electronics

MDPI

Article

Multi-Agent Deep Reinforcement Learning for Blockchain-Based Energy Trading in Decentralized Electric Vehicle Charger-Sharing Networks

Yinjie Han [1,*], Jingyi Meng [2] and Zihang Luo [3]

[1] International Business School, Xi'an Jiaotong-Liverpool University, Suzhou 215000, China
[2] Gies College of Business, University of Illinois Urbana-Champaign, Urbana, IL 61802, USA; jingyimeng0211@163.com
[3] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore; zihangluo123@163.com
* Correspondence: yinjiehan1218@gmail.com

Abstract: With The integration of renewable energy sources into smart grids and electric vehicle (EV) charger-sharing networks is essential for achieving the goal of environmental sustainability. However, the uneven distribution of distributed energy trading among EVs, fixed charging stations (FCSs), and mobile charging stations (MCSs) introduces challenges such as inadequate supply at FCSs and prolonged latencies at MCSs. In this paper, we propose a multi-agent deep reinforcement learning (MADRL)-based auction algorithm for energy trading that effectively balances charger supply with energy demand in distributed EV charging markets, while also reducing total charging latency. Specifically, this involves a MADRL-based hierarchical auction that dynamically adapts to real-time conditions, optimizing the balance of supply and demand. During energy trading, each EV, acting as a learning agent, can refine its bidding strategy to participate in various local energy trading markets, thus enhancing both individual utility and global social welfare. Furthermore, we design a cross-chain scheme to securely record and verify transaction results of energy trading in decentralized EV charger-sharing networks to ensure integrity and transparency. Finally, experimental results show that the proposed algorithm significantly outperforms both the second-price and double auctions in increasing global social welfare and reducing total charging latency.

Keywords: electric vehicles; mobile charging stations; charger sharing; blockchain; deep reinforcement learning

Citation: Han, Y.; Meng, J.; Luo, Z. Multi-Agent Deep Reinforcement Learning for Blockchain-Based Energy Trading in Decentralized Electric Vehicle Charger-Sharing Networks. *Electronics* **2024**, *13*, 4235. https://doi.org/10.3390/electronics13214235

Academic Editors: Tao Zhang, Xiangyun Tang, Jiacheng Wang and Jiqiang Liu

Received: 12 September 2024
Revised: 16 October 2024
Accepted: 17 October 2024
Published: 29 October 2024

1. Introduction

Electric vehicles (EVs), by utilizing electricity as their primary energy source within smart grids, play a pivotal role in reducing reliance on finite fossil fuels and enhancing environmental sustainability [1–5]. By 2035, it is projected that over 60% of all vehicles will be electrified, necessitating a significant expansion in both the number and capacity of EV charging infrastructures [6]. To meet the escalating energy demands of EVs, the development of an efficient and scalable charging infrastructure is imperative. While existing fixed charging stations (FCSs) provide reliable and accessible charging points in urban and suburban areas, the surge in EV adoption has placed these facilities under considerable strain, often leading to congestion and extended waiting periods for charging services. To mitigate this pressure, mobile charging stations (MCSs) have been introduced as a versatile addition to the smart grid infrastructure, offering adaptable and on-demand services, particularly in areas of high demand and remote locations. The integration of FCSs and MCSs fosters a more resilient and efficient charging ecosystem, capable of dynamically accommodating the varied demands of EVs [7–9]. However, the implementation of peer-to-peer energy trading between EVs and MCSs is poised to further alleviate system strain and enhance the overall energy efficiency of EV charger-sharing networks [10].

Despite the many benefits of this hybrid EV charging infrastructure, distributed EV charging still faces significant challenges, particularly in terms of provisioning enough incentives for FCSs and MCSs and reducing the charging latency of EVs [11]. Without well-structured incentives, FCSs and MCSs may hesitate to share their energy with EVs. In the meanwhile, the absence of methods to reduce charging latency might lead to operational inefficiencies and potential conflicts among EVs. Therefore, it is essential to ensure that FCSs and MCSs can receive adequate incentives to not only foster participation but also reduce the charging latency of EVs in the distributed EV charger-sharing networks.

The application of auction theory to the allocation of incentives in EV charger-sharing networks presents a range of unique challenges [10,12,13]. Although second-price auctions are commonly utilized to balance supply and demand, their effectiveness is often compromised in the EV charging context due to significant variability in service values among users. This variability can lead to suboptimal outcomes that fail to effectively match supply with demand. On the other hand, double auctions [14], which allow both buyers and sellers to submit bids according to their valuations, aim to maximize the income of MCSs. Yet, these auctions frequently face challenges in practical applications due to their inherent complexity and often do not achieve an ideal balance between maximizing income and ensuring fair service provision [15,16]. Consequently, it becomes imperative to develop a novel auction mechanism that addresses these issues, thereby optimizing the distributed energy trading in EV charger-sharing networks for better social welfare and operational efficacy.

In this paper, we propose a multi-agent reinforcement learning (MADRL)-based auction algorithm for distributed energy trading in EV charger-sharing networks that integrates both fixed and mobile charging stations. Utilizing MADRL [16], the proposed auction mechanism not only achieves a more efficient balance of supply and demand compared to second-price auctions but also excels at maximizing revenue for MCSs compared to existing double auctions. Specifically, this algorithm employs a multi-agent system where each agent (EV or MCS) dynamically adjusts its bids based on real-time market conditions and historical data, employing predictive analytics and adaptive learning to enhance decision-making. By integrating dynamic pricing strategies and adaptive incentive structures, our framework ensures an efficient and equitable allocation of charging resources. The EV charger-sharing market employs a cross-chain scheme that coordinates local and global markets through side and main blockchains, respectively. The process begins with local market energy trading confirmation on the side blockchain, followed by the aggregation of results into the global market on the main blockchain, utilizing a Two-Phase Commit protocol to ensure synchronization and atomicity in cross-chain transactions [17]. We validate the effectiveness and scalability of the proposed algorithm through simulations on real-world datasets, where the experimental results demonstrate significant performance improvements in global social welfare and total charging latency.

Our main contributions can be summarized as follows.

1. We propose a MADRL-based distributed energy trading algorithm for EV charger-sharing networks, where EVs can request charging services from either FCSs or MCSs. Specifically, the distributed energy trading in EV charger-sharing networks can enhance the scalability and adaptability of the charging infrastructure by allowing EVs to flexibly select their charging mode, ensuring access to necessary services under a wide range of scenarios.
2. We introduce a distributed energy trading structure that enables EVs to choose from various submarkets based on their immediate energy demands within their local area. This adaptable market design promotes efficient resource allocation and adheres to principles of individual rationality and incentive compatibility. To enhance economic efficiency, we have developed a multi-agent deep reinforcement learning (MADRL) algorithm to aid EVs and MCSs in making decisions about market participation based on real-time data and local market conditions.

3. To enhance the security and reliability of energy trading transaction recording, we propose a cross-chain scheme that leverages a Two-Phase Commit protocol-based approach. Specifically, all transaction data of each local market are first accurately and securely confirmed at the side blockchain and then aggregated into the main blockchain of the global market.
4. Experimental results confirm significant improvements in the efficiency and effectiveness of the proposed charging and trading system. Through controlled simulations and real-world tests, we have demonstrated the robustness of our auction framework under varied operational conditions, showing marked reductions in charging latency and the improvement of global social welfare.

The rest of the paper is organized as follows. A literature review identifies gaps in existing approaches in Section 2. The system model is formulated in Section 3. The MADRL-based auction is introduced in Section 4. In Section 5, experimental validations are presented. The paper is concluded in Section 6.

2. Related Work

2.1. Energy Trading for EVs in Smart Grids

With the widespread adoption of EVs, the role of smart grids in managing and optimizing charging infrastructure has become increasingly important [18]. Smart grids can effectively manage power by integrating Distributed Energy Resources (DERs) and renewable energy sources, such as solar and wind, to meet the significant power demands posed by EVs. For instance, Danial et al. [1] conduct a techno-economic analysis that emphasizes cost considerations that are crucial for promoting the widespread adoption of EVs. They highlight the economic viability threshold for EV charging stations and propose governmental interventions like subsidies and tax strategies to support the installation of an estimated 646 to 3300 charging stations in Brunei by 2035. Building on the framework of energy optimization, Kim et al. [19] design a system that enables EV owners to engage in energy trading, facilitated by an aggregator. They propose a non-cooperative game model for energy trading decisions and a coordinated charging/discharging algorithm to optimize cost efficiency at the aggregator level and minimize energy expenses for EV owners. Their experimental simulations utilize a double auction mechanism to explore market dynamics and price uncertainty, thereby enhancing the adaptability of the model to real-world conditions.

