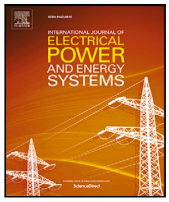
Electrical Power and Energy Systems 160 (2024) 110092

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes



Non-cooperative coupled microgrid-transportation coordination system: A deep deterministic policy gradient-based approach

Shiyao Zhang [a], Xingzheng Zhu [a], Shengyu Zhang [b], James J.Q. Yu [c,*]

[a] Research Institute for Trustworthy Autonomous Systems, Southern University of Science and Technology, Shenzhen, 518055, China
[b] Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore, 487372, Singapore
[c] Department of Computer Science, University of York, Heslington, YO10 5GH, United Kingdom

ARTICLE INFO

Keywords:
Coupled microgrid-transportation system
Connected electric vehicles
Deep deterministic policy gradient
Markov decision process
Microgrid
Non-cooperative system

ABSTRACT

The cooperation between multiple microgrids (MGs) take a substantial role in enabling the process of information and energy exchange, which can increase the reliability of multi-microgrid (MMG) systems. Since MGs tend to be selfish in real-world operations, it is vital to address the way to motivate MGs to cooperate efficiently in MMG systems. In addition, considering the participated connected electric vehicles (CEVs) in CMT systems, the synergy between the MG-transportation coupled networks shall be accounted. In this paper, we propose a Stackelberg game theoretic approach for the non-cooperative coupled microgrid-transportation (CMT) system by exploiting connected electric vehicles (CEVs). This formulated game can jointly trigger the MG energy trading and vehicle-to-grid coordination schemes in the system. To solve the CMT game with imperfect information, a Markov Decision Process (MDP) and a deep deterministic policy gradient-based methods are developed. The devised algorithm can learn a deterministic optimal response from the opponents. The game's Nash equilibrium can be approximately converged when the game is with imperfect information. Simulation results show that the proposed model can guarantee the effectiveness of the energy trading process in the CMT system. In addition, through the efficient CEV coordination strategies, the V2G regulation service is provided efficiently while it can further alleviate the grid power fluctuations. Furthermore, large economic benefits can be achieved under the CMT system.

1. Introduction

With the advanced development of modern power systems, microgrid (MG), known as a self-contained power system, is essential for managing power and load balance, which can further facilitate the sources pooling during grid islanding. In general, for power networks, one MG includes different types of loads, distributed generators (DGs), and energy storage technologies (ESSs). For transportation network, one MG owns various connected electric vehicles (CEVs) that can perform autonomous driving operations through vehicle-to-everything (V2X) techniques [1]. However, there are difficulties to coordinate these factors in one MG. On the one hand, the increasing penetration of renewable energy sources and can potentially make a huge impact on modern power systems due to their intermittency and uncertainty characteristics. Such renewables can result in inevitable random and uncorrelated power fluctuations in MG systems. On the other hand, the rapid application of CEVs can affect each MG system due to their stochastic charging behaviors. To tackle this issue, the conversion of multiple MGs into the conventional power system network becomes a promising solution for the development of future sustainable power systems [2]. Since one MG is an individual self-controlled system, the multi-microgrid (MMG) system can indeed contribute to the economical benefits and energy utilization through the interaction with other MGs [3].

Through information and resources sharing, an MMG system connected with multiple individual MGs can facilitate the energy and relevant data transfer. For instance, [4,5] have studied the feasibility and practicability of such the MG techniques. Efficient MG collaboration can increase the system's reliability when considering the utilization of renewable energy sources and CEVs in power systems, as well as lowering the power losses of end-users. Several pieces of existing research work have studied the benefit of the efficient cooperation in an MMG system, such as [6–10]. In [6], a bi-level framework was established to consider the interaction between the distribution network operator and the MMGs, which aimed to schedule power generations to compensate the load volatility and exchange power in both day-ahead and real-time ways. Furthermore, a probabilistic

* Corresponding author.
E-mail address: jqyu@ieee.org (J.J.Q. Yu).

https://doi.org/10.1016/j.ijepes.2024.110092
Received 8 October 2023; Received in revised form 28 April 2024; Accepted 13 June 2024

model was proposed in [7] to assess the controllability of both network voltages and currents in the MMG systems. Besides, considering the uncertainties of renewable energy in power systems, an interval optimization scheduling method was developed for coordinated MMG systems to study the uncertainties of photovoltaic and wind power [8]. Furthermore, to alleviate the adverse effects of uncertain renewable power generations, a distributed robust model predictive control in [9] was proposed for islanded MMGs. Specifically, this method succeeds to balance the robustness and efficiency of the single MG system operation while dealing with the unpredictability of renewable energy sources. Lastly, with the consideration of electric vehicles (EVs), [10] studied multi-period optimal energy trading and scheduling for MMG systems integrated with a double-ring urban transportation network. This work not only alleviate the system operation cost but also solve traffic congestion issues. Nonetheless, these results developed based on the grid optimization and optimistically assume that the system operator directly controls each individual MG and its attached EV fleet. In practice, each individual MG generally more concerns about their utilities rather than the overall system operations, i.e., being selfish. Moreover, since CEV charging patterns are considered in each MG, the coupled microgrid-transportation (CMT) system is modeled in a sufficient manner rather than the modeling of the traditional MMG system. Hence, how to motivate the MGs with the attached CEVs to participate in the efficient and cooperative operations of such the CMT system is still a pressing challenge.

To address the aforementioned issue, game theoretical approaches can provide an effective framework for analyzing the relationship between the individual utility and the system goals. Specifically, each decision-maker develops a unique plan that optimizes its own utility rather than following the instructions from a centralized controller. Since each individual MG can select the best strategy to maximize its utility, these methods naturally ensure the satisfaction of MG utilities in the system that motivate the participation of MGs. Several existing studies investigated the MG trading schemes under MMG systems, e.g. [11–13]. For instance, an optimal energy trading approach was proposed in [11] to lower the risk of overbidding for the energy trading of renewable resources while maximizing each individual MG's revenue. Additionally, a priority-based energy scheduling algorithm in [12] was devised in a distributed power network so as to assigning the buyer and seller roles in the system via the mismatch between energy supply and demand. In [14], a non-cooperative game model was developed for efficient energy trading scheme in MMG system. Besides, [13] proposed an MG energy trading Bayesian game model to tackle bidirectional energy trading among the stochastic EVs. Here, the installed battery packs can be regarded as a mobile power storage in power systems [15,16]. In addition, the quality of the power ancillary services can be further improved upon the availability of CEV coordination, e.g. [14,17–20]. However, these games already assumed perfect information of all participated MGs, whereas each individual MG knows the other MGs' utility functions or has the ability to exchange the strategy.

Since there are various privacy concerns in real-world operations, MGs that may be owned by different companies are reluctant to share privacy to other MGs. As a result, the game turns into one with incomplete information. For instance, games in [11–13] cannot find their Nash equilibrium under the imperfect information case. Thus, the utilization of reinforcement learning (RL) techniques can help solve the games with imperfect information. In [21], an intelligent MMG energy management technique was proposed to optimize the selling energy profit and to reduce the ratio of demand-side peak-to-average by means of neural network and model-free RL methods. However, this work solely considers the implementation of renewables and it does not account the effect of CEV charging behaviors in each MG coupled with the transportation network. The provisioning of effective CEV charging services can not only fulfill the charging requirements but also further help stabilize the urban power systems, i.e. [22,23]. Generally, the consideration of CEV charging operations in each MG inevitably increases the complexity of the entire system model since the synergy between the MG-transportation coupled networks are involved. Since CEVs can operate both the wired and wireless charging/discharging operations through the vehicle-to-grid (V2G) technique, the stability condition of the MG power network shall be further considered. These research gaps motivate us to address the imperfect information and the effect of CEV wired and wireless V2G coordination in the CMT system with game theoretical approaches.

In this paper, we develop a non-cooperative game theoretic approach for the CMT system by exploiting CEVs with the consideration of the MG imperfect information. The formulated game jointly provides MG energy trading and V2G charging services in the system. In particular, the game is reformulated as a Markov Decision Process (MDP) and a deep deterministic policy gradient (DDPG)-based method is developed to solve the game with imperfect information. The devised algorithm can learn a deterministic optimal response from the opponents. The game's Nash equilibrium can be approximately converged when the game is with imperfect. The main contributions of this work are summarized as follows:

- Different from the existing approaches, e.g. [6–10], that solely consider the efficient MG collaboration operation, we design a novel non-cooperative Stackelberg CMT game framework to motivate each individual MG and its attached CEV fleet to jointly participate in MG energy trading and V2G coordination operations. In addition, we study the effect of both the energy trading and wired/wireless V2G scheduling schemes by reacting to the decision price in the CMT system.
- Unlike the existing studies [11–13] that have assumed the perfect information of the MGs during the MG trading schemes, we devise a DDPG-based method to solve the Stackelberg CMT game with imperfect information. In this case, no MG is able to share its strategy with other MGs or access any of their private information.
- We discuss the merits and effectiveness of the proposed DDPG-based approach. We find that through a comprehensive series of case studies, both CMT energy trading and CEV charging services can be achieved in an effective manner.

The rest of this paper is organized as follows. In Section 2, we develop the system model, including CMT system, CEVs, and system operations. In Section 3, we formulate the proposed non-cooperative CMT game with well-defined constraints and utility functions. Then, Section 4 devises a DDPG-based approach to solve the game. Section 5 details the case studies that evaluate the proposed DDPG-based approach and Section 6 concludes this work.

