IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 October 2022, accepted 28 October 2022, date of publication 7 November 2022, date of current version 28 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3220251

RESEARCH ARTICLE

Joint Communication, Computation, and Control for Computational Task Offloading in Vehicle-Assisted Multi-Access Edge Computing

TRI NGUYEN DANG[1], AUNAS MANZOOR[1], YAN KYAW TUN[1], (Member, IEEE), S. M. AHSAN KAZMI[1,3], RIM HAW[1], (Member, IEEE), SANG HOON HONG[2], ZHU HAN[1,4], (Fellow, IEEE), AND CHOONG SEON HONG[1], (Senior Member, IEEE)

[1]Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, Republic of Korea
[2]Department of Electronic Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, Republic of Korea
[3]Department of Computer Science and Creative Technologies, University of the West of England, BS16 1QY Bristol, U.K.
[4]Electrical and Computer Engineering Department, University of Houston, Houston, TX 77004, USA

Corresponding author: Choong Seon Hong (cshong@khu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government [Ministry of Science and Information and Communications Technology (MSIT)] under Grant 2020R1A4A1018607, and in part by the Institute for Information and Communication Technology Planning and Evaluation (IITP) Grant funded by the Korea Government (MSIT) (Evolvable Deep Learning Model Generation Platform for Edge Computing) under Grant 2019-0-01287.

ABSTRACT Future generation of Electric Vehicles (EVs) equipped with modern technologies will impose a significant burden on computation and communication to the network due to the vast extension of onboard infotainment services. To overcome this challenge, multi-access edge computing (MEC) or Fog Computing can be employed. However, the massive adoption of novel infotainment services such as Augmented Reality, Virtual Reality-based services will make the MEC and Fog resources insufficient. To cope with this issue, we propose a system model with onboard computation offloading, where an EV can utilize its neighboring EVs resources that are not resource-constrained to enhance its computing capacity. Then, we propose to solve the problem of computational task offloading by jointly considering the communication, computation, and control in a mobile vehicular network. We formulate a mixed-integer non-linear problem (MINLP) to minimize the trade-off between latency and energy consumption subject to the network resources and the mobility of EVs. The formulated problem is solved via the block coordination descent (BCD) method. In such a way, we decompose the original MINLP problem into three subproblems which are resource block allocation (RBA), power control and interference management (PCP), and offload decision problem (ODP). We then alternatively obtain solutions of RBA and PCP via the duality theory, and the third sub-problem is solvable via the relaxation method and alternating direction Lagrangian multiplier method (ADMM). Numerical results reveal that the proposed solution BCD-based algorithm performs a fast convergence rate.

INDEX TERMS Multi-access edge computing (MEC), collaborative V2Vs-assisted MEC system, tasks offloading, resource allocation, alternating direction method of multipliers (ADMM), interference management, V2V communication.

I. INTRODUCTION

Next-generation vehicles will be equipped with advanced computing, caching, communications, and control resources to meet the stringent requirements of safe driving, automation, and infotainment services. By connecting these resourceful vehicles with the wireless network to devise the Internet of Vehicles (IoV), an efficient, cost-effective, safe, autonomous, and intelligent transportation system will be created. However, such massive connectivity with wireless networks is non-trivial and may overwhelm the network. Moreover, meeting the coverage, communication,

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

and computation demands of a massive number of vehicles in a wide geographical area is challenging. Even though the deployment of road side units (RSUs) has contributed well to alleviating the network traffic load in congested locations, they lack to provide coverage to all roads in an area. To address this problem, an independent, cellular coverage-free network of vehicles is required where the vehicles can collaborate, share, trade, and exploit redundant resources efficiently. Indeed, neighboring vehicles sharing the road vicinity can collaborate, cooperate, and share their redundant computation, power, and caching resources for their mutual benefit. The collaboration among the vehicles in a road segment uncovered by the cellular networks becomes significant to offload, process, and retrieve the computation-intensive tasks. However, such resource sharing pertains to its own challenges of managing the limited resources, meeting the QoS requirements, having very short points of contact among the vehicles, and guaranteeing task completion. Therefore, a sophisticated design of collaborative sharing of resources for task completion in the absence of cellular coverage is strongly desired.

Task offloading among the moving vehicles is particularly challenging to meet the latency, task completion, and limited point of contact among vehicles. Moreover, the unavailability of resources in the neighborhood to perform the desired task makes efficient task management challenging. Another challenge of task offloading among the vehicles is the design of a suitable incentive mechanism for the vehicles performing the task. Moreover, executing the task according to the desired order is critical to streamlining the application flow of the requesting vehicle. Moreover, selecting suitable vehicles to perform the task and optimal offloading decisions becomes challenging under dynamic channel conditions and speed variations of vehicles. In addition, to meet the strict latency requirements of the tasks for vehicular networks, another challenge is to balance the task offloading under the limited energy resources. By allocating high computation and communication resources to a task, the latency requirements can be fulfilled at the cost of high energy consummation. Therefore, it is desired to design a balanced scheme for managing the trade-off between energy and latency constraints.

In the literature, a number of task management schemes for vehicular networks have been proposed. An intuitive solution can be the reduction of task sizes instead of offloading the unified larger tasks. In this way, the slices of tasks are executed at local, neighboring vehicles or the MEC servers. The authors in [1] proposed a destination selection framework from the neighboring vehicle or the MEC server to offload the task. However, the authors do not consider energy management in task offloading. The same problem of V2X task offloading is addressed by [2] where the authors proposed a task offloading scheme in the vehicular networks under the task latency requirements. The authors in [3] proposed a task partitioning scheme to offload the slices of tasks to different MEC servers such that the latency requirements are fulfilled. However, these works do not consider the energy management of vehicles in their framework. To address the challenge of task offloading under the limited energy resources in vehicular networks, the authors in [4] proposed the selection of a MEC server according to the availability of energy resources. To manage the energy-latency trade-off efficiently, the authors in [5] proposed an energy-aware task offloading scheme for vehicular networks. Similarly, the authors in [6] and [7] investigated the problem of mobility-aware task offloading and latency without the problem of interference management and the trade-off between energy and latency. Most of the aforementioned works lack involvement of latency-energy trade-off in the task execution at local, neighboring, and MEC servers. Moreover, they have not considered the power control problem and interference management for V2V communication.

To this end, we propose a hybrid collaboration on the road scheme to efficiently manage the computation and communication resources, and interference management of the vehicular networks while optimizing the task offloading decisions to local, neighboring vehicles and MEC servers. To do this, we develop a system model containing two sets of vehicles: (i) resource-constrained vehicles and (ii) vehicles that have available resources. In which, the resource-constrained vehicle generates a latency and computation-intensive task to be executed. This can be a computing task generated by passengers for using infotainment services [8], or an image processing task for road tracking of EVs [9]. Typically, these tasks can be executed locally or can be offloaded to a resourceful neighboring vehicle, and MEC server. To formulate the optimization problem, we design the mobility model of the vehicles using the kinematic equations. Then, the corresponding communication, computation, offloading, and latency models are designed. We formulate the optimization problem with the objective of a trade-off between the latency and energy of the vehicular network. The optimization problem is solved through the block coordinate descent technique (BCD) by decomposing the problem into three subproblems. To the best of our knowledge, this is the first paper that considers joint communication, computation, and control in terms of latency-energy trade-off, including power control and interference management for V2V communication and computing task offloading in EVs-assisted MEC. The following is a summary of our main contributions:

- We propose a system model of Electric Vehicles (EVs)-assisted MEC for task offloading in which EVs are allowed to offload the computing task either to the MEC server or close vicinity EVs, which have the available resources to process the task.
- We formulate the joint optimization problem w.r.t. the trade-off between latency and energy consumption. In which we consider the problem of communication resource allocation, interference management, and task offload decision. The formulated problem falls into Mixed-Integer Non-Linear Programming (MINLP) category. It is intractably NP-hard and time-consuming for a large and practical scale setting.

- By employing the Block Coordination Descent (BCD) technique, we then decompose the formulated problem into three subproblems named Resource Block Allocation problem (RBA), Power Control problem (PCP) for Vehicle-to-Vehicle (V2V) communication and EVs offload decision problem (ODP).
- RBA problem is solvable via an alternative method named dual decomposition with duality and subgradient methods. The second subproblem of PCP is a bi-convex, and thus we employ the Lagrangian multiplier method to distributively obtain a sub-optimal solution. Furthermore, we solve the third problem via the alternating direction method of multipliers (ADMM) to obtain the optimal of the relaxed problem.
- Finally, we validate our proposed model with comprehensive numerical results. We have verified the convergence rate as well as the performance of each algorithm for solving each sub-problem.

The rest of this paper is structured as follows: Section II covers related works. Section III describes the system model. The proposed problem formulation and solution approaches are presented in Sections IV and V, respectively. Simulation results are provided in Section VI. Finally, Section VII concludes the paper.

II. RELATED WORKS

A. MULTI-ACCESS EDGE COMPUTING

The recent developments of high-performance applications have overwhelmed the resource-limited vehicular networks. A number of works tried to address this issue by proposing the task offloading to cloud servers [10], [11], [12], [13], [14], edge server, and neighboring vehicles. For instance, the authors in [10] proposed a cloud-based architecture for vehicular networks to efficiently manage the computation, storage, and spectrum resources. However, the proposed architecture covers the standard vehicular data, e.g., navigation and surveillance. The authors in [11] proposed a decentralized clustering of vehicles near the traffic signals to alleviate the load from the cloud server. However, they did not consider the edge server in their V2I architecture. The authors in [12] proposed a task scheduling scheme for vehicular clouds. They developed a polynomial-time approximation scheme for single-task scheduling. However, this work is limited to only a single task scheduling with the cloud server. The authors in [13] proposed a joint cloud and MEC-based task offloading scheme for the vehicular network. They adopted a game-theoretic approach for the task offloading decision and the Lagrange multiplier method for resource allocation. However, the authors did not consider the task of offloading to the neighboring vehicles. The authors in [14] proposed a task handover scheme to the cloud server for the vehicles, which are leaving the network without completing the task.

The cloud-based computation schemes lack to meet the latency requirements of time-intensive tasks. To address this challenge, edge server-based task offloading for vehicular networks has been proposed [15], [16], [17], [18], [19]. For instance, the authors in [15] proposed a MEC and a backup computing server-based task offloading scheme for vehicular networks. They devised a Stackelberg game for the multilevel offloading of vehicular tasks to maximize the utilities of edge servers and mobile users. The authors in [16] exploited edge computing to design a resource-sharing scheme among the vehicles. In the proposed two-stage mechanism, the resource requirements from nearby vehicles are gathered by the edge server to compute the optimal prices, and then resource allocation according to the task requirements is performed in the second stage. The authors in [17] proposed a scheme for the efficient placement of an edge server at an optimal location in an urban environment to meet the computational and latency requirements of vehicles. The authors in [18] proposed an energy management scheme for cellular users in vehicles to offload their workload to MEC. They formulated the optimization problem to minimize power consumption under latency constraints and power budgets and solved it using ADMM. However, they did not consider the allocation of computing resources for each device in their formulation. The authors in [19] devised a contract and matching-based resource allocation scheme in vehicular fog networks. They designed a contract-based incentive mechanism to increase the BS utility.

B. V2Vs-ASSISTED MULTI-ACCESS EDGE COMPUTING

A number of works have investigated the collaboration among the vehicles for content delivery and task allocation [20], [21], [22], [23], [24], [25], [26]. For instance, the authors in [20] proposed a task allocation scheme in which a generated task by a vehicle is allocated to the neighboring stationery and mobile fog nodes. They also performed experimental analysis for different mobility settings of the network. The authors in [21] proposed a probabilistic task prioritizing scheme in which the contents in the other vehicle are estimated beforehand. However, the scope of this work is limited to content sharing only. A similar content dissemination scheme is proposed in [22] where advertising content is broadcast to all vehicles in an urban area by exploiting the V2V communication. To further reap the benefits of available resources in a vehicle, a contract-based scheme is proposed in [23] to incentivize resource-sharing vehicles. The work in [24] also proposed a contract-based task allocation scheme among the vehicles where the participating vehicles are encouraged to maintain a relative acceleration to guarantee task completion. The authors in [25] proposed a communication and computation resource management scheme for autonomous vehicles. However, the scope of this work is to handle the propulsion and control of autonomous vehicles only. The authors in [26] proposed a task offloading scheme in a vehicular fog network where tasks of different priority are handled accordingly.

