Optimal demand response strategy of commercial building-based virtual power plant using reinforcement learning

Tao Chen1 | Qiushi Cui2 | Ciwei Gao1 | Qinran Hu1 | Kexing Lai3 | Jianlin Yang4 | Ran Lyu5 | Hao Zhang6 | Jinyuan Zhang7

1 School of Electrical Engineering, Southeast University, Nanjing, China
2 Department of Electrical and Computer Engineering, Arizona State University, Tempe, Arizona, USA
3 Optimization Division, Hitachi ABB Power Grids, San Jose, California, USA
4 Wind Power Industry Innovation Center, State Power Investment Corporation Limited’s (SPIC), Shanghai, China
5 Economic Technology Research Institute, State Grid Corporation of China-Shanghai Company, Shanghai, China
6 Department of Technology and Management, Tengtian Energy Saving Technology Corporation Limited’s, Shanghai, China
7 Administration Office, Shanghai Huangpu District Development and Reform Commission, Shanghai, China

Abstract

In this paper, the optimal demand response strategy of a commercial building-based virtual power plant with real-world implementation in heavily urbanised area is studied. Instead of modelling the decision-making process as an optimisation problem, a reinforcement learning method is used to seek the optimal strategy, which could update its performance with minimal manpower manipulation. Specifically, the data collection from several commercial buildings, including hotel, shopping mall and office, in Huangpu district, Shanghai city is analysed to deploy the demand response program. Compared with the conventional demand response strategy based on optimisation, the learnt strategy does not rely on the forecasting information as input and could adapt to the changing demand response incentive automatically. It may not produce the best result every time, but can guarantee the benefit in a non-deterministic way in long-term operation. The real-world deployment of the Huangpu virtual power plant involving hardware and software platform is also introduced, as well as its future development projection.

1 | INTRODUCTION

The energy consumption of commercial buildings occupies a significant portion of the total energy sector [1], especially in urbanized areas with heavy demand for air conditioners, lighting and electric vehicle charging. It is not only a challenge for distribution network operation and power supply, but also an opportunity for power system ancillary service [2] and demand response (DR) programs [3]. If these buildings could accept the external DR incentive via aggregation and automatic control, adjusting their energy consumption or load curve according to the optimal scheduling of economic benefit and demand requirements, both building itself and distribution network could benefit from such interactions [4]. Thus, the popular concept of virtual power plant (VPP) [5, 6] is by natural an ideal tool to management the commercial buildings as a whole to participate in various DR programs, with the help of advanced information and communication technology (ICT) platforms [7]. The optimal operation and demand response strategy of such a VPP should consider the electricity price, extra incentive and residents’ utility at the same time. Meanwhile, the decision-making process of scheduling the energy consumption could be in an autonomous and intelligent manner with minimal manpower involvement and programming for easy deployment with limited resources.

In recent years, many methods have been proposed to help operate VPP in an economic optimal or most energy efficient way, which mainly focus on the internal energy management...
and low-level equipment operation. Most works [8–10] use various optimisation methods to model the VPP as an equivalent energy management system (EMS) and usually depend heavily on the forecasting information to deal with the specific decision variables. For instance, Pandžić et al. [10] propose a mid-term dispatch optimisation model for VPP, which maximises the weekly virtual power plant profit subject to the long-term bilateral contracts and technical constraints. Liang et al. [11] propose a new framework for the optimal VPP energy management problem considering risk metrics and minimising the VPP operating cost while maintaining the power quality of the system. A few works [12] propose to use learning-based methods to handle the VPP operation problem, but without consideration for external demand response signal, and cannot enable a simple implementation with limited computational resource and data processing capability.

By utilizing the new paradigm of data-driven and learning-based method, specifically reinforcement learning (RL) techniques that are widely used in various decision-making problems in many fields (e.g. video game playing [13], electricity market bidding [14], battery storage control [15], electric vehicle charging [16]), the VPP economic operation problem could be solved more easily and efficiently in a self-adaptive manner. In recent years, actually reinforcement learning has already been extensively studied in the power system area, including load frequency control [17], demand response [18], community energy management [19], resilience enhancement [20] and many other applications over the power transmission and distribution. In [18], it proposes a real-time autonomous residential demand response management strategy to reduce the household’s energy costs, which is very similar to a scaled-down VPP operation problem that aggregates the control actions for various flexible loads. In [21], it also uses multi-agent reinforcement learning to solve an energy management problem, meanwhile leveraging the extreme learning machine for prediction of electricity price and solar photovoltaic generation. In this way, the demand response on consumption side and distributed energy resources on generation side are both considered for the coupled fine-tuned energy management system. However, most of these works are purely conceptual design of the VPP operation or energy management framework with simulation data, unlike [22], which lack the field experiments and real-world deployment information. Instead, this paper aims to present a realistic commercial building-based VPP pilot project in Huangpu district, Shanghai city, to illustrate the platform structure and automatic decision-making process.