2. System model

As discussed in Section 1, the purpose of the proposed CMT system is to address the open problem of joint energy trading and V2G coordination in the power grid. This section details our system model of this problem.

2.1. Multi-microgrid system

In this work, we develop the system operation model for the CMT system shown in Fig. 1. The operational time period is denoted as $\mathcal{T}$, which can be divided into a set $\mathcal{T} = \{1, \ldots, |\mathcal{T}|\}$ of $|\mathcal{T}|$ time slots. We define the set of MGs as $\mathcal{M} = \{1, \ldots, M\}$, where MG $i \in \mathcal{M}$ is operated within the CMT system. Through power and communication connections, each MG can link to some of the other MGs. In the meantime, let $\mathcal{A} = \{A_{i,j}\}$ be the adjacency matrix. In this case, if $A_{i,j} = 1$ holds, MG i is connected with MG j. Otherwise, $A_{i,j} = 0$. Meanwhile, the diagonal element is set as $A_{i,i} = 0$. In the system, all MGs are connected into the power grid with the operation of grid-connected

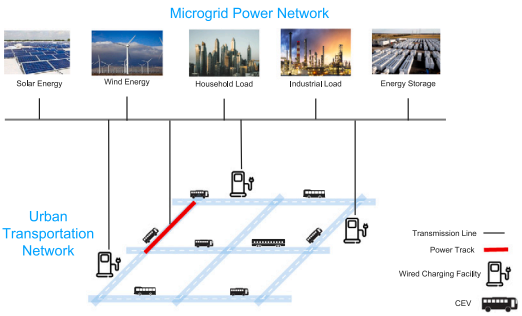


Fig. 1. Proposed coupled MG-transportation system overview.

modes. Each MG includes various components, namely, distributed generators, loads, energy storages, and its control center. Specifically, the DGs are partially related to renewable energy generations, i.e. solar panels and wind turbines. The system includes both the household and industrial loads. In this work, we consider the utilization of CEV batteries as the distributed energy storages in the system.

2.2. Transportation network

Suppose $\mathcal{N}_i$ be the index set of CEVs at MG i. By traveling between any two locations (s, v), where $s, v \in \mathcal{V}_i$, the CEVs achieve their assigned traveling objectives. To accurately depict the physical distance and linking from one location to another within a given area, the road network of each MG i can be described as a directed graph $G(\mathcal{V}_i, \mathcal{E}_i)$, where the set of nodes $\mathcal{V}_i$ represents the junction locations of the road segments, which are linked by the edges in the set $\mathcal{E}_i$. Then, we further define $d_{sv,n}$ as the travel distance of AV n from location s to location v. Due to potential asymmetries between distinct route plans, it should be noted that $d_{sv,n,t}$ may not always equal $d_{vs,n,t}$. We use Dijkstra's algorithm to get the vehicle routing plans in order to determine the shortest route between the departure point and the destination point.

2.3. Wired and wireless charging facilities

In the CMT system in Fig. 1, we denote $\mathcal{K}^{cs}$ as the set of charging stations that are installed with stationary charging plugs, which is also the set of CEV aggregators (CEVAs). Note that, for the wired V2G coordination, we monitor the available parking capacity for CEVA $k \in \mathcal{K}_i^{cs}$ at each MG i in real-time. For each CEVA k, it can assign a fleet of CEVs in MG i to park and to provide charging/discharging services. By following [24], the fast discharging and charging limits are assumed to be ± 70 kW, respectively.

Besides, we consider that the certain number of power tracks (PTs) are embedded over the specific urban transportation network of each MG for CEVs to wirelessly charge and discharge, whereas PTs are defined as the energy supply units of wireless power transfer-based electromagnetic induction roadway powering systems [25]. Some road segments of the road network in MG i are placed certain amount of PTs, which allows CEVs to wirelessly charge and discharge. We denote $\mathcal{K}_i^{pt}$ as the set of PTs in each MG i. In addition, the related wireless charging standard is assumed to be maximum 7.7 kW by following SAE J2954 standard in [26].

2.4. System operations

In Fig. 2, The bi-level system architecture includes two crucial parts, namely, the utility grid and MGs. The MGs have local renewable energy (RE) generation from solar or wind resources in addition to being connected to the utility grid through a point of common connection (PCC). The MGs can operate in both grid-connected and islanded modes by regulating PCCs. In this work, we assume that every MG operates in the grid-connected mode. Each individual operation is illustrated in the following:

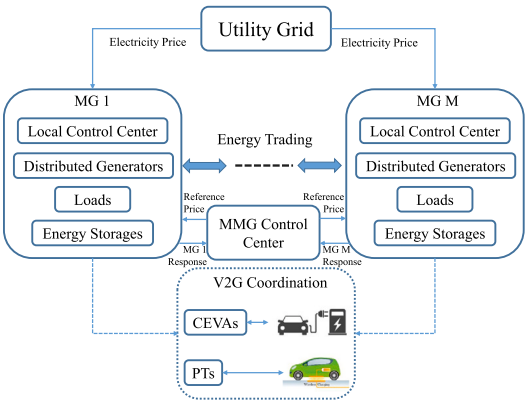


Fig. 2. Proposed CMT system model.

- **Energy trading**: The local control center for each MG aims to coordinate the internal operations, including CEV charging and discharging behaviors. In addition, it also involves the energy exchange with other MGs that are connected to this MG and the purchase of electricity from the utility grid. If extra renewable energy is generated at MG i, it can sell the excess to other MGs that are connected to it. On the other hand, when MG i does not have sufficient energy, it can purchase energy from others that are connected nearby. Meanwhile, the utility grid acts as an intermediary agent between energy buyers and sellers. For ease of representation, we denote MG to MG (M2M) energy trading as such the type of operational process.
- **V2G coordination**: V2G coordination refers to an efficient approach of utilizing CEV batteries to restore electricity back to the power system. In particular, they can be regarded as distributed energy storages that follow the power control signals from the energy market to provide a variety of auxiliary services, e.g. frequency regulation service [27]. The active power imbalance in the power system can be alleviated by the provisioning of V2G regulation services, thereby keeping the grid frequency at its nominal value.

The proposed CMT system is required to encourage and control the process for M2M energy trading. Each MG are installed with multiple smart meters to track the trading activities. The entire CMT system is responsible for identifying and storing all transaction information during the trading activities, which includes the buyer's and seller's identities, the trading energy volume, the trading time, and the unit energy price. Furthermore, the system aims to increase the utilization of the renewable energy when the energy trading price is given. However, the selected price might not ensure the M2M energy market's equilibrium, which equalizes power supply and demand. As a result, each individual MG is required to keep the market balance during the entire time period.

Different from the CMT system in [14], we further consider the coupling effect between each MG and its transportation network. In this case, the CEV coordination becomes a key factor to trigger the efficient energy management approaches, e.g. the V2G coordination scenes between CEVs and each local control center. Specifically, the process includes both wired and wireless charging or discharging of CEVs. For the wired V2G coordination, each CEV is assigned to be parked into the CEV charging stations, e.g. real-world parking lots, via the first-come-first-served strategy. These CEVs have to park in a stationary manner until their charging requirements are reached. On the other hand, for the wireless V2G coordination, it mainly involves the charging mechanisms to enable CEVs to charge while they are in motion over PTs. The CMT system broadcasts the real-time power imbalance and fixed electricity price to each individual MG. Then,

each MG shall determines its actions, including the amount of energy trading, the amount of utilized renewable energy, and the strategy to coordinate CEVs schedules, so as to maximize their rewards, known as their individual social welfare. Meanwhile, the electricity price is jointly affected by the actual power imbalance volume and the actions of these MGs in real-time. Hence, the actions of all MGs affect the reward. Such interactions between the system operator and multiple MGs are modeled as an CMT game with imperfect information, whereas neither one MG is aware of the private information of the other MGs nor does it share its strategy with others.

3. CMT game formulation

In this section, we present the system framework for the proposed Stackelberg game framework, which consists of two stages, namely, CMT operation and V2G coordination with detailed mathematical representations shown as follows.

3.1. Multi-microgrid constraints

Based on the system design, the CMT system functions during the whole time period $\mathcal{T} = \{1, \dots, T\}$, which can be divided into $|\mathcal{T}|$ time slots. In the CMT system, we consider both power generations and load demands. The power generation and load of MG i at time t are denoted as $G_{i,t}$ and $L_{i,t}$, respectively. We denote $R_{i,t}$ as the utilized renewable energy volume of MG i at time t. Since some portion of the generated renewable energy of each MG is utilized to serve the load demand or sell to other connected MGs, the constraint of the utilized renewable energy shall follow

$$R_{i,t} \le G_{i,t}, \ \forall i \in \mathcal{M}, t \in \mathcal{T}. \tag{1}$$

In this case, the unutilized renewable energy is discarded.

Then, considering M2M energy trading process in CMT system, the trading price is given for each MG to determine the amounts of buying energy and selling energy from other MGs. We define $\alpha_{ij,t}$ as the buying energy volume from MG i at MG j. Meanwhile, $\boldsymbol{\alpha}_{j,t} = \{\alpha_{ij,t} | \forall i, j \in \mathcal{M}, i \neq j\}$ stands for the MG i's buying energy vector, which denotes the total quantity of buying energy from other MGs at MG i. Besides, we denote $S_{i,t}$ as the timely amount of selling energy to the other MGs at MG i. Given the price by the system operator in the CMT operation, the connected MGs perform scheduling scheme locally, and subsequently reply their individual commitment to the system operator.

