

A Three-stage Incentive Scheme for Integrated Energy-Traffic Systems Using Deep Q-Network

Tianyu Yang
 Dept. of Electrical Engineering
 Tsinghua University
 Beijing, China
 yangty15@mails.tsinghua.edu.cn

Qinglai Guo
 Dept. of Electrical Engineering
 Tsinghua University
 Beijing, China
 guoqinglai@tsinghua.edu.cn

Hongbin Sun
 Dept. of Electrical Engineering
 Tsinghua University
 Beijing, China
 shb@tsinghua.edu.cn

Abstract—This paper solves the problem of dynamic pricing strategy in an urban integrated energy-traffic system. A three-stage incentive scheme has been introduced to enhance operational profits and mitigate the impact of system uncertainties. Pricing strategies in the three stages, namely day-ahead equilibrium references, hour-level prices, and surge prices, are used to incentivize human users. The proposed framework could yield a set of charging and service prices for IETS, which could improve in overall revenue and security.

Keywords—integrated energy-traffic system, dynamic pricing, reinforcement learning, Deep Q-Network

I. INTRODUCTION

Revolutionary transformation in future vehicles will be changing the way of traffic and transportation in human society [1]. Among the four trends of innovation, electric and shared are already happening in nowadays world [2]. As one of the promising solutions to the energy crisis and environment challenges, plug-in electric vehicles (PEVs) have been through rapid growth with a global fleet exceeding 5.1 million at the end of 2018 [3] and 1.13 million new sales for the 1st half of 2019 (H1) [4]. China, as the global leader in terms of market size and growth, is experiencing an explosive 66% growth in new energy vehicle (NEV) sales in 2019 H1 [4] and was determined to raise the target to 25% EV sales by 2025 [5]. As for the other branch, ride-hailing platforms, such as Uber, Lyft and Didi, are also steadily growing in global markets [6]. The new fashion of mobility services is bringing higher efficiency and better experiences while seeking for game-changers in future technology innovations.

As can be predicted in the near future, urban areas will be incorporating higher penetration of EVs and e-hailing vehicles. In such a scenario, urban traffic system and electric power system will be closely interconnected, which could lead to the concept of integrated energy-traffic systems (IETS) [7]. Moreover, human users, including drivers and passengers, are making an impact on IETS in ways of charging decisions and matching in ride-hailing platforms. The human-related factors can be considered as part of the Cyber-Physical-Human System (CPHS) [8], which could bring both uncertainties and potential to IETS.

Previous works on the topic of the coupling effect of power and traffic systems can be broadly categorized into two main aspects. A vast majority of studies look into

routing problems of EVs in IETS with interdisciplinary approach as is mentioned in [9], which includes coordinated modeling, optimization frameworks, and equilibrium analysis. Besides, economic incentives are also taken as an effective approach in coordinating IETS [10]. Based on the current situation, where energy system and traffic system have different ownership and have no mutual dispatch center, it could be a practical solution to consider price incentives as the handle to leverage IETS through human users. Despite the fact that energy system and traffic system are having coupling effect in the operational stage [7], on the contrary, studies on pricing methods either focus on ride-hailing markets [11] or coordinate charging [12][13], which could not deal with the on-going trend of system coupling.

In this paper, the problem of dynamic pricing strategy in urban IETS is studied. Real-time price incentives of both charging services and ride-hailing services are used to improve the system's performance of revenue and security. The pricing strategy is implemented in three stages, namely day-ahead equilibrium references, hour-level dynamic prices, and instant surge prices. These three stages work together to calculate an optimal price series for each area, which could help to mitigate unbalanced traffic load, unmet service demand and overload in electric power system. The main contribution of this work can be concluded as follows:

- 1) A three-stage incentive scheme has been introduced, which could improve revenue and security of IETS by setting proper price incentives in charging and ride-hailing services.
- 2) A Deep Q-Learning Network is trained in *Stage II* to deal with uncertainties of system states and errors in system modeling, which could outperform model-based methods in terms of robustness and computation load.