Furthermore, Alvaro-Hermana et al. [20] explore peer-to-peer energy trading among EVs to alleviate grid strain during peak tariff periods. They propose a system model employing quadratic programming to optimize energy trading and minimize grid impact during business hours. The uniqueness of their proposed solutions is validated through experiments that demonstrate the potential of peer-to-peer trading to mitigate the effects of EV charging on the grid during peak periods. Additionally, Aggarwal et al. [11] tackle the challenges of energy trading for EVs in smart grids, emphasizing the integration of blockchain technology to enhance the security and efficiency of peer-to-peer transactions. Their approach not only addresses the balancing of energy demands during peak hours but also enhances transaction security against cyber threats like false data injections. Experiments validate the effectiveness of their distributed secure energy trading scheme, showcasing its capability to facilitate reliable and secure energy exchanges within smart grids. Houda et at. [21] proposes a blockchain-based system for vehicle-to-vehicle (V2V) electricity trading using a reverse auction mechanism. This approach aims to create a decentralized, secure, and transparent electricity market among electric vehicles (EVs) within a smart grid environment. By utilizing Ethereum's smart contracts, the system facilitates automated and fair trading without the need for third-party intermediaries, thus eliminating single points of failure associated with centralized models. These studies collectively underscore the progressive efforts to integrate advanced technological solutions, including blockchain, game theory, and economic incentives to optimize energy trading and management within smart grid environments tailored for EVs.

2.2. Economic Optimization of EV Charging in Smart Grids

In smart grids, economic optimization is pivotal for effective EV charging management [22]. FCSs, while reliable, lack the flexibility required to adapt to rapidly changing power demands and price fluctuations. Recent studies have explored dynamic pricing and auction mechanisms to enhance the economics of EV charging [23]. Wang et al. [24] focus on the interactions of storage units within smart grids, particularly plug-in hybrid EVs and batteries, aimed at intelligent decision-making to maximize utility. They introduce a non-cooperative game model allowing storage units to decide strategically on energy sales, integrating dynamic pricing and double auction models to optimize trading benefits against costs. Simulation results reveal up to a 130.2% improvement in the average utility per storage unit, underscoring the model's effectiveness in enhancing trading efficiency. Building on these insights, Hou et al. [12] address EV charging scheduling in a distributed context to boost social welfare, aligning with user preferences and considering the state of charge. Their iterative bidding framework enables efficient negotiation on charging times and prices, with experiments illustrating an 85% efficiency relative to optimal solutions and highlighting the impact of information disclosure on scheduling efficiency.

Further extending the discussion on optimizing smart grid-based EV charging stations, Wang et al. [25] introduce a novel operation mechanism named JoAP. This mechanism optimizes EV admission control, pricing, and charging schedules to maximize profitability. Using a tandem queueing network model, their study analytically characterizes the influence of admission control and pricing policies on profits, with simulations showing a 531% increase in profitability compared to conventional methods. Kikusato et al. [13] propose a management scheme to optimize the use of photovoltaic power in EV charging, aiming to reduce operation costs and curtailment in low-voltage distribution systems. Their auction mechanism ensures voluntary participation, maintaining fairness and autonomy among customers. Simulations demonstrate effective cost reductions and curtailment mitigation, even under forecasting errors and unexpected disconnections. Finally, Kim et al. [10] develop an auction-based incentive mechanism for energy trading between EVs and MCSs, focusing on efficiently utilizing MCS energy resources. Their distributed auction-based mechanism ensures truthfulness, individual rationality, and budget balance. Simulation studies confirm the mechanism's efficiency, enhancing system performance by more than twice as much as existing approaches. They highlight the potential of innovative auction mechanisms and dynamic pricing to adapt to and optimize energy trading within evolving market conditions, thereby facilitating more efficient and flexible smart grid operations.

2.3. Deep Reinforcement Learning in EV Charging

DRL has emerged as a potent machine learning method, demonstrating significant potential in addressing complex decision-making problems, particularly in EV charging optimization within smart grids [22,26,27]. By engaging continuously with the environment and refining policies, DRL adapts to uncertainties and dynamic demands effectively. For instance, Zou et al. [28] tackle the unique challenges of EV charging in urban prosumer communities, characterized by varied energy generation patterns and fluctuating prices. They introduce an innovative DRL strategy, employing the Asynchronous Advantage Actor–Critic with Long Short-Term Memory (A3C-LSTM) model, a facet of multi-agent deep reinforcement learning aimed at optimizing energy purchasing decisions for EVs. Experimental results highlight significant improvements in charging rates and social welfare, surpassing traditional methods.

Further exploring the optimization of EV charging, Wang et al. [29] utilize a reinforcement learning (RL) approach to refine charging scheduling and pricing strategies at EV charging stations. Motivated by the unpredictability of EV arrivals and departures, their model-free RL algorithm focuses on maximizing profit through strategic adjustments to charging schedules and pricing. Experiments with real-world data demonstrate that their approach yields a 138.5% increase in profit over benchmark algorithms, showcasing the efficacy of RL in real-time operational environments. Fan et al. [15] leverage MADRL

to address supply and demand challenges in EV charging. They develop a distributed service allocation mechanism where Road-Side Units (RSUs) and Internet of Vehicles (IoV) [30] serve as markets for AI-generated content (AIGC). In this paper, we propose a MADRL-based distributed energy trading algorithm for EV charger-sharing networks that optimizes real-time EV charging decisions to improve global social welfare and reduce local charging latency.

3. System Model

3.1. Overview

As illustrated in Figure 1, the system aims to optimize EV charging management within smart grids by integrating three essential components: FCSs, MCSs, and EVs. Represented as $\mathcal{F} = \{1, \ldots, F\}$, the set of FCSs forms the backbone of the charging network, offering continuous and dependable charging services from their permanent locations. Each FCS f is equipped with substantial capacity to handle regular and predictable charging demands in the local market. Conversely, the set of MCSs under the coverage of FCS f is denoted as $\mathcal{M}^f = \{1, \ldots, M\}$. MCSs can relocate in response to real-time demand, effectively mitigating peak load pressures and extending services to remote areas that fixed stations cannot conveniently cover. This flexibility ensures that the system is responsive to varying charging requirements, providing timely support where necessary. EVs under the coverage of FCS f, represented by the set $\mathcal{E}^f = \{1, \ldots, E\}$, are the primary consumers within this ecosystem. Each EV requires strategic management to efficiently meet its charging demands while maximizing the utilization of both fixed and mobile charging resources. Our auction-based optimization framework orchestrates the interactions among f, $\mathcal{M}^f$, and $\mathcal{E}^f$, ensuring seamless coordination. Specifically, when an EV, denoted as e, acts as a buyer, it pays a price p_e^b that is equal to or lower than its valuation v_e, i.e., $p_e^b \leq v_e$. Conversely, when an MCS, denoted as m, acts as a seller, it receives a payment p_m^s that is equal to or higher than its cost c_m, i.e., $p_m^s \geq c_m$.

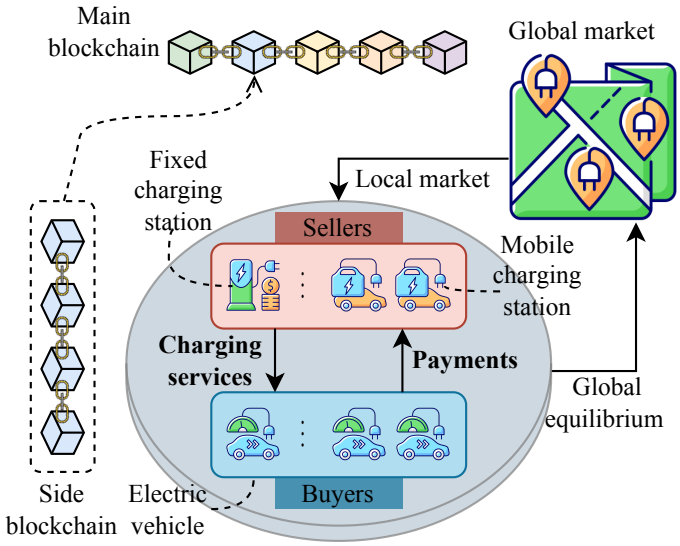


Figure 1. Blockchain-based energy trading in distributed EV charger-sharing networks, where each local market maintains its blockchain.

3.2. Charger Sharing Market

In this system, we consider a charger-sharing market with distributed energy trading where each FCS serves as an auctioneer for its local energy charging market. In each local market, MCSs act as sellers, and EVs act as buyers. In a local charger-sharing market, multiple bids from energy consumers (such as EVs) and providers (such as MCSs) are consolidated at the FCS for determining both pricing and allocation strategies.

