ReLeDP: Reinforcement Learning Assisted Dynamic Pricing for Wireless Smart Grid

Peiying Zhang, Chao Wang, Gagangeet Singh Aujla, *Senior Member, IEEE,*
Ranbir Singh Bath, *Member, IEEE*

Abstract—The smart grid must ensure that power providers can obtain substantial benefits by selling energy whereas, at the same time, they need to consider the cost of consumers. So, to realize this win-win situation, the smart grid relies on dynamic pricing mechanisms. However, most of the existing dynamic pricing schemes are based on artificial objective rules or conventional models, which cannot ensure the desired effectiveness. Thus, we apply reinforcement learning (RL) to model the supply-demand relationship between power providers and consumers in a smart grid. The dynamic pricing problem of the smart grid is modeled as a discrete Markov decision process (MDP), and the decision process is solved by Q-learning. Now, the success of any intelligent dynamic pricing scheme relies on timely data transmission. However, the scale and speed of data generation can create several network bottlenecks that can further reduce the performance of any dynamic pricing scheme. Hence, to overcome this challenge, we have proposed an Artificial Intelligence-based adaptive network architecture that adopts software-defined networking. In this architecture, we have used a self-organized map-based traffic classification approach followed by a dynamic virtual network embedding mechanism. We demonstrate the effectiveness of the dynamic pricing strategy supported through adaptive network architecture based on various performance indicators. The outcomes suggest that the proposed strategy is of great significance to realize the sustainability of power energy in the future. Lastly, we discuss various implementation challenges and future directions before concluding the article.

Index Terms—Artificial Intelligence, Adaptive Network, Dynamic Pricing, Reinforcement Learning, Smart Grid.

1 Introduction

The high energy consumption, dissipating levels of fossil fuels, and increase in global carbon emissions have triggered the need for specialized technologies to drive the energy sector. Thus, the smart grid replaces the traditional power grid by utilizing information and communication technologies. Smart grid consists of three main parties, namely power producers, providers, and consumers. Power producers are responsible for the generation of energy through distributed energy resources (such as thermal power, wind power, and solar power), store it, and sell it to the local power providers at wholesale prices. Power providers sell energy to actual consumers (like residential, commercial, or industrial loads) at retail prices based on the realistic energy demand. Figure 1 shows the demand-supply relationship between the power providers and consumers. To provide sustainable energy demand-supply, it is necessary to ensure that power providers can obtain substantial profits by selling energy. At the same time, they need to take into account the cost of consumers to achieve a win-win situation. For this, the smart grid relies on a dynamic pricing mechanism that can provide realistic pricing to consumers.

Although there are several benefits of a smart grid, the amount of uncontrolled data generated by smart devices (like a smart meter, sensors) is one of the key concerns. This data must be transferred, analyzed, and processed in real-time, which can cause serious network bottlenecks. Thus, the wireless network must be flexible and adaptive in line with the changing workload situations. In this line, the following key research questions must be answered.

Research Question 1: How to design and develop realistic dynamic pricing that can provide a win-win situation for all communicating parties in smart grid.

Nowadays, there are several latest technological developments like, artificial intelligence (AI) that can be utilized to provide more timely and realistic power forecasting and dynamic pricing in smart grid. For example, the reinforcement learning approach has provided a promising dynamic pricing scheme in the smart grid as suggested in [1].

Research Question 2: How to design and develop an adaptive and intelligent communication framework that can overcome network congestion in smart grid.

In this context, the following key question must be answered. The traditional network infrastructure may no longer meet the quality of service (QoS) requirements, so AI-based network solutions can be considered to support the communication backbone in the smart grid. The wireless smart grid (WSG) can manage the energy network through intelligent methods (machine learning, deep learning, etc). This can help to make wireless networks more sustainable and reliable energy solutions.

2 Smart Grid Dynamic Pricing

To explain the dynamic pricing of smart grid briefly, we focus on the dynamic pricing between power providers

- *P. Zhang and C. Wang are with the China University of Petroleum (East China). E-mail: zhangpeiying@upc.edu.cn, wangch_upc@qq.com.*
- *GS Aujla is with the Department of Computer Science, Durham University, Durham, DH1 3LE, UK. [Corresponding Author] E-mail: gagangeet.s.aujla@durham.ac.uk.*
- *RS Batth is with the Department of Computer Science, Lovely Professional University, India. E-mail: ranbir.21123@lpu.co.in.*

