IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 19 July 2024, accepted 5 August 2024, date of publication 7 August 2024, date of current version 19 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3440000

RESEARCH ARTICLE

Computation Offloading and Resource Allocation for Vehicle-Assisted Edge Computing Networks With Joint Access and Backhaul

BO HUANG, YAN ZHOU, XU ZHANG, JIE CHEN, AND LI SHANG
School of Electronic Information Engineering, Suzhou Vocational University, Suzhou 215104, China

Corresponding author: Bo Huang (hbo@jssvc.edu.cn)

This work was supported in part by the Startup Fund and Intramural Grants Program of Suzhou Vocational University under Grant 202105000037, in part by the Seventh Batch of Science and Technology Development Plan (Agriculture) Project of Suzhou under Grant SNG2023007, and in part by Jiangsu Province Vocational College Youth Teacher Enterprise Practice Training Project under Grant 2023QYSJ092.

ABSTRACT The edge computing utilizes vehicles as resources to assist in computational offloading can shorten the distance between users and computing servers, thereby improving the reliability of communication between them. In this paper, we investigate a Vehicle-assisted Edge Computing (VEC) model by jointly considering wireless access and backhaul links, and formulate an optimization problem that combines computational offloading and resource allocation, aiming at minimizing system delay. Further, the formulated problem is decomposed into two subproblems, e.g., computation offloading and resource allocation. In particular, we propose a new computational offloading approach that models the offloading decision for joint wireless access and backhaul as a potential game. The Nash equilibrium is guaranteed by the rational design of potential function, and the corresponding solution is solved by a backward induction method. On the other hand, the resource allocation subproblem is transformed from a nonconvex to a convex optimization problem based on equivalent transformation with successive convex approximation methods and finally derives the optimal solution satisfying Karush-Kuhn-Tucker (KKT) conditions. Simulation results show that the proposed algorithm have near-optimal performance and is superior to the state-of-the-art over a wide range of parameter settings.

INDEX TERMS Computation offloading, access/backhaul link, resource allocation, edge computing.

I. INTRODUCTION

The growing popularity of intelligent mobile applications, such as virtual reality, autonomous vehicles, and biometrics recognition, has collectively led to an exponential increase in the demand for data computation [1]. However, user equipment frequently lacks the requisite computing capacity to adopt these compute-intensive and latency-sensitive services due to their inherent resource limitations. To address this challenge, Mobile Edge Computing (MEC), which is introduced as a key technology in 5G networks [2], provides low-latency and high-reliability computation services by offloading users' computation data to the MEC servers deployed at the network edge node, such as roadside units and 5G base stations. However, the substantial expansion of access and input data will inevitably lead to congestion within the network. Consequently, the resolution of the conflict between the demand of users and the scarcity of wireless resources represents a pivotal area of future research.

The utilization of vehicles as resources has become an emerging paradigm to expand the MEC computation offloading capacity. The vehicles are capable to deploy small power base stations as the Vehicle Base Stations (VBS) with wireless transmission capabilities owing to their large size and continuous energy supply. The 3rd Generation Partnership Project (3GPP) specifies multiple communication interaction modes between vehicles and their surroundings, e.g., Vehicle-to-Network (V2N) [3], Vehicle-to-Pedestrian

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Khalil Afzal.

(V2P) [4], Vehicle-to-Infrastructure (V2I) [5], Vehicle-to-Vehicle (V2V) [6], etc. Utilizing V2I communications, the process that vehicles indirectly forward users' offloading data to MEC servers can shorten the distance between users and MEC servers. Significant research efforts have been devoted to study the computation offloading of vehicular edge computing [7], [8], [9], [10]. Karoui et al. [7] explored the impact of infrastructure deployment density (e.g., macro base station) on the performance of vehicular communication technologies in a real urban environment, confirming that denser infrastructure deployment significantly improves the performance of VBS in terms of communication range, traffic intensity, and frequency of message generation. Ho et al. [8] utilized index coding techniques to propagate 3D road map data through fog nodes to reduce the overall data load in a heterogeneous vehicular network. Luo et al. [9] proposed a algorithm to minimize the delay and cost of computation offloading for VEC from the perspective of multi-objective optimization. Guo et al. [10] coped with frequent handover problem of vehicular edge computing networks by designing an intelligent task offloading scheme based on deep learning. Consequently, the potential advantage for using mobile vehicles to assist in augmenting task processing capabilities of resource-limited user equipments by offloading their tasks to the edge servers has become an inevitable trend.

The resource scarcity problem is another challenge for vehicle-assisted edge computing network [11]. In order to obtain vehicle services, users must compete with each other to share communication resources, and transmit computation data to the MEC server through the vehicle's forwarding function. Intense competition for scarce communication resources among mobile users may significantly reduce the mobile users' transmission rate, leading to degradation of transmission performance and increase in transmission delay. Some researches endeavor concentrate on resource allocation for vehicle-assisted edge computing [12], [13], [14], [15]. Tang and Wu [12] proposed a real-time allocation method based on Stackelberg game for computational resource offloading to achieve win-win state where edge servers and vehicles are allowed to optimize their utility values. Wang et al. [13] proposed a blockchain model for secure resource sharing in VEC to solve the security issue caused the lack of incentive mechanism under asymmetric information. Zhang et al. [14] presented an end-edge-cloud collaboration paradigm by incorporating vehicles with idle resources to cope with the limited resources problem of edge servers. Gu and Zhang [15] formulated a minimum-maximum optimization problem for VEC networks that jointly optimizes transmission power, on-board processing power, and local model accuracy for worst-case case at minimum cost, considering vehicle position and speed. Nevertheless, none of the preceding researches explored the cooperative effects of resource allocation and computation offloading for communication and computing.

Significant researches have been conducted to investigate the joint optimization of task offloading and resource allocation in VEC [16], [17], [18], [19], [20]. Fan et al. [16] proposed a scheme minimized the system delay for all vehicles by jointly optimizing task offloading and resource allocation. Zhang et al. [17] investigated the problem of joint task offloading and resource allocation in vehicle networks with the aim of optimizing the system utility associated with the latency and cost of computing and communication services. Feng et al. [18] designed a dual-interface offloading and resource allocation strategy for cellular vehicle-to-everything systems to minimize the system latency with characterizing successful transmission probability. Gao et al. [19] formulated a joint problem of considering task offloading scheduling, resource allocation and time-varying channel to minimize the delay and energy consumption for vehicular edge computing. Ju et al. [20] proposed a joint security offloading and resource allocation scheme based on deep reinforcement learning, which utilizes physical layer security techniques and spectrum sharing architecture to improve the confidentiality performance and resource efficiency of multi-user VEC networks. However, existing studies have predominantly focused on data transmission for wireless access at vehicle base stations, without concurrently addressing the influence of wireless backhaul on the data offloading and communication resource allocation.

A few literatures have focused on task offloading or resource allocation by considering wireless backhaul. Chen et al. [21] proposed a joint optimization problem for offloading decision and resource allocation to minimize the total energy consumption for processing computational tasks with considering the limited capacity of the access and backhaul links. Peng et al. [22] formulated a migration optimization problem aiming to maximize the services' satisfaction degree of delay in VEC networks with the limited backhaul bandwidth resource. However, these solutions treated the wireless access and backhaul as two separate processes, lacking researches that consider both as a whole to jointly affect computing task offloading and resource allocation in vehicle networks.

After briefly summarizing the relevant studies and mentioning the motivation, this paper investigates a scheduling scheme for joint computational offloading and resource allocation in vehicle-assisted edge computing network based on the potential game theory and convex optimization theory. Specifically, we first model the offloading decision process for joint wireless access and backhaul as a potential game [23] including a well-designed potential function that ensures the existence and convergence of a Nash equilibrium (NE), which is a state of perfect competition that assumes that all rational players aim at profit maximization, where the users are rational participants to minimize communication/computation delay. According to the potential game theory, NE can be achieved by minimizing the potential of each user through the designed potential function. Furthermore, the remaining

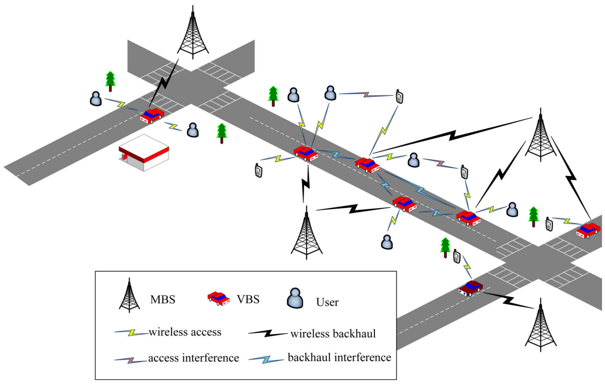


FIGURE 1. The system model.

multiple resource allocation problem is transformed from a nonconvex to a convex optimization problem based on equivalent transformation with successive convex approximation method and finally derives the optimal solution satisfying KKT conditions [24]. The main contributions of this work are outlined as follows.

