An Enhanced Dual-Currency VCG Auction Mechanism for Resource Allocation in IoV: A Value of Information Perspective

Wei Wang , *Graduate Student Member, IEEE*, Nan Cheng , *Senior Member, IEEE*, Conghao Zhou , *Member, IEEE*, Haixia Peng , *Member, IEEE*, Haibo Zhou , *Senior Member, IEEE*, Zhou Su , *Senior Member, IEEE*, and Xuemin Shen , *Fellow, IEEE*

Abstract—The Internet of Vehicles (IoV) is undergoing a transformative evolution, enabled by advancements in future 6G network technologies, to support intelligent, highly reliable, and low-latency vehicular services. However, the enhanced capabilities of IoV have heightened the demands for efficient network resource allocation while simultaneously giving rise to diverse vehicular service requirements. For network service providers (NSPs), meeting the customized resource-slicing requirements of vehicle service providers (VSPs) while maximizing social welfare has become a significant challenge. This paper proposes an innovative solution by integrating a mean-field multi-agent reinforcement learning (MFMARL) framework with an enhanced Vickrey-Clarke-Groves (VCG) auction mechanism to address the problem of social welfare maximization under the condition of unknown VSP utility functions. The core of this solution is introducing the "value of information" as a novel monetary metric to estimate the expected benefits of VSPs, thereby ensuring the effective execution of the VCG auction mechanism. MFMARL is employed to optimize resource allocation for social welfare maximization while adapting to the intelligent and dynamic requirements of IoV. The proposed enhanced VCG auction mechanism not only protects the privacy of VSPs but also reduces the likelihood of collusion among VSPs, and it is theoretically proven to be dominant-strategy incentive compatible (DSIC). The simulation results demonstrate that, compared to the VCG mechanism implemented using quantization methods, the proposed mechanism exhibits significant advantages in convergence speed, social welfare maximization, and resistance to collusion, providing new insights into resource allocation in intelligent 6G networks.

Index Terms—IoV, network function virtualization, value of information, VCG auction.

Wei Wang, Nan Cheng, and Conghao Zhou are with the State Key Laboratory of ISN and School of Telecommunications Engineering, Xidian University, Xi'an 710126, China (e-mail: weiwang_2042@stu.xidian.edu.cn; dr.nan.cheng@ieee.org; conghao.zhou@ieee.org).

Haixia Peng is with the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: haixia.peng@xjtu.edu.cn).

Haibo Zhou is with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: haibozhou@nju.edu.cn).

Zhou Su is with the School of Cyber Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: zhousu@ieee.org).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/TMC.2025.3594358

I. INTRODUCTION

WITH the ongoing research and anticipated advancements in 6G network technologies, the Internet of Vehicles (IoV), as one of its core application scenarios, is expected to evolve toward intelligence, high reliability, and low latency [1]. 6G networks not only support transmissions with low latency and high data rate but also enable intelligent management of network resources through artificial intelligence technologies [2], [3]. In this context, vehicle service providers (VSPs) can offer personalized and high-quality services to users associated with vehicles. In addition to currently available services such as navigation, traffic information prompts, and multimedia services, future services like intelligent driving, driving behavior analysis, and vehicle condition monitoring will significantly enhance traffic efficiency and the driving experience [4]. VSPs rely on communication, computation, and cache resources to provide these services. Network service providers (NSPs), with their extensively distributed physical infrastructure, are capable of delivering reliable resource support to numerous VSPs. Nevertheless, the fixed device deployment of NSPs leads to an inflexible network architecture that makes it difficult to adapt to dynamically changing service demands. Network Function Virtualization (NFV) technology can create logically isolated slices dedicated to VSPs based on their specific requirements, significantly enhancing resource utilization while increasing the flexibility of network resource allocation [5]. This capability allows the network to dynamically adapt to diverse service types and rapidly respond to evolving vehicular user requirements.

Despite the advantages of NFV, establishing adaptive resource slices for VSPs remains a challenging issue. One of the fundamental dilemmas pertains to the competitive dynamics for resources among various VSPs, necessitating that NSPs devise equitable and efficacious mechanisms for resource allocation to foster the active participation of VSPs in resource negotiation. Furthermore, given that NSPs often embody public benefit attributes, their optimization goals transcend mere equity in resource distribution, incorporating the maximization of social welfare (the sum of all VSPs' utility). However, the

pursuit of societal welfare enhancement may clash with the individual profit-maximization goals of VSPs, thereby adding layers of complexity to the allocation quandary. Such problems are typically modeled within the frameworks of game theory, and auction-based mechanism design emerges as a potential resolution, celebrated for its merits in incentive compatibility and operational efficiency. When NSPs understand all VSPs' utility functions, the social welfare maximization problem can be addressed by a classic auction mechanism known as the Vickrey-Clarke-Groves (VCG) auction [6]. VCG auctions are widely used to achieve efficient allocation in multi-item or multi-attribute auctions, incentivizing bidders to reveal their true valuations for the auctioned items truthfully. In the VCG auction, bidders submit their maximum willingness to pay for each item, and the mechanism selects an allocation that maximizes social welfare. After determining the allocation, the actual payment prices are calculated based on each bidder's impact on the welfare of other bidders.

However, one limitation of existing works on VCG auctions is that calculating the actual payment prices typically requires the utility functions of the bidders [7], [8]. Specifically, The utility functions of VSPs in IoV often involve private information that can affect the safety of vehicle operation [9], [10], making it difficult for VSPs to report their utility functions to NSPs. In addition, utility functions may not have explicit expressions for some complex services. Even if all the utility functions of VSPs possess desirable properties (e.g., convexity), the computational complexity can be significant in large-scale IoV [8]. The difficulty increases when utility functions cannot be explicitly expressed. Another concern arises from the possibility of collusion among VSPs to maximize their benefits, which makes it challenging for auctioneers to achieve social welfare maximization goals [11], [12].

To address the issue of utility functions that cannot be explicitly expressed, recent studies have introduced the concept of "Value of Information (VoI)," which reflects the impact of specific information on the performance of specific services [13], [14], [15]. VoI can serve as a transactional medium between NSPs and VSPs, exhibiting currency-like properties for resource allocation purposes. Compared to directly estimating fixed utility functions, estimating VoI based on current information characteristics can reflect the real-time service performance of VSPs, thereby facilitating the goal of maximizing the performance of all VSP services, i.e., social welfare maximization. VoI can support measuring service value without relying on explicit utility functions, facilitating the expression of utility functions.

To address the privacy and collusion issues faced by the auction mechanism, a promising approach is to integrate multi-agent reinforcement learning (MARL) with the auction mechanism. MARL can directly learn optimal strategies through environmental interactions without requiring mathematical modeling of the environment [16], [17]. The MARL-based approach is conducive to VSPs' privacy needs, as NSPs can indirectly understand the relationship between VSPs' value assessments and resource needs by training neural networks rather than directly acquiring utility functions or data. The gaming process within the MARL framework introduces randomness[1] into the auction mechanism, increasing the difficulty of bidder collusion and reducing its impact [18], [19], [20].

This paper investigates the problem of establishing VSP resource slices in IoV and proposes an improved VCG auction mechanism based on mean-field multi-agent reinforcement learning (MFMARL). Unlike the usual VCG auction mechanism, a third-party institution, known as the "Bank", is introduced to issue VoI Currency (VC), impose taxes, and reclaim VC. The mechanism can be divided into three phases: currency issuance, sealed bidding for intention value, and transaction. During the currency issuance phase, the Bank determines the amount of VC to issue based on the total resources of the NSP and historical social welfare, establishing a reserve pool. In the sealed bidding phase, VSPs submit their resource bidding intentions in VC to the NSP and borrow VC from the Bank based on their expectations. In the transaction phase, the NSP determines the resource allocation scheme, after which the VSPs pay VC to acquire resources and begin providing services. Finally, the Bank calculates taxes and reclaims residual VC based on the information disclosed by the NSP and the financial statements of the VSPs. Throughout the process, the NSP and the Bank are training their strategies within the MFMARL framework to achieve near-optimal social welfare allocation while protecting VSP privacy and reducing the impact of collusion. The main contributions of this paper are as follows:

- An improved VCG auction mechanism based on MF-MARL was designed, achieving near socially optimal resource allocation without requiring knowledge of VSP utility functions. Additionally, the mechanism can be scaled to changes in NSP available resources and the number of VSPs without the need for redeployment.
- Theoretical analyses and simulation studies assert that the proposed auction mechanism introduced achieves Dominant Strategy Incentive Compatibility (DSIC), indicating that for all bidders(VSP), strategies that optimize the social welfare equally maximize their profits. This secures the initiative of the VSPs to participate in the auction process.
- Theoretical analyses and simulation studies confirm the effectiveness of the proposed auction mechanism in addressing the collusion of bidders, reducing the negative impact of such collusion on overall social welfare, and showing enhanced robustness.

The rest structure of this paper is organized as follows: Section II reviews and summarizes the previous work on VCG auctions and MRAL-enhanced auctions. In Section III, we provide a comprehensive description of the system model. Section IV introduced the proposed auction mechanism where information value serves as currency and analyzes its theoretical performance. Section V elucidates the implementation method of the auction

---

[1]Randomness in this context is introduced by the exploration process of MARL, meaning that the allocation schemes obtained in the same state exhibit a certain level of uncertainty. This characteristic makes it more challenging for conspirators to achieve a stable allocation strategy (reaching an equilibrium point in the game).

mechanism based on MFMARL. Section VI demonstrates the results of simulations, and Section VII concludes the paper.

II. RELATED WORK

NFV brings numerous benefits to IoV, including isolation, automation, flexibility, and scalability. Many researchers have attempted to work in this research field from economics and game theory perspectives. In [21], a dynamic spectrum leasing mechanism for power control strategies was proposed, while a spectrum-sharing power allocation scheme based on a Stackelberg game for Femtocell networks, regulated by pricing, was introduced in [22]. Moreover, utility maximization in spectrum-sharing networks using game-theoretic power control strategies was explored in [23] and [24]. A scenario in which mobile users compete for wireless resources, aiming to optimize profits for both physical substrates and virtual wireless networks, was studied in [25]. Although these game-theoretic approaches drive resource competition towards equilibria such as Nash equilibrium, they often fail to achieve social welfare maximization. Consequently, researchers have further refined auction mechanisms.

A. VCG-Based Auction Mechanism

The VCG auction, a non-cooperative game strategy, is widely recognized for its ability to maximize social welfare. Specifically, consider n bidders and m homogeneous items, where each bidder i reports its valuation function $v_i(\cdot)$. The mechanism first selects an allocation scheme $\boldsymbol{a}^* = (a_1^*, \ldots, a_m^*)$ that maximizes social welfare as

$$\max_{\boldsymbol{a}} \sum_{i=1}^{n} v_i(a_i),$$

$$\text{s. t.} \sum_{i=1}^{n} a_i \leq m. \tag{1}$$

For each bidder i the payment z_i is calculated as

$$z_i = \max_{\boldsymbol{a}} \sum_{j \neq i} v_j(a_j) - \sum_{j \neq i} v_j(a_j^*), \tag{2}$$

which represents the difference in welfare caused to other bidders due to bidder i's participation. The VCG auction is an optimal auction mechanism that satisfies DSIC and maximizes social welfare. Its core idea is to have each winner pay the social opportunity cost caused by their actions, thereby achieving both efficient resource allocation and the revelation of true preferences. Its application to network slice resource allocation has been extensively studied. For example, [26] proposed a model based on VCG auctions, where NSPs act as auctioneers to allocate discrete physical resource blocks. Similarly, [27] integrated a VCG auction model to address power allocation in LTE air interface virtualization. However, these approaches often face challenges due to limitations in information exchange mechanisms and the non-convexity of bidders' utility functions, hindering the practical implementation of VCG auctions.