The rest of this paper is organized as follows: Section II

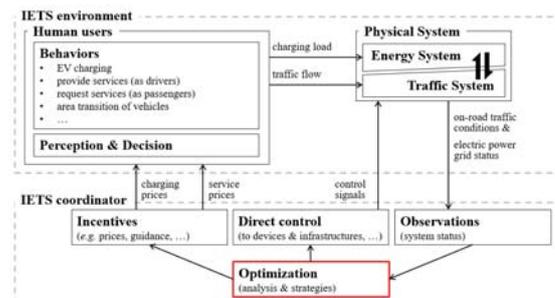


Fig. 1. System structure of IETS with information exchanges

presents the system model and dynamic of urban IETS. In Section III, the three-stage incentive scheme and corresponding formulations are introduced in detail. Next, a case study based on a real-world dataset is designed in Section IV. Finally, Section V concludes the paper.

II. SYSTEM MODELING AND DYNAMICS OF URBAN IETS

In this part, the system modeling of urban IETS is demonstrated with corresponding information exchange and system dynamics.

A. System Structure of Urban IETS

In the setting of urban IETS, mobility service vehicles and PEVs have the potential of being guided or re-allocated for better temporal or spatial distribution [7]. To better exploit this potential, a centralized IETS coordinator (e.g. a municipal operational entity or a third-party dispatching platform) would decide on incentive sent to human users, which could help to yield higher revenue, fewer congestions, and more security.

The system structure of urban IETS is shown in Fig. 1. In general, IETS consist of human users, physical part along with the information/cyber system, which could fit in the category of Cyber-Physical-Human System. Human users, including drivers and passengers, receive broadcast or notifications on prices, navigations, which is taken as their information input. With their preferences and decision criteria, they will make a decision on the next-step movement, thereby have impact on the physical part through driving or charging. The physical part, including electric power grid and road network, has internal dynamic in each sub-system while having human factors as boundary. The cyber system is in charge of data acquisition, transmission, processing and actuation. The coordinator, as part of the cyber system, will solve the optimization problem based on observed system status and send incentive signals back to human users and physical infrastructures. It should be noticed that the methods of this paper are conducted in the optimization sector of IETS coordinator (marked in red).

B. State Space Representation of System Dynamics

To describe the system dynamics in every time step, state-space equations of each sub-system are used and explained in (1)-(3).

$$\mathbf{E}^{\{k+1\}} = E\left(\mathbf{E}^{\{k\}}, u_E\left(\mathbf{T}^{\{k+1\}}\right), \theta_E^{\{k+1\}}, \pi_E^{\{k+1\}}, \mathbf{C}_E^{\{k+1\}}\right) \quad (1)$$

$$\mathbf{T}^{\{k+1\}} = T\left(u_T\left(\mathbf{E}^{\{k\}}\right), \mathbf{T}^{\{k\}}, \theta_T^{\{k\}}, \pi_T^{\{k\}}, \mathbf{C}_T^{\{k+1\}}\right) \quad (2)$$

$$\mathbf{b}^{\{k+1\}} = b\left(u_b\left(\mathbf{E}^{\{k\}}, \mathbf{T}^{\{k\}}\right), \mathbf{I}^{\{k\}}, \pi_b^{\{k+1\}}\right) \quad (3)$$

Electric power grid, as energy system in IETS, is depicted as (1). System state vector $\mathbf{E}^{\{k+1\}}$ (logical state including switching devices and analog data including power flow, power injection and other measurements) of the next time slot $\{k+1\}$ is determined by current state $\mathbf{E}^{\{k\}}$, loads of charging stations due to PEV charging $u_E\left(\mathbf{T}^{\{k+1\}}\right)$ (where $u_E\left(\right)$ represents the inter-system coupling effect as status of PEVs are part of system status vector of traffic system $\mathbf{T}^{\{k+1\}}$), parameters $\theta_E^{\{k+1\}}$ (physical structure, topology and component parameters), intra-system controlling strategy $\pi_E^{\{k+1\}}$ (including equipment-level

controlling strategy of each electric equipment and system-level controlling strategy from independent system operator, ISO or distribution system operator, DSO) and control signal $\mathbf{C}_E^{\{k+1\}}$ from IETS coordinator.

Traffic system has a similar dynamic (2) as energy system, while has a longer time scale as forming new equilibriums takes minutes. Thus, system state vector $\mathbf{T}^{\{k+1\}}$ (including vehicle number of each type in each areas/on each roads) of the next time slot is determined by current state $\mathbf{T}^{\{k\}}$, charging-related indexes $u_T\left(\mathbf{E}^{\{k\}}\right)$ (including service capacity, available charging power, electricity prices), parameters $\theta_T^{\{k\}}$ (e.g. road network structures and capacities) and intra-system controlling strategy $\pi_T^{\{k\}}$ (manual, automatic and intelligent traffic controlling signals with applications of vehicle-to-everything, V2X) control signal $\mathbf{C}_T^{\{k+1\}}$ from IETS coordinator.