In addition, the energy transmission loss of the proposed CMT system is represented by $\boldsymbol{\gamma}^T = \{\gamma_{ij,t} | \forall i, j \in \mathcal{M}, i \neq j\}$, where γ_{ij} represents the amount of energy loss during power transmission between MGs i and j. In this case, when MG j purchases a specific quantity of energy $\alpha_{ij,t}$ from MG i, the energy obtained by MG j is

$$\beta_{ij,t} = \alpha_{ij,t}(1 - \gamma_{ij}), \ \forall i, j \in \mathcal{M}, t \in \mathcal{T}. \tag{2}$$

Besides, the operational bound of the amount of energy transmitted between MGs i and j when considering the capacity constraint of the power transmission lines in the CMT system follows

$$0 \le \alpha_{ij,t} \le \overline{\alpha}_{ij}, \ \forall i, j \in \mathcal{M}, t \in \mathcal{T}, \tag{3}$$

where $\overline{\alpha}_{ij}$ denotes the maximum energy transmitted among MGs. Then, the operational limit between the utility grid and the connected MGs follows

$$0 \le E^g_{i,t} \le \overline{E}^g_i, \ \forall i, j \in \mathcal{M}, t \in \mathcal{T}, \tag{4}$$

where $\overline{E}^g_i$ denotes the maximum energy received from the utility grid of MG i.

Furthermore, taking the stability of the CMT system into consideration, we need to implement the power balance constraint. Since the energy supply in MG i should exceed the grid demand due to the energy generated from renewable sources and the total amount of energy purchased from both the utility grid and other MGs, we have

$$G_{i,t} + E^g_{i,t} + \sum_{j \neq i} \beta_{ji,t} \ge C_{i,t} + L_{i,t} + S_{i,t}, \quad \forall i, j \in \mathcal{M}, t \in \mathcal{T}, \tag{5}$$

where $C_{i,t}$ represents the amount of charging/discharging energy of MG i at time t, including two types of energy storage systems, namely, stationary energy storage devices and CEVs. In this paper, since we consider the effect of the CEV coordination in the CMT system, we have

$$C_{i,t} = \sum_{n_i \in \mathcal{N}} P_{n_i,t}, \ \forall i \in \mathcal{M}, t \in \mathcal{T}, \tag{6}$$

where $P_{n_i,t}$ is the charging/discharging power of CEV n traveling inside MG i at time t.

Moreover, based on constraints (1) and (5), the power balance constraint in each MG i is represented by

$$R_{i,t} + E^g_{i,t} + \sum_{j \neq i} \beta_{ji,t} = C_{i,t} + L_{i,t} + S_{i,t}, \ \forall i, j \in \mathcal{M}, t \in \mathcal{T}. \tag{7}$$

Since the total amount of energy purchased by other MGs from MG i must equal the amount of energy sold in MG i, it holds

$$\sum_{j \neq i} \alpha_{ij,t} = S_{i,t}, \ \forall i, j \in \mathcal{M}, t \in \mathcal{T}. \tag{8}$$

Only the excess renewable energy produced by each MG is sold to other MGs. The amount of sold energy in this instance is constrained by

$$\begin{cases} S_{i,t} = 0 & \text{if } G_{i,t} \le R_{i,t}, \\ 0 \le S_{i,t} \le G_{i,t} - R_{i,t} & \text{if } G_{i,t} \ge R_{i,t}. \end{cases} \tag{9}$$

3.2. Vehicle-to-grid scheduling constraints

Next, we tackle both the wired and wireless CEV charging scheduling schemes for those are willing to participate in the proposed system framework. First, we introduce the CEV dynamic wireless scheduling scheme. Within the time period $\mathcal{T}$, CEVs $\mathcal{N}_i \subset \mathcal{N}$ at each MG $i \in \mathcal{M}$ can perform dynamic wireless charging/discharging. In addition, the stability condition of the MG must be fulfilled by providing V2G regulation services. The mathematical representations of CEV dynamic V2G scheduling scheme are shown below. To begin with, the followings are the operation limits for CEV n's charging and discharging powers

$$P^{\text{dch}}_n \le P_{n,t} k_{n,t} \le P^{\text{ch}}_n, \ \forall n \in \mathcal{N}, t \in \mathcal{T}, \tag{10}$$

where $P^{\text{dch}}_n \le 0$ and $P^{\text{ch}}_n \ge 0$ are the discharging and charging power of CEV n, respectively. If CEV n performs charging/discharging at MG i, we can further denote its charging power as $P_{n_i,t}$. $k_{n,t}$ presents the binary control variable to indicate whether CEV n is traveling over the PTs or not. In this case, if CEVs do not participate the dynamic wireless scheduling within the specific time period $\mathcal{T}^s \subset \mathcal{T}$, $P_{n,t} = 0, \forall t \in \mathcal{T}^s$ holds. This further reflects that such CEVs solely travel on roads without dynamic wireless charging or discharging.

The CEV n is then connected to the PTs and begins its charging/discharging process. Each participating CEV's state-of-charge (SOC) at time t can be calculated as follows:

$$\text{SOC}_{n,t+\Delta t} = \text{SOC}_{n,t} + \frac{\Delta t}{C^{\text{ev}}_n} \eta\left(P_{n,t}\right) P_{n,t}, \ t \in \mathcal{T}, \tag{11}$$

where CEV n's installed battery capacity is indicated by $C^{\text{ev}}_n > 0$. Additionally, the energy efficiency of each CEV's charging and discharging power, $\eta(\cdot)$, is determined by

$$\eta(P_{n,t}) = \begin{cases} \eta^{\text{ch}} & P_{n,t} \ge 0, \\ 1/\eta^{\text{dch}} & P_{n,t} < 0, \end{cases} \tag{12}$$

where the charging and discharge efficiencies of each participating CEV n are represented by $\eta^{\text{ch}}, \eta^{\text{dch}} \in (0, 1]$.

The travel distance of CEV n from locations s to v in MG $i \in \mathcal{M}$ at time t is defined as $d_{sv,n,t} = v_{sv,n,t}\Delta t$, where $v_{sv,n,t}$ is the average travel speed across the road (s, v). The battery level of CEV n at time t can accordingly be represented as

$$\mathrm{SOC}_{n,t+\Delta t} = \mathrm{SOC}_{n,t} - \frac{\lambda^n}{C_n^{\mathrm{ev}}} d_{sv,n,t}, \tag{13}$$

where $\lambda^n \geq 0$ is the CEV n's unit energy consumption factor.

Through the connections of wireless PTs, the energy demand of V2G scheduling shall be met for each participating CEV n at every MG $i \in \mathcal{M}$. Each CEV n's required SOC level corresponds to the energy requirement. Let SOC_n^0 and $\mathrm{SOC}_n^\star$ represent the starting SOC of the CEV n and its required SOC level, respectively. Before arriving at the final destination, the charging requirement must be satisfied as indicated by

$$\mathrm{SOC}_{n,|\mathcal{T}|} \geq \mathrm{SOC}_n^0 + \mathrm{SOC}_n^\star, \ \forall n \in \mathcal{N}, \tag{14}$$

where $\mathrm{SOC}_n^\star$ is the charging requirement of CEV n.

Additionally, during the charging and discharging process, it is important to avoid over-charging and deep-discharging situations. Consequently, the SOC operational limits follow

$$\underline{\mathrm{SOC}_n} \leq \mathrm{SOC}_{n,t} \leq \overline{\mathrm{SOC}_n}, \ \forall n \in \mathcal{N}, \tag{15}$$

where the lower and upper SOC limits of CEV n are $\underline{\mathrm{SOC}_n}$ and $\overline{\mathrm{SOC}_n}$, respectively.

Besides the CEV dynamic wireless scheduling scheme, we further consider the CEV wired charging scenarios. When the CEVs are plugged in the charging stations in MG i, they are first scheduled by the system to offer the V2G regulation service by establishing real-time communications between the charging stations and CEVs. In this case, the constraints are developed based on the stationary parking time period, defined by $t \in [T_{n,in}, T_{n,out}]$ where the plug-in and plug-out times of CEV n are $T_{n,in}$ and $T_{n,out}$, respectively. Then, the CEV coordination is considered within the period for (10). It indicates that each CEV should get its own charging requirements guaranteed before departing from charging stations.