In this paper, our contribution is threefold: (1) provide the construction method and ICT platform structure for commercial buildings-based virtual power plant; (2) demonstrate the experiments of the demand response program for commercial buildings in real-world deployment with more than ten building participants; and (3) propose a reinforcement learning framework to formulate the optimal demand response strategy of commercial-building based virtual power plant with adaptive and updating solution.

FIGURE 1 The information and communication architecture of a real-world commercial building-based virtual power plant (VPP) in Huangpu district, Shanghai city

2 COMMERCIAL BUILDING-BASED VIRTUAL POWER PLANT

This paper describes the architecture of a real-world commercial building-based virtual power plant (VPP), as well as its communication channels with various monitoring and data collection functionalities as shown in Figure 1. The basic operation principles and the most important features are explained here. More details regarding the realistic deployment, field experiments, software platform and operation performance are introduced later in Section 4 and Section 5.3. The proposed reinforcement learning method will be used on top of such realistic VPP operation data samples with continuous status update.

The idea of commercial building-based VPP is out of the realistic need for efficient management tool that could exploit the huge energy saving or demand response (DR) potentials in hundreds or even thousands of commercial buildings densely situated in heavily urbanized region. Following this idea, “Shanghai Huangpu VPP Demand Response Pilot Program”, introduced in this paper, aims to (1) connect all the available buildings with large energy potential to a unified VPP platform; and (2) provide external DR incentive to each connected building for load rescheduling. Technically, as shown in Figure 1, each building infrastructure has been upgraded with control unit hardware, in which the learning-based operation logic could be embedded, and various software services that enable edge-computing like data processing, storage and immediate analysis. For example, the MQTT service module could subscribe DR incentive and
electricity price information, and publish directly to the virtual machine involving instant control actions.

The fundamental operation principles of such a real-world commercial building-based VPP are summarised as follows:

- Adjusting the energy consumption is the only way to virtually export or import energy, equivalent to “power generation”, but could incorporate the active power source, such as distributed renewables, in the future expansion;
- Each building is taken as a single VPP unit, analogous to “power plant unit”, and self-contained with the control actions, mostly for the HVAC units;
- Upper-level VPP platform is only responsible to broadcast the external DR incentive fairly to each building, without direct control authority of the low-level devices;
- The economic incentive with DR received by each building is complete by default, but also according to the overall performance and local DR policy;

3 | METHODOLOGY AND FORMULATION

3.1 | Markov decision process

The feature of reinforcement learning framework that distinguishes itself from other learning-based methods is the combination of state status update and control action execution using Markov decision process (MDP) [23]. Many other learning-based methods, like clustering and classification, are more inclined to be pattern recognition problems without direct decision-making process [24]. An MDP is usually described by state space \( S \), action space \( A \), reward function set \( R \), state transition probability matrix \( P \) and a discount factor \( \gamma \in [0, 1] \).

Thus, the target system evolution with state status update and control action choice could be systematically modelled in the compact tuple \((S, A, R, P, \gamma)\). It is also assumed that an MDP remains the Markov property, which implies the transition probabilities of a state are only affected by the previous one step with no memory.

Due to the fact that reinforcement learning has been introduced and applied in many recent power system operation problems and become more and more popular, which could be almost taken as common sense, the agent-environment interaction and MDP mapping relationship are just briefly described in Section 3.2 only with emphasis on the specific formulation of VPP demand response operation modelling using MDP. The fundamental principles of MDP and agent-environment interactions could be found in our previous works [25, 26] and other in-depth discussion [23, 27].

3.2 | Modelling VPP operation using markov decision process

The key of modelling VPP operation using MDP model is to map the VPP state status and control action associated with demand response to the aforementioned compact tuple \((S, A, R, P, \gamma)\). For reward function \( R \) in VPP modelling, it is defined as the combination of concrete financial incentive and subjective utility of the studied commercial-building as a whole.

