Residential HVAC Aggregation Based on Risk-averse Multi-armed Bandit Learning for Secondary Frequency Regulation

Xinyi Chen, Qinran Hu, Qingxin Shi, Xiangjun Quan, Zaijun Wu, and Fangxing Li

***Abstract*—As the penetration of renewable energy continues to increase, stochastic and intermittent generation resources gradually replace the conventional generators, bringing significant challenges in stabilizing power system frequency. Thus, aggregating demand-side resources for frequency regulation attracts attentions from both academia and industry. However, in practice, conventional aggregation approaches suffer from random and uncertain behaviors of the users such as opting out control signals. The risk-averse multi-armed bandit learning approach is adopted to learn the behaviors of the users and a novel aggregation strategy is developed for residential heating, ventilation, and air conditioning (HVAC) to provide reliable secondary frequency regulation. Compared with the conventional approach, the simulation results show that the risk-averse multi-armed bandit learning approach performs better in secondary frequency regulation with fewer users being selected and opting out of the control. Besides, the proposed approach is more robust to random and changing behaviors of the users.**

***Index Terms*—Heating, ventilation, and air conditioning (HVAC), load control, multi-armed bandit, online learning, secondary frequency regulation.**

I. Introduction

HIGH penetration of wind and solar has brought great challenges in stabilizing the frequency of power grids. Conventionally, the generators track system loading levels to maintain the system frequency within a safe range [1]. However, as renewable energy resources are gradually taking the place of conventional ones, it lowers the system inertia [2], deteriorates the characteristics of system frequency response [3], and jeopardies the system stability [4]. This raises interests of researchers on exploring the potential of demand-side resources to enhance the system stability.

With the rapid development of advanced metering infrastructure and communication technology, demand-side resources including electric vehicles [5], batteries [6] and thermostatically controlled loads [7] are enabled as candidates to provide frequency regulation service. As for the capacity, the residential loads account for a significant proportion among all the candidates [8]-[10]. Through proper control strategy, electric water heater (EWH) [11], heating, ventilation, and air conditioning (HVAC) [12] can be aggregated to provide frequency regulation services.

Previous research efforts have been made on achieving load aggregation with faster response time, larger flexibility, higher economic efficiency and user-friendliness [13]-[15]. For instance, a hierarchical framework is applied to control HVACs to participate in primary frequency regulation (PFR) [16]. Reference [17] investigates the ramping rate flexibility of thermostatically controlled loads with motors and compressors. From the perspective of economic efficiency, a multi-agent demand control system is proposed in [18] to maximize the total customer welfare of spinning reserves. To satisfy the thermal comfort of the users, a load grouping control strategy is presented in [19] to select thermostatically controlled loads based on temperature distance and power rating similarity. Reference [20] presents a thermostatic load control strategy for both primary and secondary frequency regulation (SFR), which considers more practical issues such as daily demand profile and load rebound. However, the demand control strategy adopted in [20] is the random switching (RS) approach that ignores opt-out behavior of the users to the regulation commands.

Demand response programs usually offer user option to opt-out while receiving the control signal. The reasons of opt-out are various such as having important events and feeling uncomfortable. Thus, due to the uncertain opt-out behavior, the aggregated demand in practice may differ from the scheduled target and can hardly serve as reliable resource for SFR. To tackle this issue, we propose a control strategy based on a risk-averse multi-armed bandit (MAB) learning approach. Through the online learning process, the load aggregator can understand the opt-out behavior of the users, so as to mitigate the influence on the uncertain response of us-

Manuscript received: August 1, 2020; accepted: October 21, 2020. Date of CrossCheck: October 21, 2020. Date of online publication: November 26, 2020.

This work was supported by the National Natural Science Foundation of China (No. 51907026), Natural Science Foundation of Jiangsu (No. BK20190361), and Jiangsu Provincial Key Laboratory of Smart Grid Technology and Equipment, and Global Energy Interconnection Research Institute (No. SGGR0000WLJS1900107).

X. Chen, Q. Hu (corresponding author), X. Quan, and Z. Wu are with the School of Electrical Engineering, Southeast University, Nanjing, China, and they are also with Jiangsu Provincial Key Laboratory of Smart Grid Technology and Equipment, Nanjing, China (e-mail: chxy@seu.edu.cn; qhu@seu.edu.cn; xquan@seu.edu.cn; zjwu@seu.edu.cn).

Q. Shi and F. Li are with the School of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, USA (e-mail: qshi1@vols.utk.edu; fli6@utk.edu).;

DOI: 10.35833/MPCE.2020.000573

ers for load aggregation. The contributions of this work can be summarized as follows. Firstly, risk-averse MAB learning approach has been applied to learn uncertain responses of the users to SFR commands of load aggregator. Secondly, the proposed approach improves the reliability of aggregating residential HVAC for SFR while reducing the number of called and opt-out behaviors of the users. Thirdly, the proposed approach is robust to the random and changing behaviors of the users.

The rest of this paper is structured as follows. Section II proposes the model of HVAC in demand aggregation. Section III presents the dynamic control strategy based on the MAB learning approach for SFR. Section IV illustrates the performance of the proposed approach with case studies. Section V concludes this paper.

II. HVAC Model Considering User Behavior

A. HVAC Model

The first-order differential equation is adopted to model the temperature variation process based on the assumption that the temperature change is almost linear within the narrow temperature deadband [20]. The heat cycle process and power consumption profile of HVAC are depicted in Fig. 1 with cooling mode [21].