The works mentioned above have not investigated the power control and interference management problem for V2V communication, which significantly affects V2V, V2I

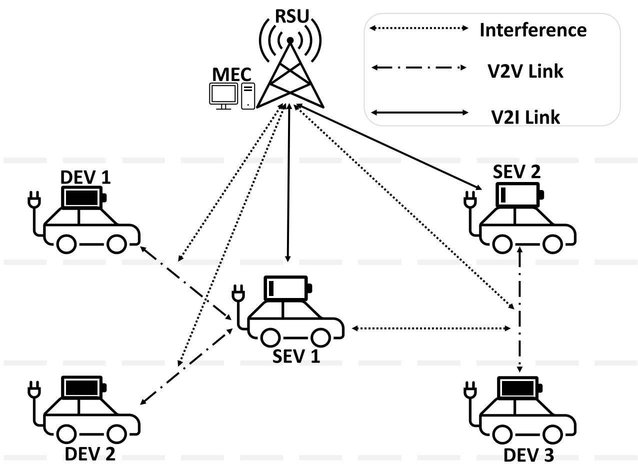

FIGURE 1. **Illustration of our system model.**

communication, and task offloading. Therefore, in this work, we combine communication which is resource block allocation, computation which is computational resource allocation, and control which is task offload decision, power control, and interference management.

III. SYSTEM MODEL

As illustrated in Fig. 1, we consider a vehicular network that consists of a MEC server and two sets of EVs; $\mathcal{O} = \{1, 2, \ldots, O\}$ is a set of resource-constrained EVs (SEVs); and $\mathcal{D} = \{1, 2, \ldots, D\}$ is a set of EVs which have available resources to assist the resource-constrained EVs (DEVs); and a Road Side Unit (RSU) enabled MEC (REC). In this work, we assume that EVs are traveling on the road under the coverage radius of the RSU. Each EV $i \in \mathcal{O}$ has a latency-sensitive computation task A_i that can be expressed by a positive tuple $A_i = (S_i, C_i, \delta_i)$, where S_i is the task size, C_i is the CPU cycles per second per bit that is requires to compute a unit data of the task, and δ_i is the worst case execution of the task. For instance, EV passengers are using onboard infotainment services, and some of the requested content is missing in its cache storage. However, it has similar content which the requested one but is stored in a different format, i.e., the requested content is $720p$, and the cached content is $360p$. Thus, EV needs to transform cached content into the requested format by using a certain method such as *transcoding*, *low-resolution*, or *supper-resolution* [8]. In this case, the size of task S_i is the size of cached content, C_i is the CPU requirements to transform the cached content to the requested one, and δ_i is the maximum latency. Due to the limited computation capacity of the EV and the latency constraint of the tasks, it is challenging for the EV to compute its tasks locally. Therefore, EVs need to offload a fraction of the task or the complete tasks to the associated RSU via wireless links to perform remote computing. However, the RSU might be overloaded due to the high number of tasks offloaded by the EVs. Therefore, we propose a hybrid offloading system model in which EVs are allowed to offload their associated tasks either to the MEC server or another close vicinity EVs that are not resource-constrained. To cope with this challenge, we propose a hybrid offloading system model in which EVs are allowed to offload their associated tasks either to the MEC server or another close vicinity EVs that are not resource-constrained.

A. EV MODEL

For simplicity we denote EV $i \in \mathcal{O}$ as source EV (SEV) and EV $j \in \mathcal{D}$ as destination EV (DEV), and the set of all EV in the system as $\mathcal{K} = \mathcal{O} \cup \mathcal{D}$. Inspired by our previous work [24], we extend the mobility modeling of EVs as follows. In this work, we assume that EVs are moving with a positive velocity denoted as $v_k, \forall k \in \mathcal{K}$, $v_k[t]$ denotes the instantaneous velocity at time t, and initial location $l_k[t_0] = \{\varphi_k[t_0],\ \phi_k[t_0]\}, \forall k \in K$, where $\varphi_k[t_0]$, and $\phi_k[t_0]$ represents the initial longitude, and latitude of EV k at time t_0, respectively. Longitude φ_k represents the location, and latitude ϕ_k represents the lane shifting of an EV k. Let $v_{i\in\mathcal{O}}[t_0],\ v_{j\in\mathcal{D}}[t_0]$ be the instantaneous velocity of SEV, and DEV at time t_0, respectively. Similarly, let $a_i[t_0]$, and $a_i[t_0]$ be the acceleration of SEV, and DEV at time t_0, respectively. Based on the kinematic equation, we can measure the location of any EV $k \in \mathcal{K}$ after some time duration $\triangle t = t - t_0$ as follows:

$$\varphi_k[\triangle t] = \varphi_k[t_0] + \frac{1}{2}\left|\frac{1}{\triangle t}\left(\int_{t_0}^{t} a_k[u|\varphi]\partial(u)\right)\right|, \quad \forall k \in \mathcal{K},$$

$$\phi_k[\triangle t] = \phi_k[t_0] + \sum_{u=t_0}^{t} \phi_k[u], \quad \forall k \in \mathcal{K}. \tag{1}$$

Based on (1), the relative distance between a pair of SEV and DEV, e.g., $i \in \mathcal{O}, j \in \mathcal{D}$ at time t is given by:

$$d_{i,j}[t] = \sqrt{(\varphi_i[t] - \varphi_j[t])^2 + (\phi_i[t] - \phi_j[t])^2}, \quad \forall i \in O, \quad \forall j \in D. \tag{2}$$

Note that, the mobility of an EV affects the performance of the offloading process due to the availability of a pair in the communication range of V2V denoted as $d_{\max}$. Therefore, it is indispensable to consider the Relative Acceleration (RA) between SEV and DEV. In this paper, we assume that SEV and DEV are moving in the same direction, and thus, RA of SEV and DEV can be modeled as follows:

$$\triangle a_{i,j}[t] = |a_j[t] - a_i[t]|, \quad \forall i \in \mathcal{O}, \quad \forall j \in \mathcal{D}. \tag{3}$$

Therefore, the relative distance between any pair of EVs after some time duration $\triangle t$ can be measured at follows:

$$d_{i,j}[\triangle t] = d_{i,j}[t_0] + \frac{1}{2} a_{i,j} t^2, \quad \forall i \in \mathcal{O}, \quad \forall j \in \mathcal{D}. \tag{4}$$

For instance, let $\triangle t$ be the offloading time period, during this time, the relative acceleration between the SEV and DEV is significantly large resulting in a larger distance between them. This will then hinder the performance of the offloading

TABLE 1. Summary of notations.

Notation	Definition	Notation	Definition
$\mathcal{O}$	Set of SEVs	$\mathcal{D}$	Set of DEVs
A_i	Offloading task of user $i \in \mathcal{O}$	S_i	Total input data size of user i's task
C_i	CPU requirement to process a bit of data in task i	δ_i	Worst case execution of task i
φ	Longitude of EV (location)	ϕ	Latitude of EV (lane shifting)
a_i	Acceleration of EV i (m/s^2)	v_i	Velocity of EV i (m/s)
$d_{i,j}$	Distance between EV i and EV j	$d_{\max}$	Maximum acceptable distance for V2V communication
Γ	Signal-to-interference-plus-noise-ratio (SINR)	p	Transmit power level
I_0	Additive Gaussian White Noise (AGWN)	g	Channel gain at the reference distance 1 m
η	Path-loss component	β	Set of Resource Blocks (RBs)
Ω_b	Set of EV that allocated in RB $b \in \beta$	R_i^b	Achievable data rate of EV i at RB b
W_b	Total available bandwidth at RB b	x_i	Decision variable of SEV $i \in \mathcal{O}$
$y_{i,j}$	Decision variable of DEV j	z_i	Decision variable of RSU
E_i^{tx}	Communication energy	E_i^{comp}	Computing energy
$H_i^{\max}$	Computational capacity of the EV i	$E_i^{\max}$	Energy capacity of EV i
T_i^{tx}	Transmission latency	θ	Computing latency
L_i	Total latency of EV i	E_i	Total energy consumption of EV i
$I_{\max}$	Interference threshold	Φ	Normalization coefficient for energy consumption

process due to violation of the maximum range of V2V communication threshold $d_{i,j}[\triangle t] > d_{\max}$. The SEV and DEV will not be able to communicate with each other, thus, the offloading service will be interrupted. Next, we present the communication model for EVs.

B. COMMUNICATION MODEL

This subsection presents our communication models that include communication between SEVs and the RSU, and the communication between SEVs and DEVs. Similar to our work in [8], [24], we consider Long-Term Evolution (LTE) and the fifth generation (5G) based communication technologies in this work. The LTE or 5G has higher stability than Dedicated short-range communications (DSRC)/IEEE802.11p in terms of bandwidth features suiting the purpose of the task offloading service as opposed to the other technologies.Then, we assume that the total communication system bandwidth W is divided into β resource blocks (RBs), each of bandwidth $W_{b\in\beta}$. To begin with, we present the communication model between the SEV and the RSU in which the SEV's task is offloaded to the RSU using the vehicle-to-infrastructure (V2I) communication followed by the resource-constrained SEV offloading its task to close vicinity DEVs via the vehicle to vehicle (V2V) communication.

1) SEV TO RSU COMMUNICATION VIA V2I

In this case, the communication takes place between a SEV and the RSU. Thus, the signal-to-interference-plus-noise ratio (SINR) of a SEV i is given by:

$$\Gamma_{i,0}^b = \frac{p_{i,0}^b g_{i,0}^b d_{i,0}^{-\alpha}}{\sum_{k\in\Omega_b, k\neq i} p_k^b g_{k,i}^b + I_0}, \tag{5}$$

where $p_{i,0}$ and $g_{i,0}$ are the transmit power, and channel power gain between SEV i and the RSU, respectively. $d_{i,0}$ is the distance between SEV i and RSU, α is path-loss coefficient. $\sum_{k\in\Omega_b, k\neq i} p_k^b g_{k,i}^b$ is the interference from other SEVs that use the same RB b, Ω_b is the set of SEVs that are allocated the same RB b. I_0 is the Additive Gaussian White Noise (AGWN). Next, we describe the scenario of SEV to DEV communication.

2) SEV TO DEV VIA V2V

In this case, the communication takes place between SEV i and DEV j. Thus, the SINR of EV pair $(i,\ j)$ is given by:

$$\Gamma_{i,j}^b = \frac{p_{i,j}^b g_{i,j}^b d_{i,j}^{-\alpha}}{\sum_{k\in\Omega_b, k\neq i} p_k^b g_{k,i}^b + p_0^b g_{0,i}^b + I_0}, \tag{6}$$

where $p_{i,j}^b$ and $g_{i,j}^b$ are the transmit power, and power channel gain between SEV i and DEV j, $d_{i,j}$ is the distance between SEV i and DEV j. $p_0^b g_{0,i}^b$ is the interference from SEV that communicate with RSU using the same RB b.

Let $\boldsymbol{\zeta}$ be the association decision variable with each element ζ_i^b representing if the RB b is either assigned to SEV i or not.

$$\zeta_i^b = \begin{cases} 1, & \text{if RB } b \text{ is allocated to SEV } i\ , \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Consequently, the achievable data rate of SEV i under both scenarios can be formulated as follows:

$$R_i^b = \begin{cases} W_b \log_2(1+\Gamma_{i,0}^b)\zeta_i^b, & \text{V2I scenario}, \\ W_b \log_2(1+\Gamma_{i,j}^b)\zeta_i^b, & \text{V2V scenario}. \end{cases} \tag{8}$$

Note that, we consider the relative distance between EV at initial slot $t = 0$, and resulting time slot $t = \tau$, thus, we omit the time slot subscription in (5), (6), and (8). Next, we present the computing model of our work.

C. COMPUTING MODEL

In this section, we present the computation model. In our work, the computation of a task can be performed either locally at the SEV or remotely by offloading the task. Next, we present both scenarios of computations.