Human users' behaviors $\mathbf{b}^{\{k+1\}}$ (actions in terms of charging and service in particular) are affected by a perception and decision process with multiple information inputs. Partially observed information of the physical system $u_b\left(\mathbf{E}^{\{k\}}, \mathbf{T}^{\{k\}}\right)$ (such as road congestion and charging queues), incentives and guidance $\mathbf{I}^{\{k\}}$ from IETS coordinator, strategy $\pi_b^{\{k+1\}}$ which represents preferences and criteria of the perception and decision process. It should be noted that as is discussed in [14], decisions are made with perceptions of utilities of each option, where errors and uncertainties should also be taken into account.

Incentives $\mathbf{I}^{\{k\}}$ sent by IETS coordinator are calculated by the proposed incentive scheme. An ideal version of incentives with perfect knowledge of IETS is shown in (4), where the incentive could be decided based on all the details of physical system and human participants. However, the structures and parameters of each sub-system cannot be easily obtained in real-world scenarios due to ownership, privacy and safety considerations. To deal with this problem, the IETS coordinator will take a step back and make decisions on partially observed system status of the abovementioned state variables with an outcome of incentive \mathbf{I} as is described in (5).

$$\begin{aligned} \mathbf{I}^{\{k\}} &= I\left(\mathbf{E}^{\{k\}}, \mathbf{T}^{\{k\}}, u_E, u_T, u_b, \theta_E^{\{k\}}, \theta_T^{\{k\}}, \pi_E^{\{k\}}, \pi_T^{\{k\}}, \pi_b^{\{k\}}\right) \\ &= I\left(\mathbf{E}^{\{k\}}, \mathbf{T}^{\{k\}}, \mathbf{u}, \theta^{\{k\}}, \boldsymbol{\pi}^{\{k\}}\right) \end{aligned} \quad (4)$$

$$\tilde{\mathbf{I}}^{\{k\}} = \tilde{I}\left(\tilde{\mathbf{E}}^{\{k\}}, \tilde{\mathbf{T}}^{\{k\}}, \tilde{\mathbf{u}}, \tilde{\theta}^{\{k\}}, \tilde{\boldsymbol{\pi}}^{\{k\}}\right) \quad (5)$$

III. THREE-STAGE INCENTIVE SCHEME FOR IETS

In this section, the proposed three-stage incentive scheme and corresponding formulations are introduced in detail. Fig. 2 gives an explicit explanation on the main considerations and outcomes of problems in each stage.

A. Stage I - Day-ahead Equilibrium References

In this stage, the coordinator tends to find the optimal state of energy and traffic system to maximize the overall revenue $R\left(\right)$ of IETS as in (6).

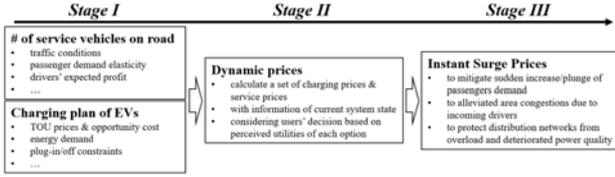


Fig. 2. Problems and considerations in each stage

$$\{E^*, T^*\} = \arg \max \sum_k R(E^{(k)}, T^{(k)}, \tilde{\theta}^{(k)}, \tilde{\pi}^{(k)}) \quad (6)$$

Compared with private-owned vehicles, mobility service vehicles have more flexibility in terms of activated hours, destinations and routes. Similarly, PEVs with energy demand have greater dispatchable potentials than traditional vehicles as charging activities could be temporally and spatially shifted. Here, in particular, T^* is the optimal time series for in-service ride-hailing vehicles while E^* accounts for the charging load of PEVs. Operational constraints in this stage mainly describe the equilibrium of drivers and passengers utilities considering traffic conditions, average waiting time, customer demand elasticity, which could be found in [15].