To address these issues, some scholars have focused on allocating infinitely divisible resources under limited information exchange mechanisms. For instance, [28] and [29] proposed Kelly's proportional fairness mechanism, where agents submit one-dimensional bids, and allocations are determined based on the ratio of bids to a congestion price. However, this approach may incur efficiency losses of up to 25% of the maximum social welfare, as noted by [30]. To mitigate this drawback, several works have combined Kelly's concept with the VCG auction, resulting in the "VCG-Kelly" mechanism [31], [32], [33]. The "VCG-Kelly" mechanism requires agents to communicate only a single scalar value, enabling the allocation of infinitely divisible resources under limited information exchange. Furthermore, [8] proposed a dominant allocation strategy for divisible network resources under constrained information by partitioning resources into indivisible units and applying the VCG mechanism to each partition, achieving near-socially optimal allocations.

B. MARL-Based Auction Mechanism

In recent years, MARL has gained attention for its potential in auction modeling. MARL provides a model-free approach to decision-making and control, enabling agents to learn optimal strategies through interaction with the environment without prior knowledge of the system. For example, [34] used MARL to reduce operating costs in double auctions within energy markets, while [35] integrated the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm with electricity market auction mechanisms to guide participants toward better bidding strategies. However, many MARL algorithms suffer from poor scalability, limiting their applicability to vehicular network resource slicing problems with rapidly changing environments. Although independent reinforcement learning approaches can mitigate this issue, they often exacerbate non-stationarity, hindering stable strategy development.

To address these challenges, extended MADDPG frameworks have been proposed. For instance, [36] introduced an order book structure that allows agents to abstract and observe the behaviors of others. However, implementing such frameworks becomes increasingly difficult as the number of agents grows. The emergence of MFMARL offers a promising solution, providing robust scalability and reducing computational requirements [37]. In IoV, MFMARL has been successfully applied to address joint spectrum and power allocation problems, significantly reducing computational complexity [38].

Despite these advancements, the adoption of the aforementioned two types of auction mechanisms still encounters certain limitations. Auction mechanisms based on the VCG principle often require substantial communication to comprehend the utility functions of participants, posing significant demands on the communication capabilities of IoV networks and failing to meet the privacy requirements of VSPs; concurrently, auction mechanisms based on MARL typically face limitations in scalability. This paper builds upon and extends these ideas, aiming to provide innovative solutions to the resource allocation challenges in vehicular network slicing.

III. SYSTEM MODEL AND PROBLEM FORMULATION

We consider an IoV composed of one NSP and N VSPs as shown in Fig. 1. Base stations (BS), mobile edge computing
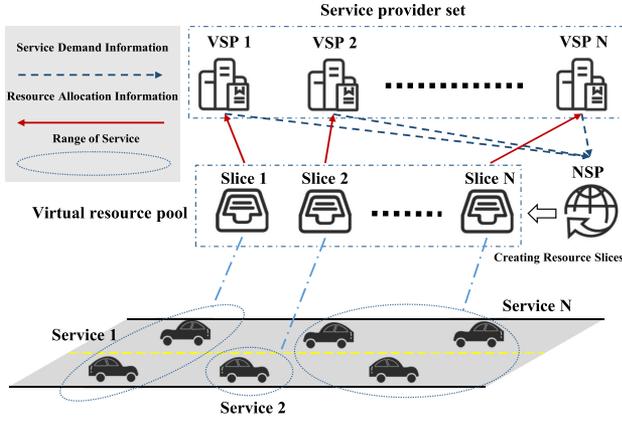
Fig. 1. Schematic illustration of how IoV resources are sliced and allocated.

(MEC) servers, and roadside units (RSU), are owned by the NSP. The communication, computation, and cache resources are virtualized into a resource pool of the NSP. These virtualized network resources can be partitioned into slices by the NSP and allocated to VSPs for a certain period. Each period, denoted as time T_s, is characterized by VSPs submitting their bids to the NSP, which in turn allocates the new resource slices to the VSPs. The VSPs aim to maximize profits, whereas the NSP allocates resources to maximize social welfare.

A. VSP Profits Model

We denote the VSP set as $\mathcal{V}$, and define VSP i as v_i. The slice set is denoted by $\mathcal{S}$. The slice vector in t-th round is denoted by $\boldsymbol{s}_t$, and $s_{i,t}$ denotes the resource slice allocated by the NSP to v_i in t-th round, constituted by bandwidth resources, computational resources, and cache resources. The terms $\boldsymbol{X}$, $\boldsymbol{x}_t$, $x_{i,t}$ signify the bid set, the bid vector in t-th round and the bid of the VSP i in t-th round. For VSP i, we employ $P_{i,t}$ to represent the profit of VSP i during the t-th round, with $P_{i,t}$ being calculated as

$$P_{i,t}(x_{i,t}) = g_i(s_{i,t}) - z_i(x_{i,t}), \qquad (3)$$

where $g_i(\cdot)$ serves as the gross profit function. Note is that the gross profit function encompasses the private information of the VSP and may not always offer a manifest expression. The term $z_i(\cdot)$ represents the cost function of deploying the resource slice based on the bid.

B. Social Welfare Model

The maximization of social welfare symbolizes the maximization of the aggregate utility of all VSPs. Momentous to note is the discrepancy between utility and profits,[2] implying that an accurate utility function is generally unobtainable directly through the gross profit function (even if the profit function is explicative). We let $U_{i,t}$ to denote the utility of VSP i in the

t-th round and employ the function $u_i(\cdot)$ to express the bond between VSP i's utility and the resource slice. For the NSP, we define the total social welfare $\mathcal{SW}_t$ as

$$\mathcal{SW}_t = \sum_{i=1}^{N} U_{i,t} = \sum_{i=1}^{N} u_i(s_{i,t}). \qquad (4)$$

Analogous to the gross profit function, the utility function may likewise be implicit; a more dire circumstance is that neither the NSP nor the VSPs might understand the specific form of the utility function.[3]

Normalization of the resources within the resource pool is performed for convenient analysis, ensuring that $\sum_{i=1}^{N} s_{i,t} \leq \mathbb{1}$. Although the precise form of the utility function is arduous to obtain via the gross profit function, an estimate can still be fabricated through the observation of gross profits. Hence, we establish the mapping from the utility function to the gross profit function as

$$G_{i,t} = \alpha_i U_{i,t} + w_i, \qquad (5)$$

where α_i is the rate coefficient of gross profit $G_{i,t}$ to utility $U_{i,t}$, possessing varied values in different VSPs, and w_i is a random variable obeying a specific distribution representing the uncertain association between profit and utility.[4] Consequently, we can rewrite the profit function as

$$P_{i,t}(x_{i,t}) = \alpha_i u_i(s_{i,t}) + w_i - z_i(x_{i,t}). \qquad (6)$$

To impel the achievement of maximal social welfare, it is requisite that a consensus is conceivably reached between VSP and NSP. In simpler terms, the socially optimal allocation result must remain dominant[5] for both VSPs and the NSP. This consensus is represented mathematically as

$$\boldsymbol{s}_t^* = \mathcal{A}(x_{1,t}^*, \ldots, x_{i,t}^*, \ldots). \qquad (7)$$

Due to both gross margin functions and utility functions being unknown to the NSP, we employ $\mathcal{A}(\cdot)$ as the mathematical characterization of the NSP's resource allocation process based on VSPs' bid. $x_{i,t}^*$ denotes the optimal bid under the circumstance of VSP maximizing its profits as

$$x_{i,t}^* = \arg\max P_{i,t}(x_{i,t}), \qquad (8)$$

while $\boldsymbol{s}_t^*$ signifies the optimal resource slice allocation under the condition of maximal social welfare in t-th round as

$$\boldsymbol{s}_t^* = \arg\max_{\boldsymbol{s}_t} \sum_{i=1}^{N} u_i(s_{i,t}). \qquad (9)$$

---

[2]The discussion of the utility notion bears unique significance in IoV as several services within these networks pertain to vehicular driving safety, such as autonomous driving, assisted driving, navigation, etc. These services may not generate substantial profits but proffer immense utility (i.e., they may prevent severe traffic accidents).

[3]This owes to the more subjective nature of utility compared to profits, usually involving insuperably abstract mapping measures such as user satisfaction.

[4]The relationship between utility and profit is based on the core logic of equity incentives in agency theory: the social welfare generated by the VSP is analogous to "equity," while the economic returns (dividends) derived from it correspond to the utility-related component. Meanwhile, the profit component unrelated to equity is abstracted as a specific probability distribution, aiming to retain the random impact of uncontrollable factors on profit in the real world.

[5]In game theory, a dominant strategy refers to a course of action that yields the optimal outcome for a participant, irrespective of the strategies employed by other players. In other words, selecting a dominant strategy is invariably the most advantageous move, regardless of the opponents' actions.

It should be noted that the allocation functions $\mathcal{A}(\cdot)$ satisfying the (7) may form a non-singleton set. Accordingly, we define $\mathbb{A}$ as the collection of all feasible allocation functions.

C. Problem Formulation

We formalize the problem of resource slicing in IoV as *Problem 1*.

$$\mathcal{P}1 : \max_{\mathcal{A}\in\mathbb{A},\mathcal{S}} \sum_{t=1}^{T}\sum_{i=1}^{N} u_i(s_{i,t}), \tag{10}$$

$$\text{s.t.} \sum_{i=1}^{N} s_{i,t} \leq \mathbb{1}, \tag{10a}$$

$$\boldsymbol{s}_t = \mathcal{A}(x_{1,t}^*, \ldots, x_{i,t}^*, \ldots), \tag{10b}$$

$$x_{i,t}^* = \arg\max P_{i,t}(x_{i,t}). \tag{10c}$$

The optimization problem requires the solution of the auction mechanism $\mathcal{A}$ and the optimal allocation $\mathcal{S}$. Constraint (10a) ensures that the slices do not exceed the available resources of the NSP, while (10b) and (10c) represent the strategic consensus that can be reached between the NSP and the VSP. The intricacy of resolving problem $\mathcal{P}1$ stems from the NSP inability to directly obtain formulations for utility or profit functions. This limitation affects the development of effective slicing allocation strategies. Fortunately, the NSP can deduce the relationship between resource slicing, social welfare, and VSP profit through bid information from the VSP and statistical information on overall utility over a while. This offers us a viable avenue for addressing the problem above.

IV. AN ENHANCED DUAL-CURRENCY VCG AUCTION MECHANISM BASED ON VoI

We propose a dual-currency auction mechanism based on the VoI, an extension of the traditional VCG auction, abbreviated as DVCG. In this section, we first elucidate the implementation of this mechanism and subsequently demonstrate that it shares a similar DSIC with VCG auctions.