1) LOCAL COMPUTING MODEL

In our model, we assume that each SEV $i \in \mathcal{O}$ is also equipped with a small computing server that can provide some light computing services. Let x_i be the decision variable indicating whether SEV i can execute the task locally or not.

$$x_i = \begin{cases} 1, & \text{if SEV } i \text{ is executing the task,} \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Moreover, we assume that the computational resource of the SEV is also limited and represented by $H_i^{\max}$ (CPU cycle per second). Thus, the EV has to guarantee the efficiency of computing capacity abiding by the following constraint:

$$C_i S_i \leq h_i x_i \leq H_i^{\max}, \quad \forall i \in \mathcal{O}. \tag{10}$$

2) REMOTE COMPUTING MODEL

In the remote computation model, we consider a hybrid offloading in which SEV $i \in \mathcal{O}$ is either offloading to DEV $j \in \mathcal{D}$ or MEC server. Therefore, we have two scenarios of offloading: i) V2V offloading; ii) V2I offloading. Next, we define each scenario of offloading.

a: V2V OFFLOADING MODEL

Let $y_{i,j}$ be the offloading decision variable of SEV i to DEV j.

$$y_{i,j} = \begin{cases} 1, & \text{if DEV } j \text{ executes the task of SEV } i, \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

In our model, if the computational resources of SEV i is insufficient, it can offload the task to close vicinity DEVs via V2V links. On the other hand, the DEV $j \in \mathcal{D}$ also has a limit on computing and energy capacity. Thus, DE j also needs to guarantee the computational capacity $H_j^{\max}$ and energy capacity $E_j^{\max}$ in order to handle the offloading task of SEV i. Then, the following constraints need to be considered for successful task offload and execution. For the energy constraint:

$$\sum_{i \in \mathcal{O}} \kappa (h_{i,j})^2 y_{i,j} \leq E_j^{\max}, \quad \forall i \in \mathcal{O}, \quad \forall j \in \mathcal{D}, \tag{12}$$

where $h_{i,j}$ is the computational resources that DEV j allocates to process the computing task of SEV i. Similarly for computational capacity constraint:

$$\sum_{i \in \mathcal{O}} C_i S_i y_{i,j} \leq \sum_{i \in \mathcal{O}} h_{i,j} y_{i,j} \leq H_j^{\max}, \quad \forall j \in \mathcal{D}. \tag{13}$$

b: V2I OFFLOADING MODEL

In this case, SEV i is offloading computing task A_i to the RSU. We assume that the RSU has to handle all of the offloading tasks of EVs. However, the latency depends on the arrival rate of the demand. The higher the demand, the more increase in latency. Moreover, the RSU is using grid electricity, therefore, we do not have the energy constraint in the case of the V2I offloading model. Then, let z_i be the decision variable for the computing task of SEV i.

$$z_i = \begin{cases} 1, & \text{if RSU executes the task of SEV } i, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

The RSU's computational capacity constraint can be formulated as follows:

$$\sum_{i \in \mathcal{O}} h_{0,i} z_{0,i} \leq H_0^{\max}. \tag{15}$$

Next, we discuss the energy and latency models in the following subsections.

D. ENERGY MODEL

Firstly, we consider the case of local energy consumption of SEV i. In the case of processing its local computing, SEV i needs to consider its energy capacity constraint given by:

$$E_i^{\text{comp}} = \kappa (h_i)^2 x_i \leq E_i^{\max}, \quad \forall i \in \mathcal{O}, \tag{16}$$

where $\kappa = 5.0 \times 10^{-27}$ is a constant which depends on the chip architecture of the server at SEV, h_i is the computational resource that SEV i allocates to process the task A_i. And, the required transmission energy when SEV i offloads its task to DEV j is given by:

$$E_{i,j}^{\text{tx}} = p_{i,j} \frac{S_i}{R_{i,j}^b} y_{i,j}, \quad \forall j \in \mathcal{D}. \tag{17}$$

Similarly, the energy consumption when SEV i offloads its task to the RSU is given by:

$$E_{i,0}^{\text{tx}} = p_{i,0} \frac{S_i}{R_{i,0}^b} z_i, \quad \forall i \in \mathcal{O}. \tag{18}$$

Next, we define the energy consumption of DEV j. The total energy consumption of DEV j is calculated based on the total computing task of SEVs that are being executed at its server. Thus, the energy consumption of DEV j is given by:

$$E_j^{\text{comp}} = \sum_{i \in \mathcal{O}} \kappa (h_{i,j})^2 y_{i,j}, \quad \forall j \in \mathcal{D}. \tag{19}$$

Consequently, the energy constrain for any EV $k \in \mathcal{K}$ is given by

$$E_k^{\text{comp}} + E_k^{tx} + E_k^{\text{base}} \leq E_k^{\max}, \tag{20}$$

where E_k^{base} is the base load energy of EV k.

E. LATENCY MODEL

In this work, we assume that the output of a computing task in terms of task size is always smaller compared with its input size. Therefore, it is negligible in terms of latency impact on feedback the results from DEV to SEV [24]. The latency in our model considers both transmission and computation latency. Communication latency of SEV i:

$$L_i^{\text{tx}} = \sum_{j \in \mathcal{D}} \frac{S_i}{R_i^b} y_{i,j} + \frac{S_i}{R_i^b} z_i, \quad \forall i \in \mathcal{O}, \quad \forall j \in \mathcal{D}. \tag{21}$$

Let L_i^{comp} be the latency of local computing of SEV i:

$$L_i^{\text{comp}} = \frac{S_i C_i}{h_i} x_i, \quad \forall i \in \mathcal{O}, \tag{22}$$

Similarly, the latency for remote computing at neighbor DEV be given as: $L_{i,j}^{\text{comp}}$ is given by:

$$L_{i,j}^{\text{comp}} = \sum_{j \in \mathcal{D}} \frac{S_i C_i}{h_i} y_{i,j}, \quad \forall i \in \mathcal{O}. \tag{23}$$

If the computational resources of SEVs and DEVs are insufficient. The task is forwarded to the RSU. Let $L_{i,0}^{\text{comp}}$ be the computing latency at RSU. $L_{i,0}^{\text{comp}}$ is represented as follows:

$$L_{i,0}^{\text{comp}} = \frac{S_i C_i}{h_i} z_{i,0}, \quad \forall i \in \mathcal{O}. \tag{24}$$

Consequently, the total latency of a computing task of SEV i can be formulated using the following:

$$L_i = L_i^{\text{tx}} + L_i^{\text{comp}} + \sum_{j \in \mathcal{D}} L_{i,j}^{\text{comp}} + L_{i,0}^{\text{comp}}, \quad \forall i \in \mathcal{O}. \tag{25}$$

On the other hand, we also consider the mobility of SEVs and DEVs in order to guarantee the success of offloading services while moving. Therefore, the latency is bound to either the worst-case execution of offloading task or the time that SEV and DEV are moving out of the acceptable communication range $d_{\max}$. Consequently, let $\tau_{i,j}$ be the maximum latency threshold for EV pair i, j. $\tau_{i,j}$ can be modeled as follows:

$$\tau_{i,j} = \min\{\delta_i, \frac{2d_{\max}}{a_{i,j}}\}, \quad \forall i \in \mathcal{O}, \quad \forall j \in \mathcal{D}. \tag{26}$$

Intuitively, the V2V offloading latency is bounded using the following constraints:

$$L_{i,j} \triangleq \sum_{j \in \mathcal{D}} \left(\frac{S_i}{R_i^b} + \frac{S_i C_i}{h_i} \right) y_{i,j} \leq \tau_{i,j}. \tag{27}$$

As shown in (26), if the relative acceleration between SEV i and DEV j is too high, thus, the minimum latency is strictly low. Consequently, DEV j must allocate more computational resources to guarantee the feasibility of latency constraint in (27). As a result, total energy consumption is significantly increasing. Therefore, we consider mobility as a constraint in our optimization problem to guarantee the feasibility of our solution. It must be noted that, in practice, an EV's velocity varies over time, which might lead to the problem of interruption due to inter EV's distance exceeding the maximum acceptable range $d_{\max}$ for V2V communication. One viable solution is that both SEV and DEV can synchronously increase or decrease their velocity. This has been done in our previous work in [24], where the RSU offers a reward to motivate SEV and DEV synchronously take the same action, such as increase or decrease velocity to keep the inner distance at an acceptable value. By doing so, we can avoid interruption during the offloading process.

Next, we present our problem formulation and proposed solution approach.

IV. PROBLEM FORMULATION

The goal of task offloading is typically to minimize the total latency and energy consumption. However, there is a conflict between latency and energy consumption. For instance, to minimize latency, we need to reduce transmission latency by increasing transmit power to achieve a higher data rate; and thus, it is increasing the total energy consumption. Similarly, reducing computation latency by allocating more computational resources to process the offloading tasks increases energy for computing. To overcome this challenge, we propose an objective function that is a trade-off between energy consumption and latency. The detail of the objective function is presented in (28), as shown at the bottom of the page, where ψ is a trade-off coefficient, and Φ is the normalization parameter due to the different scale of latency in milliseconds (ms) and energy in milliwatt (mW).

To the best of our knowledge, this is the first study to take into account the trade-off of latency and energy consumption minimization problem in the collaborative EVs-assisted RSU-enabled MEC server, by jointly optimizing offloading decisions, communication resource allocation, power control, and computational resources allocation.

Our optimization problem is formulated as follows:

$$\min_{\boldsymbol{\zeta}, \boldsymbol{p}, \boldsymbol{h}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}} \quad \mathcal{F}(\boldsymbol{\zeta}, \boldsymbol{p}, \boldsymbol{h}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \tag{29a}$$

$$\text{s.t.:} \quad \sum_{b=1}^{B} \zeta_{i,b} \leq 1, \quad \forall i \in \mathcal{O}, \tag{29b}$$

$$0 \leq p_i \leq p_i^{\max}, \quad \forall i \in \mathcal{O}, \tag{29c}$$

$$R_i \geq R_{\min}, \quad \forall i \in \mathcal{O}, \tag{29d}$$

$$\begin{aligned}\mathcal{F}(\boldsymbol{\zeta}, \boldsymbol{p}, \boldsymbol{h}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \sum_{i \in \mathcal{O}} \Bigg[\psi \Phi \Bigg(& \underbrace{\kappa_i h_i^2 x_i}_{local\ energy} + \underbrace{\sum_{j \in \mathcal{D}} \Bigg(\kappa_j h_{i,j}^2 + p_{i,j} \frac{S_i}{\sum_{b \in \beta} R_{i,j}^b \zeta_{i,j}^b} \Bigg) y_{i,j}}_{V2V\ energy} + \underbrace{\Bigg(\kappa_0 h_{i,0}^2 + p_{i,0} \frac{S_i}{\sum_{b \in \beta} R_{i,0}^b \zeta_{i,0}^b} \Bigg) z_{i,0}}_{V2I\ energy} \Bigg) \\ + (1-\psi) \Bigg(& \underbrace{\frac{S_i}{h_i} x_i}_{local\ latency} + \underbrace{\sum_{j \in \mathcal{D}} \Bigg(\frac{S_i}{h_{i,j}} + \frac{S_i}{\sum_{b \in \beta} R_{i,j}^b \zeta_{i,j}^b} \Bigg) y_{i,j}}_{V2V\ latency} + \underbrace{\Bigg(\frac{S_i}{h_{i,0}} + \frac{S_i}{\sum_{b \in \beta} R_{i,0}^b \zeta_{i,0}^b} \Bigg) z_{i,0}}_{V2I\ latency} \Bigg) \Bigg]. \end{aligned} \tag{28}$$

$$\sum_{i \in O} p_i^b g_{i,0}^b \zeta_{i,b} \leq I_b^{\max}, \forall b \in \beta, \quad (29e)$$
$$0 \leq h_i x_i \leq H_i^{\max}, \quad \forall i \in \mathcal{O}, \quad (29f)$$
$$\sum_{i \in \mathcal{O}} h_{i,0} z_{i,0} \leq H_0^{\max}, \quad \forall i \in \mathcal{O}, \quad (29g)$$
$$\sum_{i \in \mathcal{O}} h_{i,j} y_{i,j} \leq H_j^{\max}, \quad \forall j \in \mathcal{D}, \quad (29h)$$
$$E_i \leq E_i^{\max}, \quad \forall i \in \mathcal{O}, \quad (29i)$$
$$E_j \leq E_j^{\max}, \quad \forall j \in \mathcal{D}, \quad (29j)$$
$$L_i \leq L_i^{\max}, \quad \forall i \in \mathcal{O}, \quad (29k)$$
$$L_{i,j} \leq \tau_{i,j}, \quad \forall i \in \mathcal{O}, \quad \forall j \in \mathcal{D}, \quad (29l)$$
$$x_i + \sum_{j \in \mathcal{D}} y_{i,j} + z_{i,0} = 1, \quad (29m)$$
$$h_{i,j}, h_{i,0} \geq 0, \quad \forall i \in \mathcal{O}, \quad \forall j \in \mathcal{D}, \quad (29n)$$
$$x_i, y_{i,j}, z_{i,0} \in \{0, 1\}, \quad \forall i \in \mathcal{O}, \quad (29o)$$
$$\zeta_{i,b} \in \{0, 1\}, \quad \forall i \in \mathcal{O}, \quad \forall b \in \beta, \quad (29p)$$

where $\boldsymbol{\zeta} \in \{0, 1\}^{|\beta| \times |O|}$ is the RBs allocation variable, $\mathbf{p} \in \mathbf{R}^{|O|}$ is the transmit power variable, $\mathbf{h} \in \mathbf{R}^{|O|}$ is the computational resource allocation variable, $\boldsymbol{x} \in \{0, 1\}^{|O|}$ is the decision variable local EVs, e.g., SEVs, $\boldsymbol{y} \in \{0, 1\}^{|O|}$ is decision variable of neighboring EVs, e.g., DEVs, and $\boldsymbol{z} \in \{0, 1\}^{|O|}$ is the decision variable of MEC. Constraint (29b) guarantees that an EV is allocated in at most one RB b. Constraint (29c) represents that the transmitting power level of EVs does not exceed the maximum transmit power level. Constraint (29d) guarantees the QoS for a SEV with a minimum achievable data rate threshold $R_{\min}$. Constraint (29e) represents the protection of the RSU's user that is allocated in RB b with maximum interference threshold $I_b^{\max}$. Constraints (29f), (29g), and (29h) are computational capacity constraints of SEVs, DEVs, and RSU, respectively. Similarly, constraints (29i) and (29j) are the energy constraint of SEVs and DEVs, respectively. Constraint (29k) represents the upper bound of latency with $L_i^{\max}$ while the constraint (29l) is representing latency constraint concerning mobility between SEV i, and DEV j. Constraint (29m) guarantees the success of offloading tasks. The resource allocation variable is linear in (29n) while offload decision and RB allocation variables are binaries in (29o), and (29p).

