IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 20 May 2023, accepted 13 June 2023, date of publication 16 June 2023, date of current version 3 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3286851

RESEARCH ARTICLE

Joint Crowdsensing and Offloading Algorithms for Edge-Assisted Internet of Intelligent Vehicles

SUNGWOOK KIM
Department of Computer Science, Sogang University, Mapo-gu, Seoul 04107, South Korea
e-mail: swkim01@sogang.ac.kr

This work was supported in part by the Ministry of Science and ICT (MSIT), South Korea, through the Information Technology Research Center (ITRC) Support Program, Supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP), under Grant IITP-2023-2018-0-01799; and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2021R1F1A1045472.

ABSTRACT In this paper, we aim to propose a new joint crowdsensing and offloading scheme that considers the benefits of social welfare. To induce the sensing participation, we adopt the ideas of *cooperative multi-agent reinforcement learning* (*CMARL*) to develop a novel crowdsensing algorithm. Due to the limitation of computation and communication resources in the IoIV system, the *Lozano, Hinojosa, and Mármol solution* (*LHMS*) is applied to solve the IoIV resource allocation problem. Our proposed crowdsensing and offloading algorithms are tightly coupled and work together to reach a consensus with reciprocal advantages. The main merits possessed by our hybrid approach are its flexibility and adaptability to current IoIV system situations. Performance evaluations on the proposed scheme show the superiority of our joint approach by comparing it with three existing baseline protocols.

INDEX TERMS Internet of Intelligent Vehicles, vehicular crowdsensing, vehicular offloading, cooperative multi-agent reinforcement learning, Lozano, Hinojosa, Mármol solution.

I. INTRODUCTION

With the advance of Internet of Thing (IoT), communication technology, and the increasing ability of data collection, all smart objects would have embedded processors and capability to communicate with each other through public Internet. IP addresses can be assigned to these smart objects, so that they can collect, refine, process, and communicate important information over a network. Recent years have witnessed the rising popularity of IoT paradigm that leverages to transform the business and consumer world in an unforeseen manner and is driving a new industrial revolution. Within the objectives of IoT, vehicles play an important role for safe and convenient travel while realizing the highly efficient intelligent transportation system. Traditionally, vehicular ad hoc networks (VANETs) are created by applying the principles of mobile ad hoc networks to insure safe drive by improving the traffic flow. For the new era of the IoT, the vehicular ad-hoc networks will further evolve into the Internet of Intelligent Vehicles (IoIV) [1].

As driving becomes more automatic in recent days, the IoIV comes to people's daily life. Recently, intelligent vehicles evolve into smart objects equipped with a multi-sensory platform as well as being able to be equipped with on-board units (OBUs). In this way, they will be able to communicate with each other and with the local mobile network infrastructure, which consists of road-side units (RSUs). As Internet access points deployed along roadsides, RSUs mainly play roles in collecting and analyzing traffic messages given from intelligent vehicles, or providing drivers with Internet access. Inaddition, the steep grow in IoIV boosts the proliferation of computation-intensive and bandwidth-hungry applications. Nowadays, industry research predicts that these services will become the dominant applications for IoIV traffic.

An on-board computer in vehicles is an electronic device with a microprocessor, made on the principle of a home PC. This device can register, calculate and display all kinds of IoIV applications. Even though the performance of on-board computer has improved due to the technological progress, the requirements of IoIV applications have increased more rapidly. Usually, IoIV applications are delay-sensitive and resource-intensive with complex computations. However,

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

individual vehicles are still of limited computing resources and insufficient storage capacity. Therefore, on-board computers in vehicles are not enough to support IoIV application requirements. Due to this reason, the IoIV has ushered in a new research direction, called mobile edge computing (MEC) with the rapid growth of the number of computing oriented applications [2], [3].

To facilitate the efficient and stable execution of onboard applications, MEC helps vehicles achieve real-time data processing by employing the computing capabilities of the edge of the network. Therefore, smart vehicles offload computing-intensive application tasks to the MEC server to meet the demands of multiple onboard applications; MEC can assist in expanding the computing capacity provided by the existing IoIV platform. Thisapproach aligns with the vision of high bandwidth, computation intensive, and real-time interaction for future networks. Therefore, RSUs equipped with MEC servers can contribute to a more effective solution to the computing offloading problem in the IoIV infrastructure [3], [4].

In recent years, intelligent vehicles have been equipped with more and more components, which are considered programmable and powerful mobile sensors. Because of vehicles' mobility along roads day after day, they have the potential to collect data for the traffic management, and environment monitoring. Based on these outsourcing related sensing tasks in mobile vehicles, mobile crowdsensing has emerged as a new sensing and computing paradigm. In a crowdsensing system, RSUs and mobile vehicles with a number of sensors collaboratively perform large-scale sensing tasks. This approach is ideal to sense up-to-date and fine-grained information for large areas. However, sending sensing data to a remote cloud server consumes a heavy bandwidth and incurs an increased transmission delay. Therefore, the MEC brings new opportunities for mobile crowdsensing applications. Benefiting from its proximity to vehicles, the main advantage of MEC-assisted crowdsensing (MECC) is to largely relieve access overheads [5], [6].

Even though there are some technical merits, the MECC system faces three challenging issues to be addressed. First, smart vehicles are self-interested and have no obligation to use their resources for crowdsensing services. In general, most vehicles have limited resources, and they want to save these resources for their own services. However, participating in crowdsensing needs to spend time and energy for collecting and uploading data. Therefore, a carefully designed incentive mechanism with reasonable rewards is indispensable. Second, the interactions among independent decision makers, such as individual vehicles, RSUs, and MEC servers, are complicated and very difficult to formulate. Due to their rationality and selfishness, it is really hard to give a blanket answer to this kind of question. Third, for the joint problem of crowdsensing and offloading services, an effective control algorithm is essential. Until now, several algorithms have been developed to deal with the crowdsensing, vehicular offloading, and MECC control problems. However, to the best of our knowledge, most existing works ignore the interconnected relationships of above mentioned challenging issues [7].

Game theory is a branch of modern mathematics for analyzing the strategic interactions among multiple decision makers. The strategic interaction activities are referred to as games where decision-makers, i.e., game players, select the strategy that will give maximum possible outcome for themselves while predicting the rational decisions taken by other players. It was originally adopted in economics to analyze the behavior of rational players in a multi-agent decision making process. But nowadays, game theory is widely applied to other areas, such as biology, sociology, politics, computer science, and engineering, especially in the telecommunication environment. In general, game theory can be categorized into two major classes, i.e., static and dynamic games. Static games, also called strategic, one-shot, single stage or simultaneous games, formulate interactions among players when they take actions only once in a single period. By contrast, dynamic games, also called sequential or repeated games, are applied when players take actions over multiple time rounds. Therefore, the players in the dynamic games can observe the behaviors of the other players in the past and are able to adjust their strategies to achieve their goals. In particular, dynamic games can be successfully applied in the field of network management [8].

In recent years, there has been increased interest in decentralized approaches to solving complex network control problems. Many such approaches fall into the area of*cooperative multi-agent reinforcement learning*(*CMARL*), which has attained substantial attention in the machine learning community. It facilitates each agent to perform objectives by interacting with its neighbor agents while improving the robustness and efficiency of system. In the *CMARL*, agents can actively sense the environment via different actions, evaluate the actions and adjust subsequent actions. Therefore, an individual agent can be mapping different environmental states into specific actions. As a result, the *CMARL*shares better performance in cooperative action acquisition. This approach concerns how agents choose actions to avoid cooperation dilemmas in an environment, and make all agents come to a joint agreement [9].

In this study,we propose a new joint control scheme for the crowdsensing and offloading control problems in the MECC platform. Motivated by the above discussion, the main principles of dynamic game theory and *CMARL*are employed to develop our joint crowdsensing and offloading control process. To strike an appropriate performance balance among different service requirements, the proposed scheme consists of crowdsensing and offloading algorithms. In the crowdsensing algorithm, individual vehicles decide their contributions for the crowdsensing process. Based on the *CMARL*, each vehicle can learn his best sensing action in a distributed manner. In the offloading algorithm, RSUs and MEC servers work together to handle the offload services. According to the dynamic game model, RSUs provide communication resources, and MEC servers offer computing resources in a centralized manner. Our two control algorithms are tightly coupled with each other to successfully operate crowdsensing and offloading services in the MECC system infrastructure.