- We formulate a computational offloading cooperative resource allocation problem in a VEC network to minimize the system delay by jointly optimizing offloading decision and bandwidth allocation in wireless access/backhaul links, transmit power and task computation frequency, guaranteeing sufficient backhaul capacity for VBSs. Specifically, we derive offloading decision model resulting from the joint impact of wireless access and backhaul. Then, we prove that the considered problem belongs to the difficult mixed integer non-convex optimization.
- We decompose the proposed problem into two subproblems, i.e., computation offloading and resource allocation. Specifically, the computation offloading subproblem is modeled as a noncooperative game among users and further prove it as a potential game with NE existence and convergence. Then, the nonconvex resource allocation subproblem is solved using equivalent transformations and successive convex approximation methods, and finally satisfies KKT conditions and derives the optimal solution.
- Simulation results verify the convergence effect of the proposed algorithm, and have been demonstrate that, our proposal is closer to the global optimization solution and can significantly reduce delay in comparison with edge computation server only scheme, local computation only scheme and random offloading scheme in various scenarios.

The rest of this paper is organized as follows. Section II introduces the proposed vehicle-assisted edge computing system model and formulates the delay-minimization computation offloading and resource allocation problem. Section III details the solution process of the proposed problem. Section IV presents the performance evaluation. Section V concludes the paper.

TABLE 1. Notations.

Notation	Description
U	Number of users
A	Number of VBSs
B	Number of MBSs
$\mathbf{X}$	Association matrix between users and VBSs
$\mathbf{Y}$	Association matrix between VBSs and MBSs
$g_{x,y}$	Channel gain between network nodes $x \in \{i, a\}$ and $y \in \{a, b\}$
$\beta_{x,y}$	Log-normal shadowing path loss components between network nodes $x \in \{i, a\}$ and $y \in \{a, b\}$
$h_{x,y}$	Small-scale fading components between network nodes $x \in \{i, a\}$ and $y \in \{a, b\}$
α	Path-loss exponent
v	Velocity of vehicles
W^A	Wireless access channel bandwidth
W^B	Wireless backhaul channel bandwidth
N_0	Noise power
T_i	Computation task of user i
$\mathbf{m}, \mathbf{n}$	Vector of bandwidth fraction allocated to access link and backhaul link, respectively
$\mathbf{p}$	Vector of transmit power
$\mathbf{f}$	Vector of computation frequency
$t_i^{\max}$	Delay constraint
$p_i^{\max}$	Maximum power of user i
$p_{i_a}^{\max}$	Maximum power of VBS a
α	Path-loss exponent

II. SYSTEM MODEL

In this paper, we investigate a vehicles-assisted edge computation network consisting of U users, A VBSs moving on the bidirectional roads, and B Macro Base Stations (MBS), as shown in Fig. 1. Each MBS is equipped with a edge computation server to provide computing service to users. The set of the user, VBS and MBS are denoted by $\mathcal{U} = \{1, \ldots, i, \ldots, U\}$, $\mathcal{A} = \{1, \ldots, a, \ldots, A\}$ and $\mathcal{B} = \{1, \ldots, b, \ldots, B\}$, respectively. The task requesting user firstly offloads its computational task to a VBS via the wireless access link, and denote an incident matrix $\mathbf{X} = [x_{i,a}]_{i\in\mathcal{U},a\in\mathcal{A}} \in \{0, 1\}^{U\times A}$ where $x_{i,a} = 1$ represents user i is associated with vehicle a, and $x_{i,a} = 0$ otherwise. Then, the VBS forward the offloading task to an associated MBS via wireless backhaul to complete the task computation, and the backhaul offloading decision incident matrix denoted as $\mathbf{Y} = [y_{i_a,b}]_{i\in\mathcal{U},b\in\mathcal{B}} \in \{0, 1\}^{A\times B}$, where $y_{i_a,b} = 1$ if the offloading data of user i is offloaded to MBS b through VBS a, and $y_{i_a,b} = 0$, otherwise. The main notations used in this paper are appeared in Table 1.

A. CHANNEL MODEL

The channel gain between network nodes $x \in \{i, a\}$ and $y \in \{a, b\}$ is denoted by $g_{x,y}$, and it can be present as

$$g_{x,y} = |h_{x,y}|^2 \beta_{x,y}, \tag{1}$$

where $\beta_{x,y} = -120.9 + 10\alpha log_{10}(\frac{d_{x,y}}{d_0}) + X_\sigma$ is the log-normal shadowing path loss components that considers both geometric attenuation and shadow fading [25], $d_{x,y}$ is the distance between network nodes x and y, the remain constant for a fixed distance between the two associated nodes in which X_σ follows $X_\sigma \sim \mathcal{N}(0, \sigma^2)$, and α is the path-loss exponent; N_0 is the background noise power; $|\cdot|$ is the modules operation; $h_{x,y}$ is the small-scale fading components between network nodes x and y, which varies in different frames and is represented by Jakes' model as [26]

$$h_{x,y} = \rho h_{x,y}^{pre} + n_{x,y}, \tag{2}$$

where $n_{x,y} \sim \mathcal{CN}(0, 1-\rho^2)$ and $h_{x,y} \sim \mathcal{CN}(0, 1)$, $h_{x,y}^{pre}$ is the previous channel small-scale fading; $\rho = J_0(2\pi T_s v f_c/c)$ is the correlation coefficie where $J_0(\cdot)$ is the first kind zero-order Bessel function, T_s is the time interval between adjacent frames, v is the relative speed between vehicles, f_c is the carrier frequency of the signal, and c is the speed of light.

B. COMMUNICATION MODEL

1) WIRELESS ACCESS LINK

When user i is associated to VBS a, the VBS allocates the portion of bandwidth $m_{i,a}W^A$ to user i where $m_{i,a} \in [0, 1]$ and W^A are the bandwidth allocation ratio between user i and VBS a, and bandwidth on the wireless access link, respectively. The uplink achievable data rate of the wireless access link between user i and VBS a is defined as:

$$r_{i,a} = m_{i,a}W^A \log(1 + SINR_{i,a}), \tag{3}$$

$$SINR_{i,a} = \frac{x_{i,a}p_i g_{i,a}}{I_{i,a} + N_0}, \tag{4}$$

where $SINR_{i,a}$ is the signal-to-interference-plus-noise-ratio (SINR) of user i associated with VBS a; p_i is the transmit power of user i; $I_{i,a} = \sum\limits_{i' \in \mathcal{U}\backslash\{i\}} \sum\limits_{a' \in \mathcal{A}\backslash\{a\}} x_{i',a'}p_{i'}g_{i',a'}$ is the interference from other users competing for the shared wireless access resource; N_0 is the noise power.

2) WIRELESS BACKHAUL LINK

After receiving the computation task from user i, VBS a forwards it to associated MBS $b \in \mathcal{B}$, which is allocated a portion of bandwidth $n_{i_a,b}W^B$ for the process, where $n_{i_a,b} \in [0, 1]$ and $W^B$are the bandwidth allocation ratio and bandwidth for the wireless backhaul, respectively. The uplink achievable data rate of the backhaul link between VBS a and MBS b is defined as

$$r_{i_a,b} = n_{i_a,b}W^B \log(1 + SINR_{i_a,b}), \tag{5}$$

$$SINR_{i_a,b} = \frac{y_{i_a,b}p_{i_a}g_{i_a,b}}{I_{i_a,b} + N_0}, \tag{6}$$

where $SINR_{i_a,b}$ is the backhaul SINR of VBS a sending the computing data task of user i to MBS b; p_{i_a} is the transmit power of VBS a for delivering task of user i; $I_{i_a,b} = \sum\limits_{i' \in \mathcal{U}\backslash\{i\}} \sum\limits_{a' \in \mathcal{A}\backslash\{a\}} \sum\limits_{b' \in \mathcal{B}\backslash\{b\}} y_{i'_{a'},b'}p_{i'_{a'}}g_{i'_{a'},b'}$ is the interference from other VBSs competing for the shared wireless backhaul resources.

C. COMPUTATION MODEL

Assuming that user $i, i \in \mathcal{U}$ has a computation task that is expressed as $T_i = \left\{d_i, c_i, t_i^{\max}\right\}$. d_i is the data size of each computation task, c_i denotes the number of Central Processing Unit (CPU) cycles necessary to accomplish the computation task, and $t_i^{\max}$ is the maximum tolerable delay for the user.

When the computation task of user i is sent to MBS b for execution via VBS a, the total offloading delay of edge computing comes primarily from the transmission and computation of computation tasks. Specifically, the transmission delay of this computation task is described by

$$t_{i_a,b}^{tras} = \frac{d_i}{r_{i,a}} + \frac{d_i}{r_{i_a,b}}. \tag{7}$$

The computational delay of a computation task on MBS b is derived by

$$t_{i_a,b}^{comp} = \frac{d_i c_i}{f_{i_a,b}}, \tag{8}$$

where $f_{i_a,b}$ denotes the computation capability (i.e., CPU cycles per second) of MBS b allocated to vehicle a.