Upon detecting an energy-sharing request from EVs, the system prompts MCSs in the local market, which may possess surplus charging capacity, to participate in energy trading. To engage in the market, both EVs and MCSs submit their bids to their FCSs. Before allocations and pricing can be determined, auctioneers at FCSs must compile detailed data from the local market to develop an exhaustive supply and demand matrix. The defined rules for allocation, termed as Π, include a supply matrix $X(t) \subseteq \{0,1\}^{|\mathcal{E}^f(t)|\times|\mathcal{M}^f(t)|} = \{x_1(t), \ldots, x_{|\mathcal{E}^f(t)|}(t)\}$ and a corresponding demand matrix $Y(t) \subseteq \{0,1\}^{|\mathcal{E}^f(t)|\times|\mathcal{M}^f(t)|} = \{y_1(t), \ldots, y_{|\mathcal{M}^f(t)|}(t)\}$. The supply vector of MCS m is depicted as $x_m \subseteq \{x_{m,1}(t), \ldots, x_{m,|\mathcal{E}^f(t)|}(t)\}$, and for EV e, the demand vector is shown as $y_e \subseteq \{y_{e,1}(t), \ldots, y_{e,|\mathcal{M}^f(t)|}(t)\}$. Then, each FCS calculates allocations and prices according to the auction protocol $M = (\Pi, \Psi)$ for its local charger-sharing market, where $\Pi = (X, Y)$ indicates the allocation rules and $\Psi = (p_b, p_s)$ describes the pricing rules. In this model, a trustworthy bid $b_{t,e}$ from buyer e represents their true valuation v_e, ensuring that its payment $p_e^b(t)$ remains within its budget, i.e., $p_e^b(t) \leq v_e$. Similarly, a seller m offers a selling bid $a_{t,m}$ reflecting its real cost c_m, ensuring its earning $p_m^s(t)$ meets or exceeds these costs, i.e., $p_m^s(t) \geq c_m$. These methods guarantee the pricing framework, defined by the vectors $p_b(t)$ and $p_s(t)$ for buyers and sellers, respectively.

3.3. Problem Formulation

Meanwhile, the mechanism needs to allocate and price energy to meet the demand and supply conditions and at the same time minimize the total charging time of both buyers and sellers under imperfect information conditions that any buyer or seller will act truthfully under individual rationality (IR). It is strategically rational for each participant, whether as a buyer or as a seller, to gain something out of the auction and therefore must prevent a utility that is not less than zero from engaging in the auction. Truthfulness in auctions can be referred to as a notion that reflects the fact that in an auction, buyers and sellers tender bids that are real regarding their appraisal or cost. This characteristic makes the auctions accurate and efficient, since all those interested can give an estimation of the value of the good or service that is being auctioned. In the marketplace, each buyer's bid $b_e(t)$ aligns precisely with their actual valuation v_e, ensuring $b_e(t) = v_e$. Consequently, the auction design ensures that the amount paid by buyers, $p_e^b(t)$, remains within their evaluated value, namely $p_e^b(t) \leq v_e$. Conversely, each seller's offer $a_m(t)$ mirrors the actual expenses incurred c_m, establishing $a_m(t) = c_m$. This arrangement ensures that the income received by sellers, $p_m^s(t)$, matches or exceeds their expenses, i.e., $p_m^s(t) \geq c_m$. In this bidirectional market, to preserve integrity, the buyer's valuation must match or exceed what they eventually pay, and similarly, the seller's listed prices reflect or exceed their perceived value of the goods.

Furthermore, the total societal benefit, labeled as $SW(t)$ at time t, is the aggregate of values accumulated from both sellers and buyers involved, which can be calculated as

$$SW(t) = \sum_{e\in\mathcal{E}^f(t)} \sum_{m\in\mathcal{M}^f(t)} y_{e,m}(t) v_e(t) + \sum_{e\in\mathcal{E}^f(t)} \sum_{m\in\mathcal{M}^f(t)} x_{m,e}(t) c_m(t). \tag{1}$$

Furthermore, the total charging latency at time slot t is denoted as $L(t)$, which can be computed as

$$\begin{aligned} L(t) = \sum_{e\in\mathcal{E}^f(t)} \Bigg[\Bigg(1 - \sum_{m\in\mathcal{M}^f(t)} y_{e,m}(t)\Bigg) \Big(T_e^{\text{charge}} + T_e^{\text{travel}}(t) + T_f^{\text{wait}}(t)\Big) \\ + \sum_{m\in\mathcal{M}^f(t)} y_{e,m}(t) \Big(T_e^{\text{charge}} + T_m^{\text{travel}}(t)\Big) \Bigg], \end{aligned} \tag{2}$$

where T^{charge_e} is the charging time of EV e, T_e^{travel} is the traveling time of EV e, and T_f^{wait} is the waiting time at FCS f. This formula accounts for the total charging time, which includes the energy transfer time and the travel time to MCSs for winning EVs in the auction. For losing EVs, the charging time involves energy transfer, travel to FCSs, and additional waiting times at these stations, thus highlighting the impact of auction outcomes on charging efficiency.

4. Multi-Agent Deep-Reinforcement-Learning-Based Energy Trading Algorithm

4.1. Distributed Incentive Mechanism

In this subsection, we discuss the details of distributed energy trading in terms of the following points: Each MCS creates a local market in the region they cover; this includes all EVs within that region. As illustrated in Figure 2, there are two submarkets in each local market, i.e., the single-side urgent submarket and the double-side mundane submarket. In the urgent submarket, there is transacting immediately into clearing since bids for buying and selling are matched directly. On the other hand, in the mundane submarket, the transactions require procedures to be cleared at certain points to enable those with non-urgent energy requirements to participate in the market at certain times. At time t, EV e places a bid $b_e(t)$ to acquire energy, and MCS m offers a bid $a_m(t)$ catering to these energy requests.

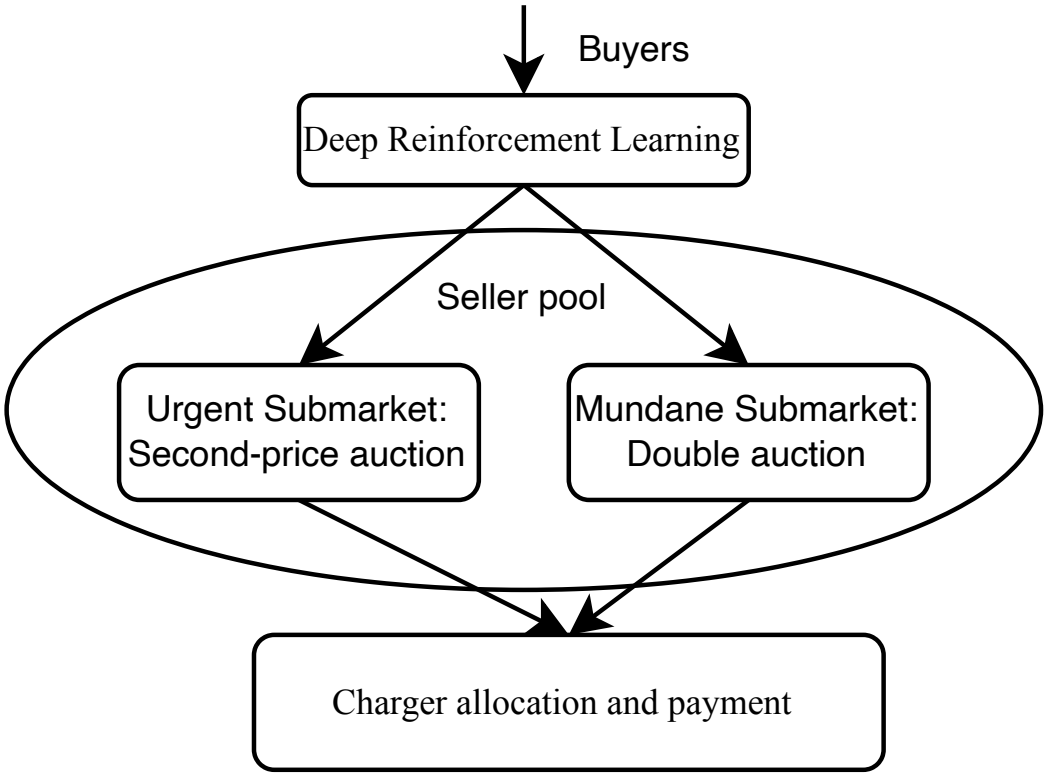


Figure 2. The workflow of the MADRL-based energy trading algorithm.