The rest of this paper is organized as follows. In Section II, we introduce the basic technical concepts and main contributions of our proposed scheme. Section III reviews the related work. Section IV describes the scenario of MECC platform, and formulate our proposed control model. In addition, we introduce the necessary background of the dynamic game theory and multi-agent learning model. And then, our proposed scheme is presented in detail. To help readers understand better, we also provide the primary steps of the proposed algorithm. Section V evaluates the performance of the proposed scheme through extensive simulations by comparing the state-of-the-art benchmark protocols. Finally, the conclusion is presented and future research plans are shown in Section VI.

II. TECHNICAL CONCEPTS AND STUDY MOTIVATION

In dynamic games, players choose and implement strategies from their respective possible strategy sets and obtain corresponding benefits successively for many times.Dynamic bargaining game is a branch of dynamic gamesto deal the decision-making problem in complex and interactive strategies. Original, the idea of bargaining game theory was initiated by J. Nash. His bargaining game model is formulated by an expected utility function over the set of feasible agreements and the outcome which would result in case of disagreement. For a specification, the solution concept of Nash bargaining game is to determine an outcome that consists of players' payoffs. Recently, there are multiple alternative approaches to classical bargaining problems [10].

In 2019, Lozano, Hinojosa, and Mármol extended the original Nash bargaining solution for the data envelopment analysis. The developed *Lozano, Hinojosa, and Mármol solution*(*LHMS*) has some properties, which are derived as a consequence of original Nash solution. Explicitly, the *LHMS*approach considers the projection of an inefficient decision making unit onto the efficient frontier as a bargaining problem, in which the players are the input and output dimensions. The *LHMS*offers a new perspective on projecting the decision making unit, one in which the input and output target values are not decided globally from a system point of view, but as the result of a bargaining process between the different input and output variables, using the observed values as disagreement point. Therefore, the solution is decided by the result of an impartial, rational arbitration between the inputs and output variables [10].

Reinforcement learning is a prevalent method to optimize the agent's strategy in a Markov decision process. The majority of the success in reinforcement learning has been in single-agent environments. However, many real-world problems require multiple agents to interact and cooperate.With the popularity of game theory, multi-agent learning has garnered attention and become a prevalent method for solving cooperative problems. To ensure the capability of improving robustness and efficiency, the *CMARL* facilitates multiple agents to perform objectives by interacting with their neighbor agents. This method can obtain the optimal joint strategy that maximizes the expected common cumulative reward for all agents. Based on the game-theoretic approach, the analysis of *CMARL* dynamics has focused on repeated games with few agents and actions [11].

In this study, the study motivation is to optimize the MECC system performance by using the proposed crowdsensing and offloading algorithms. Based on the*LHMS* and *CMARL*, we intelligently encourage vehicles to participate in the crowdsensing process while providing the adaptive offloading services. For the efficient operation of MECC system infrastructure, our crowdsensing and offloading algorithms are sophisticatedly combined, and operated in a step-by-step repeated manner. Therefore, we harness the synergies between the cooperative learning and bargaining game to achieve a socially optimal solution. The significant major contributions of our study are summarized as follows:

- Based on the basic concepts of the *CMARL* and *LHMS*, we design a new joint crowdsensing and offloading scheme to strike the appropriate performance balance between efficiency and social welfare.
- For the crowdsensing, individual vehicles cooperatively learn their sensing strategies in a distributed manner. For the task offloading, the RSU and MEC server address the resource allocation problem, and reach a consensus in a centralized manner.
- Our learning and bargaining processes are jointly combined to get reciprocal advantages. This hybrid approach explores the sequential interactions among RSUs, MEC servers and vehicles. The simulation results reveal the effectiveness of our joint control scheme by comparing to the existing control protocols.

III. RELATED WORK

In this section, we touch on research literatures relevant to the research topic of this study. Recently, there have been a few papers focusing on the incentive design for mobile crowdsensing, and handling the joint resource allocation problem for offloading services [5], [7], [12], [15]. In [5], L. Liu et al. have studied the problem of participant recruitment in edge-assisted crowdsensing system, and propose the *Incentive-aware Vehicle Recruitment for Mobile Crowdsensing* (*IVRMC*) scheme. Due to the selfishness of intelligent vehicles, the design of vehicle recruitment in edge-assisted mobile crowdsensing is very challenging. To deal with this challenge, the *IVRMC* scheme designs a new incentive mechanism to motivate cooperation among intelligent vehicles, and implements the Nash bargaining theory to obtain the optimal cooperation decision. To weigh the contribution of vehicles, a practical and efficient algorithm is also proposed based on the vehicular spatiotemporal availability, the vehicular reputation, and the priority of regions. In addition, the participant recruitment is formulated as an optimization problem, and an effective heuristic algorithm is also proposed by leveraging the property in submodular optimization. To validate the superiority of the *IVRMC* scheme, extensive simulations are conducted and numerical results show that the *IVRMC* scheme outperforms the other methods [5].

The paper [12] proposes the*Combination of Auction and Incentive for Mobile Crowdsensing* (*CAIMC*) scheme that considers the benefits of both the platform and participants. For vehicular crowdsensing systems, the *CAIMC* scheme designs an effective incentive method to solve the problem of low willingness to participate the sensing process. As a novel incentive method, the reputation-based reverse combination auction is developed and the conversion between virtual currency and monetary rewards is considered for vehicle participations in crowdsensing activities. This method divides perception tasks and vehicle data, and adopts a reverse auction by using a winner selection algorithm. Under a limited budget, the reverse auction approach can increase the task coverage rate and the profit of the platform. In addition, the *CAIMC* scheme adds the vehicle reputation to the incentive method while ensuring the quality of data uploaded by intelligent vehicles. Finally, simulation results show that the *CAIMC* scheme has good performance in crowdsensing participation, budget controlling and auction winner selections [12].

The*Contract Stackelberg based Offloading Resource Allocation*(*CSORA*) scheme formulates a vehicular fog-edge computing paradigm as a multistage Stackelberg game [7]. To utilize individual vehicles for computation offloading, a contract-based incentive mechanism is developed in the first stage to manage the idle computing resources. In order to overcome the information asymmetry, the *CSORA* scheme can satisfy individual rationality and incentive compatibility of vehicles while maximizing the system utility. In the second stage, the edge server determines the quantity of computing resources and broadcasts the price of resources to all mobile users. In the third stage, each mobile user decides whether to offload its task and the quantity of computing resources purchased from the edge server. The *CSORA* scheme can solve this three-stage control problem by using the backward induction method, and obtains effective price and computing resource demand strategies in each stage. Numerical results demonstrate the effectiveness of the *CSORA* scheme while showing the performance comparison with the existing vehicular offloading protocols [7].

The paper [15] proposes a cloud-based MEC offloading scheme in vehicular networks. This scheme investigates the computation offloading mechanism. Both the resource limitation of the MEC servers and the latency tolerance of the computation tasks are considered in the design of the mechanism. Furthermore, this study designs a contract-based offloading and computation resource allocation algorithms. The developed method aims to maximize the utility of the MEC service provider while satisfying the offloading requirements of the tasks [15].

As discussed above, the earlier schemes in [5], [7], and [12] have been studied on the traditional crowdsensing systems and offloading services for the use in vehicular networks. Although these studies tackled the crowdsensing and offloading control problems, they did not consider the interconnected relationship of control issues for the IoIV platform. Unlike the aforementioned the *IVRMC*, *CAIMC* and *CSORA* schemes, our proposed scheme combines ideas of *CMARL* and *LHMS* for controlling the cooperative activities of IoIV agents, and guides them toward a socially optimal outcome. To the best of our knowledge, our joint control paradigm based on the bargaining game solution and multi-agent learning algorithm is the first in the literature to ensure a well-balanced performance for the edge assisted IoIV system infrastructure.

IV. CROWDSENSING AND OFFLOADING SERVICES FOR THE MECC SYSTEM

In this section, we first describe the MECC based IoIV scenario that we focus on. Then, we introduce the fundamental concepts of bargaining game and cooperative learning model in the MECC platform. After that, we present our proposed scheme in detail for the crowdsensing and offloading services.