The problem in (29) is a Mixed-Integer Non-Linear Programming (MINLP) problem due to binary decision variable, e.g, $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, and $\boldsymbol{\zeta}$. Thus, it falls into the NP-hard category. Obtaining a solution for such kind of problem requires a huge amount of computing resources and results in huge time complexity. Moreover, due to the huge dimension of the feasible space in problem (29), it is unsolvable via theoretical analysis, and thus, there is no optimality condition guarantee for the optimal solution. Therefore, to cope with this issue, we employ the Block Coordinate Descent (BCD) technique [27], [28] to obtain an approximate solution for the problem (29). BCD is also known as the Gauss-Seidel method, an iterative algorithm for non-convex, block multi-convex, and smooth objective functions under constraints optimization problem [27]. In such a way, the original problem is decomposed into multiple sub-problems and solved sequentially, then composed solutions to the original problem.

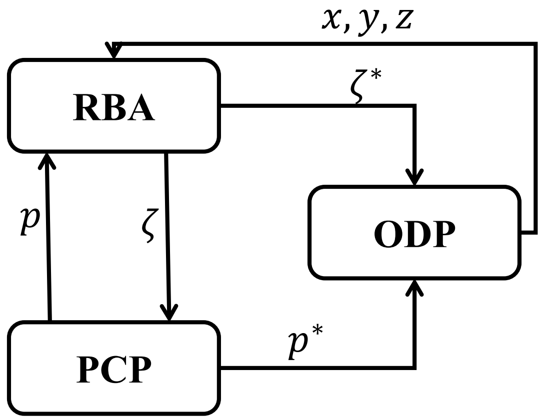

FIGURE 2. Proposed framework.

Intuitively, we decompose our original problem (29) into three subproblems: Resource Block Allocation problem (RBA), Power Control problem (PCP) for Vehicle-to-Vehicle (V2V) communication and EVs offload decision problem (ODP).

Then, these decomposed subproblems provide us the flexibility through which they can be either bi-convex, e.g., PCP problem, or transformed into convex problems, e.g. ODP, and solved alternatively. In which, solutions of RBA and PCP are exchanged with each other to obtain a stationary solution. Then, the final results of RBA and PCP are input for the ODP problem.

Next, we present our proposed solution approaches.

V. PROPOSED SOLUTION APPROACHES

This section presents our solution approach in which we decompose and solve the aforementioned problems. To begin with, we derive the solution approaches for RBA and PCP that couple with each other. Then, we design the solution approach for ODP problem based on the output of RBA and PCP. Our proposed approach is sequentially updating the solution at each EV and exchanging it with the others until each reaches a stationary solution. This can be done by fixing two blocks and solving the remaining block. This proposed algorithm is alternatively processed until reached a stationary solution. Furthermore, we have concrete our proposed model with the convergence condition in [27] and [28]. Therefore, it always exists at least a stationary solution. The detail of these processes is described in Fig. 2. Moreover, our approach aims to obtain solutions in a distributed manner. In such a way, each EV makes a decision locally and exchanges it with others. Therefore, in order to aggregate the solution of EVs, the RSU is assumed to be the global controller, where decisions of EVs are exchanged and balanced via some auxiliary variables [29], [30]. Furthermore, to successfully exchange this information, we assume that EVs are able to communicate with the RSU via wireless links. Moreover, the resources allocation variable $\boldsymbol{h}$ is binding as constraint in (29f), (29g), and (29h), and is considered as a projection function for the feasible solution [8].

Next, we present our solution approach for the first block RBA.

A. RESOURCE BLOCK ALLOCATION PROBLEM (RBA)

In this subsection, we present a solution approach for the RBA block, while fixing the other two blocks, e.g., PCP and ODP. For a given transmit power levels vector $\boldsymbol{p}$, EVs offload decision $\boldsymbol{x}, z$, and DEVs decision y, the RSU have to allocate RB for EVs that maximize the total achievable data rate of RSU users (RUEs) and V2V users (VUEs). In which, RUE and VUE are the set of SEVs that preferred to offload to the MEC or offload to nearby DEVs, respectively. The RB allocation for RUE has been proposed in many existing works [31], [32], [33], [34], [35]. Moreover, the task transmission scheduling has been considered in [36], [37], and [38]. Therefore, in this work, we focus more on the VUEs resource allocation problem. Let Ω_S be the set of SEVs that preferred to use V2V offloading. The optimization problem for V2V resource allocation can be formulated as follows:

$$\text{RBA}: \max_{\zeta} \ \mathcal{F}(\boldsymbol{\zeta}) = \sum_{i\in\Omega_S}\sum_{b\in B} R_{i,b}\zeta_{i,b} \tag{30a}$$

$$\text{s.t.} \ \sum_{b=1}^{B}\zeta_{i,b} \le 1, \quad \forall i \in \Omega_S, \tag{30b}$$

$$\sum_{i\in\Omega_S} p_{i,0,b}g_{i,0,b}\zeta_{i,b} \le I_b^{\max}, \quad \forall b \in \beta, \tag{30c}$$

$$\zeta_{i,b} \in \{0,1\}, \quad \forall b \in \beta. \tag{30d}$$

The problem in (30) falls out as a combinatorial problem due to the binary variable ζ. Therefore, it needs to use an alternative algorithm to solve this optimization on the RSU side. Based on the proof of convergence using a sub-gradient based technique in [39] and [29]. We propose an algorithm based on duality theory to solve RBA in (30). The detail of the algorithm is stated in Alg. 1.

B. POWER CONTROL PROBLEM FOR V2V OFFLOADING (PCP)

For a given RB b allocated to EV i, EV i has to carefully choose a transmit power level by considering the interference effect on RSU's users and other V2V users that use the same RB b. The optimization problem can be formulated as follows:

Algorithm 1 Duality-Based Resource Block Allocation for V2V Communication

1: **Initialize:** $t = 0$; $\gamma_{i,b}^{(0)} \ge 0$, step-size $\theta_b^{(0)} > 0$, $\epsilon_S = 1e^{-4}$,
2: **repeat**
3: $\quad t \leftarrow t+1$;
4: $\quad$The RSU updates $\zeta_{i,b}$ for SEV i and dual variable $\gamma_{i,b}$ as follows:
5: $\quad$Finding the optimal RB index for each SEV:

$$b^* = \arg\max_b \left\{\sum_{i\in\Omega_S}\sum_{b\in B}\left(R_{i,b} - \gamma_{i,b}^{(t)}g_{i,b}p_{i,b}\right)\right\}. \tag{31}$$

6: $\quad$RB allocation decision:

$$\zeta_{i,b} = \begin{cases}1, & \text{if } b = b^*,\\ 0, & \text{otherwise.}\end{cases} \tag{32}$$

7: $\quad$The RSU updates the dual variable according to step-size $\theta_b^{(t)}$:

$$\gamma_{i,b}^{(t+1)} = \left[\gamma_{i,b}^{(t)} - \theta_b^{(t)}\left(\sum_{i\in\Omega_S} p_{i,b}g_{i,0,b}\zeta_{i,b} - I_b^{\max}\right)\right]^+. \tag{33}$$

8: **until** $|\theta^{(t+1)} - \theta^{(t-1)}| \le \epsilon_S$;
9: Then, set $\boldsymbol{\zeta}$ as the desired solution.

as follows:

$$\text{PCP}: \max_{\boldsymbol{p}} \ \mathcal{F}(\boldsymbol{p}) = \sum_{i\in\mathcal{O}}\sum_{b\in\beta} F(p_{i,b}) = R_i^b \tag{34a}$$

$$\text{s.t.:} \ 0 \le p_i^b \le p_i^{\max}, \tag{34b}$$

$$R_i^b \ge R_{\min}, \tag{34c}$$

$$\sum_{i\in O} p_i^b g_{i,0}^b \zeta_{i,b} \le I_b^{\max}. \tag{34d}$$

The PCP problem is non-convex nor concave due to the properties of the objective function (5), and (6). Therefore, we approximate the objective function of problem PCP into an equivalent function as follows:

$$F(p_{i,b}) = \alpha\frac{R_{i,b}}{R_{i,b}^{\max}} - (1-\alpha)\frac{p_{i,b}}{p_{i,b}^{\max}}, \tag{35}$$

where α is a trade-off coefficient between transmit power and achievable rate in an RB. Then, the problem PCP can be rewritten as follows:

$$\text{PCPE}: \max_{\boldsymbol{p}} \ \mathcal{F}(\boldsymbol{p}) = \sum_{i\in\mathcal{O}}\sum_{b\in\beta} F(p_{i,b}) \tag{36a}$$

$$\text{s.t.:} \ 0 \le p_i^b \le p_i^{\max}, \quad \forall i \in \mathcal{O}, \tag{36b}$$

$$R_i^b \ge R_{\min}, \quad \forall i \in \mathcal{O}, \tag{36c}$$

$$\sum_{i \in O} p_i^b g_{i,0}^b \zeta_{i,b} \leq I_b^{\max}, \quad \forall b \in \beta. \tag{36d}$$

Lemma 1: The problem PCPE is a bi-convex problem.

Since PCPE is a bi-convex problem. It always exists at least a stationary solution. Therefore, we have derived the general form of solution for (36) in the following proposition.

Proposition 2: According to the Karush-Kuhn-Tucketer (KKT) conditions [29], the optimal solution of transmit power level for EV i for PCPE is as follows:

$$p_{i,b}^* = \frac{\left(\Lambda_{1,i} + \lambda_{3,i}^*\right) W_b \ln(2)}{\Lambda_{2,i} + \lambda_{1,i}^* + \lambda_{2,i}^* - \lambda_{4,i}^* g_{i,0}^b} - \frac{\sum_{j \in \Omega_b, j \neq i} p_j^b g_{j,i}^b + I_{RSU} + I_0}{g_i^b}. \tag{37}$$

Proof: Please refer to appendix. ■

Based on aforementioned lemma, and proposition, we now apply dual decomposition method to alternatively get a approximate solution [29] for PCPE stated in (36). The detail of the proposed algorithm is described in Alg. 2.

Algorithm 2 Lagrangian Multiplier-Based Distributed Power Control for V2V Communication

1: **Initialize:** $t = 0, \alpha = 0.5,$ step-size $\theta_{2,i}, \theta_{3,i}, \theta_{4,i} \geq 0,$
2: **repeat**
3: Each SEV $i \in \mathcal{O}$ alternatively updates transmit power according to (37).
4: After receiving the transmit power level of all SEVs, the RSU updates the dual variable as followings:

$$\lambda_{2,i}^{(t+1)} = \left[\lambda_{2,i}^{(t)} + \theta_{2,i}\left(p_{i,b}^{(t+1)} - p_{\max}\right)\right]^+ \tag{38a}$$

$$\lambda_{3,i}^{(t+1)} = \left[\lambda_{3,i}^{(t)} + \theta_{3,i}\left(W_b \log_2(1 + \xi_{i,b} p_{i,b}^{(t+1)}) - R_{\min}\right)\right]^+ \tag{38b}$$

$$\lambda_{4,i}^{(t+1)} = \left[\lambda_{4,i}^{(t)} + \theta_{4,i}\left(\sum_{i \in \Omega_b} p_{i,b}^{(t+1)} g_{i,0,b} - I_b^{\max}\right)\right]^+ \tag{38c}$$

5: After updating all of the variables, the RSU then informs all SEVs by broadcasting the value of dual variables.
6: **until** $|\boldsymbol{p}^{(t+1)} - \boldsymbol{p}^{(t)}| \leq \epsilon_p$;
7: Then, set $\boldsymbol{p}$ as the desired solution.

