veh. Technol. **69**(2), 2071–2078 (2020)
7. Du, J., Yu, F.R., Chu, X., Feng, J., Lu, G.: Computation offloading and resource allocation in vehicular networks based on dual-side cost minimization. IEEE Trans. Veh. Technol. **68**(2), 1079–1092 (2019)
8. Liu, Y., Yu, H., Xie, S., Zhang, Y.: Deep reinforcement learning for offloading and resource allocation in vehicle edge computing and networks. IEEE Trans. Veh. Technol. **68**(11), 11158–11168 (2019)
9. Zhang, K., Mao, Y., Leng, S., He, Y., Zhang, Y.: Mobile-edge computing for vehicular networks: a promising network paradigm with predictive off-loading. IEEE Veh. Technol. Mag. **12**(2), 36–44 (2017)
10. Bi, S., Zhang, Y.J.: Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading. IEEE Trans. Wirel. Commun. **17**(6), 4177–4190 (2018)
11. Saleem, U., Liu, Y., Jangsher, S., Tao, X., Li, Y.: Latency minimization for D2D-enabled partial computation offloading in mobile edge computing. IEEE Trans. Veh. Technol. **69**(4), 4472–4486 (2020)
12. You, C., Zeng, Y., Zhang, R., Huang, K.: Asynchronous mobile-edge computation offloading: energy-efficient resource management. IEEE Trans. Wirel. Commun. **17**(11), 7590–7605 (2018)
13. Wang, F., Xu, J., Wang, X., Cui, S.: Joint offloading and computing optimization in wireless powered mobile-edge computing systems. IEEE Trans. Wirel. Commun. **17**(3), 1784–1797 (2018)
14. Zhou, Z., Feng, J., Chang, Z., Shen, X.: Energy-efficient edge computing service provisioning for vehicular networks: a consensus ADMM approach. IEEE Trans. Veh. Technol. **68**(5), 5087–5099 (2019)
15. Gupta, A., Singh, K., Sellathurai, M.: Time-switching eh-based joint relay selection and resource allocation algorithms for multi-user multi-carrier AF relay networks. IEEE Trans. Green Commun. Netw. **3**(2), 505–522 (2019)
16. Xu, Y., Gu, B., Hu, R.Q., Li, D., Zhang, H.: Joint computation offloading and radio resource allocation in MEC-based wireless-powered backscatter communication networks. IEEE Trans. Veh. Technol. **70**(6), 6200–6205 (2021)

17. Song, Z., Liu, Y., Sun, X.: Joint radio and computational resource allocation for NOMA-based mobile edge computing in heterogeneous networks. IEEE Commun. Lett. **22**(12), 2559–2562 (2018)
18. Dai, Y., Xu, D., Maharjan, S., Zhang, Y.: Joint load balancing and offloading in vehicular edge computing and networks. IEEE Internet Things J. **6**(3), 4377–4387 (2019)
19. Dinkelbach, W.: On nonlinear fractional programming. Manage. Sci. **13**, 492–498 (1967)
20. Papandriopoulos, J., Evans, J.S.: SCALE: a low-complexity distributed protocol for spectrum balancing in multiuser dsl networks. IEEE Trans. Inf. Theory **55**(8), 3711–3724 (2009)
21. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, U.K. (2004)
22. Fang, F., Cheng, J., Ding, Z.: Joint energy efficient subchannel and power optimization for a downlink NOMA heterogeneous network. IEEE Trans. Veh. Technol. **68**(2), 1351–1364 (2019)
23. Lan, Y., Wang, X., Wang, D., Zhang, Y., Wang, W.: Mobile-edge computation offloading and resource allocation in heterogeneous wireless networks. In: 2019 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6 (2019)