A. IoIV SYSTEM INFRASTRUCTURE AND BARGAINING AND LEARNING MODELS

We consider a typical edge assisted IoIV system platform, which consists of n RSUs, i.e., $\mathbb{R} = \{\mathcal{R}_1, \ldots, \mathcal{R}_n\}$, n MEC servers, i.e., $\mathbb{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_n\}$ and m intelligent vehicles, i.e., $\mathbb{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_m\}$. Each RSU is equipped with one MEC server where $\mathcal{R}_i \in \mathbb{R}$ and $\mathcal{M}_i \in \mathbb{M}$ are tightly coupled as a pair $[\mathcal{R}_i, \mathcal{M}_i]$. $\mathcal{R}_{1\leq i\leq n}$ has a coverage area and a fixed spectrum band $\left(\mathfrak{M}_{\mathcal{R}_i}\right)$, and its equipped $\mathcal{M}_i$ has its own computation capacity $\left(\mathfrak{N}_{\mathcal{M}_i}\right)$. Each vehicle $\left(\mathcal{V}_{1\leq j\leq m}\right)$ is covered by its corresponding RSU, and has the offload task $\mathcal{T}_{\mathcal{V}_j} = \left\{\mathcal{B}_{\mathcal{V}_j}, \mathcal{C}_{\mathcal{V}_j}\right\}$ where $\mathcal{B}_{\mathcal{V}_j}$ and $\mathcal{C}_{\mathcal{V}_j}$ are required communication and computation resources. When the $\mathcal{V}_j$ runs within the $\mathcal{R}_i$'s covering area, the $\mathcal{V}_j$ offloads his $\mathcal{T}_{\mathcal{V}_j}$ to the $\mathcal{M}_i$ through $\mathcal{R}_i$. A general edge assisted IoIV infrastructure is shown in Fig.1 [5].

In this paper, individual vehicles are considered as employees for the mobile crowdsensing; they use their sensors to perform the sensing tasks and upload the sensing data to their corresponding RSUs. While performing crowdsensing tasks, vehicles consume their own resources and energy. Hence, ordinary vehicles are reluctant to perform crowdsensing tasks unless they receive satisfactory rewards to compensate their efforts. To compensate the cost incurred by the crowdsensing participants, we provide incentives to the participating vehicles for their offloading services. For the participant recruitment in crowdsensing, the $\mathfrak{M}_{\mathcal{R}}$ and $\mathfrak{N}_{\mathcal{M}}$ resources are adaptively distributed in terms of compensations; they are used for vehicles' offloading services. From the viewpoint of $\mathcal{R}$ and $\mathcal{M}$, their resources proportionally allocate based on the sensing contribution of each vehicle. From the viewpoint of vehicles, they act their crowdsensing work to maximize the profits for their offloading services.

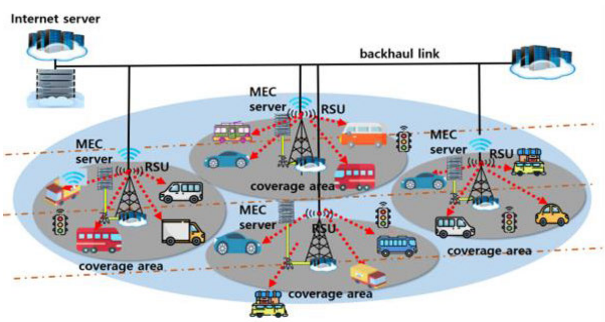

FIGURE 1. The edge assisted IoIV platform infrastructure.

In the proposed scheme, we adopt the joint control paradigm using learning model($\mathbb{L}$) and bargaining game ($\mathbb{G}$). Through the $\mathbb{L}$ and $\mathbb{G}$, multiple RSUs, MEC servers and individual vehicles aresequentially interacted with each other to reach a desired outcome, which is called social optimum. It is noteworthy that we formulate the $\mathcal{R}-\mathcal{M}-\mathcal{V}$ association in an iterative manner. Formally, we define the tuple entities in our proposed $\mathbb{L}$ and $\mathbb{G}$, such as shown in the equation at the bottom of the next page.

- $\mathbb{R}$, $\mathbb{M}$ and $\mathbb{V}$ represent the sets of RSUs, MEC servers and vehicles, respectively.
- Each RSU has its fixed spectrum band ($\mathfrak{M}_{\mathcal{R}}$) and is equipped with one MEC server. Each embedded MEC server has its computation capacity ($\mathfrak{N}_{\mathcal{M}}$).
- Each vehicle in $\mathbb{V}$ participates in its learning model ($\mathbb{L}_{\mathcal{V}}$). Let $\mathbb{V}_{\mathcal{R}_i}$ be the set of vehicles, which are covered by the $\mathcal{R}_i$.
- In the $\mathbb{L}_{\mathcal{V}_j}$, $\mathcal{V}_j \in \mathbb{V}_{\mathcal{R}_i}$ has hisaction set $\left(\mathcal{L}_{\mathcal{V}_j}\right)$, which consists of total l actions $\left(a^{\mathcal{V}_j}_{1\leq k\leq l}\right)$. $\mathcal{R}^{\mathcal{V}_j}_{\boldsymbol{a}}$ is the $\mathcal{V}_j$'s reward function with the joint action $\boldsymbol{a}$, and $\mathcal{H}_{\mathcal{V}_j}(\boldsymbol{a})$ is a heuristic function to accelerate the learning process.
- Based on the *CMARL*model, the $\mathbb{L}_{\mathcal{V}_j}$ is operated in a distributed manner to learn the best action of $\mathcal{V}_j$.
- Each pair of RSU and MEC server operate their bargaining game $\left(\mathbb{G}^{\mathcal{M}}_{\mathcal{R}}\right)$. In the $\mathbb{G}^{\mathcal{M}_i}_{\mathcal{R}_i}$, $\mathfrak{X}_{\mathcal{V}_j}$ is the $\mathcal{V}_j$'s contribution to perform crowdsensing tasks where $\mathcal{V}_j \in \mathbb{V}_{\mathcal{R}_i}$.
- In the $\mathbb{G}^{\mathcal{M}_i}_{\mathcal{R}_i}$, the $\mathcal{V}_j$ is a game player, and $S_{\mathcal{V}_j}$ and $\mathcal{U}_{\mathcal{V}_j}(\cdot)$ are his strategy and utility function, respectively, to share the $\mathfrak{M}_{\mathcal{R}_i}$ and $\mathfrak{N}_{\mathcal{M}_i}$. According to the *LHMS*, the $\mathbb{G}^{\mathcal{M}_i}_{\mathcal{R}_i}$ is operated in a centralized manner.
- The $\mathbb{L}_{\mathcal{V}_j}$ and $\mathbb{G}^{\mathcal{M}_i}_{\mathcal{R}_i}$ are reciprocally interdependent each other based on $\mathfrak{X}_{\mathcal{V}_j}$, $\mathcal{H}_{\mathcal{V}_j}(\boldsymbol{a})$ and $\mathcal{U}_{\mathcal{V}_j}(\cdot)$, and work together, iteratively.
- Discrete time model $T \in \{t_1, \ldots, t_c, t_{c+1}, \ldots\}$ is represented by a sequence of time steps. The length of t_c matches the event time-scale of $\mathbb{L}_{\mathcal{V}_j}$ and $\mathbb{G}^{\mathcal{M}_i}_{\mathcal{R}_i}$.

B. TECHNICAL CONCEPTS AND IDEAS OF LHMS AND CMARL

In this subsection, we quickly review the fundamental concepts of *LHMS*, and the cooperative multi-agent reinforcement learning based on the joint action.

1) LOZANO, HINOJOSA, AND MÁRMOL SOLUTION FOR BARGAINING GAMES

Data envelopment analysis (DEA) is a non-parametric methodology mostly used for assessing and improving the relative efficiency of a set of homogeneous operating units. DEA uses the observed data about the input consumption and the output production of the decision making units to infer a production possibility set.Based on the ideas underlying the theory of bargaining, the *LHMS* is proposed to handle the DEA model. To characterize the idea of *LHMS*, the following notations will be used. J is the set of operating units, and I, O are the sets of inputs and outputs, respectively. For a given operating unit j, consider I^-_j and O^+_j are sets of inputs and outputs, respectively. For the cases in which several inputs and/or outputs could improve their values, the classical Nash bargaining theory can be adapted. Assume that inputs in I^-_j are input players, and outputs in O^+_j are output players. The bargaining set of the game, denoted by S_j, is a $\left|I^-_j\right| \times \left|O^+_j\right|$ $-$dimensional feasible set in the utility space [10].

For the output players in O^+_j, the utilities coincide with the corresponding output variable. These utilities are represented by v, where for a given amount of output y_r, $r \in O^+_j$, $v_r(y_r)$ is its utility. However, for the input players in I^-_j, the utilities decrease with the amount of input consumed. Therefore, we can assume the input utilities with respect to the worst case. These utilities are represented by u, where for a given amount of input x_k, $k \in I^-_j$, its utility is defined as $u_k(x_k) = \left(x^*_k = \max_{j\in J} x_{kj}\right) - x_k$, if the feature of linearity is adopted; x_{kj} is the amount of input k, consumed by operating unit j. Usually, the utilities coincide with the corresponding input and output variable. It means that there is a one-to-one correspondence between the operating points $\left(\hat{x_k}, \hat{y_r}\right)$ and their corresponding utility vectors $\left(\left(u_k\left(\hat{x_k}\right)\right)_{k\in I^-_j}, \left(v_r\left(\hat{y_r}\right)\right)_{r\in O^+_j}\right)$. The disagreement point corresponds to the utility of the players is $d_j = \left(d^x_j, d^y_j\right) = \left(\left(d^x_{1j}, \ldots d^x_{kj}, \ldots, d^x_{\left|I^-_j\right|j}\right), \left(d^y_{1j}, \ldots d^y_{rj}, \ldots, d^y_{\left|O^+_j\right|j}\right)\right)$ [10].