Therefore, the total offloading delay of this computation task in the edge computing is defined as

$$t_{i_a,b} = t_{i_a,b}^{tras} + t_{i_a,b}^{comp}. \tag{9}$$

In the paper, we do not consider the downlink delay incurred in transmitting the computation results performed by MBSs back to users, because the size of computation result is generally smaller than that of the input.

D. PROBLEM FORMULATION

Latency is one of the key measures for computation offloading. In this paper, transmitting offloading tasks with assisted-vehicles require deciding which VBSs are as access nodes and which MBSs are as backhaul nodes, as well as these offloading selection decisions will have a direct impact on the system latency. Thus, this paper aims to minimize system delay incurred by offloading process via jointly optimizing the offloading decision $\{\mathbf{X}, \mathbf{Y}\}$ and the resource allocation in terms of the bandwidth allocation $\{\mathbf{m}, \mathbf{n}\} = \left\{\left[m_{i,a}\right]_{i\in\mathcal{U}, a\in\mathcal{A}}, \left[n_{i_a,b}\right]_{a\in A, b\in B}\right\}$ in both access and backhaul links, the transmit power $\mathbf{p} = \left\{[p_i]_{i\in\mathcal{U}}, \left[p_{i_a}\right]_{a\in A}\right\}$, and the edge computation frequency $\mathbf{f} = \left[f_{i_a,b}\right]_{a\in A, b\in B}$, taking the backhaul capacity constraint, latency constraint, and offloading decision constraint into account. We formulate the Delay-minimization Computation offloading and Resource allocation (DCR) problem as

$$\min_{\mathbf{X},\mathbf{Y},\mathbf{p},\mathbf{m},\mathbf{n},\mathbf{f}} \sum_{i\in\mathcal{U}} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} t_{i_a,b}$$

$$\text{s.t.} \sum_{i\in\mathcal{U}} r_{i,a} \leq \sum_{b\in\mathcal{B}} r_{i_a,b}, \forall a \in \mathcal{A}, \tag{10a}$$

$$\sum_{a\in\mathcal{A}} x_{i,a} \leq 1, \sum_{b\in\mathcal{B}} y_{i_a,b} \leq 1, \forall i \in \mathcal{U}, \tag{10b}$$

$$x_{i,a} \in \{0, 1\}, y_{i_a,b} \in \{0, 1\} \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \tag{10c}$$

$$t_{i_a,b} \leq t_i^{\max}, \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \tag{10d}$$

$$0 \le f_{i_a,b} \le f_{i_a,b}^{\max}, \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \quad (10e)$$

$$0 \le p_i \le p_i^{\max}, 0 \le p_{i_a} \le p_{i_a}^{\max}, \forall i \in \mathcal{U}, a \in \mathcal{A}, \quad (10f)$$

$$0 < m_{i,a} < 1, 0 < n_{i_a,b} < 1, \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}. \quad (10g)$$

Constraint (10a) ensures that the backhaul rate can support the total rate from all users associated with each VBS. Constraint (10b) denotes that each user or VBS can be only associate with one VBS or MBS at a given time respectively, Constraint (10c) indicates that user i decides whether to offload its computation task from VBS a to MBS b or not, whereas $t_i^{\max}$ of constraint (10d) represents the maximum delay constraints for individual users. Finally, constraints (10e)-(10g) specify allowable ranges in terms of computation frequency, transmit power, and bandwidth allocation ratio, respectively.

The DCR is a mixed-integer nonlinear programming problem, recognized as NP-hard [27], because the objective function and constraints of (10) are non-convex and contain integer variables. To address the problem, traditional centralized algorithms, such as exhaustive search, need to collect all the necessary information (including channel state, traffic characteristics, and interference state of all users), and need to assign computation offloading and resource allocation decisions to all VBSs and MBSs, resulting in extremely high control and signaling overhead among network nodes, which may not be tolerated by latency-sensitive tasks. To make the problem tractable and simplify it, we divide the original problem into two subproblems and solve them accordingly.

III. PROBLEM SOLUTION

The DCR problem can be addressed by decoupling the two subproblems, i.e., task offloading ($P1$) and resource allocation ($P2$). Specifically, the subproblem $P1$ is modeled as a noncooperative game among users and further prove it as a potential game with NE existence and convergence. On the other hand, the subproblem $P2$ is a nonconvex problem that are solved using equivalent transformations and successive convex approximation methods in order to drive the optimal solutions for resource allocation that satisfies KKT conditions.

A. FORMULATION AND SOLUTION OF TASK OFFLOADING SUBPROBLEM

Given the resource allocation variables $\{\mathbf{p}, \mathbf{m}, \mathbf{n}, \mathbf{f}\}$, the computational offloading subproblem P1 with respect with $\mathbf{X}$ and $\mathbf{Y}$ is formulated as

$$P1: \min_{\mathbf{X},\mathbf{Y}} \sum_{i\in\mathcal{U}} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} t_{i_a,b}$$

$$\text{s.t.} \quad \sum_{i\in\mathcal{U}} r_{i,a} \le \sum_{b\in\mathcal{B}} r_{i_a,b}, \forall a \in \mathcal{A}, \quad (11a)$$

$$\sum_{a\in\mathcal{A}} x_{i,a} \le 1, \sum_{b\in\mathcal{B}} y_{i_a,b} \le 1, \forall i \in \mathcal{U} \quad (11b)$$

$$x_{i,a} \in \{0, 1\}, y_{i_a,b} \in \{0, 1\} \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B} \quad (11c)$$

We then model subproblem $P1$ as a noncooperative game among users, who as players independently determine jointly access and backhaul offloading strategy for computational tasks. Denote $\mathcal{G} = \{\mathcal{U}, \mathbb{S}, \{u_i\}_{i\in\mathcal{U}}\}$ as the game model, where $\mathcal{U}$ represents the set of players; $\mathbb{S}$ represents the strategy space of the game, defined as the Cartesian product of all individual strategy sets of players, i.e., $\mathbb{S} = \mathbf{s}_1 \times \cdots \times \mathbf{s}_i \times \cdots \times \mathbf{s}_U$, where $\mathbf{s}_i = [s_{i_a,b}]_{a\in A, b\in B} \in \{0, 1\}^{A\times B}$, $s_{i_a,b} = x_{i,\alpha} \cdot y_{i_a,b}$ is the set of all selection strategies for user i, if $s_{i_a,b} = 1$ implies $x_{i,a} = 1$ and $y_{i_a,b} = 1$ which means user i chooses VBS a as the access association and then VBS a chooses MBS b as the backhaul association. A strategy selection profile is denoted as $\mathbf{s} = \{s_1, \ldots, s_i, \ldots, s_U\} \in \mathbb{S}$, which can be also rewritten as $\mathbf{s} = (s_i, \mathbf{s}_{-i})$, where$s_i = \{s_{i_a,b} | i \in \mathcal{U}\}$ is the strategy of user i and $\mathbf{s}_{-i}$ represents the strategies set of $U - 1$ users except user i; u_i denotes the utility function of play i. Considering all users adopt the combined wireless access and backhaul offloading strategy, the task offloading mode is susceptible to suffer from the SINR effect of both access and backhaul links. Thus, we model a utility function that jointly takes the SINR of both access and backhaul links into account to determine the task offloading strategy. The utility function can be defined as follows.

Definition 1: The utility function of user i is defined as the sum of the inverse of access and backhaul SINR of the users under the strategy profile S as follows

$$u_i(s_i, \mathbf{s}_{-i}) = \sum_{i\in\mathcal{U}} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} s_{i_a,b} \left(\frac{1}{SINR_{i,a}} + \frac{1}{SINR_{i_a,b}} \right). \quad (12)$$

By providing a potential function as follows, we further prove that the noncooperative game model $\mathcal{G}$ is a potential game with NE existence and convergence.

Definition 2: A game is defined a potential game if and only if exists a potential function $\phi(S)$that satisfies: ϕ:$\mathbb{R}$ for user $i \in \mathcal{U}$, $S \in \mathbb{S}$, and $\mathbf{s}_{-i} \in \prod_{j\ne i} \mathbf{s}_j$, such that:

$$u_i(s_i, \mathbf{s}_{-i}) - u_i(s'_i, \mathbf{s}_{-i}) > 0 \Rightarrow \phi(s_i, \mathbf{s}_{-i}) - \phi(s'_i, \mathbf{s}_{-i}) > 0. \quad (13)$$

Theorem 1: The proposed computational offloading game $\mathcal{G}$ is a potential game with potential function as follows:

$$\phi(s_i, \mathbf{s}_{-i}) = \frac{1}{2} \sum_{i\in\mathcal{U}} \sum_{j\in\mathcal{U}\setminus\{i\}} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} s_{i_a,b} s_{j_a,b} C_{i_a,b} C_{j_a,b}, \quad (14)$$

where $C_{i_a,b} = \frac{1}{SINR_{i,a}} + \frac{1}{SINR_{i_a,b}}$.