C. SEV OFFLOAD DECISION PROBLEM (ODP)

For a given RB allocation $\boldsymbol{\zeta}$, and transmit power level $\boldsymbol{p}$ the ODP can be formulated as follows:

$$\begin{aligned}
\text{ODP}: \min_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} \ & \mathcal{F}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) && (39a)\\
\text{s.t:} \ & 0 \leq h_i \leq H_i^{\max}, && (39b)\\
& \sum_{i \in \mathcal{O}} h_{i,0} \leq H_0^{\max}, && (39c)\\
& \sum_{i \in \mathcal{O}} h_{i,j} y_{i,j} \leq H_j^{\max}, \quad \forall j \in \mathcal{D}, && (39d)\\
& E_i \leq E_i^{\max}, \quad \forall i \in \mathcal{O}, && (39e)\\
& E_j \leq E_j^{\max}, \quad \forall j \in \mathcal{D}, && (39f)\\
& L_i \leq L_i^{\max}, && (39g)\\
& L_{i,j} \leq \tau_{i,j}, \quad \forall i \in \mathcal{O}, \quad \forall j \in \mathcal{D}, && (39h)\\
& x_i + \sum_{j \in \mathcal{D}} y_{i,j} + z_{i,0} = 1, && (39i)\\
& h_{i,j} \geq 0, && (39j)\\
& h_{i,0} \geq 0, && (39k)\\
& x_i \in \{0,1\}, && (39l)\\
& y_{i,j} \in \{0,1\}, && (39m)\\
& z_{i,0} \in \{0,1\}, . && (39n)
\end{aligned}$$

The problem in (39) falls into the combinatorial category due to the binary variables $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{z}$. Therefore, it is an NP-hard problem. In such a case, obtaining solution for ODP is time consuming and intractable. Thus, in order to solve ODP, we firstly relax the binary variables into continuous variables. Note that this yields an interesting scenario in which the task now can be offloaded distributively between DEVs and RSU, thus, forming partial offloading between the remote servers. In other words, a part of the task can be processed at DEVs and the remaining parts are processed at the RSU. Intuitively, we reformulate the problem ODP to an equivalent problem ODPE as follows:

$$\begin{aligned}
\text{ODPE}: \min_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} \ & \mathcal{F}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) && (40a)\\
\text{s.t:} \ & (39b)-(39k), && (40b)\\
& 0 \leq x_i \leq 1, && (40c)\\
& 0 \leq y_{i,j} \leq 1, && (40d)\\
& 0 \leq z_{i,0} \leq 1. && (40e)
\end{aligned}$$

The problem in (40) turns out to be convex due to linear objective, and either linear or close convex set constraints [29]. However, due to the coupling constraint (39i), it is required an alternative method to obtain the solution for ODPE. Therefore, we apply ADMM [30], to obtain the solution for ODPE. Firstly, we define a feasible set of solutions at each SEV, DEV, and RSU, respectively, which aims to reduce the number of information exchanges between EVs, and RSU. By taking the constraint that respects each EV such as computational capacity, and energy capacity. Let Ω_i be the feasible set of

SEV i. Ω_i is defined as follows:

$$\Omega_i \triangleq \Big\{x_i \in \mathbf{R} \ \mathcal{V}\ 0 \le x_i \le 1, h_i x_i \le H_i^{\max}, E_i \le E_i^{\max}, \times L_i \le L_i^{\max}, L_{i,j} \le \tau_{i,j}\Big\}. \tag{41}$$

Similarly, let Ω_j be the feasible set of DEV j. Ω_j is defined as follows:

$$\Omega_j \triangleq \Big\{\boldsymbol{y}_j \in \mathbf{R}^{|\mathcal{O}|} \ \mathcal{V}\ \boldsymbol{y}_j \succeq 0, \boldsymbol{h}_j^T \boldsymbol{y}_j \le H_j^{\max}, E_j \le E_j^{\max}\Big\}. \tag{42}$$

Finally, the feasible set of the RSU Ω_0 is defined as follows:

$$\Omega_0 \triangleq \Big\{z \in \mathbf{R}^{|\mathcal{O}|} \ \mathcal{V}\ z \succeq 0, \boldsymbol{h}_0^T z \le H_0^{\max}\Big\}. \tag{43}$$

Consequently, the problem in (40) can be rewritten as follows:

$$\text{ODPE}: \min_{\boldsymbol{x},\boldsymbol{y},z} \ \mathcal{F}(\boldsymbol{x},\boldsymbol{y},z) \tag{44a}$$

$$\text{s.t.} \quad \boldsymbol{x}_i + \sum_{j\in\mathcal{D}} \boldsymbol{y}_{i,j} + z_{i,0} = \mathbf{1}, \quad \forall i \in \mathcal{O}, \tag{44b}$$

$$\boldsymbol{x}_i \in \Omega_i, \quad \boldsymbol{y}_i \in \Omega_j, \quad z_i \in \Omega_z, \tag{44c}$$

where $\mathbf{1}$ is a -vector with all elements equal to one. Following the ADMM method, we can drive the augmented Lagrangian function of the problem in (44) as follows:

$$\mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{y},z,\boldsymbol{\lambda}) = \mathcal{F}(\boldsymbol{x},\boldsymbol{y},z) + \boldsymbol{\lambda}^T(\boldsymbol{x} + \sum_{j\in\mathcal{D}} \boldsymbol{y}_j + z - \mathbf{1}) + \frac{\rho}{2}\Big|\Big|\boldsymbol{x} + \sum_{j\in\mathcal{D}} \boldsymbol{y}_j + z - \mathbf{1}\Big|\Big|_2^2, \tag{45}$$

where $||\cdot||_2^2$ is the norm-2 squared, $\boldsymbol{\lambda}$ is a Lagrangian multiplier of the constrain (44b), and ρ is any positive number considered as a penalty term for the Lagrangian function. By taking the partial derivative w.r.t. each variables $\boldsymbol{x}, \boldsymbol{y}$ and z, based on the result of [29]. The solution of ODPE can be obtain by sequentially updating each primal and dual variable. The problem ODPE is always guaranteed an optimal solution due to the convexity of $\mathbf{F}(\boldsymbol{x},\boldsymbol{y},z)$, and either linear or closed convex set constraints [29]. In this work, we assume that SEV i will be the first one who makes the decision for offloading and informing its neighbors. Next, after receiving the decision of SEVs, DEVs are going to make decisions w.r.t to its feasible set and feedback on the result to SEVs, and RSU. The RSU will now have the information about SEVs and DEVs decisions. Then, it might make a decision to accept and process the task or not. Then, it will update the dual variable $\boldsymbol{\lambda}$. The detail of updating scheme is stated as follows. Each SEV $i \in \mathcal{O}$ in parallel update the primal variable $\boldsymbol{x}$ as follows:

$$\boldsymbol{x}_i^{(t+1)} = \arg\min_{\boldsymbol{x}_i} \Big\{\mathcal{F}(\boldsymbol{x}_i) + \boldsymbol{\lambda}_i^{(t)}\Big(\boldsymbol{x}_i + \sum_{j\in\mathcal{D}} \boldsymbol{y}_{ij}^{(t)} + z_i^{(t)} - \mathbf{1}\Big) + \frac{\rho}{2}\Big|\Big|\boldsymbol{x}_i + \sum_{j\in\mathcal{D}} \boldsymbol{y}_{ij}^{(t)} + z_i^{(t)} - \mathbf{1}\Big|\Big|_2^2\Big\}, \quad \forall \boldsymbol{x}_i \in \Omega_i. \tag{46}$$

Algorithm 3 ADMM-Based EV-Assisted MEC Decision

Input: $\mathcal{O}$, $\mathcal{D}$;
Output: $\boldsymbol{x}, \boldsymbol{y}, z$;
Initialize: $t \leftarrow 0; \boldsymbol{x}^{(0)} \leftarrow 0, \boldsymbol{y}^{\mathbf{0}} \leftarrow 0, z^{(\mathbf{0})} \leftarrow 0, \boldsymbol{\lambda}^{(\mathbf{0})} \leftarrow 0$, and $\rho = 1.0$;
repeat
  SEV i updates $\boldsymbol{x}^{(t+1)}$, $\forall i \in \mathcal{O}$ according to (46) ;
  Each neighbors EV $j \in \mathcal{O}$ update its decision $\boldsymbol{y}_{j\in\mathcal{D}}$ paralelly according to (47)
  The RSU update the value of its decision $z^{(t+1)}$ according to (48), and Lagrangian multiplier $\boldsymbol{\lambda}$ according to (49) ;
  The RSU and Neighbor EVs feedback the information about its decision to EV i;
  After receiving the information about the neighbors and RSU, the EV i then update the dual Lagrangian variable according to (49);
  $t \leftarrow t + 1$;
until $||(\boldsymbol{x}^{(t+1)} + \boldsymbol{y}^{(t+1)} + z^{(t+1)}) - (\boldsymbol{x}^{(t)} + \boldsymbol{y}^{(t)} + z^{(t)})|| \le \epsilon_{pri} \cap ||\boldsymbol{\lambda}^{(t+1)} - \boldsymbol{\lambda}^{(t)}|| \le \epsilon_{dual}$;
Then, set $\boldsymbol{x}^*, \boldsymbol{y}^*, z^*$ as the desired solution.

After receiving the information of SEVs, neighboring DEVs $j \in \mathcal{D}$ will update the variable y_j as follows:

$$\boldsymbol{y}_j^{(t+1)} = \arg\min_{\boldsymbol{y}_j} \Big\{\mathcal{F}(\boldsymbol{y}_j) + \boldsymbol{\lambda}^{(t),T}\Big(\boldsymbol{x}^{(t+1)} + \boldsymbol{y}_j + \sum_{j'\in\mathcal{D}, j'\neq j} \boldsymbol{y}_{j'}^{(t)} + z^{(t)} - \mathbf{1}\Big) + \frac{\rho}{2}\Big|\Big|\boldsymbol{x}^{(t+1)} + \boldsymbol{y}_j + \sum_{j'\in\mathcal{D}, j'\neq j} \boldsymbol{y}_{j'}^{(t)} + z^{(t)} - \mathbf{1}\Big|\Big|_2^2\Big\}, \tag{47}$$

Finally, the RSU will make decision and update dual variable and then broadcast it to all EVs in the system. The update primal variable of RSU is given as:

$$z^{(t+1)} = \arg\min_{z} \Big\{\mathcal{F}(z) + \boldsymbol{\lambda}^{(t),T}\Big(\boldsymbol{x}^{(t+1)} + \sum_{j\in\mathcal{D}} \boldsymbol{y}_j^{(t+1)} + z - \mathbf{1}\Big) + \frac{\rho}{2}\Big|\Big|\boldsymbol{x}^{(t+1)} + \sum_{j\in\mathcal{D}} \boldsymbol{y}_j^{(t+1)} + z - \mathbf{1}\Big|\Big|_2^2\Big\}, \tag{48}$$

and the update of dual variable $\boldsymbol{\lambda}$ is given by:

$$\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)} + \rho\Big(\boldsymbol{x}^{(t+1)} + \sum_{j\in\mathcal{D}} \boldsymbol{y}_j^{(t+1)} + z^{(t+1)} - \mathbf{1}\Big). \tag{49}$$

The detail of proposed algorithm is stated in Alg. 3.

Based on the aforementioned solution approach, we then integrated them together in order to compose the solution equivalent problem of the problem stated in (29). The detail of the proposed algorithm is presented in Alg. 4.