3.3. Non-cooperative microgrid model

The proposed CMT system framework is formulated to tackle the joint MG energy trading and CEV coordination operations. For the MG operation, the related objective of the entire CMT system is to minimize the quantity of the total discarded renewable energy. Meanwhile, by following the aggregator model in [28], the cost of CEV charging/discharging is incurred by the V2G scheduling mechanism. Additionally, since we take the proposed CMT system's stability requirement into consideration, each MG is capable of providing V2G regulation services by means of the participated CEVs in its individual MG system. The aim of regulation services is to smooth out active power fluctuations in order to maintain a stable range for grid frequency. As a result, based on these aforementioned aspects, for each microgrid i at time t, the utility function is defined as:

$$\begin{aligned} U_{i,t} = & -p_t^{\mathrm{tr}}\Big(\sum_{j\neq i} S_{ij,t} - \sum_{j\neq i} S_{ji,t}\Big) + \sum_{j\neq i}(S_{ji,t})^2 \\ & + p_i^{\mathrm{re}}(R_{i,t}) - p_{i,t}^{\mathrm{g}} E_{i,t}^{\mathrm{g}} + \sum_{n\in\mathcal{N}} p_{n,t}^{\mathrm{ch}} P_{n,t} + w\left(\frac{F}{|\mathcal{T}|}\right. \\ & \left. - \xi|P_t^{\mathrm{tot}} - P_{t-1}^{\mathrm{tot}}|\right) + \sigma_1 \max\left(C_{i,t} + L_{i,t} + S_{i,t}\right. \\ & \left. - G_{i,t} - E_{i,t}^{\mathrm{g}} - \sum_{j\neq i}\beta_{ji,t}, 0\right) + \sigma_2 \max\left(\mathrm{SOC}_n^* \right. \\ & \left. + \mathrm{SOC}_{n,t}^{\mathrm{ms}} - \mathrm{SOC}_{n,t} - \eta(P_{n,t})P_{n,t}\Delta t \right. \\ & \left. - \eta^{\mathrm{ch}} P_n^{\mathrm{ch}}(|\mathcal{T}| - t - 1)\Delta t, 0\right), \end{aligned} \tag{16}$$

where p_t^{tr} is the energy trading prices provided by the CMT system operator and p_i^{re} denotes the unit price for renewable energy generations. w denotes the weighting factor for the provisioning of the V2G regulation service. In addition, $p_{n,t}^{\mathrm{ch}}$ is the unit charging/discharging price for CEV scheduling with the power of all CEVs $P_{\mathcal{N},t} = \{P_{1,t}, P_{2,t}, \ldots, P_{N,t}\}$, while the base reward for rendering the regulating services is F. Due to the battery degradation effect, the penalty factor is set as $\xi \in (0, 1]$. Lastly, the active power profile of the city smart grid, which is determined as P_t^{tot}, is obtained by

$$P_t^{\mathrm{tot}} = P_t^{\mathrm{imb}} + \sum_{n\in\mathcal{N}} P_{n,t}, \tag{17}$$

where $P^{\mathrm{imb}}(t)$ denotes the timely power imbalance volume of the CMT system. By focusing on the utility function in (16), the first term represents the energy selling revenue or buying cost and the second term is related to the incurred operational cost for energy trading process. Moreover, the third term presents the operational cost on renewable energy generations while the fourth term is associated with the electricity payment from the utility grid. Besides, the fifth and sixth terms are the participating CEVs' charging expenses and rewards for providing the V2G regulation services, respectively. These two terms are directly related to the effectiveness of the V2G scheduling scheme in each MG i. Furthermore, the seventh term introduces the constraint penalty function related to CMT energy trading process, where σ_1 penalizes the reward if the grid demand exceeds the upper bound based on (5). For the last term, σ_2 penalizes the reward within each MG i if the charging requirement is not satisfied based on (18). In this case, since the non-cooperative MG model is developed in a real-time manner, based on constraints (11), (14), and (15), the charging requirement of the proposed online V2G scheduling of CEV n shall follow

$$\begin{aligned} & \mathrm{SOC}_{n,t} + \eta(P_{n,t})P_{n,t}\Delta t + \eta^{\mathrm{ch}} P_n^{\mathrm{ch}}|(\mathcal{T}| - t - 1)\Delta t \\ & \qquad \geq \mathrm{SOC}_n^* + \mathrm{SOC}_{n,t}^{\mathrm{ms}}, \ \forall t \leq |\mathcal{T}| - 1. \end{aligned} \tag{18}$$

On the left-hand side of the inequality above, the maximum SOC that can be charged for CEV n and its charging or discharging power $P_{n,t}$ are displayed, while the safety margin, denoted as $\mu \in [0, 1]$, is utilized to deal with the uncertainty of real-time regulation signals, which is shown on the right-hand side. The aim is to let the participating CEV n have a large energy buffer to deal with the uncertainty of real-time power imbalance. Then, we have

$$\mathrm{SOC}_{n,t}^{\mathrm{ms}} = \mu(\overline{\mathrm{SOC}_n} - \mathrm{SOC}_n^*). \tag{19}$$

Besides, we introduce $\boldsymbol{H}_{i,t} = \{\{S_{ij,t}|i \neq j\}, R_{i,t}, E_{i,t}^{\mathrm{g}}, P_{\mathcal{N}_i,t}\}$ as the feasible actions of MG i, where it satisfies (4), (5), (6), (8), (9), (10), and (15). Thus, the proposed non-cooperative CMT game can be formulated as:

$$\textbf{maximize} \quad U_{i,t}, \tag{20a}$$

$$\textbf{subject to} \quad \boldsymbol{H}_{i,t} \in \mathcal{A}_i, \tag{20b}$$

where $\mathcal{X}_i$ denotes the feasible strategy set of MG i for the non-cooperative game, which includes all the feasible sequence of related variables.

We denote the proposed CMT game $\mathcal{W}(\mathcal{A}, \mathcal{F})$ as the set of online coupled optimization problem (20) of all CEVs, which includes the strategy set $\mathcal{A} \triangleq \prod_{i=1}^{M} \mathcal{A}_i$. The related payoff function is denoted as $\mathcal{F}(\mathcal{M}) = \sum_{i\in\mathcal{M}}(U_{i,t})$. Generally, a Nash equilibrium solution is a feasible value of $\mathcal{A}_i^*$ under the condition that no MG may unilaterally deviate and obtain more benefit if the other MGs' strategies do not change. In the meantime, the best response of MG i refers to the strategy at the equilibrium.

4. Deep deterministic policy gradient

The game $\mathcal{W}(\mathcal{Z}, \mathcal{F})$ is hard to solve due to the imperfect information. Hence, in order to attain the equilibrium, we reformulate the proposed CMT game as an MDP. In this way, the devised DDPG method is capable of finding the solution.

4.1. MDP reformulation of proposed CMT game

State Space. Recall that the total time period is divided into T time slots. For the information state of the proposed CMT game, it consists of seven sub-states, namely, the power generation of MG i $G_{i,t}$, the load demand of MG i $L_{i,t}$, the quantity of energy transmitted between MGs i and j $\alpha_{ij,t}$, the timely amount of energy sold at MG i $S_{i,t}$, the regulation signals P_t^{reg}, and the residual SOC $SOC_{\mathcal{N}_i,t}$, and the number of participated CEVs in each MG i $N_{i,t}$. $G_{i,t}$ and $L_{i,t}$ are the basic elements for the MG i. $\alpha_{ij,t}$ presents the amount of energy transmitted and $S_{i,t}$ the timely amount of energy sold at MG i. In addition, P_t^{reg} is the regulation signals that affect the charging/discharging price of N_i participated CEVs at MG i and the reward function of MG i. $SOC_{\mathcal{N}_i,t}$ denotes the residual SOC of CEVs at MG i so as to fulfill the CEV fleets' charging needs $SOC^*_{\mathcal{N}_i}$, while $N_{i,t}$ is the real-time number of participated CEVs in MG i. Therefore, the information state of MG i can be described as $\boldsymbol{s}_{i,t} = [G_{i,t}, L_{i,t}, \alpha_{ij,t}, S_{i,t}, P_t^{\text{reg}}, SOC_{\mathcal{N}_i,t}, N_{i,t}, t]$.

Action Space. Referring to the constraint (20), we execute the action space $\boldsymbol{a}_t^i$ at each time t and it is related to the MG i's strategy $\boldsymbol{H}_{i,t}$. Different from most of the existing works that consider these elements to be discrete variables, in this work, we exploit the characteristics of latest technology in both MG operations and CEV charging schemes through continuous rate control. In other words, we define $\boldsymbol{a}_{i,t}$ as a continuous variable vector, where $\boldsymbol{a}_{i,t} = \boldsymbol{H}_{i,t}$. In the meantime, we should notice that the following requirements are met during the operation process: (i) the utilized renewable $R_{i,t}$ should not exceed the total power generation in (1), (ii) the quantity of energy received at MG j at time t in (2), (iii) the volume of energy drawn from the electric grid should not be greater than the maximum volume in (4), and (iv) the charging/discharging rate $P_{\mathcal{N}_i,t}$ should not exceed the maximum charging/discharging rate mentioned in (10).

State Transition. The next information state of the MDP reformulation is denoted as $\boldsymbol{s}_{i,t+1}$, which depends on both the action vector at the current time slot t and the MG i operation pattern at the time $t+1$. The transition probability from $\boldsymbol{s}_{i,t}$ to $\boldsymbol{s}_{t+1}^i$ is denoted as $\mathcal{P}(\boldsymbol{s}_{i,t+1}|\boldsymbol{s}_{i,t}, \boldsymbol{a}_{i,t})$

Reward. By performing the action $\boldsymbol{a}_t^i$ that is related to the strategy of MG i and the charging/discharging powers of the CEVs at MG i, the proposed CMT system can harvest a reward at the beginning of time $t+1$. Recall that the utility objective of the proposed CMT game, formulated in (20), aims to tackle the joint MG energy trading and CEV coordination operations. Hence, we define $R_{i,t+1}$ as the reward of MG i by following the objective function in (20). Besides, we introduce the long-term return J to correspond with the information state $\boldsymbol{s}_t^i \in \mathcal{S}_i$ and the action $\boldsymbol{a}_t^i \in \mathcal{A}_i$, where $\mathcal{S}_i$ and $\mathcal{A}_i$ are the state and action spaces, respectively.