Proof: According to Definition 2, we should prove that the potential function increases or decreases as $C_{i_a,b}(s_i, \mathbf{s}_{-i})$ increases or decreases. Assuming that user i

satisfies the decision update condition under the current decision s_i, there exists a decision ${s'}_i$ that is able to gain smaller total offloading delay, i.e., $\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} t_{i_a,b}\left(s_i,\mathbf{s}_{-i}\right) > \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} t_{i_a,b}\left({s'}_i,\mathbf{s}_{-i}\right)$. Since the inverse of the access and backhaul SINR decreases monotonously as the offloading delay function decreases, user i will update its current decision s_i to ${s'}_i$, and $\sum_{i\in\mathcal{U}} u_{i_a,b}\left(s_i,\mathbf{s}_{-i}\right) > \sum_{i\in\mathcal{U}} u_{i_a,b}\left({s'}_i,\mathbf{s}_{-i}\right)$ is satisfied.

Suppose that an improving strategy for user i is from s_i to ${s'}_i$, and the other users' task offloading selection strategies hold unchanged, thus remain unchanged, then $s_{i_a,b}$=1 and ${s'}_{i_{a'},b'}$=1, based on the potential function (14), we have

$$\begin{aligned}
&\phi(s_i,\mathbf{s}_{-i})-\phi({s'}_i,\mathbf{s}_{-i})\\
&=\frac{1}{2}s_{i_a,b}C_{i_a,b}\sum_{j\in\mathcal{U}\backslash\{i\}} s_{j_a,b}C_{j_a,b}+\frac{1}{2}s_{i_a,b}C_{i_a,b}\sum_{j\in\mathcal{U}\backslash\{i\}} s_{j_a,b}C_{j_a,b}\\
&\quad+\frac{1}{2}\sum_{j'\in\mathcal{U}\backslash\{i\}} s_{{j'}_a,b}C_{{j'}_a,b}\sum_{j\in\mathcal{U}\backslash\{i,j'\}}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} s_{j_a,b}C_{j_a,b}\\
&\quad-\frac{1}{2}\sum_{j'\in\mathcal{U}\backslash\{i\}} {s'}_{{j'}_{a'},b'}C_{{j'}_{a'},b'}\sum_{j\in\mathcal{U}\backslash\{i,j'\}}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} s_{j_{a'},b'}C_{j_{a'},b'}\\
&\quad-\frac{1}{2}{s'}_{i_{a'},b'}C_{i_{a'},b'}\sum_{j\in\mathcal{U}\backslash\{i\}} s_{j_{a'},b'}C_{j_{a'},b'}\\
&\quad-\frac{1}{2}{s'}_{i_{a'},b'}C_{i_{a'},b'}\sum_{j\in\mathcal{U}\backslash\{i\}} s_{j_{a'},b'}C_{j_{a'},b'}\\
&=C_{i_a,b}\sum_{j\in\mathcal{U}\backslash\{i\}} s_{j_a,b}C_{j_a,b}-C_{i_{a'},b'}\sum_{j\in\mathcal{U}\backslash\{i\}} s_{j_{a'},b'}C_{j_{a'},b'}>0.
\end{aligned}\tag{15}$$

■

1) EXISTENCE OF THE NASH EQUILIBRIUM

Definition 3: The strategy profile $\mathbf{s}^* \in \mathbb{S}$ is a Nash equilibrium of the task offloading game $\mathcal{G}$, i.e., no user could further change its strategy to obtain smaller delay, if and only if [28]

$$u_i\left(s_i,\mathbf{s}^*_{-i}\right)\geq u_i\left(s^*_i,\mathbf{s}^*_{-i}\right)\tag{16}$$

Theorem 2: The game $\mathcal{G}$ has at least one NE If there is a potential function $\phi\left(s_i,\mathbf{s}_{-i}\right)$.

Proof: In each iteration, the game $\mathcal{G}$ selects strategy combinations from the strategy space $\mathbb{S}$ based on the best response dynamics, and forms an improvement sequence after multiple iterations. Since the user's strategy space $\mathbb{S}$ is closed and bounded, the strategy combinations are also bounded and consequently the improvement sequence is also bounded. The game $\mathcal{G}$ has finite improvement properties, thus necessitating the presence of at least one finite improvement path. The potential function is required to exhibit a decrease along the improvement path and ultimately to attain the NE. ■

B. FORMULATION AND SOLUTION OF RESOURCE ALLOCATION SUBPROBLEM

Given the offloading decision strategies $\{\mathbf{X},\mathbf{Y}\}$, the second subproblem $P2$ with respect to $\{\mathbf{p},\mathbf{m},\mathbf{n},\mathbf{f}\}$ considers transmission power, bandwidth allocation in both wireless access and backhaul links, and computation resource allocation respectively, which is formulated as follows

$$P2:\min_{\mathbf{p},\mathbf{m},\mathbf{n},\mathbf{f}}\ g=\sum_{i\in\mathcal{U}}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} t_{i_a,b}$$

$$\text{s.t.}\ \sum_{i\in\mathcal{U}} r_{i,a}\leq\sum_{b\in\mathcal{B}} r_{i_a,b},\forall a\in\mathcal{A},\tag{17a}$$

$$t_{i_a,b}\leq t_i^{\max},\forall i\in\mathcal{U},a\in\mathcal{A},b\in\mathcal{B}\tag{17b}$$

$$0\leq f_{i_a,b}\leq f_{i_a,b}^{\max},\forall i\in\mathcal{U},a\in\mathcal{A},b\in\mathcal{B}\tag{17c}$$

$$0\leq p_i\leq p_i^{\max},0\leq p_{i_a}\leq p_{i_a}^{\max},\forall i\in\mathcal{U},a\in\mathcal{A},\tag{17d}$$

$$0<m_{i,a}<1,0<n_{i_a,b}<1,\forall i\in\mathcal{U},a\in\mathcal{A},b\in\mathcal{B}\tag{17e}$$

Due to the coupling and nonconvexity, several equivalent transformation with convex approximation methods are utilized to solve the second subproblem $P2$. Firstly, we transform the subproblem $P2$ into a more tractable form through a series of equivalent transformations. Subsequently, the convergence of convex upper bound for subproblem $P2$ is guaranteed by applying the successive convex approximation method [29] to iteratively solve a set of convex optimization problems.

The first step is to deal with the coupling among variables. By introducing slack variables $\left\{z_{i,a},q_{i_a,b}\right\}\geq 0,\forall i\in\mathcal{U}$, $a\in\mathcal{A}$, $b\in\mathcal{B}$, the subproblem $P2$ is rewritten as

$$P2.1:\min_{\substack{\mathbf{p},\mathbf{m},\mathbf{n},\mathbf{f},\\ \mathbf{z}\geq 0,\mathbf{q}\geq 0}}\ \sum_{i\in\mathcal{U}}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}}\frac{z_{i,a}}{d_i}+\frac{q_{i_a,b}}{d_i}+\frac{f_{i_a,b}}{c_i}$$

$$\begin{aligned}\text{s.t.}\ &\sum_{i\in\mathcal{U}} m_{i,a}W^A\log(1+SINR_{i,a})\\ &\leq\sum_{b\in\mathcal{B}} n_{i_a,b}W^B\log(1+SINR_{i_a,b}),\forall a\in\mathcal{A},\end{aligned}\tag{18a}$$

$$SINR_{i,a}\leq 2^{z_{i,a}/m_{i,a}W^A}-1,\forall i\in\mathcal{U},a\in\mathcal{A},\tag{18b}$$

$$SINR_{i_a,b}\leq 2^{q_{i_a,b}/n_{i_a,b}W^B}-1,\ \forall i\in\mathcal{U},a\in\mathcal{A},b\in\mathcal{B},\tag{18c}$$

$$\text{(17b), (17c), (17d), (17e)},\tag{18d}$$

where $\{\mathbf{z},\mathbf{q}\}=\left\{\left[z_{i,a}\right]_{i\in\mathcal{U},a\in\mathcal{A}},\left[q_{i_a,b}\right]_{a\in A,b\in B}\right\}$. Indeed, although the right hand sides of (18b) and (18c) are nonconvex, they are in convex exponential cone form by multiplying $z_{i,a}$ and $q_{i_a,b}$ on both sides respectively. However, the left hand sides of (18b) and (18c) are still nonconcave functions, the additional slack variables $\left\{d_{i,a},e_{i_a,b},k_{i,a},l_{i,a},o_{i_a,b},v_{i_a,b},w_{i_a,b}\right\}\geq 0$ are introduced to further relax these constraints as

$$P2.2: \min_{\Omega} \sum_{i\in\mathcal{U}}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} \frac{z_{i,a}}{d_i} + \frac{q_{i_a,b}}{d_i} + \frac{f_{i_a,b}}{c_i}$$

$$\text{s.t. } W^B \sum_{b\in\mathcal{B}} n_{i_a,b} e_{i_a,b} \le W^A \sum_{i\in\mathcal{U}} m_{i,a} d_{i,a}, \forall a \in \mathcal{A}, \tag{19a}$$

$$m_{i,a} p_i g_{i,a} \le k_{i,a} l_{i,a}, \forall i \in \mathcal{U}, a \in \mathcal{A}, \tag{19b}$$