Algorithm 4 BCD-Based Algorithm for Joint Communication, Computation, Interference Management for Vehicle-Assisted MEC

Initialize: $t \leftarrow 0$; $\boldsymbol{x}^{(0)} \leftarrow 0$, $\boldsymbol{y}^{\mathbf{0}} \leftarrow 0$, $\boldsymbol{z}^{(\mathbf{0})} \leftarrow 0$, $\boldsymbol{p} = p_{\max}$, $\boldsymbol{z} = \mathbf{1}$, $\boldsymbol{\lambda}^{(\mathbf{0})} \leftarrow 0$, and $\rho = 1.0$, $\epsilon_{bcd} = 1e-4$;
repeat
  $t \leftarrow t+1$;
  Observe $\boldsymbol{x}^{(t+1)}, \boldsymbol{y}^{(t+1)}, \boldsymbol{z}^{(t+1)}$ by solving ODP
  Observe $\boldsymbol{\zeta}^{t+1}$ by solving RBA
  Observe $\boldsymbol{p}^{t+1}$ by solving PCP
until $|\mathbf{F}(\boldsymbol{x}^{(t\mathcal{C}1)}, \boldsymbol{y}^{(t\mathcal{C}1)}, \boldsymbol{z}^{(t\mathcal{C}1)}, \boldsymbol{p}^{(t\mathcal{C}1)}, \boldsymbol{\zeta}^{(t\mathcal{C}1)}) - \mathbf{F}(\boldsymbol{x}^{(t)}, \boldsymbol{y}^{(t)}, \boldsymbol{z}^{(t)}, \boldsymbol{p}^{(t)}, \boldsymbol{\zeta}^{(t)})| \le \epsilon_{bcd}$;
Then, set $\boldsymbol{x}^{*}, \boldsymbol{y}^{*}, \boldsymbol{z}^{*}$ as the desired solution.

VI. NUMERICAL RESULTS

In this section, we evaluate the performance of our proposed algorithm in the considered collaborative EV-assisted MEC system. The proposed system model aims to use parallel processing. However, due to the limitation on physical devices, we evaluate our proposed framework on a single computer with specifications: Intel Core i5-4690, 16 (GB), GPU GTX 1060 - 3(GB). We use Python3.8 as our simulation tool combined with CVXPY.

A. SIMULATION SETUP

In order to demonstrate the numerical results for our proposed approach, we choose a network consisting RSU-enabled MEC server with a computational capacity of 3.9 (GHz). We let the number of SEVs be in the range [10 $\sim$ 40], and the number of DEVs is in the range [5 $\sim$ 20]. Each SEV has a task to offload which represents a tuple containing the task's size is random in the range [100.0 $\sim$ 500.0] (MB) with the median at 265.5 (MB) and mode at 271.5 (MB). Similarly, the CPU requirement of the task is random in range [1.0 $\sim$ 4.0] (kHz) with median at 2.2 (kHz) and mode at 2.5 (kHz). The location of EVs is assumed to follow the Homogeneous Poison Point Process (HPPP) in this work. The main parameter is presented in Table. 2.

Furthermore, in our simulation, we set the convergence criteria of each algorithm at 10^{-4}, and the acceptable convergence rate of a primal variable at 10^{-3} [8]. Moreover, there is no existing real dataset to demonstrate the practicality of our approach and limitation on physical devices to deploy a test-based performance evaluation, we can only provide comprehensive numerical results by taking multiple runs such as at least 50 runs per result in our works, where input data are totally random for each run, and the final results are taking the average value.

B. PERFORMANCE BENCHMARKS

In this paper, we compare our proposed Alg. 3 which deployed in distributed manner with below mentioned baseline schemes named as *Greedy Approach*, *Exhaustive Search*,

TABLE 2. Simulation parameters.

Parameter	Value
Number of SEVs (O)	$10 \sim 40$
Number of DEVs (D)	$5 \sim 20$
Path loss (η)	2
Maximum transmit power (p)	23 (dBm) [31]
Shadow fading standard deviation	3 (dB) [40]
Path loss (cellular link)	$128.1 + 37.6\log(d)$ [40]
System bandwidth	3 (MHz)
Task size (S_i)	$[100.0 \sim 500.0]$(MB) $\bar{S}_i = 265.5, \tilde{S}_i = 271.5$ (MB)
Task requirement (C_i)	$[1.0 \times 10^3 \sim 4.0 \times 10^3]$ $\bar{C}_i = 2.2 \times 10^3$, $\tilde{C}_i = 2.5 \times 10^3$
Thermal noise for 1 Hz at 20°C	-174(dBm)
Penalty parameter (ρ)	$\rho = \{1.0, 5.0, 10.0, 15.0\}$
Stopping criteria ($\epsilon_p, \epsilon_S, \epsilon_{dual}$)	10^{-4}
Stopping criteria (ϵ_{primal})	10^{-3}
Minimum achievable data rate ($R_{\min}$)	$\{2.0 \sim 4.0\}$ (Mbps)
Interference threshold ($I_b^{\max}$)	$-\{80., 100., 120.\}$ (dBm) [41]
Computational capacity of EVs (H_i)	$1.2 \sim 1.5$ (GHz)
Energy capacity of EVs ($E^{\max}$)	$[1.0 \times 10^3 \sim 2.0 \times 10^3]$ (W)
Velocity of EVs (v_i)	$[10.0 \sim 20.0]$ (m/s)
Normalization parameter (Φ)	0.5×10^{-2}

and *Centralized*. The summarization of these schemes is as follows:

Greedy Approach(GA): This approach required the information of all EVs to be available at the RSU server which can solve (39) in a centralized manner. In such a way, the RSU is using *the best first search* strategy to make decisions for each EV. Furthermore, the solution is mostly achieved as a locally optimal solution but has a huge gap compared to the global optimal. In some trivial cases, the greedy approach might achieve an optimal solution. This approach has an $\mathcal{O}(N^2)$ complexity.

Exhaustive Search(ES): Similar to the GA approach, ES requires the information of all EVs at the RSU side to solve the problem in a centralized manner. The RSU might try to search all of the possible solutions that satisfy all of the constraints and conditions posed in (39). It always guarantees the globally optimal solution for MINLP but is time-consuming and computationally intensive. Moreover, the original problem in (29) is NP-hard, thus, the ES scheme could not obtain the optimal solution on a large-scale setting in polynomial time. Therefore, to quantify the gap between the optimal and proposed solution, we have taken a small-scale network to compare our proposed approach with the optimal solution, where the number of DEVs in the range of [10 $\sim$ 20], and the number of SEV in the range [10 $\sim$ 40].

This approach has an $\mathcal{O}(2^N)$ complexity.

Centralized Algorithm: This algorithm requires the RSU with complete information as inputs for solving the problem (40) in a centralized manner and has a complexity of at least $\mathcal{O}(N\log(N))$ [29].

It must be noted that ES and GA schemes are designed to solve the problem of MINLP state in (29), where the goal of ES is to obtain an optimal solution of our formulated problem in a strictly small scale-setting such as number of EVs is very small. Moreover, GA is a scheme that can be used to obtain a solution of a lower complexity compared to

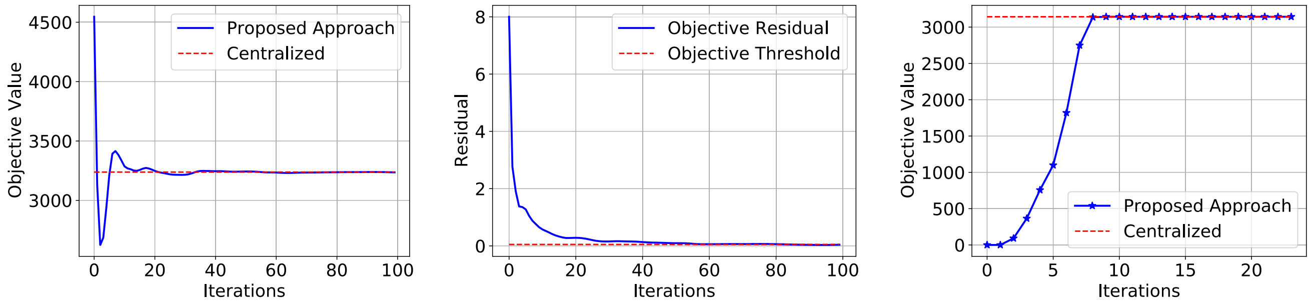

FIGURE 3. **Performance of proposed framework: (a) Convergence of BCD-based algorithm, (b) Convergence rate of BCD-base algorithm, and (c) Convergence of ODP algorithm.**

ES, where the solution is obtained via the best first search strategy. On the other hand, the *Centralized*, and our proposed approach are designed to solve the relaxed problem that is the convex problem state in (29) and (40). This approach requires complete information on EVs in our model such as task profiles, transmit power, EVs' velocity, location, etc.

Furthermore, our proposed approach is performed in a distributed manner which does not require complete information in any agent such as EVs or the RSU. In which, each EV makes decisions simultaneously and sent them to the RSU. After gathering all decisions of EVs, the RSU balances the decision of all EVs via auxiliary variables such as γ, and λ to avoid the violation of feasible constraints. This will incur an overhead communication of message exchange among EVs and the RSU. However, this information is typically very small in the scale of just a few bits and can be ignored, e.g., $8 + 4|O|$ bytes when EV send feedback to the RSU, and 12 bytes when the RSU share information to EVs.

It must be noted that our proposed approach sequentially optimizes each block in turn, where RBA will be optimized first, then optimized PCP, and ODP is optimized afterward. Moreover, the iteration complexity of BCD-based algorithm is $\mathcal{O}(\epsilon_{bcd}^{-1}(\log(\epsilon_{bcd}^{-1}))^2)$ [28]. Furthermore, each block have complexity of $\mathcal{O}(N)$ in RBA, $\mathcal{O}(1)$ in PCP, and $\mathcal{O}(N)$ in ODP. In the worst scenario, with the coherence time is one second, and the criteria condition $\epsilon_{bcd} = 10^{-4}$, our proposed approach can handle up to $N = 100$ which is equivalent to $|O| = 100$ EVs to achieve the final solution.

C. NUMERICAL RESULTS

This subsection mainly focuses on the performance improvement of our proposed algorithm.

1) CONVERGENCE PERFORMANCE OF BCD-BASED ALGORITHM

The convergence of the proposed algorithm based on the BCD technique is presented in Fig. 3(a). As shown in the figure, we can see that the proposed approach achieved a stationary solution within 40 iterations. Moreover, we also observed that our proposed approach is stable after reached to the stationary without any fluctuation. On the other hand, we verify the convergence rate of the BCD-based algorithm by capturing the residual of the objective function (28) between each iteration in Fig. 3(b). It means that with up to 40 iterations, our solution is stable without any change even if we increase the number of iterations.

Furthermore, based on theoretical analysis in [28], our proposed approach remains in the strong convexity or bi-convex category; and thus, the convergence rate is always better than the non-convex category. Therefore, our proposed approach can be deployed in large-scale settings with a significant improvement in convergence performance and reasonable for particle settings. However, to successfully deploy in a realistic environment, we need to improve our mechanism further for uncertain conditions, such as accidents, traffic jams, or natural uncertainty.

2) CONVERGENCE PERFORMANCE OF OPD

The convergence of the ODP algorithm is presented in Fig. 3(c). From the figure, we observe that the proposed algorithm has converged to the optimal solution within less than 10 iterations. In the initial state, the objective function is smaller than the centralized solution due to the violation of the constraint (39i). Moreover, the solution in earlier iterations does not satisfy the stopping conditions ϵ_{pri} and ϵ_{dual}, thus, the algorithm keeps repeating. Therefore, when the number of iterations is increasing up to 10, our proposed approach achieved the same performance with the *Centralized* scheme due to the convexity of problem (40). This convergence always guarantees a stationary point which is known as the globally optimum point for problem ODP. Moreover, we have analysed the convergence performance with various simulation setting such as $\rho = \{0.5, 1.0, 2.0\}$ parameter,the number of SEVs $O = \{10, 20, 30, 40\}$ and the number of DEVs $D = \{5, 10, 15, 20\}$. We obverse that our proposed algorithm always guarantee the stationary solution at most within 10 iteration with $\rho = 1, D = 10$ and $O = 40$.

3) CONVERGENCE PERFORMANCE OF PCP AND RBA ALGORITHM

The convergence of the PCP and RBA algorithm is presented in Fig. 4(a), and Fig. 4(b), respectively. For a better visualization in a small figure, we set the number of V2V pairs equal to 6. Indeed, the proposed approach can operate on a

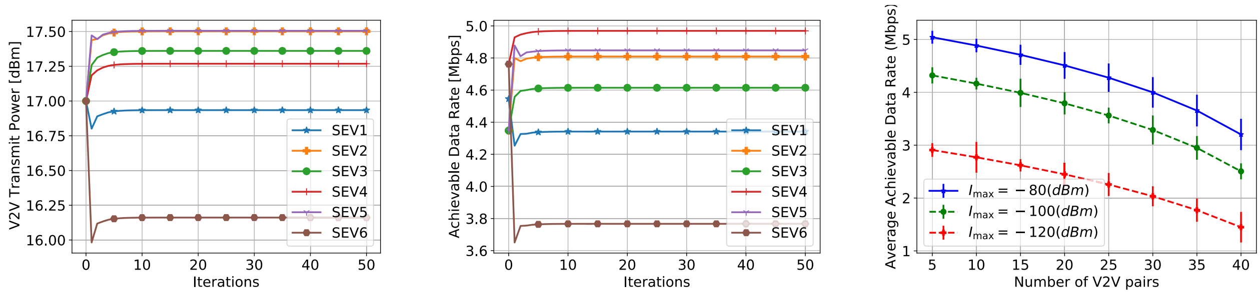

FIGURE 4. Performance of PCP and RBA: (a) Convergence of PCP algorithm, (b) Convergence of RBA algorithm, (c) Average achievable data rate.