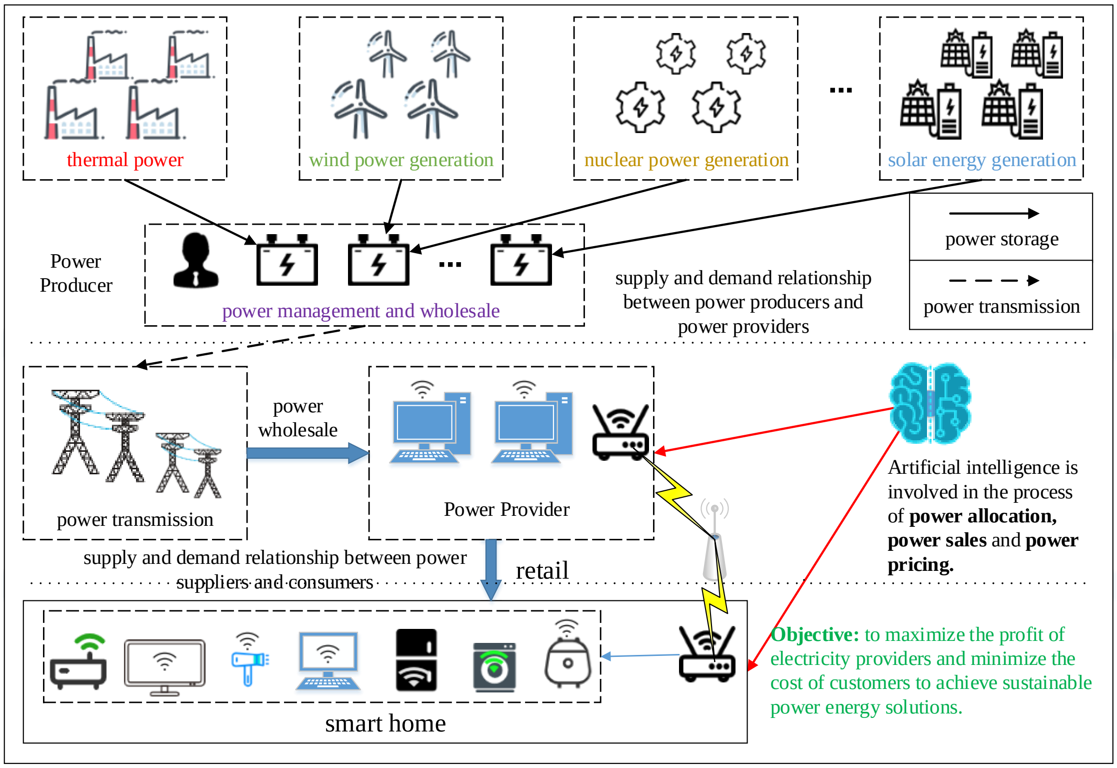


Fig. 1: Artificial intelligence assisted wireless smart grid system.

and consumers, i.e., how power providers dynamically adjust the retail price according to the power consumption status of users, while taking into account the profits of power providers and the cost of users. Power providers can dynamically set the retail price of electricity according to the power demand and potential attitude of consumers. Potential attitude refers to the demand that decreases with the increase of electricity price, which largely determines the common interests of power providers and consumers. Power providers can dynamically adjust the retail price as per consumer's power consumption status to maximize the profits and achieve sustainable power production results, i.e., the goal of smart grid dynamic pricing. A relatively competitive price can empower the energy infrastructure and allow the consumers to make choices and extract value out of its relationship with the energy supplier. The sustainable demand and supply of energy have to face several challenges related to the fluctuations in energy generation and hidden uncertainties in the energy price. This has a direct relation with the energy infrastructure and a viable pricing scheme can empower the energy infrastructure by tacking the uncertainties and establishing a win-win situation between energy providers (including the energy infrastructure) and the energy consumers.

Several state-of-the-art algorithms for dynamic pricing in the smart grid have emerged. For example, Jin *et al.* [2] discussed the pricing and operating strategy of microgrid retailers in the integrated energy system. The authors modeled the pricing and operating strategy as a mixed-integer quadratic programming problem and proposed an optimal scheduling plan that uses the spark difference and promotes the integration of renewable energy. Nojavan *et al.* [3] studied the bilateral contracts and sales prices of smart grid electricity retailers under uncertain conditions. The authors conducted uncertainty modeling based on a random frame of scenarios. Under this model, the energy retailer compared the fixed pricing, time-of-use energy price, and real-time energy price with the sales price in the smart grid.

Wang *et al.* [4] studied the pricing strategy of the electric vehicle (EV) market, and designs a pricing mechanism for charging station operation and joint admission. This strategy maximizes the profits of charging stations by jointly optimizing the charging scheduling, acceptance, and pricing of EVs. Bhatti *et al.* [5] proposed a rule-based power management scheme in the application scenario of EVs. Liang *et al.* [6] studied the interrelationship between the energy-saving and emission-reduction performance of EVs and the charging method. They established a battery exchange pricing model that included swap station operation, stakeholder evaluation, and generator set schedule.

2.1 Key Challenges and Motivation

The existing dynamic pricing algorithms in the smart grid are mainly designed from the perspective of power providers and consumers. The former adjusts the electricity price through the response of the electricity provider to the user's electricity consumption pattern. The latter adopts incentive measures based on users' electricity consumption, encouraging users to reduce energy consumption during peak electricity consumption. Most of the existing proposals on the dynamic pricing of the smart grid are based on artificial objective rules or conventional models such as mixed-integer linear programming and game theory, which cannot ensure the effectiveness of these methods applied to the non-stationary operation system. In addition, these methods are abstracted from the real situation and cannot fully reflect the dynamic changing smart grid environment.