A. Auction Mechanism

Firstly, we explain the concept of the VoI. As previously discussed, one of the challenges of problem $\mathcal{P}1$ lies in the uncertainty of the utility function, which precludes an accurate mapping to profit. The VoI metric, as described in [13], is defined based on the impact of specific information on the performance of a particular task, which aptly reflects the statistical information of task performance (i.e., utility in this context). We extend this concept in this paper, where VoI represents the expected utility of VSP under the social welfare maximization strategy π. The specific formulation is as

$$\text{VoI}_{i,t} = \mathbf{E}\left[U_{i,t}|\pi \simeq \max\sum_{i=1}^{N} u_i(s_{i,t})\right]. \tag{11}$$

By understanding the VoI distribution of the VSP, the NSP can indirectly comprehend its utility function. To this end, we design a dual-currency auction mechanism based on VoI, as illustrated in Fig. 2. Unlike traditional auction mechanisms, we introduce a third-party entity called the "bank". This bank can be stationed in neutral regulatory department devices. Irrespective of its deployment location, the bank must ensure that the information it collects can be directly reviewed by the corresponding VSPs, thereby mitigating the risk of fraudulent information. It is particularly noteworthy that each VSP's audit authority is subject to well-defined scope constraints: limited exclusively to verifying data directly related to its operations, specifically including its own gross profits, utility, and complete resource allocation schemes. In tandem, the bank collects available information $\mathcal{B}_t = (U_t, G_t)$ from VSPs' gross profits and feedback on user satisfaction, which we term "financial reports". It is important to note that this information exhibits a delay, which means that in the t-th round of the auction, only the information from the previous $t-1$ rounds can be observed. We presume the bank's historical bills, constituting the prior information, are represented as $\mathcal{BI}$. The specifics of the auction mechanism can thus be articulated as follows:

- During the currency issuance phase, the bank estimates the mean total VoI of all VSPs based on historical bill information $\mathcal{BI}$. It forms a pool $\mathcal{F}_\mathcal{P}(t)$ of VoI currency (VC) as

$$\mathcal{F}_\mathcal{P}(t) = \sum_{i=1}^{N} \mathbf{E}\left[U_{i,t}|\pi \simeq \max\sum_{i=1}^{N} u_i(s_{i,t})\right]$$
$$= \sum_{i=1}^{N} f_i(\mathcal{BI}; \hat{\phi}_i), \tag{12}$$

where the function $f_i(\cdot)$ signifies the bank's projection of the utility achievable by the VSP i under the stratagem of optimizing its profit, with $\hat{\phi}_i$ representing the inferred of the function parameters.

- In the sealed bid for intention value phase, each VSP sends proposals $x_{i,t}$ to the NSP in units of VC, reflective of their service profit estimates, and applies for a loan of equal VC amount from the bank. The bank allows loans $\mathcal{K}_t$ to VSPs based on the $\mathcal{F}_\mathcal{P}(t)$. $\mathcal{K}_t$ can be calculated as

$$\mathcal{K}_t = (\mathcal{K}_{1,t}, \ldots, \mathcal{K}_{N,t}) = \mathcal{F}_\mathcal{P}(t) * \frac{x_{i,t}}{\sum_{j=1}^{N} \boldsymbol{x}_{j,t}}. \tag{13}$$

- In the phase of transaction, the NSP allocates slices and estimates the VSP's utility based on the submitted proposals, parameter $\hat{\theta}_i$, and Δ. The method for achieving the optimal slicing policy and optimal sum utility $\mathcal{SW}_t^*$ in the t-th round is

$$(\mathcal{SW}_t^*, \boldsymbol{s}_t^*) = \arg\max_{\boldsymbol{s}_t} \sum_{i=1}^{N} u_i(s_{i,t}; \hat{\theta}_i), \tag{14}$$

$$\tilde{\boldsymbol{s}}_t = \text{norm}\left(\boldsymbol{s}_t^* + \mathcal{N}(0, \Delta^2 \boldsymbol{I})\right), \tag{15}$$

$$\Delta = \frac{\max\left(\sum_{i=1}^{N} x_{i,t} - \mathcal{SW}_t^*, 0\right)}{N}. \tag{16}$$
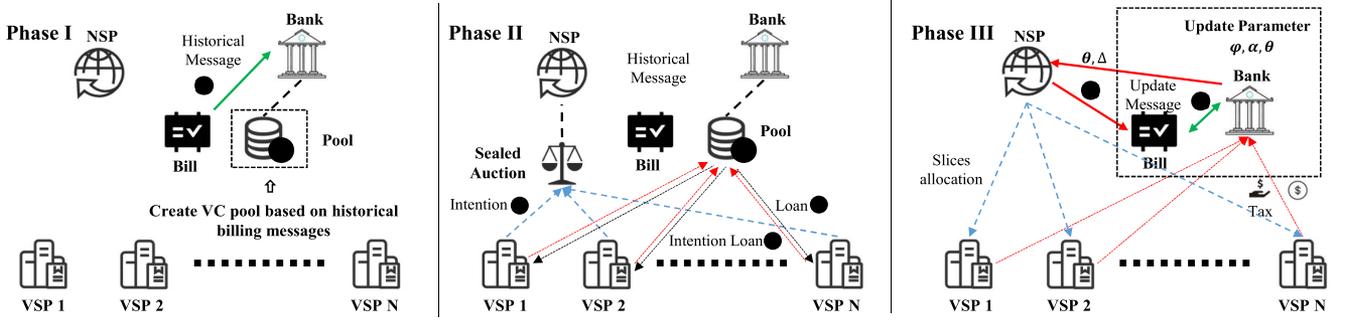
Fig. 2.   Three phases of VCG auction mechanism for dual currency.

Where $\hat{\theta}_i$ represents the estimated parameters of the function $u_i(\cdot)$, Δ represents the amount of additional resources to be allocated to each VSP in instances where the aggregated bids from all VSPs exceed the maximum social welfare. If the aggregated bids do not exceed this threshold, $\Delta = 0$. The expression $\mathcal{N}(0, \Delta^2 \boldsymbol{I})$ denotes a multivariate Gaussian random variable with a mean of zero and a variance of Δ^2, utilized to add stochastic elements to each allocation of resources. The function $\mathrm{norm}(\cdot)$ ensures that the total of the resource allocation vector is kept within the permitted bounds, hence adhering to the established resource constraints. It is important to note that the above two parameters $\hat{\theta}_i$ and Δ cannot be independently obtained by the NSP and must be updated with the assistance of the bank. During this process, the NSP does not need to directly access the original billing; the specific implementation method can be referred to in the detailed explanation provided in Section V.

- The bank computes the VC interest that VSP i is required to pay during round t as $|\mathcal{K}_{i,t} - U_{i,t}^*|$, where $U_{i,t}^* = u_i(s_{i,t}^*; \hat{\theta}_i)$. In addition, the social welfare impact of VSP i on other participants can be calculated as

$$\mathcal{I}_{i,t} = \left(\max_{\boldsymbol{s}_t} \sum_{j=1, j \neq i}^{N} u_j(s_{j,t}; \hat{\theta}_j) \right) - \sum_{j=1, j \neq i}^{N} u_j(\tilde{s}_{j,t}; \hat{\theta}_j). \tag{17}$$

- Subsequently, the bank amalgamates the VC interests of the VSP with their repercussions on societal welfare, imposing a tangible financial tax Z_i on the VSP. The methodology for computing this taxation is delineated as

$$Z_{i,t} = \hat{\alpha}_i(\mathcal{I}_{i,t} + |\mathcal{K}_{i,t} - U_{i,t}^*|). \tag{18}$$

- Finally, the bank will collect all bill information from the current auctioning round, appending it to pre-existing bills denoted as $\mathcal{BI}^+$ while reevaluating all parameters. The specifics of the collected information and updated parameters include the following parts

$$\mathcal{BI}^+ = \mathcal{BI} \cup (\tilde{\boldsymbol{s}}_t, \boldsymbol{x}_t, \mathcal{K}_t, U_t^*, \mathcal{I}_t, \mathcal{B}_t), \tag{19}$$

$$(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) \to \underset{\boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\theta}}{\arg\min} \mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathcal{BI}^+), \tag{20}$$

where parameters $\boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\theta}$ exemplify the parameter vectors constituted by corresponding parameters, while $\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}$ denote the inference pertaining to the aforementioned parameter vector. The function $\mathcal{L}(\cdot)$ is the loss function used for parameter estimation.

B. Dominant Strategy Incentive Compatible Analysis

DSIC, an essential concept originating from game theory and mechanism design, entails a circumstance in which the architect's goal is to establish an environment wherein participants' optimal actions, or their "dominant strategies", are in unison with the objectives of the design architect [6]. In this paper, our design target for this auction mechanism is to optimize social welfare. Hence, an auction mechanism possessing the DSIC attribute implies that the maximization of social welfare aligns with the VSP's aim of profit maximization.

As articulated in Section III, our problem essentially maps to a multi-round auction scenario. For an analysis of the DSIC properties of the auction mechanism, it's beneficial to consider the VSP's expected profit during the T-round auction when $T \to \infty$. Here, we operate under the assumption that the mean value of the random variable w_i is designated as $\mu(w_i)$, and the estimator of the parameters $(\boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\theta})$ are unbiased. The expected profit for a VSP's bid can thus be represented as

$$\mathbf{E}[P_i | x_i] = \alpha_i u_i(\tilde{s}_i; \theta) + \mu(w_i)$$
$$- \mathbf{E}[\hat{\alpha}_i(|\mathcal{K}_i - U_i^*| + \mathcal{SW}_{-i}^* - \mathcal{SW}_{-i}(\tilde{\boldsymbol{s}}))|x_i], \tag{22}$$

where $\mathcal{SW}_{-i}^* = \max_{\boldsymbol{s}} \sum_{j=1, j \neq i}^{N} u_j(s_j; \hat{\theta}_j)$ represents optimal social welfare allocation result except VSP i, and $\mathcal{SW}_{-i}(\tilde{\boldsymbol{s}}) = \sum_{j=1, j \neq i}^{N} u_j(\tilde{s}_j; \hat{\theta}_j)$ represents current social welfare except VSP i. For brevity in our exposition, we have intentionally omitted the round marker, t, from the formula above. As the auction rounds converge towards the endless, we can assume that the sample set $\mathcal{BI}$ from historical records assumes a large magnitude. This ensures the estimation of parameters $(\boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\theta})$ independent of the VSP bidding stratagem because of our auction mechanism. Therefore, the expected profit for VSP can be simplified to the following form as (21) shown at the bottom of the next page.

Term 1 of the (21) remains irrelevant to VSP's bid. In term 2, since the parameters θ and ϕ are unbiased, as the number

of auction rounds tends to infinity, the VoI of VSP i which corresponds to the function $f_i(\mathcal{BI}; \phi_i)$ will tend to $u_i(s_i^*; \theta_i)$. Then, the term 2 can be rewritten as

$$\mathcal{SW}(\tilde{s}) - \left| \sum_{j=1}^{N} u_j(s_j^*; \theta_i) * \frac{x_i}{\sum_{j=1}^{N} x_j} - u_i(s_i^*; \theta_i) \right|, \quad (23)$$

and we discern term 2 under two cases.

Case I: Given a VSP i, when other VSPs disclose their genuine VoI as the bid. Term 2 can be rewritten as

$$\mathcal{SW}(\tilde{s}) - \left| \frac{x_i * \sum_{j=1}^{N} u_j(s_j^*; \theta_i)}{x_i + \sum_{j=1, j \neq i}^{N} u_j(s_j^*; \theta_i)} - u_i(s_i^*; \theta_i) \right|. \quad (24)$$

The dominant strategy for VSP i lies in bidding equivalent to its authentic value of VoI. In this case, the SW depends on the bid x_i of VSP. When VSP i's bid exceeds the true VoI, $\mathcal{SW}(\tilde{s})$ is admittedly less than the optimum $\mathcal{SW}^*$, and when x_i is lower than the true VoI, $\mathcal{SW}(\tilde{s})$ is equivalent to $\mathcal{SW}^*$. Due to the presence of the absolute value terms, $x_i = u_i(s_i^*; \theta_i)$ is the dominant strategy.

