Bulletin of Networking, Computing, Systems, and Software – www.bncss.org, ISSN 2186-5140
Volume 9, Number 1, pages 58–62, January 2020

Multi-agent Based Energy Balancing Management Algorithm for Smart Grid System

Malla Dinesh Bahadur
Grid, Inc.
Tokyo, Japan
Engineering department,
The University of Electro-Communications
Tokyo, Japan

Katsuyoshi Sakamoto
Engineering deptpartment,
The University of Electro-Communications
Tokyo, Japan

Tomah Sogabe*
i-PERC &
Engineering department,
The University of Electro-Communications
Tokyo, Japan
Grid, Inc.
sogabe@uec.ac.jp

***Abstract*—The smart power system focuses on renewable energy sources, which has the potential to reduce the dependence of residential buildings on electricity systems. However, their integration into existing systems increases instability, supply insecurity. Optimizing output schedules and regular forecasting of electricity demand can improve power system stability. But, constantly changing demand for electric power creates problems in scheduling and forecast. For this, it is essential to prioritize how to exploit the immediate deployment of operating units and storage resources online. These processes include optimization and forecasting processes that can address under the umbrella of a multi-agent learning process. The purpose of the proposed multi-agent algorithm in the centralized controller is to learn the policy of maximizing the performance of each agent by ordering it to perform the average of each step based on each agent's reward. In this way, the multi-agent learns to solve and optimize problems. In this paper, we use a multi-agent DQN algorithm by introducing two global states, which play role to communicate among each individual agent to minimize the electricity peak problem and electricity balance between houses by optimal use of storage utilities.**

***Keywords*—smart grid, reinforcement leaning, multi-agent**

I. Introduction

The smart power system focuses on renewable energy sources, which has a great potential for reducing the dependence of a residual building on electricity systems [1]. We expect these to benefit planning and operation of the future power systems and to help customers transition from a passive to an active role [11]. The emerging smart power system uses digital technology to meet the end user's expected value for bilateral communication between utilities and affiliated consumers. Integrating sufficient amounts of renewable energy sources at the residential and power system levels reduces the environmental impact of electrical infrastructure [2]. However, their integration into existing systems increases instability, supply insecurity. Smart power systems enable two-way information flow, a power grid status, and real-time reporting of outages and effective interaction of renewable energy sources. These technologies allow monitoring of power generation, automatic control of the power consumption of smart devices, error detection in the system and so on. Integrating residential-level renewable power generation such as photovoltaics (PV) and power storage into smart grids can be useful in reducing power outages that empowering residential consumers during peak periods of the day. Therefore, the design of the dynamic balancing control algorithm is an important task for the smart grid to deliver on its promises [3].

Although many researchers consider human comfort and satisfaction [12], many of them focus on a single-agent system with unparalleled demand-free electricity prices and a stable environment. A single-agent problem-solving strategy with the purpose of optimization is very popular in many real-world problems [4]. But solving multipurpose problems requires sharing agent competence, which is challenging through a separate agent teaching process. We understand the need to explore enforcement education in coordination with multi agent [13] systems that can take part in demand response programs driven by demand-driven power systems.

In this paper, we use the multi-agent reinforcement learning (MARL) technique, where each agent derives the optimal control policy for the residential energy storage module, which only possess partial knowledge of system modeling. More specifically, the reinforcement learning-based storage control does not need the information of power conversion efficiencies of various DC/DC converters and DC/AC inverters but needs the information precisely estimate the remaining energy in the storage module. For reinforcement learning purposes, time sequence data are very important. In our work, PV generation and demand are time dependent data where battery state of charge (SOC) is varying with each action implementation. Over the years, MARL has attracted extensive investigations and real-world applications because of its distributed nature of the multi-agent solution [5], in which each agent can maximize its payoff by competing or cooperating with other agents, e.g., a deep communication for getting a high-quality optimal solution of the multi-agent system (MAS). In terms of Q- learning approach, getting a better result, each agent consumes a large amount of computation time to gain the optimal Q-value matrix, especially for the equilibrium computation [6]. And action space is also proportional with the computation time [7], i.e., the higher the control accuracy, the longer computation time and vice versa..