$$n_{i_a,b} p_{i_a} g_{i_a,b} \le o_{i_a,b} v_{i_a,b}, \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \tag{19c}$$

$$d_{i,a} \le \log(1 + SINR_{i,a}), \forall i \in \mathcal{U}, a \in \mathcal{A}, \tag{19d}$$

$$p_{i_a} g_{i_a,b} \le v_{i_a,b} w_{i_a,b}, \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \tag{19e}$$

$$k_{i,a} \le m_{i,a} 2^{z_{i,a}/m_{i,a} W^A} - m_{i,a} \forall i \in \mathcal{U}, a \in \mathcal{A}, \tag{19f}$$

$$l_{i,a} \le \sum_{i'\in\mathcal{U}\setminus\{i\}} \sum_{a'\in\mathcal{A}\setminus\{a\}} x_{i',a'} p_{i'} g_{i',a'} + N_0 \forall i \in \mathcal{U}, a \in \mathcal{A}, \tag{19g}$$

$$w_{i_a,b} \le \left(2^{e_{i_a,b}} - 1\right) \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \tag{19h}$$

$$v_{i_a,b} \le \sum_{i'\in\mathcal{U}\setminus\{i\}} \sum_{a'\in\mathcal{A}\setminus\{a\}} \sum_{b'\in\mathcal{B}\setminus\{b\}} y_{i'_{a'},b'} p_{i'_{a'}} g_{i'_{a'},b'} + N_0 \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \tag{19i}$$

$$o_{i_a,b} \le n_{i_a,b} 2^{q_{i_a,b}/n_{i_a,b} W^B} - n_{i_a,b}, \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \tag{19j}$$

$$(17b), (17c), (17d), (17e), \tag{19k}$$

where $\Omega = \{\mathbf{p}, \mathbf{m}, \mathbf{n}, \mathbf{f}, \mathbf{z} \ge 0, \mathbf{q} \ge 0, \mathbf{d} \ge 0, \mathbf{e} \ge 0, \mathbf{k} \ge 0, \mathbf{l} \ge 0, \mathbf{o} \ge 0, \mathbf{v} \ge 0, \mathbf{w} \ge 0\}$. Indeed, the problem (18) and (19) are equivalent.

The transformed optimization subproblem (19) is still nonconvex due to the nonconvex constraints (19a)-(19e). In order to overcome these obstacles, the convex upper bounds of the nonconvex terms are derived by applying successive convex approximation method so that obtain a global upper bound of the original problem (17). First, the nonconvex constraint form present in (19a) - (19c) have same formula form, e.g., $\zeta\psi \ge \mu\nu$. It is obvious that both sides of this inequality are neither convex nor concave function with respect to all variables. According to the result of [30], the right-hand side of the inequality is firstly substituted by its convex upper bound $\mu\nu \le \frac{\theta}{2}\mu^2 + \frac{1}{2\theta}\nu^2$ where $\theta \ge 0$. By defining define $\theta^{(n)} = \nu^{(n)}/\mu^{(n)}$ given a feasible point $\left(\mu^{(n)}, \nu^{(n)}\right)$, the variable $\theta^{(n)}$ is updated in an iterative form to solve the function $z \ge \frac{\theta^{(n)}}{2}x^2 + \frac{1}{2\theta^{(n)}}y^2$ until convergence to a limiting point satisfying KKT constraints. Based on the second-order cone function $\zeta\psi = \frac{(\zeta+\psi)^2-(\zeta-\psi)^2}{4}$, the approximate convex constraint can be rewritten as

$$\frac{(\zeta+\psi)}{2} \ge \sqrt{\frac{\theta^{(n)}}{2}x}\sqrt{\frac{1}{2\theta^{(n)}}y}, \frac{(\zeta-\psi)}{2}. \tag{20}$$

Next, we rewrite the nonconvex constraint (19d) as

$$d_{i,a} + \log\left(\sum_{i'\in\mathcal{U}\setminus\{i\}} \sum_{a'\in\mathcal{A}\setminus\{a\}} p_{i'} g_{i',a'} + N_0\right) \le \log\left(\sum_{i\in\mathcal{U}}\sum_{a\in\mathcal{A}} p_i g_{i,a} + N_0\right). \tag{21}$$

Obviously, the right-hand side of the nonconvex function (21) can be convexly approximated around the point $p_i^{(n)}$ using the first-order Taylor approximation,

$$\log\left(\sum_{i\in\mathcal{U}}\sum_{a\in\mathcal{A}} p_i^{(n)} g_{i,a} + N_0\right) + \frac{\sum_{i\in\mathcal{U}}\sum_{a\in\mathcal{A}}\left(p_i - p_i^{(n)}\right) g_{i,a}}{\sum_{i\in\mathcal{U}}\sum_{a\in\mathcal{A}} p_i^{(n)} g_{i,a} + N_0} \ge d_{i,a} + \log\left(\sum_{i'\in\mathcal{U}\setminus\{i\}} \sum_{a'\in\mathcal{A}\setminus\{a\}} p_{i'} g_{i',a'} + N_0\right). \tag{22}$$

Substituting the approximations (20) and (22) into corresponding nonconvex term in (19a) - (19d) and (19e) respectively, the approximated convex problem (23), as shown at the bottom of the next page, is a global upper bound for problem (17).

Theorem 3: the sequence $\Omega^{(n)}$ obtained by iteratively solving (23) finally converges to a KKT point.

Proof: We first obtain an initialization point $\Omega^{(0)}$ through (23), which satisfies the Slater conditions [31] because (23) is a convex problem. Similarly, the optimization point $\Omega^{(n)}$ generated by the previous iteration $\Omega^{(n-1)}$ also satisfies the Slater conditions. Repeat this step to conclude that the optimization point Ω^* of problem (23) holds the Slater conditions. The slack variables $\{\mathbf{z}, \mathbf{q}\}$ are the upper bound of the objective function (17a), the function (23a) is the upper bound of the function (17a) as follow

$$g\left(\Omega^{(n)}\right) = \tilde{g}\left(\Omega^{(n)}, \Omega^{(n)}\right) \ge \tilde{g}\left(\Omega^{(n+1)}, \Omega^{(n)}\right) \ge g\left(\Omega^{(n+1)}\right) \tag{24}$$

This means that the current iteration point $\Omega^{(n+1)}$ can present lower delay performance for the problem (23) than the previous iteration point $\Omega^{(n)}$ on condition that the $\Omega^{(n+1)} \ne \Omega^{(n)}$ is satisfied. Based on the boundedness of the sequence $\Omega^{(n)}$, these exists a convergent subsequence $\left\{\Omega^{(n_v)}\right\}_{v=1}^{\infty}$ satisfying condition $\lim_{v\to+\infty}\left[g\left(\Omega^{(n_{v+1})}\right) - g\left(\Omega^{(n_v)}\right)\right] = 0$ according to Cauchy theorem. For each iteration n, there is v such that $n_v \le n$ and $n+1 \le n_{v+1}$. Since the objective function is nonincreasing according to (23), $0 \ge \lim_{v\to+\infty}\left[g\left(\Omega^{(n+1)}\right) - g\left(\Omega^{(n)}\right)\right] \ge \lim_{v\to+\infty}\left[g\left(\Omega^{(n_{v+1})}\right) - g\left(\Omega^{(n_v)}\right)\right] \ge 0$, showing that $\lim_{v\to+\infty}\left[g\left(\Omega^{(n+1)}\right) - g\left(\Omega^{(n)}\right)\right] = 0$. Therefore, the convergent subsequence $\left\{\Omega^{(n_v)}\right\}_{v=1}^{\infty}$ is the KKT point [24]. ■

C. PROPOSED ALGORITHM FOR SOLVING THE DCR PROBLEM

The details of proposed algorithm for solving the DCR problem are briefly described in Algorithm 1. We firstly initialize the decision slot, the offloading strategies of all users, and the original point from the feasible solution of (23), respectively, and then determine whether these values satisfy the constraint conditions of the loop. In each iteration, collect information about the game environment and computes the current wireless channel interference, all users calculate the corresponding utility function, determine whether update their own offloading strategies to grantee the delay minimization by the given previous resource allocation strategy, and will then broadcast these messages to the wireless network. Users will stick to their original offloading decisions $s_i^{(t)} = s_i^{(t-1)}$ if they do not update their offloading decisions. Finally, the VBSs traverse all MBS to solve convex program (23) to obtain the current optimal resource allocation strategies by the given current offloading strategies. All users reach a mutual NE state after a finite number of iterations because the potential game admits an NE within finite improvements, i.e. no user can increase their own revenue by updating strategies without reducing the revenue of other users. This indicates that the potential game reaches an NE, and the algorithm declares the end. At each iteration of Algorithm 1, the computation complexity is polynomial with respect to the subproblems (11) and (23). For the subproblem (11), the time complexity is $O(B)$ based on step 6 of Algorithm 1. The time complexity of the subproblem (23) is $O\left(A^2B^3\right)$ according to [29].