Case II: When some VSPs abstain from tendering their actual VoI as a bid, it is construed that a state of collusion exists amongst the VSPs. Without loss of universality, we postulate that the assembly of VSPs participating in such collusion is denominated as set $\mathcal{M}$. It's essential to underscore that for the $\mathcal{SW}(\tilde{s})$, even under collusive conditions, the VSP has no incentive to elevate the bid beyond the true VoI. Because under parameter-unbiased circumstances, the $\mathcal{SW}(\tilde{s})$ cannot transcend the optimum $\mathcal{SW}^*$. Under such conditions, the absolute value part for any given VSP i in $\mathcal{M}$ can be expressed as

$$\left| \frac{x_i * \sum_{j=1}^{N} u_j(s_j^*; \theta_i)}{x_i + \sum_{m \in \mathcal{M}, m \neq i} x_m + \sum_{j=1, j \notin \mathcal{M}}^{N} u_j(s_j^*; \theta_i)} - u_i(s_i^*; \theta_i) \right|. \quad (25)$$

Drawing upon the characteristics of absolute values, it becomes apparent that the lowest point of the term occurs at 0, under which conditions x_i is determined as

$$x_i = \frac{u_i(s_i^*; \theta_i) \left(\sum_{m \in \mathcal{M}, m \neq i} x_m + \sum_{j=1, j \notin \mathcal{M}}^{N} u_j(s_j^*; \theta_i) \right)}{\sum_{j=1, j \neq i}^{N} u_j(s_j^*; \theta_i)} \quad (26)$$

Within the set $\mathcal{M}$, for VSPs to achieve collusion, a unified depreciation of the term by all members is essential. Without such depreciation, any VSP that fails to reduce the term to its minimal value finds no incentive to engage in collusive activities. Consequently, the bidding behavior of each VSP in set $\mathcal{M}$ must conform to (26). This stipulation leads to a system of linear equations characterized by $\mathcal{M}$ variables, expressed as $\mathbf{A}x_{\mathcal{M}} = \mathbf{b}$, where $\mathbf{A}$ is a matrix of full rank. This system, therefore, yields a unique solution.

From Case I, we observe that $x_i = u_i(s_i^*; \theta_i)$ presents itself as a non-zero solution within the system. Together with the specified conditions, it also stands as the sole non-zero solution. Therefore, it can be concluded that in scenarios where VSPs collude, tendering their genuine VoI in bids retains its status as a dominant strategy.

C. Other Discussion

In addition to DISC, there are two auction mechanism-related issues worth discussing, with the necessity of component design being the first. The introduction of the bank as a distinct role in the DVCG auction mechanism brings about a series of new components, making it necessary to discuss its design rationale. In the DVCG mechanism, the role of the bank is primarily designed to assist in handling privacy-related information. More specifically, it ensures that the sensitive information collected (i.e., the financial information of VSPs) is always processed within a trusted zone (i.e., the bank). Here, "trusted" refers to the bank being a public organization or government agency that is completely unrelated to auction interests, thus having no incentive to collude with other VSPs or NSPs. The bank ensures that the profit and utility functions of VSPs are not disclosed to NSPs and other VSPs by sharing parameters rather than actual billing information. Compared to the original VCG auction, the introduction of the bank reduces the risk of privacy leakage, thereby enhancing the willingness of VSPs to participate in the auction. Furthermore, the bank can generate VC interest through funding pools and lending mechanisms, making truthful bidding a dominant strategy for VSPs, even in the presence of collusion.

Another issue that requires further discussion is the fairness of resource pricing. Similar to the original VCG, the DVCG mechanism possesses pricing rationality but typically cannot achieve completely fair pricing. This implies that the unit prices at which VSPs obtain resources are inconsistent. On the positive

$$\mathbf{E}[P_i|x_i] = \alpha_i u_i(\tilde{s}_i; \theta) + \mu(w_i) + \mathbf{E}\left[\hat{\alpha}_i|x_i\right] \left(\sum_{j=1, j \neq i}^{N} u_j \left(\mathbf{E}\left[\tilde{s}_j|x_i\right]; \mathbf{E}\left[\hat{\theta}_j|x_i\right] \right) \right) - \mathbf{E}\left[\hat{\alpha}_i|x_i\right] \left(\max_{s} \sum_{j=1, j \neq i}^{N} u_j \left(s_j; \mathbf{E}\left[\hat{\theta}_j|x_i\right] \right) \right)$$

$$- \mathbf{E}\left[\hat{\alpha}_i|x_i\right] \left| \sum_{i=1}^{N} f_i \left(\mathcal{BI}; \mathbf{E}\left[\hat{\phi}_i|x_i\right] \right) * \frac{x_i}{\sum_{j=1}^{N} \boldsymbol{x}_j} - u \left(s_i^*; \mathbf{E}\left[\hat{\theta}_i|x_i\right] \right) \right|$$

$$= \underbrace{\mu(w_i) - \alpha_i \mathcal{SW}_{-i}^*}_{\textbf{Term1}} + \underbrace{\alpha_i \left(\mathcal{SW}(\tilde{s}) - \left| \sum_{j=1}^{N} f_j(\mathcal{BI}; \phi_j) * \frac{x_i}{\sum_{j=1}^{N} \boldsymbol{x}_j} - u_i(s_i^*; \theta_i) \right| \right)}_{\textbf{Term2}} \quad (21)$$

side, although lacking pricing fairness, the resource pricing in DVCG remains rational, as its pricing rules adhere to the principle that the winner pays the social cost, with the amount paid by the winner equaling the loss incurred by displacing other bidders' opportunities. Furthermore, the unfairness in resource pricing is a reasonable trade-off made to achieve the objective of maximizing social welfare; maintaining pricing fairness could inevitably harm overall social welfare.

V. MARL ALGORITHM

To enact the mechanism articulated in Section IV, a cardinal issue is estimating parameters and resolving the optimization problem delineated in (14). The parameter α_i can be estimated unbiasedly using the least squares method. The parameters θ, ϕ and the optimization problem (14) can be interpreted as a variant of the multi-armed bandit problem. The VSPs can be seen as several arms, where the payoff of each arm corresponds to the societal welfare provided in any given slice allocation scenario. Thus, the problem can be modeled as a Markov Decision Process (MDP) problem, wherein its states, actions, and rewards are defined as follows:

State: We designate the resource slice allocation s_{t-1} as the observation state o_t for round t. Hence, the s_{t-1} consists of a $3 \times N$ matrix, where N stands for the collective total of VSPs, and 3 denotes the varieties of resources.

Action: We stipulate the action as the decision on adjusting resource allocation based on the current observation state o_t, i.e., a_t - also a $3 \times N$ matrix. Each element of this matrix has a discrete value in the $[-0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3]$, illustrating the proportion of the corresponding VSP's resource allocation either increased or decreased when juxtaposed with the last round. To adhere to constraint (10a), a normalization map is necessitated on the outcome of the adjusted resource allocation.

Reward: After a round of slicing segmentation, the reward of the MDP is the cumulative bid sum gained by all VSPs as

$$\mathcal{R}_t = \sum_{i=1}^{N} x_{i,t}, \qquad (27)$$

Consequently, we are capable of employing RL methodologies to train a network, designed for the optimization of resource slicing partitioning, to stand instead of (14) and parameter θ. However, in the design phase of said reinforcement learning strategies, we are confronted with a pair of issues. The first pertains to scalability; as previously mentioned in Section I, the quantity of VSPs is subject to frequent modification. We aspire to craft an auction mechanism that adaptively accommodates the variability in the number of VSPs. For conventional neural networks (such as MLPs), this implies transformative changes to the network's width, necessitating a process of retraining and redeployment. The second predicament is found within (17), in which there is a requisite to obtain the socially optimal allocation devoid of VSP i. This, by implication, necessitates not a unified neural network, but rather N individual networks that are amenable to arbitrary combinations. On the other hand, due to the actions of the MDP being enshrined within a $3 \times N$

matrix, this insinuates the possibility of the action space becoming exceedingly vast, particularly in circumstances wherein the count of VSPs is considerable, thus giving rise to the dimensional curse amid learning and optimization procedures.

To address the above complexities, the MFMARL framework permits the training of N neural networks as agents, and employing these agents to infer socially optimal allocation schemes of any combination. For convenience, we still use the VSP numbers to label agents. Concurrently, with the aid of mean field theory, we can substantially mitigate the dimensionality of the joint action space. Within MFMARL, the joint action-value function of agent i can be deconstructed as follows:

$$\mathcal{Q}_i(o_t, a_t) = \frac{1}{\mathcal{N}^i} \sum_{j \in \mathcal{N}(i)} \mathcal{Q}_i(o_{i,t}, a_{i,t}, m_{j,t})$$

$$\approx Q_i(o_{i,t}, a_{i,t}, \overline{m}_{i,t}) \qquad (28)$$

where $\overline{m}_{i,t} = \frac{1}{\mathcal{N}^i} \sum_{j \in \mathcal{N}(i)} m_{j,t}$, $m_{j,t} = \overline{m}_{i,t} + \delta m^{i,j}$ and $m_{j,t} = (o_{j,t}, a_{j,t})$ represents the one-hot encoding of agent j's discrete action $a_{j,t}$ and state $o_{i,t}$, that is, if $a_{j,t} = -0.3$ $\overline{m}_{j,t} \triangleq [1, 0, 0, 0, 0, 0, 0, o_{i,t}]$. $\delta m^{i,j}$ is a small fluctuation between m_j and mean-field $\overline{m}_i$. $\mathcal{N}(i)$ denotes the neighborhood set of agent i. $\overline{m}_i$ can be interpreted as the empirical distribution of the state and the actions taken by agent j's neighbors, equivalently, an estimation of the action distribution within the cohort of agents in the mean-field. The proof for (28) is detailedly expounded in [37].

The agents' networks employ a dual Actor-Critic architecture, with each agent possessing a target and online network. Both the online actor and target actor are deployed at the decision center of the NSP, whereas the online critic and the target critic are positioned within the bank. The auction mechanism under the MFMARL framework manifests as illustrated in Fig. 3. Consistent with our descriptions in Section IV, the entire auction process is divided into three distinct phases.

In the initial phase, the bank estimates the funding pool for the current round based on the allocation from the previous auction. To achieve this, we trained an MLP network called the VoI network. The input to the network is the allocation scheme, and the output is the VoI. The network is represented as $f_i(\cdot)$, where $\text{VoI}_{i,t} = f(o_{i,t-1}, a_{i,t-1})$. The update of the VC pool is represented as

$$\mathcal{F}_{\mathcal{P}}(t) = \sum_{i=1}^{N} f_i(o_{i,t-1}, a_{i,t-1}), \qquad (29)$$

In the second phase, the VSP dispatches a bid x_t to the bank and secures the loan $\mathcal{K}_t$. Meanwhile, the NSP, guided by the observed state o_t, employs the online actor to acquire an action a_t. We employ $\mu_i(o_{i,t}, \overline{o}_{i,t})$ and $\mu_i'(o_{i,t}, \overline{o}_{i,t})$ to represent the online actor and target actor networks respectively, the input to the actor-network being the observed state o_i to wit, the resource slices apportioned to VSP i during previous rounds, and the average slices allocated to each VSP excluding i, notated as $\overline{o}_{i,t}$.