Charges of taxi services and ride-hailing services in a certain area should remain stable. That means pricing rules in a day should be preset to a sequence of values that could vary in a day but should remain unchanged each day. With this preset price sequence, the IETS coordinator could solve the problem of service market equilibrium with traffic considerations, which will get the optimal time series T^* to maximize service revenue. Accordingly, the optimal charging schedule could then be solved, where EV users decide on charging plan with time-of-use electricity prices as basic monetary cost and hourly profit of services as the opportunity cost of charging. With proper utilization rate in car-hailing services and expected energy consumption in the electricity market, E^* and T^* could then serve as a reference for future stages.

B. Stage II - Hour-level Dynamic Prices

In this stage, the IETS coordinator decides on extra incentives other than preset prices to balance of being guided or re-allocated for better temporal or spatial distribution. However, previous work used a model-based method, which has certain requirements on detailed parameters of human users' decision processes [14], which could not be easily obtained with limited observation on each individual due to privacy concerns. In addition, this multi-step optimization problem is also faced with problems of the curse of dimension and cannot deal with uncertainties in boundary input values and human decisions.

Alternatively, reinforcement learning (RL) method is used to find the incentive strategy for the IETS with uncertainties. The RL agent observes the environment and takes actions to maximize the rewards, while in this problem it set incentives to maximize system overall revenue minus the penalty for security problems such as system overload in the distribution network. With inference to the original system states, RL methods could upgrade the previous observation-based method from the first level to the second level of the Ladder of Causation [16]. Double deep Q-network (Double-DQN) [17], which is a model-free, online, off-policy RL method, is a proper method to get a better incentive strategy with discrete action space.

The algorithm used in *Stage II* is shown in Fig. 3. Observation s is a vector that consists of current number in each state of each type of vehicles (service/non-service, PEV/non-PEV), expected injections in the following time step and the offset compared with the reference value from *Stage I*. Action a is the changes on the base of PEV charging prices and ride-hailing service prices, namely incentives. The incentive is set to be conducted on only one area at a time and is discretized into normal (0), low incentive ($\pm 10\%$) and high incentive ($\pm 30\%$). Reward r is the operational revenue from PEV charging and ride-hailing services minus the penalty related to traffic congestion and overload in electric power system.

Through the training process of the Q-network, DQN agent of the IETS coordinator could gradually find the optimal strategy. At each time step during learning, the agent updates parameters θ of main net $Q(\cdot)$ and explores the action space using epsilon-greedy exploration. With the incentive $a^* = \arg \max_a Q(s, a; \theta)$ sent to IETS environment, the new observation is then stored in a circular experience buffer for future training. Practically, the training process can either be conducted on generated data based on behavioral experiments or be embedded in real-world operations.

C. Stage III - Instant Surge Prices

In real-time operation, sudden changes in users' demand might exceed the service capacity of IETS, which may lead to traffic congestions or even overload risks in power system. To close the gap of unbalanced demand and supply, surge price is the primary lever.

In this stage, prices in the IETS is temporarily intervened through instant surge prices. In ride-hailing markets, riders and drivers responses to a surge price up to several times higher than usual, which attracts more drivers to this area and filter riders with rigid demand and higher willingness to pay

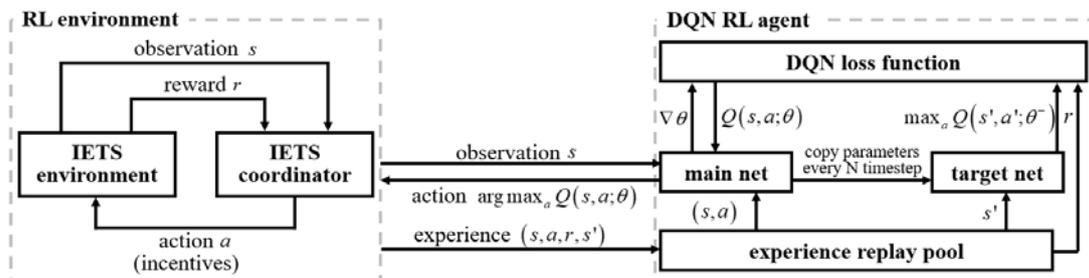


Fig. 3. Structure of DQN algorithm in dynamic pricing of IETS

[11]. In cases of PEV charging, the flexibility of charging could postpone the plug-in of PEVs or deviate vehicles to other charging stations with lower prices. Users' excessive traveling and charging demand would thus be alleviated. Strategies in this stage could be explicitly calculated based on equilibrium analysis, which is discussed in [11] and is not the focus of this paper.