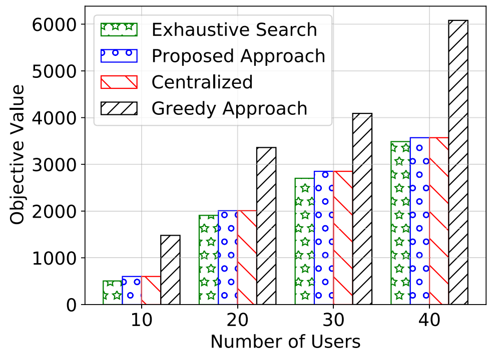

FIGURE 5. Performance comparison between proposed approach and benchmarks.

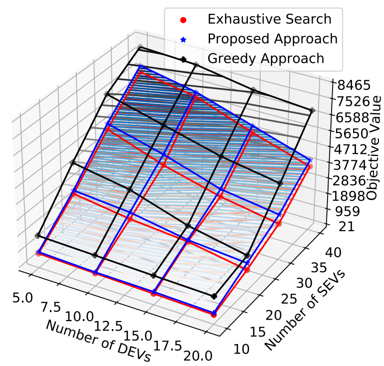

FIGURE 6. Performance of proposed approach versus the number of SEVs and DEVs vary.

large scale since our proposed method is based on gradient theory and the convergence rate is dependent on the property of the objective function and constraints. We can see that our proposed approach achieves a stationary solution at most after 10 iteration for both algorithms while they couple each other. This algorithm is stable since the solution does not fluctuate after getting to the stationary point. On the other hand, we have to also capture the effect of our proposal in dense settings, therefore, we only choose to have 15 RBs [31], [40]. By setting the maximum interference experienced by RUE, $I_b^{\max} = \{-80.0, -100.0, -200.0\}$ (dBm), explicitly. We can see that the system still achieves the highest achievable data rate at $I_b^{\max} = -80.0$ (dBm) due to the tight protection for RUE. In such a case, every V2V user must carefully control its power level to keep the interference of RUE as low as possible. On the other hand, the RUEs are transmitting with fixed power levels such that the average achievable rate is higher than the case of $I_b^{\max} = \{-100.0, -120.0\}$ (dBm) as shown in Fig. 4(c).

4) PERFORMANCE OF PROPOSED SCHEME

As show in Fig. 5, we can see that our proposed approach has achieved the same performance with *Centralized* approach, while outperforms the *Greedy Approach*, and close to the *Exhausted Search* which is considered as an optimal solution. Moreover, the proposed approach is higher than the optimal solution 3.5% at the number of SEVs $O = 10$ and 3.9% at $O = 40$. Meanwhile, it is lower than the *Greedy Approach*

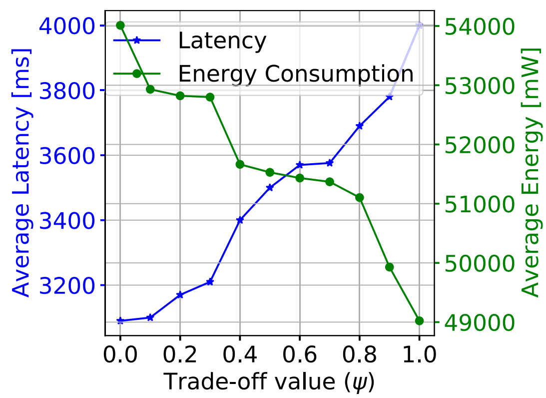

FIGURE 7. System performance versus trade-off value vary.

65% at $O = 10$ and 49.5% at $O = 40$. Furthermore, we have varied the number of SEVs and DEVs, respectively. The result is shown in Fig. 6 in which our proposed approach always achieves a performance with the lowest 2.5% and highest 6.0% compare to the performance of the *Exhausted Search*.

5) THE IMPACT OF TRADE-OFF VALUE ψ

The results of the aforementioned sections are conducted at $\psi = 0.5$. To show the impact of trade-off value on the objective value, we have varied the value of $\psi = \{0.0, 0.1, \ldots, 1.0\}$. We can see that when $\psi = 1.0$, the objective function in (28) is focused on minimizing the total energy consumption such that the power is minimized at 49, 000 (mW) while the latency has the highest value 4, 000 (ms). Similarly, when we set $\psi = 0.0$ that means that the optimization is focused on minimizing the latency only. In such a case, the energy consumption is achieved at the highest 54, 000 (mW) while the latency is minimized at 3115 (ms). Thus, we choose $\psi = 0.5$ to balance between latency and energy consumption for our numerical results.

VII. CONCLUSION

In this paper, we propose a novel solution that minimizes and balances between latency and energy consumption. Since the formulated problem was the MINLP category, we have decomposed the original problem into three subproblems by using the BCD technique. We then proposed an algorithm based on duality theory to get the sub-optimal solution for the first sub-problem RBA. And, the sub-optimal for the second sub-problem PCP by using distributed power control based on the Lagrangian multiplier method. Especially, we have achieved a global optimum for the third sub-problem ODP. Through a comprehensive numerical analysis, the results show that the final solution of the original problem converges to the sub-optimal solution with an average gap of 5%. In addition, our proposal reduces the complexity into a polynomial complexity compared to the exponential complexity of *Exhaustive Search* and quadratic complexity of *Greedy Approach*. We have considered the mobility of EVs as a time constraint for the formulated problem as well as wireless channel aspects such as shadowing, fast fading, and interference management for offering more adequate offloading service in the vehicular network.

APPENDIX

PROOF PCPE

Proof: The objective function in (36) can be rewrite as follows:

$$\begin{aligned}F(p_i^b) &= \sum_{i\in\Omega_b} R_i^b \\ &= \sum_{i\in\Omega_b} \alpha\frac{R_i^b}{R_i^{b,\max}} - (1-\alpha)\frac{p_i^b}{p_i^{\max}} \\ &= \sum_{i\in\Omega_b} \alpha\frac{W_b\log_2(1+\frac{p_i^b g_i^b}{I_{V2V}+I_{RSU}+I_0})}{R_i^{b,\max}} - (1-\alpha)\frac{p_i^b}{p_i^{\max}} \\ &= \sum_{i\in\Omega_b} \Lambda_{1,i}W_b\log_2(1+\xi_i^b p_i^b) - \Lambda_{2,i}p_i^b, \end{aligned} \tag{50}$$

where $\Lambda_{1,i} = \frac{\alpha}{R_i^{b,\max}}$, $\Lambda_{2,i} = \frac{1-\alpha}{p_i^{\max}}$, $\xi_i^b = \frac{g_i^b}{I_{V2V}+I_{RSU}+I_0}$. The Lagrangian function is given by:

$$\begin{aligned}\mathcal{L}(p,\lambda) = &\sum_{i\in\Omega_b}\Big(\Lambda_{2,i}p_i^b - \Lambda_{1,i}W_b\log_2(1+\xi_i^b p_i^b)\Big) \\ &- \sum_{i\in\Omega_b}\lambda_{1,i}p_i^b + \sum_{i\in\Omega_b}\lambda_{2,i}(p_i^b - p_i^{\max}) \\ &- \sum_{i\in\Omega_b}\lambda_{3,i}\Big(W_b\log_2(1+\xi_i^b) - R_i^{\min}\Big) \\ &- \lambda_4\Big(\sum_{i\in\Omega_b}p_i^b g_{i,0}^b - I_b^{\max}\Big). \end{aligned} \tag{51}$$

The derivative w.r.t. p_i^b is given by:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial p_i^b} &= \Lambda_{2,i} - \Lambda_{1,i}W_b\ln(2)\frac{\xi_i^b}{1+\xi_i^b p_i^b} - \lambda_{1,i} + \lambda_{2,i} \\ &\quad - \lambda_{3,i}W_b\ln(2)\frac{\xi_i^b}{1+\xi_i^b p_i^b} - \lambda_4 g_{i,0}^b \\ &= \Lambda_{2,i} + \lambda_{1,i} + \lambda_{2,i} - \lambda_4 g_{i,0}^b \\ &\quad - \Big(\Lambda_{1,i} + \lambda_{3,i}\Big)W_b\ln(2)\frac{\xi_i^b}{1+\xi_i^b p_i^b}. \end{aligned} \tag{52}$$

By setting the first order condition equal to zero, e.g., $\frac{\partial\mathcal{L}}{\partial p_i^b} = 0$, then

$$\begin{aligned}&\Lambda_{2,i} + \lambda_{1,i} + \lambda_{2,i} - \lambda_4 g_{i,0}^b \\ &\quad = \Big(\Lambda_{1,i} + \lambda_{3,i}\Big)W_b\ln(2)\frac{\xi_i^b}{1+\xi_i^b p_i^b}. \end{aligned} \tag{53}$$

It is equivalent to

$$\begin{aligned}\frac{\xi_i^b}{1+\xi_i^b p_i^{b,*}} &= \frac{\Lambda_{2,i} + \lambda_{1,i} + \lambda_{2,i} - \lambda_4 g_{i,0}^b}{\Big(\Lambda_{1,i} + \lambda_{3,i}\Big)W_b\ln(2)} \\ \Rightarrow p_i^{b,*} &= \frac{\Big(\Lambda_{1,i} + \lambda_{3,i}\Big)W_b\ln(2)}{\Lambda_{2,i} + \lambda_{1,i} + \lambda_{2,i} - \lambda_4 g_{i,0}^b} - \frac{1}{\xi_i^b}. \end{aligned} \tag{54}$$

Base on (54), the optimal transmit power for EV i must satisfy the following conditions(complementary slackness):

1) $\lambda_{1,i}^* p_i^{b,*} = 0, \forall i \in \Omega_b$,
2) $\lambda_{2,i}^*(p_i^{b,*} - p_i^{\max}) = 0, \in \Omega_b$,
3) $\lambda_{3,i}^*(W_b\log_2(1+\xi_i^b p_i^{b,*}) - R_{\min}) = 0, \forall \in \Omega_b$,
4) $\lambda_4(\sum_{i\in\Omega_b} p_i^b g_{i,0}^b - I_b^{\max}) = 0$.

Therefore, based on the KKT conditions, the Lagrangian multiplier must be non-negative $\lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^* \succeq 0$, we can have a general form for transmit power of EV i at RB b given by:

$$\begin{aligned}p_i^{b,*} &= \frac{\Big(\Lambda_{1,i} + \lambda_{3,i}^*\Big)W_b\ln(2)}{\Lambda_{2,i} + \lambda_{1,i}^* + \lambda_{2,i}^* - \lambda_{4,i}^* g_{i,0}^b} - \frac{1}{\xi_i^b} \\ &= \frac{\Big(\Lambda_{1,i} + \lambda_{3,i}^*\Big)W_b\ln(2)}{\Lambda_{2,i} + \lambda_{1,i}^* + \lambda_{2,i}^* - \lambda_{4,i}^* g_{i,0}^b} \\ &\quad - \frac{\sum_{j\in\Omega_b, j\neq i} p_j^b g_{j,i}^b + I_{RSU} + I_0}{g_i^b}. \end{aligned} \tag{55}$$

■

REFERENCES

[1] X. Xu, X. Zhang, X. Liu, J. Jiang, L. Qi, and M. Z. A. Bhuiyan, "Adaptive computation offloading with edge for 5G-envisioned internet of connected vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 8, pp. 5213–5222, Aug. 2021.

[2] K. Xiong, S. Leng, C. Huang, C. Yuen, and Y. L. Guan, "Intelligent task offloading for heterogeneous V2X communications," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 4, pp. 2226–2238, Apr. 2021.

[3] M. Feng, M. Krunz, and W. Zhang, "Joint task partitioning and user association for latency minimization in mobile edge computing networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 8108–8121, Aug. 2021.

[4] Y. Zhai, W. Sun, J. Wu, L. Zhu, J. Shen, X. Du, and M. Guizani, "An energy aware offloading scheme for interdependent applications in software-defined IoV with fog computing architecture," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3813–3823, Jun. 2021.