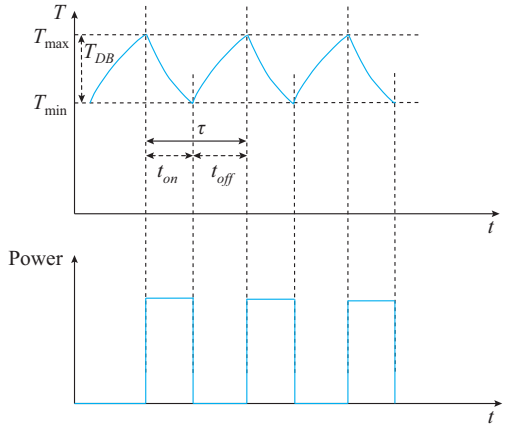

Fig. 1. On/off curve of HVDC and power consumption.

In Fig. 1, T_{max} and T_{min} are the upper and lower bounds for room temperature, respectively; and T_{DB} is the temperature deadband. The HVAC unit operates with a cycling time τ, the falling curves indicate "on" state of the HVAC during time period t_{on}, and the rising curves represent the opposite during time period t_{off}. Concrete equation of room temperature T_{in} is as follows [22]:

$$C\frac{\mathrm{d}T_{in}(t)}{\mathrm{d}t}=\frac{T_{out}(t)-T_{in}(t)}{R}-s(t)Q \tag{1}$$

where C and R are the thermal capacitance and resistance, respectively; T_{out} is the outdoor temperature; Q is the thermal power; and $s(t)$ is the on/off state of HVAC.

The discrete form of (1) is:

$$T_{in}(t+\Delta t)=T_{in}(t)+\frac{\Delta t}{C}\left(\frac{T_{out}-T_{in}(t)}{R}-s(t)Q\right) \tag{2}$$

$s(t)$ is governed by a switching law [23]:

$$s(t)=\begin{cases}0 & s(t-\Delta t)=1, T_{in}(t)\le T_{\min}\\ 1 & s(t-\Delta t)=0, T_{in}(t)\ge T_{\max}\\ s(t-\Delta t) & \text{otherwise}\end{cases} \tag{3}$$

In this case, the total HVAC load profile can be obtained:

$$P(t)=\frac{1}{\eta}\sum_{i=1}^{N}s_i(t)Q_i \tag{4}$$

where η is the performance of coefficient of an HVAC; N is the total number of HVACs; and Q_i is the heat output of HVAC i.

The parameters of HVAC are given in Table I [24].

TABLE I
PARAMETERS OF HVAC THERMAL MODEL

Symbol	Value
R	2.0 ℃/kW
C	2.0 kWh/℃
Q	6.25 kW
T_{DB}	1.0 ℃
T_{min}	23.5 ℃
T_{max}	24.5 ℃
η	2.5

On the basis of thermal model of 1 HVAC, the potential of 50000 HVACs in SFR on a typical hot summer day in Houston [20] can be derived, as shown in Fig. 2. The maximal load that can be aggregated is P_{max}; the load baseline is P_{base}; and the upward and downward frequency regulation capacities are P_{up} and P_{down}, respectively. It can be seen that large-scale HVACs have great flexibility and adjustment capability to participate in SFR.

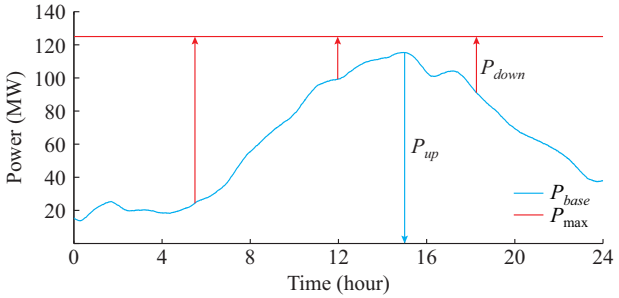

Fig. 2. Aggregated load profile of 50000 HVACs.

B. User Behavior Model in Real-world Practice

Based on the aggregated load profile of HVACs, the reserve capacity for SFR at different time can be easily estimated. Furthermore, when a disturbance occurs, the dynamic demand control strategy determines which HVAC to switch on/off according to the SFR requirement and reserve. An RS approach is carried out in the previous research, where the users are stochastically called with a probability of p_{off} [20].

$$p_{off}=\frac{P_{SFR}(t)}{P_{AC}(t)} \tag{5}$$

where $P_{SFR}(t)$ is the expected power aggregation for SFR; and $P_{AC}(t)$ is the available reserve.

However, the above approach optimistically estimates the behaviors of the users. Usually, in the residential demand response contracts, users always have the option to opt-out from the regulation commands. The complexity and randomness of their behaviors in the real-world increase the difficulty of performing reliable load aggregation.

In this paper, we model the response of each user to frequency regulation command as a probability function which follows Bernoulli distribution $X_i \sim Bern(p_i)$:

$$X_i = \begin{cases} 1 & p_i \\ 0 & 1-p_i \end{cases} \tag{6}$$

where $X_i = 1$ means the user i follows the command and $X_i = 0$ means user i opts out the command; and p_i is the participation probability that user i follows the command, thus the expected load change is $E(X_i)=p_i$, and the variance is $\sigma_i^2 = p_i(1-p_i)$.

Practically, we can hardly obtain accurate value of p_i, which affects the performance of SFR with HVACs. Therefore, we adopt an online learning approach to estimate p_i and aggegate the optimal set of users to improve the performance of their HVACs in SFR.

III. Demand Control Strategy Based on MAB

A. Frequency Regulation Scheme

The flowchart of frequency regulation scheme is shown in Fig. 3, in which HVACs are aggregated for SFR.