For the MDP reformulation, the target is to maximize the expected total reward yield by each MG i, which has

$$J = \mathbb{E}[\frac{1}{T}\sum_{t\in\mathcal{T}} R_{i,t+1}(\boldsymbol{a}_t^i)]. \tag{21}$$

4.2. Deep deterministic policy gradient algorithm

Referring to [29], DDPG is a model-free, actor-critical algorithm that works in continuous state and action spaces. The action-value function acts as the critic in the actor–critic algorithm to evaluate the actor's performance and directing the actor's upcoming actions, while the policy function acts as the actor to generate actions and interact with the environment.

The actor–critic algorithm's two approximation functions were developed by the DDPG using deep neural networks (DNNs). A policy function $\mu(\boldsymbol{s}_{i,t}|\theta_\mu)$ can be used to define the actor network, where θ_μ is the parameter in the network μ. The critic network is defined as an action-value function $Q(\boldsymbol{s}_{i,t}, \boldsymbol{a}_{i,t}|\theta_Q)$, where θ_Q represents the parameter in the network Q. The actor target network μ' with the parameter $\theta_{\mu'}$ and the critic target network Q' with the parameter $\theta_{Q'}$ are also built with a copy to facilitate the training process.

The aim is to determine an action $\boldsymbol{a}_{i,t}$ so as to maximize the action-value function $Q(\boldsymbol{s}_{i,t}, \boldsymbol{a}_{i,t})$ by

$$\boldsymbol{a}^*_{i,t} = \arg\max_{\boldsymbol{a}_{i,t}} Q(\boldsymbol{s}_{i,t}, \boldsymbol{a}_{i,t}). \tag{22}$$

Based on the state-value function (20), the objective function in the DDPG can be redefined as J_{θ_μ}. Since the state-value function is differentiable with the continuous action space, the parameter θ_μ can be updated through gradient descent. Hence, the chain rule of such gradient search in the devised algorithm follows

$$\nabla_{\theta_\mu} J = \nabla_{\boldsymbol{a}_t^i} Q(\boldsymbol{s}_{i,t}, \boldsymbol{a}_{i,t}|\theta_Q)\nabla_{\theta_\mu}\mu(\boldsymbol{s}_{i,t}|\theta_\mu). \tag{23}$$

The entire process for the deep deterministic policy gradient algorithm is shown in Algorithm 1. The following procedure describes how to train a DDPG-based agent. For the state $\boldsymbol{s}_{i,t}$, the network μ produces $\mu(\boldsymbol{s}_{i,t})$ with the consideration of the noise $\boldsymbol{v}_t$ so as to obtain the action $\boldsymbol{a}_{i,t} = \mu(\boldsymbol{s}_{i,t}) + \boldsymbol{v}_t$. Then, at a new state $\boldsymbol{s}_{i,t+1}$, the agent obtains a reward $R_{i,t}$. After that, the agent then stores the tuple $(\boldsymbol{s}_{i,t}, \boldsymbol{a}_{i,t}, R_{i,t}, \boldsymbol{s}_{i,t+1})$ in the experience replay buffer. A mini-batch $(\boldsymbol{s}_{i,\tau}, \boldsymbol{a}_{i,\tau}, R_{i,\tau}, \boldsymbol{s}_{i,\tau+1})$ is then created by the agent when selecting N_τ tuples from the buffer.

Based on the chain rule in (21), the actor network must be updated when it is being trained by the above procedure. In the meantime, the critic network also follow the same process. After that, the target critic network considers the output of the target actor network as its input, and the target value for the state-value function is given by

$$y_{i,t} = R_{i,t} + \epsilon Q_t(\boldsymbol{s'}_{i,t}, \mu_t(\boldsymbol{s'}_{i,t}|\theta_{\mu'_t})|\theta_{Q'_t}), \tag{24}$$

where ϵ is the discount parameter.

Through the mini-batch method, the network μ' outputs the action $\mu'(\boldsymbol{s}_{i,\tau+1})$ to the network Q'. In order to compute the loss L using the mini-batch and $\boldsymbol{s}_{i,\tau+1}$, the network Q' can calculate $y_{i,\tau}$ based on (20). In this case, the loss of the network should be minimized for the critic network. Thus, we have

$$\min L = \min(y_{i,t} Q(\boldsymbol{s}_{i,t}, \boldsymbol{a}_{i,t}|\theta_Q))^2. \tag{25}$$

The two target actor and critic networks' corresponding parameters are updated by

$$\theta_{\mu_t} = \delta\theta_{\mu_t} + (1-\delta)\theta_{\mu_t}, \tag{26}$$

$$\theta_{Q_t} = \delta\theta_{Q_t} + (1-\delta)\theta_{Q_t}, \tag{27}$$

where δ denotes the soft updating parameter of the networks.

Algorithm 1 Deep deterministic policy gradient algorithm

```
Initialize network parameters μ and θ.
for each iteration z = 1: Z do
  Initialize action space and observation state.
  for t = 1: |T| do
    Choose action a_{i,t} by considering noise v_t.
    Observe state s_{i,t}, generate action a_{i,t}, and approximate the
    state-value function.
    Calculate reward R_{i,t} and observe the state s_{i,t+1}.
    Store the transition (s_{i,t}, a_{i,t}, R_{i,t}, s_{i,t+1}).
    Sample a random mini-batch with N_τ tuples and calculate (24).
    Update the actor policy by using the gradient in (23).
    Update the critic by following (25).
    Update the target network Q' based on (26) and (27).
  end for
  if cross-validation is required then
    Validate the current network parameter μ in the policy
    network.
  end if
end for
```

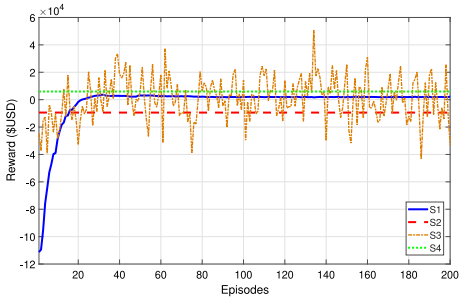


Fig. 3. Comparison of reward among different scenarios.

5. Case studies

In this section, the effectiveness of the proposed CMT system is evaluated. Firstly, the simulation setup is presented. Secondly, the relevant performance metrics and the baseline scenarios for comparison are included. Thirdly, the energy trading process is studied. Fourthly, the economic feasibility of the proposed model is covered. Lastly, the impact of different CEV fleets is examined.

5.1. Simulation setup

The proposed CMT system is studied on both the MG operations and CEV coordination. The time horizon for the system's operational period is one day, divided into $|\mathcal{T}| = 96$ slots with a length of 15 min. In each MG, the power load demand and the renewable energy generated are obtained from real power system data in [30], which includes time-varying household load and solar and wind power generations. The utility grid's adopted electricity tariff is set according to the time-of-use (TOU) pricing plan [31]. By following the PJM market standard [32], a regulation capacity of each MG is set to 300 kW with base reward $40 per MWh. For MMG power network, a fully connected MMG system is considered and it includes four individual MGs. For each MG, the urban transportation network is obtained as the real-world traffic network in the town of Carolina, Alabama,[1] in which it includes 26 nodes and 56 edges with a different length of road segment.

The settings for CEVs are presented in the following. There are two groups of CEV fleets with $|\mathcal{N}| = 80$ that are equally distributed in each MG, e.g. Nissan Leaf with 40 kWh [33] and Chevrolet Volt with 18.4 kWh [34]. We assume that each CEV travels with a fixed velocity of 30 kilometers (18.64 miles) per hour referring to the city average traffic speed. In the meantime, each CEV's λ^n (unit energy consumption) is set at 1.112 kWh per kilometers. The wired and wireless charging standards follow [24,26], respectively, as mentioned in Section 2.1. In the meantime, the energy efficiency is set to 0.9 for both wired and wireless charging/discharging schemes. Besides, we model the uniform distribution-based SOC settings of CEVs, which is denoted as U[·] [35]. Specifically, we set the CEV n's initial SOC to be U[40%, 50%], and the charging requirement to be U[0%, 10%]. The safety condition of the charging/discharging mechanism is considered when the minimum and maximum SOCs, respectively, are U[10%, 20%] and U[90%, 99%]. Last but not least, the charge price is set to $0.59 per kWh.

5.2. Algorithms for comparison

In the simulation, we consider four different algorithms in comparison so as to assess the performance of the proposed CMT system, which are shown as:

1. S1: Proposed DDPG-based approach with imperfect information,
2. S2: Constant charging policy approach adopted in [35],
3. S3: Random policy with equal probability approach adopted in [36],
4. S4: Non-cooperative CMT game approach in (20) with perfect information.
5. S5: Non-cooperative MMG game approach without CEV fleets [14].

S1 denotes the proposed CEV system with DDPG-based approach with imperfect information. For CEV scheduling schemes in each MG, we assume that CEVs perform dynamic wireless scheduling when they travel to the CEVAs and then they operate fast charging mechanisms through wired chargers. Meanwhile, all road segments are installed with PTs in the transportation network of MG i. Then, S2 modifies the charging policy during the entire time period, wherein each CEV is charged to meet its charging needs at a constant power rate. In addition, for S3, we assume that 50% of road segments are embedded with PTs in a random manner. For S3, we assess the random policy approach. In this case, each MG with the participated CEVs can choose its action with equal probability and this method does not involve the implementation of dynamic V2G scheduling. Furthermore, S4 considers solving the proposed non-cooperative CMT game approach shown in (20), where it covers the perfect information. Last but not least, S5 considers the non-cooperative MMG game approach without the consideration of CEV fleets, in which it does not include the model of V2G scheduling scheme. In the following performance evaluations, we mainly consider four MGs in the proposed CMT system, which also follows the settings in [14]. In practice, the proposed model can be deployed in different CMT systems with various typologies.