Based on the Nash solution, the *LHMS*is defined by taking into account the one-to-one correspondence between the utility vectors $(u, v) \in S_{d_j}$ and their corresponding operating points $(\hat{x}, \hat{y})$. Mathematically, the *LHMS*for (u, v), i.e., *LHMS* (u, v), is given by [10];

$$
\begin{aligned}
&LHMS\,(u, v) = \\
&\arg\max_{(\hat{x},\hat{y})} \left\{ \prod_{k\in I^-_j} \left(u_k\left(\hat{x_k}\right) - d^x_{kj}\right) \times \prod_{r\in O^+_j} \left(v_r\left(\hat{y_r}\right) - d^y_{rj}\right) \right\} \\
&\text{s.t.}, \left(\left(u_k\left(\hat{x_k}\right)\right)_{k\in I^-_j}, \left(v_r\left(\hat{y_r}\right)\right)_{r\in O^+_j}\right) \in S_{d_j} \qquad (1)
\end{aligned}
$$

2) MULTI-AGENT REINFORCEMENT LEARNING FOR GAME THEORY

Traditionally, reinforcement learning is a model-free type of machine learning which is aimed at learning the desirability of taking any available action in any state of the environment through online learning framework. This desirability of an action is represented by the expected cumulative reward, as known as $Q-$value; for taking a particular action in a particular state. Therefore, $Q-$values are stored in a two-dimensional array for every action in every state. However, in some cases, an environment does not necessary to be represented by states. Only one-dimensional action array is enough to learn the best action. For learning agents, the main goal of stateless learning approach is to estimate an expected $Q-$value of a single reward for each action. Due to its simplicity of implementation, the stateless learning method is practically applicable to the *CMARL*; it can dramatically reduce the complexity of the traditional multi-agent learning algorithm [13], [14].

Under the multi-agent environment, each agent has to consider not only the effects of his own actions, but also the influence by the actions of the other agents. From this perspective, agents are assumed as joint action learners, and they are able to observe all actions taken by any agent. Therefore, the multi-agent learning is intrinsically linked to the game theory, which analyzes the strategic interactions among multiple decision makers. In our stateless*CMARL* model, we assume the $Q-$value of agent i, i.e., $Q_i(a_i, \boldsymbol{a}_{-i})$, that provides an estimate of the value of performing joint action $\boldsymbol{a} = (a_i, \boldsymbol{a}_{-i})$. The sample $\langle (a_i, \boldsymbol{a}_{-i}), \mathcal{R} \rangle$ is the experience obtained by the agent i where the joint action $\boldsymbol{a}$ was performed resulting in the reward $\mathcal{R}$. At the beginning, all $Q-$values of all actions are initialized to zero, so all agents start learning with equal choice among all available actions. Each $Q-$value is updated by an agent each time. Based on the $\langle (a_i, \boldsymbol{a}_{-i}), \mathcal{R} \rangle$, the recursive update equation for $Q_i(a_i, \boldsymbol{a}_{-i})$ is defined as follows [13], [14]:

$$Q_i(a_i, \boldsymbol{a}_{-i}) = [(1-\alpha) \times Q_i(a_i, \boldsymbol{a}_{-i})] + (\alpha \times \mathcal{R}) \tag{2}$$

where $\alpha \in [0, 1]$ is the learning rate parameter which weights recent experience with respect to previous estimates of the $Q-$values. In addition, we can take a heuristic acceleration technique for the reinforcement learning process. From external or internal additional knowledge, a heuristic function $(\mathcal{H}(\boldsymbol{a}))$ is derived; it is not included in the learning process. Generally, The $\mathcal{H}(\boldsymbol{a})$ can influence the action selections of a learning agent to accelerate the learning process. The format and dimensions of $\mathcal{H}(\boldsymbol{a})$ should be compliant with the $Q-$table; used by the given learning agent. By employing the $\mathcal{H}(\boldsymbol{a})$, we can create the heuristically guided $Q-$value of agent i's action, i.e., $Q_i^{\mathcal{H}}(a_i, \boldsymbol{a}_{-i})$, while a classical learning process takes place using the $Q_i(a_i, \boldsymbol{a}_{-i})$, as defined in (2) [13].

$$Q_i^{\mathcal{H}}(a_i, \boldsymbol{a}_{-i}) = Q_i(a_i, \boldsymbol{a}_{-i}) + \mathcal{H}(\boldsymbol{a}) \tag{3}$$

Let $\mathcal{L}_i$ be the action set of agent i, and $\mathbb{P}_i = \left(p_1^i, p_2^i, \ldots, p_{|\mathcal{L}_i|}^i\right)$ is the probability set for action selections where $p_{1\le j\le|\mathcal{L}_i|}^i$ is the probability of selecting the j^{th} action of agent i. The p_j^i is updated as follows:

$$p_j^i = exp\left(\frac{Q_i^{\mathcal{H}}(j)}{\tau}\right) / \sum_{k=1}^{|\mathcal{L}_i|} exp\left(\frac{Q_i^{\mathcal{H}}(k)}{\tau}\right) \tag{4}$$

where $Q_i^{\mathcal{H}}(j)$ is the $Q_i^{\mathcal{H}}(\cdot)$ value for the j^{th} action, and τ is the temperature parameter that balances exploration and exploitation [11].

C. THE PROPOSED JOINT CONTROL SCHEME FOR THE MECC BASED IOIV PARADIGM

To develop our joint control scheme for the MECC based IoIV paradigm, we construct the learning ($\mathbb{L}$) and game ($\mathbb{G}$) models for each vehicle. In the $\mathbb{L}$, a stateless *CMARL* process is implemented in a distributed manner to learn the best sensing action. For the crowdsensing process, the incentive method plays an important role in ensuring the quantity and quality of sensed data. To recruit sensing participants, our IoIV platform provides offloading services to compensate vehicles' sensing costs based on the sensing contribution $(\mathfrak{X}_{\mathcal{V}})$. Therefore, intelligent vehicles are motivated to participate in the crowdsensing process in the long term. From the viewpoint of vehicles, major challenge is to maximize its reward while minimizing the crowdsensing cost. In the learning model of $\mathcal{V}_j$ $\left(\mathbb{L}_{\mathcal{V}_j}\right)$, the crowdsensing activity of $\mathcal{V}_j$ is discretely varied, and each activity level is defined as the $\mathcal{V}_j$'s action. We assume that there are l different actions where $a_{1\le k\le l}^{\mathcal{V}_j} \in \mathcal{L}_{\mathcal{V}_j}$. After an action in the $\mathcal{L}_{\mathcal{V}_j}$ is selected, it is estimated by receiving a reward, and what is the best action is gradually learned.

In our stateless*CMARL*setting, we assume that the $\mathcal{V}_j$ has his action set $\mathcal{L}_{\mathcal{V}_j} = \left\{a_1^{\mathcal{V}_j}, \ldots a_k^{\mathcal{V}_j} \ldots, a_l^{\mathcal{V}_j}\right\}$ where each action represents its crowdsensing participation level. For example, the $Q-$value of $\mathcal{V}_j$'s k^{th} action, i.e., $Q_{1\le k\le l}^{\mathcal{V}_j}\left(a_k^{\mathcal{V}_j}, \boldsymbol{a}_{-\mathcal{V}_j}\right)$, provides an estimate value of performing the joint action $\mathbf{a} = \left(a_k^{\mathcal{V}_j}, \boldsymbol{a}_{-\mathcal{V}_j}\right)$. Simply, sensing actions are mapped to