IV. PERFORMANCE EVALUATION

In this section, we perform simulations to evaluate the performance of proposed scheme. The simulation parameter settings are as follows. An area with coverage radius of 1000 m is used to simulate a city block and surrounding streets in which $U = 30$ users and $A = 10$ VBSs are uniformly distributed surrounding streets and on bidirectional roads respectively. There are also $B = 5$ MBSs, each of which is equipped with an MEC, to provide computational

$$P2.3: \min_{\Omega} \tilde{g} = \sum_{i\in\mathcal{U}}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} \frac{z_{i,a}}{d_i} + \frac{q_{i_a,b}}{d_i} + \frac{f_{i_a,b}}{c_i}$$

$$\text{s.t. } W^B \sum_{b\in\mathcal{B}} \frac{\left(n_{i_a,b}+e_{i_a,b}\right)^2 - \left(n_{i_a,b}-e_{i_a,b}\right)^2}{4} \geq W^A \sum_{i\in\mathcal{U}} \left\| \sqrt{\frac{d_{i,a}}{2m_{i,a}}} m_{i,a}, \sqrt{\frac{m_{i,a}}{2d_{i,a}}} d_{i,a} \right\|, \forall a \in \mathcal{A}, \tag{23a}$$

$$\left\| \sqrt{\frac{m_{i,a}}{2p_i}} p_i, \sqrt{\frac{p_i}{2m_{i,a}}} m_{i,a} \right\| g_{i,a} \leq \frac{\left(k_{i,a}+p_i\right)^2 - \left(k_{i,a}-p_i\right)^2}{4}, \forall i \in \mathcal{U}, a \in \mathcal{A}, \tag{23b}$$

$$\left\| \sqrt{\frac{n_{i_a,b}}{2p_{i_a}}} p_{i_a}, \sqrt{\frac{p_{i_a}}{2n_{i_a,b}}} n_{i_a,b} \right\| g_{i_a,b} \leq \frac{\left(o_{i_a,b}+v_{i_a,b}\right)^2 - \left(o_{i_a,b}-v_{i_a,b}\right)^2}{4}, \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \tag{23c}$$

$$\log\left(\sum_{i\in\mathcal{U}}\sum_{a\in\mathcal{A}} p_i^{(n)} g_{i,a} + N_0\right) + \frac{\sum_{i\in\mathcal{U}}\sum_{a\in\mathcal{A}}\left(p_i - p_i^{(n)}\right) g_{i,a}}{\sum_{i\in\mathcal{U}}\sum_{a\in\mathcal{A}} p_i^{(n)} g_{i,a} + N_0} ged_{i,a}$$
$$+ \log\left(\sum_{i'\in\mathcal{U}\setminus\{i\}}\sum_{a'\in\mathcal{A}\setminus\{a\}} p_{i'} g_{i',a'} + N_0\right), \forall i \in \mathcal{U}, a \in \mathcal{A}, \tag{23d}$$

$$p_{i_a} g_{i_a,b} \leq \frac{\left(v_{i_a,b}+w_{i_a,b}\right)^2 - \left(v_{i_a,b}-w_{i_a,b}\right)^2}{4}, \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B} \tag{23e}$$

$$k_{i,a} \leq m_{i,a} 2^{z_{i,a}/m_{i,a}W^A} - m_{i,a} \forall i \in \mathcal{U}, a \in \mathcal{A}, \tag{23f}$$

$$l_{i,a} \leq \sum_{i'\in\mathcal{U}\setminus\{i\}}\sum_{a'\in\mathcal{A}\setminus\{a\}} p_{i'} g_{i',a'} + N_0 \forall i \in \mathcal{U}, a \in \mathcal{A}, \tag{23g}$$

$$w_{i_a,b} \leq \left(2^{e_{i_a,b}} - 1\right) \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \tag{23h}$$

$$v_{i_a,b} \leq \sum_{i'\in\mathcal{U}\setminus\{i\}}\sum_{a'\in\mathcal{A}\setminus\{a\}}\sum_{b'\in\mathcal{B}\setminus\{b\}} y_{i'_{a',b'}} p_{i'_{a'}} g_{i'_{a',b'}} + N_0 \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \tag{23i}$$

$$o_{i_a,b} \leq n_{i_a,b} 2^{q_{i_a,b}/n_{i_a,b}W^B} - n_{i_a,b} \forall i \in \mathcal{U}, a \in \mathcal{A}, b \in \mathcal{B}, \tag{23j}$$

$$\text{(17b), (17c), (17d), (17e)}, \tag{23k}$$

Algorithm 1 Proposed Algorithm for Solving the DCR Problem

Initialization: Set initial decision slot: $t = 0$; each user $i \in \mathcal{U}$ chooses the initial offloading decision that offloads its task to a edge computation server; the original point $\left\{\mathbf{z}^{(0)}, \mathbf{q}^{(0)}, \mathbf{d}^{(0)}, \mathbf{e}^{(0)}, \mathbf{k}^{(0)}, \mathbf{l}^{(0)}, \mathbf{o}^{(0)}, \mathbf{v}^{(0)}, \mathbf{w}^{(0)}\right\}$ obtained from (23);

for each slot t **do**
  get the current game environment and calculate the current wireless channel interference; given resource allocation strategy $\left\{\mathbf{p}^{(t-1)}, \mathbf{m}^{(t-1)}, \mathbf{n}^{(t-1)}, \mathbf{f}^{(t-1)}\right\}$;
  for each user $i \in \mathcal{U}$ **do**
    Calculate the corresponding utility function $u_i\left(s_i, \mathbf{s}_{-i}\right)$;
    Select new strategy s'_i such that

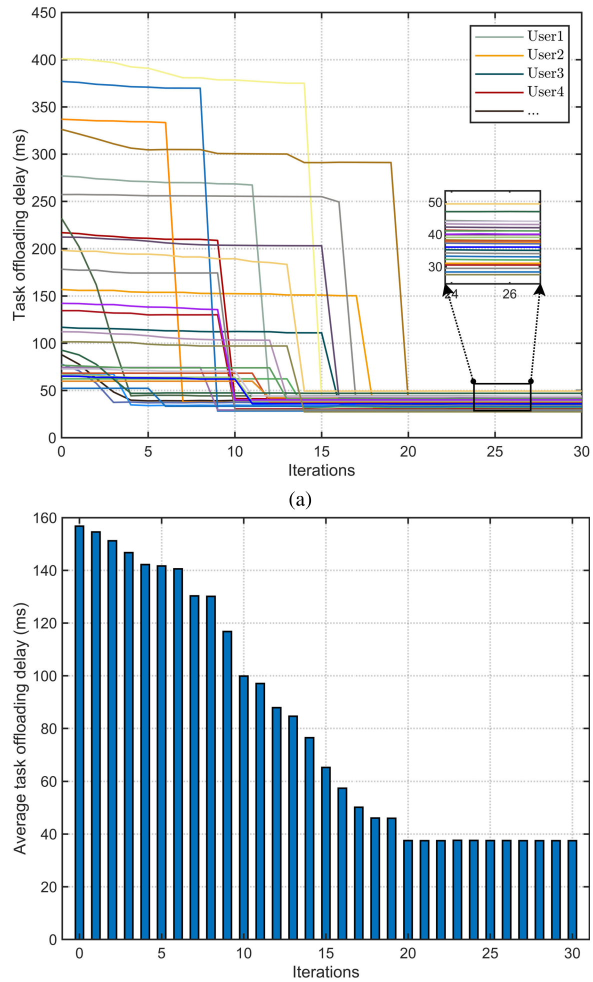
$$u_i\left(s'_i, s_{-i}\right) = \underset{a\in A, b\in B}{\arg\min} \left\{s_{i_a,b}\right\} \tag{25}$$

    if $u_i\left(s_i, s_{-i}\right) < u_i\left(s'_i, s_{-i}\right)$ **then**
      Keep the old strategy $s_i^{(t)} = s_i^{(t-1)}$ unchanged and report $\{s_i, \mathbf{p}\}$ to the network;
    else
      Update the new strategy $s_i^{(t)} = s'_i$ and report $\left\{s'_i, \mathbf{p}\right\}$ the update to the network;
    end if
  end for
  given offloading strategy $\mathbf{s}^{(t)}$;
  for each VBS $a \in \mathcal{A}$ **do**
    for each MBS $b \in \mathcal{B}$ **do**
      obtain the optimal resource allocation strategy $\left\{\mathbf{p}^{(t)}, \mathbf{m}^{(t)}, \mathbf{n}^{(t)}, \mathbf{f}^{(t)}\right\}$ by solving convex program (23);
    end for
  end for
end for
return optimization values $\{\mathbf{s}^*, \mathbf{p}^*, \mathbf{m}^*, \mathbf{n}^*, \mathbf{f}^*\}$;

TABLE 2. System parameters.