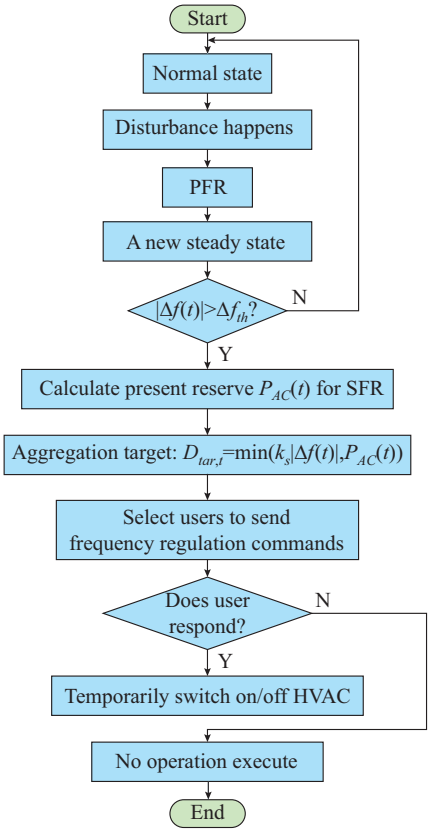

Fig. 3. Flowchart of frequency regulation scheme.

Suppose a disturbance happens. After PFR, the system frequency reaches a new steady state. Calculate the present frequency deviation $|\Delta f(t)|$ from the rated value. Once the deviation exceeds the frequency threshold Δf_{th}, SFR would be performed. Since the demand side is required to simulate the droop characteristic of the generation side to restore the system frequency to the normal range, the SFR droop coefficient is estimated by:

$$k_S = \frac{\max(P_{AC})}{|\Delta f_{m,SFR}|} \tag{7}$$

where $\max(P_{AC})$ is the maximum SFR reserve of HVAC; and $|\Delta f_{m,SFR}|$ is the maximum frequency deviation that the system can sustain for SFR. Then, at time t, the expected aggregated demand target for SFR can be calculated by:

$$D_{tar,t} = \min\left(k_s |\Delta f(t)|, P_{AC}(t)\right) \tag{8}$$

The smaller value is picked between the theoretical expectation and the actual reserve on demand side for SFR at time t in case the HVAC capacity is insufficient. The same formula can be used for upward and downward frequency regulation.

Next, the load aggregator selects the users to send the frequency regulation commands according to the target $D_{tar,t}$. If the user responds to the request, corresponding HVAC will be temporarily switched on/off, otherwise no action will be performed. For achieving reliable frequency regulation, our objective is to minimize the square difference between the actual aggregated power of HVAC after user response $D_{agg,t}$ and the target $D_{tar,t}$ in expectation:

$$\min E\left(D_{agg,t} - D_{tar,t}\right)^2 \tag{9}$$

B. Risk-averse MAB

During each frequency regulation event, the load aggregator receives $D_{tar,t}$ from the dispatching center, and then selects a group of users $S(t) \subseteq [N]$ with unknown participation probability profile $p_1, p_2, \ldots, p_n$ to minimize the aggregated power deviation from the target. It is similar to combinatorial MAB (CMAB) problem: the decision-maker is allowed to choose some revenue-generating arms from a set each time, where the distribution of rewards for each machine is unknown. The objective for decision-maker is to maximize the rewards after a limited number of trials. Different from maximizing revenue, we aim to minimize the expected mismatch between the actual aggregated power and $D_{tar,t}$ in order to achieve reliable frequency regulation.

Response probabilities of the users are unknown in realities, the aggregator should learn response behavior of the users online and adjust selection strategy according to the feedbacks of previous events. Since the feedback can only be obtained when the user is selected within a limited number of events, there is a trade-off between the exploitation of the known information and exploration of more information. Specifically, the exploitation means selecting users based on present estimated participation probability $\hat{p}$, and the exploration means selecting users that have not been adequately called for more information.

For the CMAB problem, there exist many related algorithms such as ϵ-greedy, Thompson sampling [25], Exp3 algorithm [26], combinatorial upper confidence bound (CUCB) [27], contextual CUCB [28]. Most algorithms are designed to achieve the objective of maximizing rewards or minimizing average costs. However, unlike both objectives, for frequency regulation, it is critical to minimize the gap between the actual power aggregation and the target. Otherwise, it will cause additional power imbalance and frequency fluctuation. A learning algorithm CUCB-avg is proposed to handle this problem, which is demonstrated to perform better than the classic CUCB algorithm [29].

Meanwhile, the algorithm introduced above ignores the risk together with the rewards. However, the behaviors of the users are random and changing. The load aggregators should not only consider the expected aggregation power that the users can achieve but also avoid the risk of users with high power expectation but erratic performance. Similar consideration often arises in financial investments. For example, in the stock market, investors may not only consider whether stocks will bring high returns but also their variation to avoid risk when making long-term investments. Therefore, to ensure reliable frequency regulation, load aggregators may prefer users that provide stable responses. In this paper, we adopt risk-averse MAB learning approach [30] for load aggregation, and the algorithm is designed as follows:

Algorithm 1

1. Input: ρ_1, ρ_2, $P_1, P_2, \ldots, P_n$, $D_{tar,t}$
2. Initialization:
   for each user $i \in [N]$ do
   Initialize estimate of participation probability $\hat{p}_{t,i}$, estimate of variance $\hat{\sigma}^2_{t,i}$, and historical called times $n_{t,i}$.
   end for
3. for each time $t = 1$ to T do
   Calculate priority index for each user i:
   $$v_{t,i} = P_i \hat{p}_{t,i} - \rho_1 P_i^2 \hat{\sigma}^2_{t,i} + \rho_2 \sqrt{\ln(t)/n_{t,i}}$$
4. Rank users in descending order with $v_{t,i}$:
   $$v_{t,\varepsilon_1} \geq v_{t,\varepsilon_2} \geq \ldots \geq v_{t,\varepsilon_n}$$
5. Select m ($m \geq 0$) users in sequence until:
   $$\sum_{i=1}^{m} P_{\varepsilon_i} \hat{p}_{t,\varepsilon_i} \geq D_{tar,t}$$
   Let $m = n$ if $\sum_{i=1}^{m} P_{\varepsilon_i} \hat{p}_{t,\varepsilon_i} \leq D_{tar,t}$.
6. Output: $S_t = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m\}$.
7. Update:
   for each called user $j \in S_t$ do
   $$\begin{cases} \hat{p}_{t+1,j} = \dfrac{\hat{p}_{t,j} n_{t,j} + X_{t,j}}{n_{t,j} + 1} \\ \hat{\sigma}^2_{t+1,j} = \hat{p}_{t+1,j}(1 - \hat{p}_{t+1,j}) \\ n_{t+1,j} = n_{t,j} + 1 \end{cases}$$
8. end for
9. end for