Within the coverage of FCS f, sellers are arranged into a local seller pool $\mathcal{P}_f^S(t) = \{m \mid \sum_{e \in \mathcal{E}^f(t)} x_{m,e}(t) = 0, m \in \mathcal{M}^f(t)\}$, where they react to incoming requests from buyers. The bids submitted by these sellers are accumulated in a designated value pool $\mathcal{A}_f^{\mathcal{S}}(t) = \{a_m(t) \mid \sum_{e \in \mathcal{E}^f(t)} x_{m,e}(t) = 0, m \in \mathcal{M}^f(t)\}$ and are arranged to ensure $a_m(t) \leq a_{m+1}(t), \forall m, m+1 \in \mathcal{P}_f^{\mathcal{S}}(t)$. Concurrently, EVs engaging in the routine submarket are sorted into a buyers' pool $\mathcal{P}_f^B(t)$, where their bids are organized in $\mathcal{O}_f^B(t)$, ranked such that $b_e(t) \geq b_{e+1}(t), \forall e, e+1 \in \mathcal{P}_f^B(t)$. For the urgent submarket on the buyers'

side, a second-price auction protocol is utilized, whereas the routine submarket adopts McAfee's method.

In the single-sided urgent submarket: The auctioneer evaluates the seller pool and adjusts the supply and demand dynamics along with the monetary stipulations of MCS m. The logic of the selection of the highest bid from the given number of qualified sellers for EV e is established by setting up

$$x_{m,e}(t) = y_{e,m}(t) = 1(a_m(t) > \max\{\mathcal{A}^{\mathcal{S}}_{f,-m}(t)\}), \tag{3}$$

where $1(\cdot)$ is the indicator function and $\mathcal{A}^{\mathcal{S}}_{f,-m}(t)$ represents the highest bid in the seller pool excluding MCS m. A second-price sealed-bid auction determines the transaction price for buying EV e in the urgent market:

$$p^b_e(t) = \sum_{m \in \mathcal{P}^S_f(t)} x_{m,e}(t) \cdot \max\{\mathcal{A}^{\mathcal{S}}_{f,-m}(t)\}, \tag{4}$$

and the revenue for the selling MCS m is

$$p^s_m(t) = y_{e,m}(t) \cdot \max\{\mathcal{A}^{\mathcal{S}}_f(t)\}. \tag{5}$$

Double-side mundane submarket: This submarket comprises the buyer pool $\mathcal{P}^B_f(t)$ and untraded sellers in the seller pool $\mathcal{P}^S_f(t)$, which are cleared periodically. Using McAfee's mechanism, the auctioneer sorts buyers and sellers and determines the allocation and pricing rules by finding the breakeven index k in $\mathcal{A}^B_f(t)$ and $\mathcal{O}^B_f(t)$. The average price is calculated as

$$p^b_e(t) = p^s_m(t) = \frac{b_{k+1} + s_{k+1}}{2}, \tag{6}$$

ensuring the first k buyers and sellers trade at this price, with adjustments made for trade reduction if necessary.

After all local markets have cleared, the local auctioneers of MCSs calculate the local budget cost $\beta_f(t)$ for their market, and the aggregate budget cost across the globally distributed market is computed as

$$\beta(t) = \sum_{f \in \mathcal{F}} \beta_f(t) = \sum_{f \in \mathcal{F}} \left[\sum_{e \in \mathcal{E}^f} |p^b_e(t) - \bar{p}(t)| + \sum_{m \in \mathcal{M}^f} |p^s_m(t) - \bar{p}(t)| \right], \tag{7}$$

where the market clearning price is $\bar{p}(t) = \frac{\sum_{e \in \mathcal{E}^f(t)} p^b_e(t) + \sum_{m \in \mathcal{M}^f(t)} p^s_m(t)}{\sum_{m \in \mathcal{M}^f(t)} \sum_{e \in \mathcal{E}^f(t)} x_{m,e}(t) + \sum_{m \in \mathcal{M}^f(t)} \sum_{e \in \mathcal{E}^f(t)} y_{e,m}(t)}$.

To minimize the total budget cost while optimizing social welfare, the buying of EVs acts as a learning agent, adopting strategies that ensure equilibrium between supply and demand across the submarkets.

4.2. Partially Observable Markov Decision Process

Within this framework, every purchaser functions as a cognitive entity governed by a Partially Observable Markov Decision Process (POMDP), which comprises the components outlined below:

1. Observation: The observation of EV e at each time step t, denoted as $O_e(t)$, encompasses several key market dynamics. These include the number of buyers and sellers in each local market, represented as $|\mathcal{E}^f(t)|, \forall f \in \mathcal{F}$ and $|\mathcal{M}^f(t)|, \forall f \in \mathcal{F}$, respectively; the price of the last transaction denoted by $\bar{p}(t-1)$; and the charging latency, traveling time, and waiting time of EV e, represented by T^{charge}_e, $T^{\text{travel}}_e(t)$, $T^{\text{wait}}_n(t)$, respectively. The comprehensive observation set is formalized as

$$O_e(t) = \{|\mathcal{E}^1(t)|, \ldots, |\mathcal{E}^f(t)|, |\mathcal{M}^1(t)|, \ldots, |\mathcal{M}^f(t)|, \bar{p}(t-1), T^{\text{charge}}_e, T^{\text{travel}}_e(t), T^{\text{wait}}_f(t)\}. \tag{8}$$

2. Action: The action $A_e(t)$ for EV e at time slot t is defined as the market selection strategy, where $A_e(t) = 0, 1$. Here, $A_e(t) = 0$ implies participation in the urgent submarket, while $A_e(t) = 1$ indicates entry into the mundane submarket within the buyer pool.
3. Reward: The reward function, $R_e(O_e(t), A_e(t))$, incorporates the social welfare, budget, and total charging time at the current time slot, calculated as

$$R_e(O_e(t), A_e(t)) = SW(t) - \alpha\beta(t)^2 - L(t), \tag{9}$$

where α is a coefficient that scales the budget cost. Social welfare, denoted by $SW(t)$, reflects the total utility of all buyers and sellers based on transaction prices; the budget, represented by $\beta(t)$, measures the net fiscal impact of transactions; and the total charging time, $L(t)$, accounts for delays due to energy transfer and travel times.
4. Value Function: The value function $V_{\pi_e}(O_e(t))$ for the policy π_e of EV e is the expected return starting from state S and following policy π_e, which is defined as

$$V_{\pi_e}(S) := E_\pi[\sum_{t=0}^{T} \gamma^k R_e(O_e(t), A_e(t))|S_0 = S], \tag{10}$$

where $E_\pi(\cdot)$ represents the expected value under policy π, and $\gamma \in [0, 1]$ is the discount factor that progressively reduces the weight of future rewards.

4.3. Policy Iteration

The POMDP framework offers a comprehensive mathematical model to enhance decision-making for buyers. By evaluating system outcomes and executing suitable actions, agents can maximize their long-term expected rewards. This model facilitates the development of learning agents engaged in P2P energy trading, optimizing their utility while improving overall system welfare. To refine the valuation function, the Multi-agent Proximal Policy Optimization (MAPPO) [31], an algorithm based on reinforcement learning, facilitates the training of numerous agents within settings where rewards serve as feedback. Within the POMDP framework, each agent assesses the current state of the system and strategizes to enhance rewards over an extended period.

Define θ_e as the settings for the policy network specific to EV e and ϕ_e as those for the value network. Within MAPPO, every agent adheres to a policy $\pi_e(\theta_e)$, undergoing periodic adjustments via a centralized critic $V(s; \phi_e)$, which appraises the overarching state S. Both the critic and policy frameworks undergo simultaneous training with a clipped loss metric, expressed as $L_{\text{CLIP}}(\theta_e, \phi_e) = L_P(\theta_e) + L_V(\phi_e)$, integrating losses from both networks to reflect their collective efficacy in the global scenario. This methodology supports effective learning and adaptation within a multi-agent environment, ensuring that each agent's actions contribute optimally to the collective goals of the energy trading system. Algorithm 1 provides a summary of the proposed MADRL algorithm.

The complexity of the proposed MADRL-based energy trading algorithm increases with the number of agents, as each agent (EV or MCS) operates within a dynamic state-action space, which grows exponentially with the number of participants. This scaling challenge impacts the learning algorithm, as more agents require additional computational resources to update policies and value functions through deep learning iterations. Communication overheads also rise with the number of agents, particularly in decentralized systems where data exchange for coordination becomes necessary across both local and global markets. Auction mechanisms further introduce complexity, particularly in double auction scenarios where both buyer and seller bids need to be evaluated. Despite these challenges, the system is designed to maintain scalability through decentralized learning and localized markets, which effectively mitigate complexity, making the algorithm feasible for large-scale deployment. However, the careful management of computational resources and communication protocols will be critical to ensure efficiency as the number of agents increases.