Nowadays, intelligent learning technology is deeply integrated with the smart grid and AI-based communication technologies support the movement of data across various entities. Hence, it is quite interesting to utilize AI technologies [7] to solve the dynamic pricing problem of the smart grid. For example, Ali *et al.* [8] proposed an adaptive model known as *SynergyChain* based on blockchain technology for peer-to-peer energy trading. This approach used smart contracts for transaction management and RL for enhancing performance and profit through a self-adaptive mechanism. Thus, we adopt RL to solve the dynamic pricing problem of WSG. RL focuses on how agents interact with the environment and accumulate high profits. To better apply the RL method to the dynamic pricing problem, it is necessary to model the power service providers and users.

2.2 Problem Modelling

RL is a representative machine learning algorithm, which is inspired by behaviorism and psychology. The implementation of RL is to adjust the strategy through the interaction between intelligent agent and environment, and finally make the agent obtain high revenue, which is the same as the original intention of power provider [9].

Power provider: It is assumed that the power provider purchases energy from the power producer (every hour of the day) at a certain wholesale price, then sells it to consumers at its retail price. The purpose of the power provider is to maximize its profit. At the time t, the retail price $\alpha_{t,n}$ for user n should not be lower than the wholesale price β_t, but should not be higher than the maximum wholesale price $\beta_{t,user}$ stipulated in the agreement with the user.

Users: The power consumption of users is mainly composed of critical load part $e_{t,n}^{crux}$ and potential load part $e_{t,n}^{potential}$. Critical load refers to the power that must be consumed to maintain normal function, such as the power consumption of a residential area. Potential load refers to the demand for energy that changes with a change in the energy price except for critical loads, such as air conditioning and television, which is usually a negative correlation. The decrease of potential load supply will cause user dissatisfaction, which can be expressed by the dissatisfaction function $sat_{t,n}$. This function represents the degree of discomfort experienced by users in reducing the power demand. The degree of dissatisfaction increases with a decrease in the power supply at the load end. The larger the dissatisfaction function, the more sensitive it is to change the potential load.

2.3 Key Contributions

Based on the above discussion, this article proposes, **ReLwDP**, an RL-assisted dynamic pricing framework for WSG supported by AI-based network architecture. The main contributions of this paper are as follows.

- A dynamic pricing algorithm for WSG assisted by RL is designed. We establish a model for smart grid elements and define the dynamic pricing problem of WSG as a discrete MDP. Then the typical Q-learning algorithm is used to solve the problem.
- We propose an AI-assisted WSG network architecture for sustainable demand and supply management in a smart grid.
- The proposed work is validated using a case study to show how it can intuitively reflect the dynamic changes of the relationship between user demand and energy price, as well as the supply-demand relationship between different power grid entities. Also, we validate the proposed approach based on accuracy and communication latency.

3 Reinforcement Learning in Smart Grid

In this section, we discuss the preliminaries related to **ReLwDP**. Moreover, the dynamic pricing of the smart grid is defined as a discrete MDP from the perspective of RL, and the application of Q-learning to solve the decision-making process is also discussed comprehensively.

3.1 Reinforcement Learning

Using intelligent learning technologies to solve the pricing problem of WSG has significant advantages. First, RL is an adaptive learning method, hence RL agents can fully learn the uncertainty and flexibility of the smart grid, to quickly adapt to the dynamic power demand of consumers. Secondly, RL is a model-free intelligent learning algorithm and can perform an efficient learning process without relying on a hand-made smart grid model. This is achieved through constant interaction between the agent and the environment. We treat the power provider as an RL agent and the user as an environment for interacting with the agent. The retail price of electricity presented by the power provider to the user acts as an agent's action on the environment. The energy fee paid by the user to the power provider serves as a reward. The status is the user's power demand and power consumption. The process of exploring how to make the RL agent accumulate to obtain substantial rewards is the process of maximizing the profits of the power provider.

3.2 Markov Decision Processing

Dynamic pricing of the smart grid is a dynamic decision problem in a random environment. Markov decision is faced with optimization of the stochastic dynamic system controlled by a series of decision-making. So, based on the modeling of power suppliers and users, we define the dynamic pricing of WSG as discrete MDP. The typical step of a Markov decision is that at each time step, the agent takes specific actions in a certain state (environment). Next, the old state will change to a new state. According to the effect of the state, the environment will give the agent corresponding rewards. In contrast, in discrete MDP, agents cannot directly transform the state, but they must transform the state as per legitimate state distribution based on the observation set. The typical MDP and discrete MDP are shown in Fig. 2.