In Algorithm 1, ρ_1 and ρ_2 are the constants; $P_1, P_2, \ldots, P_n$ are the HVAC power of each user; n is the total number of users; and $S_t = \{\varepsilon_1, \varepsilon_1, \ldots, \varepsilon_m\}$ is the set of selected users.

The core of the above algorithm is to sort users in descending order as:

$$v_{t,i} = P_i \hat{p}_{t,i} - \rho_1 P_i^2 \hat{\sigma}^2_{t,i} + \rho_2 \sqrt{\ln(t)/n_{t,i}} \tag{10}$$

Apparently, this formula gives the priority to choose those users with high estimated participation probability and fewer times of receiving frequency regulation signals, where the trade-off between the exploitation and exploration is reflected. The second term of (10) is to lower users' priority whose responses exhibit great variability. Thus, the risk brought from uncertain behaviors of the users can be mitigated.

Note that the highest order term of time complexity is the log-linear time, which scales well in risk-averse MAB learning approach.

C. Control Framework

The schematic of control framework is shown in Fig. 4. Load aggregator aggregates users to participate in power system operation [31], [32]. The demand controller of the load aggregator is responsible for measuring the bus frequency and collecting the information about how many users are registered for frequency regulation. Once the frequency deviation above the threshold is detected or receives commands from the power grid dispatching center, the demand controller sends regulation commands to users for frequency regulation based on the risk-averse MAB learning approach. If the user responds to the control signals, on/off switching action will be executed on connected appliances. The response of each user is fed back to the demand controller [21].

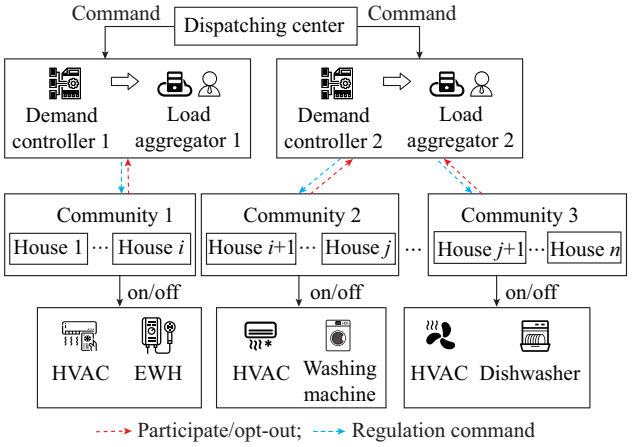

Fig. 4. Schematic of control framework.

IV. Case Study

In this section, the modified IEEE RTS 24-bus, 10-machine system [20] is employed to test the performance of the proposed control strategy for SFR, as shown in Fig. 5. The simulation environment is built on MATLAB PSAT toolbox (V.2.1.11) [33] with 100 MW system base power.

A. Comparison of Different Approaches

We compare the risk-averse MAB based load aggregation approach with RS and offline approaches. Since the previous study [20] does not consider the opt-out behavior of the users, for the sake of fairness of comparison, we improve RS to some extent, and the number of users called is calculated by:

$$N_{RS} = D_{tar,t} / E(\bar{P}) \tag{11}$$

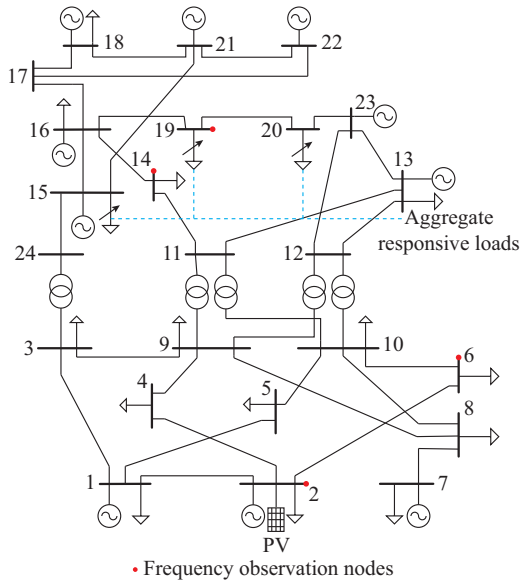

Fig. 5. IEEE RTS 24-bus system.

where $E(\bar{P})$ is the expected mean of power of HVACs based on the estimated participation probability of $\hat{p}$ of the user.

Meanwhile, the offline approach is often used to verify the effectiveness of online learning approaches, where the response probability of all users to the regulation commands is pretended to be known.