We use the utility function in [28] that is customised for different commercial building and satisfies non-decreasing but marginal non-increasing characteristics [29].

\[
U_i(p^L_{t,i}) = \begin{cases} 
\omega_i p^L_{t,i} - b_i (p^L_{t,i})^2 & p^L_{t,i} \leq \omega_i / 2b_i \\
\omega_i / 4b_i & p^L_{t,i} > \omega_i / 2b_i 
\end{cases} \tag{1}
\]

where, \( U_i(\cdot) \) indicates the utility function and \( p^L_{t,i} \) is the load level of \( i \)th building at time \( t \); \( \omega_i \) and \( b_i \) are the utility coefficient balancing the quantitative gain. For simplification, since then we drop the subscript \( i \) indicating different building and formulate the total utility or monetary economic benefit (EB) for each individual building as the following:

\[
EB = \sum_i (-b_i p^L_{t,i}) + \sum_i \eta^{DR} f^{DR}(p^L_{t,i} - p^L_{t observers}) + U_i(p^L_{t,i}) \tag{2}
\]

\[
f^{DR}(p^L_{t,i} - p^L_{t observers}) = \begin{cases} 
\frac{\eta^{IR}}{2^\gamma} \max \left( 0, \frac{p^L_{t,i} - p^L_{t observers}}{\omega_i} \right) & \text{peak shaving} \\
\frac{\eta^{IR}}{2^\gamma} \max \left( 0, \frac{p^L_{t observers} - p^L_{t,i}}{\omega_i} \right) & \text{valley filling} 
\end{cases} \tag{3}
\]

where, \( p^L_{t observers} \) and \( p^L_{t observers} \) are load commitment and actual load at time \( t \), respectively; therefore, the load demand adjustment ratio is defined as \( \%\Delta p^{DR}_{t,i} = (p^L_{t observers} - p^L_{t,i}) / p^L_{t observers} \), in which the positive value indicates peak shaving and negative for valley filling; \( \eta^{DR} \) is the DR incentive value, which differs in peak shaving and valley filling time according to Table 2 and DR policy executed specifically in Huangpu district, Shanghai city; \( f^{DR}(\cdot) \) is a parametric function that adjust the certified demand response capacity using coefficient \( \frac{\eta^{IR}}{2^\gamma} \) of incentive ratio, according to the actual DR performance listed in different cases in Table 3. The details of these coefficient tables and DR certification process executed in realistic field experiments are also further explained in Section 4.

For state status space \( S \) and action space \( A \) in VPP modelling, they are composed of time stamp plus load commitment \((t, p^L_{t observers})\) and demand adjustment ratio \( \%\Delta p^{DR}_{t,i} \), respectively. Furthermore, they are also processed via discretisation or binned [30] according to different levels with consideration for acceptable computational complexity. More accurate granulation implies more computational burden and calculation time, which is further explained in Section 5.

3.3 | Action clipped Q-learning algorithm

Although Q-learning [27] is a simple algorithm that has been used for solving many reinforcement learning problems, bearing some limitations (e.g. discrete action/state space), its
variations are still popular in many current applications [16, 31] due to the straightforward implementation and simple programming embedded in most engineering modules. To save the page and recognize the fact that its convergence and theoretical derivation are easily found in many existing works, including our previous work on the episode-dependent algorithm [25], we mean to neglect the $Q$-learning algorithm formulation of agent-environment interaction principle and value or policy functions.

The pseudo-code of the proposed action clipped $Q$-learning algorithm is just briefly presented in Algorithm 1. It is noteworthy that the only change made to conventional $Q$-learning algorithm is the action choice are bounded or clipped by the backup power limitation extracted from empirical statistic information of actual demand response actions.

Another reason that we stick to such a simple algorithm is because of the fact that the real-world installed control unit for end-user only possesses a limited computational capability, which better deal with self-contained computation. Compared with implementation of deep learning-derivative deep reinforcement learning (DRL) algorithm, $Q$-learning is free from many software dependencies. For instance, it waives the complex neural network architecture that depends on back propagation operation and third-party software package [26], like TensorFlow or PyTorch. Therefore, the proposed method could be easily implemented in an edge-computing like portable hardware environment with affordable chips and less communication channel issues. In the future, we would test more powerful methods using upgraded hardware in the field experiments.