Under RL method, the system parameters of Markov

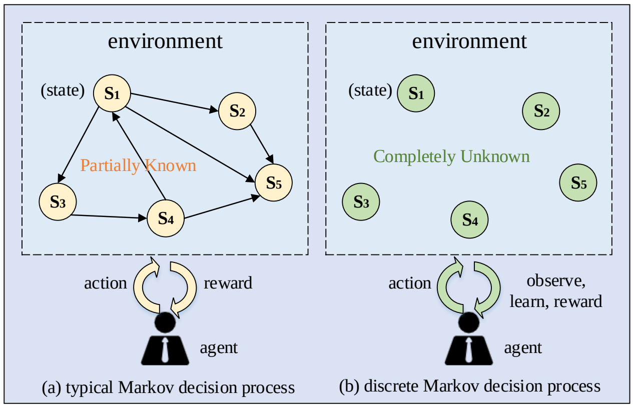


Fig. 2: Typical Markov decision process and discrete Markov decision process.

decision model mainly include action $A(\alpha_{t,n})$, state $S(P_{t,n}, p_{t,n})$ and reward $R(r_{t,n})$. t is a time period (in whole hours) of the day. $\alpha_{t,n}$ is the retail price of electricity charged by the power provider to customer $n$during time period t. $P_{t,n}$ represents the power demand of user n in time period t. $p_{t,n}$ is the actual power consumption of user n in time period t. $r_{t,n}$ means that the power provider collects the electricity charge of user n in time t.

3.3 Q-learning

Q-learning is a model-free RL method, which can learn based on experience in an unknown environment and explore the optimal strategy. So, it can be applied to discrete MDP to find the optimal action strategy. The main parameters of the Q-learning model include action $A(\alpha_{t,n})$, state $S(P_{t,n}, p_{t,n})$ and reward $R(r_{t,n})$, and an agent participates in the Q-learning process. The agent will act in a specific state, i.e., impose a certain operation on the environment. The environment will change from the current state to the next stage, at which time the agent will also receive a reward. The ultimate goal of the agent is to maximize the accumulation of reward signals.

The application of Q-learning in the smart grid has been proved to be effective as it contains a large number of private information. For example, Shimotakahara *et al.* [10] studied the resource scheduling problem of smart grid and proposed a resource allocation algorithm based on multi-agent Q-learning. The algorithm realized the function of a long-term evolution device to device communication agent to generate orthogonal transmission scheduling beyond network coverage. The results showed that the packet loss rate was significantly reduced and the throughput of network traffic was effectively improved. Wang *et al.* [11] proposed a two-stage load modeling framework of the comprehensive load model of the Western Electricity Coordination Committee based on the double depth Q-learning network.

Based on the excellent decision performance of the Q-learning method, the dynamic pricing problem which is modeled as discrete MDP is suitable. Q-learning continuously updates the retail price of electricity in the way of iterative calculation according to the relevant parameters such as electricity demand and wholesale price of energy. Then, the epsilon greedy strategy is used to select the random optimal retail price within the optional retail price, i.e., the maximum Q-value. It should be noted that the Q-value is not fixed, it can be updated with the iteration. After determining the retail price of a period, the power provider will get immediate revenue, and then the next iteration process will be carried out according to the user requirements of the next period. Through the decision-making process of Q-learning, the retail price with the maximum profit to power suppliers can be achieved and the cost of power consumption of consumers can be reduced.

4 ReLeDP: Proposed Case Study

Based on the above preliminaries, we design **ReLeDP** that comprises two aspects, a) RL-based dynamic pricing mechanism, and b) AI-based WSG network architecture supporting the data communication between different WSG entities. The proposed **ReLeDP** is validated in a realistic case study to understand the effectiveness and performance.

4.1 Dynamic Pricing Algorithm

In the WSG environment, there are supply and demand relationships between two pairs, i.e., a) power producers and power providers, and b) power providers and consumers. We assume that the wholesale price of electricity sold by power producers to power providers is fixed in each period t, and only the dynamic pricing problem between power suppliers and consumers is considered. Therefore, the RL method is applied to the supply-demand relationship between power providers and consumers. We assume that the wholesale price of electricity is fixed so that we can study the changing relationship between the retail price of electricity and consumer demand. As RL agents, power providers participate in the interaction process with the environment and take consumers as the environment. The interactive process of the two is the process of dynamic decision-making of optimal electricity prices. Among them, the energy retail price set by the power provider is the action imposed on the environment by the agent. The electricity fee charged by the power provider to the user is regarded as the reward, and the power demand and actual power consumption of the user are regarded as the state of the environment. RL-assisted dynamic pricing algorithm is essentially a process based on look-up table and updating decision-making strategy.

The running time of the Q-learning algorithm is set to $T = 1$ to $T = 24$ to represent 24 hours in a day. The agent (power provider) continuously explores the optimal action (retail price) iteratively. The agent can get the current status (user demand, power consumption of the last period, etc.) from the environment. Because the random choice of retail price may not get a globally optimal solution, we choose the optimal retail price between the wholesale price and the maximum retail price through the ϵ greedy strategy. The advantage of the ϵ greedy strategy is that it can assign a probability evenly to a group of random retail prices. In each iteration, the retail price with the largest probability is selected as the optimal retail price. The current optimal retail price can be continuously updated through iteration. Since the reward of the agent in the next iteration may be greater than the current optimal retail price, the optimal retail price needs to be changed to a larger one. If there is no greater reward through iteration, the current maximum retail price is still used as the optimal retail price. The algorithm complexity is determined by the number of iterations I and the running time slot T, so the algorithm complexity can be expressed as $O(I \cdot T)$.