Assume that a disturbance of power supply shortage of 74 MW occurs at bus 2 at 15:00. There are 50000 users (one HVAC per user) who have signed up for providing frequency regulation service distributed at load buses 15, 19, and 20, and the average power consumption for each HVAC is $\bar{P} \approx 2.5$. According to the load profile presented in Fig. 2, the upward regulation capacity is 115.3160 MW for SFR, and the maximum reserve is 115.3753 MW. The maximal frequency deviation $\left|\Delta f_{m,SFR}\right|$ that system can sustain for SFR is set to be 0.2 Hz, and the frequency droop k_s is calculated as:

$$k_s = \frac{\max(P_{AC})}{\left|\Delta f_{m,SFR}\right|} = \frac{115.3753}{0.2} = 576.8764 \tag{12}$$

Meanwhile, the number of HVAC available for SFR is calculated as:

$$n = \frac{115.3160 \times 10^6}{2.5 \times 10^3} = 46126 \tag{13}$$

We consider that the actual participation probabilities of users $p_1, p_2, \ldots, p_n$ obey the uniform distribution $[0,1]$. Since these values are unknown in practice, the initial estimates of them in RS and risk-averse MAB learning approaches are $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n$, and the mean $\hat{p}$ is set to be 0.65.

When the system frequency falls below the threshold of 59.97 Hz, PFR is activated. And the responsive load capacity and frequency droop for PFR are 34.40 MW and 200.0 MW/Hz, respectively [20]. After that, the system frequency reaches the new steady state of about 59.9514 Hz at 19 s, as shown in Fig. 6.

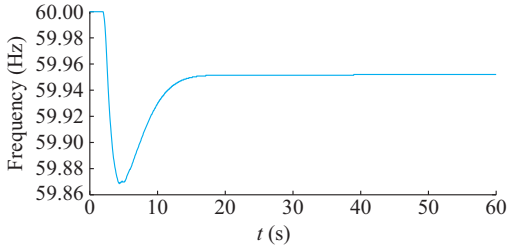

Fig. 6. System frequency after PFR.

The system frequency should be further brought back to the normal level (60 ± 0.04 Hz) through SFR by turning off the HVACs. According to (8), the target reduction of HVACs is 28.09 MW. Figure 7 plots the frequency regulation results of offline, RS, and risk-averse MAB approaches. The results of risk-averse MAB approach at the 20th, 80th, and 200th events are provided respectively. It can be observed that the risk-averse MAB based demand control strategy always performs better than RS in raising the frequency level, and gradually gets closer to the offline approach with the accumulation of learning process about the behaviors of the users.

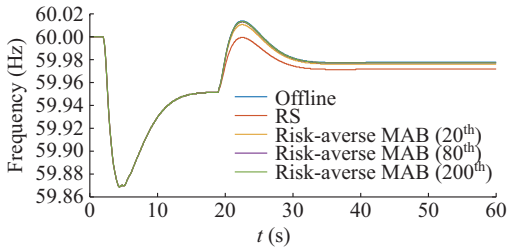

Fig. 7. Comparison of frequency regulation results.

Figure 8 shows 95% confidence intervals of actual aggregated power and relative reduction deviations from the target of different approaches during 200 frequency regulation events. The power value aggregated by the proposed risk-averse MAB based control strategy gradually approaches the results of the offline approach after learning the behaviors of the users, outperforming RS strategy which remains a mismatch of about 6.49 MW from the target. It is also observed in Fig. 8(b) that the relative deviation of the risk-averse MAB learning approach falls to less than ±5% after dozens of events while the relative deviation of RS always remains about 23%. The results verify the regulation reliability of our proposed approach and its solid performance in multiple Monte Carlo cases.

Besides, Fig. 9 reveals that RS sends frequency regulation commands to more users and nearly half of them opt out, while the risk-averse MAB has better results in both aspects. Even both of them have very limited knowledge of users at the very beginning, MAB calls fewer users because it selects users with high expectations until reaching the adjustment target.

In practical application, the selection for each time comes with a cost and user fatigue. Therefore, the risk-averse MAB approach not only ensure the reliability of the load aggregation but also guarantee the economy and user-friendliness, which achieves the win-win results for both the load aggre-

gators and users.

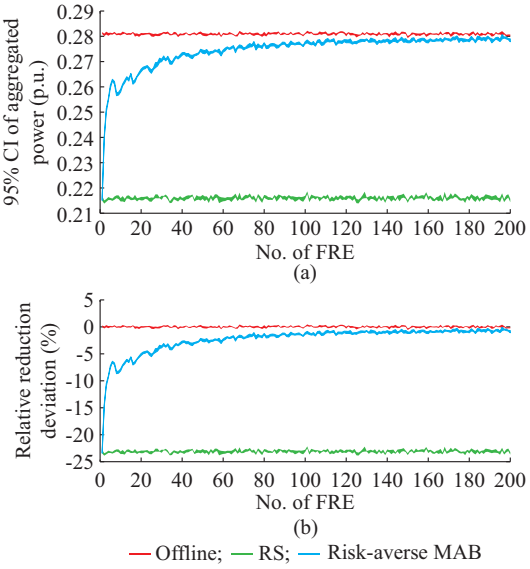

Fig. 8. 95% confidence intervals of load reduction results of different approaches. (a) Actual aggregated power. (b) Relative reduction deviations from target.

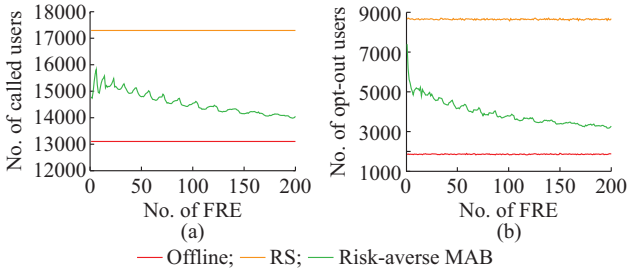

Fig. 9. Number of called and opt-out users of different approaches. (a) Called users. (b) Opt-out users.