5.3. Reward evaluation under proposed DDPG-based approach

In this part, we evaluate the episodic reward under different scenarios. In the hypothetical game $\mathcal{W}(\mathcal{A}, \mathcal{F})$, the social utility of MGs can be maximized at the equilibrium point. Thus, in order to assess the convergence of the proposed DDPG-based approach, the performance metric that we use is social welfare, also referred to as the sum of all MG rewards. The relevant results reflected by the four comparative scenarios are shown in Fig. 3. Owing to the utilization of random policy, S3 performs worst because of the largest volatility occurred in the reward profile. Besides, for S2, due to the implementation of constant charging policy, the reward seems to be nearly a straight line since the difference between the minimum and maximum of reward under every episode during the training process is relatively small. Through the utilization of DDPG-based reinforcement learning approach, we can see that S1 performs well on social welfare since the curve approximately converges to the equilibrium as the episode lengthens in S1. It is demonstrated that with imperfect information, S1 can still nearly converge to the Nash equilibrium attained by S4.

5.4. Energy trading results of CMT system

We assess how the proposed CMT system's MGs trade energy among themselves by means of Algorithm 1 in this section. The process of energy exchange between MGs and the utility grid is shown in Fig. 3, along with a description of the MGs' energy trading behaviors. Here we focus on the cumulative hourly performance for the representation of these MGs. It is apparent that MG 1 sells relatively large amount of energy to other MGs due to its amount of the generated renewable energy and V2G power. Meanwhile, the amount of the V2G power is defined as $P_{\mathcal{T}}^{\text{tot}}$ in (17). In this case, MG 1 can earn more income by selling energy to other MGs with the generated renewable energy surplus. Considering the characteristics of hourly power imbalance in the system, MGs 2, 3, and 4 can either be an energy buyer or seller at particular times. In addition, for the CEV scheduling schemes, we

[1] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/CUWWYJ

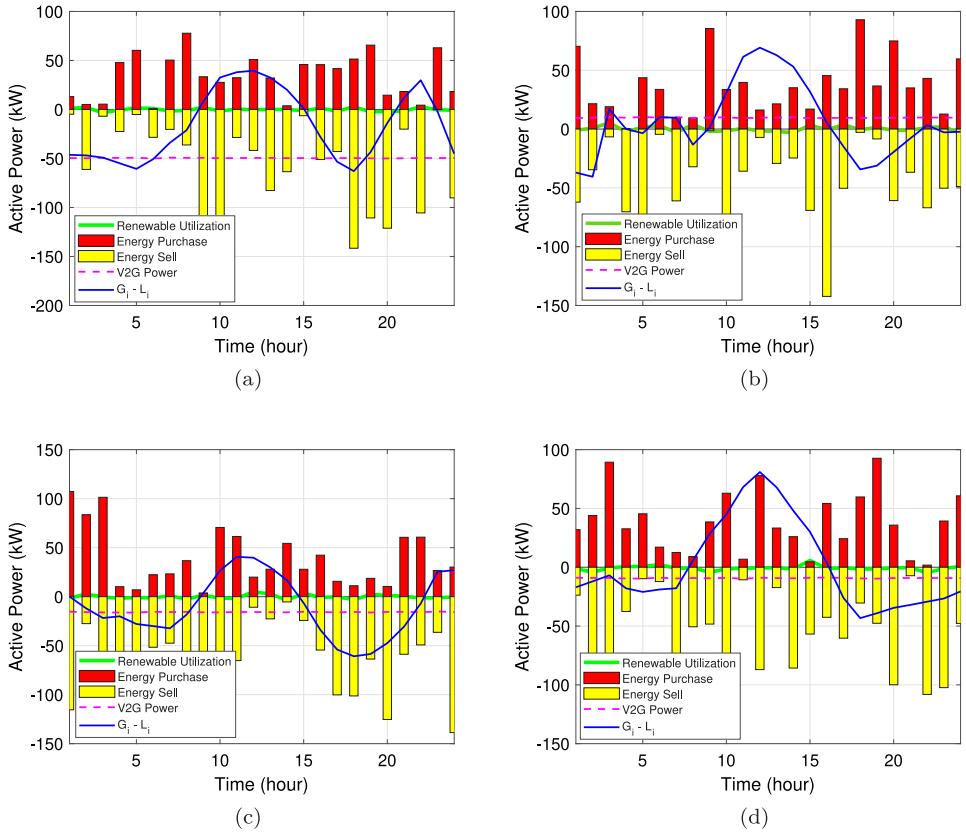


Fig. 4. Energy trading process in proposed CMT system: (a) MG 1, (b) MG 2, (c) MG 3, (d) MG 4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

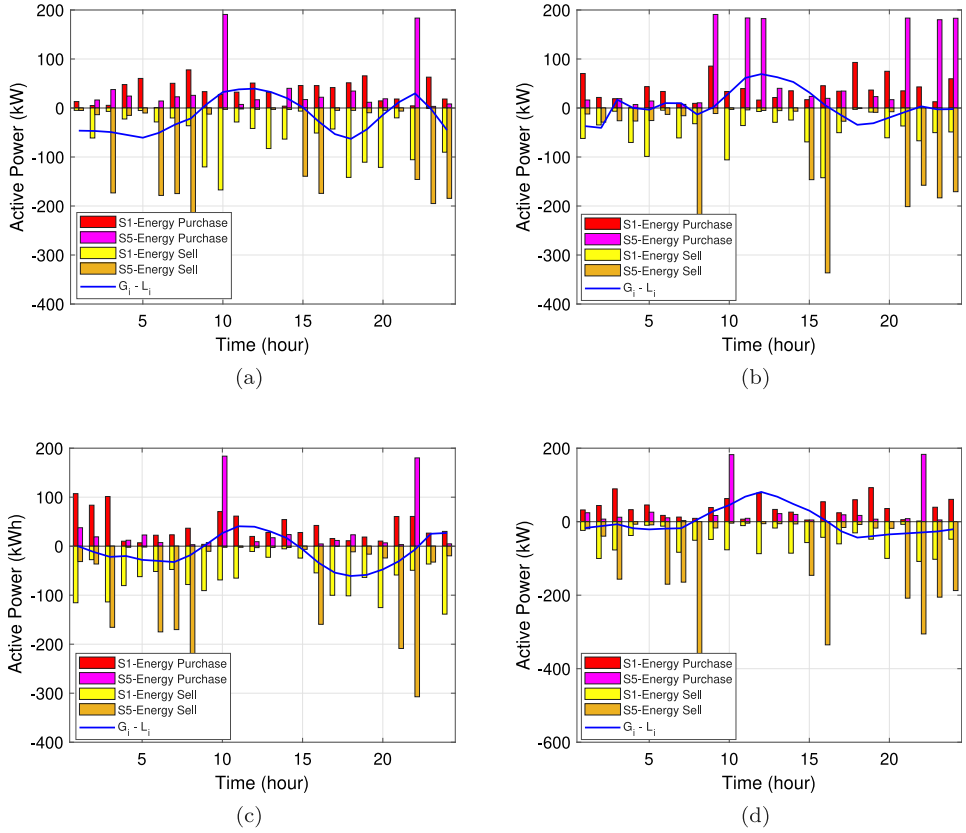


Fig. 5. Comparison of different approaches in energy trading process: (a) MG 1, (b) MG 2, (c) MG 3, (d) MG 4.

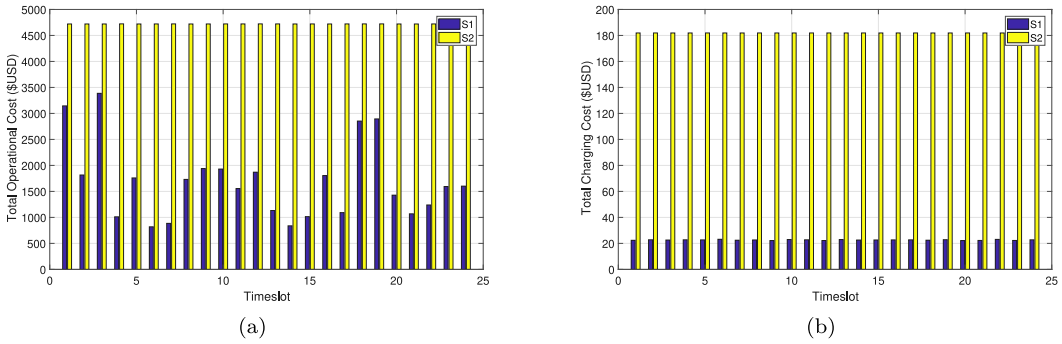


Fig. 6. Total economical cost in proposed CMT system: (a) operational cost, (b) charging cost.