### 3.4 Flowchart of training environment

In order to clarify the training environment of algorithm implementation and building-VPP platform interaction process, a flowchart is also presented in Figure 2 to emphasise the key variables and stages of the overall VPP demand response procedure. Similar to the classic reinforcement learning framework [23], the learning agent (building) continuously executes the learning actions via exploration and exploitation to maximise its expected obtained rewards. On the other hand, the training environment (VPP platform) automatically updates the system state status according to the chosen actions and releases the certified reward information, channelling the external price and demand response signal in this specific building-VPP environment.

It is also noteworthy that, as shown in Figure 2 and Algorithm 1, the agent-environment training process usually includes two loops for algorithm implementation. In this paper, the internal loop refers to the real implementation of one VPP experimental event with limited time horizon and terminal state. The outside loop refers to the multiple episode simulation of reusing the historic information and accelerates the agent performance improvement. More experiment setup and time schedule information are discussed in Section 4.

### 4 EXPERIMENT SETUP

The testing buildings for “Shanghai Huangpu VPP Demand Response Pilot Program” are located in Huangpu district, Shanghai city, China. By end of year 2020, there are about 300 buildings participating in this pilot project that allocate demand response incentives to all the responded buildings in emergent events, such as extreme high temperature with large HVAC energy consumption. Each of them is taken as a VPP unit, similar to power plant unit, which is connected to a upper-level VPP communication platform that monitors the overall operation status and DR potential. The real-world implementation and statistic information are further introduced in Section 5.3. In this paper, we select 20 typical commercial buildings for the experiments and provide in-depth analysis of
TABLE 1 General information for the selected twenty commercial buildings

| No. | Type                  | Construction year | Occupancy ratio | Area (m²)  | Height (m) | Floors | Cooling source    | Heating source |
|-----|-----------------------|-------------------|-----------------|------------|-----------|--------|------------------|----------------|
| 1   | Office building       | 2007              | 95%             | 64,648     | 161.5     | B5 - 35F | Water chilling    | Electric boiler |
| 2   | Office & shopping     | 2001              | 90%             | 94,940     | 150.0     | B3 - 37F | Water chilling    | Electric boiler |
| 3   | Office building       | 2003              | 96%             | 98,000     | 134.6     | B2 - 21F | Water chilling    | -              |
| 4   | Office & shopping     | 2013              | 95%             | 100,000    | 134.6     | B4 - 27F | Water chilling    | Gas boiler     |
| 5   | Office                | 1997              | 90%             | 52,124     | -         | B3 - 23F | Air cooled pump   | Electric boiler |
| 6   | Office & shopping     | 2010              | 60%             | 28,290     | 142.0     | B3 - 32F | Water chilling    | Gas boiler     |
| 7   | Office building       | 2002              | 95%             | 95,000     | 198.0     | B3 - 50F | Water chilling    | Gas boiler     |
| 8   | Office                | 1999              | 100%            | 58,900     | 107.0     | B2 - 30F | Air cooled pump   | Air heat pump  |
| 9   | Hotel                 | 1999              | 70%             | 46,400     | 84.0      | B2 - 27F | Air cooled pump   | Air heat pump  |
| 10  | Shopping mall         | 2001              | 100%            | 22,500     | -         | B2 - 6F  | Water chilling    | Air heat pump  |
| 11  | Office & shopping     | 2008              | 96%             | 29,531     | 99.0      | B2 - 20F | Air cooled pump   | Air heat pump  |
| 12  | Office & shopping     | 2010              | 95%             | 100,000    | 99.0      | B4 - 22F | Air cooled pump   | Electric boiler |
| 13  | Office                | 1996              | 98%             | 33,625     | 92.0      | B2 - 28F | Split air conditioner | Air heat pump |
| 14  | Hotel                 | 2007              | 80%             | 28,290     | 79.0      | B2 - 22F | Air cooled pump   | Gas boiler     |
| 15  | Office                | 1997              | 90%             | 52,100     | 150.0     | B3 - 36F | Water chilling    | Electric boiler |
| 16  | Hotel & restaurant    | 2007              | -               | 101,046    | 223.0     | B2 - 52F | Water chilling    | Gas boiler     |
| 17  | Office                | 2011              | 65%             | 26,100     | 66.0      | B2 - 18F | VRV              | VRV            |
| 18  | Office                | 1995              | 98%             | 20,480     | 79.0      | B1 - 21F | Air cooled pump   | -              |
| 19  | Office                | 2010              | 95%             | 76,848     | 163.6     | B3 - 39F | Water chilling    | Gas boiler     |
| 20  | Office                | 1996              | 92%             | 35,000     | 120.0     | B2 - 34F | Water chilling    | Electric boiler |

their demand response performance. The general information for these twenty selected buildings are summarised in Table 1, in which “B5 - 35F” indicates underground 5 floors and over-ground 35 floors, and “VRV” indicates variable refrigerant volume-air conditioner. The different cooling source and heating source of different buildings are the main driving force to the total energy consumption or load level.