II. SYSTEM ARCHITECTURE AND ALGORITHM

A. system architecture

In this paper, we consider a microgrid system with an external power supply which is shown in Fig.1. We consider microgrid as a balancing model by using a fixed time and fixed external power supply through grid where unbalanced power is settled between the user's storage supply and PV production. The microgrid contains several houses which consider residential consumer equipped with PV power generation and energy storage modules which is plotted in Fig.1. The primary purpose of energy storage modules in this system is to serve critical loads during a utility outage and offer power to a residential consumer during the peak period of the day for peak saving. It connects the PV and storage modules to a residential DC bus via DC-DC converters. It connects the smart grid and the residential AC load to the AC bus, which is further connected to the residential DC bus via an AC/DC inverter and a rectifier (Fig.2). The energy storage controller can supply energy to the energy storage modules when utility-supplied electricity suffices to store and can convert the output of an energy storage module to AC load during the peak period of the day. It could operate both in coordination with PV energy generation. We adopt a slotted time model, i.e., it provides all system constraints and decisions for discrete-time intervals of equal and fixed length. We divide each day into T time slots, each with duration D. Hence, we use T = 48 and D = 30 minutes. All house contains the same capacity battery which has fixed discharge and charge quantity power per hour. The battery power is not used for sale(supply) purpose so using battery power agent can supply PV generation to the external grid. The multi-agent environment is shown in Fig.3. Each agent demand is E^h_{demand} and supply is E^h_{supply} ,here h = [house1, house2, house3]. E^h_{demand} is different at each time with each agent and it is scalar value is defined in equation (1). E_{demand} is total supply to the grid by the agent which is in equation (3). E^h_{supply} is supply by the agent which is equal to the PV production of agent is defined in equation (2).

Fig.1 Sketch of overall micro-grid system

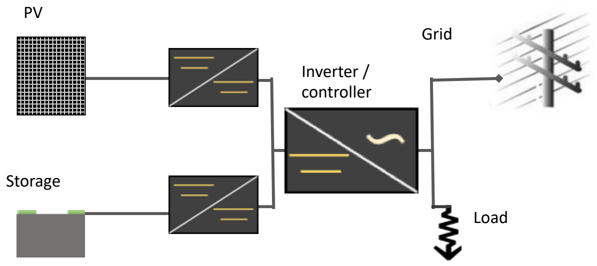


Fig.2 Sketch of electricity flow and agent-based battery scheduling

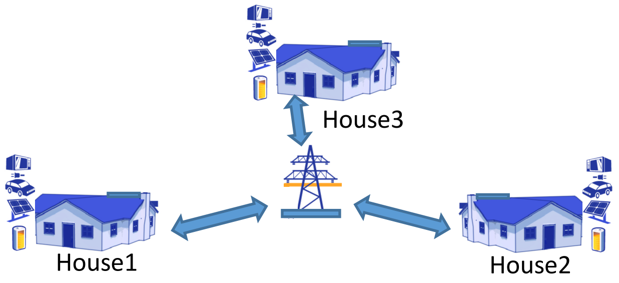


Fig.3 Diagram of multi-agent environment

$$E^h_{demand,t} \geq E^h_{B_d,t} + E^h_{PV,t} + E_{supply,t} \qquad (1)$$
$$E^h_{supply,t} \leq E^h_{PV,t} \qquad (2)$$
$$E_{demand,t} = \sum_{h=1}^{H} E^h_{supply,t} \qquad (3)$$

B. Learning algorithm

DQN [9] is popular method in reinforcement learning and has been previously applied to multi-agent settings [8]. Q-learning makes use of an action-value function for policy π as $Q^{\pi}(s,a) = \mathbb{E}[R|s^t = s, a^t = a]$. This Q function can be recursively rewritten as $Q^{\pi}(s,a) = \mathbb{E}_{s'}[r(s,a) + \gamma(\mathbb{E}_{a'\sim\pi}[\,Q^{\pi}(s',a')]]$. DQN learns the action-value function Q^* corresponding to the optimal policy by minimizing the loss:

$$\mathcal{L}(\theta) = \mathbb{E}_{s,a,r,s'}[(\,Q^*\,(s,a|\theta) - y)^2] \qquad (4)$$