B. Impacts of Initial Estimated Probability Average $\hat{p}$

The mean of initial estimate of $\hat{p}$ is 0.65, which is higher than the actual average value $\bar{p}$. Since $\hat{p}$ is one indispensable input of user-selection approaches, its impact on the results of different approaches under diverse settings is discussed. We consider two more cases: ① $\hat{p}$ is about 0.5, so the guess about the behavior of the user is very close to the truth; ② $\hat{p}=0.4$, which underestimates the actual response of the user.

Figure 10 demonstrates that the power aggregation performance of RS is highly sensitive to the initial guess of user participation probability. However, no matter how the initial setting of $\hat{p}$ changes, MAB can always gradually approximate the optimization results of the offline approach by learning the behavior of the user, indicating that the proposed approach is reasonably robust to the initial estimate of $\hat{p}$.

Figure 11 shows that the number of users called and opt-out based on the risk-averse MAB learning approach is always lower than those of RS in 200 events under different initial estimated probability settings. Besides, with the increase of $\hat{p}$, RS calls fewer users, and the number of opt-out users also decreases, which is reasonable because RS overestimates the power aggregation expectation when sending commands. The performance of the risk-averse MAB learning approach once again confirms its robustness and indicates its reliability, economy and user-friendliness for load aggregation.

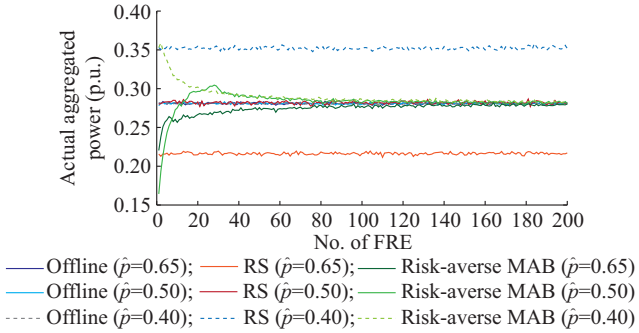

Fig. 10. Impacts of initial estimated probability to actual aggregated power.

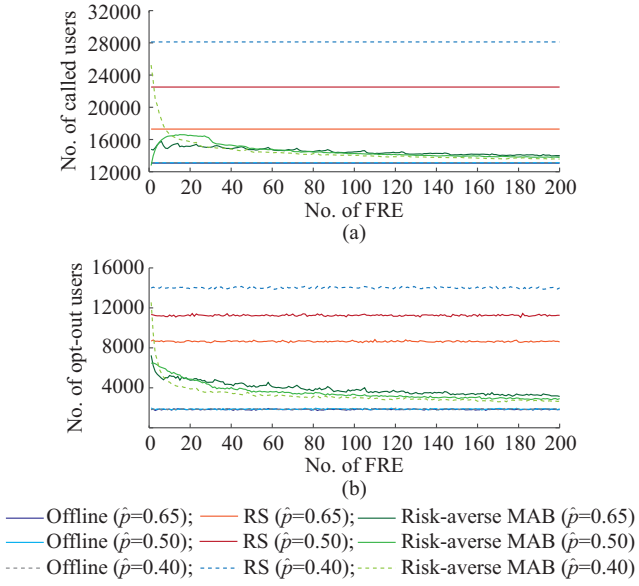

Fig. 11. Impacts of initial estimated probability to number of called and opt-out users of different approaches. (a) Called users. (b) Opt-out users.

C. Impacts of Changes in Behaviors of Users

In practice, the behaviors of the users have great uncertainty and randomness. Hence, the response probability is not always a constant value. It may be affected by the satisfaction level of frequency regulation program, temperature tolerance, outdoor temperature, personal lifestyle, and neighborhood effect. In this case, a variation ratio for the behaviors of the users is introduced. We assume that 10% of users would change their response probability for every 20 events. Other settings are consistent with Section IV-A.

Figure 12 shows that even if the actual response patterns of the users change, the aggregated power of the risk-averse MAB can still gradually approximate that of the offline approach with a small fluctuation, whereas RS still remains a mismatch of about 6.50 MW from the target. It is also observed that the relative reduction deviation of the proposed approach is reduced to below 5% after dozens of events. Figure 13 shows the frequency regulation results of offline, RS, and risk-averse MAB learning approaches at the 20th, 80th,

and 200th events, demonstrating that the MAB-based load aggregation strategy can maintain better performance when facing the changing behaviors of the users.

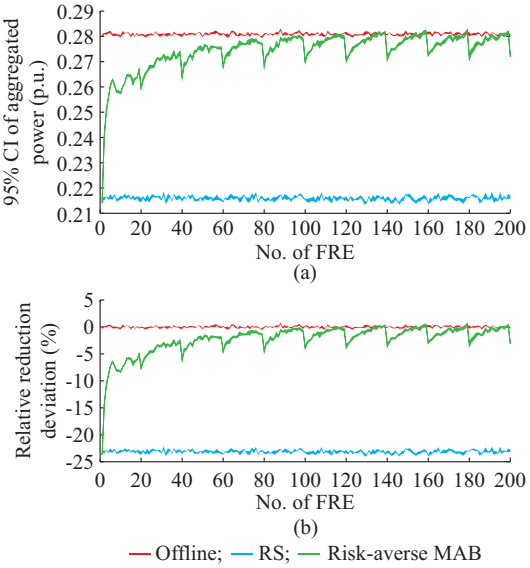

Fig. 12. 95% confidence intervals of load reduction results of different approaches with changing behaviors of users. (a) Actual aggregated power. (b) Relative reduction deviations from target.

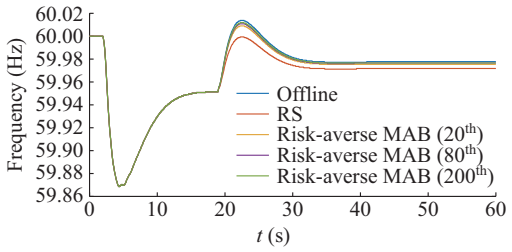

Fig. 13. Comparison of frequency regulation results with changing behaviors of users.