[5] C. Zheng, S. Liu, Y. Huang, and L. Yang, "Hybrid policy learning for energy-latency tradeoff in MEC-assisted VR video service," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9006–9021, Sep. 2021.

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

[7] X. Huang, L. He, X. Chen, L. Wang, and F. Li, "Revenue and energy efficiency-driven delay-constrained computing task offloading and resource allocation in a vehicular edge computing network: A deep reinforcement learning approach," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8852–8868, Jun. 2022.

[8] S. M. A. Kazmi, T. N. Dang, I. Yaqoob, A. Ndikumana, E. Ahmed, R. Hussain, and C. S. Hong, "Infotainment enabled smart cars: A joint communication, caching, and computation approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8408–8420, Sep. 2019.

[9] A. Baumgartner, S. Hinz, and C. Wiedemann, "Efficient methods and interfaces for road tracking," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 34, pp. 28–31, 2002.

[10] R. Yu, Y. Zhang, S. Gjessing, W. Xia, and K. Yang, "Toward cloud-based vehicular networks with efficient resource management," *IEEE Netw.*, vol. 27, no. 5, pp. 48–55, Sep. 2013.

[11] A. A. Alahmadi, A. Q. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Distributed processing in vehicular cloud networks," in *Proc. 8th Int. Conf. Netw. Future (NOF)*, London, U.K., Nov. 2017, pp. 22–26.

[12] M. Nabi, R. Benkoczi, S. Abdelhamid, and H. S. Hassanein, "Resource assignment in vehicular clouds," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

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

[14] Q. Wu, H. Ge, H. Liu, Q. Fan, Z. Li, and Z. Wang, "A task offloading scheme in vehicular fog and cloud computing system," *IEEE Access*, vol. 8, pp. 1173–1184, 2020.

[15] K. Zhang, Y. Mao, S. Leng, S. Maharjan, and Y. Zhang, "Optimal delay constrained offloading for vehicular edge computing networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[16] F. Lin, X. Lü, I. You, and X. Zhou, "A novel utility based resource management scheme in vehicular social edge computing," *IEEE Access*, vol. 6, pp. 66673–66684, 2018.

[17] G. Premsankar, B. Ghaddar, M. Di Francesco, and R. Verago, "Efficient placement of edge computing devices for vehicular applications in smart cities," in *Proc. IEEE/IFIP Netw. Oper. Manag. Symp. (NOMS)*, Apr. 2018, pp. 1–9.

[18] Z. Zhou, P. Liu, Z. Chang, C. Xu, and Y. Zhang, "Energy-efficient workload offloading and power control in vehicular edge computing," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Barcelona, Spain, Apr. 2018, pp. 191–196.

[19] Z. Zhou, P. Liu, J. Feng, Y. Zhang, S. Mumtaz, and J. Rodriguez, "Computation resource allocation and task assignment optimization in vehicular fog computing: A contract-matching approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3113–3125, Apr. 2019.

[20] C. Zhu, G. Pastor, Y. Xiao, Y. Li, and A. Ylae-Jaeaeski, "Fog following me: Latency and quality balanced task allocation in vehicular fog computing," in *Proc. 15th Annu. IEEE Int. Conf. Sens., Commun., Netw.*, Hong Kong, Jun. 2018, pp. 1–9.

[21] B. Yu and F. Bai, "PYRAMID: Probabilistic content reconciliation and prioritization for V2V communications," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6615–6626, Jul. 2018.

[22] Y. Hui, Z. Su, T. H. Luan, and J. Cai, "Content in motion: An edge computing based relay scheme for content dissemination in urban vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 3115–3128, Aug. 2018.

[23] J. Zhao, M. Kong, Q. Li, and X. Sun, "Contract-based computing resource management via deep reinforcement learning in vehicular fog computing," *IEEE Access*, vol. 8, pp. 3319–3329, 2020.

[24] S. M. A. Kazmi, T. N. Dang, I. Yaqoob, A. Manzoor, R. Hussain, A. Khan, C. S. Hong, and K. Salah, "A novel contract theory-based incentive mechanism for cooperative task-offloading in electrical vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8380–8395, Jul. 2022.

[25] K. Xiong, S. Leng, X. Chen, C. Huang, C. Yuen, and Y. L. Guan, "Communication and computing resource optimization for connected autonomous driving," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 12652–12663, Nov. 2020.

[26] J. Shi, J. Du, J. Wang, J. Wang, and J. Yuan, "Priority-aware task offloading in vehicular fog computing based on deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16067–16081, Dec. 2020.

[27] P. Tseng, "Convergence of a block coordinate descent method for non-differentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, 2001.

[28] A. Beck and L. Tetruashvili, "On the convergence of block coordinate descent type methods," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2037–2060, 2013.

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

[30] S. Boyd, N. Parikh, and E. Chu, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, 2011.

[31] S. A. Kazmi, N. H. Tran, W. Saad, Z. Han, T. M. Ho, T. Z. Oo, and C. S. Hong, "Mode selection and resource allocation in device-to-device communications: A matching game approach," *IEEE Trans. Mobile Comput.*, vol. 16, no. 11, pp. 3126–3141, Mar. 2017.

[32] Q. Pham, T. LeAnh, N. H. Tran, B. J. Park, and C. S. Hong, "Decentralized computation offloading and resource allocation for mobile-edge computing: A matching game approach," *IEEE Access*, vol. 6, pp. 75868–75885, 2018.

[33] Q. Wei, W. Sun, B. Bai, L. Wang, E. G. Ström, and M. Song, "Resource allocation for V2X communications: A local search based 3D matching approach," in *Proc. IEEE Int. Conf. Commun.*, Paris, France, May 2017, pp. 1–6.

[34] M. Hasan and E. Hossain, "Distributed resource allocation for relay-aided device-to-device communication under channel uncertainties: A stable matching approach," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3882–3897, Oct. 2015.

[35] J. Wang, K. Zhu, B. Chen, and Z. Han, "Distributed clustering-based cooperative vehicular edge computing for real-time offloading requests," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 653–669, Jan. 2022.

[36] F. Sun, F. Hou, N. Cheng, M. Wang, H. Zhou, L. Gui, and X. Shen, "Cooperative task scheduling for computation offloading in vehicular cloud," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11049–11061, Aug. 2018.

[37] Z. Ning, P. Dong, X. Wang, X. Hu, J. Liu, L. Guo, B. Hu, R. Y. K. Kwok, and V. C. M. Leung, "Partial computation offloading and adaptive task scheduling for 5G-enabled vehicular networks," *IEEE Trans. Mobile Comput.*, vol. 21, no. 4, pp. 1319–1333, Apr. 2022.

[38] J. Sun, Q. Gu, T. Zheng, P. Dong, A. Valera, and Y. Qin, "Joint optimization of computation offloading and task scheduling in vehicular edge computing networks," *IEEE Access*, vol. 8, pp. 10466–10477, 2020.

[39] T. M. Ho, N. H. Tran, C. T. Do, S. M. A. Kazmi, E.-N. Huh, and C. S. Hong, "Power control for interference management and QoS guarantee in heterogeneous networks," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1402–1405, Aug. 2015.

[40] T. N. Dang, K. Kim, L. U. Khan, S. M. A. Kazmi, Z. Han, and C. S. Hong, "On-device computational caching-enabled augmented reality for 5G and beyond: A contract-theory-based incentive mechanism," *IEEE Internet Things J.*, vol. 8, no. 24, pp. 17382–17394, Dec. 2021.

[41] S. M. Ahsan Kazmi, N. H. Tran, M. H. Tai, K. L. Dong, and C. S. Hong, "Decentralized spectrum allocation in D2D underlying cellular networks," in *Proc. 18th Asia–Pacific Netw. Oper. Manag. Symp. (APNOMS)*, Kanazawa, Japan, Oct. 2016, pp. 1–6.

TRI NGUYEN DANG received the B.S. degree in information teaching from the Hue University's College of Education, Vietnam, in 2014. He is currently pursuing the Ph.D. degree in computer science and engineering under fully funded scholarship with Kyung Hee University, South Korea. His professional experience includes mobile application developer and middleware programming. His research interests include network optimization, mobile cloud computing, multi-access edge computing, and the Internet of Things. He had won Consolation Prize of Olympiad in ACM ICPC Vietnam, in 2013.

AUNAS MANZOOR received the M.S. degree in electrical engineering with a specialization in telecommunication from the National University of Science and Technology, Pakistan, in 2015, and the Ph.D. degree in computer science and engineering from Kyung Hee University, South Korea, in 2021. He is currently working as a Postdoctoral Researcher with the Department of Computer Science and Engineering, Kyung Hee University. His research interest includes applying analytical techniques for resource management in mobile cellular networks.

YAN KYAW TUN (Member, IEEE) received the B.E. degree in marine electrical systems and electronics engineering from Myanmar Maritime University, Thanlyin, Myanmar, in 2014, and the Ph.D. degree in computer science and engineering from Kyung Hee University, Seoul, South Korea, in 2021. He is currently working as a Postdoctoral Researcher with the Intelligent Networking Laboratory. His research interests include network economics, game theory, network optimization, wireless communications, wireless network virtualization, mobile edge computing, and wireless resource slicing for 5G. He received the Best Ph.D. Thesis Award in Engineering, in 2020.

S. M. AHSAN KAZMI received the Ph.D. degree from Kyung Hee University, South Korea, in 2017. He worked as a Postdoctoral Fellow at the Department of Computer Science and Engineering, Kyung Hee University, from 2017 to 2018. He is currently a Senior Lecturer with the Department of Computer Science and Creative Technologies, University of the West of England, Bristol, U.K. Prior to this, he worked as an Assistant Professor with the Institute of Information Security and Cyber-Physical System, Innopolis University, Russia, for three years. His research interests include applying analytical techniques of optimization, machine learning, and game theory to radio resource management for future networks. He received the Best KHU Thesis Award in Engineering, in 2017, and several best paper awards from prestigious conferences.

RIM HAW (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer engineering from Kyung Hee University, Seoul, South Korea, in 2008, 2010, and 2019, respectively. He is currently a Research Professor of computer science and engineering at Kyung Hee University. He holds several national and international patents. His research interests include ambient intelligent living, advanced wireless network protocols, and P2P networks. He is a member of KIISE.

SANG HOON HONG received the B.S. degree in electronic engineering from Yonsei University, Seoul, South Korea, in 1993, and the S.M. and Ph.D. degrees in electrical engineering from Harvard University, in 1998 and 2001, respectively. From 1998 to 2005, he was a Senior Member of Technical Staff at the Research and Development Laboratory, SK Hynix Semiconductor, where he designed high performance memory circuits. In 2006, he joined the Faculty at Kyung Hee University. His research interests include processing in memory circuits, light field sensors, and neuromorphic computations.

ZHU HAN (Fellow, IEEE) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 1999 and 2003, respectively. From 2000 to 2002, he was a Research and Development Engineer of JDSU, Germantown, MD. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an Assistant Professor at Boise State University, Boise, ID, USA. He is currently a John and Rebecca Moores Professor with the Electrical and Computer Engineering Department as well as with the Computer Science Department, University of Houston, Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid. He has been an AAAS Fellow, since 2019. He has been an ACM Distinguished Member, since 2019. He received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the *Journal on Advances in Signal Processing* in 2015, the IEEE Leonard G. Abraham Prize in the field of Communications Systems (Best Paper Award in IEEE Journal on Selected Areas in Communications) in 2016, and several best paper awards in IEEE conferences. He is also the Winner of 2021 IEEE Kiyo Tomiyasu Award, for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks." He was an IEEE Communications Society Distinguished Lecturer from 2015 to 2018. Since 2017, he has been a 1% Highly Cited Researcher according to Web of Science.

CHOONG SEON HONG (Senior Member, IEEE) received the B.S. and M.S. degrees in electronic engineering from Kyung Hee University, Seoul, South Korea, in 1983 and 1985, respectively, and the Ph.D. degree from Keio University, Japan, in 1997. In 1988, he joined KT, where he was involved in broadband networks as a Member of Technical Staff. Since 1993, he has been with Keio University. He was with the Telecommunications Network Laboratory, KT, as a Senior Member of Technical Staff and as the Director of the Networking Research Team, until 1999. Since 1999, he has been a Professor with the Department of Computer Science and Engineering, Kyung Hee University. His research interests include future internet, ad hoc networks, network management, and network security. He was an Associate Editor of the IEEE Transactions on Network and Service Management and the IEEE Journal of Communications and Networks. He currently serves as an Associate Editor of the *International Journal of Network Management*, and an Associate Technical Editor of the *IEEE Communications Magazine*.