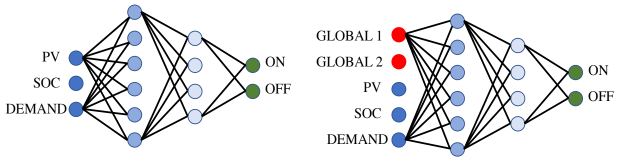
where $y = r + \gamma \max_{a'} \bar{Q}^*\,(s',a')$.

where $\bar{Q}$ is a target Q function, whose parameters are periodically updated with the most recent θ, which helps stabilize learning. Another crucial component of stabilizing DQN is the use of an experience replay buffer R containing tuples (s,a,r,s'). Q-Learning can be directly applied to multi-agent settings by having each agent i learn an independently optimal function Q_i [10]. However, because agents are independently updating their policies as learning progresses, the environment appears non-stationary from the view of anyone agent, violating Markov assumptions required for convergence of Q-learning.

We used the information sharing method for Agent's Global objectives [10]. Each time, the computing agent receives the latest information from the environment after the last agent's policy execution. But, only individual states available from the environment are not sufficient to take corrective action. As such, it is shared as a new state (global state) to inform the cooperative agent interaction with the environment, which is shown in Fig.4. where GLOBAL1 is total demand of total agent each time steps which is expressed in equation(5). And GLOBAL2 is external grid capacity after learned agent external power used, as shown in equation (6).

$$GLOBAL1 = \sum_{h=1}^{n} E^{h}_{demand,t} \tag{5}$$

$$GLOBAL2 = E^{t}_{supply} - \sum_{h=1}^{n} E^{h}_{demand,t} \tag{6}$$



Normal state based learning agent

Global state based learning agent

Fig.4 Multi-agent information sharing using global state

We consider two reward functions for each agent updating process, the first is a local reward received by the agent's actions for local progress and the other is a reward for global balance, which considers all the agent's actions for the global balance process. The purpose of the learning process is to reduce the external grid's energy consumption which can be defined as follows:

$$\text{Minimize } E(t) = \sum_{t=1}^{T} \sum_{h=1}^{n} E^{h}_{supply,t} - E^{h}_{demand,t} \tag{7}$$

Constraints:

$$E^{t}_{supply} = supply\ list[t] \tag{8}$$

$$E^{t}_{supply} \leq \sum_{h=1}^{n} E^{h}_{demand,t} \tag{9}$$

$$E^{t}_{demand} = \sum_{h=1}^{n} E^{h}_{supply,t} \geq 0 \tag{10}$$

Here, E(t) is the power of external grid and it contains grid supply to the house agent and power received from these agents. Each house agents have balancing objectives and meanwhile they have to fulfill the constraints [(1), (2), (3)] at each time step. Supply list is the fixed supply power from external grid, in this work we assume that there is no supply in the day time and in the morning. Since there is not sufficient power supply, all agent cannot take the same action to get powered from external grid.

III. RESULT AND ANALYSIS

In this work, we used a microgrid to consider residential consumers equipped with PV power generation and energy storage modules. Each house has its own demand profile, which they have to fulfill by using storage, PV production, or external grid supply. The external grid supply is fixed and each agent demand differs from a time step. So, each time step has a balance constraint, which is a global objective for the microgrid agent.