To compute the taxes payable by VSP i, the NSP presents the joint observed states o_t and actions a_t to the bank. The bank, in
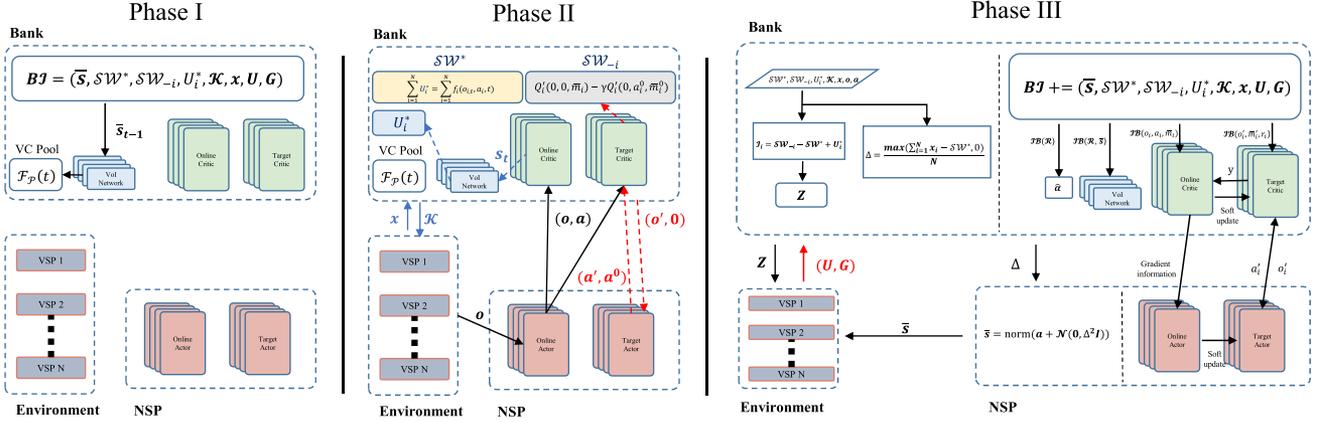
Fig. 3.   Three phases of VCG auction mechanism for dual currency under MFMARL framework.

response, employs the critic network nested within it to estimate the optimal societal welfare $\mathcal{SW}_{-i}$ without VSP i. Corresponding to the actor network, the critic network also encompasses online and target versions, signified as $Q_i(o_{i,t}, a_{i,t}, \overline{m}_{i,t})$ and $Q'_i(o_{i,t}, a_{i,t}, \overline{m}_{i,t})$ respectively. $\mathcal{SW}_{-i}$ can be estimated as follows:

$$\mathcal{SW}_{-i} = \min\left(\mathcal{F}_{\mathcal{P}max}, \max\left(\mathcal{SW}'_{-i}, \mathcal{F}_{\mathcal{P}max}^{-i}\right)\right), \quad (30)$$

$$\mathcal{SW}'_{-i} = Q'_i(\mathbf{0}, \mathbf{0}, \overline{m}_{i,t}^0) - \gamma Q'_i(\mathbf{0}, a_{i,t+1}^0, \overline{m}_{i,t+1}^0), \quad (31)$$

$$\overline{m}_{i,t}^0 = \frac{1}{\mathcal{N}^i} \sum_{j \in \mathcal{N}(i)} m_{j,t}\big|_{a_{j,t} = \mu'(o_{j,t}, \overline{o}_{j,t}|o_{i,t}=0)}, \quad (32)$$

$$\overline{m}_{i,t+1}^0 = \frac{1}{\mathcal{N}^i} \sum_{j \in \mathcal{N}(i)} m_{j,t+1}\big|_{a_{j,t+1} = \mu'(o_{j,t+1}, \overline{o}_{j,t+1}|o_{i,t+1}=0)}, \quad (33)$$

where $\mathcal{F}_{\mathcal{P}max}$ and $\mathcal{F}_{\mathcal{P}max}^{-i}$ is the maximum value of $\sum_{i=0}^{N} u_{j,t}$ and $\sum_{j=0, j \neq i}^{N} u_{i,t}$ in $\mathcal{BI}$, γ is the discount factor, observation state $\mathbf{0}$ implies that no resources are allocated to VSP i, and action $\mathbf{0}$ implies that there will be no adjustment for VSP's resource slices in this round. $\overline{m}_{i,t}^0 = \frac{1}{\mathcal{N}^i} \sum_{j \in \mathcal{N}(i)} m_{j,t}\big|_{a_{j,t} = \mu'(o_{j,t}, \overline{o}_{j,t}|o_{i,t}=0)}$, $a_{i,t+1}^0 = \mu'_i(\mathbf{0}, \overline{o}_{i,t+1})$ and $\overline{m}_{i,t+1}^0 = \frac{1}{\mathcal{N}^i} \sum_{j \in \mathcal{N}(i)} m_{j,t+1}\big|_{a_{i,t+1} = a_{i,t+1}^0}$ can be calculated using o_t and a_t.

In the third phase, the bank calculates the amount of tax the VSP must pay according to the currently known information. As stated in the (15), by comparing the size of $\mathcal{SW}^*$ and the sum of inverse bids x, the allocation of resource slices in this round is determined. Noticeably, to meet constraint (10a), the resource slice adjusted by the actor-network needs linear normalization, ensuring the total of each resource does not exceed the physical resource limit. Subsequently, according to (17) and (18), tax is calculated and the resource allocation is completed.

Then, the bank incorporates the financial report information $\mathcal{R}$ collected in the current round, compounded with the intermediate information generated during the auction process of the round into the bill, i.e., into the replay buffer. The involved parameters in the auction are also updated through the replay buffer. With α being estimated via the least squares method, its loss function is as follows

$$\nabla_{\alpha_i} \mathcal{L}(\alpha_i) = \mathbb{E}_{\mathcal{R}_i \sim \mathcal{BI}} \left[\left(\frac{G_i}{U_i} - \alpha_i \right)^2 \right]. \quad (34)$$

The loss function for the VoI network is formulated as

$$\nabla_{\phi_i} \mathcal{L}(\phi_i) = \mathbb{E}_{o_i, a_i, U_i \sim \mathcal{BI}} \left[(f_i(o_i, a_i) - U_i)^2 \right]. \quad (35)$$

For each agent's online actor network, the anticipated gradient update $\nabla_{\theta_i} J(\mu_i)$ can be expressed as

$$\mathbb{E}_{o \sim \mathcal{BI}, a_i \sim \mu_i}[\nabla_{\theta_i} \mu_i(o_i, \overline{o}_i) \nabla_{a_i} Q_i(o_i, a_i, \overline{m}_i)|_{a_i = \mu_i(o_i, \overline{o}_i)}].$$

Herein, μ_i is the policy of agent i, Q_i^μ is agent i's online critic network, and the bill $\mathcal{BI}$ is also considered as the experience replay buffer. For the update of the online critic network, we typically employ Temporal Difference error to calculate the discrepancy between the target Q value and the current Q value, then employ back-propagation to update the parameters of the critic network. This can be depicted as:

$$\nabla_{\psi_i} \mathcal{L}(\psi_i) = \mathbb{E}_{o, a, r, o^\dagger \sim \mathcal{BI}}[(Q_i(o_i, a_i, \overline{m}_i) - y)^2]. \quad (36)$$

In this case, $o^\dagger$ is the next round observation state and $y = r_i + \gamma Q'_i(o_i^\dagger, a_i^\dagger, \overline{m}_i^\dagger)|_{a_i^\dagger = \mu'(o_i^\dagger, \overline{o}_i^\dagger)}$ is the target Q-value. Finally, the parameters of the target network are updated using a soft update mechanism. Specifically, for both the actor and critic network parameters, the update process can be expressed as:

$$\theta'_i = \tau \theta_i + (1 - \tau) \theta'_i, \quad (37)$$

$$\psi'_i = \tau \psi_i + (1 - \tau) \psi'_i, \quad (38)$$

where θ_i and ψ_i are the parameters of the online actor and critic networks, θ'_i and ψ'_i are the parameters of the target actor and critic networks, and τ is the ratio of the soft update. The entire auction and training process is summarised in Algorithm 1.

Through the aforementioned MFMRAL framework, the DVCG mechanism is implemented. At the end of this chapter, we would like to appropriately discuss three related issues in the implementation process of the DVCG mechanism. The first issue is the additional communication and computational overhead generated by the implementation mechanism. This additional

---

Algorithm 1: Three-Phase Auction Mechanism With MF-MARL Architecture.

Require: VSP set $\mathcal{V} = \{1, \ldots, N\}$, VoI networks $\{f_i(\cdot; \phi_i)\}_{i=1}^N$, Actor-Critic networks $\{\mu_i, Q_i\}_{i=1}^N$,

Ensure: Allocation scheme $\mathbf{s}_t$, Payment vector Z_t, Updated networks $\{f_i, \mu_i, Q_i\}_{i=1}^N$

1: **Phase 1: Funding Pool Estimation**
2: **for** each VSP $i \in \mathcal{V}$ **do**
3:     Compute VoI: $\text{VoI}_{i,t} \leftarrow f_i(o_{i,t-1}, a_{i,t-1}; \phi_i)$
4: **end for**
5: Update VC pool: $\mathcal{F}_\mathcal{P}(t) \leftarrow \sum_{i=1}^N \text{VoI}_{i,t}$
6: **Phase 2: Sealed Bidding**
7: **for** each VSP $i \in \mathcal{V}$ **do**
8:     Submit sealed bid $x_{i,t}$ to the bank and NSP
9: **end for**
10: Bank computes loan $\mathcal{K}$ based on Eq. (13)
11: NSP computes action: $\mathbf{a}_t \leftarrow \{\mu_i(o_{i,t}, \overline{o}_{i,t}; \theta_i)\}_{i=1}^N$
12: **Phase 3: Tax Calculation & Update**
13: Bank computes $\mathcal{SW}^* \leftarrow \sum_{i=1}^N f_i(o_{i,t}, a_{i,t}; \phi_i)$
14: **for** each VSP $i \in \mathcal{V}$ **do**
15:     Compute $\mathcal{SW}_{-i} \leftarrow \min(\mathcal{F}_{\mathcal{P}max}, \max(\mathcal{SW}'_{-i}, \mathcal{F}_{\mathcal{P}max}^{-i}))$
16:     Compute tax: $Z_{i,t}$ based on Eq. (18)
17: **end for**
18: NSP computes $\mathbf{s}_t$ based on Eq. (15)
19: **Network Updates**
20: Sample batch $\mathcal{B} \sim \mathcal{BI}$
21: **for** each VSP $i \in \mathcal{V}$ **do**
22:     Update VoI net: $\nabla_{\phi_i} \mathcal{L} = \mathbb{E}_\mathcal{B}[(f_i(o_i, a_i) - U_i)^2]$
23:     Update Critic: $\nabla_{\psi_i} \mathcal{L} = \mathbb{E}_\mathcal{B}[(Q_i(o_i, a_i, \overline{m}_i) - y_i)^2]$
24:     Update Actor: $\nabla_{\theta_i} J = \mathbb{E}_\mathcal{B}[\nabla_{\theta_i} \mu_i \nabla_a Q_i|_{a=\mu_i}]$
25:     Soft update targets: $\theta'_i \leftarrow \tau\theta_i + (1-\tau)\theta'_i$
26:         $\psi'_i \leftarrow \tau\psi_i + (1-\tau)\psi'_i$
27: **end for**
28: **return** $\mathbf{S}_t, Z_t, \{f_i, \mu_i, Q_i\}_{i=1}^N$

TABLE I
PARAMETER SETTING

Parameter	Detail
Discount Factor γ	0.99
Learning Rate	0.001
Number of Episodes	1000
Batch Size	128
Replay Buffer Size	10000
Soft Update Ratio τ	0.005

making the ledger update process complicated and costly. More importantly, the ledger information in a blockchain is completely public, which means that the privacy of the VSP may not be adequately protected.