IV. CASE STUDIES

To verify the proposed incentive scheme, a case study on a toy IETS model (Fig. 4) is conducted with behavioral patterns extracted from real-world datasets. Datasheet of case settings can be found at [18].

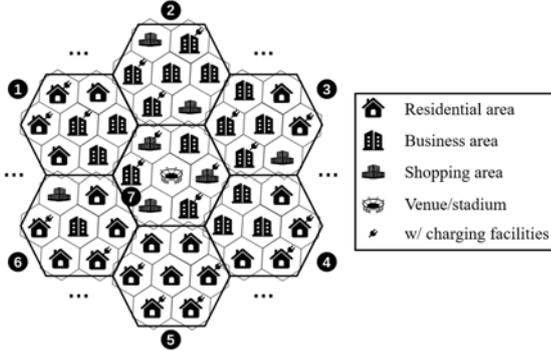


Fig. 4. Area function and charging facilities of IETS

A. Basic Settings of IETS

The IETS model used in this paper consists of 7 areas, each of which has 7 subareas with different functions. Each function areas have different patterns of traveling demand, vehicle density, and electricity baseload. Besides, subareas with plug signs are equipped with charging facilities, where residential subareas have less fast charging poles than public subareas. Data including baseload, traffic conditions, demand of traveling and charging could be found in [18].

B. Problem Settings of Incentive Scheme

The three-stage incentive scheme is solved on a time span of 96 time steps in a day. The 7 areas hold a daily maximum of 6,500 vehicles in total with a peak traveling demand of 7,000 passengers. *Stage I* first solve the day-ahead equilibrium of the ride-hailing market based on predicted traveling demand and traffic conditions. With

knowledge of energy consumption of on-road vehicles, the charging plan of PEVs and expected charging loads are then calculated based on charging prices and expected profit of drivers. As for *Stage II*, with continuous intervene-response process of the IETS (or an equivalent simulation environment as is used in this paper), the coordinator could train the Q-network. In hour-ahead decision, the IETS coordinator could get a price incentive on charging and service prices based on the current state of vehicle distribution and system operational margins. If the IETS's status goes beyond adjustment ability of the first two *Stages*, surge prices in *Stage III* will be activated for both riders and drivers to respond.

C. Results and Discussions

The results of *Stage I* are shown in Fig. 5. Total traveling demand in a day has an obvious tidal pattern with two peaks as in Fig. 5a. Despite the spikes during morning and evening peak hours, the number of riders (Fig. 5b) and active vehicles (Fig. 5c) on hailing platforms has relatively flatter peaks due to rush hour congestions and other public transportation alternatives. In most congested scenarios (e.g. Area 1, 4, 5 around 7:00 and Area 2, 3 around 17:00), trips and in-service vehicles may decline to avoid the jammed traffic. This can be explained by the constraints where trips during the extreme traffic conditions could lead to a higher payment of riders but lower average profit for drivers, which could discourage their participants. One exception happens at the central area (Area 7) during evening peaks, where ride-hailing vehicles gathered for traveling demand to outskirt areas in all directions and thus could be away from congestions.

The charging plan of ride-hailing PEVs is illustrated in Fig. 5d, where a reversed trend could be observed compared with in-service vehicles. Service PEVs tend to get charged either with a prepare-in-advance strategy or a make-up-afterward strategy, i.e. before and after time step with peaks in the number of transactions. This effect shows the flexibility of PEVs' energy storage capacity. As for spatial distribution, as the four areas, namely Area 1, 2, 3 and 7, are supporting all the 7 areas, areas are taking more fast charging loads the closer they are to the areas without fast charging facilities. Meanwhile, as traveling demand is comparably higher in Area 2 than in Area 1 and 3, charging demand is, therefore, less due to higher opportunity cost of not making money from ride-hailing platforms. This demonstrates the coupling effect of how traffic conditions could affect the energy system in IETS.