4.2 AI-assisted Adaptive WSG Network Architecture

The AI-assisted WSG network architecture is shown in Fig. 3. It includes three layers described below.

Data Layer: This layer is composed of real power scenarios, mainly including power producers, power providers, and consumers. This layer generates a large amount of data that is forwarded using the network layer and utilized by the price control layer to realize dynamic pricing.

Network Layer: This layer supports adaptive network architecture for different application scenarios, and realizes the communication, control, and routing.

Price Control Layer: The price control layer is the layer where AI technology plays a key role. It receives the power user scenario data uploaded from the data layer forwarded through the network layer, conducts integrated training on it, and makes the optimal power pricing strategy.

The communication between different parties in WSG is based on an AI-based adaptive routing scheme using a software-defined networking architecture (SDN), network function virtualization [12] and virtual network embedding. In our previous work [13], we tried to optimize the QoS on the Internet of Vehicles scenario. However, this work is not supported by any intelligent technique, that limits its flexibility. So, we extend it and utilize a self-organized map (SOM)-based algorithm to design an adaptive network architecture. This approach work in two phases elaborated as below.

Phase 1: SOM-based classification based on priority In phase 1, SOM (based on the artificial neural network) is used to differentiate the incoming traffic based on their inherent characteristics. This approach classifies the data traffic into high priority, guaranteed forwarding, and best-effort forwarding based on different packet metrics, i.e., port number, type of the service (ToS), time to live (TTL), etc. SOM creates a vector set that correlates the similar input patterns. In the training step, the weights are assigned tentatively by eigenvectors to expedite the learning process. In the mapping step, the first step is to initialize the vector weights from a randomly selected sample vector. Then a map search is initiated to find a weight that relates to the sample in the best manner. Here, the similar weights and best machine unit are selected using euclidean distance, and the neighboring weights are rewarded as they are similar to the randomly selected weight. This process classifies data traffic into different priority-based queues.

Phase 2: Resource-aware virtual network embedding scheme The phase is responsible for assigning the network resources to the data traffic using the concept of virtual network embedding. The global network controller receives the request and divides it into subgraphs, that are sent to the local controller to check the resource requirement condition. It selects the candidate nodes that maximize the condition and forwards its information to the global controller for pre-mapping. The global controller selects the pre-mapping scheme and issues the same to the local controller. Now, the local controller checks the pre-mapping condition and uploads the candidate node information to the global controller (if it is complete), that the physical link for embedding.