The third issue is the delayed rewards problem in the implementation of the mechanism. The occurrence of delayed rewards mainly arises from the fact that the actual rewards of the DVCG mechanism are transmitted through the chain of financial reporting, the VoI network, and VSP bidding. Considering that the VoI network needs to be trained and the VSP strategy requires response time, the existence of rewards is delayed. Fortunately, the utility function of the VSP typically does not exhibit sparse limit characteristics, meaning the function itself is relatively smooth (generally, performance increases with the increase of resources). This leads to the DVCG mechanism causing a delayed rewards issue, but it does not result in sparse rewards, making the delayed rewards problem not a significant influencing factor on the stability of the strategy. One possible direction for improvement is to use a reinforcement learning framework specifically designed for the delayed rewards problem, providing more stable training results.

communication overhead primarily occurs between the NSP and the bank (i.e., the gradient information transmitted during the training of the network), while the communication overhead for the VSP is significantly reduced. Generally, the NSP and the bank can effectively reduce additional communication costs by establishing dedicated lines, which is more economically viable compared to setting up dedicated communication lines between multiple VSPs and the NSP. The introduced computational overhead mainly comes from the training costs of the three neural networks. Considering that the training process of neural networks can be further parallelized, the DVCG mechanism remains competitive in terms of efficiency and cost-effectiveness.

The second issue is the trade-off regarding security and privacy in the design process of the DVCG. The current version of DVCG has potential security risks. If the bank suffers from a malicious attack, it will severely impact the performance of the auction mechanism. Theoretically, a more secure solution would be to establish a decentralized ledger maintenance mechanism jointly by the NSP, VSP, and the bank. For example, using consortium blockchain technology, multiple nodes could jointly maintain the ledger information to achieve higher security. However, decentralized mechanisms represented by blockchain often require the design of complex consensus mechanisms,

VI. SIMULATION RESULTS

In the simulation section, we consider a multi-VSP network environment. The utility function of each VSP is set as

$$u(s_0, s_1, s_2) = \beta \left(1 - e^{-(\xi_0 s_0 + \xi_1 s_1 + \xi_2 s_2)}\right), \quad (39)$$

where $\beta \sim U(100, 200)$ and $\xi_0, \xi_1, \xi_2 \sim U(0, 10)$. s_0, s_1, s_2 denote the quantities of three distinct types of resource slices. The profit of VSPs is formulated as indicated in (5). Without loss of generality, $w \sim \mathcal{N}(\mu, \sigma^2)$ follows a normal distribution, unknown to the NSP. The total amount of all types of resources is normalized. In the simulation environment, the total number of VSPs is set between 10 and 50, and all algorithms are deployed under the same scenario with the same conditions. All experiments are conducted using Python 3.8 with PyTorch 2.0 framework on an Intel i7-10700 CPU platform. To ensure fair comparison, all neural network training processes are performed without parallelization. The involved additional parameters are consolidated in Table I. The algorithms involved in the simulation are as follows:

- **DVCG-MFMARL**(Proposed): A dual-currency VCG mechanism based on the MFMARL framework employs the dual-currency auction mechanism introduced in Section IV and the MFMARL to solve the slicing problem.
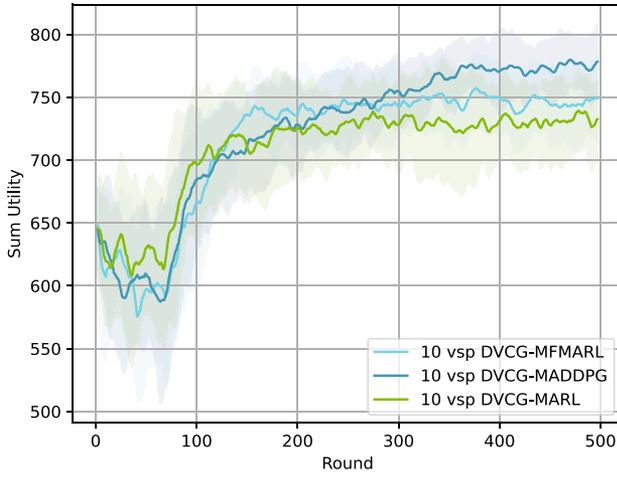
Fig. 4. Comparative convergence of the total social welfare under different multi-agent reinforcement learning algorithms when the number of VSPs is 10.
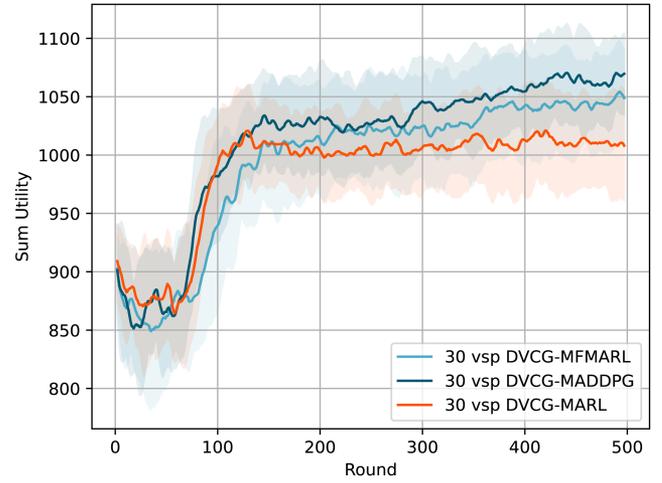

Fig. 5. Comparative convergence of the total social welfare under different multi-agent reinforcement learning algorithms when the number of VSPs is 30.

- **DVCG-MADDPG**: Similarly, it applies the auction mechanism presented in Section IV and utilizes the MADDPG framework to solve the slicing problem. Unlike the proposed algorithm, it uses joint actions and observed states as inputs for each agent network.
- **DVCG-MARL**: It also uses the auction mechanism introduced in Section IV, and employs the MARL framework to settle the slicing problem. It differs from the proposed algorithm by using only the observed states and actions of the corresponding VSP as inputs for each Agent's network.
- **QVCG**: The VCG mechanism proposed in [8] under limited communication conditions, posits resource fragmentation into small quantized units. Each auction participant reports a bid vector instead of a utility function to achieve maximum social welfare allocation. This bid vector $X = \{x_1, x_2, \ldots, x_M\}$ comprises of individual bids where $x_{im} = u(\frac{m}{M}) - u(\frac{m-1}{M})$, with $\frac{1}{M}$ being the quantized resource unit. Relevant to our proposed scenario, due to three types of resources, a matrix of dimensions $M \times M \times M$ is used instead of a bid vector. The QVCG mechanism completes the allocation by solving a constrained linear programming problem with a variable dimension of $N \times M \times M \times M$, where N is the number of participating VSPs.

A. Convergence Evaluation

We first show the convergence of the auction algorithm. As depicted in Figs. 4 and 5, we present the convergence curve graphs for the DVCG-MFMARL, DVCG-MADDPG, and DVCG-MARL algorithms under 10 and 30 VSPs. QVCG is not an online learning algorithm, so its convergence is not discussed. Please note that in the above comparative simulations, VSP employs a bidding strategy to maximize net profits. Its strategy utilizes a DDPG architecture AC network as a substitute, with the reward function being the net profit of the current round. As illustrated, DVCG-MFMARL has a faster convergence rate compared to DVCG-MADDPG and DVCG-MARL,

a consequence of DVCG-MFMARL having a smaller network size whilst witnessing more comprehensive global information. This advantage in the convergence speed of DVCG-MFMARL becomes more apparent as the number of intelligent agents increases. Simultaneously, DVCG-MADDPG has a higher total social welfare than DVCG-MFMARL, stemming from the inevitable information loss when DVCG-MFMARL simplifies observation and action dimensions using Mean Field Theory.

B. Performance Evaluation

Subsequently, we exhibit the cumulative social welfare achieved by four algorithms, wherein the auction was conducted for a total of 500 rounds. It is worth noting that for the QVCG algorithm, the precision of quantization, i.e., the size of the smallest resource unit, affects the attainable social welfare. Specifically, when the quantization precision is infinitely high, the allocation result theoretically approaches the optimal distribution. However, the dimension of the bid matrix also tends to be infinitely large. Moreover, the QVCG algorithm didn't consider the situation where bidders are not aware of their utility functions during its design. For the mechanism to be implemented, we conceded the QVCG algorithm, under the assumption that bidders are aware of parameters α, which is different from the settings of the other three online algorithms. For the QVCG algorithm, we introduced versions with quantization parameters of 0.01 and 0.05, respectively corresponding to different quantization accuracy.

As shown in Fig. 6, where the x-axis represents the number of intelligent agents and the y-axis represents the total social welfare, the DVCG-MADDPG, DVCG-MFMARL, and QVCG-0.01 algorithms more closely approach the optimal social welfare, outperforming the QVCG-0.01 and QVCG-0.05 algorithms. This performance gap narrows as the number of VSPs increases, mainly because a larger number of agents makes it more difficult for them to reach a tacit understanding. It should be noted that DVCG and QVCG mechanisms differ
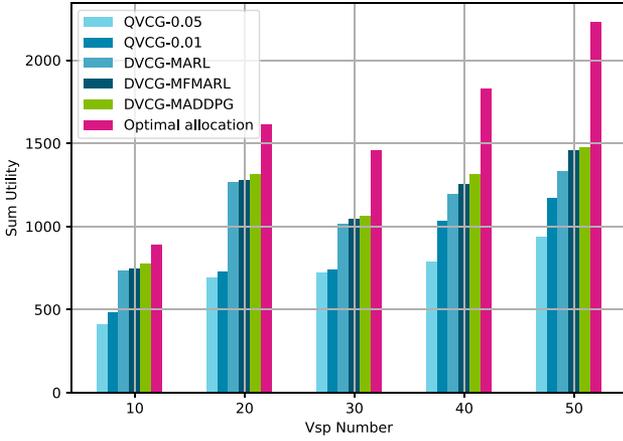
Fig. 6. The relationship between the number of VSPs and the total social welfare under various resource slicing algorithms.

TABLE II
INFORMATION EXCHANGE VOLUME (FLOAT $\times 10^6$)

N	10	20	30	40	50
DVCG-MFMARL	0.005	0.01	0.015	0.02	0.025
QVCG-0.05	0.08	0.16	0.24	0.32	0.40
QVCG-0.01	10.0	20.0	30.0	40.0	50.0

TABLE III
RUNNING TIME ($\times 10^3$ S)

N	10	20	30	40	50
DVCG-MFMARL	1.57	5.87	12.9	22.8	36.6
QVCG-0.05	0.19	1.82	3.68	9.35	30.4
QVCG-0.01	3.40	33.6	117	225	1140

significantly in the amount of information exchange required, with DVCG-MFMARL demonstrating higher efficiency in this regard.