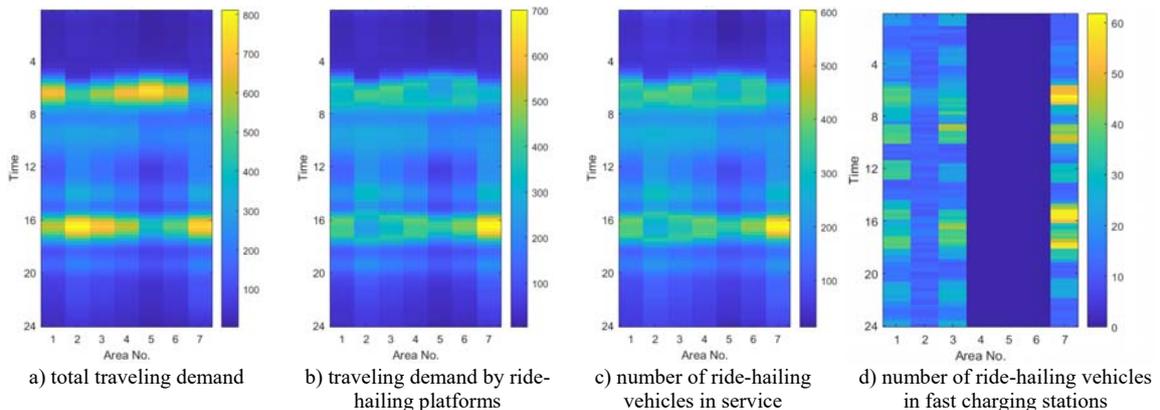


Fig. 5. Results of system equilibrium in *Stage I*

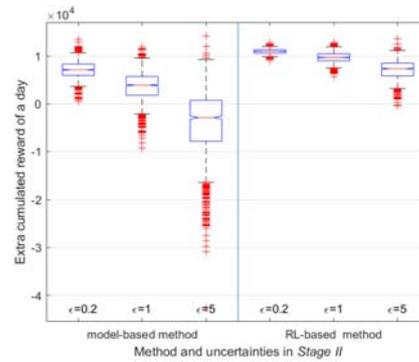
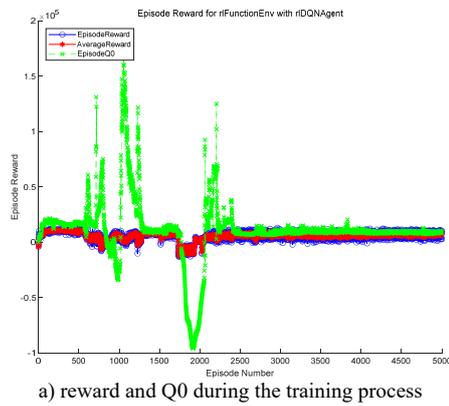


Fig. 6. Training process and performance of double DQN in *Stage II*

Fig. 6 shows the training process and performance of double DQN in *Stage II*. Each training episode is on the time span of one day with 96 time steps. Episode reward is defined to be how much the adjusted prices outperform the original prices in terms of daily cumulated reward r as is described in Section IIIB. It can be seen that the Q-network converged after around 2,500 episodes of trial and learning, which could take 8 hours on a device with Intel Core i7-10510U CPU and 16GB RAM. As for performance in 1,500 Monte Carlo simulations of operational scenarios with different uncertainties in human decision process and predictions in both traffic and charging load, RL-based method is proved to be more robust than model-based method when considering system uncertainties as it can have adaptive price incentives in different situations. On the contrary, model-based methods can only make the optimal decision based on a fixed situation, which will have a deteriorated performance in situations with higher uncertainties and may even be worse than the original prices. Besides, despite the long training time, RL-based method could meet the need of hour-level operational decisions as it has an extremely short time of application compared with model-based method, which needs to solve the optimization problem with newly-received information.

V. CONCLUSIONS

The proposed three-stage incentive scheme is verified to be effective with a day-ahead reference for the equilibrium number of vehicles and a set of pricing strategies to incentivize human users in IETS. On a scale of 24-hour operation, there could be an average of 8% to 12% increase in overall reward with system uncertainties. The proposed method could be used to improve the performance of IETS with higher revenue in the ride-hailing market and charging services together with better security in distribution networks. With higher penetration of PEVs and ride-hailing expected in the future, broader implementation of this work could be benefiting urban areas from a social coordinator's perspective. Future works will be extended to problems with connected or autonomous vehicles, which will be less human-related and controllable.

ACKNOWLEDGMENT

This work is supported in part by Beijing Natural Science Foundation (JQ18008) and in part by The National Key Research and Development Program of China (Basic Research Class 2017YFB0903000).