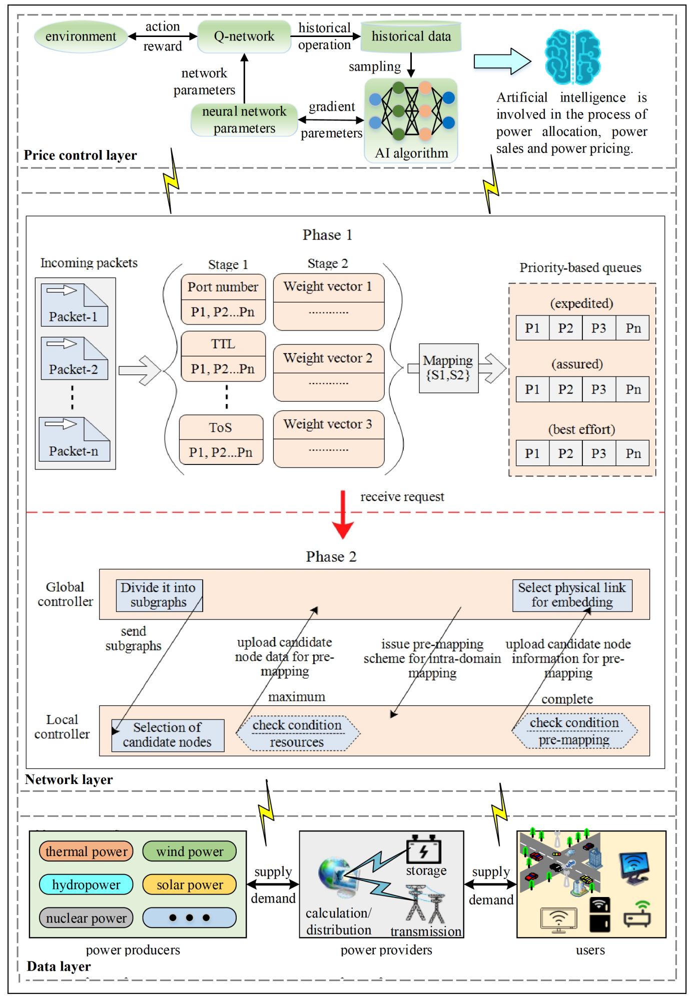
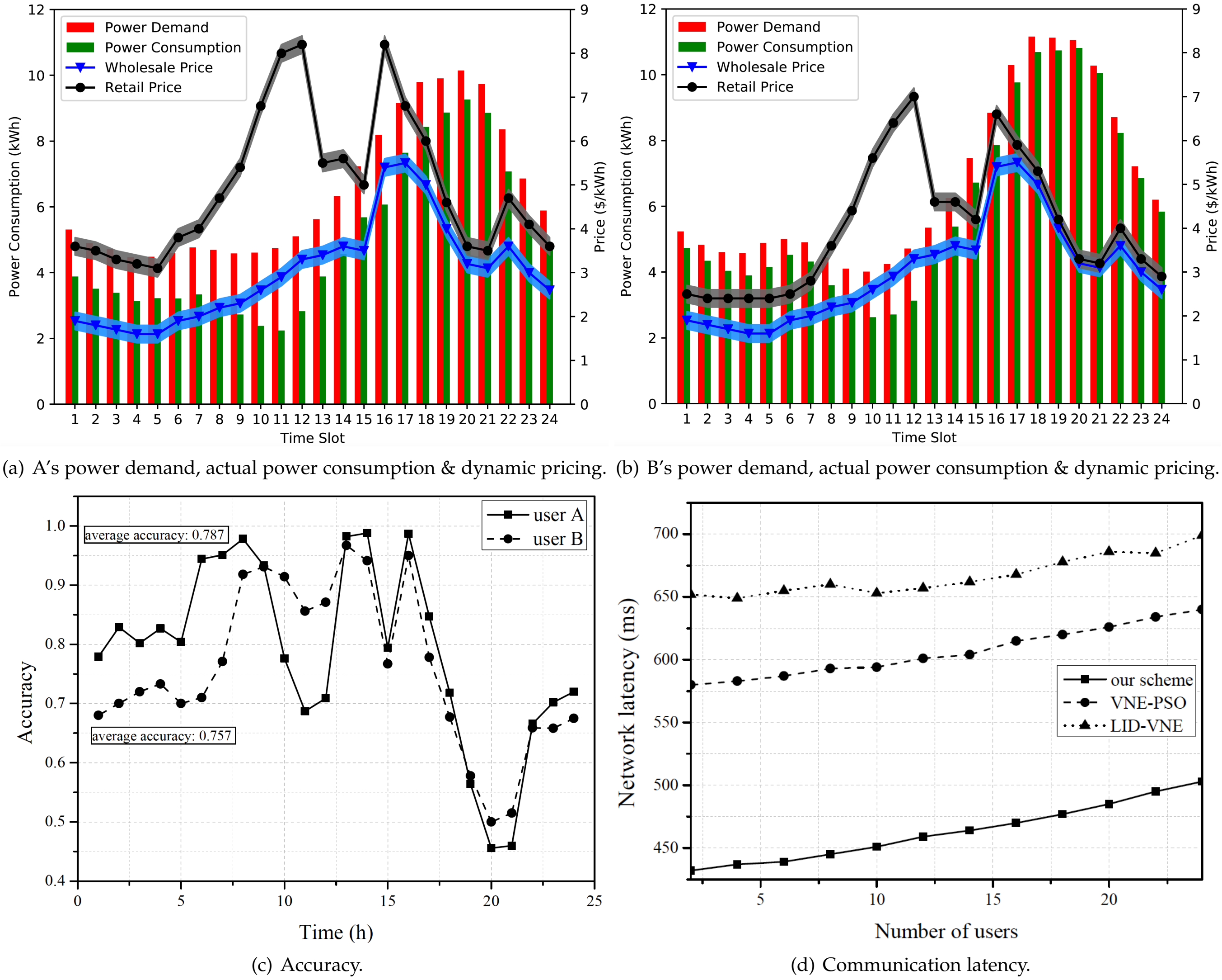


Fig. 3: AI-assisted WSG network architecture.

4.3 Experimental Validation and Analysis

To reflect the effect of the proposed work, we conduct lightweight simulation experiments based on one power provider and two consumers. The difference between the two consumers is their dissatisfaction degree, i.e., dissatisfaction coefficient of user A is $sat_{t,A} = 0.3$, and user B is $sat_{t,B} = 0.8$. The power demand of two consumers is taken from the actual demand of two consumers in a certain area in a day. Figures 4(a) and 4(b) show the power demand, power consumption, wholesale price of power, and retail price of power for users A and B at different times in a day.

It can be seen from the figures that the changing trend is the same, which is in line with the actual situation. First, the retail price of the power provider should be higher than the



(a) A's power demand, actual power consumption & dynamic pricing. (b) B's power demand, actual power consumption & dynamic pricing.

(c) Accuracy. (d) Communication latency.