In terms of economy and user-friendliness, Fig. 14 also demonstrates the robustness of the risk-averse MAB concerning the number of called and opt-out users with changes in response probabilities.

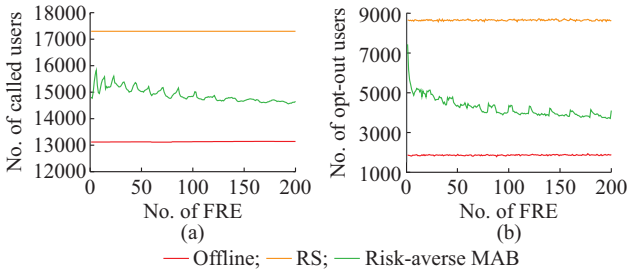

Fig. 14. Number of called and opt-out users of different approaches with changing behaviors of users. (a) Called users. (b) Opt-out users.

V. Conclusion

This paper presents a control strategy for aggregating residential HVACs to participate in SFR based on the risk-averse MAB. Based on the thermal model of individual HVAC, the aggregated load profile is estimated. Then, the frequency regulation reserve of HVACs for the up-down regulation and droop coefficient of SFR can be determined. In the aggregation process, the risk-averse MAB learning approach is implemented to understand the opt-out behavior of the users to frequency regulation commands. Through the online learning process of risk-averse MAB, the load aggregator can mitigate the uncertainty of aggregated demand and provide better SFR service.

Compared with conventional approach, the proposed MAB-based approach can achieve a better frequency regulation performance while fewer users are called and opt-out. The simulation results verify that the proposed approach is robust to the random and changing behaviors of the users. These advantages are beneficial for load aggregators to provide efficient and economical SFR service.

In the future, we plan to consider the impact of the fatigue effect of the users in responding to repeated demand aggregation control signals.

References

[1] P. J. Douglass, R. Garcia-Valle, P. Nyeng *et al*., "Smart demand for frequency regulation: experimental results," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1713-1720, Sept. 2013.

[2] Y. Bian, H. Wyman-Pain, F. Li *et al*., "Demand side contributions for system inertia in the GB power system," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 3521-3530, Jul. 2018.

[3] H. Bevrani, A. Ghosh, and G. Ledwich, "Renewable energy sources and frequency regulation: survey and new perspectives," *IET Renewable Power Generation*, vol. 4, no. 5, pp. 438-457, Sept. 2010.

[4] A. Palomino and M. Parvania, "Data-driven risk analysis of joint electric vehicle and solar operation in distribution networks," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 141-150, Mar. 2020.

[5] H. Liu, Z. Hu, Y. Song *et al*., "Vehicle-to-grid control for supplementary frequency regulation considering charging demands," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3110-3119, Nov. 2015.

[6] Q. Zhai, K. Meng, Z. Dong *et al*., "Modeling and analysis of lithium battery operations in spot and frequency regulation service markets in Australia electricity market," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2576-2586, Oct. 2017.

[7] L. Zhao, W. Zhang, H. Hao *et al*., "A geometric approach to aggregate flexibility modeling of thermostatically controlled loads," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4721-4731, Nov. 2017.

[8] R. D'hulst, W. Labeeuw, B. Beusen *et al*., "Demand response flexibility and flexibility potential of residential smart appliances: experiences from large pilot test in Belgium," *Applied Energy*, vol. 155, no. 1, pp. 79-90, Oct. 2015.

[9] S. Nistor, J. Wu, M. Sooriyabandara *et al*., "Capability of smart appliances to provide reserve services," *Applied Energy*, vol. 138, no. 15, pp. 590-597, Jan. 2015.

[10] M. Afzalan and F. Jazizadeh, "Residential loads flexibility potential for demand response using energy consumption patterns and user segments," *Applied Energy*. doi: 10.1016/j.apenergy.2019.113693

[11] T. Clarke, T. Slay, C. Eustis *et al*., "Aggregation of residential water heaters for peak shifting and frequency response services," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 22-30, Nov. 2020.

[12] O. Erdinç, A. Taşcıkaraoğlu, N. G. Paterakis *et al*., "End-user comfort oriented day-ahead planning for responsive residential HVAC demand aggregation considering weather forecasts," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 362-372, Jan. 2017.

[13] Q. Shi, F. Li, Q. Hu *et al*., "Dynamic demand control for system frequency regulation: concept review, algorithm comparison, and future vision," *Electric Power Systems Research*, vol. 154, pp. 75-87, Jan. 2018.

[14] F. Pallonetto, M. De Rosa, F. D'Ettorre *et al*., "On the assessment and control optimisation of demand response programs in residential buildings," *Renewable and Sustainable Energy Reviews*. doi: 10.1016/j.rser.2020.109861