$$\{\mathbb{L}, \mathbb{G}\} = \left\{\begin{array}{c} \mathbb{R}, \mathbb{M}, \mathbb{V}, \\ \left(\mathbb{L}_{\mathcal{V}_{1\le j\le m}} = \left\{\mathbb{L}_{\mathcal{V}_j} | \mathcal{V}_j \in \mathbb{V}_{\mathcal{R}_i}, a_{1\le k\le l}^{\mathcal{V}_j} \in \mathcal{L}_{\mathcal{V}_j}, \mathcal{R}_{\boldsymbol{a}}^{\mathcal{V}_j}, \mathcal{H}_{\mathcal{V}_j}(\boldsymbol{a})\right\}\right), \\ \left(\mathbb{G}_{\mathcal{R}_{1\le i\le n}}^{\mathcal{M}_{1\le i\le n}} = \left\{\mathbb{G}_{\mathcal{R}_i}^{\mathcal{M}_i} | \mathcal{V}_j \in \mathbb{V}_{\mathcal{R}_i}, \mathfrak{X}_{\mathcal{V}_j}, \mathfrak{M}_{\mathcal{R}_i}, \mathfrak{N}_{\mathcal{M}_i}, S_{\mathcal{V}_j}, \mathcal{U}_{\mathcal{V}_j}(\cdot)\right\}\right), \\ T \end{array}\right\}$$

real numbers, which represent vehicles' contributions. The $\mathcal{V}_j$ updates its estimate $Q_k^{\mathcal{V}_j}(\cdot)$ value based on the experience sample $\left\langle \boldsymbol{a},\ \mathcal{R}_{\boldsymbol{a}}^{\mathcal{V}_j}, \mathcal{H}_{\mathcal{V}_j}(\boldsymbol{a})\right\rangle$ where $\mathcal{R}_{\boldsymbol{a}}^{\mathcal{V}_j}$ and $\mathcal{H}_{\mathcal{V}_j}(\boldsymbol{a})$ arethe $\mathcal{V}_j$'s reward and heuristic functions of the joint action $\boldsymbol{a}$.The $\mathcal{R}_{\boldsymbol{a}}^{\mathcal{V}_j}$ corresponds to the received $\mathcal{V}_j$'s benefit $\left(B_{\mathcal{V}_j}(\cdot)\right)$ minus the incurred sensing cost $\left(C_{\mathcal{V}_j}(\cdot)\right)$. The $B_{\mathcal{V}_j}(\cdot)$ is defined by considering the $\mathcal{V}_j$'s currently generating offload services and sensing action. The $C_{\mathcal{V}_j}(\cdot)$ increases monotonically with the sensing contribution. When the $\mathcal{V}_j$ selects the action $a_k^{\mathcal{V}_j}$, the $\mathcal{R}_{\boldsymbol{a}}^{\mathcal{V}_j}$ is given by;

$$\mathcal{R}_{\boldsymbol{a}}^{\mathcal{V}_j}\left(a_k^{\mathcal{V}_j}\right) = B_{\mathcal{V}_j}\left(a_k^{\mathcal{V}_j}, \sigma_{\mathcal{V}_j}\right) - C_{\mathcal{V}_j}\left(a_k^{\mathcal{V}_j}\right)$$
$$\text{s.t.,}\begin{cases} B_{\mathcal{V}_j}\left(a_k^{\mathcal{V}_j}, \sigma_{\mathcal{V}_j}\right) = \log\left(\lambda + \left(\dfrac{\sigma_{\mathcal{V}_j}}{\mathcal{X}_{\mathcal{V}_j}} \times a_k^{\mathcal{V}_j}\right)\right) \\ C_{\mathcal{V}_j}\left(a_k^{\mathcal{V}_j}\right) = \left(\beta \times \dfrac{\sigma_{\mathcal{V}_j}}{\mathcal{X}_{\mathcal{V}_j}} \times a_k^{\mathcal{V}_j}\right)^{\varepsilon} \end{cases} \tag{5}$$

where λ is an adjustment parameterfor the $B_{\mathcal{V}_j}(\cdot)$, $\sigma_{\mathcal{V}_j}$ is the $\mathcal{V}_j$'s current offload task amount, $\mathcal{X}_{\mathcal{V}_j}$ is the $\mathcal{V}_j$'s local computation capacity, respectively. β and ε are cost control factors for the $C_{\mathcal{V}_j}(\cdot)$. In the proposed scheme, we adopt the heuristic function $\left(\mathcal{H}_{\mathcal{V}_j}(\boldsymbol{a})\right)$ to effectively estimate its incentive for the cooperative crowdsensing process. The $\mathcal{H}_{\mathcal{V}_j}(\boldsymbol{a})$ is strongly related to the $\mathcal{U}_{\mathcal{V}_j}(\cdot)$, which is defined in the game model $\mathbb{G}$. Mathematically, the $\mathcal{H}_{\mathcal{V}_j}(\boldsymbol{a})$ is defined according to the rate of $\mathcal{V}_j$'s relative utility change.

$$\mathcal{H}_{\mathcal{V}_j}(\boldsymbol{a}) = \frac{\mathcal{U}_{\mathcal{V}_j}(\boldsymbol{a}) - \mathcal{A}\_\mathcal{U}_{\mathcal{V}_j}(\boldsymbol{a})}{\mathcal{A}\_\mathcal{U}_{\mathcal{V}_j}(\boldsymbol{a})} \tag{6}$$

where $\mathcal{A}\_\mathcal{U}_{\mathcal{V}_j}(\boldsymbol{a})$ is the accumulated average utility of $\mathcal{V}_j$, which is obtained from its the offloading services. Finally, our stateless *CMARL* $Q-$function; is defined according to (2)-(3),(5)-(6). From this $Q-$value, the $a_k^{\mathcal{V}_j}$'s selection probability by the $\mathcal{V}_j$, i.e., $p_{a_k}^{\mathcal{V}_j}$, is updated as follows:

$$p_{a_k}^{\mathcal{V}_j} = exp\left(\frac{Q_{a_k}^{\mathcal{V}_j}\left(a_k^{\mathcal{V}_j}, \boldsymbol{a}_{-\mathcal{V}_j}\right)}{\tau}\right) \times \Bigg/ \sum_{l=1}^{\left|\mathcal{L}_{\mathcal{V}_j}\right|} exp\left(\frac{Q_{a_l}^{\mathcal{V}_j}\left(a_l^{\mathcal{V}_j}, \boldsymbol{a}_{-\mathcal{V}_j}\right)}{\tau}\right)$$
$$\text{s.t.,} Q_{a_k}^{\mathcal{V}_j}\left(a_k^{\mathcal{V}_j}, \boldsymbol{a}_{-\mathcal{V}_j}\right) = \mathcal{R}_{\boldsymbol{a}}^{\mathcal{V}_j}\left(a_k^{\mathcal{V}_j}\right) + \mathcal{H}_{\mathcal{V}_j}(\boldsymbol{a}) \tag{7}$$

Based on its sensing action, the $\mathcal{V}_j$ can adaptively obtain the communication and computation resources for its offloading service. Usually, multiple vehicles generate different offloading tasks, and request different amounts of resources. To effectively share the limited $\mathfrak{M}_{\mathcal{R}}$ and $\mathfrak{N}_{\mathcal{M}}$ resources, we develop a cooperative game model $\mathbb{G}_{\mathcal{R}}^{\mathcal{M}}$. According to the resource type, the utility function for the $\mathcal{V}_j \in \mathbb{V}_{\mathcal{R}_i}$ $\left(\mathcal{U}_{\mathcal{V}_j}(\cdot)\right)$ consists of two sub-utility functions, i.e., $\mathcal{U}_{\mathcal{V}_j}^{\mathcal{R}_i}(\cdot)$ for $\mathfrak{M}_{\mathcal{R}_i}$ and $\mathcal{U}_{\mathcal{V}_j}^{\mathcal{M}_i}(\cdot)$ for $\mathfrak{N}_{\mathcal{M}_i}$. They are defined as follows:

$$\mathcal{U}_{\mathcal{V}_j}\left(\mathfrak{M}_{\mathcal{R}_i}, \mathfrak{N}_{\mathcal{M}_i}, \boldsymbol{a}, \mathfrak{X}_{\mathcal{V}_j}, \mathcal{Q}_{\mathcal{V}_j}^{\mathcal{R}}, \mathcal{S}_{\mathcal{V}_j}^{\mathcal{R}}\right) = \mathcal{U}_{\mathcal{V}_j}^{\mathcal{R}_i}(\cdot) \times \mathcal{U}_{\mathcal{V}_j}^{\mathcal{M}_i}(\cdot)$$
$$\text{s.t.,}\begin{cases} \mathcal{U}_{\mathcal{V}_j}^{\mathcal{R}_i}\left(\mathfrak{M}_{\mathcal{R}_i}, \boldsymbol{a}, \mathfrak{X}_{\mathcal{V}_j}, \mathcal{Q}_{\mathcal{V}_j}^{\mathcal{R}_i}, \mathcal{S}_{\mathcal{V}_j}^{\mathcal{R}_i}\right) = \\ \left(Y \times \log\left(\dfrac{\min\left(\mathcal{Q}_{\mathcal{V}_j}^{\mathcal{R}_i}, \mathcal{S}_{\mathcal{V}_j}^{\mathcal{R}_i}\right)}{\mathcal{Q}_{\mathcal{V}_j}^{\mathcal{R}_i}} + \eta\right) + \psi\right)^{\mathcal{W}_{\mathcal{V}_j}} \\ \mathcal{U}_{\mathcal{V}_j}^{\mathcal{M}_i}\left(\mathfrak{N}_{\mathcal{M}_i}, \boldsymbol{a}, \mathfrak{X}_{\mathcal{V}_j}, \mathcal{Q}_{\mathcal{V}_j}^{\mathcal{M}_i}, \mathcal{S}_{\mathcal{V}_j}^{\mathcal{M}_i}\right) = \\ \left(\dfrac{\varphi}{\xi + \exp\left(\mu \times \dfrac{\min\left(\mathcal{Q}_{\mathcal{V}_j}^{\mathcal{M}_i}, \mathcal{S}_{\mathcal{V}_j}^{\mathcal{M}_i}\right)}{\mathcal{Q}_{\mathcal{V}_j}^{\mathcal{M}_i}}\right)}\right)^{\mathcal{W}_{\mathcal{V}_j}} \\ \mathcal{W}_{\mathcal{V}_j} = \dfrac{\mathfrak{X}_{\mathcal{V}_j}}{\sum_{\mathcal{V}_g \in \mathbb{V}_{\mathcal{R}_i}} \mathfrak{X}_{\mathcal{V}_g}} \end{cases} \tag{8}$$

where Y, η, ψ are adjustment parameters for the $\mathcal{U}_{\mathcal{V}_j}^{\mathcal{R}_i}(\cdot)$, and φ, ξ and μ are adjustment parameters for the $\mathcal{U}_{\mathcal{V}_j}^{\mathcal{M}_i}(\cdot)$. $\mathcal{Q}_{\mathcal{V}_j}^{\mathcal{R}_i}(or\,\mathcal{Q}_{\mathcal{V}_j}^{\mathcal{M}_i})$ is the requested bandwidth (*or* computing power) amount from the $\mathcal{V}_j$, and $\mathcal{S}_{\mathcal{V}_j}^{\mathcal{R}_i}(or\,\mathcal{S}_{\mathcal{V}_j}^{\mathcal{M}_i})$ is the allocated bandwidth (*or* computing power) amount from the $\mathcal{R}_i(or\,\mathcal{M}_i)$. To adaptively decide the $\mathcal{S}_{\mathcal{V}_j}^{\mathcal{R}_i}$ and $\mathcal{S}_{\mathcal{V}_j}^{\mathcal{M}_i}$, we adopt the basic idea of *LHMS*. According to (1), the *LHMS*for $\left(\mathfrak{M}_{\mathcal{R}_i}, \mathfrak{N}_{\mathcal{M}_i}\right)$, i.e., *LHMS* $\left(\mathfrak{M}_{\mathcal{R}_i}, \mathfrak{N}_{\mathcal{M}_i}\right)$, is given by;

$$LHMS\left(\mathfrak{M}_{\mathcal{R}_i}, \mathfrak{N}_{\mathcal{M}_i}\right) = \arg\max_{\left(\widehat{\mathcal{S}_{\mathcal{V}}^{\mathcal{R}}}, \widehat{\mathcal{S}_{\mathcal{V}}^{\mathcal{M}}}\right)} \left\{\begin{array}{l} \prod_{\mathcal{V}_j \in \mathbb{V}_{\mathcal{R}_i}} \left(\mathcal{U}_{\mathcal{V}_j}^{\mathcal{R}_i}(\cdot) - d_{\mathcal{V}_j}^{\mathcal{R}}\right) \\ \times \prod_{\mathcal{V}_j \in \mathbb{V}_{\mathcal{R}_i}} \left(\mathcal{U}_{\mathcal{V}_j}^{\mathcal{M}_i}(\cdot) - d_{\mathcal{V}_j}^{\mathcal{M}}\right) \end{array}\right\}$$
$$s.t., \widehat{\mathcal{S}_{\mathcal{V}}^{\mathcal{R}}} = \left\langle \mathcal{V}_j \in \mathbb{V}_{\mathcal{R}_i} \middle| \ldots, \mathcal{S}_{\mathcal{V}_j}^{\mathcal{R}_i}, \ldots\right\rangle \text{ and}$$
$$\widehat{\mathcal{S}_{\mathcal{V}}^{\mathcal{M}}} = \left\langle \mathcal{V}_j \in \mathbb{V}_{\mathcal{R}_i} \middle| \ldots, \mathcal{S}_{\mathcal{V}_j}^{\mathcal{M}_i}, \ldots\right\rangle \tag{9}$$

where $d_{\mathcal{V}_j}^{\mathcal{R}}$ and $d_{\mathcal{V}_j}^{\mathcal{M}}$ are $\mathcal{V}_j$'s disagree points to share the $\mathfrak{M}_{\mathcal{R}_i}$ and $\mathfrak{N}_{\mathcal{M}_i}$, respectively.

D. MAIN STEPS OF OUR JOINT LEARNING AND BARGAINING ALGORITHMS

In the IoIV platform, crowdsensing provides a new mode to address the demands of data sensing, with the advantages of low cost and strong service flexibility. It has been used both in industry and multiple research studies, including traffic and road monitoring, urban mobility, smart cities, and ecological monitoring, and others. With the popularity of vehicular crowdsensing technique, vehicular edge computing can realize many applications. It can help the resource-limited vehicles to augment their task processing capabilities by offloading their tasks to the edge servers equipped on or near the RSUs. In this study, we investigate the intersection of crowdsensing and offloading service algorithms to maximize the IoIV system performance. Based on the *CMARL*, each vehicle learns his best action for the crowdsensing process. To participate in the crowdsensing, we give bargaining powers to vehicles according to their relative contribution ratios. According to the *LHMS*, individual vehicles fair-efficiently share the limited $\mathfrak{M}_{\mathcal{R}}$ and $\mathfrak{N}_{\mathcal{M}}$ resources based on their bargaining powers. In our jointly designed control scheme, the *CMARL*and *LHMS*are strongly related, and work together to strike a performance balance between conflicting requirements. With the stateless learning paradigm, our hybrid approach is especially appropriate to practically implement in dynamically changing IoIV system environments. The primary steps of our proposed scheme are described as follows.

Step 1: The parameter settings of our simulations are shown in the Table 1 to carry out the numerical experiments, and simulation scenario is determined in Section V.

Step 2: At a sequence of time steps, learning ($\mathbb{L}$) and game ($\mathbb{G}$) algorithms are operated interactively, and act cooperatively with each other.

Step 3: In the $\mathbb{L}_{\mathcal{V}_j}$, the $\mathcal{V}_j$ in the $\mathbb{V}_{\mathcal{R}_i}$ selects its action to participate in the crowdsensing process. Based on the experience of joint action, the action reward is obtained using (5), and the heuristic function is given by (6).

Step 4: During the *CMARL* process, the $Q-$value of each action is estimated based on the equations (5) and (6). By using this $Q-$value, the selection probability of each action is given by using (7).

Step 5: As a result of $\mathbb{L}_{\mathcal{V}_j}$, the $\mathcal{V}_j$'s contribution is defined as his selected sensing action where $\mathfrak{X}_{\mathcal{V}_j} = a^{\mathcal{V}_j}$. This information is used as the $\mathcal{V}_j$'s bargaining power $\left(\mathcal{W}_{\mathcal{V}_j}\right)$. Our proposed *CMARL* algorithm is implemented as a stateless learning mode, and it is operated in a distributed fashion.

Step 6: In the $\mathbb{G}_{\mathcal{R}_i}^{\mathcal{M}_i}$, the $\mathcal{V}_j$ in the $\mathbb{V}_{\mathcal{R}_i}$ share the $\mathfrak{M}_{\mathcal{R}_i}$ and $\mathfrak{N}_{\mathcal{M}_i}$ based on the *LHMS*. The $\mathcal{V}_j$'s sub-utility functions. i.e., $\mathcal{U}_{\mathcal{V}_j}^{\mathcal{R}_i}(\cdot)$ and $\mathcal{U}_{\mathcal{V}_j}^{\mathcal{M}_i}(\cdot)$, are defined according to (8).

Step 7: Based on the equation (9), the *LHMS*is obtained for vehicles in the $\mathbb{V}_{\mathcal{R}_i}$. The $\mathbb{G}_{\mathcal{R}_i}^{\mathcal{M}_i}$ is executed in a centralized manner.