Tables II and III present the information exchange volume and runtime of three algorithms over 500 auction rounds with the number of VSPs ranging from 10 to 50. The information exchange volume measures the total amount of information exchanged between VSPs and NSPs, quantified in the number of floating-point values. Specifically, the DVCG mechanism requires the exchange of bid information in each auction round, so its information exchange volume is calculated as the number of VSPs multiplied by the number of rounds. In contrast, the QVCG mechanism requires each VSP to exchange a matrix of size M^3 (where M represents the quantization precision), and this exchange occurs only once regardless of the number of auction rounds. The results show that in realistic scenarios (i.e., with a limited number of auction rounds), the information exchange volume of the DVCG mechanism is significantly lower than that of the QVCG-0.01 and QVCG-0.05 mechanisms. Additionally, the runtime of the DVCG mechanism falls between that of the two QVCG mechanisms with different precision levels.

Furthermore, we elucidated the DSIC properties of the DVCG mechanism and its robustness in counteracting collusions among VSPs through our simulation. In Fig. 7, we demonstrate the DSIC characteristic of DVCG, where the VSP number is set to
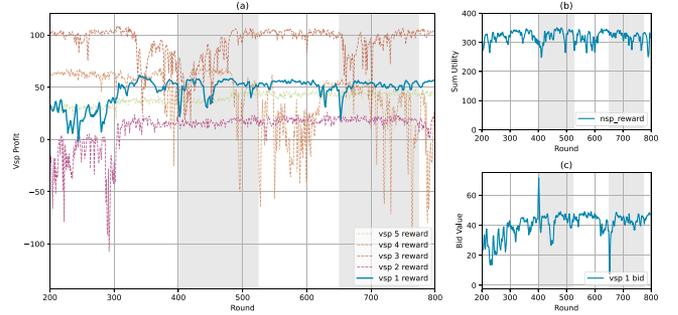

Fig. 7. Verification of the DSIC properties in the DVCG-MFMARL algorithm: (a) Oscillations in VSP profits; (b) Variations in the total social welfare; (c) Shifts in the bidding behavior of disruptive VSP.
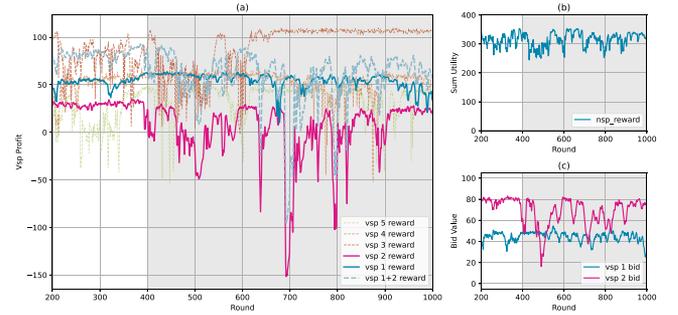

Fig. 8. Evaluation of the DVCG-MFMARL algorithm's resilience against collusive behavior: (a) Oscillations in VSP profits; (b) Variations in the total social welfare; (c) Shifts in the bidding behavior of disruptive VSPs.

5. The x-axis represents the number of auction rounds, while the y-axis records the net profit of the VSPs. Specifically, VSP 1 is a unique auction participant used to illustrate the DISC properties of the DVCG mechanism. It adjusts its bidding strategy in the 400th and 650th rounds. With the 400th round, VSP 1 commences inflating its bid above its true VoI. Starting with the 650th round, VSP 1 commences deflating its bid below the authentic VoI. It is observable that, regardless of the strategy employed, the net earnings for VSP 1 can never surpass those achieved when the true VoI is reported. However, it is worth noting that VSP 1's reporting of the bid can impact the net earnings of other VSPs. This necessitates the validation of the DVCG's robustness in resisting collusion.

With the same parameters, we demonstrated the robustness of the DVCG mechanism for collusion among auction participants. In this simulation, we selected two particular individuals from the five VSPs, denoted VSP 0 and VSP 1. VSP 0 and VSP 1 form a colluding set, which means that they use the same joint action agent, with the reward function of the bidding agents of the VSP 0 and VSP 1 set as $r = P_{VSP0} + P_{VSP1}$. We trained the VSP agents under the DVCG-FMMARL mechanism and illustrated the bids after the strategy of the VSP agents had converged, as depicted in Fig. 8. The blue and pink lines represent the net profits of VSP 0 and VSP 1. It can be observed that, even in collusion, the VSP 0 and VSP 1 agents still use the real bid as their dominant strategy after adequate exploration, proving the robustness of the DVCG mechanism against collusion.
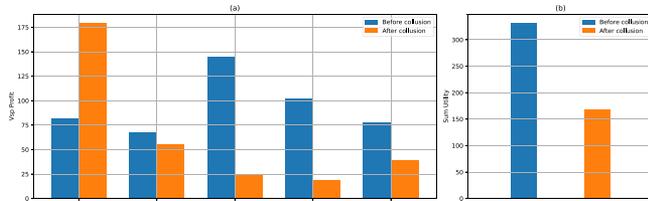
Fig. 9. Assessment of the QVCG algorithm's vulnerability to collusive behavior: (a) Comparative analysis of VSP profits pre- and post-collusion; (b) Contrast of total social welfare before and after the collision.
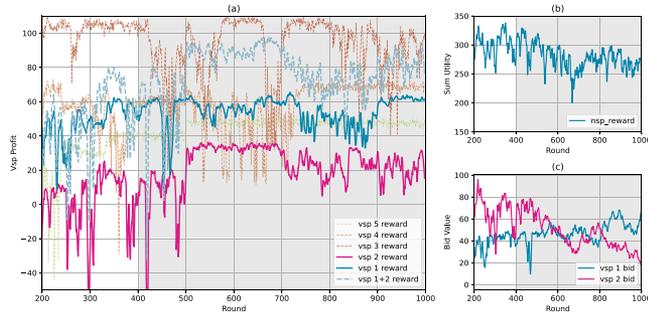

Fig. 10. Assessment of the NVI-DVCG algorithm's vulnerability to collusive behavior: (a) Oscillations in VSP profits; (b) Variations in the total social welfare; (c) Shifts in the bidding behavior of disruptive VSPs.

Simultaneously, we show the vulnerability of the QVCG mechanism when faced with collusion among VSPs. Due to the large dimensionality of the bid matrix under the QVCG mechanism, it is challenging to affirm its robustness against collusion through agent training. Therefore, we used an exhaustive search to demonstrate that under the QVCG mechanism, there is a colluding bidding scheme that, under the condition of collusion, increases the net profit of the colluding set (i.e., VSP 0 and VSP 1), as shown in Fig. 9.

C. Ablation Study

Finally, we conducted ablation studies on the proposed mechanisms and introduced a variant mechanism based on DVCG-FMMRAL, termed NVI-DVCG, which eliminates the VC interest mechanisms from DVCG. The simulation setup is consistent with the simulation used to verify serial robustness. As shown in Fig. 10, the NVI-DVCG mechanism significantly reduces the auction mechanism's resistance to serial stability compared to the DVCG mechanism. This diminished robustness allows VSP0 and VSP1 to collude, resulting in higher aggregate profits for the colluding parties but a decline in total social welfare.

The reason for this result is that for the NVI-DVCG auction mechanism, there exists a strategy that allows one participant to benefit while others manipulate their bids to skew the system's parameter estimates. This requires precisely suppressing certain utility values near optimal allocation that excludes VSP i without affecting other regions. However, when verification costs (VC interest) exist, any bid adjustment becomes prohibitively expensive, making such manipulation impractical as the required fine-tuning cannot be achieved. Thus, verification costs naturally

prevent this form of profit manipulation through coordinated bidding.

VII. CONCLUSION

This paper addresses the resource-slicing allocation problem in IoV by proposing a novel VoI-based dual-currency VCG auction mechanism. The mechanism achieves near-optimal social welfare improvement through optimized resource allocation without requiring knowledge of VSPs' revenue models. We theoretically and experimentally demonstrate that the mechanism satisfies DSIC properties, and can effectively mitigate potential collusion risks among bidders. Additionally, the mechanism significantly reduces the volume of information exchange per auction iteration while maintaining acceptable computational complexity, making it suitable for practical applications. In the future, we plan to extend our research to broader practical network scenarios and real-world optimization challenges. A representative focus will be solving multidimensional resource allocation problems in MIMO systems under complex constraints across frequency, temporal, spatial, and power domains, thereby broadening the horizons of this current discourse.

REFERENCES

[1] D.C. Nguyen et al., "6G Internet of Things: A comprehensive survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, Jan. 2022.
[2] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.
[3] R. Sun et al., "A comprehensive survey of knowledge-driven deep learning for intelligent wireless network optimization in 6G," *IEEE Commun. Surveys Tuts.*, early access, May 29, 2025, doi: 10.1109/COMST.2025.3574765.
[4] B. Ji et al., "Survey on the Internet of Vehicles: Network architectures and applications," *IEEE Commun. Standards Mag.*, vol. 4, no. 1, pp. 34–41, Mar. 2020.
[5] W. Zhuang, Q. Ye, F. Lyu, N. Cheng, and J. Ren, "SDN/NFV-empowered future IoV with enhanced communication, computing, and caching," *Proc. IEEE*, vol. 108, no. 2, pp. 274–291, Feb. 2020.
[6] L. Makowski and J. M. Ostroy, "Vickrey-Clarke-Groves mechanisms and perfect competition," *J. Econ. Theory*, vol. 42, no. 2, pp. 244–261, 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022053187900871
[7] M. H. Rothkopf, "Thirteen reasons why the Vickrey-Clarke-Groves process is not practical," *Operations Res.*, vol. 55, no. 2, pp. 191–197, 2007.
[8] H. Ge and R. A. Berry, "Dominant strategy allocation of divisible network resources with limited information exchange," in *Proc. IEEE INFOCOM 2018 – IEEE Conf. Comput. Commun.*, 2018, pp. 2753–2761.
[9] T. Alladi, V. Chamola, N. Sahu, V. Venkatesh, A. Goyal, and M. Guizani, "A comprehensive survey on the applications of blockchain for securing vehicular networks," *IEEE Commun. Surveys Tut.*, vol. 24, no. 2, pp. 1212–1239, Secondquarter 2022.
[10] J. Posner, L. Tseng, M. Aloqaily, and Y. Jararweh, "Federated learning in vehicular networks: Opportunities and solutions," *IEEE Netw.*, vol. 35, no. 2, pp. 152–159, Mar./Apr. 2021.
[11] Y. Bachrach, P. Key, and M. Zadimoghaddam, "Collusion in VCG path procurement auctions," in *Proc. Internet Netw. Econ.*, A. Saberi, Ed., Berlin, Germany: Springer, 2010, pp. 38–49.
[12] V. Conitzer and T. Sandholm, "Failures of the VCG mechanism in combinatorial auctions and exchanges," in *Proc. 5th Int. Joint Conf. Auton. Agents Multiagent Syst.*, in Series AAMAS'06. New York, NY, USA: Assoc. Comput. Machinery, 2006, pp. 521–528, doi: 10.1145/1160633.1160729.
[13] T. Soleymani, J. S. Baras, and S. Hirche, "Value of information in feedback control: Quantification," *IEEE Trans. Autom. Control*, vol. 67, no. 7, pp. 3730–3737, Jul. 2022.