Algorithm 1: Multi-agent deep reinforcement learning-based auction algorithm

Initialize policy networks $\pi_e(\theta_e)$ and value networks and $V(\phi_e)$ for each EV e;
for *episode in range(max_episodes)* **do**
  Initialize state $O_e(t)$;
  for *t in range(max_timesteps)* **do**
    for *each EV e in EV_set* **do**
      Observe the current state for $O_e(t)$;
      Select an action $A_e(t)$ based on its policy network $\pi_e(O_e(t); \theta_e)$;
    end
    Execute actions and observe the next state $O_e(t+1)$ and reward $R_e(t)$;
    for *each EV e in* $\mathcal{E}^f$ **do**
      Store the experience tuple $(O_e(t), A_e(t), R_e(t), O_e(t+1))$ in replay buffer $\mathcal{B}$;
      Sample a mini-batch of experiences from replay buffers;
      Update the value network $V(\phi_e)$ using the sampled mini-batch;
      for *Each* $(O_e(t), A_e(t), R_e(t), O_e(t+1))$ *in minibatch* **do**
        Value_loss = $(V(O_e(t); \phi_e) - \text{target})^2$;
        Update ϕ_e using gradient descent on value_loss;
      end
    end
    Update the policy network $\pi_e(\theta_e)$ using the sampled mini-batch;
    for *each experience in minibatch* **do**
      $(O_e(t), A_e(t), R_e(t), O_e(t+1)) =$ experience;
      Advantage = $R_e(t) + \gamma \cdot V(O_e(t+1); \phi_e) - V(O_e(t); \phi_e)$;
      Policy_loss = $-\log(\pi_e(A_e(t); \theta_e)) \cdot$ advantage;
      Update θ_e using gradient descent on policy_loss;
    end
  end
  Update the centralized critic $V(s; \phi_e)$ using global state $O_e(t)$;
  for *each global state* $O_e(t)$ *in minibatch* **do**
    Global_target = $R(t) + \gamma \cdot V(O_e(t+1); \phi_e)$;
    Critic_loss = $(V(O_e(t); \phi_e) - \text{global\_target})^2$;
    Update ϕ_e using gradient descent on critic_loss;
  end
end
Store results and statistics for the episode.

4.4. Property Analysis

Utilizing the principles of second-price and McAfee's double auctions, we demonstrate that our innovative distributed hierarchical auction structure promotes both IR and truthfulness across a distributed market setting. It is essential initially to verify that this framework supports IR and truthfulness consistently in local markets.

Lemma 1. *With the established market entry strategies $A_e, \forall e \in \mathcal{E}^f(t)$, this system sustains IR and truthfulness locally.*

This principle is derived from the fundamental characteristics of second-price auctions and the dominant strategies used in double auctions. Expanding from this foundation, the local maintenance of IR and truthfulness is projected onto the broader, globally distributed market.

Theorem 1. *Across a diverse network of local markets that make up the global market, our framework reliably preserves individual rationality and truthfulness.*

Proof. Assessing individual rationality and truthfulness involves examining both urgent and routine submarkets, following the set market entry strategies. Specifically, an EV e with $A_e(t) = 0$ consistently opts for urgent submarkets, whereas one with $A_e(t) = 1$ chooses routine submarkets. □

We then demonstrate that our auction model guarantees IR for all participants. In urgent submarkets facilitated by a second-price auction, the top bidder pays only the second-highest bid, not exceeding their actual valuation. Consequently, the net benefit for winners remains positive, as shown by

$$\mu_e(t) = v_e(t) - p_e^b(t) \geq 0, \tag{11}$$

where $\sum_{m \in \mathcal{M}^f(t)} x_{m,e} = 1$. For those not winning, their utility remains zero as no payment is made:

$$\mu_e(t) = v_e(t) - p_e^b(t) = 0. \tag{12}$$

In the mundane submarkets, McAfee's mechanism ensures that winning buyers pay the average of the lowest winning bid and the highest losing bid, which is also less than or equal to their valuation. Hence, the utility for these buyers remains non-negative. For buyers not succeeding, and for sellers, both losing and winning, the mechanism ensures their utility is aligned with their actions, maintaining non-negative utility for winners and zero utility for non-participants.

Next, we establish that the mechanism is truthful for all involved. In the urgent submarket's second price auction, the optimal strategy for buyers is to bid their true valuation since the payment is the second-highest bid, not their own. In the mundane submarket, McAfee's mechanism incentivizes both buyers and sellers to bid truthfully, making truth-telling a dominant strategy. In conclusion, the proposed distributed hierarchical auction mechanism consistently upholds individual rationality and truthfulness across a globally distributed market comprising multiple local markets. This ensures efficient and equitable distributed energy sharing within EV charging networks, thereby benefiting all participants in their involvement.

5. Cross-Chain Scheme for Decentralized EV Charger Sharing

In EV charger-sharing networks, the cross-chain scheme for energy trading employs a structured workflow that coordinates local and global markets through side and main blockchains, respectively. This technical process can be broken down as follows.

5.1. Local Market Energy Trading Confirmation in Side Blockchain

The energy trading process begins at the local market, where each side blockchain, which manages local market transactions among charging stations and electric vehicles, ensuring decentralized verification of energy trades, is maintained by FCS, MCSs, and EVs.

1. Trading Request and Auction Mechanism: In the energy trading process of the local market, EVs submit charging requests based on their energy needs, while MCSs offer available energy for sale through the proposed MADRL-based auction mechanism. The FCS acts as the auctioneer, managing the auction and matching EVs' demands with MCSs' supplies to determine charger allocation and payment.
2. Recording in the Side Blockchain: After the auction concludes, energy trading results, including successful bids, prices, and energy quantities, are confirmed and recorded in the side blockchain. Therefore, the side blockchain manages local market transactions among FCSs, MCSs, and EVs, ensuring decentralized verification of energy trades and transparency and reliability in the trading process.

5.2. Aggregation into Global Market in Main Blockchain

The next phase aggregates results from local markets into the global market through the main blockchain, which aggregates confirmed energy trading results from side blockchains to

maintain data integrity with FCSs serving as recording nodes, which can ensure consistency and universal coordination between local and global energy demands and supplies.

1. Cross-Chain Interaction: The cross-chain interaction between the side blockchains of local markets and the main blockchain of global market involves a two-step process. First, the side blockchain communicates with the main blockchain through cross-chain interaction protocols. Then, confirmed energy trading results from the side blockchain are aggregated and submitted to the main blockchain, with FCSs acting as gateways between the two blockchains. This process ensures the seamless integration of local market data into the global energy market while maintaining data integrity and security.
2. Main Blockchain Record Update: On the main blockchain, representing the global energy market, FCSs act as validators to record aggregated results from various local markets. These results update the global market's ledger, reflecting the total energy exchanged, pricing trends, and participation of MCSs and EVs across all local markets.
3. Final Confirmation: Validators on the main blockchain confirm the legitimacy of aggregated results. Once confirmed, the global ledger is updated, ensuring consolidation and accurate reflection of all energy trading data across multiple local markets.

5.3. Cross-Chain Scheme with Two-Phase Commit Protocol

The cross-chain interaction uses the Two-Phase Commit (2PC) protocol, which ensures synchronization in cross-chain transactions through a prepare phase for result collection and a commit phase for result aggregation, to ensure synchronization and atomicity in cross-chain transactions:

1. Phase 1 (Prepare): Collators in the side blockchain collect local market results (energy trading outcomes) and submit them to main blockchain validators. Validators check result validity through lightweight verification, using simplified payment verification to ensure legitimate and accurate transaction data.
2. Phase 2 (Commit): After verification, the main blockchain validators approve the transaction. The cross-chain event is recorded, and the results are aggregated into the main blockchain, completing the global energy market update.

This cross-chain process guarantees decentralized, tamper-resistant verification of energy trading, ensuring seamless collaboration between local and global markets without compromising blockchain data integrity or security.

6. Numerical Results

6.1. Experimental Settings

The experimental environment is set up with 50 rounds and includes four FCSs, each representing a submarket. The total number of vehicles ranges from 20 to 80, and each FCS has a coverage area of 500 m. In the experiment, we utilize the data proposed in [32], which consist of 3395 high-resolution charging sessions involving 85 EV drivers across 105 stations at 25 workplace locations, meticulously recorded to analyze the effects of pricing strategies and workplace norms on optimizing shared EV charging resources. As shown in Figure 3, the EV charging data statistics for the experiments are provided as follows: the total kWh required has a mean of 5.81, a standard deviation of 2.89, and a range from 0 to 23.68, with a median of 6.23 and a peak range between 6 and 7. The charge time has a mean of 2.84 h, a standard deviation of 1.51 h, and a range from 0.0125 to 55.24 h, with a median of 2.81 h and a peak range between 2 and 3.5 h. The distance traveled has a mean of 18.65 km, a standard deviation of 11.42 km, and a range from 0.86 to 43.06 km, with a median of 21.02 km. We implemented a prototype system of our cross-chain-empowered charger-sharing market using Hyperledger Fabric for both the side and the main blockchains, with Chaincode smart contracts in Go, DPoS consensus, and Hyperledger Cactus for cross-chain interactions. The algorithm settings involve several parameters crucial for the MADRL approach. The replay buffer size is set to 10,000, with a learning rate of 1×10^{-4} and a discount factor (γ) of 0.95. The training is set to 1000 epochs, with each epoch consisting of 2000 steps. During each epoch, 16 episodes are collected, and each episode is repeated four times.