Out of these twenty selected commercial buildings, the No.1 building is chosen for demonstration purpose, which is an office building and labelled as No.1 VPP unit in the upper-level communication platform. Specifically, the construction year of this No.1 building is 2003, and it has 21 floors and 2 floors underground for parking. The total building area is 98,000 m². The main energy consumption is dedicated to chilled water pump and cooling purpose. The different categorical loads are monitored by 121 monitoring units with data channelled to the Huangpu-VPP platform. The field experiment setup and device installation for this building is as shown in Figure 3.

Besides these building information, the electricity price scheme executed currently in Shanghai is also presented in Table 2 and used for the following studied cases. Table 3 provides another important policy and DR rules executed particularly in Huangpu district, Shanghai. The response ratio measures the ratio of actually successful DR quantity provided by the building compared with the original scheduled DR commitment. According to the rule, different incentive ratios are also associated with different response ratios, which implies more accurate response could produce larger portion of complete DR incentive received. However, as indicated in Table 3, maximal 140% and minimal 60% response ratio are accepted. In practice, some communication channel issues may cause problems for data processing and synchronisation of computational results as observed in the field experiments. In order to minimise the impact of such issues, the whole computation steps using reinforcement learning could be capsulated into the embedded smart unit near the end-users and processed instantly, while ex-post results uploaded later on.

In the following numerical studies, unless specifically indicated, the study time horizon consists of 13–15 days (according to available real building measurements), and each day consists of 96 time intervals. The chosen time interval is 15-min for a moderate operation with consideration for the trade-off of the accuracy and computational burden. The method performance is firstly tested on an individual building, and the economic analysis is based on the method application for all the available buildings.

5 | NUMERICAL RESULTS

5.1 | Method performance

In order to evaluate and showcase the proposed method performance with detailed analysis, we firstly focus on a single building No.1 that is selected from all the twenty buildings and owns typical commercial building energy consumption characteristics as
introduced in Section 4. The study of more building samples, as well as the overall economic analysis for all the twenty available commercial buildings, are further analysed in Section 5.2. To present the ordinary load level of building No.1 and its normal DR performance with manual operation, its load curve is shown in Figure 4, as well as its manual DR statistics in Figure 5.

In the reinforcement learning-based DR strategy, the used key hype-parameters are as follows: the learning rate $\alpha = 0.9$, discount rate $\gamma = 0.9$, greedy policy $\epsilon = 0.05$. The action space is divided into five intervals from $-0.2$ to $0.2$, namely $[-0.2, -0.1, 0, 0.1, 0.2]$, considering the trade-off of accuracy and computational burden. We tested the algorithm up to 1000 episodes with the observation that for most buildings, such as No.1 building shown in Figure 6, the RL-based DR strategy could easily outperform the manual DR strategy since 200–300th episode. In this way, the VPP operation can gradually improve its performance, automatically adapting to the external price and incentive signals with accumulated historical experience.

In Figure 7, the comparison of load baseline and actual load with demand response using reinforcement learning is also presented for No.1 building, in which a moving average with 24 h-time window is used for clear illustration purpose. We can observe that by using RL method and DR incentive, the load level of this building fluctuates less from 1120 to 690 kW, which implies about 38.4% reduction for the gap between load peak and valley. This result can help flatten or re-scheduled the load shape to ease the upper-level VPP operation, and furthermore facilitate the general power system operation.

### 5.2 Comprehensive analysis

The economic analysis aims to demonstrate the economic benefit of using such an automatic and learning-based adaptive DR

![FIGURE 3](image3.png)

**FIGURE 3** Experiment setup for a real-world commercial building

---

### TABLE 1

| Case | Response ratio | Incentive ratio | Note |
|------|----------------|----------------|------|
| 1    | <0.6           | 0              | -    |
| 2    | [0.6, 0.8)     | 0.8            | -    |
| 3    | [0.8, 1.2]     | 1              | -    |
| 4    | [1.2, 1.4]     | 0.8            | -    |
| 5    | >1.4           | 0.8            | The DR capacity is certified and limited to 140% of the committed capacity. |