[14] C. Song, C. Zhang, A. Shafieezadeh, and R. Xiao, "Value of information analysis in non-stationary stochastic decision environments: A reliability-assisted POMDP approach," *Rel. Eng. Syst. Saf.*, vol. 217, 2022, Art. no. 108034. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095183202100541X

[15] F. Alawad and F. A. Kraemer, "Value of information in wireless sensor network applications and the IoT: A review," *IEEE Sensors J.*, vol. 22, no. 10, pp. 9228–9245, May 2022.

[16] K. Zhang, Z. Yang, and T. Başar, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," in *Handbook of Reinforcement Learning and Control*, Cham, Switzerland: Springer, 2021, pp. 321–384, doi: 10.1007/978-3-030-60990-0_12.

[17] L. Busoniu, R. Babuska, and B. D. Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Trans. Syst., Man, Cybern., Part C (Appl. Rev.)*, vol. 38, no. 2, pp. 156–172, Mar. 2008.

[18] S. S. Padhi, S. M. Wagner, and P. K. J. Mohapatra, "Design of auction parameters to reduce the effect of collusion," *Decis. Sci.*, vol. 47, no. 6, pp. 1016–1047, 2016. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/deci.12159

[19] Y.-K. Che and J. Kim, "Optimal collusion-proof auctions," *J. Econ. Theory*, vol. 144, no. 2, pp. 565–603, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022053108001270

[20] A. Hu, T. Offerman, and S. Onderstal, "Fighting collusion in auctions: An experimental investigation," *Int. J. Ind. Org.*, vol. 29, no. 1, pp. 84–96, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167718709000575

[21] G. Vazquez-Vilar, C. Mosquera, and S. K. Jayaweera, "Primary user enters the game: Performance of dynamic spectrum leasing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 1–5, Dec. 2010.

[22] X. Kang, R. Zhang, and M. Motani, "Price-based resource allocation for spectrum-sharing femtocell networks: A stackelberg game approach," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 538–549, Apr. 2012.

[23] D. Niyato and E. Hossain, "Competitive spectrum sharing in cognitive radio networks: A dynamic game approach," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2651–2660, Jul. 2008.

[24] D. Niyato, E. Hossain, and Z. Han, "Dynamics of multiple-seller and multiple-buyer spectrum trading in cognitive radio networks: A game-theoretic modeling approach," *IEEE Trans. Mobile Comput.*, vol. 8, no. 8, pp. 1009–1022, Aug. 2009.

[25] Q. Zhu and X. Zhang, "Game-theory based power and spectrum virtualization for maximizing spectrum efficiency over mobile cloud-computing wireless networks," in *Proc. 49th Annu. Conf. Inf. Sci. Syst.*, 2015, pp. 1–6.

[26] L. Gao, P. Li, Z. Pan, N. Liu, and X. You, "Virtualization framework and VCG based resource block allocation scheme for lte virtualization," in *Proc. IEEE 83rd Veh. Technol. Conf.*, 2016, pp. 1–6.

[27] B. Fan, H. Tian, and B. Liu, "Game theory based power allocation in LTE air interface virtualization," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2015, pp. 972–976.

[28] F. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. Telecommun.*, vol. 8, no. 1, pp. 33–37, 1997. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.4460080106

[29] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, 1998, doi: 10.1057/palgrave.jors.2600523.

[30] R. Johari and J. N. Tsitsiklis, "Efficiency loss in a network resource allocation game," *Math. Operations Res.*, vol. 29, no. 3, pp. 407–435, 2004, doi: 10.1287/moor.1040.0091.

[31] S. Yang and B. Hajek, "VCG-Kelly mechanisms for allocation of divisible goods: Adapting VCG mechanisms to one-dimensional signals," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, pp. 1237–1243, Aug. 2007.

[32] R. Ma, D. M. Chiu, J. Lui, V. Misra, and D. Rubenstein, "On resource management for cloud users: A generalized Kelly mechanism approach," Elect. Eng., Tech. Rep., 2010. [Online]. Available: https://www.researchgate.net/publication/267418244_On_Resource_Management_for_Cloud_Users_A_Generalized_Kelly_Mechanism_Approach

[33] R. Johari and J. N. Tsitsiklis, "Efficiency of scalar-parameterized mechanisms," *Operations Res.*, vol. 57, no. 4, pp. 823–839, 2009, doi: 10.1287/opre.1080.0638.

[34] J. Zheng, Z.-T. Liang, Y. Li, Z. Li, and Q.-H. Wu, "Multi-agent reinforcement learning with privacy preservation for continuous double auction-based P2P energy trading," *IEEE Trans. Ind. Informat.*, vol. 20, no. 4, pp. 6582–6590, Apr. 2024.

[35] B. Yin, H. Weng, Y. Hu, J. Xi, P. Ding, and J. Liu, "Multi-agent deep reinforcement learning for simulating centralized double-sided auction electricity market," *IEEE Trans. Power Syst.*, vol. 40, no. 1, pp. 518–529, Jan. 2025.

[36] D. Qiu, J. Wang, J. Wang, and G. Strbac, "Multi-agent reinforcement learning for automated peer-to-peer energy trading in double-side auction market," in *Proc. 30th Int. Joint Conf. Artif. Intell.*, Z. H. Zhou, Ed., Int. Joint Conf. Artif. Intell. Org., Aug. 2021, pp. 2913–2920, doi: 10.24963/ijcai.2021/401.

[37] Y. Yang, R. Luo, M. Li, M. Zhou, W. Zhang, and J. Wang, "Mean field multi-agent reinforcement learning," in *Proc. 35th Int. Conf. Mach. Learn.*, in Series Proc. Mach. Learn. Res., J. Dy and A. Krause, Eds., Jul. 2018, vol. 80, pp. 5571–5580. [Online]. Available: https://proceedings.mlr.press/v80/yang18d.html

[38] H. Zhang et al., "Mean-field-aided multiagent reinforcement learning for resource allocation in vehicular networks," *IEEE Internet Things J.*, vol. 10, no. 3, pp. 2667–2679, Feb. 2023.

Wei Wang (Graduate Student Member, IEEE) received the BS degree from Sichuan University, Chengdu, China, in 2016. He is currently working toward the PhD degree from Xidian University, Xi'an, China. His research interests include resource scheduling in wireless networks and mobile edge computing.

Nan Cheng (Senior Member, IEEE) received the BE and MS degrees from the Department of Electronics and Information Engineering, Tongji University, Shanghai, China, in 2009 and 2012, respectively, and the PhD degree from the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2016. From 2017 to 2019, he was a postdoctoral fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. He is currently a professor with the State Key Laboratory of ISN and School of Telecommunication Engineering, Xidian University, Shaanxi, China. His research interests include B5G/6G, space-air-ground integrated network, Big Data in vehicular networks, and self-driving systems, performance analysis, MAC, opportunistic communication, and application of AI for vehicular networks.

Conghao Zhou (Member, IEEE) received the PhD degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada. He is currently a professor with the School of Telecommunications Engineering, Xidian University, Xi'an, China. He was a postdoctoral research fellow with the University of Waterloo. His research interests include immersive communications, AI for networking, and space-air-ground integrated networks.

Haixia Peng (Member, IEEE) received the PhD degree in computer science and electrical & computer engineering from Northeastern University, Shenyang, China, in 2017, and the Second PhD degree from the University of Waterloo, Waterloo, ON, Canada, in 2021. She is currently a full professor with the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, China. From 2021 to 2022, she was an assistant professor with the Department of Computer Engineering and Computer Science, California State University Long Beach, USA. Her research interests include satellite-terrestrial vehicular networks, multi-access edge computing, resource management, artificial intelligence, and reinforcement learning. She is/was a TPC member in IEEE VTC-fall 2016&2017, IEEE ICCEREC 2018, IEEE GlobeCom 2016-2024, and IEEE ICC 2017-2024 conferences, and is also an associate editor for the Peer-to-Peer Networking and Applications and IEEE IoT Magazine.

Haibo Zhou (Senior Member, IEEE) received the PhD degree in information and communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. From 2014 to 2017, he was a postdoctoral fellow with the Broadband Communications Research Group, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He is currently a full professor with the School of Electronic Science and Engineering, Nanjing University, Nanjing, China. His research interests include resource management and protocol design in B5G/6G networks, vehicular ad hoc networks, and space-air-ground integrated networks. He was a recipient of the 2019 IEEE ComSoc Asia–Pacific Outstanding Young Researcher Award, 2023-2024 IEEE ComSoc Distinguished Lecturer, 2023-2025 IEEE VTS Distinguished Lecturer, and 2023 IEEE ComSoc WTC Young Researcher Award. He was Track/Symposium CoChair for IEEE/CIC ICCC, IEEE VTC-Fall, WCSP, IEEE GLOBECOM, and IEEE ICC. He is also an associate editor of *IEEE Transactions on Wireless Communications*, *IEEE Internet of Things Journal*, *IEEE Network Magazine*, and *Journal of Communications and Information Networks*. In 2022, he was elected as an IET fellow and a highly cited Researcher by Clarivate Analytics in 2022 & 2020.

Zhou Su (Senior Member, IEEE) received the PhD degree from Waseda University, Tokyo, Japan, in 2003. His research interests include multimedia communications, wireless communications, and network traffic. He is currently a TPC Member of some flagship conferences, including IEEE INFOCOM, IEEE ICC, and IEEE Globecom. He was the recipient of the Best Paper Award of the International Conference IEEE BigdataSE2019, IEEE ComSocGCCTC2018, IEEECyber-SciTech2017, and the Funai Information Technology Award for Young Researchers, in 2009. He is also the chair of the Multimedia Services and Applications Over Emerging Networks Interest Group (MENIG), IEEE ComSoc Society, and Multimedia Communications Technical Committee. He also was the co-chair for several international conferences, including the IEEE Vehicular Technology Conference (VTC) Spring 2016 and the IEEE Consumer Communications and Networking Conference (CCNC) 2011.

Xuemin (Sherman) Shen (Fellow, IEEE) received the PhD degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is also a University professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include network resource management, wireless network security, Internet of Things, 5G and beyond, and vehicular networks. Dr. Shen was the recipient of the "West Lake Friendship Award" from Zhejiang Province in 2023, President's Excellence in Research from the University of Waterloo in 2022, Canadian Award for Telecommunications Research from the Canadian Society of Information Theory (CSIT) in 2021, R.A. Fessenden Award in 2019 from IEEE, Canada, Award of Merit from the Federation of Chinese Canadian Professionals (Ontario) in 2019, James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, Joseph LoCicero Award in 2015 and Education Award in 2017 from the IEEE Communications Society (ComSoc), and Technical Recognition Award from Wireless Communications Technical Committee (2019) and AHSN Technical Committee (2013), Excellent Graduate Supervision Award in 2006 from the University of Waterloo, and the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada. He is/was the General Chair for the 6G Global Conference'23, and ACM Mobihoc'15, Technical Program Committee Chair/Co-Chair for IEEE Globecom'24, 16 and 07, IEEE Infocom'14, IEEE VTC'10 Fall, and the Chair for the IEEE ComSoc Technical Committee on Wireless Communications. He is also the past president of the IEEE ComSoc, vice president for Technical & Educational Activities, vice president for Publications, member at-large on the Board of Governors, chair of the Distinguished Lecturer Selection Committee, and member of IEEE Fellow Selection Committee of the ComSoc. He was the editor-in-chief of *IEEE IoT Journal*, *IEEE Network*, and *IET Communications*. He is a registered professional engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, Canadian Academy of Engineering Fellow, Royal Society of Canada Fellow, Chinese Academy of Engineering Foreign Member, and an International Fellow of the Engineering Academy of Japan.