[15] H. Hao, B. M. Sanandaji, K. Poolla *et al*., "Potentials and economics

of residential thermal loads providing regulation reserve," *Energy Policy*, vol. 79, pp. 115-126, Apr. 2015.
[16] X. Wu, J. He, Y. Xu *et al*., "Hierarchical control of residential HVAC units for primary frequency regulation," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3844-3856, Jul. 2018.
[17] B. M. Sanandaji, T. L. Vincent, and K. Poolla, "Ramping rate flexibility of residential HVAC loads," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 2, pp. 865-874, Apr. 2016.
[18] S. Weckx, R. D'Hulst, and J. Driesen, "Primary and secondary frequency support by a multi-agent demand control system," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1394-1404, May 2015.
[19] S. Lin, D. Liu, F. Hu *et al*., "Grouping control strategy for aggregated thermostatically controlled loads," *Electric Power Systems Research*, vol. 171, pp. 97-104, Jun. 2019.
[20] Q. Shi, F. Li, G. Liu *et al*., "Thermostatic load control for system frequency regulation considering daily demand profile and progressive recovery," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6259-6270, Nov. 2019.
[21] Y. Shen, Y. Li, Q. Zhang *et al*., "State-shift priority based progressive load control of residential HVAC units for frequency regulation," *Electric Power Systems Research*. doi: 10.1016/j.epsr.2020.106194
[22] D. S. Callaway, "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy," *Energy Conversion and Management*, vol. 50, no. 5, pp. 1389-1400, May 2009.
[23] Q. Shi, C. Chen, A. Mammoli *et al*., "Estimating the profile of incentive-based demand response (IBDR) considering technical models and social-psychological factors," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 171-183, Jan. 2020.
[24] J. L. Mathieu, M. Dyson, and D. S. Callaway, "Using residential electric loads for fast demand response: the potential resource and revenues, the costs, and policy recommendations," in *Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings*, Pacific Grove, USA, Aug. 2012, pp. 189-203.
[25] D. J. Russo, B. Van Roy, A. Kazerouni *et al*., "A tutorial on thompson sampling," *Foundations and Trends in Machine Learning*, vol. 11, no. 1, pp. 1-6, Nov. 2017.
[26] A. Mohamed, A. Lesage-Landry, and J. A. Taylor, "Dispatching thermostatically controlled loads for frequency regulation using adversarial multi-armed bandits," in *Proceedings of 2017 IEEE Electrical Power and Energy Conference (EPEC)*, Saskatoon, Canada, Oct. 2017, pp. 338-343.
[27] W. Chen, Y. Wang, Y. Yuan *et al*., "Combinatorial multi-armed bandit and its extension to probabilistically triggered arms," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1746-1778, Jan. 2016.
[28] S. Li, B. Wang, S. Zhang *et al*., "Contextual combinatorial cascading bandits," in *Proceedings of the 33rd International Conference on Machine Learning*, New York, USA, Jun. 2016, pp. 1245-1253.
[29] Y. Li, Q. Hu, and N. Li, "Learning and selecting the right customers for reliability: a multi-armed bandit approach," in *Proceedings of the IEEE Conference on Decision and Control*, Miami Beach, USA, Dec. 2018, pp. 4869-4874.
[30] S. Vakili and Q. Zhao, "Risk-averse multi-armed bandit problems under mean-variance measure," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 6, pp. 1093-1111, Sept. 2016.
[31] Y. Xu, L. Xie, and C. Singh, "Optimal scheduling and operation of load aggregators with electric energy storage facing price and demand uncertainties," in *Proceedings of 2011 North American Power Symposium*, Boston, USA, Aug. 2011, pp. 1-7.
[32] W. Lyu, J. Wu, L. Zhao *et al*., "Load aggregator-based integrated demand response for residential smart energy hubs," *Mathematical Problems in Engineering*, vol. 2019, pp. 1-14, Apr. 2019.
[33] F. Milano. (2006, Mar.). Power system analysis toolbox documentation for PSAT version 2.0.0_1. [Online]. Available: http://faraday1.ucd.ie/psat.html

Xinyi Chen received the B.S. degree in electrical engineering and automation from China Three Gorges University, Yichang, China, in 2018. She is currently pursuing the Ph.D. degree in the School of Electrical Engineering, Southeast University, Nanjing, China. Her current research interests include distributed energy resources and power system optimization.

Qinran Hu received the B.S. degree in electrical engineering from Chien-Shiung Wu College, Southeast University, Nanjing, China, in 2010, and the M.S. and Ph.D. degrees in electrical engineering from the University of Tennessee, Knoxville, USA, in 2013 and 2015, respectively. He was a Postdoctoral Fellow with Harvard University, Cambridge, USA, from 2015 to 2018. He joined the School of Electrical Engineering, Southeast University, in 2018. His research interests include power system optimization, demand aggregation, and virtual power plant.

Qingxin Shi received the B.S. and M.Sc. degrees in Zhejiang University, Hangzhou, China, and University of Alberta, Alberta, Canada, in 2011 and 2014, respectively. He received the Ph.D. degree in University of Tennessee, Knoxville, USA, in 2019, where he is working as a Research Assistant Professor. His research interests include demand response and distribution system resilience.

Xiangjun Quan received the B.S.E.E. and the M.S. degrees in electrical engineering from Chongqing University, Chongqing, China, in 2007 and Southeast University, Nanjing, China, in 2014. In 2018, he received his Ph. D. degree in electrical engineering from Southeast University. From February 2017 to August 2017, he had studied at North Carolina State University, Raleigh, USA. From September 2017 to August 2018, he had also studied in University of Texas, Austin, USA, as an exchange student. He was an Engineer with Huawei Technologies from 2011 to 2012. Since 2018, he has been an Assistant Professor with Southeast University. His current research interests include digital control technique for converters, renewable energy generation systems and microgrid.

Zaijun Wu received his B.S.E.E. degree from Hefei University of Technology, Hefei, China, in 1996, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2004. He is currently a Professor with the School of Electrical Engineering, Southeast University. His research interests include substation automation, microgrid, and power quality.

Fangxing Li received the B.S.E.E. and M.S.E.E. degrees in electrical engineering from Southeast University, Nanjing, China, in 1994 and 1997, respectively, and the Ph.D. degree in electrical engineering from Virginia Polytechnic Institute and State University, Blacksburg, USA, in 2001. Currently, he is the James McConnell Professor at The University of Tennessee, Knoxville, USA. He is a Fellow of IEEE (Class of 2017) and a recipient of the R&D 100 Award in 2020. His research interests include renewable energy integration, demand response, power markets, power system control, and power system computing.