Fig. 4: Experimental results.

wholesale price to ensure sufficient profit, but not greater than the maximum retail price agreed with the user. An obvious characteristic of the two graphs is that the power consumption of consumers keeps at a low level from 1:00 to 11:00, which indicates that the potential demand of users is high at this time. Because the user's power demand is low at this time, the power provider will sell power at a lower price to encourage users to increase power consumption. It is worth noting that around 12:00 in Figure 4(a), the retail price of electricity has fluctuated significantly. The reason may be that the power provider wants to increase the retail price to obtain high profits. The increase in retail price may lead to the reduction of power consumption, so the users' demand for electric energy is not high at this time. Every day from 12:00 to 20:00, the peak energy consumption period is witnessed. At this time, the price of power is high, which indicates that the potential demand of users is low and is not sensitive to the change of retail price. In addition, it can be observed that the average retail price in Figure 4(a) is slightly higher than that in Figure 4(b) because the dissatisfaction coefficient of user A is lower than that of user B. The large-scale fluctuation of the retail price will cause strong dissatisfaction of user B. for the common interests of power providers and users, the overall price of user B is maintained at a relatively stable level. On the contrary, user a is not sensitive to the change of retail price and the retail price may fluctuate greatly. Overall, the simulation experiment can effectively reflect the actual power consumption of consumers.

It may be noted that we first calculate the average value of the sample data, and then use the sample error rate to calculate the confidence interval of the predicted price. From the confidence interval, the retail price and wholesale price of electricity fluctuate in a reasonable range.

4.3.1 *Accuracy*

We test the accuracy of the retail electricity price predicted by the proposed scheme. By comparing with the actual electricity price, we get the accuracy of the forecast result, as shown in Figure 4(c). Although the dissatisfaction factors of user A and user B are different, the prediction results have achieved high accuracy. The average accuracy rate of the predicted retail electricity price of user A is 78.7%, and user B is 78.7%, and the highest prediction accuracy rate can reach more than 98%. Therefore, the proposed algorithm has good accuracy in predicting electricity prices.

4.3.2 Network Latency

The proposed algorithm optimizes the latency from the perspective of network resource allocation. By setting up an objective function to select the network links with a sufficiently small latency for communication, effectively guarantees the service needs of the users. As shown in Figure 4(d), with the increase of the number of users, the latency optimization effect of our proposed scheme is the best, while the latency of the other two schemes (VNE-PSO [14] and LID-VNE [15]) is much higher. So, the proposed adaptive network architecture can effectively guarantee communication quality.

5 Challenges and Future Research Directions

After analysis of the different perspectives of dynamic pricing in smart grid, the following challenges, and future research directions are proposed as below.

- **Algorithm deployment:** From a theoretical perspective, we propose a dynamic pricing strategy for WSG based on RL and provided a lightweight validation. The deployment and application effects of this strategy in a large-scale real-world smart grid application are still not validated or tested.
- **Scalability:** The impact of scaling the number of providers and consumers should be analyzed as it has a strong impact on the performance of any proposed framework in a realistic environment.
- **Multi-power provider competition:** In the real smart grid environment, multiple power providers compete with each other. Competing among multiple power providers may have varying degrees of impact on the wholesale price and retail price of power, which is not reflected in the algorithm.
- **Integration with other wireless technologies:** With the widespread popularity of advanced applications such as cloud computing and big data, it is necessary to consider the integration of smart grid pricing and other wireless network technologies.
- **QoS and QoE:** The ultimate goal of reasonable pricing of electric energy is to better serve users, and it is necessary to improve users' QoS and QoE. Therefore, it is necessary to consider how to use QoS and QoE metrics to enhance the performance of existing work.
- **Other supply and demand relations:** We only consider the supply and demand relationship between power providers and users. In the actual power market, there is also a supply-demand relationship between power producers and power providers. Next, we will consider the impact of supply and demand between different entities on electricity pricing.

6 Conclusion

The sustainable energy system supported by adaptive network technology is expected to replace the traditional energy infrastructure. In this article, we proposed a reinforcement learning-assisted dynamic pricing scheme for smart grids. Additionally, we proposed an AI-based adaptive network architecture that ensures the timely movement of data generated in the smart grid. Through the experiment and simulation, we have realized the dynamic price-setting based on the user's electricity demand and consumption. The experimental results show that the dynamic pricing scheme is effective in supporting power energy sustainability, so it can achieve a win-win effect. We also analyzed the impact of the proposed AI-based network architecture and the outcomes show a significant reduction in the network latency in contrast to other comparative variants.

Acknowledgement

This work is partially supported through Startup Fund provided by the Durham University, UK. This work is also partially supported by the Major Scientific and Technological Projects of CNPC under Grant ZD2019-183-006, partially supported by Shandong Provincial Natural Science Foundation under Grant ZR2020MF006, and partially supported by "the Fundamental Research Funds for the Central Universities" of China University of Petroleum (East China) under Grant 20CX05017A.

References

[1] R. Lu, S. H. Hong and X. Zhang, "A Dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach," *Applied Energy*, vol. 220, pp. 220-230, 2018.