Step 8: In our proposed joint control scheme, the $\mathbb{L}$ and $\mathbb{G}$ are sophisticatedly combined, and work together to reach a consensus with reciprocal advantages. Especially, the vehicle's crowdsensing contribution in the $\mathbb{L}$ is set up to its bargaining power in the $\mathbb{G}$, and the vehicle'sutility function in the $\mathbb{G}$ is set up to its heuristic function in the $\mathbb{L}$.

Step 9: Constantly, individual vehicles and RSUs are self-monitoring the current IoIV system environment, and sequentially interact with each other in the both distributed and centralized fashions. For the next iteration, it proceeds to Step 2.

V. PERFORMANCE EVALUATION

In this section, the representative numerical results are presented to identify the advantages of our joint approach while comparing our proposed scheme to the existing *IVRMC*, *CSORA* and *CAIMC* schemes [5], [7], [12]. In this study, we have used the simulation tool MATLAB to develop our simulation model. In particular, MATLAB's high-level syntax and dynamic types are ideal for model prototyping. Therefore, our simulator can adaptively build appropriate crowdsensing and offloading operation scenarios. The simulation results are recorded by averaging the results of 100 simulations in the scenario. Simulation parameters and their values are summarized in Table 1, and the simulation environment and system scenario are given as follows:

- Simulated the edge assisted IoIVsystem platform consists of five RSUsand fifty intelligent vehicles, i.e., $|\mathbb{R}| = 5$, and $|\mathbb{V}| = 50$.
- Five RSUsare deployed in the roadside area, and individual vehiclesare randomly distributed over there.
- Each vehicle $\mathcal{V}_{1\leq j\leq m}$ generates four different type offloading tasks $\left(\mathcal{T}_{\mathcal{V}_j}\right)$, which consists of $\mathcal{B}_{\mathcal{V}_j}$ and $\mathcal{C}_{\mathcal{V}_j}$.
- The arrival process of $\mathcal{T}_{\mathcal{V}_j}$ is the rate of Poisson process (ρ). The offered range is varied from 0 to 3.0.
- Individual vehicles participate in the crowdsensing through sensing actions $a_{1\leq k\leq l}^{\mathcal{V}}$ in the action set $\mathcal{L}_{\mathcal{V}}$ where $\mathcal{L}_{\mathcal{V}} = \{0.15, 0.3, 0.45, 0.6, 0.75, 0.9\}$.
- The total spectrum resource of each RSU ($\mathfrak{M}_{\mathcal{R}}$) is 2 Tbps, and the total computing capacity of each MEC server ($\mathfrak{N}_{\mathcal{M}}$) is 2 THz.
- Each RSUand its corresponding MEC server are connected without delay.
- The disagreement points for bargaining process, i.e., $d_{\mathcal{V}}^{\mathcal{R}}$ and $d_{\mathcal{V}}^{\mathcal{M}}$, are zeros.
- We assume the absence of physical obstacles in the RSU's coverage area.
- The communication and computing assignment process through the *LHMS*is specified in terms of basic allocation units (BAUs) where one BAU for $\mathfrak{M}_{\mathcal{R}}$ is 64 Mbpsand one BAU for $\mathfrak{N}_{\mathcal{M}}$ is 64 MHzin this study.
- The MECC based IoIVsystem performance measures obtained on the basis of 100 simulation runs are plotted as functions of the Poisson process (ρ).

To evaluate the proposed scheme, we compare its performance in terms of crowdsensing participation ratio, offloading throughput and vehicle payoff over offered task generation ratios. Table 1 shows the control parameters and system factors used in the simulation.

Fig.2 compares the crowdsensing participation ratio of vehicles in the IoIV platform. It is seen that the trend of participation ratios when implementing our scheme and the *IVRMC*, *CAIMC*and *CSORA* protocols. The results show that the proposed scheme can induce many vehicles to participate in the crowdsensing process than the other protocols. The reason is that we adopt a joint control paradigm to compensate the vehicle's sensing cost. In the viewpoint of vehicles, the

TABLE 1. System parameters used in the simulation experiments.

Parameter	Value	Description
n	5	total number of RSUs
m	50	total number of intelligent vehicles
$\mathfrak{M}_{\mathcal{R}}$	2 Tbps	wireless spectrum capacity of each RSU
$\mathfrak{N}_{\mathcal{M}}$	2 THz	computation capacity of each MEC server
l	6	the number of available sensing actions of each vehicle
α	0.2	a learning rate for the *CMARL*
τ	1	a temperature parameter for probability selection
λ	1	an adjustment parameter for the $B_{\mathcal{V}}(\cdot)$
β, ε	0.3 , 1.5	cost control factors for the $C_{\mathcal{V}}(\cdot)$
$\mathcal{X}_{\mathcal{V}}$	100 GHz	the $\mathcal{V}$'s local computation capacity
Υ, η, ψ	3 , 1, 1	adjustment parameters for the $\mathcal{U}_{\mathcal{V}}^{\mathcal{R}}(\cdot)$
ϕ, ξ, μ	2, 1, -2.5	adjustment parameters for the $\mathcal{U}_{\mathcal{V}}^{\mathcal{M}}(\cdot)$

Set	Values	Description
$\mathcal{L}_{\mathcal{V}}$	{0.15, 0.3, 0.45, 0.6, 0.75, 0.9}	the set of each device's actions for co-operative sensing
$d_{\mathcal{V}}^{\mathcal{R}}$	{… ,0, … }	disagreement point values for $\mathcal{V}$'s $\mathcal{R}$ bargaining
$d_{\mathcal{V}}^{\mathcal{M}}$	{… ,0, … }	disagreement point values for $\mathcal{V}$'s $\mathcal{M}$ bargaining

Task type	Requested $\mathcal{B}_{\mathcal{V}}$	Requested $\mathcal{C}_{\mathcal{V}}$	Service duration /*t*
I	256 Mbps	320 MHz	45 time-periods
II	640 Mbps	192 MHz	50 time-periods
III	192 Mbps	640 MHz	25 time-periods
IV	320 Mbps	256 MHz	15 time-periods

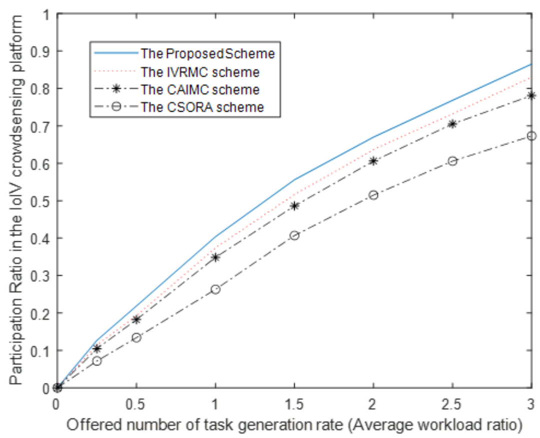

FIGURE 2. The crowdsensing participation ratio.

more the quantity of offloading tasks, the higher the longing for incentives to process offloading services. In the proposed scheme, vehicles can get the feedback from the resource allocation process; this information is directly considered to select the crowdsensing action. Based on the dynamic feedback learning process through *CMARL*, vehicles can capture how to select their actions to achieve the better benefit. Simulation result confirms that our proposed method gains a significant better performance in the crowdsensing process in the IoIV platform.

We depict the normalized vehicle payoff in Fig.3; it is obtained from the offloaded task services. From Fig.3, it can be observed that when the task generation rate increases, the vehicle payoffs of all schemes also increase. It is intuitively correct. However, we can maintain a performance superiority than the *IVRMC*, *CAIMC*and *CSORA* schemes. In our proposed scheme, the limited IoIV resources are effectively bargained in a step-by-step dynamic game iteration. Based on the *LHMS*, we can get the temporarily converged and optimal solution at the end of each time epoch. This approach can ensure the excellent control flexibility and responsiveness to current IoIV system situations. Due to this reason, we can get a higher vehicle payoff.

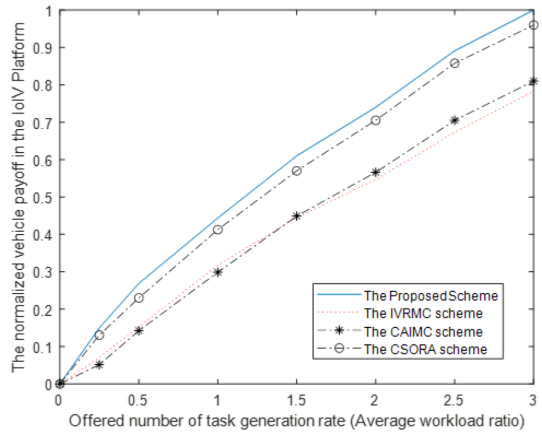

FIGURE 3. Normalized payoff of vehicles.