REFERENCES

- [1] A. Jentzsch, J. Janda, G. Xu, et al., "The Future of Commercial Vehicles - How New Technologies Are Transforming the Industry," Boston Consulting Group Publications, Oct. 2019. <https://www.bcg.com/publications/2019/future-commercial-vehicles.aspx>. [accessed 2020-02-02].
- [2] Daimler AG, "CASE - Intuitive Mobility of Connected, Autonomous, Shared, Electric", <https://www.daimler.com/case/en/>, [accessed 2020-02-02].
- [3] International Energy Agency, "Global EV Outlook 2019," <https://www.iea.org/gevo2019/>, May 2019.
- [4] R. Irle, "Global EV Sales for the 1st Half of 2019," <http://www.ev-volumes.com/country/total-world-plug-in-vehicle-volumes/>, EV Volumes. [accessed 2020-02-02].
- [5] Y. Tian and F. Shen, "China Raises 2025 Electrified-Car Sales Target to About 25%," Bloomberg, Dec. 2019. <https://www.bloomberg.com/news/articles/2019-12-03/china-raises-2025-sales-target-for-electrified-cars-to-about-25>. [accessed 2020-02-02].
- [6] R. Tsang, D. Cai and H. Liu, "The Bumpy Road to Profits in Developing Asia's Mobility Industry," Bain & Company Report. Aug. 2019. <https://www.bain.com/insights/the-bumpy-road-to-profits-in-developing-asias-mobility-industry/>. [accessed 2020-02-02].
- [7] T. Yang, Q. Guo, L. Xu, et al. "Leveraging Urban Integrated Energy-Traffic Systems with Vehicle Re-allocation," 2019 IEEE Power & Energy Society General Meeting (PES GM), Atlanta, GA, USA, Aug. 2019.
- [8] G. Xiong, F. Zhu, X. Liu, et al. "Cyber-Physical-Social System in Intelligent Transportation," IEEE/CAA Journal of Automatica Sinica, 2015, vol. 2, no. 3, pp. 320-333, July 2015.
- [9] H. Zhang, Z. Hu and Y. Song, "Power and Transport Nexus: Routing Electric Vehicles to Promote Renewable Power Integration," in IEEE Trans. on Smart Grid. Early Access.
- [10] W. Shuai, P. Maillé and A. Pelov, "Charging Electric Vehicles in the Smart City: A Survey of Economy-Driven Approaches," in IEEE Trans. on Intelligent Transportation Systems, vol. 17, no. 8, pp. 2089-2106, Aug. 2016.
- [11] B. Hu, M. Hu and H. Zhu, "Surge Pricing and Two-Sided Temporal Responses in Ride-Hailing," in SSRN, Oct. 2019.
- [12] Q. Chen, F. Wang, B.-M. Hodge, et al. "Dynamic Price Vector Formation Model-Based Automatic Demand Response Strategy for PV-Assisted EV Charging Stations," in IEEE Transactions on Smart Grid, vol. 8, no. 6, pp. 2903-2915, Nov. 2017.
- [13] H. Yang, S. Yang, Y. Xu, et al. "Electric Vehicle Route Optimization Considering Time-of-Use Electricity Price by Learnable Partheno-Genetic Algorithm," in IEEE Transactions on Smart Grid, vol. 6, no. 2, pp. 657-666, Mar. 2015.
- [14] T. Yang, Q. Guo, L. Xu. "Guiding Vehicles in Urban Integrated Energy-Traffic System with Price Incentives," CIGRE Chengdu 2019 Symposium, Chengdu, China, Sept. 2019.
- [15] J. Gan, B. An, H. Wang, et al., "Optimal Pricing for Improving Efficiency of Taxi Systems," 23rd International Joint Conference on Artificial Intelligence, Aug. 2013.
- [16] J. Pearl and M. Dana. "The Book of Why: the New Science of Cause and Effect," Basic Books, 2018.

[17] H. Hasselt, A. Guez and D. Silver, "Deep Reinforcement Learning with Double Q-learning," 30th AAAI Conference on Artificial Intelligence, Feb. 2016.

[18] T. Yang, "Datasheet for MITAB 2020," https://drive.google.com/file/d/1bZAZntDPXHCpiPiBKZ_EM-oFONrmKKB-/