The batch size for training is set to 64, and the hidden layers in the neural network have sizes of 256 and 256. The training involves eight parallel environments, while testing uses a single environment. Specific parameters for PPO include a value function coefficient of 0.5, an entropy coefficient of 0.05, an epsilon clip of 0.2, a maximum gradient norm of 0.5, and a Generalized Advantage Estimation (GAE) lambda of 0.95. Additionally, reward normalization is enabled, and several clipping and normalization parameters are set to ensure stable training. The environment settings include 20 users and 4 servers, simulating the interactions within the distributed markets.

The distribution of valuations for buyers and sellers is depicted in Figure 4. The left histogram shows the distribution of valuations for buyers, ranging from approximately 0.6 to 1.4. The distribution is centered around 1.2, indicating that the most frequent valuations are close to this value. The frequency peaks at around 1.2, with the highest frequency just above 140. The distribution appears to be approximately normal, with fewer buyers having valuations significantly lower or higher than 1.2. The right histogram shows the distribution of costs for sellers, ranging from 0 to 4. The distribution is centered around 2, indicating that the most frequent costs are close to this value. The frequency peaks at around 2, with the highest frequency just above 80. The distribution appears normal, with fewer sellers having costs significantly lower or higher than 2. Overall, both distributions exhibit a bell-shaped curve, suggesting a normal distribution with a central tendency around 1.2 for buyers' valuations and 2 for sellers' costs.

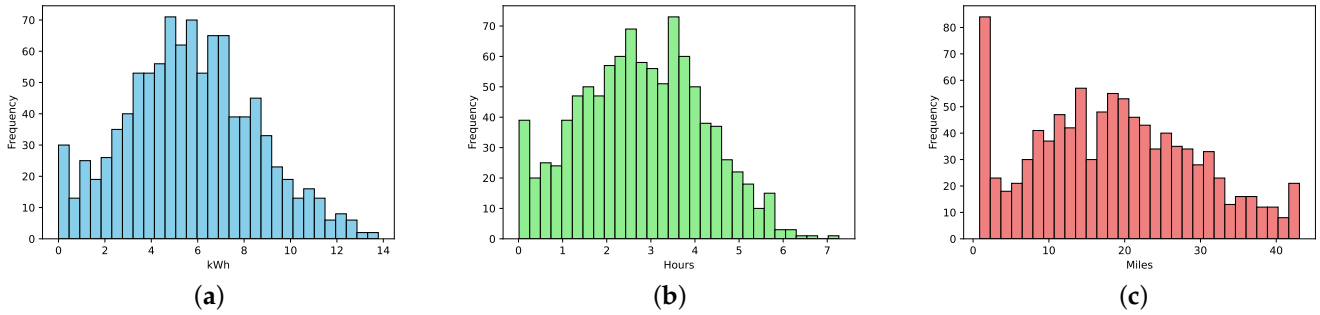


Figure 3. The data distribution of energy demands, charging time, and distance in the dataset. (**a**) Energy demands. (**b**) Charging time. (**c**) Distance.

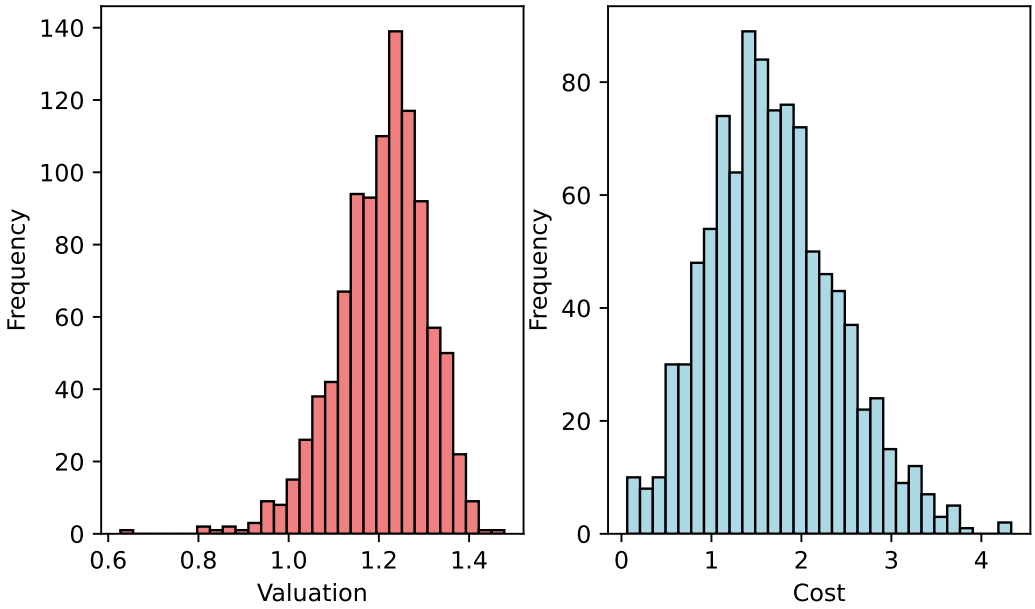
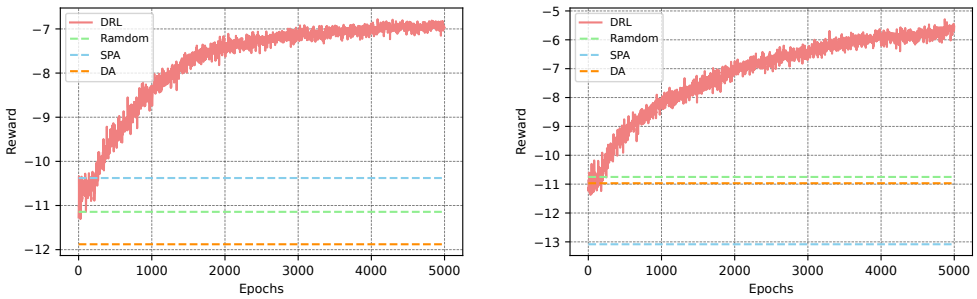


Figure 4. The distribution of valuation of buyers and sellers.

6.2. Convergence Analysis

The convergence performance of the proposed MADRL-based mechanism is analyzed under different market sizes: 20, 40, 60, and 80. Figure 5 illustrates four subplots comparing the performance of the DRL-based mechanism with Random, Single Price Auction (SPA), and Double Auction (DA) mechanisms across 5000 epochs. For a market size of 20, the DRL mechanism exhibits a steady increase in rewards, initiating at approximately −11 and converging to around −7. It significantly outperforms the Random, SPA, and DA mechanisms, which display relatively static reward trends at lower levels. The Random mechanism stabilizes around −11, SPA around −10, and DA around −12. In the scenario with a market size of 40, the DRL mechanism demonstrates consistent improvement, initiating at approximately −12 and converging to around −5. It markedly surpasses the other mechanisms. Random mechanism stabilizes around −11, SPA around −9, and DA around −12. With a market size of 60, the DRL mechanism begins with an initial dip followed by a steady increase, starting from around −12 and converging to approximately −4. The reward trajectory of the DRL mechanism surpasses those of the Random, SPA, and DA mechanisms. The Random mechanism stabilizes around −11, SPA around −9, and DA around −12. For the largest tested market size of 80, the DRL mechanism displays rapid initial improvement, beginning from around −12 and converging to approximately −4 within 2500 epochs. It consistently outshines the other mechanisms throughout the training process. The Random mechanism stabilizes around −11, SPA around −9, and DA around −12. In summary, across all market sizes, the DRL-based mechanism demonstrates superior convergence performance compared to the Random, SPA, and DA mechanisms. It consistently achieves higher rewards, indicating enhanced optimization and learning efficacy. The DRL mechanism exhibits a clear upward trend, steadily improving over time and converging to higher reward levels, while the other mechanisms maintain relatively constant and lower reward levels throughout the epochs. This underscores the effectiveness of the DRL-based approach in adapting to and optimizing the dynamics of EV charging market scenarios. MADRL shows a clear upward trend, steadily improving over time and converging to higher reward levels compared to SPA and DA. The consistent increase suggests that MADRL effectively learns over time, optimizing both the buyers' and sellers' utilities. The rewards for MADRL start relatively low, indicating a learning curve where the system first gathers data, adjusts bidding strategies, and progressively improves its decision-making. This is expected for reinforcement learning algorithms that require sufficient exploration of actions before optimizing for rewards. Meanwhile, the rapid initial improvements for market sizes of 40 and 80 suggest that the system's complexity and available data are more suited for larger market scenarios. This scalability is a crucial feature of MADRL, ensuring that the algorithm can perform effectively as the number of vehicles increases.