Symbol	Parameter	Value
c_i	Computation cycles of task T_i	520 cycles/bit
d_i	Offloading data size of task T_i	300 KB bytes
W^A	Wireless access channel bandwidth	10 MHz
W^B	Wireless backhaul channel bandwidth	20 MHz
$f_{i_a,b}$	Computing resource of edge computing	80 GHz
v	Velocity of vehicles	$2-20$ m/s
N_0	Noise power	-114 dBm
f_c	Carrier Frequency	3.6 GHz
$t_i^{\max}$	Delay constraint	$0.1-2$ s
$p_i^{\max}$	Maximum power of user	23 dBm
$p_{i_a}^{\max}$	Maximum power of VBS	33 dBm
α	Path-loss exponent	-3.4

resources. The other simulation parameters are listed in Table 2. The following simulation results are obtained on a computer with Intel Core i7 14700F 16-core 3.4 GHz CPU and 64 GB memory.



(a)

(b)

FIGURE 2. Convergence performance of proposed algorithm; (a) Dynamics of task delay. (b) Dynamic of average task delay.

A. CONVERGENCE EVALUATION

Fig. 2 shows the convergence performance of the proposed algorithm. The curves in Fig. 2(a) represent the offloading delay of 30 users located in the network. At the beginning iteration, the offloading delay of some users is very high, e.g., user 7 has an offloading delay of 408 ms, which is due to the fact that these users randomly selected VBSs or MBSs with poor channel quality during initialization process. As the number of iterations increases, the offloading delay shows a decreasing trend, i.e., the selected offloading strategy of each user satisfies the lower computational delay requirement with the iteration of the proposed algorithm. After 22 iterations, all the users' offloading delay finally achieves the lowest value and remains unchanged, which indicated that none of users in the network can decrease their offloading delay by changing their strategies. The Fig. 2(b) shows the dynamic of average task delay in the network for the proposed algorithm.

In order to evaluate the delay performance, the proposed scheme compares with four baselines as follows.

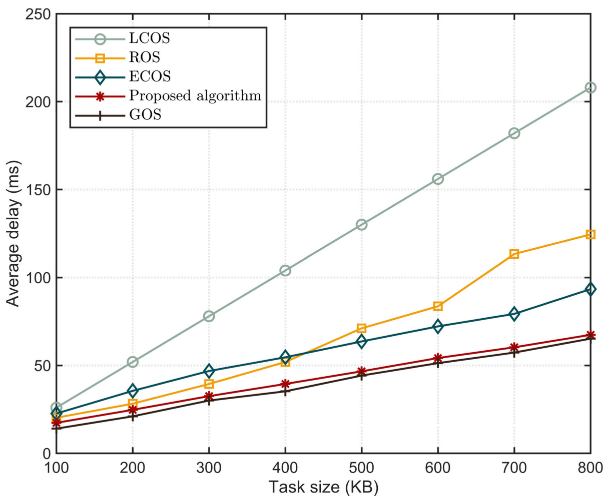


FIGURE 3. Comparison of the average delay with different task sizes.

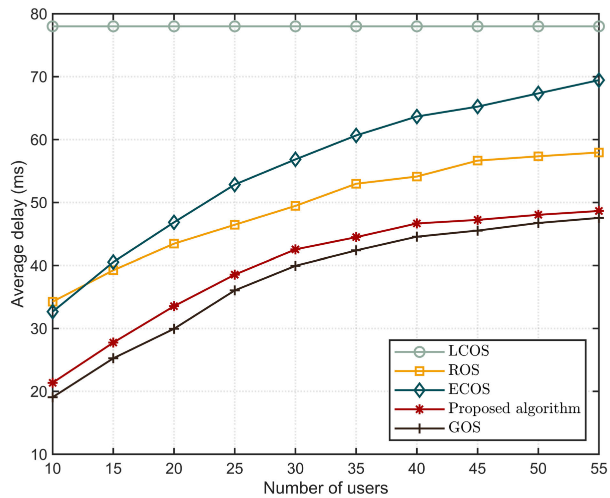


FIGURE 4. Comparison of the average delay with different number of users.

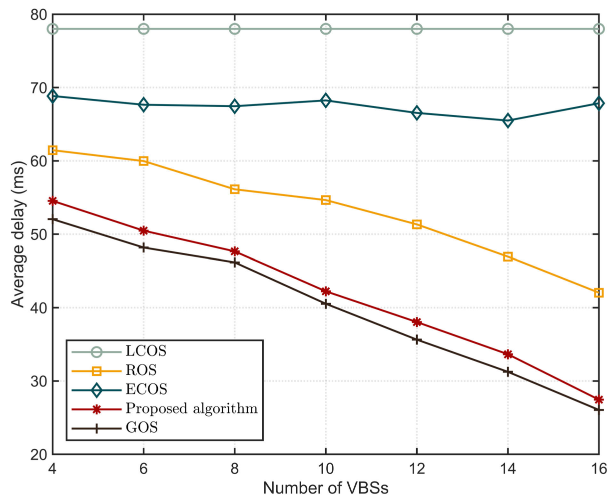


FIGURE 5. Comparison of the average delay with different number of VBSs.

- *Edge Computation Server Only Scheme* (ECOS): each user offloads its tasks directly to the edge computation server of MBSs without forwarding these via VSBs, then the offloading strategies $\mathbf{s} = \left[x_{i,b}\right]_{i\in\mathcal{U},b\in B} \in \{0, 1\}^{U\times B}$ and resource allocation strategies $\{\mathbf{p}, \mathbf{f}\}$ are determined according to the method in Section III.
- *Local Computation Only Scheme* (LCOS): users' computation tasks are only processed by their own computation capacities without offloading these to edge computation server, the computing resource of local computation is defined 2GHz.
- *Random Offloading Scheme* (ROS): a user randomly chooses a VBS as relay to forward its computation task to a random MBS.
- *Globally Optimal Solution* (GOS): the exhaustive search is used to obtain all the possible user association decisions, and then get the optimal resource power allocation for each possible case by the Lagrangian dual method.

We investigate the impact of computation task size on the delay performance using the network average delay, defined as the average delay of offloading delay of all users located in the network. As shown in Fig. 3, the network average delays of all the investigated schemes increase with the number of the task size. Particularly, the GOS has the lowest average delay. The average delay performance of the proposed algorithm matches the GOS and significantly outperform the other competing schemes. This is because each user minimizes its own offloading delay by jointly optimizing wireless access and backhaul task offloading decision, and resource allocation in terms of transmit power, bandwidth and frequency allocation. In this case, the LCOS has worst delay performance which indicates that limited computation resources increase the computation delay and may not get much benefit.

Fig. 4 shows the comparison of the impact of users' number on the average latency under different schemes. We can see that the average delay of other schemes increase with the growth of users' number except for the LCOS scheme. This is because an increase in users' number increases the amount of offloading tasks in the network, which increases the VBSs' communication delay and MBSs' computation delay leading to an increase in the average delay. However, the LCOS scheme does not use the communication/computation resource, so the average delay remains constant. For the proposed algorithm, it is close to the GOS scheme. It indicates that an increase in the number of users has no effect on the gap between the proposed algorithm and GOS. In addition, the ROS scheme has better system utility than the ECOS scheme. It also indicates that vehicle-assisted edge computing has a positive effect on the offloading tasks of users in comparison with local computing.

Fig. 5 shows the performance of average delay versus the number of VBSs for different schemes. Without the help of VBSs, the LCOS and ECOS are not affected by their number, and therefore tend to be constant with higher average

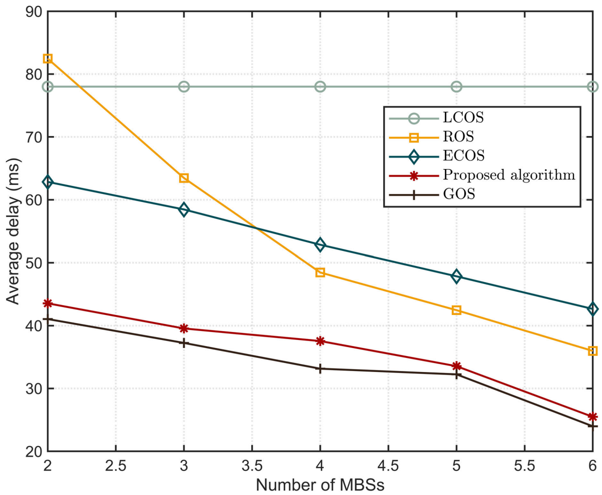


FIGURE 6. Comparison of the average delay with different number of MBSs.

delay. Additionally, we can also see that the average delay of proposed Algorithm is the closet to the GOS scheme and significantly outperform other schemes. This is due to the fact that the proposed algorithm's offloading strategy includes both radio access and backhaul components, and the offloading decisions on both links can be dynamically selected based on the number of VBSs, thus achieving NE and minimizing average delay.

Fig. 6 indicates the performance of the average latency of different schemes with respect to different number of users. We can observed that the average delay of the other schemes decreases with the increase in the MSBs' number with the exception of LCOS. The figure shows that the proposed algorithm matches the GOS, this is because the increase in the MBSs' number improves the communication capacity of the wireless backhaul link and shortens the communication distance of the wireless access link with the help of VBSs, both of which simultaneously reduce the average delay of the proposed algorithm.