[2] M. Jin, W. Feng, C. Marnay and C. Spanos, "Microgrid to enable optimal distributed energy retail and end-user demand response," *Applied Energy*, vol. 210, pp. 1321-1335, 2018.

[3] S. Nojavan, K. Zare and B. Mohammadi-lvatloo, "Optimal stochastic energy management of retailer based on selling price determination under smart grid environment in the presence of demand response program," *Applied Energy*, vol. 187, pp. 449-464, 2017.

[4] S. Wang, S. Bi, Y. J. Zhang and J. Huang, "Electrical Vehicle Charging Station Profit Maximization: Admission, Pricing, and Online Scheduling," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1722-1731, 2018.

[5] A. R. Bhatti and Z. Salam, "A Rule-based Energy Management Scheme for Uninterrupted Electric Vehicles Charging at Constant Price Using Photovoltaic-grid System," *Renewable Energy*, vol. 125, pp. 384-400, 2018.

[6] Y. Liang and X. Zhang, "Battery swap pricing and charging strategy for electric taxis in china," *Energy*, vol. 147, pp. 561–577, 2018.

[7] I. AI Ridhawi, S. Otoum, M. Aloqaily, and A. Boukerche, Generalizing AI: challenges and opportunities for plug and play AI solutions. IEEE Network, vol. 35, no. 1, pp. 372-379, 2020.

[8] F.S. Ali, O. Bouachir, Ö. Özkasap, and M. Aloqaily. "SynergyChain: Blockchain-Assisted Adaptive Cyber-Physical P2P Energy Trading." IEEE Transactions on Industrial Informatics, vol. 17, no. 8, pp. 5769-5778, 2020.

[9] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. Chen and L. Hanzo, "Machine Learning Paradigms for Next-Generation Wireless Networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98-105, 2017.

[10] K. Shimotakahara, M. Elsayed, K. Hinzer and M. Erol-Kantarci, "High-Reliability Multi-Agent Q-Learning-Based Scheduling for D2D Microgrid Communications," *IEEE Access*, vol. 7, pp. 74412-74421, 2019.

[11] X. Wang, Y. Wang, D. Shi, J. Wang and Z. Wang, "Two-Stage WECC Composite Load Modeling: A Double Deep Q-Learning Networks Approach," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4331-4344, 2020.

[12] Z. Alomari, M. F. Zhani, M.Aloqaily, and O. Bouachir. ”On minimizing synchronization cost in NFV-based environments.” In 2020 16th International Conference on Network and Service Management (CNSM), pp. 1-9. IEEE, 2020.
[13] P. Zhang, C. Wang, G. S. Aujla, N. Kumar and M. Guizani, “IoV Scenario: Implementation of a Bandwidth Aware Algorithm in Wireless Network Communication Mode,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15774-15785, 2020.
[14] R. Geng and H. Lu, “Multi-domain SDN virtual network mapping algorithm,” *A Small Microcomputer System*, vol. 37, no. 12, pp. 2593-2597, 2016.
[15] D. Dietrich, A. Rizk and P. Papadimitriou, “Multi-domain virtual network embedding with limited information disclosure,” *2013 IFIP Networking Conference, Brooklyn, NY*, pp. 1-9, 2013.

BIOGRAPHIES

Peiying Zhang (zhangpeiying@upc.edu.cn) received his Ph.D. degree in information and communication engineering from the State Key Laboratory of Networking and Switching Technology at Beijing University of Posts and Telecommunications. He is currently an associate professor with the College of Computer Science and Technology, China University of Petroleum (East China). His research interests include artificial intelligence for networking, network virtualization, and future network architecture.

Chao Wang (wangch_upc@qq.com) is a graduate student in the College of Computer Science and Technology, China University of Petroleum (East China). His research interests include network virtualization and network artificial intelligence.

Gagangeet Singh Aujla [S’15, M’18, SM’19] (gagangeet.s.aujla@durham.ac.uk) is working as an Assistant Professor of Computer Science at Durham University, Durham, UK. Prior to this he was a Post Doctorate Research Associate with the School of Computing, Newcastle University, UK. He received his Ph.D in Computer Science from Thapar University, India in 2018. He received the 2018 IEEE TCSC Outstanding Ph.D. Dissertation Award and 2021 IEEE System Journal Best Paper Award, which recognized his leading expertise in the application of scalable and sustainable algorithms for cloud data centers, SDN and smart grid. He is Area Editor of Ad hoc Networks Elsevier.

Ranbir Singh Batth (ranbir.21123@lpu.co.in) is working as an Associate Professor in the School of Computer Science and Engineering and he also serves as a coordinator for International relations, at Lovely Professional University, Punjab, India. In 2018. He has received his M.Tech. degree in Computer Engineering from Punjabi University, Patiala. His research interests include Wireless Sensor Networks, Cloud Computing, Network Security, Adhoc Networks, Machine Learning, Deep Learning, Wireless Communications and Mobile computing.