In Fig.5, the total PV and demand as well as each house's individual PV production and demand are plotted respectively. External power represents the power supply by

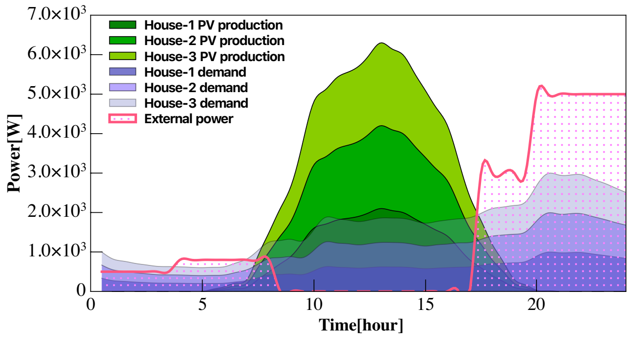


Fig.5. PV production, demand, external supply for multi-agent

the external power grid. We used a single PV and demand data profile as baseline data and the training sample were prepared by adding noise to the baseline data to create random data at every episode of learning time. Introducing the diversified trained data increases the robustness and generalization ability during the agent learning process thus improve the prediction accuracy. Battery manipulation and scheduling are the main objective during learning, based on the instant information from the environment agents and determine the action policy which favors to balance the electricity demand. Each agent has the same battery capacity, so they can discharge and charge under the same power

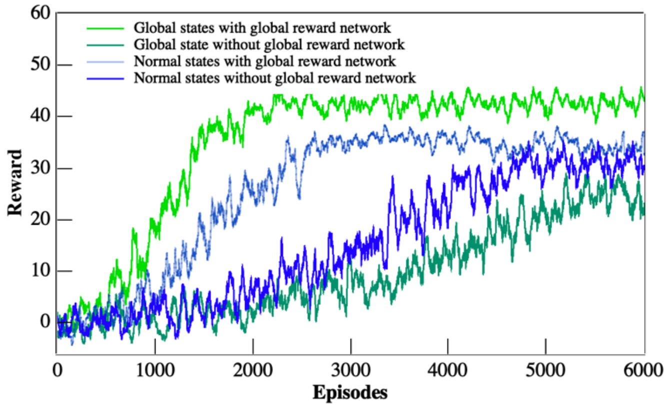


Fig.6. Effect of Global state and global reward on the multi-agent learning process

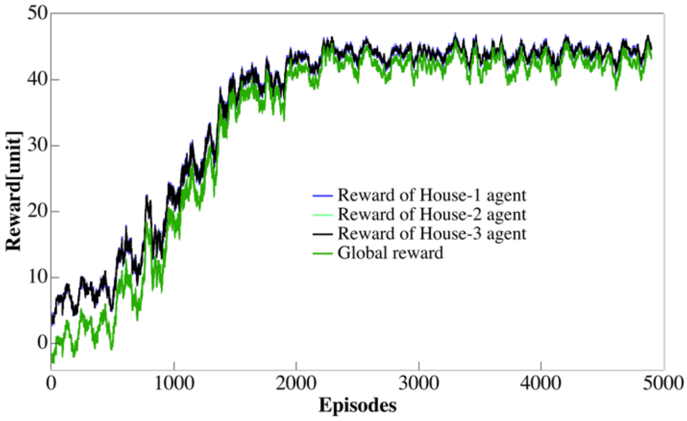


Fig.7. Global and individual agent rewards for all house agents

quantity if the agent selects the same action.

Using a global state and rewards, house agents can select the best action at each time step. We have added all the remaining demand as a global state and the other is the power balance after the previous agent utilize power an external grid. This globally combined information can provide the state of information which can not be covered by the battery production, demand, and battery SOC information in a MAS environment. We used two types of rewards in updating agents, which is very fruitful for the multi-purpose acquisition process. We invested the impact of a global reward and state on the learning process in Fig.6. The learning process with global state and reward is colored green and locates at the top of the reward plot profile. From this learning results, we can see that collective information is not an dispensible factor for an effective multi-agent learning process. Reward profile for three individual agents in MAS learning environment is also plotted in Fig.7, where agents reward is quite similar because we terminate the learning process if one of the three agents selects one bad action.