(**a**) (**b**)

Figure 5. *Cont.*

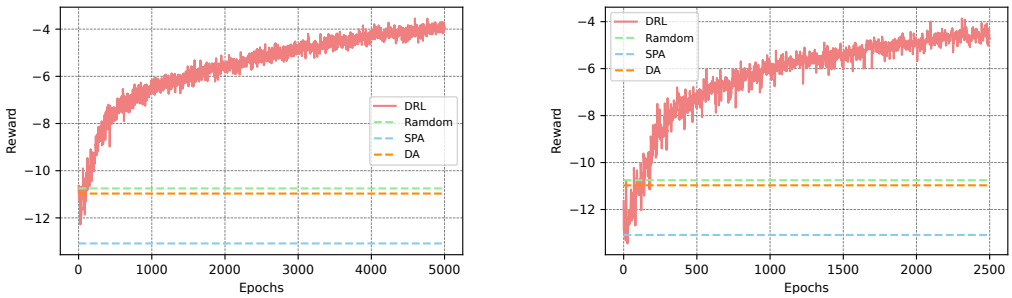


Figure 5. Convergence analysis of the proposed learning-based mechanism under different sizes of EV charging markets. (**a**) Market size = 20. (**b**) Market size = 40. (**c**) Market size = 60. (**d**) Market size = 80.

6.3. Performance Comparison

Figure 6 presents a performance comparison of the proposed DRL-based mechanism with the Random, SPA, and DA mechanisms across various sizes of EV charging markets (number of vehicles). The DRL mechanism consistently achieves the highest rewards across all market sizes, maintaining a stable reward close to -5 as the number of vehicles increases from 20 to 80. In contrast, the Random and SPA mechanisms exhibit relatively constant rewards around -10 and -11, respectively, while the DA mechanism shows a significant decrease in reward as the market size increases, dropping from around -10 to nearly -40. In terms of social welfare, the DRL, SPA, and DA mechanisms display a similar increasing trend as the number of vehicles increases. The DRL mechanism slightly outperforms the other mechanisms, particularly in larger market sizes, achieving social welfare close to 55 with 80 vehicles. The Random mechanism, however, lags, demonstrating lower social welfare across all market sizes. The DRL mechanism also maintains a relatively low and stable budget cost across all market sizes, approximately 10. The Random and SPA mechanisms show low budget costs that increase slightly with market size but remain below 20. Conversely, the DA mechanism experiences a substantial increase in budget cost as the number of vehicles increases, reaching nearly 50 with 80 vehicles. Finally, the charging latency performance of the DRL mechanism demonstrates competitiveness, increasing from about 20 to 55 as the number of vehicles grows. The SPA mechanism exhibits a similar trend but with slightly lower latency compared to DRL. The Random mechanism consistently displays higher latency than both DRL and SPA, while the DA mechanism shows the highest charging latency across all market sizes, peaking at 60 with 80 vehicles. Across all market sizes (20 to 80 vehicles), MADRL consistently achieves the highest rewards, maintaining stability near -5. The other mechanisms either stay at lower reward levels (e.g., around -10 for SPA) or decline significantly (e.g., DA's reward dropping to -40 as market size increases). This stability in rewards highlights MADRL's ability to maintain an optimal balance of supply and demand over time, even in larger markets. MADRL's ability to handle increasing market sizes while maintaining low latency, high social welfare, and low budget cost underscores its potential for scalability, making it a promising solution for the future of EV energy trading. Further analysis could focus on how MADRL adapts to even larger market sizes, different types of auction designs, or integration with renewable energy sources to improve overall system sustainability.

During the experimentation, the complexity of implementing a MADRL framework and cross-chain interactions using Hyperledger Fabric could imply potential computational challenges.

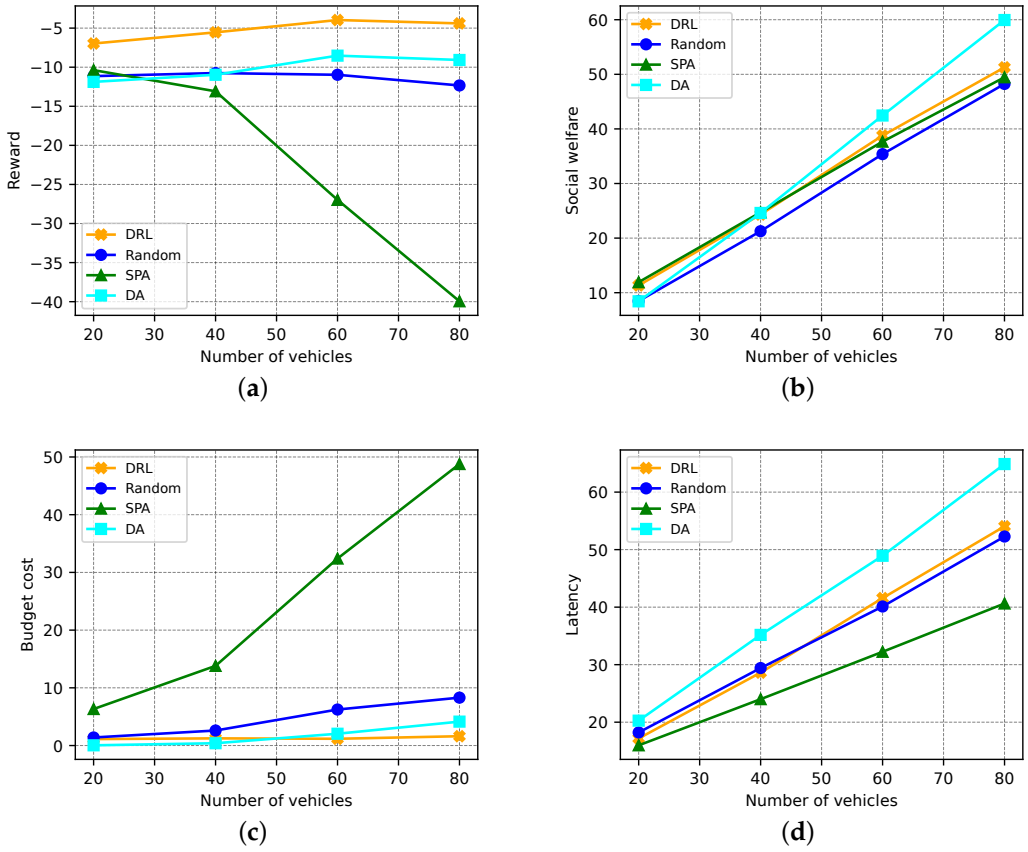


Figure 6. Performance comparison of the proposed learning-based mechanism under different sizes of EV charging markets. (**a**) Reward. (**b**) Social welfare. (**c**) Budget cost. (**d**) Charging latency.

7. Conclusions

In this paper, we introduce a MADRL-based auction algorithm for distributed energy trading in EV charger-sharing networks for achieving environmental sustainability. The proposed algorithm significantly enhances the scalability and efficiency of EV charging infrastructures by dynamically balancing supply and demand across various market sizes, demonstrating robust adaptability and superior performance in extensive simulations. By reducing charging latency and improving global social welfare, the proposed algorithm offers a sustainable solution that is adaptable to smart grid technologies globally. The proposed blockchain framework enables decentralized, secure, and scalable coordination for energy trading in EV charger-sharing markets through cross-chain collaboration, ensuring efficient resource management and transparent transactions between local and global markets. In future work, we will aim to benchmark our approach against a broader range of models, potentially including other machine learning techniques, to further validate its effectiveness and scalability. Moreover, we plan to perform real-world testing to validate the proposed algorithm in the future.