---

### TABLE 2

| Case | Response ratio | Incentive ratio | Note |
|------|----------------|----------------|------|
| 1    | <0.6           | 0              | -    |
| 2    | [0.6, 0.8)     | 0.8            | -    |
| 3    | [0.8, 1.2]     | 1              | -    |
| 4    | [1.2, 1.4]     | 0.8            | -    |
| 5    | >1.4           | 0.8            | The DR capacity is certified and limited to 140% of the committed capacity. |

---

### TABLE 3

| Case | Response ratio | Incentive ratio | Note |
|------|----------------|----------------|------|
| 1    | <0.6           | 0              | -    |
| 2    | [0.6, 0.8)     | 0.8            | -    |
| 3    | [0.8, 1.2]     | 1              | -    |
| 4    | [1.2, 1.4]     | 0.8            | -    |
| 5    | >1.4           | 0.8            | The DR capacity is certified and limited to 140% of the committed capacity. |

---

### TABLE 4

| Case | Response ratio | Incentive ratio | Note |
|------|----------------|----------------|------|
| 1    | <0.6           | 0              | -    |
| 2    | [0.6, 0.8)     | 0.8            | -    |
| 3    | [0.8, 1.2]     | 1              | -    |
| 4    | [1.2, 1.4]     | 0.8            | -    |
| 5    | >1.4           | 0.8            | The DR capacity is certified and limited to 140% of the committed capacity. |

---

### FIGURE 4

Load curve of No.1 building with empirical demand response strategy

---

### FIGURE 5

Load curve of No.1 building with anticipated demand response strategy

---

### FIGURE 6

Load curve of No.1 building with reinforcement learning demand response strategy

---

### FIGURE 7

Comparison of load baseline and actual load with demand response using reinforcement learning.
strategy for all the tested buildings over long-term operation. Due to the page limit and for concise purpose, the load curves with DR using RL are only presented for even number buildings as shown in Figure 10. However, the long-term economic benefit is estimated for overall all the available buildings. The experimental period for demand response strategy execution event is associated with different days mostly from March 2018 to December 2019, in which the DR events occur 6 times (days) to 15 times (days). The indication of the day number is meaningless for the DR events since it just reflects their occurrence order. For example, in Figure 10 for No.2 building, Day 1 is 2018.05.17, and for No.4 building, Day 1 is 2018.04.12.

The economic benefit result for all the twenty buildings is also presented in Figure 8, in which we can concludes all the participating VPP buildings increase their economic benefit that mostly range from additional 1 to 3 k CNY. Although the percentage of such extra economic benefit compared with original gain is only about 1.2%∼12.5%, the accumulated benefit over long-term operation is still considerable, especially for huge buildings, such as No.4 building (office) and No.16 building (hotel & restaurant).

Considering the fact that in most commercial buildings cooling or heating demand for keeping temperature is the main driving force to the energy consumption and also the key source of demand flexibility, we categorize all the studied twenty buildings with eight labels, \{WE, WG, AE, AA, WA, SA, AG, VRV\} indicating different types of cooling and heating source combination. For example, “WE” stands for Water chilling (cooling) and Electric boiler (heating), as referred in Table 1. In Figure 9, it can be observed that “WE” is associated with 22% buildings and 22% working time by statistics, however contributes to 26% of the total demand response rate (peak shaving plus valley filling) and 29% of the pure peak shaving DR. Of course, the building type, construction year and other factors might also affect the final DR potential estimation, which should be verified in the long-term operation and more real-world experiments.

5.3 Real-world and long-term implementation

During the period January 2018-October 2020, the Huangpu commercial-building based VPP demand response pilot project has totally initialized 2,196 demand response incentive events for 283 building connected to the Huangpu VPP software platform (Figure 11). The total accumulated demand response capacity is 312.6 MWh. The centrifugal chiller, air cooled heat pump and cooling water pump are the infrastructure reported to provide most demand response capability during the experimental period. As shown in Figure 12, the demand response potential could be estimated using a temperature-power relationship through empirical study. By mapping the
FIGURE 8  Economic benefit result for all the twenty buildings

(a) Source type  (b) Working time

(c) DR contribution  (d) DR (peak shaving) contribution

FIGURE 9  Pie chart for demand response potential of different combination types of cooling and heating sources

learning-based DR strategy and Q-learning policy function, extracted from aforementioned short-term and sampled building experiments, to the long-term Huangpu VPP operation with consideration of development plan, the total demand response potential could be estimated in Table 4. It’s worth mentioning that realizing reinforcement learning-based control methods in practice is usually meaningful for less accurate scenarios acceptable with trial-and-error. Thus, the estimation is essentially a trend analysis and can only guarantee a statistic gain in a non-deterministic manner.