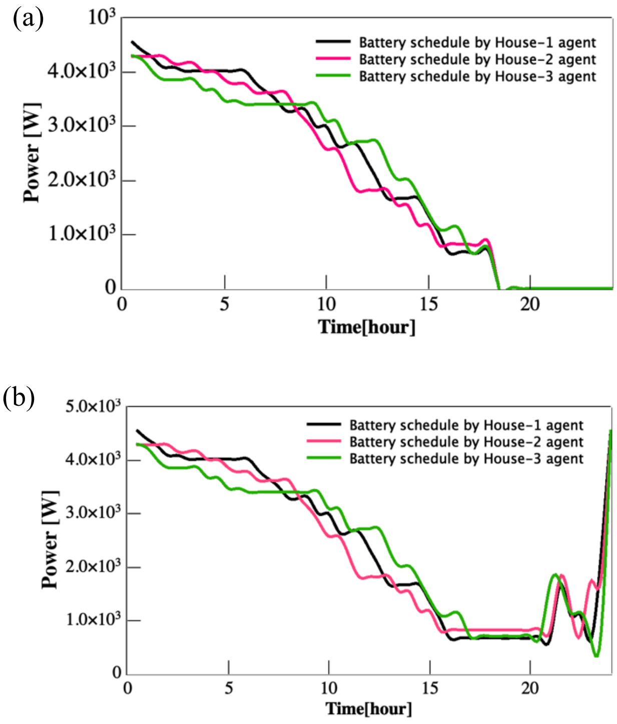


Fig.8. (a) Battery schedule at unbalance learning (initial learning period); (b) Battery schedule at balanced learning (final learning period)

After increasing the learning episodes, all agents can schedule the battery in optimal way. They can discharge the battery and use external power for the demand balance process. They properly use an external power supply at the early hours and manage their PV production at day time for demand fulfillment and sell purpose. The battery charge and discharge schedule is plotted in Fig.8. In Fig.8(a) we show that agent learnt how to discharge but failed to charge the battery at the final stage of learning process. In contrast, Fig.8(b) shows that the agent learns to charge the battery near the terminal period in order to fulfill the constraint requiring the SOC gets recovered to its initial value of 4.65KW. PV production is a source of electricity for a house. Managing PV power increases a house's income and also helps to decrease the electricity peak at day time. In this work, our aim is to increase the PV sale, which helps to decrease the peak power problem. After learning, house agents manages to sell PV when the demand is at maximization level, which is plotted in Fig.10. In this plot it shows the PV used at quite lower portion for house demand and the middle shows the total electricity powered from battery and the upper parts are the PV power sold by each agent.