Author Contributions: Conceptualization, Y.H.; methodology, Y.H.; software, J.M.; validation, Y.H., J.M. and Z.L.; formal analysis, Y.H.; investigation, Y.H.; resources, Z.L.; data curation, Z.L.; writing—original draft preparation, Y.H.; writing—review and editing, Y.H.; visualization, J.M.; supervision, Y.H., J.M. and Z.L.; project administration, Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Danial, M.; Azis, F.A.; Abas, P.E. Techno-economic analysis and feasibility studies of electric vehicle charging station. *World Electr. Veh. J.* **2021**, *12*, 264. [CrossRef]
2. Ma, J.; Zhang, Y.; Duan, Z.; Tang, L. PROLIFIC: Deep Reinforcement Learning for Efficient EV Fleet Scheduling and Charging. *Sustainability* **2023**, *15*, 13553. [CrossRef]
3. Shi, C.; Yu, M. Flexible solid-state lithium-sulfur batteries based on structural designs. *Energy Storage Mater.* **2023**, *57*, 429–459. [CrossRef]
4. Wang, J.; Du, H.; Niyato, D.; Zhou, M.; Kang, J.; Poor, H.V. Acceleration estimation of signal propagation path length changes for wireless sensing. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 11476–11492. [CrossRef]
5. Wang, J.; Du, H.; Liu, Y.; Sun, G.; Niyato, D.; Mao, S.; Kim, D.I.; Shen, X. Generative AI based Secure Wireless Sensing for ISAC Networks. *arXiv* **2024**, arXiv:2408.11398.
6. Lai, P.; Fan, R.; Zhang, X.; Zhang, W.; Liu, F.; Zhou, J.T. Utility optimal thread assignment and resource allocation in multi-server systems. *IEEE/ACM Trans. Netw.* **2021**, *30*, 735–748. [CrossRef]
7. Wang, L.; Hou, L.; Liu, S.; Han, Z.; Wu, J. Reinforcement contract design for vehicular-edge computing scheduling and energy trading via deep Q-network with hybrid action space. *IEEE Trans. Mob. Comput.* **2023**, *23*, 6770–6784. [CrossRef]
8. Zhang, T.; Xu, C.; Lian, Y.; Tian, H.; Kang, J.; Kuang, X.; Niyato, D. When moving target defense meets attack prediction in digital twins: A convolutional and hierarchical reinforcement learning approach. *IEEE J. Sel. Areas Commun.* **2023**, *41*, 3293–3305. [CrossRef]
9. Zhang, T.; Xu, C.; Shen, J.; Kuang, X.; Grieco, L.A. How to disturb network reconnaissance: A moving target defense approach based on deep reinforcement learning. *IEEE Trans. Inf. Forensics Secur.* **2023**, *18*, 5735–5748. [CrossRef]
10. Kim, O.T.T.; Le, T.H.T.; Shin, M.J.; Nguyen, V.; Han, Z.; Hong, C.S. Distributed auction-based incentive mechanism for energy trading between electric vehicles and mobile charging stations. *IEEE Access* **2022**, *10*, 56331–56347. [CrossRef]
11. Aggarwal, S.; Kumar, N. Pets: P2p energy trading scheduling scheme for electric vehicles in smart grid systems. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 14361–14374. [CrossRef]
12. Hou, L.; Wang, C.; Yan, J. Bidding for preferred timing: An auction design for electric vehicle charging station scheduling. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3332–3343. [CrossRef]
13. Kikusato, H.; Fujimoto, Y.; Hanada, S.I.; Isogawa, D.; Yoshizawa, S.; Ohashi, H.; Hayashi, Y. Electric vehicle charging management using auction mechanism for reducing PV curtailment in distribution systems. *IEEE Trans. Sustain. Energy* **2019**, *11*, 1394–1403. [CrossRef]
14. Gao, J.; Wong, T.; Wang, C.; Yu, J.Y. A price-based iterative double auction for charger sharing markets. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 5116–5127. [CrossRef]
15. Fan, J.; Xu, M.; Liu, Z.; Ye, H.; Gu, C.; Niyato, D.; Lam, K.Y. A Learning-based Incentive Mechanism for Mobile AIGC Service in Decentralized Internet of Vehicles. In Proceedings of the 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall), Hong Kong, China, 10–13 October 2023; pp. 1–5.
16. Fan, J.; Xu, M.; Guo, J.; Shar, L.K.; Kang, J.; Niyato, D.; Lam, K.Y. Decentralized Multimedia Data Sharing in IoV: A Learning-based Equilibrium of Supply and Demand. *IEEE Trans. Veh. Technol.* **2023**, *73*, 4035–4050. [CrossRef]
17. Kang, J.; Li, X.; Nie, J.; Liu, Y.; Xu, M.; Xiong, Z.; Niyato, D.; Yan, Q. Communication-efficient and cross-chain empowered federated learning for artificial intelligence of things. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 2966–2977. [CrossRef]
18. Aggarwal, S.; Kumar, N.; Tanwar, S.; Alazab, M. A survey on energy trading in the smart grid: Taxonomy, research challenges and solutions. *IEEE Access* **2021**, *9*, 116231–116253. [CrossRef]
19. Kim, B.G.; Ren, S.; Van Der Schaar, M.; Lee, J.W. Bidirectional energy trading and residential load scheduling with electric vehicles in the smart grid. *IEEE J. Sel. Areas Commun.* **2013**, *31*, 1219–1234. [CrossRef]
20. Alvaro-Hermana, R.; Fraile-Ardanuy, J.; Zufiria, P.J.; Knapen, L.; Janssens, D. Peer to peer energy trading with electric vehicles. *IEEE Intell. Transp. Syst. Mag.* **2016**, *8*, 33–44. [CrossRef]
21. Abou El Houda, Z.; Hafid, A.S.; Khoukhi, L. Blockchain-based reverse auction for v2v charging in smart grid environment. In Proceedings of the ICC 2021-IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021; pp. 1–6.
22. Osaba, E.; Villar-Rodriguez, E.; Del Ser, J.; Nebro, A.J.; Molina, D.; LaTorre, A.; Suganthan, P.N.; Coello, C.A.C.; Herrera, F. A tutorial on the design, experimentation and application of metaheuristic algorithms to real-world optimization problems. *Swarm Evol. Comput.* **2021**, *64*, 100888. [CrossRef]
23. Wang, K.; Zhang, X.; Duan, L.; Tie, J. Multi-UAV cooperative trajectory for servicing dynamic demands and charging battery. *IEEE Trans. Mob. Comput.* **2021**, *22*, 1599–1614. [CrossRef]
24. Wang, Y.; Saad, W.; Han, Z.; Poor, H.V.; Başar, T. A game-theoretic approach to energy trading in the smart grid. *IEEE Trans. Smart Grid* **2014**, *5*, 1439–1450. [CrossRef]
25. Wang, S.; Bi, S.; Zhang, Y.J.A.; Huang, J. Electrical vehicle charging station profit maximization: Admission, pricing, and online scheduling. *IEEE Trans. Sustain. Energy* **2018**, *9*, 1722–1731. [CrossRef]
26. Zhou, Y.; Yang, Z.; Zhang, X.; Wang, Y. A hybrid attention-based deep neural network for simultaneous multi-sensor pruning and human activity recognition. *IEEE Internet Things J.* **2022**, *9*, 25363–25372. [CrossRef]

27. Xie, Y.; Chan, T.T.; Zhang, X.; Lai, P.; Pan, H. Reflection-Optimized Covert Communication for Jammer-Aided Ambient Backscatter Systems. In Proceedings of the GLOBECOM 2023-2023 IEEE Global Communications Conference, Kuala Lumpur, Malaysia, 4–8 December 2023; pp. 4277–4282.
28. Zou, L.; Munir, M.S.; Tun, Y.K.; Kang, S.; Hong, C.S. Intelligent EV charging for urban prosumer communities: An auction and multi-agent deep reinforcement learning approach. *IEEE Trans. Netw. Serv. Manag.* **2022**, *19*, 4384–4407. [CrossRef]
29. Wang, S.; Bi, S.; Zhang, Y.A. Reinforcement learning for real-time pricing and scheduling control in EV charging stations. *IEEE Trans. Ind. Inform.* **2019**, *17*, 849–859. [CrossRef]
30. Zhang, T.; Xu, C.; Zou, P.; Tian, H.; Kuang, X.; Yang, S.; Zhong, L.; Niyato, D. How to mitigate DDoS intelligently in SD-IoV: A moving target defense approach. *IEEE Trans. Ind. Inform.* **2022**, *19*, 1097–1106. [CrossRef]
31. Zhang, T.; Xu, C.; Zhang, B.; Li, X.; Kuang, X.; Grieco, L.A. Towards attack-resistant service function chain migration: A model-based adaptive proximal policy optimization approach. *IEEE Trans. Dependable Secur. Comput.* **2023**, *20*, 4913–4927. [CrossRef]
32. Asensio, O.I.; Apablaza, C.Z.; Lawson, M.C.; Walsh, S.E. A field experiment on workplace norms and electric vehicle charging etiquette. *J. Ind. Ecol.* **2022**, *26*, 183–196. [CrossRef]