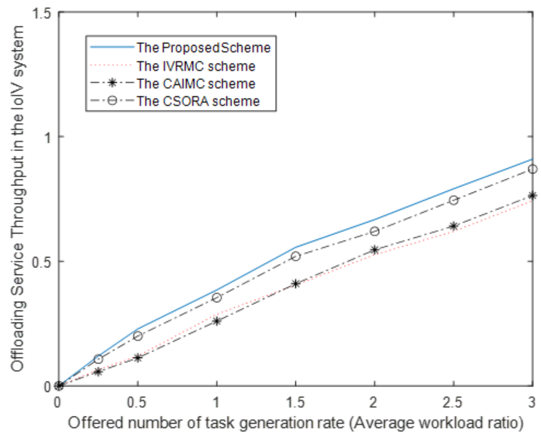

FIGURE 4. System throughput in the IoIV system platform.

Fig.4 compares the system throughput of IoIV system. In this study, the throughput refers to the complete rate of offloading task services over the communication and computation resources in the IoIV platform. As a main performance criterion, the throughput is essentially synonymous to system resource utilization. Simulation result is seen that our proposed scheme outperforms the existing schemes. That is because each vehicleon our scheme can share the limited resources in a coordinated manner. Based on the desirable bargaining features, we can achieve a fair-efficient resource allocation solution through the *LHMS*.Under dynamic changing IoIV environments, the idea of the *LHMS*can guarantee a mutually acceptable agreement based on each vehicle's sensing contribution. Compared with the other protocols, our approach is quite adaptable to maximize the system throughput while handling different offload requirements.

Fig.2 to Fig.4, it is observed that our proposed scheme not only provides the benefit for the IoIV system operator through the crowdsensing process, but also improves the payoff of individual vehicles by addressing their offloading task services. To capture dynamic interactions among RSUs, MEC servers and vehicles, the combination of *CMARL* and *LHMS* significantly improves the effectiveness of the IoIV system. The numerical simulation results confirm that the proposed learning-bargaining hybrid paradigm can get a desirable solution in the MECC based IoIV platform than that of *IVRMC*, *CAIMC* and *CSORA* schemes.

VI. SUMMARY AND CONCLUSION

This study investigates the joint crowdsensing and offloading problem in the edge assisted IoIV system. By integrating the cooperative multi-agent learning and bargaining game theory, our major goal is to strike an appropriate balance between vehicles' sensing contribution and offloading benefit. Specifically, we first design a multi-agent learning algorithm to participate in the crowdsensing process. In this algorithm, each individual vehicle learns the best sensing action to maximize his payoff. Based on the *CMARL*, this learning process is operated in a distributed manner. In addition, our stateless learning method is suitable from the viewpoint of practical system management.Second, we develop a bargaining game model to share the limited $\mathfrak{M}_{\mathcal{R}}$ and $\mathfrak{N}_{\mathcal{M}}$ resources.To support vehicles' offload services, the idea of *LHMS* is adopted to distribute the $\mathfrak{M}_{\mathcal{R}}$ and $\mathfrak{N}_{\mathcal{M}}$ in a centralized manner. In this resource sharing process, vehicles' sensing contributions are used as their bargaining powers. Therefore, our proposed crowdsensing and offloading control algorithms are tightly coupled with each other, and work together in a coordinated manner. Under diverse MECC based IoIV platform environment changes, the main novelty of our joint approach is its adaptability, which guarantees the responsiveness to current system conditions. Finally, we conduct extensive simulations to validate the performance of our proposed scheme. Through numerical analysis, we demonstrate the effectiveness of our joint control approach in terms of sensing ratio, throughput and vehicle payoff compared to the existing *IVRMC*, *CAIMC* and *CSORA* schemes.

As a future work, we will investigate the impact of communications between connected vehicles and consider adding privacy protection to the incentive method. To take the heterogeneity of different vehicles, a finer-grained task assignment and resource scheduling algorithm will be studied. Moreover, we plan to investigate the system's reliability with the consideration of disturbances and stochastic behaviors. In addition, considering three dimensional coverage area covered by drones in the sensing recruitment process is an interesting research topic.

AVAILABILITY OF DATA AND MATERIAL

Please contact the corresponding author at swkim01@sogang.ac.kr.

COMPETING INTERESTS

The author, Sungwook Kim, declares that there are no competing interests regarding the publication of this article.

REFERENCES

[1] K. Guan, D. He, B. Ai, D. W. Matolak, Q. Wang, Z. Zhong, and T. Kürner, "5-GHz obstructed vehicle-to-vehicle channel characterization for internet of intelligent vehicles," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 100–110, Feb. 2019.

[2] L. Zhu, D. He, B. Ai, K. Guan, S. Dang, J. Kim, H. Chung, and Z. Zhong, "A ray tracing and joint spectrum based clustering and tracking algorithm for internet of intelligent vehicles," *J. Commun. Inf. Netw.*, vol. 5, no. 3, pp. 265–281, Sep. 2020.

[3] C. Chen, H. Li, H. Li, R. Fu, Y. Liu, and S. Wan, "Efficiency and fairness oriented dynamic task offloading in Internet of Vehicles," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 3, pp. 1481–1493, Sep. 2022.

[4] P. Lang, D. Tian, X. Duan, J. Zhou, Z. Sheng, and V. C. M. Leung, "Cooperative computation offloading in blockchain-based vehicular edge computing networks," *IEEE Trans. Intell. Vehicles*, vol. 7, no. 3, pp. 783–798, Sep. 2022.

[5] L. Liu, X. Wen, L. Wang, Z. Lu, W. Jing, and Y. Chen, "Incentive-aware recruitment of intelligent vehicles for edge-assisted mobile crowdsensing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12085–12097, Oct. 2020.

[6] G. Gao, M. Xiao, J. Wu, L. Huang, and C. Hu, "Truthful incentive mechanism for nondeterministic crowdsensing with vehicles," *IEEE Trans. Mobile Comput.*, vol. 17, no. 12, pp. 2982–2997, Dec. 2018.

[7] Y. Li, B. Yang, H. Wu, Q. Han, C. Chen, and X. Guan, "Joint offloading decision and resource allocation for vehicular fog-edge computing networks: A contract-Stackelberg approach," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 15969–15982, Sep. 2022.

[8] S. Kim, *Game Theory Applications in Network Design*. Hershey, PA, USA: IGI Global, 2014.

[9] X. Wang, C. Zhao, T. Huang, P. Chakrabarti, and J. Kurths, "Cooperative learning of multi-agent systems via reinforcement learning," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 9, pp. 13–23, 2023.

[10] S. Lozano, M. A. Hinojosa, and A. M. Mármol, "Extending the bargaining approach to DEA target setting," *Omega*, vol. 85, pp. 94–102, Jun. 2019.

[11] H. Cui and Z. Zhang, "A cooperative multi-agent reinforcement learning method based on coordination degree," *IEEE Access*, vol. 9, pp. 123805–123814, 2021.

[12] H. Yu, Y. Yang, H. Zhang, R. Liu, and Y. Ren, "Reputation-based reverse combination auction incentive method to encourage vehicles to participate in the VCS system," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 3, pp. 2469–2481, Jul. 2021.

[13] N. Morozs, T. Clarke, and D. Grace, "Distributed heuristically accelerated Q-learning for robust cognitive spectrum management in LTE cellular systems," *IEEE Trans. Mobile Comput.*, vol. 15, no. 4, pp. 817–825, Apr. 2016.

[14] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems," in *Proc. AAAI/IAAI*, 1998, pp. 746–752.

[15] K. Zhang, Y. Mao, S. Leng, A. Vinel, and Y. Zhang, "Delay constrained offloading for mobile edge computing in cloud-enabled vehicular networks," in *Proc. 8th Int. Workshop Resilient Netw. Design Modeling (RNDM)*, Sep. 2016, pp. 1–7.

SUNGWOOK KIM received the B.S. and M.S. degrees in computer science from Sogang University, Seoul, Republic of Korea, in 1993 and 1995, respectively, and the Ph.D. degree in computer science from Syracuse University, Syracuse, NY, USA, in 2003, supervised by Prof. Pramod K. Varshney. He was a Faculty Member with the Department of Computer Science, Chung-Ang University, Seoul. In 2006, he returned to Sogang University, where he is currently a Professor with the Department of Computer Science and Engineering and the Research Director of the Network Research Laboratory. His research interests include resource management, online algorithms, adaptive quality-of-service control, and game theory for network design.