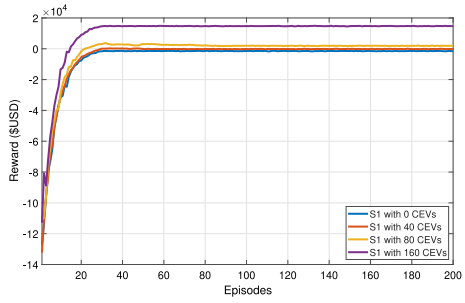


Fig. 7. Reward among different number of CEV fleets.

consider the combination of both wired and wireless CEV scheduling operations. In particular, when each CEV travels to the assigned CEVA by following the routing plan in each MG i, it mainly performs dynamic wireless charging scheme since all road segments in MG i's transportation network are embedded with PTs. After each CEV reaches the CEVA to park, it operates wired fast charging schemes. As shown in Fig. 4, given the time period, the total amount of energy sold by all MGs (shown by the yellow bar) is typically more than the total amount of energy purchased (represented by the red bar). The main reason is related to the transmission losses among these MGs in the system. Besides, since the volume of the power imbalance cannot be met by the renewable energy that has been generated, the V2G power is utilized for compensated the active power imbalance. It can be observed that MG 2 choose to charge more power by means of the CEV fleets while the CEV fleets of the other three MGs operate discharging modes.

Besides, we assess the performance of the energy trading process between S1 and S5. The result is presented in Fig. 5. It is apparent that the amount of purchasing or selling energy in S1 is a bit less than the amount of S5. The reason is that the use of V2G services can help the certain amount of energy fulfilling process since S5 does not consider the implementation of the CEV fleets. In addition, since S5 can generate more RE for energy trading process among these four MGs, more operational cost can be calculated due to the relatively large amount of RE generations. Therefore, the use of the CEV fleets in S1 can help reduce the system operational cost by using more CEV powers rather than using RE powers. Thus, comparing with the baseline approach S5, the proposed method can incur a lower system operational cost during the energy trading procedure.

5.5. Cost evaluation

Based on the same simulation settings, we assess the cost evaluation of the utilities of the CMT system. As shown in Fig. 6, it depicts the entire economical cost of the proposed CMT system, which covers the operational cost for energy transfer among MGs and the charging cost for CEVs' charging behaviors. As can be observed from Fig. 6(a), the total operation cost obtained by S1 is much lower when compared to S2. This further incurs that the efficient V2G schedules of CEVs in these MGs can contribute to a lower system operational cost. In addition, according to the charging cost in Fig. 6(b), the deployment of the effective V2G scheduling operation in S1 can indeed lower the total charging cost over the entire day. Therefore, we can conclude that the effective V2G coordination contributes to a higher economical benefit by utilizing CEVs in the CMT system.

5.6. Impact of different CEVs

Lastly, we study the impact of various CEV fleet sizes on the non-cooperative CMT system, which is associated with the scalability of the proposed game. In this part, we evaluate the system performance of four cases with 0, 40, 80, and 160 CEVs, respectively, which are evenly distributed to the four MGs in the system. Fig. 7 shows the result under these four cases. As it can be seen, the increase of CEV fleet size contributes to a higher reward in the system. In addition, when the amount of total participated CEVs is greater than 80, the reward profile reaches to the positive equilibrium level quickly. Hence, it is demonstrated that more participated CEVs in the CMT system contribute to a higher system reward. Besides, by tracing the performance of the active power profile, we can obtain that the standard deviations of total active power profile (in kW) under these four cases are 64.64, 3.90×10^{-1}, 2.01×10^{-2}, and 7.04×10^{-4}, which is shown in Table 1. It is apparent that this metric decreases with the increasing number of CEVs, indicating a better performance in providing the V2G regulation service in the CMT system. Furthermore, the average charging cost increases with the fleet size while and the average operation cost nearly remains the same, as reflected in Table 1.

5.7. Tractability for large system

In this section, we investigate the numerical performance of six MGs in the proposed CMT system. Specifically, we consider the original four fully-connected MGs and two other MGs with attached ring topology for the grid. The number of CEVs are evenly separated into these MGs with $N_i = 20$, where $i \in \mathcal{M}$. The main purpose is to investigate the effect of energy trading process for a relatively large CMT system. The result is shown in Fig. 8. We can observe that MGs 1, 2, 3, and 6 sells more energy than the buying amount compared to other MGs. Besides, MG 5 buys a relatively large amount of energy from other MGs. It is demonstrated that to fulfill the power imbalance, all MG buys/sells sufficient energy during the peak and valley of the power profile. In addition, considering the implementation of V2G operations in every MG, the CEV fleets in MGs 4 and 6 select to charge more power during the entire operation period while the CEV fleets in MGs 1, 2, 3, and 5 prefer to discharge more power back to the system so as to alleviate the power imbalance volume.

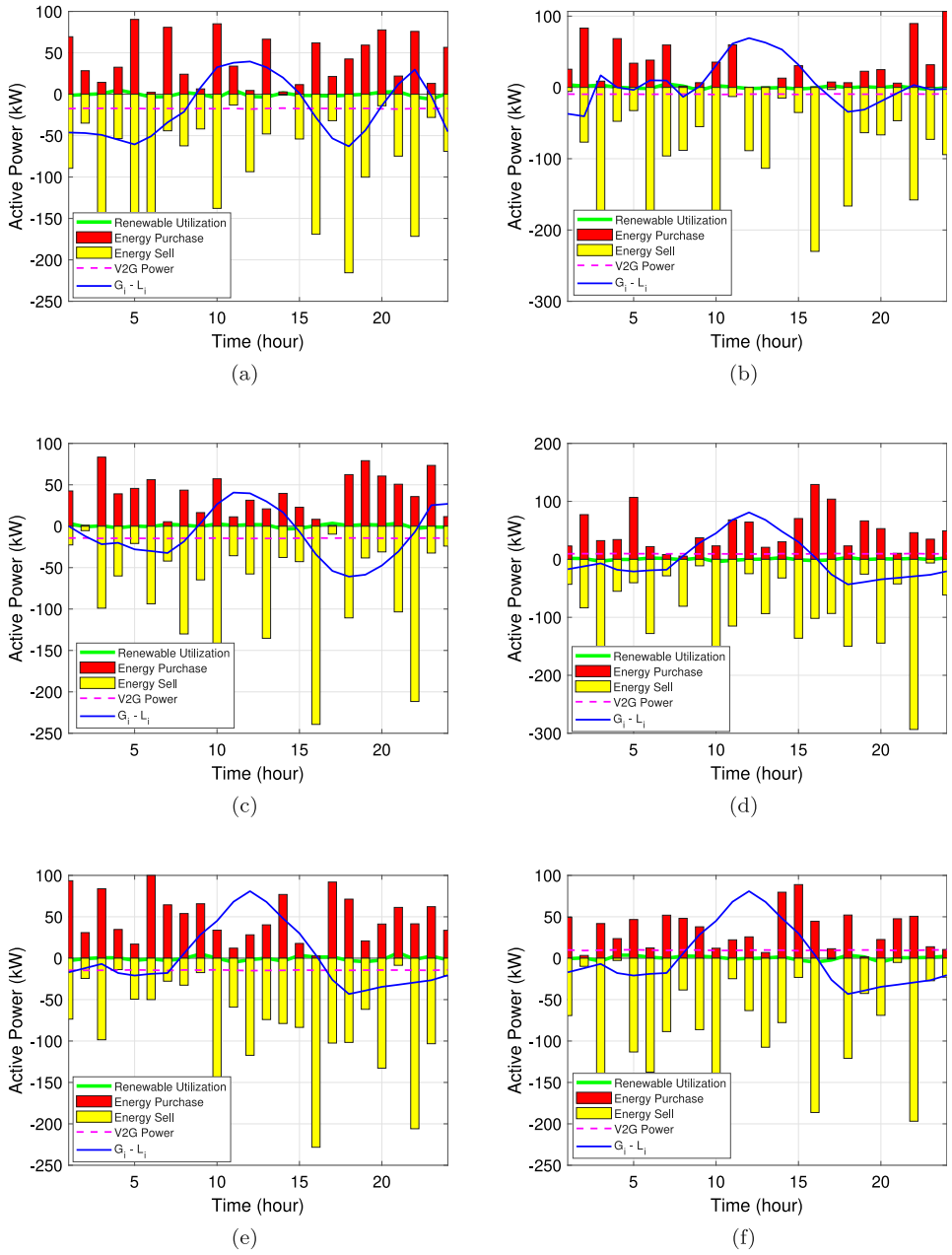


Fig. 8. Energy trading process in large CMT system: (a) MG 1, (b) MG 2, (c) MG 3, (d) MG 4, (e) MG 5, (f) MG 6.

Table 1
Impact of different fleet size of CEVs in S1.

Number of CEVs	0	40	80	160
Standard deviation of total active power profile (kW)	64.64	3.90×10^{-1}	2.01×10^{-2}	7.04×10^{-4}
Average charging cost (US$)	0	4.53	22.65	134.79
Average operation cost (US$)	1.48×10^{3}	1.49×10^{5}	1.55×10^{3}	1.49×10^{3}

6. Conclusions

In this paper, we propose a game theoretic approach for the non-cooperative CMT system by exploiting CEVs in smart grid. The combined operations can jointly trigger the MG energy trading and vehicle-to-grid coordination schemes in the system. Considering the multiple interactions between the grid operator and MGs, we formulated a non-cooperative game approach based on a Stackelberg game, where the grid determines the energy trading price as the leader while the participated MGs decide their strategies as followers. To deal with the imperfect information incurred by the CMT game, we develop an MDP and DDPG-based method. The devised algorithm can learn a deterministic optimal response from the opponents. In addition, the game's Nash equilibrium can be approximately converged even under imperfect information. Simulation results show the reward curve of the proposed approach can successfully converge to the equilibrium. In addition, the deployment of the V2G scheduling can further help contributes a higher economical benefit in the system. Besides, more CEVs participated in the V2G regulation service can lower the system operational cost significantly. Furthermore, the proposed approach can

ensure the effectiveness of the energy trading process in different scales of the CMT system. In the future, we will extend the system model with the implementation of multiple operations for heterogeneous CMT elements in the system, as well as coping with their potential uncertainties.