IV. CONCLUSION

In this paper, we proposed the use of multi-agent

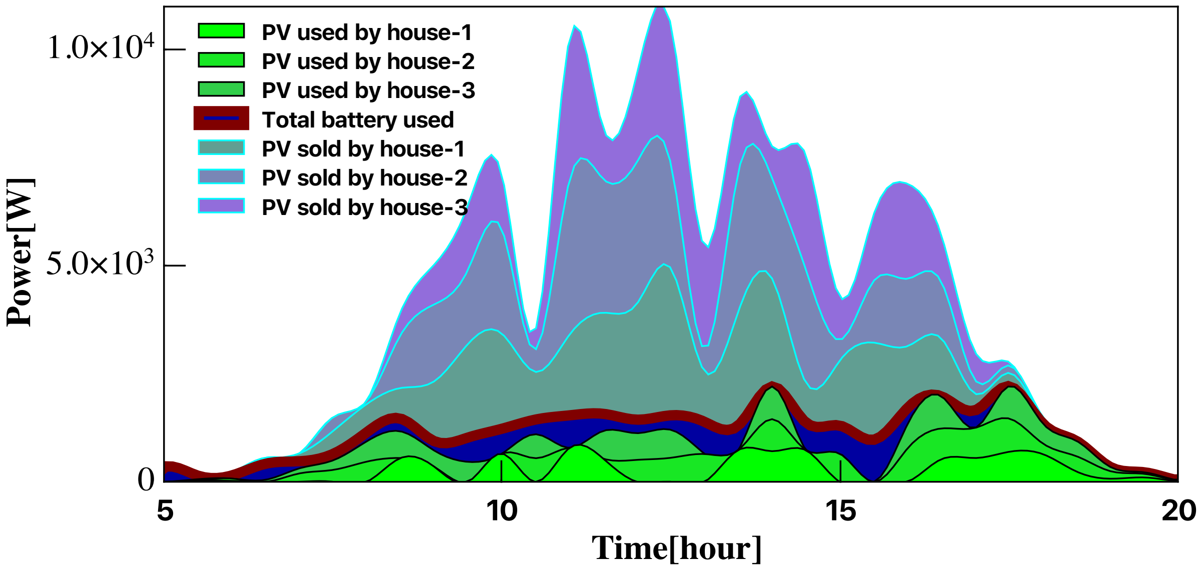


Fig.9. PV used for house and sold to an external grid

reinforcement learning, to minimize power consumption from the power grid by the optimally scheduling electricity storage devices in residential buildings and aggregations of buildings. A multi-agent, empowered with a global state and reward, is able to solve various balancing tasks. Deep Q-learning was adopted to solve the discrete action policy decision problems at both the building level and the aggregate level. Also, the advantages of multi-agent were analyzed in solving complex tasks in comparison with a in a single agent based normal state and reward reinforcement learning methods. In the further investigation, we will introduce the price of electricity into the learning process in order to minimize the energy cost. The profile gained by optimal pricing strategy will finally incentivize customers to shift their consumption behavior to lower price, off-peak periods, which is a vital for the realization of large scale virtual power plant.

ACKKNOWLEDGEMENT

This work is supported by New Energy and Industrial Technology Development Organization (NEDO), and Ministry of Economy, Trade and Industry (METI), Japan.

REFERENCES

[1] UNEP. Buildings and climate change, summary for decision-makers; 2009.

[2] St´ephane Caron, George Kesidis, "Incentive-based energy consumption scheduling algorithms for the smart grid," in *Proc. Smart Grid Commun. Conf.*, 2010.

[3] Chenxiao Guan, Yanzhi Wang, Xue Lin, Shahin Nazarian, and Massoud Pedram, "Reinforcement learning-based control of residential energy storage systems for electric bill minimization," In Consumer Communications and Networking Conference (CCNC), 2015 12th Annual IEEE, pages 637–642. IEEE, 2015.

[4] Hurtado LA, Mocanu E, Nguyen PH, Gibescu M, Kamphuis IG, "Enabling co-operative behavior for building demand response based on extended joint action learning," 1 1 IEEE Trans Ind Informatics 2018;3203.

[5] Lucian Busoniu, Robert Babuska, and Bart De Schutter, ''A comprehensive survey of multiagent reinforcement learning,'' *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 2, pp. 156–172, Mar. 2008.

[6] Yujing Hu, Yang Gao, and Bo An, ''Accelerating multiagent reinforcement learning by equilibrium transfer,'' *IEEE Trans. Cybern.*, vol. 45, no. 7, pp. 1289–1302, Jul. 2015.

[7] José del R. Millán, Daniele Posenato, Eric Dedieu, ''Continuous action Q-learning,'' *Mach. Learn.*, vol. 49, nos. 2–3, pp. 247–265, 2002.

[8] Jakob N. Foerster, Yannis M. Assael, Nando de Freitas, and Shimon Whiteson, "Learning to communicate with deep multi-agent reinforcement learning," CoRR, abs/1605.06676, 2016.

[9] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, "Human-level control through deep reinforcement learning," Nature, 518(7540):529–533, 2015.

[10] Ming Tan, "Multi-agent reinforcement learning: Independent vs. cooperative agents," In Proceedings of the tenth international conference on machine learning, pages 330–337, 1993.

[11] Elena Mocanu, Decebal Constantin Mocanu, Phuong H. Nguyen, Antonio Liotta, Madeleine Gibescu,"On-Line Building Energy Optimization Using Deep Reinforcement Learning," IEEE transactions on smart grid, vol. 10, no. 4, july 2019

[12] Hussain Kazmi, Johan Suykens, Attila Balint, Johan Driesen, "Multi-agent reinforcement learning for modeling and control of T thermostatically controlled loads," Applied Energy 238 (2019) 1022–1035

[13] Shuyue Hu, Ho-fung Leung, "Achieving Coordination in Multi-Agent Systems by Stable Local Conventions under Community Networks,"Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)