By some pioneering communication experiments, it is also reported that through fully automatic demand response functionality and protocols (e.g., OpenADR) supported by mixed 4G/5G communication modules, almost 80% buildings, which are connected to the Huangpu VPP platform, could respond to the demand response control command in less than 60 s, and the rest 20% buildings could respond within 15 min. Thus, the Huangpu commercial-building based VPP development can benefit a lot from advances in information and communication technology, including learning-based self-adaptive and autonomous decision-making software (algorithm) and ultrafast communication hardware. In return, the Huangpu VPP can also support reliable power system operation, especially for nearby distribution network, enabling a more flexible power supply-demand balance. Meanwhile, it reduces or delays capital investment for thermal power plants and helps realizing meaningful carbon neutrality for industry intensive metropolis.

TABLE 4  The future estimation of demand response potential for long-term Huangpu VPP operation

| No. building connected | Summer | Non-summer |
|------------------------|--------|------------|
| Year 2021–2025 (∼350 buildings) | 450.2 MWh | 365.8 MWh |
| Year 2026–2030 (∼600 buildings) | 982.6 MWh | 758.3 MWh |
| Year 2031–2035 (∼1000 buildings) | 1496.7 MWh | 1199.2 MWh |

FIGURE 10  Actual load with demand response for different commercial buildings (red - load baseline; blue - actual load with DR using RL)
CONCLUSION

In this paper, we proposed a reinforcement learning framework to deal with the demand response problems of commercial building-based virtual power plant in real world. The field experiments and simulation results demonstrate that the commercial building-based VPP has a considerable potential and enough load flexibility to participate in the well-designed demand response programs, especially aided by the intelligent methods with minimal manpower and programming effect. The proposed reinforcement learning method may not produce the best result every time, but can guarantee the benefit in a non-deterministic way for the long-term operation of building-based virtual power plant. In the future work, we will test more powerful reinforcement learning algorithms that are tailored for the micro-control unit with limited computational capability, and focus on the demand response command decomposition solution to handle low-level device control problems. More practical software and hardware concerns, such as communication delay effect, will be tested in field experiments. The demand response policy or business model should also be carefully re-designed to enable a long-term sustainable operation of similar kinds of commercial building-based virtual power plants.

ACKNOWLEDGEMENT

This work was supported by the Shanghai Science and Technology Committee, under the project “The study of public building-based virtual power plant modeling and dynamic construction technology, with the award No. 20DZ1206202. The authors would like to thank Shenghun Tu and Xia Zhou from (Shanghai) Tengtian Energy Saving Technology Corporation Limited’s to provide the valuable data and several field experiments. The authors also thank Economic and Technical Research Institute (ETRI) in State Grid Corporation of China-Shanghai Company and Shanghai Huangpu District Development and Reform Commission’s support to the “Shanghai Huangpu VPP Demand Response Pilot Program”.

ORCID

Tao Chen https://orcid.org/0000-0003-2616-6247
Qinran Hu https://orcid.org/0000-0002-5398-5718

REFERENCES

1. Cai, M., Pipattanasomporn, M., Rahman, S.: Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. Appl. Energy 236, 1078-1088 (2019)
2. Pavlak, G.S., Henze, G.P., Cushing, V.J.: Optimizing commercial building participation in energy and ancillary service markets. Energy Build. 81, 115–126 (2014)
3. Zhang, X., et al.: An iot-based thermal model learning framework for smart buildings. IEEE Internet Things J. 7(1), 518–527 (2019)

4. Arun, S.L., Selvan, M.P.: Dynamic demand response in smart buildings using an intelligent residential load management system. IET Gener. Transm. Distrib. 11(17), 4348–4357 (2017)

5. Pudjianto, D., Ramsay, C., Strbac, G.: Virtual power plant and system integration of distributed energy resources. IET Renewable Power Gener. 1(1), 10–16 (2007)

6. Zhang, G., Jiang, C., Wang, X.: Comprehensive review on structure and operation of virtual power plant in electrical system. IET Gener. Transm. Distrib. 13(2), 145–156 (2018)