CRediT authorship contribution statement

Shiyao Zhang: Writing – original draft, Software, Methodology. **Xingzheng Zhu:** Writing – review & editing, Conceptualization. **Shengyu Zhang:** Validation, Software. **James J.Q. Yu:** Writing – review & editing, Supervision.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by the Stable Support Plan Program of Shenzhen Natural Science Fund, China under Grant 2022081511 1111002.

References

[1] Chen S, Hu J, Shi Y, Peng Y, Fang J, Zhao R, et al. Vehicle-to-everything (v2x) services supported by LTE-based systems and 5G. IEEE Commun Stand Mag 2017;1(2):70–6. http://dx.doi.org/10.1109/MCOMSTD.2017.1700015.

[2] Huang AQ, Crow ML, Heydt GT, Zheng JP, Dale SJ. The future renewable electric energy delivery and management (FREEDM) system: The energy internet. Proc IEEE 2011;99(1):133–48. http://dx.doi.org/10.1109/JPROC.2010.2081330.

[3] Zou H, Mao S, Wang Y, Zhang F, Chen X, Cheng L. A survey of energy management in interconnected multi-microgrids. IEEE Access 2019;7:72158–69. http://dx.doi.org/10.1109/ACCESS.2019.2920008.

[4] Awad H, Bayoumi EHE, Soliman HM, De Santis M. Robust tracker of hybrid microgrids by the invariant-ellipsoid set. Electronics 2021;10(15). http://dx.doi.org/10.3390/electronics10151794.

[5] Microgrids: A review, outstanding issues and future trends. Energy Strategy Rev 2023;49:101127. http://dx.doi.org/10.1016/j.esr.2023.101127.

[6] Toutounchi AN, Seyedshenava S, Contreras J, Akbarimajd A. A stochastic bilevel model to manage active distribution networks with multi-microgrids. IEEE Syst J 2019;13(4):4190–9. http://dx.doi.org/10.1109/JSYST.2018.2890062.

[7] Arefifar SA, Ordonez M, Mohamed YA-RI. Voltage and current controllability in multi-microgrid smart distribution systems. IEEE Trans Smart Grid 2018;9(2):817–26. http://dx.doi.org/10.1109/TSG.2016.2568999.

[8] Yang D, Zhang C, Jiang C, Liu X, Shen Y. Interval method based optimal scheduling of regional multi-microgrids with uncertainties of renewable energy. IEEE Access 2021;9:53292–305. http://dx.doi.org/10.1109/ACCESS.2021.3070592.

[9] Zhao Z, Guo J, Luo X, Lai CS, Yang P, Lai LL, et al. Distributed robust model predictive control-based energy management strategy for islanded multi-microgrids considering uncertainty. IEEE Trans Smart Grid 2022;13(3):2107–20. http://dx.doi.org/10.1109/TSG.2022.3147370.

[10] Liu Y, Wang Y, Li Y, Gooi HB, Xin H. Multi-agent based optimal scheduling and trading for multi-microgrids integrated with urban transportation networks. IEEE Trans Power Syst 2021;36(3):2197–210. http://dx.doi.org/10.1109/TPWRS.2020.3040310.

[11] Li C, Xu Y, Yu X, Ryan C, Huang T. Risk-averse energy trading in multi-energy microgrids: A two-stage stochastic game approach. IEEE Trans Ind Inf 2017;13(5):2620–30. http://dx.doi.org/10.1109/TII.2017.2739339.

[12] Jadhav AM, Patne NR. Priority-based energy scheduling in a smart distributed network with multiple microgrids. IEEE Trans Ind Inf 2017;13(6):3134–43. http://dx.doi.org/10.1109/TII.2017.2671923.

[13] Yu Y, Li G, Li Z. A game theoretical pricing mechanism for multi-microgrid energy trading considering electric vehicles uncertainty. IEEE Access 2020;8:156519–29. http://dx.doi.org/10.1109/ACCESS.2020.3019815.

[14] Zhu X, Chen X, Leung K-C. A game theoretic approach to energy trading in multi-microgrid systems. In: 2019 IEEE international systems conference. 2019, p. 1–7. http://dx.doi.org/10.1109/SYSCON.2019.8836956.

[15] Gallo J-B. Electric truck & bus grid integration, opportunities, challenges & recommendations. World Electr Veh J 2016;8:45–56. http://dx.doi.org/10.3390/wevj8010045.

[16] Noel L, McCormack R. A cost benefit analysis of a V2G-capable electric school bus compared to a traditional diesel school bus. Appl Energy 2014;126:246–55.

[17] Lam AYS, Leung K-C, Li VOK. Capacity estimation for vehicle-to-grid frequency regulation services with smart charging mechanism. IEEE Trans Smart Grid 2016;7(1):156–66. http://dx.doi.org/10.1109/TSG.2015.2436901.

[18] Ko H, Pack S, Leung VCM. Mobility-aware vehicle-to-grid control algorithm in microgrids. IEEE Trans Intell Transp Syst 2018;19(7):2165–74. http://dx.doi.org/10.1109/TITS.2018.2816935.

[19] Jafari M, Gauchia A, Zhao S, Zhang K, Gauchia L. Electric vehicle battery cycle aging evaluation in real-world daily driving and vehicle-to-grid services. IEEE Trans Transp Electrif 2018;4(1):122–34. http://dx.doi.org/10.1109/TTE.2017.2764320.

[20] Han S, Lee D, Park J-B. Optimal bidding and operation strategies for EV aggegators by regrouping aggregated EV batteries. IEEE Trans Smart Grid 2020;11(6):4928–37. http://dx.doi.org/10.1109/TSG.2020.2999887.

[21] Du Y, Li F. Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning. IEEE Trans Smart Grid 2020;11(2):1066–76. http://dx.doi.org/10.1109/TSG.2019.2930299.

[22] Gümrükcü E, Klemets JRA, Suul JA, Ponci F, Monti A. Decentralized energy management concept for urban charging hubs with multiple V2G aggregators. IEEE Trans Transp Electrif 2023;9(2):2367–81. http://dx.doi.org/10.1109/TTE.2022.3208627.

[23] Chandra Mouli GR, Kefayati M, Baldick R, Bauer P. Integrated PV charging of EV fleet based on energy prices, V2G, and offer of reserves. IEEE Trans Smart Grid 2019;10(2):1313–25. http://dx.doi.org/10.1109/TSG.2017.2763683.

[24] "Poway: epic 150 automotive inverter," epc power. 2013, [Online]. Available: https://www.epcpower.com/epic-automotive/.

[25] Dai J, Ludois DC. A survey of wireless power transfer and a critical comparison of inductive and capacitive coupling for small gap applications. IEEE Trans Power Electron 2015;30(11):6017–29. http://dx.doi.org/10.1109/TPEL.2015.2415253.

[26] Wireless power transfer for light-duty plug-in/electric vehicles and alignment methodology. 2017, SAE Standard J2954.

[27] Kempton W, Tomić J. Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy. J Power Sources 2005;144:280–94. http://dx.doi.org/10.1016/j.jpowsour.2004.12.022.

[28] Chen X, Leung K-C. Non-cooperative and cooperative optimization of scheduling with vehicle-to-grid regulation services. IEEE Trans Veh Technol 2020;69(1):114–30. http://dx.doi.org/10.1109/TVT.2019.2952712.

[29] Silver D, Lever G, Heess N, Degris T, Wierstra D, Riedmiller M. Deterministic policy gradient algorithms. In: Proceedings of the 31st international conference on international conference on machine learning - vol. 32. 2014, p. I–387–I–395.

[30] Data platform: time series, Open Power System Data, [Online]. Available: http://data.open-power-system-data.org/.

[31] "Regulatory information-sce load profiles-2011 static load profiles," sce. 2018, [Online]. Available: http://www.sce.com/005_regul_info/eca/DOMSM11.DLP.

[32] "Pjm regulation zone preliminary billing data," pjm. 2020, [Online]. Available: https://dataminer2.pjm.com/feed/reg_zone_prelim_bill/definition.

[33] Nissan. *Leaf Specifications*. 2014, [Online] Available: http://www.nissanusa.com/electric-cars/leaf/versions-specs/.

[34] Chevrolet. *Volt Specifications*. 2018, [Online]. Available: http://www.chevrolet.com/electric/volt-plug-in-hybrid.

[35] Zhang S, Leung K-C. Joint optimal power flow routing and vehicle-to-grid scheduling: Theory and algorithms. IEEE Trans Intell Transp Syst 2022;23(1):499–512. http://dx.doi.org/10.1109/TITS.2020.3012489.

[36] Chen X, Leung K-C. Fictitious self-play for vehicle-to-grid game with imperfect information. In: ICC 2019 - 2019 IEEE international conference on communications. 2019, p. 1–6. http://dx.doi.org/10.1109/ICC.2019.8761320.