V. CONCLUSION

This paper presented a vehicle-assisted edge computing model for cooperative wireless access and backhaul links. Based on this, an optimization problem combining computational offloading and resource allocation was formulated and aimed at minimizing system delay. Then, we decomposed the formulated problem into two subproblems, namely, computation offloading and resource allocation. Specifically, the computational offloading subproblem was model as a potential game with the existence and convergence of NE, in which users act as players for determining the offloading decision with joint wireless access and backhaul to achieve the NE. Furthermore, the optimal solution of resource allocation subproblem was obtained by utilizing equivalent transformation with successive convex approximation methods and satisfied KKT conditions. Finally, the simulation results have shown that our proposal has a performance advantage in terms of delay reduction compared with the benchmark schemes.

REFERENCES

[1] R. Dangi, P. Lalwani, G. Choudhary, I. You, and G. Pau, "Study and investigation on 5G technology: A systematic review," *Sensors*, vol. 22, no. 1, p. 26, Dec. 2021.
[2] Y. Siriwardhana, P. Porambage, M. Liyanage, and M. Ylianttila, "A survey on mobile augmented reality with 5G mobile edge computing: Architectures, applications, and technical aspects," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1160–1192, 2nd Quart., 2021.
[3] T. Petrov, P. Pocta, and T. Kovacikova, "Benchmarking 4G and 5G-based cellular-V2X for vehicle-to-infrastructure communication and urban scenarios in cooperative intelligent transportation systems," *Appl. Sci.*, vol. 12, no. 19, p. 9677, Sep. 2022.
[4] V. Maglogiannis, D. Naudts, S. Hadiwardoyo, D. van den Akker, J. Marquez-Barja, and I. Moerman, "Experimental V2X evaluation for C-V2X and ITS-G5 technologies in a real-life highway environment," *IEEE Trans. Netw. Service Manage.*, vol. 19, no. 2, pp. 1521–1538, Jun. 2022.
[5] J. Zhao, X. Gai, and X. Luo, "Performance comparison of vehicle networking based on DSRC and LTE technology," in *Proc. Int. Conf. Intell. Transp. Eng.* Singapore: Springer, 2021.
[6] T. Petrov, L. Sevcik, P. Pocta, and M. Dado, "A performance benchmark for dedicated short-range communications and LTE-based cellular-V2X in the context of vehicle-to-infrastructure communication and urban scenarios," *Sensors*, vol. 21, no. 15, p. 5095, Jul. 2021.
[7] M. Karoui, A. Freitas, and G. Chalhoub, "Performance comparison between LTE-V2X and ITS-G5 under realistic urban scenarios," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–7.
[8] I. W. Ho, S. C. Chau, E. R. Magsino, and K. Jia, "Efficient 3D road map data exchange for intelligent vehicles in vehicular fog networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3151–3165, Mar. 2020.
[9] Q. Luo, C. Li, T. H. Luan, and W. Shi, "Minimizing the delay and cost of computation offloading for vehicular edge computing," *IEEE Trans. Services Comput.*, vol. 15, no. 5, pp. 2897–2909, Sep. 2022.
[10] H. Guo, J. Liu, J. Ren, and Y. Zhang, "Intelligent task offloading in vehicular edge computing networks," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 126–132, Aug. 2020.
[11] C. Tang, G. Yan, H. Wu, and C. Zhu, "Computation offloading and resource allocation in failure-aware vehicular edge computing," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 1877–1888, Feb. 2024.
[12] C. Tang and H. Wu, "Optimal computational resource pricing in vehicular edge computing: A Stackelberg game approach," *J. Syst. Archit.*, vol. 121, Dec. 2021, Art. no. 102331.
[13] S. Wang, D. Ye, X. Huang, R. Yu, Y. Wang, and Y. Zhang, "Consortium blockchain for secure resource sharing in vehicular edge computing: A contract-based approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1189–1201, Apr. 2021.
[14] X. Zhang, M. Peng, S. Yan, and Y. Sun, "Joint communication and computation resource allocation in fog-based vehicular networks," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13195–13208, Aug. 2022.
[15] X. Gu and G. Zhang, "Energy-efficient computation offloading for vehicular edge computing networks," *Comput. Commun.*, vol. 166, pp. 244–253, Jan. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140366420320168
[16] W. Fan, Y. Su, J. Liu, S. Li, W. Huang, F. Wu, and Y. Liu, "Joint task offloading and resource allocation for vehicular edge computing based on V2I and V2V modes," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 4, pp. 4277–4292, Apr. 2023.
[17] Y. Zhang, X. Qin, and X. Song, "Mobility-aware cooperative task offloading and resource allocation in vehicular edge computing," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2020, pp. 1–6.
[18] W. Feng, S. Lin, N. Zhang, G. Wang, B. Ai, and L. Cai, "Joint C-V2X based offloading and resource allocation in multi-tier vehicular edge computing system," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 432–445, Feb. 2023.
[19] J. Gao, Z. Kuang, J. Gao, and L. Zhao, "Joint offloading scheduling and resource allocation in vehicular edge computing: A two layer solution," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3999–4009, Mar. 2023.

[20] Y. Ju, Y. Chen, Z. Cao, L. Liu, Q. Pei, M. Xiao, K. Ota, M. Dong, and V. C. M. Leung, "Joint secure offloading and resource allocation for vehicular edge computing network: A multi-agent deep reinforcement learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 5, pp. 5555–5569, May 2023.

[21] J. Chen, Z. Chang, X. Guo, R. Li, Z. Han, and T. Hämäläinen, "Resource allocation and computation offloading for multi-access edge computing with fronthaul and backhaul constraints," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8037–8049, Aug. 2021.

[22] Y. Peng, X. Tang, Y. Zhou, J. Li, Y. Qi, L. Liu, and H. Lin, "Computing and communication cost-aware service migration enabled by transfer reinforcement learning for dynamic vehicular edge computing networks," *IEEE Trans. Mobile Comput.*, vol. 23, no. 1, pp. 257–269, Nov. 2024.

[23] D. Monderer and L. S. Shapley, "Potential games," *Games Econ. Behav.*, vol. 14, no. 1, pp. 124–143, May 1996.

[24] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[25] *Further Advancements for E-UTRA Physical Layer Aspects*, document 3GPP TR 36.814, Version 2, 2010.

[26] J. Park, H.-H. Choi, and J.-R. Lee, "Flocking-inspired transmission power control for fair resource allocation in vehicle-mounted mobile relay networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 754–764, Jan. 2019.

[27] S. Burer and A. N. Letchford, "Non-convex mixed-integer nonlinear programming: A survey," *Surv. Oper. Res. Manage. Sci.*, vol. 17, no. 2, pp. 97–106, Jul. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1876735412000037

[28] Q. D. Lã, Y. H. Chew, and B.-H. Soong, *Potential Game Theory*. Cham, Switzerland: Springer, 2016.

[29] M. Razaviyayn, "Successive convex approximation: Analysis and applications," Ph.D. dissertation, Univ. Minnesota, 2014.

[30] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optim.*, vol. 47, no. 1, pp. 29–51, May 2010.

[31] G. Scutari, F. Facchinei, and L. Lampariello, "Parallel and distributed methods for constrained nonconvex optimization—Part I: Theory," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 1929–1944, Apr. 2017.

BO HUANG received the Ph.D. degree from the School of Electronics and Information Engineering, Tongji University, Shanghai, China, in 2020. He is currently working as a Lecturer with the School of Electronic Information Engineering, Suzhou Vocational University. His research interests include wireless communication, wireless backhaul, stochastic geometry, and mobile edge computing.

YAN ZHOU received the B.Sc. and M.Sc. degrees from the China University of Geosciences, in June 2003 and June 2006, respectively, and the Ph.D. degree in signal and information processing from the School of Electronics and Information Engineering, Soochow University, Suzhou, China. Since 2006, she has been with the Department of Communication Technology, Electronic Information Engineering College, Suzhou Vocational University. Her current research interests include wireless communication, speech signal processing, and intelligent computing.

XU ZHANG received the Ph.D. degree from the University of Science and Technology of China, Hefei, in 2016. He is currently an Associate Professor with the College of Electronic Information Engineering, Suzhou Vocational University. His research interests include signal processing, computer vision, machine learning, and pattern recognition.

JIE CHEN received the master's degree from Zhejiang University, Hangzhou, China. He is currently working as an Associate Professor with the School of Electronic Information Engineering, Suzhou Vocational University, Suzhou, China. His research interests include signal processing, image processing, and pattern recognition.

LI SHANG received the Ph.D. degree in pattern recognition and intelligent system from the University of Science and Technology of China (USTC), Hefei, China, in June 2006. From September 1999 to July 2006, she worked at the Department of Automation, USTC. Since August 2006, she has been with the College of Electronic information engineering, Suzhou Vocational University, Jiangsu, China. She is currently a Professor. Her research interests include image processing, pattern recognition, artificial neural networks, and intelligent computing.