7. Zhang, X., et al.: An edge-cloud integrated solution for buildings demand response using reinforcement learning. IEEE Trans. Smart Grid 12(1), 420–431 (2020)

8. Othman, M.M., Hegazy, Y., Abdelaziz, A.Y.: Electrical energy management in unbalanced distribution networks using virtual power plant concept. Electr. Power Syst. Res. 145, 157–165 (2017)

9. Yin, S., et al.: Energy management for aggregate prosumers in a virtual power plant: A robust stackelberg game approach. Int. J. Electr. Power Energy Syst. 117, 105605 (2020)

10. Pandžić, H., Kuzle, I., Capuder, T.: Virtual power plant mid-term dispatch optimization. Appl. Energy 101, 134–141 (2013)

11. Liang, Z., et al.: Risk-constrained optimal energy management for virtual power plants considering correlated demand response. IEEE Trans. Smart Grid 10(2), 1577–1587 (2017)

12. Lin, L., et al.: Deep reinforcement learning for economic dispatch of virtual power plant in internet of energy. IEEE Internet Things J. 7(7), 6288–6301 (2020)

13. Mnih, V., et al.: Human-level control through deep reinforcement learning. Nature 518(7540), 529–533 (2015)

14. Ye, Y., et al.: Deep reinforcement learning for strategic bidding in electricity markets. IEEE Trans. Smart Grid 11(2), 1343–1355 (2019)

15. Yan, Z., et al.: Deep reinforcement learning-based optimal data-driven control of battery energy storage for power system frequency support. IET Gener. Transm. Distrib. 14(25), 6071–6078 (2020)

16. Chang, F., et al.: Control of battery charging based on reinforcement learning and long short-term memory networks. Comput. Electr. Eng. 85, 106670 (2020)

17. Daneshfar, F., Bevrani, H.: Load frequency control: a GA-based multi-agent reinforcement learning. IET Gener. Transm. Distrib. 4(1), 13–26 (2010)

18. Ye, Y., et al.: Real-Time Autonomous Residential Demand Response Management Based on Twin Delayed Deep Deterministic Policy Gradient Learning. Energies 14(3), 531 (2021)

19. Zhou, S., et al.: Artificial intelligence based smart energy community management: A reinforcement learning approach. CSEE J. Power Energy Syst. 5(1), 1–10 (2019)

20. Zhou, Z., et al.: Deep reinforcement learning framework for resilience enhancement of distribution systems under extreme weather events. Int. J. Electr. Power Energy Syst. 128, 106676 (2021)

21. Xu, X., et al.: A multi-agent reinforcement learning-based data-driven method for home energy management. IEEE Trans. Smart Grid 11(4), 3201–3211 (2020)

22. Christensen, M.H., Li, R., Pinson, P.: Demand side management of heat in smart homes: living-lab experiments. Energy 195, 116993 (2020)

23. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA (2018)

24. Mutanen, A., et al.: Customer classification and load profiling method for distribution systems. IEEE Trans. Power Delivery 26(3), 1755–1763 (2011)

25. Chen, T., Su, W.: Indirect customer-to-customer energy trading with reinforcement learning. IEEE Trans. Smart Grid 10(4), 4338–4348 (2018)

26. Chen, T.: Local energy trading behavior modeling with deep reinforcement learning. IEEE Access 6, 62806–62814 (2018)

27. Watkins, C.J., Dayan, P.: Q-learning. Mach. Learn. 8(3–4), 279–292 (1992)

28. Pourbabak, H., et al.: A novel consensus-based distributed algorithm for economic dispatch based on local estimation of power mismatch. IEEE Trans. Smart Grid 9(6), 5930–5942 (2017)

29. Mohajerami, S., et al.: A novel economic model for price-based demand response. Electr. Power Syst. Res. 135, 1–9 (2016)

30. Claessens, B.J., Vrancx, P., Ruelens, F.: Convolutional neural networks for automatic state-time feature extraction in reinforcement learning applied to residential load control. IEEE Trans. Smart Grid 9(4), 3259–3269 (2016)

31. Cui, Q., Weng, Y.: An environment-adaptive protection scheme with long-term reward for distribution networks. Int. J. Electr. Power Energy Syst. 124, 106350 (2021)

How to cite this article: Chen, T., et al.: Optimal demand response strategy of commercial building-based virtual power plant using reinforcement learning. IET Gener. Transm. Distrib. 2021;15:2309–2318. https://doi.org/10.1049/gtd2.12179